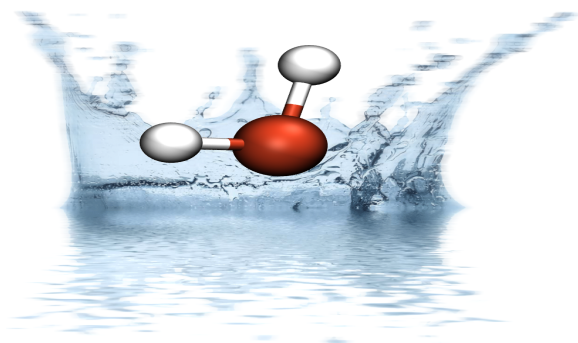


Universidad Andrés Bello
Facultad de Ciencias Biológicas
Center for Bionformatics & Integrative Biology
Chile

Simulaciones Moleculares de Sistemas Líquidos

Uso de VMD y NAMD para análisis de
propiedades estructurales utilizando
dinámica molecular

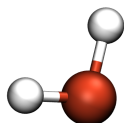


Daniel Aguayo V.

Introducción al Práctico

Este práctico introduce al estudiante al uso del programa VMD para analizar simulaciones de dinámica molecular. Comprenda y realice todas las instrucciones, conteste las distintas preguntas, y entregue un informe escrito con los resultados y discusión de los temas abordados.

A lo largo del texto se entregará material complementario en forma de “cuadros” independientes, los que incluyen información y preguntas de distintos aspectos como importancia biológica, parámetros a profundizar y atajos para VMD.



Agua. En este práctico nos enfocaremos en el análisis de una caja de agua utilizando el programa VMD. Se analizarán simulaciones de dinámica molecular a distintas temperaturas, visualizando el cambio en la estructura molecular del agua utilizando la función de distribución radial de las moléculas del sistema.

0.1. Programas Requeridos

Los siguientes programas serán requeridos en este práctico

- **VMD:** Disponible en <http://www.ks.uiuc.edu/Research/vmd/> (para todas las plataformas)
- **NAMD:** Disponible en <http://www.ks.uiuc.edu/Research/namd/> (para todas las plataformas)
- **Programa de graficación matemática:** Será necesario utilizar programas para gráficar las salidas de VMD y NAMD. VMD tiene un programa básico de graficación incorporado. Algunos ejemplos de otros programas son:
 - Unix/Linux: xmgrace, <http://plasma-gate.weizmann.ac.il/Grace/>
 - Windows: Excel, <http://office.microsoft.com/en-us/FX010858001033.aspx> (Pago)
 - Mac/Multiples Plataformas: Mathematica, <http://www.wolfram.com/> (Pago); gnuplot, <http://www.gnuplot.info/> (Descarga gratuita)

Los archivos de este práctico se encuentran en el directorio **practico-vmd**.

Introducción

1. La molécula de Agua

El agua es indispensable para la vida en la tierra, ya que es el medio (solvente) en el que ocurren la gran mayoría de los procesos biológicos. Unas de las propiedades fundamentales del agua es la de absorber largas cantidades de calor (tiene una alta capacidad calorífica), lo que le permite mantener controlada la temperatura de distintos procesos que permiten la homeostasis. Entre otras propiedades del agua se encuentran la de actuar como agente ionizante de distintas especies químicas, favoreciendo la formación de gradientes electrostáticos, o como medio para el transporte de distintas especies químicas, permitiendo el transporte de moléculas a los distintos sitios de interés. La rápida reorientación del moléculas de agua, en el orden de femtosegundos, impiden la directa

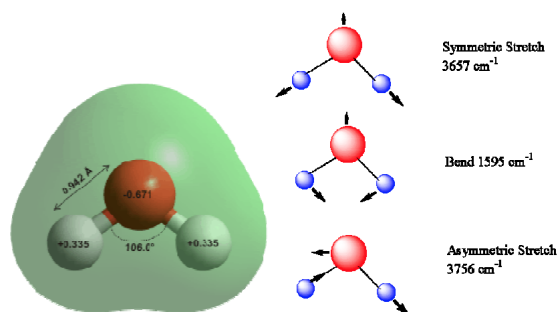
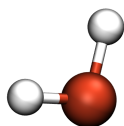
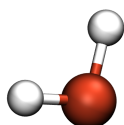


Figura 1: Propiedades estructurales, superficie y modos vibracionales del agua

observación de la estructura en fase líquida del agua. Los estudios experimentales y teóricos se centran, por lo general, en propiedades macroscópicas descritas principalmente por modelos continuos, sin embargo, la estructura de la fase líquida del agua esta dominada por la formación de los enlaces de hidrógeno entre las distintas moléculas. La importancia del agua en la función biológica y procesos químicos explican el gran interés y esfuerzo realizado para lograr simular sus propiedades a escalas microscópicas.



Punto de ebullición y congelamiento. ¿Cómo la estructura y capacidad de formar puentes de hidrógeno del agua permiten explicar sus puntos congelación y de ebullición?



Agregados de moléculas de agua. ¿Cómo se pueden distinguir "agregados" de aguas de las moléculas que se encuentran adyacentes y no forman parte de un agregado?. Un agregado es una estructura transiente que no perdura en el tiempo. La posición de una molécula de agua cerca de otra esta gobernada por consideraciones energéticas y geométricas. La formación de enlaces de hidrógenos de manera ordenada da origen a estructuras de pequeño tamaño, sin embargo, el tiempo en que estas permanecen es aun una interrogante que puede ser resulta a través de métodos de simulación molecular.

El agua es una de las pocas sustancias que en su estado sólido tiene una menor densidad que en estado líquido. La figura 2 muestra como su volumen varía con respecto a la temperatura, con un aumento del 9 % cuando desciende de los 0 °C y 16.8 % bajo los -20°C. Cuando la temperatura se encuentra entre -4 y < 0°C la densidad aumenta debido a la formación de agregados de moléculas por la formación de redes de enlaces de hidrogeno. Sobre 4 °C, los enlaces O-H vibran debido al aumento de la temperatura, aumentando el movimiento y choque de las moléculas, separandolas y disminuyendo la densidad.

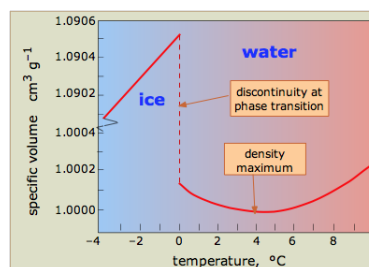


Figura 2: Estado de fase del agua con respecto a la temperatura.

Introducción a VMD

2. Uso Básico de VMD

2.1. Inicio

En esta unidad utilizaremos VMD para crear y visualizar una caja de agua, la cuál servirá en el resto de este practico. Solo se abordarán aspectos básicos de uso, por lo que se recomienda leer el manual de usuario disponible en <http://www.ks.uiuc.edu/Research/vmd/current/ug/>

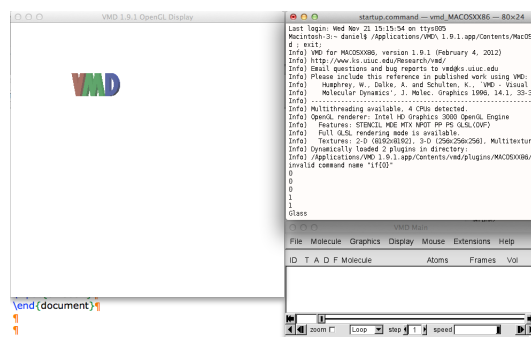


Figura 3: Ventanas abiertas al ejecutar VMD.

- 1 Para iniciar VMD escriba `vmd` en una ventana de terminal a Unix, doble-click en el icono de la aplicación de VMD en la carpeta **Applications** en Mac OS X, o haga click en el menu Inicio → Programas → VMD en Windows. En Unix escriba `vmd` en una ventana de terminal.

2.2. Cargando el sistema

Para iniciar se deben cargar las coordenadas de los distintos átomos del sistema, las cuales se pueden obtener desde distintas fuentes como el *protein data bank* en el caso de proteínas o directamente con VMD. El archivo `BOX.pdb` contiene las coordenadas atómicas de un sistema compuesto por moléculas de agua.

- 1 Elija el menú **File** → **New Molecule...** Fig. 4 en la ventana **VMD Main**. Aparecerá otra ventana **Molecule File Browser**, la que permite elegir los archivos dentro de los directorios disponibles.
- 2 Use el botón **Browse...** para localizar el archivo de coordenadas en formato PDB de nombre **BOX.pdb** en **vmd-tutorial-files** en el directorio **practico-difusion**. Note que luego de seleccionar este archivo se encuentra nuevamente en la ventana **Molecule File Browser**. Para cargar el archivo, presione el botón **Load**. Recuerde revisar que esta seleccionado **New Molecule** en la sección **Load Files for**. Si lo desea, puede cerrar la ventana **Molecule File Browser**.

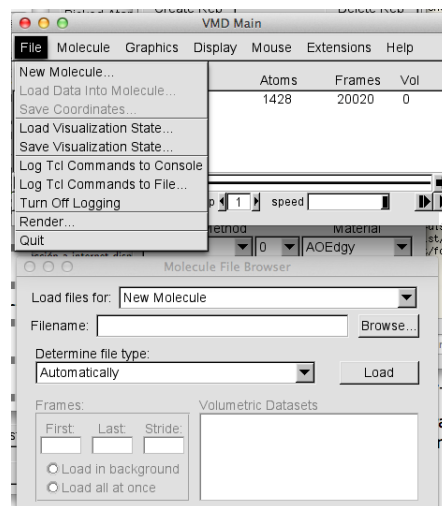


Figura 4: Cargando una molécula.

Una caja de agua se visualiza en la ventana **OpenGL Display**. Utilice el ratón para rotar el sistema, para escalar la representación utilice el tercer botón del ratón o haga click la tecla “S” en el teclado y luego mueva el cursor en la ventana **OpenGL Display**. Para volver a rotar haga click en la tecla “R” en el teclado. Para trasladar el sistema, utilice la tecla “T”.

- 3 Para volver a la visualización inicial presione **Display** → **Reset View**.
- 4 En el menú **Menu** → **Graphics** elija **Representations**. Se abrirá la ventana **Graphical Representations**, en amarillo se destaca la representación que se está utilizando para mostrar el sistema (fig. 5).

Selección	Acción
water	muestra las moléculas de agua si están presente.
name OH2	muestra el átomo de oxígeno de las moléculas de agua.
protein	muestra una proteína si esta presente.
resid 1	el primer residuo
x >0	átomos que tienen la coordenada x positiva
same residue as x>0	átomos de residuos que tienen un átomo cuya coordenada x es positiva
name OH2 and x>0	átomos de nombre OH2 y coordenada x es positiva
within 5 of resid 1	átomos a 5 Angstrom del residuo 1

Cuadro 1: Ejemplo de selecciones

Con el ratón se pueden crear y seleccionar diferentes representaciones. Además, estas pueden activarse o desactivarse haciendo doble-click sobre ellas, borrarse utilizando el botón **Delete Rep button**, etc. Diferentes combinaciones de operadores pueden utilizarse para crear selecciones de mayor complejidad. El cuadro 1 muestra algunas selecciones posibles.

- 5 La entrada de texto **Selected Atoms** de la ventana **Graphical Representations** permite modificar los átomos que se están visualizando. Reemplace la palabra *all* por *water* y presione **Apply** (o presione **Enter** en el teclado), lo cual debe realizar cada vez que modifica un campo en la ventana **Graphical Representations**.
- 6 En la ventana **Graphical Representations** elija la viñeta **Selections**. En la sección **Single-words** se encuentran los diferentes campos por los cuales se pueden realizar selecciones. En este ejemplo, escriba *water* en el campo de texto para seleccionar las moléculas de aguas automáticamente.

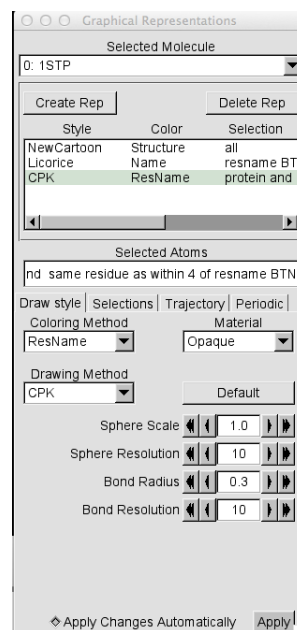
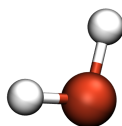


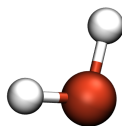
Figura 5: Ventana Graphical Representations



Selecciones geométricas complejas. Distintos tipos de geométricas se pueden lograr utilizando selecciones complejas en relación al espacio cartesiano. Cuál es la selección apropiada para obtener una esfera a partir de las moléculas presentes. Recuerde que una molécula no puede quedar incompleta!

- 7 En la ventana **Graphical Representations** elija la viñeta **Draw Style**, esta permite cambiar el estilo de visualización de cada representación creada. El estilo predeterminado es *Lines*, que muestra en forma de líneas los enlaces entre átomos. Cada método de dibujo tiene sus propios parámetros. Por ejemplo, modifique el campo **Thickness** para aumentar el ancho de las líneas que representan los enlaces atómicos.
- 8 Cambie el método de dibujo a **VDW** (van der Waals).
- 9 Utilice el ratón para orientar la visualización. Para volver a la visualización inicial presione **Display** → **Reset View**.

Cada átomo es ahora representado como una esfera. En el menú **Coloring Method** es posible elegir el color de la representación. Diferentes estilos se encuentran disponibles: Nombre, tipo de átomo, nombre del residuo, estructura, etc.



Otras métodos de dibujo. CPK y Licorice son métodos de dibujo que permiten observar estructuras con facilidad. La primera emula los sistemas de bolas de radio variable y tubos utilizados en química para visualizar moléculas, mientras que la segunda representa los átomos como esferas de radio fijo que no puede ser modificado.

- 10 Cree una representación para moléculas de agua que tengan un campo *beta* > 0 y otra para aquellas que tienen un valor de *beta* < 0, con colores azul y verde respectivamente. Seleccione solo los átomos de nombre "OH2", representandolas con esferas de van der Waals. El resultado debe asemejar a la figura 6

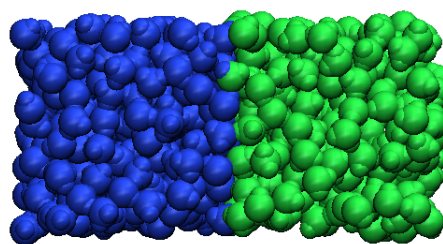


Figura 6: Representación Final

- 11 Antes de continuar explore distintas selecciones, representaciones



Operadores lógicos. Habitualmente utilizamos los operadores booleanos "Y", "O" y "NO". El primer es un operador "incluyente", el segundo de "exclusión", mientras que el tercer operador es considerado de "reunión" o suma lógica. Podemos ver la diferencia entre ellos observando la cantidad de átomos de selecciones que las incluyan como: `'name OH2 or x>0'`], `'name OH2 and x>0'`] y `'name OH2 and not x>0'`].

2.3. Guardando su trabajo

La imagen creada puede ser almacenadas en archivos utilizando VMD. Además, las representaciones creadas pueden ser almacenadas utilizando un estado de VMD. El archivo de estado contienen la información necesaria para reiniciar las sesiones de VMD, sin perder el trabajo realizado.

- 1 En la ventana VMD Main, elija el menú File → Save State. Escriba un nombre apropiado para el archivo (i.e, `aguas.vmd`) y guárdelo en el directorio de trabajo.
- 2 cierre y reinicie VMD para cargar el estado guardado

El menú File → Load State permite cargar el estado guardado previamente. Para guardar en imágenes las representaciones realizadas, VMD puede generar un archivo imagen que puede ser utilizado en otros programas.

- 3 Usando lo aprendido, genere una imagen de calidad, que muestre alguna característica de interés del sistema (por ejemplo la interacción entre dos moléculas de agua). Ponga cuidado en la resolución de cada representación.
- 4 Cambie el color del fondo seleccionando el menú Graphics → Colors. Elija la categoría Display → Background → 8 white. El fondo debe estar en blanco ahora.
- 5 Elija el menú File → Render.... Se abrirá una ventana llamada File Render Controls.
- 6 Puede utilizar distintos métodos para generar la imagen, seleccione Snapshot o Tachyon.



Troubleshooting. Si el directorio donde guardara los archivos contienen espacios, es necesario que antes de utilizar File Render Controls, seleccionar el directorio de trabajo utilizando la Console Tk. Para esto seleccione el menú Extensions → Tk Console. Use cremillas para navegar a la dirección, i.e., `cd "C:\ Documents and Settings"`.

- 7 Escriba el nombre del archivo a generar en el campo de texto Filename , i.e., `imagen.bmp`.

- 8 Presione el botón **Start Rendering** y se generará el archivo con la imagen que aparece en la ventana **OpenGL**. Esto puede tardar algún tiempo dependiendo de las capacidades del computador utilizado. Al finalizar, el archivo se encuentra en el directorio seleccionado o en el directorio de trabajo.
- 9 Antes de comenzar con la siguiente sección explore VMD libremente. Al terminar, reinicie VMD.

2.4. Trabajando con trayectorias



Archivos de Coordenadas y de Conectividad. Los programas de visualización como VMD, infieren los distintos enlaces a partir de las distancias entre átomos, por lo que habitualmente se pueden encontrar átomos con una mayor cantidad de enlaces que los permitidos, ie. carbonos con 5 átomos asociados, moléculas de agua con 4 o más hidrógenos. Para solucionar esto se hace uso de un archivo extra que contiene la conectividad de cada átomo, además de otros parámetros necesarios para realizar una simulación, por ejemplo las cargas atómicas.

- 1 En la ventana **VMD Main** elija el menú **File** → **New Molecule**
- 2 Use el botón **Browse...** para localizar en el directorio de trabajo el archivo de conectividad **BOX.psf**. Luego de seleccionar este archivo y cargarlo, en la ventana **Main** se encuentra la información de nuestro sistema, sin embargo, no aparecen átomos en la ventana de visualización.

Note que a diferencia de la sección 2, se ha cargado primero un archivo de extensión **psf**, el cual contiene la conectividad del sistema

- 3 Cargaremos ahora las coordenadas atómicas. Primero debemos asegurarnos que en la ventana **Molecule File Browser**, en la sección **Load Files for** se encuentre seleccionado **0:BOX.psf**, luego, use el botón **Browse...** para localizar el archivo **BOX.pdb** en el directorio de trabajo. Para cargar el archivo, presione el botón **Load**.

Ahora, se visualiza una caja de agua en la ventana **OpenGL Display**. En el directorio **practico-difusion/ejemplo** encontrará la trayectoria de una simulación de dinámica molecular realizada con el programa **NAMD**. En esta sección utilizaremos esta trayectoria en VMD para realizar distintos análisis.

- 4 Elija el menú **File** → **New Molecule...** en la ventana **VMD Main**. Aparecerá otra ventana **Molecule File Browser**, la que permite elegir los archivos dentro de los directorios disponibles.
- 5 En la ventana **Molecule File Browser**, en la sección **Load Files for** asegúrese que se encuentre seleccionado **0:BOX.psf**. Luego, use el botón **Browse...** para localizar el archivo **BOX.dcd** en el directorio de trabajo. Para reducir la cantidad de configuraciones cargadas en el campo **STRIDE** coloque 1000. Para cargar el archivo, presione el botón **Load**.
- 6 cree una representación para aquellas moléculas de agua que tengan un campo *beta* > 0 y otra para aquellas que tienen un valor de *beta* < 0, coloreandolas de color azul y verde cada una. Seleccione solo los átomos de nombre **OH2** en un método de dibujo de tipo **VDW**. El resultado debe asemejar la figura 6 para el primer cuadro de la simulación
- 8 Utilice los controles de la ventana **Main** para recorrer la trayectoria, observe como la posición de la molécula de agua varía. ¿Qué información podemos obtener de ella?.

- 7 En la ventana Main se puede observar la cantidad de átomos y configuraciones (*frames*) que se han cargado para esta molécula (Fig 7). Además, se pueden utilizar los controles de reproducción (flechas hacia adelante, hacia atrás,...,etc. para controlar la configuración que se esta visualizando.

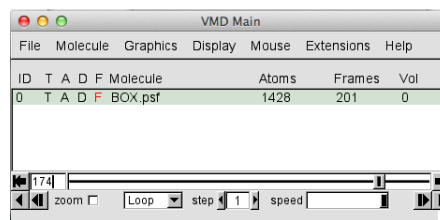


Figura 7: Información de la molécula y trayectoria cargada

- 9 Para reconocer una molécula específica es posible utilizar el ratón y la ventana OpenGL para conocer su identificador. Para ello en la ventana VMD Main, elija el menú Mouse → Query y haga click sobre un átomo de interés. Al hacerlo, aparecerá en pantalla información básica del átomo seleccionado. Para obtener mayor información en la ventana VMD Main elija el menú Graphics → Label, que abrirá una ventana con la descripción del átomo seleccionado.

- 10 Cree una representación que destaque la molécula de agua a la que pertenece el átomo de interés. Por ejemplo, representación de VDW, coloreada por posición en la coordenada x (recuerde seleccionar actualizar el color en la viñeta *trajectory*). El resultado debe asemejar la figura 9.

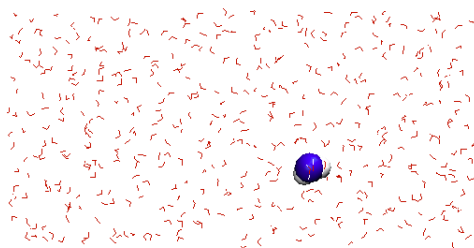
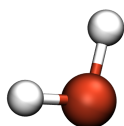


Figura 8: Representación final

- 11 Seleccionando en la ventana VMD Main el menú Mouse → Labels → Bond es posible conocer la distancia entre dos átomos de interés. Al hacerlo, aparecerá en la ventana Label bajo la etiqueta Bonds → value la distancia entre los átomos seleccionados en la configuración actual. La opción *preview* de la viñeta Graph permite observar la distancia seleccionada a lo largo de la trayectoria.

2.4.1. Red de enlaces o puentes de hidrógeno



Enlaces de Hidrógeno. La diferencia de electronegatividad entre los átomos de hidrógeno y oxígeno permite que las moléculas de agua puedan compartir H entre ellas, formando lo que se denomina puente o enlace de hidrógeno, ¿Cuales son los parámetros necesarios para que se formen? Tiene relación la formación de puentes de hidrógeno con las propiedades del agua?

- 12 Utilizando el método de dibujo Hbond en la ventana de representaciones genere una visualización para observar los puentes de hidrógeno. Tip: utilice una representación Licorice para las moléculas de agua y otra para los enlaces de hidrógeno.

- 13** Para reconocer una molécula específica es posible utilizar el ratón y la ventana OpenGL para conocer su identificador. Para ello en la ventana VMD Main, elija el menú **Mouse** → **Query** y haga click sobre un átomo de interés. Al hacerlo, aparecerá en pantalla información básica del átomo seleccionado. Para obtener mayor información en la ventana VMD Main elija el menú **Graphics** → **Label**, que abrirá una ventana con la descripción del átomo seleccionado.



Actualización automática de una selección. Por defecto las selecciones realizadas con VMD no se actualizan para cada frame. Para que se actualicen, se deben activar las casillas **Update selection every frame** o **Update color every frame** en la viñeta **trajectory** de la ventana **Graphical Representation**. Pruebe esto realizando una selección en el cuadro inicial para los átomos que tengan coordenada x positiva.

- 14** Utilizando las selecciones descritas en el cuadro 1 genere una representación que destaque la molécula de agua a la que pertenece el átomo de interés. Por ejemplo, representación de VDW, coloreada por posición en la coordenada x (recuerde seleccionar actualizar el color en la viñeta **trajectory**). El resultado debe asemejar la figura 9.

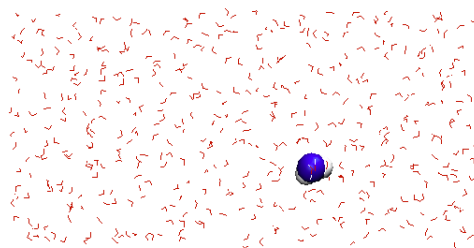
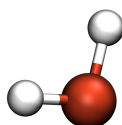


Figura 9: Representación final

Seleccionando en la ventana VMD Main el menú **Mouse** → **Labels** → **Bond** es posible conocer la distancia entre dos átomos de interés. Al hacerlo, aparecerá en la ventana **Label** bajo la etiqueta **Bonds** → **value** la distancia entre los átomos seleccionados en la configuración actual. La opción **preview** de la viñeta **Graph** permite observar la distancia seleccionada a lo largo de la trayectoria.



Configuración molecular del agua. ¿Que sucede con las moléculas de aguas que se encuentran adyacentes a la molécula seleccionada? ¿A que distancia se encuentran las moléculas de agua que forman enlaces de hidrógeno? ¿Adoptan alguna configuración tridimensional entre ellas?

Efecto de la temperatura en la estructura del Agua

Este práctico buscamos observar como se modifica la estructura del agua bajo distintas condiciones de simulación. El cuerpo humano mantiene una temperatura cercana a 36.5 °C, donde la capacidad calorífica del agua es fundamental para mantener la homeostasis. A 100 °C el agua se encuentra cerca de su punto de ebullición, mientras que a 0 °C hablamos del punto de congelamiento. ¿Que significa esto desde el punto de vista molecular? ¿Cómo se ve afectada la estructura del agua cuando cambia la temperatura? ¿Que efecto puede tener sobre la actividad biológicas?

Para responder estás preguntas estudiaremos trayectorias de simulaciones de dinámica molecular bajo distintas condiciones de temperatura (ver cuadro 2).

Modelo de Agua	Temperatura	Archivo
TIP3P	300 K	300.dcd
TIP3P	10 K	10.dcd

Cuadro 2: Trayectorias suministradas en este práctico

3. Análisis de la estructura molecular del agua

3.1. Función de distribución radial

Para caracterizar la forma en que interactúan las partículas, habitualmente, se utiliza la función de distribución de pares o función de distribución radial “ $g(r)$ ”, definida formalmente como “*la probabilidad de encontrar un par atómico a una distancia r , normalizado por la probabilidad esperada para una distribución aleatoria en la misma densidad*”.

La figura 10 muestra como se calcula $g(r)$, en ella se observa que a partir de una partícula cualquiera, los partículas adyacentes se ordenan en distinta densidad (partículas por área o volumen) a medida que nos alejamos. De esta forma, en sistemas que se encuentran altamente estructurados presentará una función de distribución radial diferente a uno en que el sistema se encuentra en un mayor desorden.

Durante este práctico utilizaremos la función de distribución radial y otros observables para ca-

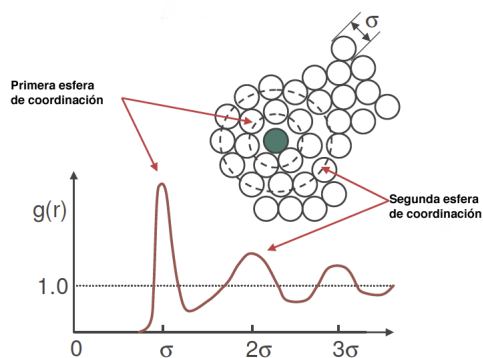
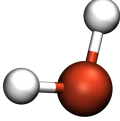


Figura 10: Función de distribución radial.

racterizar diferentes sistemas moleculares, los cuales son representativos de sistemas de menor y mayor complejidad.



función de distribución radial. En un ensamble canónico la función de distribución de pares esta dada por

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{\rho^2 Z_{NVT}} \int d\mathbf{r}_3 d\mathbf{r}_4 \dots \exp(-\beta \mathcal{V}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)) \quad (1)$$

Siendo N el número total de átomos, ρ la densidad del sistema, Z la función de partición, $\beta = 1/kT$ y \mathcal{V} es la energía potencial que gobierna el movimiento los átomos. De modo conveniente, es posible escribirla como

$$g(\mathbf{r}) = \frac{V}{N^2} \langle \sum_i \sum_j \delta(\mathbf{r} - \mathbf{r}_{i,j}) \rangle \quad (2)$$

La ecuación anterior corresponde a la función de distribución de pares a una distancia lineal, luego al incorporar una elemento de volumen esférico (radial en \mathcal{R}^3) puede escribirse de manera práctica como

$$N(\mathbf{r}) = \rho g(\mathbf{r}) 4\pi r^2 dr \quad (3)$$

Donde $N(\mathbf{r})$ es la cantidad de átomos a una distancia acotada por $[\mathbf{r}, \mathbf{r} + d\mathbf{r}]$ para un átomo dado.

3.1.1. Caracterización de la estructura del Agua

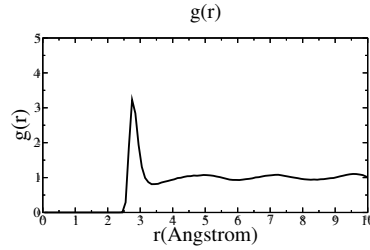
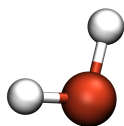


Figura 11: Función de distribución radial del agua, calculada utilizando los átomos de oxígeno (modelo TIP3P) utilizando VMD y NAMD.

La figura 11 muestra la función de distribución radial oxígeno-oxígeno del agua a 300 K. El valor de $g(r)$ es 0 en el intervalo $0 \leq 2,5 \text{ \AA}$, luego existe un pico cercano a $r = 3,8 \text{ \AA}$. ¿Cuál es el significado de esto? En palabras simples, si seleccionamos un átomo de oxígeno cualquiera, la cantidad de oxígenos a una distancia $r = 3,8 \text{ \AA}$ es cercana a tres veces la cantidad de átomos que encontraríamos en una distribución aleatoria de los átomos bajo la misma densidad ($r = 10 \text{ \AA}$)

A ¿A que corresponde la distancia para la cual se observa el primer pico de la función $g(r)$ de las simulación del del agua a 300 K?



PMF. PMF es la energía libre a lo largo de uno o más grados de libertad del sistema. El nombre proviene de la forma habitual en que se calcula el PMF, donde se integra la media de la fuerza que ejerce trabajo sobre un cierto grado de libertad, también llamada coordenada de reacción. También es posible obtener el PMF utilizando la relación inversa de Boltzmann y el valor de $g(r)$

$$\text{PMF} = -k_B T \log(g(r)) + C \quad (4)$$

En la práctica es realizar un gráfico $y = -0,75 \ln x + 0,0001$ en nuestra gráfica de $g(r)$. Puede a partir del valor del PMF calcular el valor de la energía de hidratación del agua en agua de $-6,32$ kcal/mol

B ¿Esperaría ud que la distancia r a la cual aparece el primer pico de la función $g(r)$ varíe si se simula el agua a 200 y 600 K?

La figura 12 muestra como se ordenan las moléculas de agua a 10 y 300 K. En ella es posible observar que a medida que nos alejamos de una molécula cualquiera, las moléculas adyacentes se distribuyen de manera distinta dependiendo de la temperatura, a 300 K hay una mayor cantidad de átomos a distancias cercanas a la molécula central, mientras que a 10 K la cantidad de átomos adyacentes es menor.

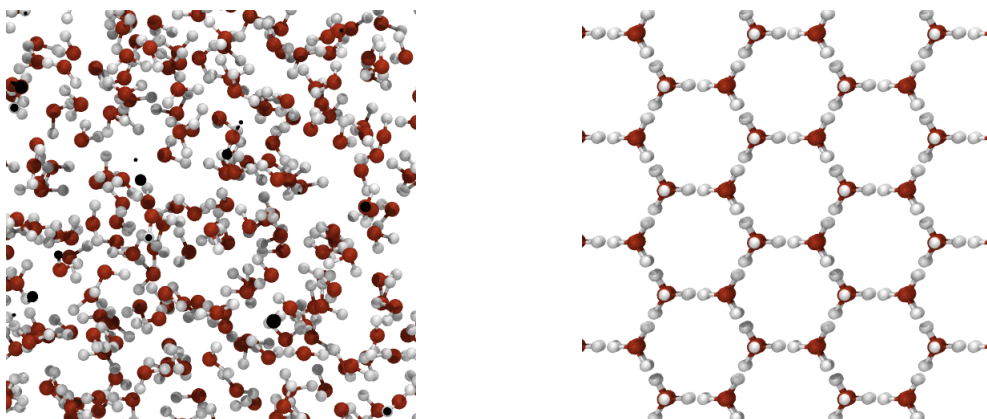
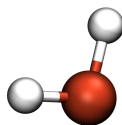


Figura 12: Estructura del agua a 300 y 10 K.

A continuación caracterizaremos la estructura del agua utilizando la función de distribución radial. Para este análisis utilizaremos las trayectorias de ejemplo de una caja de agua simuladas a 10 (`practico-vmd/files/10.dcd`) y 300 K (`practico-vmd/files/300.dcd`) y la herramienta de cálculo de función de distribución radial de VMD.

- 1 Cargue cada trayectoria según lo aprendido en la sección 2.4. Utilice el archivo de conectividad `agua.psf`.
- 2 Una vez cargada la trayectoria seleccione el menú `Extensions → Analysis → Radial Pair Distribution Function`. Al hacerlo se abrirá la ventana `RPDF g(r)`.
- 3 Complete los cuadros de texto `Selection 1` y `Selection 2` según los átomos sobre los que se calculará la función de distribución radial. En nuestro caso utilizaremos una selección de los átomos de oxígeno, cuyo nombre es `OH2`. Escriba en cada campo `name OH2`.

- 4 Utilizaremos condiciones periódicas de borde para este calculo, seleccione el campo **Use PBC**. Seleccione también la opción **Display $g(r)$** para generar un gráfico de $g(r)$. El valor del campo delta r corresponde al espaciado del elemento de volumen a utilizar (dr) y el campo max r a la distancia máxima desde el punto de referencia para el cálculo correspondiente.
- 5 Ejecute el cálculo seleccionando la opción **Compute $g(r)$** en la parte inferior de ventana. Luego de algunos minutos se abrirá una ventana que contiene el gráfico de la función de distribución radial (Fig. 11) para la trayectoria seleccionada.

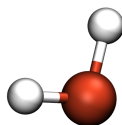


Análisis de $g(r)$. Como explicaría el primer, segundo y tercer pico presentes en el gráfico? ¿Que sucede si realizamos este estudio analizando solo los átomos de hidrógeno? Se ve modificada esta función cuando se varía la temperatura? ¿Por qué?

- C Determine y compare $g(r)$ para la simulación a 10 K y 300 K.
- D Explique la naturaleza del pico observado a 1.8 Å en la pdf del agua a 300 K.
- E ¿Que puede inferir del efecto de la temperatura sobre la estructura del agua a partir del análisis de $g(r)$ de las simulaciones realizadas?

3.1.2. Número de coordinación

Una manera simple de interpretar la función de distribución radial es considerar la posición r de los picos como la distancia de mayor probabilidad de encontrar átomos vecinos, mientras que el ancho da cuenta de la dispersión las distancias medidas ya que se calcula sobre todos los átomos del sistema. Otro dato importante a obtener de ella es el número de átomos que se ubican a una distancia determinada, obtenido mediante el cálculo del área bajo la curva de $g(r)$, lo cual permite calcular el número de partículas presentes en las llamadas esfera o capa de coordinación. (figuras 10 y 13).



Esfera o capa de solvatación. Consiste en las moléculas de solvente que se agrupan al rededor del soluto en un orde relativo. El ejemplo más clásico es la llamada esfera de hidratación de los iones, que esta formada por aguas que adoptan una geometria especifica o cluster que forman la interface entre el ion y el resto de las moléculas de solvente, en este ejemplo, el agua.

La función de distribución radial también puede ser calculada entre átomos de diferente tipo, llamándose función de distribución de pares o “pdf”. Por ejemplo la distribución del agua oxígeno-hidrógeno provee información relevante respecto a la formación de puentes de hidrógenos entre moléculas de agua (interacciones interMoleculares).

El área bajo la curva entre el primer pico y el primer mínimo de $g(r)$ entre los átomos de oxigeno e hidrógeno permite estimar la cantidad de puentes de hidrógeno que se generan con moléculas adyacentes. Por ejemplo la integración para el primer mínimo en el agua a 300 K es 2.

- 6 Utilice el modulo de $g(r)$ seleccionando la opción **Compute Int($g(r)$)** en la parte inferior de ventana. Luego de algunos minutos se abrirá una ventana que contiene el gráfico que contiene la integral de la función de distribución radial (Fig. 11) para la trayectoria seleccionada.

F Calcule la cantidad de moléculas de agua en la segunda esfera de solvatación para el agua a 300 K y 10 K.

G ¿A que distancia se pierde el orden el orden local? ¿Con cuantas moléculas de agua interactúa entonces una molécula de agua?.

La estructura tridimensional del agua se ilustra en la figura 13, la cual le permite un mejor entendimiento de como las moléculas se orientan unas a otras. En particular que las primeras 4 moléculas interactúan con el agua formando una estructura tetrahédrica.

H ¿Cree Ud. que este número de agua que rodea a las moléculas de agua en estado líquido varía con la temperatura, por ejemplo cercano a 400 K?.

I Discuta sobre el origen de la estructura adopta la primera y segunda esfera de solvatación del agua.



Figura 13: 1^a y 2^a Esfera de coordinación de una molécula de agua a 10 (*Izq.*) y 300 K (*Der.*).

J Con los datos obtenidos discuta las diferencias y consecuencias biológicas de la estructura molecular del agua a 10 y 300 K. Como material de apoyo puede referirse a la investigación de Bhattacharjee y cols [Biophys Chem. 2011 Sep;158(1):73-80. doi: 10.1016/j.bpc.2011.05.009]

Análisis avanzado de trayectorias con VMD

3.2. Programando la función de distribución radial

Anteriormente utilizamos un módulo de VMD para calcular la función de distribución radial y caracterizar la estructura del agua a distintas temperaturas.

K Ahora, utilice sus conocimientos de programación para generar un script para VMD que le permita calcular la función de distribución radial tridimensional del agua en una configuración estática. Como ayuda le entregamos las tareas que debe realizar para obtener el resultado.

1. Elegir un valor de δr .
2. Iterar sobre todos los valores r a los cuales ud. quiere caracterizar
3. Considerar cada partícula que corresponda. Cuente todas las partículas que están a una distancia r y $r + \delta r$ de la partícula de interés. Considere que estas partículas están en una capa esférica (como la cascara de cebolla) al rededor de la partícula de interés, esta capa tiene un grosor δr .

4. Divida la cuenta total en cada volumen por N , este último el número de partículas consideradas del sistema.
5. Divida este resultado por $4\pi r^2 \delta r$, que corresponde al volumen de la capa esférica. Esto considera que mientras r crece se encuentran un mayor número de partículas en cada volumen δr .
6. Divida el resultado por la densidad del sistema. Esto delimita un valor de $g(r) = 1$ para un sistema desestructurado. En otras palabras, si ud. coloca una capa esférica de radio r y un radio externo $r + \delta r$, ud. espera encontrar $\rho * V$ partículas en este volumen, donde ρ es la densidad y V es el volumen de la capa. En 2D este paso correspondería a dividir por $2\pi r \delta r$.

Como puede ver de las tareas, se requiere de un conocimiento específico del sistema y de los átomos que lo componen.

1 No desespere

Para su ayuda, en la siguiente sección hemos “resumido” los comandos más relevante de VMD que le permitirán obtener estos parámetros y poder generar el script solicitado.

L Escriba una lista de los parámetros y valores que necesita conocer para calcular $g(r)$.

M Identifique y/o caracterize los comandos que le permitirán obtener los parámetros y valores que necesarios para calcular $g(r)$.

3.2.1. Comandos básicos de programación con VMD

VMD permite realizar estudios avanzados a través del uso de las funciones internas disponibles y del lenguaje de programación Tcl/Tk. A continuación introduciremos al uso de la consola de programación y las funciones básicas de VMD.

Comando	descripción
ls	lista archivos y directorios
mkdir	crea un directorio
cd directorio	cambia de posición al directorio
cd	cambia el directorio <i>home</i>
cd ..	cambia a directorio padre
pwd	muestra la ruta del directorio actual

Cuadro 3: Comando básicos de Unix

- 2 Reinicie VMD y seleccione el menú **Extensions** → **Tk Console**. Al hacerlo se abrirá la ventana Tk Console. En ella se pueden utilizar los comandos básicos de Unix para manejo de archivos (ver cuadro 3), Tcl y propios de VMD

El cuadro 4 resume los comandos de mayor uso. A continuación mostraremos como pueden ser utilizados para realizar distintas tareas. Para ello, presentaremos cuadros donde las líneas que comienzas con “<Main>%” representan las instrucciones que Ud. debe escribir en la consola Tk y la línea siguiente (si están presentes) el resultado de su ejecución.

Una de las tareas habituales es asignar un valor a una variable para luego leerla en pantalla o utilizarla en una expresión matemática o generar una selección.

Comando	descripción
<code>set nombre valor</code>	asigna el <i>valor</i> a la variable <i>nombre</i>
<code>puts \$nombre</code>	muestra el valor de la variable <i>nombre</i>
<code>expr expresión</code>	evalúa la <i>expresión</i> matemática
<code>[comandos]</code>	Permite unir múltiples operaciones
<code>open</code>	permite abrir un archivo para lectura y/o escritura
<code>foreach</code>	recorre una lista elemento por elemento
<code>atomselect molid selección</code>	crea una nueva selección de átomos.
<code>measure propiedad selección</code>	permite realizar mediciones como distancia, posición, rmsd y otras.

Cuadro 4: Comando básicos de VMD

```

<Main>0%   set x 10
10
<Main>1%   puts "el valor de x es:$x"
el valor de x es 10
<Main>2%   set text "Hola mundo!"
Hola mundo!
<Main>3%   puts "el valor de text es:$text."
el valor de text es:Hola mundo!
<Main>4%   set result [ expr -3 * $x ]
-30
<Main>5%   puts $result
-30

```

Habitualmente es necesario realizar operaciones secuenciales siguiendo un valor de indexación. Para esto hacemos uso del comando `for {inicialización} {test} {incremento} {comandos}`. Un ejemplo de su utilización se muestra a continuación

```

<Main>6%   set file [open "mioutput.dat" w]
<Main>7%   for {set x 0} {$x <= 10} {incr x} {
<Main>8%       puts $file [ expr -3 * $x ]
<Main>9%       }
<Main>10%  close $file

```

En la primera instrucción asignamos a la variable `file` un *puntero* o direccionamiento que permite acceder al archivo *mioutput.dat*, el cuál se ha creado con el comando `file` y habilitado para escribir contenido en él. En la segunda instrucción hacemos uso de un ciclo `for` que nos permitirá escribir en el archivo el resultado de una operación matemática asociada a las veces en que se ejecuta la instrucción, delimitado a valores de la variable `file`, a la cuál se le asigna un valor que va desde 0 a 10, en incrementos de 1 unidad (función `incr`). Finalmente, el comando `close $file` cierra el archivo, habilitándolo para su lectura.



Listas en Tcl. Las listas en Tcl son agrupaciones de elementos de cualquier tipo, incluso listas. Cada elemento posee un indicador numérico único (índice) que inicia en 0. El comando `lindex $lista i` devuelve el elemento que ocupa el lugar $i - 1$ de la lista. De esta forma para acceder al primer elemento de la lista $i = 0$, para el segundo $i = 1$, etc.

El lenguaje Tcl permite el manejo fácil de listas (arreglos de datos). Múltiples funciones están disponibles para esta tarea. Para crear una lista utilizamos el `list`, para obtener la cantidad de

elementos de la lista el comando `llength` y para recorrerla podemos utilizar la función `foreach`.

```
<Main>11% set lista {1 2 3 4}
A B C D
<Main>12% llength $lista
4
<Main>13% foreach elemento $lista {
    puts "el elemento es $elemento"
}
el elemento es 1
el elemento es 2
el elemento es 3
el elemento es 4
```

También es posible combinar múltiples listas

```
<Main>14% set lista2 {A B C D}
A B C D
<Main>15% foreach elemento_1 $lista elemento_2 $lista2 {
    puts " * $elemento_1 $elemento_2 *"
}
1 A
2 B
3 C
4 D
```

y ejecutar una acción específica combinando los diferentes comandos y estructuras de programación. ¿Cuál es el resultado de ejecutar el siguiente código?

```
<Main>16% set lista {1 2 3 4 5 6 7 8 9 10}
<Main>17% set n [llength $lista]
<Main>18% set file [open "mioutput.dat" w]
<Main>19% for {set x 0} {$x < $n} {incr x} {
    set numero [expr $n - [lindex $lista $x]]
    puts $file $numero
}
<Main>20% close $file
```

3.2.2. Obtención de coordenadas y propiedades estáticas de una selección

Los comandos básicos de Tcl pueden ser combinados con las funciones de selección y análisis de VMD para obtener información del sistema de estudio.

- 3 El inicio de este práctico es cargar los archivos de topología y coordenadas utilizando la consola de comandos. Para eso en la consola Tk utilizamos los siguientes comandos

```
<Main>21% set mol [mol new agua.psf type psf waitfor all]
0
<Main>22% mol addfile 300.pdb type pdb waitfor all
```

La variable `mol` nos permite almacenar el identificador del sistema que ha sido cargado con el comando `mol new`. En la ventana `Main` se observa que se ha cargado el sistema. El valor de ID corresponde con el valor de la variable `$mol`. Por otro lado, en la ventana `vmd console window` hay información de la cantidad de átomos, residuos, enlaces del sistema.

- 4 Una vez cargado el sistema es posible realizar selecciones de los átomos de interés utilizando el comando `atomselect`. En este ejemplo se obtendrá la posición de una molécula de agua que se le asigne un valor de *ocupancia* igual a 1 en el archivo `pdb`, para ello escriba en la consola

```
<Main>23% set AguaSel [atomselect top "occupancy 1"]
```

Esto crea una selección de los átomos que cumplen con la condición de tener “`occupancy 1`” que es almacenada en la variable `$AguaSel`. En vez de utilizar el valor de ID de la molécula, se utilizó la palabra `top`, que hace referencia a la molécula marcada en la columna T en la ventana `Main` de VMD.

El resultado del comando `atomselect` es una función (`$AguaSel`) que permite operar sobre los átomos seleccionados a través de la aplicación de comandos. Una amplia gama de comandos se encuentran disponibles¹, a continuación se ilustrará el uso de alguno de ellos.

- 5 Para conocer la cantidad de átomos seleccionados se utiliza el comando `num`.

```
<Main>24% $AguaSel num
3
```

El resultado es la aplicación de la función `num` sobre la selección `$AguaSel`, que entrega el número de átomos en la selección.

- 6 Para conocer el nombre del o los residuos a los cuales pertenecen los átomos seleccionados y otros datos de interés se utiliza el comando `get`.

```
<Main>25% $AguaSel get resname
<Main>26% $AguaSel get resid
```

la primera instrucción permite conocer el nombre (*resname*) y la segunda el identificador (*resid*) de los residuos a los que pertenecen las partículas seleccionadas.²

- 7 Para identificar una partícula disponemos de distintos campos, escriba en la consola los siguientes comandos y asegure de entender a que corresponde cada campo solicitado.³

```
<Main>27% $AguaSel get index
<Main>28% $AguaSel get serial
<Main>29% $AguaSel get type
<Main>30% $AguaSel get name
<Main>31% $AguaSel get beta
<Main>32% $AguaSel get occupancy
<Main>33% $AguaSel get user
```

- 8 Para obtener una mayor cantidad de datos simultáneamente es posible combinar distintos campos de la función `get`, por ejemplo, para conocer la posición en el espacio cartesiano de las partículas escribiremos

¹ver el manual de usuario de VMD <http://www.ks.uiuc.edu/Research/vmd/current/ug/>,

²Como aparecen en el archivo `pdb`

³ http://deposit.rcsb.org/adit/docs/pdb_atom_format.html,

```
<Main>34%   $AguaSel get {x y z}
{1.5039819478988647 -3.501293897628784 1.5630073547363281}
{1.3331609964370728 -2.644653558731079 2.0667877197265625}
{0.6170657277107239 -3.953375816345215 1.5228157043457031}
```

que nos entrega una lista de datos, donde cada elemento entre paréntesis (`{}`) son las coordenadas de cada átomo de la selección utilizada. También es posible combinar los distintos campos para identificar cada partícula, por ejemplo

```
<Main>35%   $AguaSel get {index name serial x y z}
{1257 OH2 1258 1.50398194789 -3.50129389762 1.5630073547363281}
{1258 H1 1259 1.333160996437 -2.64465355873 2.0667877197265625}
{1259 H2 1260 0.617065727710 -3.95337581634 1.5228157043457031}
```

- 9 Los comando utilizados entregan los datos en pantalla, para ejecutar tareas de mayor complejidad es útil almacenar los datos obtenidos en variables para acceder a ellos posteriormente. Esto se realiza combinando las funciones `get` y `[]`

```
<Main>36%   set PosicionAguas [$AguaSel get {x y z}]
```

Para recuperar la información almacenada utilizamos el comando `puts`

```
<Main>37%   puts $PosicionAguas
```

el cuál muestra el contenido de la variable en pantalla. También es posible utilizar nuevamente esta variable en una nueva operación, por ejemplo, las siguientes instrucciones

```
<Main>38%   lindex [$AguaSel get {x y z}] 0
{1.5039819478988647 -3.501293897628784 1.5630073547363281}
<Main>39%   set PosX [lindex [lindex [$AguaSel get {x y z}] 0] 0]
1.5039819478988647
```

utilizan el comando `lindex` para recuperar el primer elemento de la lista generada con el comando `get`, para este ejemplo las coordenadas del primer átomo en la selección `$AguaSel` y en la segunda instrucción solo la coordenada x del mismo átomo.

- 10 También es posible modificar los distintos campos (`resname`, `resid`, `name`, `type`, x , y , z , β , `occupancy`, `user`, etc) disponibles para cada partícula. Para ello hacemos uso del comando `set` de la función `atomselect`. Las siguientes líneas de comandos modifican el valor del campo β de todas las partículas, asignándoles un valor $\beta = 0$.

```
<Main>40%   set Todos [atomselect $mol all]
<Main>41%   $Todos set beta 0
```

A continuación se modificará el campo β de los átomos de oxígeno que tienen una coordenada x positiva, para ello se realizará una nueva selección.

```
<Main>42%   set AguasX_1 [atomselect top "name OH2 and x>0"]
<Main>43%   $AguasX_1 num
<Main>44%   $AguasX_1 set beta 1
```

- 11 Las operaciones realizadas no generan una representación en la ventana OpenGL Display. Para visualizar los átomos a los cuales se les asignó el nuevo valor de β cree una representación de tipo vdW con la selección antes utilizada (“name OH2 and x>0”), coloreando la representación por el campo *beta* (fig. 14, ver sección ??).

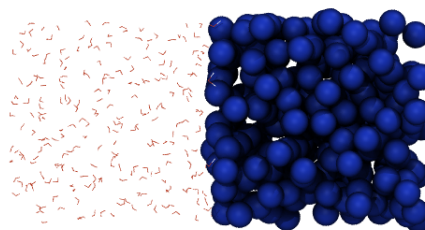


Figura 14: Representación obtenida

- 12 Anteriormente se almacenó en la variable \$PosicionesAguas, en forma de lista, las coordenadas de los átomos que tienen un valor de ocupancia = 1. Para utilizar estos datos es necesario acceder a cada elemento de la lista de manera ordenada. El siguiente código utiliza el comando `foreach` para esta tarea

```
<Main>45%   foreach PosAtomo $PosicionesAguas {
              puts "coordenadas $PosAtomo"
              }
coordenadas 18.53900715 2.6208207 6.898287
coordenadas 17.86873817 3.022436 6.4451613
coordenadas 18.59943001 1.7897097 6.286479
```

el que recorre la lista \$PosicionesAguas un elemento a la vez, almacenando el valor correspondiente en la variable PosAtomo y permite operar sobre esta última siguiendo las instrucciones delimitadas entre los paréntesis.

La correcta asignación de valores en los distintos campos permite reconocer las partículas para su seguimiento durante una trayectoria, colorear a partir de parámetros específicos y otros.



Operadores lógicos. Habitualmente utilizamos los operadores booleanos “Y”, “O” y “NO”. El primer es un operador “incluyente”, el segundo de “exclusión”, mientras que el tercer operador es considerado de “reunión” o suma lógica. Podemos ver la diferencia entre ellos observando la cantidad de átomos de selecciones que las incluyan como: “name OH2 or x>0’], “name OH2 and x>0’] y “name OH2 and not x>0’].

- 13 Antes de continuar explore distintas selecciones, representaciones y el uso de comandos en la consola Tk. Al terminar, reinicie VMD.



Actualización automática de una selección. Por defecto las selecciones realizadas con VMD no se actualizan para cada frame. Para que se actualicen, se deben activar las casillas Update selection every frame o Update color every frame en la viñeta trajectory de la ventana Graphical Representation. Pruebe esto realizando una selección en el cuadro inicial para los átomos que tengan coordenada *x* positiva.

3.2.3. Propiedades dinámicas

El análisis de una trayectoria nos entrega una gran cantidad de posibles propiedades de estudio, que se traducen en una alta cantidad de datos que deben ser manejados de manera ordenada o almacenados para su estudio posterior. A continuación se ilustrará como realizar este tipo de tareas obteniendo los datos de posición para una partícula determinada. Para esto haremos uso de lo aprendido hasta el momento en conjunto con nuevos comandos. Realice un programa en Tcl/VMD siguiendo las siguientes instrucciones. En caso de no poder realizarlo, al final de esta sección, se encuentra el código necesario paso por paso.

14 Seleccione el menú Extensions → Tk Console.

15 Cargue el archivo de conectividad y coordenadas utilizados como punto de inicio de la simulación de dinámica molecular realizada en la sección anterior.

```
<Main>46%  set mol [mol new agua.psf type psf waitfor all]
0
<Main>47%  mol addfile 300.pdb type pdb waitfor all
```

16 La trayectoria calculada se encuentra almacenada en el archivo `dcd`, cuyo nombre se encuentra definido en la línea `outputname` del archivo de configuración de NAMD. Para cargar el archivo de trayectoria utilizamos el comando `mol addfile`.

```
<Main>48%  mol addfile 300.dcd type dcd waitfor all
```

17 Una trayectoria consiste de diferentes configuraciones ordenadas de manera secuencial, para determinar el número de configuraciones disponibles utilizamos el comando `molinfo` en conjunto con la función `get`

```
<Main>49%  set frames [molinfo top get numframes]
1000
```

18 Utilice los controles de la ventana Main para recorrer la trayectoria, observe como la posición de la molécula de agua varía. ¿Qué información podemos obtener de ella?. Cada cuadro de la trayectoria corresponde a una configuración del sistema. El primer cuadro de simulación es el número 0 y el último es el total de cuadros es el valor almacenado en la variable `$frames - 1`.

En la sección anterior se obtuvo la posición de una partícula en una configuración estática. Ahora utilizaremos las herramientas de VMD para obtener las coordenadas de una partícula a lo largo de la simulación. A continuación se hará uso de lo aprendido anteriormente, no continúe esta sección sin haber realizado y comprendido las distintas tareas de las secciones anteriores.

19 En la consola Tk realice una selección, utilizando el comando `atomselect`, para el átomo de oxígeno de una molécula de su elección.



Trick: Obtención de coordenadas. Para obtener las coordenadas a partir de una variable (ie. `set Pos [$Agua get {x y z}]`) es necesario realizar tres operaciones: `set coordX [lindex $Pos 0]`; `set coordY [lindex $Pos 1]`; `set coordZ [lindex $Pos 2]`. Para optimizar el trabajo de cálculo podemos hacer uso del manejo de listas en Tcl y ocupar el comando `foreach` para realizar el trabajo en un solo paso: `foreach {x y z} $Pos {break}`

- 20 Actualice la posición de los átomos de la selección utilizando el comando `$seleccion frame cuadro`, donde `cuadro` es el número que identifica a un cuadro de simulación. Por ejemplo para el cuadro 1 y el cuadro 99

```
<Main>50%   set Agua1 [atomselect top "resid 4857 and name OH2"]
<Main>51%   $Agua1 num
<Main>52%   $Agua1 frame 0
<Main>53%   $Aguas get {x y z}
<Main>54%   $Agua1 frame 99
<Main>55%   $Aguas get {x y z}
```

Generalmente, las trayectorias de una simulación son suficientemente largas para impedir la obtención manual de la información de las partículas, entonces, para realizar los análisis debemos combinar las distintas estructuras de programación con las funciones de VMD para almacenar esta información de manera automática, ya sea en memoria o archivos, para su posterior análisis.

- 21 Para generar un archivo para almacenar los datos de la trayectoria haga uso del comando `open`, recuerde guardar el puntero al archivo en una variable.
- 22 Genere un bloque de instrucciones utilizando la instrucción `for` que permita recorrer la trayectoria cuadro a cuadro. Tip: recuerde la estructura de esta instrucción en las secciones anteriores, modifique para avanzar la variable de iteración desde 0 hasta el identificador del último cuadro de la trayectoria.
- 23 Obtenga las coordenadas $\{x, y, z\}$ de los átomos de la selección utilizando el comando `get`. Recuerde actualizar la selección para obtener los datos de cada cuadro de simulación. Guarde cada coordenada en variables independientes utilizando el comando `lindex`.



Actualización de selecciones. Una selección varía durante la trayectoria, por ejemplo, la posición de una partícula puede no ser la misma al cambiar de configuración. Las selecciones pueden tener un comportamiento dinámico que implica recalcular que partículas cumplen con la propiedad en cada marco de la trayectoria. Para esto hacemos uso de los comandos `frame` y `update`, mientras que el primero actualiza las propiedades de las partículas de una selección, el segundo actualiza que partículas pertenecen a una selección dada.

- 24 Muestre en pantalla las coordenadas $\{x, y, z\}$ utilizando el comando `puts`.
- 25 Guarde en un archivo, una línea a la vez, el número del cuadro de trayectoria actual y las coordenadas $\{x, y, z\}$ de la selección en el cuadro seleccionado. Utilice el comando `puts` dirigiendo la salida hacia el archivo utilizando el puntero antes creado con el comando `open`.
- 26 Cierre el bloque de instrucciones (bloque `for`). Luego cierre el archivo que utilizo para almacenar las coordenadas con el comando `close`.
- 27 Compruebe que el archivo creado contiene la información esperada. Puede utilizar el editor nativo de Tcl con el comando `edit nombre de archivo`.

Si Ud. no ha logrado el resultado esperado, a modo de ayuda, a continuación se muestra el código que permite obtener las coordenadas del átomo de oxígeno de la molécula de agua cuyo id es 4857, las coordenadas son almacenadas en el archivo *coordenadas.dat* y visualizadas en el editor de texto de la consola Tk.

```

<Main>56%   set Agua [atomselect top "resid 4857 and name OH2"]
<Main>57%   set frames [molinfo top get numframes]
<Main>58%   set file [open coordenadas.dat w]
<Main>59%   for {set i 0} {$i <= $frames} {incr i} {
<Main>60%       $Agua frame $i
<Main>61%       set Pos [$Agua get {x y z}]
<Main>62%       foreach {x y z} $Pos {break}
<Main>63%       puts "coordenadas x $x"
<Main>64%       puts "coordenadas y $y"
<Main>65%       puts "coordenadas z $z"
<Main>66%       puts $file "$i $x $y $z"
<Main>67%   }
<Main>68%   close $file
<Main>69%   edit coordenadas.dat

```

3.2.4. g(r): Implementación computacional

N Con los comandos y ejemplos entregados Ud. dispone de toda la información necesaria para calcular $g(r)$. Luego es tiempo de sentarse y crear el script que le fue solicitado en la sección 3.2. Considere que las tareas entregadas corresponden a las necesarias para determinar $g(r)$ sobre una configuración y partícula, por lo cual Ud. debe extender su solución para realizar el cálculo sobre todas las partículas y configuraciones disponibles.

Ñ Escriba un script para VMD que le permita calcular la función de distribución de pares.

O Compare la función de distribución radial calculada por ud. con los resultados obtenidos con VMD. Discuta el por qué de las diferencias obtenidas (el resultado de VMD es correcto)

P Explique si los sistemas de una caja de partículas de agua de tamaño finito es útil o no para representar un sistema real compuesto de agua.



Recorrer pares de átomos. La mayor parte de los cálculos a partir de una trayectoria incluyen determinar una función $F(i, j)$ que depende de los pares existentes entre N las partículas del sistema, para lo cual se utiliza una estructura del tipo $\sum_{i=1}^{N-1} \sum_{j=i+1}^N F(i, j)$

Periodic Boundary Conditions

Si ud. observa las distintas trayectorias de las cajas de aguas de este práctico se dará cuenta que a lo largo de la simulación el tamaño de la caja no varía, además que hay moléculas de agua que parecen desaparecer y aparecer en los bordes del sistema. Es más, hemos caracterizado estos sistemas líquidos a través de solo un pequeño número de partículas.

Una manera habitual de minimizar efectos de superficie y simular las propiedades de un sistema molecular es usar condiciones periódicas de borde. Esta técnica se utiliza ya que en usualmente utilizamos una pequeña cantidad de partículas para predecir las propiedades de los sistemas a un bajo costo computacional.

Bajo condiciones periódicas de borde, una caja de simulación ortorómbica es replicada en una grilla infinita. En el transcurso de la simulación, cuando una molécula se mueve en la llamada celda central (que corresponde a la celda “verdadera”), la imagen periódica de esta molécula se mueve en la misma dirección. De esta forma, si una molécula sale de la celda central, una de sus imágenes entrará a esta última por la cara contraria de la celda, manteniendo la cantidad de partículas en la celda central constante.

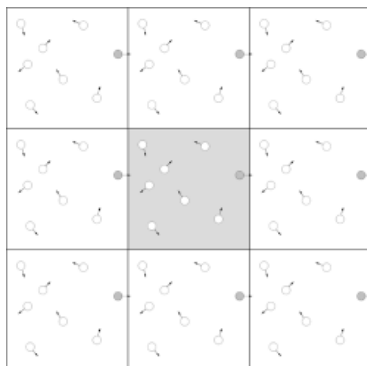


Figura 15: Representación de condiciones periódicas de borde. En gris la celda central.

No es necesario guardar las coordenadas de las partículas en cada una de las celdas de la grilla, sino solo las de la celda central. Para una “caja” de dimension L (una línea) con el origen en su centro, la condición de borde es equivalente a considerar esta línea como si fuera un círculo. La distancia entre las partículas puede ser medida a través del arco generado, donde la distancia que maximiza la separación entre las partículas es $L/2$.

De esta forma, el código para representar la condición periódica es simple. Si una partícula abandona la caja central cruzando uno de sus bordes, solo debemos sustraer o sumar L al valor de la coordenada resultante. Esto es fácil de lograr usando una instrucción IF luego se han calculado las nuevas posiciones de las partículas, por ejemplo :

```

if (X > Lx)
    X = X - Lx;
else if (X < 0)
    X = X + Lx;
# X es la coordenada de la partícula.
# Lx es el largo de la caja en la dimensión X

```

Q Genere un código que permita considerar PBC para su uso en el espacio cartesiano, utilice las herramientas de tcl y VMD

Otra forma de calcular las coordenadas de las partículas bajo PBC es actualizar las coordenadas del sistema utilizando la relación $x_i = (x_i + 2L) \bmod L$. Al igual que la aproximación anterior se conserva el número de partículas en la celda, ya que cada partícula es remplazada por una partícula ingresando a la celda en la dirección opuesta. El valor de $2L$ en este caso es utilizado para considerar las moléculas que cumplan con $x_i < -L$.

R Demuestre que ambas soluciones para aplicar PBC son correctas.

Convención de la imagen más cercana

Si analizó correctamente el significado y como calcular la función de distribución radial, ud. debe haber observado que existen problemas para calcular el número de partículas que rodean una partícula que esta cerca de los bordes de la celda central. Para solucionar esta problemática, inherente al uso de PBC, debemos utilizar la llamada “convención de la imagen más cercana” (“nearest image

convention”).

Imagine dos partículas i y j que se encuentran en lados opuesto de la celda central. La convención dicta que se debe considerar la menor distancia calculable entre bordes o entre imágenes adyacentes (que son replicas desplazadas de la celda central). De esta forma, si $\Delta x_{ij} = x_j - x_i$ es mayor que $L/2$, la coordenada de la partícula j no debe ser considerada con para calcular la distancia Δx_{ij} , debiendo utilizar la distancia con su partícula imagen de la celda “a su izquierda” en su lugar, la cual tiene como coordenada $x_j - L$.

En la práctica la aplicación de la “convención de la imagen más cercana” es simple, ya que para calcular cualquier propiedad $F(x_{ij})$, dependiente de la distancia entre dos partículas i y j , solo debemos calcular la cantidad $\Delta x_{ij} - L$ en vez de Δx_{ij} . Lo cual también rige para $\Delta x_{ij} \leq -L/2$ y para otras coordenadas.

De esta forma, según la regla de la “convención de la imagen más cercana” el valor Δx_{ij} para calcular cualquier propiedad $F(x_{ij})$ queda definida como:

$$\Delta x'_{ij} = \Delta x_{ij} - L * \text{int} \left(\frac{\Delta x_{ij}}{L} \right).$$

S ¿Que sucede cuando $\frac{\Delta x}{L}$ toma los valores de -0.49 y -0.51? ¿Que característica debe tener entonces la función “int” para entregar el resultado de manera correcta?



Unidades reducidas. Algunos programas optimizan el tiempo de cálculo o reducen el error de punto flotante utilizando las llamadas unidades reducidas. Por ejemplo, llevando el valor de $L = 1$ para el cálculo de la “convención de la imagen más cercana” reduce drásticamente el tiempo de cálculo. ¿Por qué? Por ejemplo, una caja de Argón en un sistema LJ esta definida como (El lado derecho contiene el nombre que le hemos dado a estas unidades):

- $\sigma = \varepsilon = m = k_B = 1$
- $3,405 \cdot 10^{-10} \text{ m} = 1 \text{ LU (unidad de longitud)}$
- $1,654 \cdot 10^{-21} \text{ J} = 1 \text{ EU (unidad de energía)}$
- $6,69 \cdot 10^{-26} \text{ kg} = 1 \text{ MU (unidad de masa)}$
- $1,199 \cdot 10^2 \text{ K} = 1 \text{ TU (unidad de temperatura)}$
- $2,17 \cdot 10^{-12} \text{ s} = 1 \text{ tU (unidad de tiempo)}$

De esta forma el sistema de Argón de número de densidad $n = 0,9 \text{ LU}^{-3}$. Su densidad de masa $\rho = nm = 0,9 \text{ MU/LU}^3 = 1,5 \text{ g/cm}^3$.

T Verifique e incorpore la “convención de la imagen más cercana” a la función de distribución radial calculada por ud. Cabe destacar que todas estas expresiones pueden ser reemplazadas con expresiones del tipo IF.

U Calcule la función de distribución radial para toda su trayectoria. Para ello le dejamos unas líneas de código que le pueden ser útil.

```
for {set m 0} {$m < [expr $NF]} {incr m} {
  for {set i 0} {$i < [expr $indl-1]} {incr i} {
    for {set j [expr $i+1]} {$j < [expr $indl]} {incr j} {
      set rxij [expr [$atom1 get x] - [$atom2 get x]]
      set rxijm [expr $rxij - $L*round(double($rxij)/$L)]
    }
  }
}
```

```

for {set i 0} {$i< $maxbin} {incr i} {
  set nideal [expr 4.0*3.141592654*$rho/3.0* \
    (pow($rupper,3)-pow($rlower,3))]
}

```

3.3. Coeficiente de difusión

Difusión es la migración de gases o materia en estado líquido en un gradiente de concentración. La migración de materia esta descrita cuantitativamente por el número de partículas atravesando una unidad de área en un tiempo determinado (flujo). La primera ley de Ficks indica que el flujo, J , es proporcional al gradiente de concentración c :

$$J = -D\nabla c \quad (5)$$

El coeficiente de proporcionalidad, D , es llamado *coeficiente de difusión* y es un parámetro importante en procesos químicos y distintos fenómenos de transporte. En general, el coeficiente de difusión es una propiedad macroscópica, sin embargo, es posible calcularlo a través del análisis de las velocidades y posiciones de las partículas, las cuales pueden ser obtenidas a partir de una trayectoria de DM.

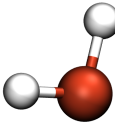
La expresión de Einstein

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle |r(t) - r(0)|^2 \rangle \quad (6)$$

relaciona el desplazamiento de las partículas con el coeficiente de difusión, a través de la media de la distancia recorrida por cada partícula por unidad de tiempo $\langle |r(t) - r(0)|^2 \rangle$ (*MSD*, mean square distance). Esta relación nos permite determinar la constante de difusión de una partícula a partir del cálculo de las posiciones en cada paso o “cuadro” de la trayectoria de DM antes realizada. Para ello, podemos reescribir esta relación como

$$\langle |r(t) - r(0)|^2 \rangle = \lim_{t \rightarrow \infty} 6tD \quad (7)$$

La cuál nos permite observar que existe una relación lineal entre los valores de *MSD* y D cuando el tiempo es infinito, luego al graficar los valores de *MSD* en el tiempo, es posible calcular el valor de D a partir de la pendiente de la curva generada (ver cuadro anexo).



Relación de Green-Kubo. El coeficiente de difusión también puede ser calculado partir de las velocidades de las partículas utilizando la relación de Green-Kubo:

$$D = \frac{1}{3} \int_0^\infty \langle v(t) \cdot v(0) \rangle dt \quad (8)$$

¿Podemos obtener las velocidades directamente desde de una trayectoria? En NAMD podemos utilizar el comando velDCDfile.

3.4. Cálculando el coeficiente de difusión

Algoritmo computacional

Para calcular la constante de difusión es necesario computar la posición de cada partícula N_i durante los M cuadros que tiene la trayectoria obtenida tras la simulación y, a partir de ellas, calcular el desplazamiento en el tiempo para cada partícula.

El algoritmo 1 muestra una aproximación para calcular el valor de *MSD* durante los distintos periodos temporales de una simulación. En el, N representa el número de partículas en una trayectoria de M “cuadros” de simulación. Cada cuadro puede ser considerado como un inicio de la

Algorithm 1 MSD N número de partículas $R_i(n)$ posición partícula i a tiempo n M cuadros totales en la trayectoria t origen dt variación de tiempo

```

1: for  $j = 0 \rightarrow M; j + 1$  do
2:   for  $t = 0 \rightarrow M; t + dt$  do
3:      $\text{cnt}(dt) = \text{cnt}(dt) + 1$ 
4:     for  $i = 1 \rightarrow N; i + 1$  do
5:        $\mathbf{r}(dt) += [\mathbf{R}_i(t + dt) - \mathbf{R}_i(t)] * [\mathbf{R}_i(t + dt) - \mathbf{R}_i(t)]$ 
6:     end for
7:   end for
8: end for
9: for  $t = 0 \rightarrow M; t + dt$  do
10:   $\text{MSD}(t) = \mathbf{r}(t) / [\text{cnt}(t) * N]$ 
11: end for

```

trayectoria, por lo cuál debemos promediar por sobre todos los orígenes posibles, de esta forma, el valor de MSD se calcula considerando el cambio en la posición de la partícula en un intervalo temporal de largo $M - t$. dt es una variable que itera sobre los tiempos de simulación permitidos, $\text{cnt}()$ es un arreglo que contiene el número de orígenes de tiempo para cada intervalo, $\mathbf{r}()$ es un arreglo que almacena el valor de desplazamiento en cada dimensión i (x, y, z). Note que cada arreglo tiene M elementos, uno por cada intervalo de tiempo, incluyendo el intervalo de largo 0. El último ciclo de este algoritmo, permite calcular la media del desplazamiento de las partículas en todos los segmentos de tiempo permitidos, con el cuál podemos calcular la constante de difusión.

Implementación de calculo de MSD con VMD

Cada linea del archivo `difusion.tcl` corresponde a una operación (escrita en el lenguaje de VMD) que se realizará para completar una tarea específica. A continuación se describen las operaciones de mayor importancia.

- 1 El primer paso es cargar los archivos de conectividad y trayectoria correspondiente

```

<Main>70%      set mol [mol load psf BOX.psf ]
<Main>71%      mol addfile BOX.dcd step 100 waitfor all

```

- 2 Si se utilizaron condiciones periódicas de borde durante la simulación es necesario reescribir las coordenadas para considerar el desplazamiento real cada partícula del sistema. Para esto podemos utilizar el modulo PBCtools de VMD o simplemente entregar un archivo de coordenadas con las coordenadas reales.

```

      ;# recuperando coordenadas desde condicones periodicas
<Main>72%      pbc unwrap

```

- 3 Para calcular las coordenadas es necesario recuperar la cantidad de cuadros del sistema, como también la cantidad de partículas de la selección deseada

```

      ;# cantidad de cuadros
<Main>73%      set frames [molinfo $mol get numframes]

```

```

        ;# seleccion del átomo 80
<Main>74%   set atoms [atomselect $mol "resid 4857 and name OH2"]
<Main>75%   set n [$atoms num]

```

4 Las siguientes líneas de código permiten iterar por cada cuadro de simulación recuperando las coordenadas de cada partícula.

```

        ;# iterando sobre todos los cuadros
<Main>76%   for {set f 0} {$f < $frames } {incr f} {
<Main>77%     $atoms frame $f
        ;# coordenadas de todos los atomos
<Main>78%     set coordinates [$shell get x y z]
        ;# guardo las coordenadas
<Main>79%         for {set p 0} {$p<$n} {incr p} {
<Main>80%             ;# virtual array a(frame,particle) <--- {x,y,z}
<Main>81%             set a($f,$p) [lindex $coordinates $p]
<Main>82%         }
<Main>83%     }

```

El resultado de las operaciones anteriores es un arreglo virtual del tipo `Arreglo(f,i)` que contiene las coordenadas $\{x, y, z\}$ de la partícula i en el cuadro f de la trayectoria.

Una vez almacenadas las coordenadas en memoria es posible utilizarlas para cualquier objetivo deseado, en esta sección se utilizarán para calcular el valor de *MSD*. El cálculo del valor de *MSD* se realiza en un intervalo de tiempo $[t_0, t]$, donde el valor t_0 es un origen aleatorio del análisis, luego, debemos considerar todos los intervalos posibles durante la simulación.

5 Calculamos el desplazamiento de cada partícula en un periodo de tiempo dt , considerando intervalos de tiempo de igual duración durante la simulación

```

<Main>84%   for {set f 0} {$f < $times } {incr f} {
<Main>85%       for {set dt 1} {[expr $f + $dt] < $frames} {incr dt} {
<Main>86%           incr cnt($dt)
<Main>87%       for {set p 0} {$p<$n} {incr p} {
<Main>88%           set msd($dt) [expr $msd($dt) +
<Main>89%               [veclength2 [vecsub $a([expr $f + $dt], $p) $a($f, $p)]]]
<Main>90%       }
<Main>91%     }
<Main>92%   }

```

6 Finalmente se calcula el desplazamiento promedio de cada partícula, almacenando este valor en una lista

```

<Main>93%   for {set dt 1} {$dt< $times} {incr dt} {
<Main>94%       lappend msdval [expr $msd($dt)/($n*$cnt($dt))]
<Main>95%       lappend tval $dt
<Main>96%   }

```

7 Es posible almacenar en un archivo el valor de *msd* y el tiempo para su posterior uso en el análisis.

```

<Main>97%   set fp [open msd.dat w]
<Main>98%   foreach t $tval m $msdval {
<Main>99%       puts $fp "$t $m"
<Main>100%  }
<Main>101%  close $fp

```

- 8 Utilizando la herramienta multiplot de VMD se puede graficar la evolución del valor de MSD en el tiempo.

```
<Main>102% set plothandle [multiplot -x $tval -y $msdval \\  
-title "Example plot" \\  
-lines -linewidth 3 -marker point -plot]  
<Main>103% $plothandle replot;
```

- 9 Escriba y ejecute este script utilizando el comando `play` en la consola Tk.

```
<Main>104% prompt play difusion.tcl
```

- 10 En pantalla se muestra el gráfico de *MSD vs tiempo*, modifique este script para graficar el valor de MSD en el tiempo para las distintas simulaciones realizadas.

- 11 La figura 16 muestra la dependencia del valor de MSD con el tiempo. La relación lineal es evidente. ¿Como se comparan sus gráficos al mostrado en la imagen? ¿A que se deben las diferencias?

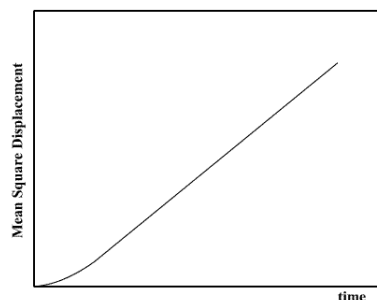
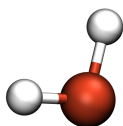


Figura 16: MSD en el tiempo



¿Régimen balístico o browniano?. En periodos cortos de simulación el gráfico pierde su linealidad. Cuando una partícula recorre un trayectoria de manera libre, sin colisiones, esta se encuentra realizando un desplazamiento a velocidad constante; una vez que se efectúa una primera colisión con otra partícula su movimiento se asemeja a desplazamiento aleatorio. Que regimen debemos utilizar para calcular la constante de difusión.

Simulación Molecular de Líquidos, Sólidos y Gases

La densidad de un líquido es levemente diferente a la de su estado sólido, por lo cual, la separación media de las partículas en ambos estados no debieran diferir en gran medida. Sin embargo existe una diferencia fundamental entre ambos estados: el estado cristalino sólido presenta un orden de largo alcance, mientras que el estado líquido solo presenta un orden local. Por otro lado, el estado sólido amorfo, por ejemplo el vidrio, no tiene un orden de largo alcance y presenta diferencias con respecto a la estructura de un líquido. Otra diferencia entre estos estados está dada por la velocidad con que difunden sus partículas. Una partícula puede acceder a todo el espacio ocupado por un líquido en un “corto” periodo de tiempo, mientras que en los otros estados el movimiento está impedido por la mayor cantidad de moléculas adyacentes, haciendo su movimiento más lento, casi nulo. De esta forma, el comportamiento de un líquido, sólido o un gas está directamente relacionado con su estructura, y tener las herramientas que permitan describirlas es fundamental.

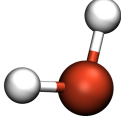
En esta unidad utilizaremos NAMD para generar una trayectoria temporal de un sistema compuesto por partículas y moléculas bajo distintas condiciones. Para esto, le entregamos los conocimientos básicos para utilizar los archivos de configuración suministrados para realizar simulaciones de dinámica molecular bajo las condiciones deseadas.

Los ejemplos a utilizar incluyen un sistema de moléculas de agua rígidas, un sistema monoatómico (Argón) y un sistema de moléculas flexibles (etanol). Algunas referencias teóricas y experimentales se encuentran disponibles para la comparación y corroboración de sus resultados. Como puede observarse de la literatura asociada a este práctico, los sistemas planteados parecen ser respuesta a problemas antiguos, sin embargo, aun permiten resolver problemáticas de interés actual. En todos los casos a analizar el punto de inicio es un archivo que contiene las posiciones atómicas de las partículas de interés.

4. Dinámica Molecular

En simulación molecular se desarrollan modelos desde la teoría o desde nuestro conocimiento experimental. Luego estos modelos son sometidos a pruebas en sistemas controlados, lo que permite refinar las teorías y representar de manera más certera lo que ocurre en nuestras observaciones experimentales. A su vez, este proceso de desarrollo permite realizar predicciones, las cuales pueden ser observadas en nuevos experimentos o análisis de datos previos. De esta forma la simulación molecular es el puente entre la teoría y la experimentación.

Dos técnicas resaltan entre los estudios de sistemas moleculares, Monte-Carlo y Dinámica Molecular. Ambos métodos requieren el cálculo de las interacciones entre las diferentes partículas que componen el sistema, sin embargo, la progresión de las simulaciones ocurre de maneras diferentes. Monte-Carlo utiliza una búsqueda aleatoria de las configuraciones del sistema, las cuales corresponden con ensambles específicos de la mecánica estadística, por otro lado, dinámica molecular está basada principalmente en el resolver las ecuaciones de movimiento, lo cual permite muestrear de forma secuencial las configuraciones que conforman un ensamble. Ambas técnicas permiten calcular las propiedades termodinámicas, sin embargo, solo dinámica molecular permite investigar el comportamiento dinámico de las diferentes cantidades que permiten caracterizar el sistema de estudio.



La historia de la Dinámica Molecular. El método de dinámica molecular fue introducido por primera vez en 1950 por Alder and Wainwright en sus estudios sobre el movimiento e interacción de esferas rígidas. En el 64, Rahman reportó su estudio en que por primera vez se utilizó un potencial “realístico” para simular el comportamiento del Argón en estado líquido. Luego, en 1974 Stillinger y Rahman incorporaron un mayor nivel de complejidad, simulando un sistema de moléculas de agua en estado líquido. Ya en 1977 McCammon, perteneciente al grupo de M. Karplus, simuló el primer sistema proteico: el inhibidor de la tripsina pancreática (BPTI).

- Alder, B. J. and Wainwright, T. E. J. Chem. Phys. 27, 1208 (1957)
- Alder, B. J. and Wainwright, T. E. J. Chem. Phys. 31, 459 (1959)
- Rahman, A. Phys. Rev. A136, 405 (1964)
- Stillinger, F. H. and Rahman, A. J. Chem. Phys. 60, 1545 (1974)
- McCammon, J. A., Gelin, B. R., and Karplus, M. Nature (Lond.) 267, 585 (1977)

4.1. Ecuaciones del Movimiento

Como fue mencionado anteriormente, dinámica molecular se enfoca principalmente en la evolución temporal de las configuraciones moleculares. Como herramienta, es útil para entender experimentos o poner a prueba teorías, sin embargo, para poder realizarlo es necesario que la evolución temporal de las configuraciones se realiza de una forma que “imite” la evolución real de un sistema molecular. Para satisfacer esta condición debemos generar una metodología que se base en lo que conocemos de la configuración de un sistema de partículas a un tiempo t_1 , que básicamente es que cada partícula en la configuración tiene una posición (\mathbf{q}) y una velocidad ($\dot{\mathbf{q}}$), y a partir de estos datos, conocer qué configuración adopta a un tiempo t_2 . La respuesta a este desafío metodológico se encuentra en las ecuaciones clásicas del movimiento, las que se pueden reescribir para mayor utilidad usando la formulación Lagrangiana.

Si se considera el principio de Hamilton, desarrollado a partir de la observación que cuando la naturaleza realiza una acción S esta ocurre minimizando el gasto de energía, de tal manera que

$$S = \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}) dt \quad (9)$$

Hamilton observó que la acción “ S ” debía ser mínima para todos los sistemas de partículas, definida como la integral de una función Lagrangiana “ L ” desconocida entre dos instantes de tiempo dados, que cumple que: Donde la función Lagrangiana ($L(\mathbf{q}, \dot{\mathbf{q}})$) depende de la posición y velocidades y satisface que

$$L(\mathbf{q}, \dot{\mathbf{q}}) = K - V, \quad (10)$$

donde K es la energía cinética y V es la energía potencial.

Bajo estas condiciones, la integral del “Lagrangiano” o “Lagrangiana” en el tiempo tiene un valor estacionario cuando la acción recorre una trayectoria de movimiento (path of motion), es decir, la variación de la integral del Lagrangiano en el tiempo debe ser cero. We

$$\delta \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}) dt = 0. \quad (11)$$

Esta variación puede ser reescrita en términos de las variables del Lagrangiano,

$$\int_{t_1}^{t_2} \sum_{i=1}^{3N} \left(\frac{\partial L}{\partial \mathbf{q}_i} \delta \mathbf{q}_i + \frac{\partial L}{\partial \dot{\mathbf{q}}_i} \delta \dot{\mathbf{q}}_i \right) dt = 0. \quad (12)$$

Ya que $\dot{\mathbf{q}}$ es la derivada de \mathbf{q} con respecto al tiempo e integrando la segunda derivada parcial en el paréntesis por partes, esta ecuación se simplifica a

$$\int_{t_1}^{t_2} \sum_{i=1}^{3N} \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}_i} - \frac{\partial L}{\partial \mathbf{q}_i} \right) \delta \mathbf{q}_i dt = 0, \quad (13)$$

y como cada variable es independiente, es posible separar la contribución de cada una de ellas

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}_i} - \frac{\partial L}{\partial \mathbf{q}_i} = 0 \quad (i = 1, 2, \dots, N). \quad (14)$$

A partir de esta ecuación es posible obtener las ecuaciones clásicas del movimiento de las partículas, para esto se debe sustituir la ecuación (10) en la ecuación anterior (14) considerando $m\dot{\mathbf{r}}^2/2$ para la energía cinética, lo cual da como resultado

$$\frac{d}{dt}(m\dot{\mathbf{r}}) + \frac{dV}{d\mathbf{r}} = 0, \quad (15)$$

o en una forma más reconocible,

$$\mathbf{f} = m\mathbf{a}, \quad (16)$$

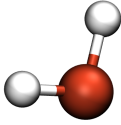
donde además se cumple que

$$\mathbf{f} = -dV/d\mathbf{r}$$

y

$$\mathbf{a} = d^2\mathbf{r}/dt^2$$

Esta formulación Lagrangiana (ecuación (14)) es una forma generalizada y puede ser usada para determinar las ecuaciones del movimiento incluso fuera del espacio Cartesiano.



Espacio de fase y el espacio de estados. El espacio de fase es el espacio en el cual residen los microestados de un sistema. Algunas veces es usado solo para problemas que pueden ser descritos en terminos de cordenadas espaciales y momentos, como también para los problemas en los cuales se pueden aplicar las ecuaciones de movimiento en su forma descrita por Hamilton. En otras oportunidades se utiliza el término espacio de estados para describir el espacio de microestados que no son posibles de describir solo con sus momentos y cordenadas.

Integración de las ecuaciones de movimiento

En dinámica molecular el objetivo es obtener la evolución temporal de las configuraciones del sistema, para esto es necesario un algoritmo que permita integrar las ecuaciones de movimiento del sistema descritas en la sección anterior. Un algoritmo ideal debe ser simple, preciso y certero en la descripción del movimiento, como a su vez cumplir con restricciones como la conservación de la energía del sistema. Existen diferentes algoritmos que cumplen con estas restricciones, sin embargo, por razones históricas nos centraremos en el llamado algoritmo de integración de Verlet.

En su búsqueda de realizar “experimentos computacionales” del comportamiento de partículas, Verlet integró las derivadas de segundo orden de las ecuaciones del movimiento de Newton

$$m \frac{d^2 \mathbf{r}_i}{dt^2} = \sum_{j \neq i} \mathbf{f}(r_{ij}), \quad (17)$$

a través de escribir la expansion de Taylor de segundo y tercer orden las las posiciones $\mathbf{r}(t)$, una hacia adelante y otra hacia atrás en el tiempo. Si consideramos a \mathbf{v} las velocidades, \mathbf{a} las aceleraciones, y \mathbf{b} la tercera derivada de \mathbf{r} con respecto a t , Entonces se puede describir el movimiento del movimiento hacia adelante y atras en el tiempo de las partículas de un sistema como:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + (1/6)\mathbf{b}(t)\Delta t^3 + O(\Delta t^4) \quad (18)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 - (1/6)\mathbf{b}(t)\Delta t^3 + O(\Delta t^4) \quad (19)$$

Al sumar ambas expresiones obtenemos que:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + O(\Delta t^4) \quad (20)$$

La que corresponde a la forma más simple del algoritmo de Verlet. Ya que estamos integrando las ecuaciones del movimiento de Newton, $\mathbf{a}(t)$ es nada más que la fuerza dividida por la masa, y la fuerza en función de las posiciones corresponde a $\mathbf{r}(t)$

$$\mathbf{a}(t) = -(1/m)\nabla V(\mathbf{r}(t)) \quad (21)$$

El algoritmo de Verlet tiene la ventaja de ser reversible en el tiempo, lo cual significa que uno puede integrar el movimiento configuracional hacia adelante o hacia atrás en el tiempo hasta la configuración inicial. Otra propiedad de este algoritmo es que es “symplectico” (symplectic), lo en palabras simples significa que el algoritmo mantiene la evolución del sistema en la región del espacio de fase correspondiente al ensamble elegido y que permite un alto grado de la conservación de la energía.

Inmediatamente es posible percibir que el error del algoritmo mientras el sistema se propaga o evoluciona en las posiciones es pequeño ($\mathcal{O}(\delta t^4)$), mientras que en las velocidades es sustancialmente largo ($\mathcal{O}(\delta t^2)$).

Un problema de esta versión del algoritmo de Verlet es que las velocidades del sistema no aparecen de manera explícita. Estas últimas, a pesar de no ser utilizadas en el cálculo de la evolución del sistema, muchas veces son útiles para calcular otras propiedades. En particular, son necesarias para calcular la energía cinética del sistema K , cuya evaluación es necesaria para corroborar que la energía total se conserva en el sistema en estudio, $E=K+V$. Esto es importante ya que es indicativo que la simulación procede de manera correcta.

Si uno desea calcular las velocidades, esto es posible mediante la ecuación

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t}. \quad (22)$$

Sin embargo, el error asociado a esta expresión es del orden Δt^2 y no de Δt^4 como el resto del algoritmo. Para solucionar este problema se hace uso de variantes del algoritmo de Verlet, como el algoritmo *leap-frog*. Una mejor implementación es el esquema *velocity Verlet*, en el cual posiciones, velocidades, aceleraciones y el tiempo $t + \Delta t$ son obtenidos de la siguiente manera

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + (1/2)\mathbf{a}(t)\Delta t^2 \quad (23)$$

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + (1/2)\mathbf{a}(t)\Delta t \quad (24)$$

$$\mathbf{a}(t + \Delta t) = -(1/m)\nabla V(\mathbf{r}(t + \Delta t)) \quad (25)$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t/2) + (1/2)\mathbf{a}(t + \Delta t)\Delta t \quad (26)$$

Donde solo se necesitan $9N$ posiciones de memoria para almacenar las $3N$ posiciones, velocidades y aceleraciones.

En su aplicación, los pasos de integración o **time-step** Δt normalmente es 1 femtosegundo ($\sim \frac{1}{20}$ periodo de vibración), sin embargo, el uso de algoritmos del estilo SHAKE permiten “congelar” los enlaces más cortos (como hidrógenos) y pueden permitir aumentar el Δt hasta 2 fs.

4.2. Ensambles

La densidad de probabilidad en el espacio de fase no puede ser calculada a partir de una única conformación obtenida en un tiempo dado. De ser así, el sistema se encontraría en un solo microestado, obtenido de manera aleatoria, cuando lo que queremos observar es la estadística de los diferentes microestados a los que puede acceder el sistema. Este problema fue solucionado por Josiah Willard Gibbs (Yale University), quien generó el concepto de ensamble, colectivo o conjunto termodinámico, el cual consiste en un número grande de sistemas idénticos en diferentes microestados. Los microestados de un sistema de M moléculas con f grados de libertad son puntos en un espacio de fase de $2fM$ dimensiones. Si logramos obtener información de la densidad de probabilidad asignada a esos puntos, esta puede ser utilizada para calcular las funciones de estado termodinámicas a través de las teorías de la probabilidad.

Dinámica molecular permite obtener el valor promedio del observable del ensamble (ensamble average o valor promedio del colectivo) a partir de una trayectoria larga de la evolución del sistema en el espacio de fase, en vez de un ensamble del sistema obtenido a un tiempo dado. Durante esta trayectoria, el sistema recorrerá diferentes microestados que están relacionados entre ellos debido a la resolución de las ecuaciones del movimiento, sin embargo, si se observa una trayectoria suficientemente larga es probable que se recorran todos los puntos disponibles del espacio de fase, es más, recorrer estos puntos con una frecuencia que corresponda a la densidad de probabilidad asociada del sistema. Esta idea es la base del análisis de las trayectorias de DM en términos de las funciones de estado termodinámicas, donde el valor promedio del observable del ensamble $\langle \mathcal{A} \rangle$ es reemplazado por el promedio temporal (time average) de \mathcal{A} , de tal forma que

$$\langle \mathcal{A} \rangle = \bar{\mathcal{A}} = \sum \mathcal{A}_i P_i \quad (27)$$

donde P_i representa la probabilidad de observar el estado i , los paréntesis (brackets) $\langle \rangle$ señalan el valor promedio del observable del ensamble, independiente del tiempo y $\bar{}$ el promedio temporal.

En el contexto de dinámica molecular, la hipótesis fundamental es que el promedio de una propiedad \mathcal{A} de un ensamble corresponde a la propiedad promedio de las diferentes conformaciones de la trayectoria analizada

$$\langle \hat{\mathcal{A}} \rangle = \frac{1}{t_{tot}} \int_0^{t_{tot}} \mathcal{A}(t) dt = \frac{1}{N} \sum_i \mathcal{A}(t_i)$$

volveremos a ello más adelante.

Los sistemas que cumplen esta condición se llaman “ergódicos”. Si un sistema es ergódico o no depende de las accesibilidad cinética del sistema a todos los “estados termodinámicamente accesibles” del espacio de fase.

A continuación veremos como se relacionar la “accesibilidad termodinámica” con la temperatura y la energía asociada a cada punto en el espacio de fase. En general, los puntos son accesibles si su energía no es relativamente más alta que el mínimo del espacio de fase. Si un sistema dinámico es capaz de visitar todos los puntos del espacio de fase a una temperatura dada – y el tiempo que requiere para ello – depende directamente de la magnitud de las barreras energéticas a las que se ve enfrentado, por lo que habitualmente en dinámica molecular no se logra muestrear todos las conformaciones termodinámicamente accesibles (microestados) dentro de los tiempos de simulación.

4.3. Probabilidad Termodinámica.

Ludwig Boltzmann (Austriaco) y James Clerk Maxwell (Escocés) son considerados los fundadores de la física teórica. A pesar que el trabajo de Boltzmann es basada en la estadística, logró conectar las funciones termodinámicas macroscópicas con las propiedades moleculares a través de la ecuación

$$S = k \ln W, \quad (28)$$

donde S es la entropía de una colección de moléculas distribuidas en una colección de niveles energético, k es la constante de la ley de los gases ideales R por molécula, y W es la llamada probabilidad termodinámica. Para calcular la entropía S de un sistema macroscópico, por ejemplo un gas, es necesario determinar el valor de la probabilidad termodinámica.

La probabilidad termodinámica es una medida del número de configuraciones que pueden adoptar las moléculas al distribuirse por los diferentes niveles energéticos, de esta forma, para las moléculas $W > 1$, además de crecer rápidamente con el número de niveles energéticos disponibles y, especialmente, con el número de partículas que conforman el sistema.

1. Se obtiene una probabilidad $W = 1$, cuando todas las moléculas comparten el mismo nivel de energía, llamado el estado basal (ground state) o nivel de mínima energía.
2. Si 100 moléculas se distribuyen a lo largo de 10 niveles energéticos de acuerdo a los siguientes números de ocupación 30, 25, 20, 15, 5, 5, 0, 0, 0, 0, la probabilidad de encontrar esta distribución es

$$W = \frac{N!}{n_1!n_2!\dots n_{10}!} = \frac{100!}{30!25!20!15!5!5!0!0!0!0!} \quad (29)$$

donde n_1, n_2, \dots, n_i son la cantidad de partículas n en el nivel energético i .

Gibbs fue el primero en reconocer la relación entre $S = k \ln W$, lo cual fue redescubierto por Boltzmann y Einstein de manera independiente. Ellos observaron que para un número N de partículas, la distribución más probable de estas en los diferentes niveles de energía es aquella que maximiza el valor de W , por lo cual esa distribución es la única que debe ser considerada ya que su aporte es significativamente mayor que todas las otras distribuciones posibles. A partir de esta observación, se entiende que el valor de la entropía es función del valor máximo que puede alcanzar la probabilidad termodinámica, es decir, maximiza la cantidad de configuraciones posibles del sistema al utilizar la mayor cantidad de niveles energéticos disponibles. Luego, para un material en particular se obtiene que,

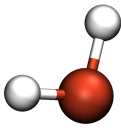
$$S = k \ln W_{max}. \quad (30)$$

Por su parte, Boltzmann demostró que la configuración más probable adoptada por un conjunto de partículas distribuidas en diferentes estados energéticos, bajo la restricción de un número de partículas N y energía E constante y niveles de energía ϵ_i , esta dada por

$$P_i = \frac{n_i}{N} = \frac{e^{-\epsilon_i/kT}}{\sum e^{-\epsilon_i/kT}} \quad (31)$$

Note que $0 \leq \frac{n_i}{N} < 1$ es la probabilidad verdadera de encontrar una partícula específica en el i -ésimo nivel de energía. Lo cual puede ser definido como “La probabilidad $P_i = n_i/N$ de encontrar una partícula con energía ϵ_i esta asociada con la densidad de probabilidad del ensamble microcanónico con una energía ϵ_i ”.

La distribución de Boltzmann (ec. 31) muestra que la cantidad de moléculas en un nivel de energía particular ϵ_i declina cuando ϵ_i aumenta. Este resulta de la restricción de mantener la energía total del sistema E constante, ya que las moléculas en un nivel energético alto fuerzan a las otras moléculas a niveles inferiores donde el número de distribuciones disponibles son menores.



Función de partición canónica. La suma $Z(N, V, T) = \sum e^{-\epsilon_i/kT}$ requerida para la normalización en la ecuación ec. 31 es llamada Función de partición canónica. Esta función termodinámica es una función de estado. La dependencia de N y V surge ya que los niveles energéticos disponibles están influenciados por el número de partículas y el volumen que estas tienen accesibles.

4.4. Ensamble termodinámico

En la sección anterior se estudió la distribución de los estados energéticos que puede adoptar un sistema de N partículas, donde el conjunto de microestados accesibles del sistema definen el llamado ensamble o colectivo, a partir del cual podemos calcular el promedio de una propiedad (ensamble average) del sistema en estudio.

Una simulación molecular se desarrolla bajo condiciones termodinámicas bien definidas, por ejemplo un número constante de partículas N , volumen V y/o temperatura T . Estas condiciones especifican el llamado “ensamble termodinámico” del sistema. Dado que en dinámica molecular se basa en la integración de las ecuaciones del movimiento de Newton que llevan a la conservación de la energía, su ensamble natural es el denominado microcanónico (NVE), ya que en el se mantiene esta condición. En cambio, para otras técnicas como el muestreo tipo Monte-Carlo, el ensamble natural es el NVT o canónico, el cual no requiere mantener la energía constante ya que en el no se considera el aporte cinético a la energía total del sistema.

La característica principal del ensamble microcanónico es representar un sistema aislado de N partículas y un volumen fijo. Al ser un sistema aislado la energía total del sistema E también se encuentra fija, de esta forma, el sistema está confinado a recorrer el espacio de fase correspondiente a ese nivel de energía ε_i (llamada hipersuperficie de Energía). De esta forma, la esencia del ensamble microcanónico es que todos los microestados accesibles para generar el ensamble tienen la misma energía $E = \varepsilon_i$.

Si el sistema puede intercambiar energía, pero no está habilitado para cambiar su número de partículas ni su volumen, el ensamble pasa a ser llamado “ensamble canónico”. En el, la energía del sistema E fluctúa centrada en el promedio de la energía del ensamble en un intervalo $[E, \Delta E]$, en un “equilibrio termodinámico”. Una mirada alternativa es considerar que el sistema se encuentra en un, es decir en un equilibrio térmico, donde las variaciones de energía del sistema son amortiguadas por el contacto del mismo con un baño térmico (por ejemplo, un volumen de agua de temperatura constante). Debido a esto, las energías de cada microestado que conforma el ensamble pueden diferir, sin embargo la temperatura T del ensamble debe ser consistente con la distribución de la densidad de probabilidad en el espacio de fase. Cabe destacar que T corresponde entonces al valor de la temperatura del baño térmico, habitualmente la temperatura del ambiente. Ya que la energía de un ensamble canónico fluctúa en el intervalo $[E, \Delta E]$, este ensamble puede ser considerado como un ensamble compuesto por subensambles microcanónicos de energías ε_i , de esta forma la densidad de probabilidad de este ensamble obedece una distribución de Boltzmann (ec. 31).

Temperatura y el teorema de la Equipartición de la Energía

La definición de temperatura a volumen constante es $T_v = \delta U / \delta S$, por lo cual la Temperatura es un concepto macroscópico, ya que esta derivada no tiene sentido para sistemas microscópicos. Una definición más relevante es que “La temperatura es la cantidad que se iguala cuando dos sistemas tienen permitido compartir energía entre ellos”. Imagine que tiene dos sistemas cerrados A y B , de volumen constante y compuestos por igual cantidad de átomos de gases nobles. Considere además que los sistemas se encuentran a temperaturas diferentes, con $T_A > T_B$. Las leyes de la termodinámica indican que al colocar ambos sistemas en contacto, los sistemas alcanzarán el llamado el equilibrio termodinámico, donde $T_A = T_B$. Si los sistemas no pueden intercambiar materia y no ocurren reacciones químicas en ellos, entonces la única opción disponible para que el sistema alcanzará el equilibrio es que el sistema de menor temperatura es aumentar su energía cinética.

A partir de lo anterior, y de manera simple, podemos decir que la temperatura es una medida de la energía cinética media del sistema, es decir depende de los movimientos y vibraciones submicroscópicas aleatorias de las partículas que constituyen un sistema. Considerando lo anterior, la temperatura depende de los grados de libertad de movimiento de las partículas y como ellos aportan a la energía cinética. La energía cinética es la suma de todas las energías asociadas con los movimientos moleculares que pueden almacenar energía: translación, rotación y vibración, luego

En términos simple un grado de libertad es cada forma independiente en el cual se almacena la energía, por ejemplo

- Translación a través de un eje de coordenada
- Rotación sobre un eje de coordenada
- Oscilación sobre el eje de coordenada
- Energía potencial de un oscilador sobre un eje de coordenada

El llamado teorema de equipartición, ley de equipartición, equipartición de la energía o simplemente equipartición, indica que cada grado de libertad que aparece solo en forma cuadrática en la energía total del sistema, contribuye en promedio con $\frac{1}{2}kT$ unidades de energía cuando el sistema se encuentra en equilibrio y aporta con $\frac{1}{2}k$ a la capacidad calorífica del sistema.

De esta forma, y recordando que un átomo tiene 3 grados de libertad (los cuales le permiten absorber energía), correspondiente a los componentes x, y y z en coordenadas cartesianas, la energía cinética de un átomo queda definida como

$$E_{\text{kin}} = \frac{1}{2}m|\mathbf{v}|^2 = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)$$

y ya que los componentes del momentum aparecen en forma cuadrática en la energía cinética, cada átomo aporta en promedio $\frac{3}{2}kT$ a la energía cinética en equilibrio térmico.

En términos generales es posible definir que

- Un gas monoatómico ideal, solo tiene translación, por lo que su energía intera es For a monatomic ideal gas, there is only translation and the internal energy is: $E = \frac{3}{2}NkT = \frac{3}{2}nRT$ donde N es el número de partículas.
- Una molécula diatómica posee 3 grados de libertad traslacionales y 2 rotacionales, luego $E = \frac{5}{2}NkT$

Cuando los movimientos de las partículas pueden en términos traslacionales, por ejemplo un movimiento en la dimensión x del eje cartesiano, la definición de temperatura es

$$\frac{1}{2}kT = \left\langle \frac{1}{2}mv_x^2 \right\rangle$$

Lo anterior puede ser extendido a las diferentes dimensiones de un sistema, por ejemplo, en el espacio bidimensional se define la temperatura como ,

$$\left\langle \frac{1}{2}m|\vec{v}^2| \right\rangle = \left\langle \frac{1}{2}m(v_x^2 + v_y^2) \right\rangle = 2 \left\langle \frac{1}{2}mv_x^2 \right\rangle = kT$$

Por otro lado, para el espacio tridimensional, la temperatura esta definida por

$$\left\langle \frac{1}{2}m|\vec{v}^2| \right\rangle = \left\langle \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) \right\rangle = 3 \left\langle \frac{1}{2}mv_x^2 \right\rangle = \frac{3}{2}kT$$

Para todos los casos anteriores hemos utilizado la restricción $v_x = v_y = v_z$ para la partícula estudiada. Sin embargo, los grados de libertad de una partícula suelen ser de mayor complejidad.

Es posible establecer los grados de libertad para moléculas más complejas. Para moléculas lineales existen dos ejes de rotación y dos grados de libertad asociados con su rotación, en cambio, moléculas no lineales hay que considerar sus ejes y momentos de inercia. De esta forma, una molécula con N átomos tendrá $3N$ grados de libertad, distribuidos como

Cuadro 5: default

Tipo	Translación	Rotación	Vibración
Monoatómica	3	0	0
Linear	3	2	3n-5
No Lineal	3	3	3n -6

Termostatos

Cabe recordar que en las ecuaciones de MD se conserva la energía (colectivo micro-canónico NVE). Si queremos *muestrear* el colectivo NPT o NVT hay que modificar artificialmente las velocidades. Para esto hacemos uso de termostatos, los que mantienen la temperatura del sistema constante (no su energía).

La forma más simple de controlar las velocidades del sistema es utilizando factores de escalamiento cada cierto periodo de tiempo. De esta forma las velocidades de todos los átomos deben ser escalados para alcanzar la temperatura deseada satisfaciendo la relación básica

$$\left(\frac{\mathbf{v}_{nueva}}{\mathbf{v}_{actual}}\right)^2 = \left(\frac{T_{objetivo}}{T_{actual}}\right)$$

Un ejemplo de termóstato es el llamado “Termóstato de Berendsen”. En este se escalan las velocidades por un factor λ que satisface la relación

$$\lambda^2 = 1 + \frac{\Delta t}{\tau} \left(\frac{T_{requerida}}{T_{actual}} - 1 \right) \quad (32)$$

Donde τ es un parámetro acoplado a un baño externo.

En algunos casos para simular el efecto térmico de un entorno inexistente es útil usar una combinación de rozamiento sobre la velocidad (Langevin) y una fuerza estocástica o aleatoria (Browniana) (Dinámica de Langevin y Browniana. Las ecuaciones de movimiento en este caso son:

$$m_i \ddot{q}_i(t) = F_i - m_i \beta_i \dot{q}_i(t) + f_i(t)$$

Si a esta ecuación le añadimos el término de restricción

$$-m_i \Omega_i^2 [q_i(t) - q_i^{ref}]$$

tenemos el método Stochastic Boundary Molecular Dynamics (SBMD) que se usa habitualmente en QM/MM para simular un entorno inexistente.

El termóstato más popular es el de Nose-Hoover. Es un método con un Lagrangiano extendido (como en el método CPMD). Para mantener constante la presión (colectivo *e.g* NPT) se puede aplicar el mismo factor de escalaje o también con un Lagrangiano extendido.

Constricciones y restricciones

Si queremos limitar el movimiento de algún grado de libertad tenemos dos posibilidades

- Constricciones: Se modifican las ecuaciones de movimiento para congelar completamente el grado de libertad. El ejemplo por excelencia es el SHAKE. También se usa en algunos cálculos de energía libre y en Targeted MD.

- Restricciones: Se añade un potencial adicional, normalmente parabólico, que penaliza cuando el grado de libertad se aleja del valor requerido.

$$V_{tot} = V_0 + \frac{1}{2}k(q - q_{ref})^2$$

Se usa en Umbrella Sampling, ABF y cuando se quiere calentar un sistema poco a poco.

4.5. Campo de Fuerza, Topologías y parámetros

Como fue mencionado anteriormente, dinámica molecular se basa en la integración de las ecuaciones de movimiento del sistema definido por

$$\frac{d}{dt}(m\dot{\mathbf{r}}) + \frac{dV}{d\mathbf{r}} = 0$$

De ellas se desprende que es necesario describir la variación de la energía potencial del sistema, lo hacemos a través de los denominados ForceFields o campos de fuerzas, los cuales serán presentados a continuación.

En su forma más simple, una mezcla de partículas pueden ser modeladas por un potencial de interacción del tipo Lennard-Jones

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (33)$$

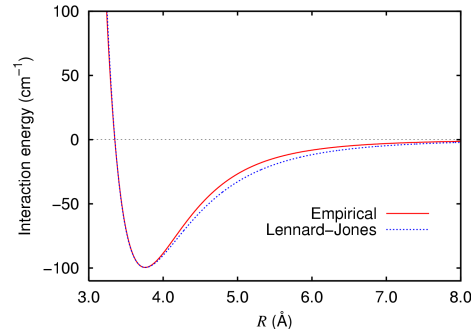
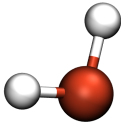


Figura 17: Potencial de Lennard-Jones de dos átomos de Argón

Donde ϵ y σ son los parámetros de LJ y r_{ij} es la distancia entre las partículas i y j . El potencial de LJ es, a pesar de su simplicidad, adecuado para capturar las propiedades físicas de fluidos.



Potencial de Lennard-Jones. La figura 17 muestra la energía de interacción de dos átomos de argón a medida que se acercan entre ellos. En ella se puede observar que el potencial de Lennard-Jones se ajusta a lo observado empíricamente, sin embargo, ¿A que corresponden los parámetros de Lennard-Jones, cómo se puede explicar la forma de la curva?

Por otro lado, los modelos moleculares de mayor complejidad utilizados en simulación molecular se basan, por lo general, en describir las moléculas como un conjunto de partículas de carga, masa y

volumen puntual, en cuyos centros se definen potenciales que dan cuenta de la energía de dispersión, repulsión y atracción atómica.

Bajo la descripción de la Mecánica Molecular, la energía potencial puede ser particionada de manera funcional en

$$\begin{aligned} E_{banded} &= E_{stretching} + E_{bending} + E_{torsion} + E_{improper} \\ E_{non-banded} &= E_{vanderWaals} + E_{electrostatic} \\ E_{total} &= E_{banded} + E_{non-banded} \end{aligned}$$

donde los términos enlazantes se describen como

$$\begin{aligned} E_{stretching} &= \sum_i^{bonds} k_{ri}(r - r_{eq})^2 \\ E_{bending} &= \sum_i^{angles} k_{\theta i}(\theta - \theta_{eq})^2 \\ E_{torsion} &= \sum_i^{dihedr} \frac{V_i}{2}[1 + \cos(n\phi - \gamma)] \\ E_{improper} &= \sum_i^{impropers} k_{\gamma i}(\gamma - \gamma_{eq})^2 \end{aligned}$$

y los términos no enlazantes

$$\begin{aligned} E_{vanderWaals} &= \sum_i^{atoms} \sum_{j>i}^{atoms} \epsilon_{ij} \left[\left(\frac{r_{min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{min_{ij}}}{r_{ij}} \right)^6 \right] w_{ij} \\ \text{con } r_{min} &= \sigma 2^{1/6} \\ &= \sum_i^{atoms} \sum_{j>i}^{atoms} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] w_{ij} \\ &= \sum_i^{atoms} \sum_{j>i}^{atoms} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] w_{ij} \end{aligned}$$

y la componente electrostatica como

$$E_{electrostatic} = \sum_i^{atoms} \sum_{j>i}^{atoms} \frac{q_i q_j}{r_{ij}} w_{ij}$$

donde el potencial $\frac{1}{r}$ decae muy lentamente y es fundamental en biomoléculas.

En este práctico hemos utilizado el programa **NAMD** para realizar el cálculo de las trayectorias de moléculas de agua. Este programa hace uso de una descripción de la energía potencial utilizando el campo de fuerza **CHARMM** (ec. 34), el cual permite el cálculo de las interacciones intra y intermoleculares con mayor detalle que un potencial de tipo LJ. La principal diferencia entre los distintos modelos de campos de fuerza existentes similares a **CHARMM**, como lo son **AMBER**, **GAFF**, **GROMOS** y otros, son los parámetros y forma de la función de energía potencial, los cuales son seleccionados para representar las propiedades de interés.

$$\begin{aligned} \mathcal{V}(\mathbf{r}) &= \sum k_r(r - r_0)^2 + \sum k_\theta(\theta - \theta_0)^2 + \sum B_n(1 + s \cos n\omega) \\ &\quad + \sum \frac{q_i q_j}{D r_{ij}} + \sum \epsilon \left[\left(\frac{r_{m_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{m_{ij}}}{r_{ij}} \right)^6 \right] \end{aligned} \tag{34}$$

Como se puede observar de las ecuaciones 34,34 y 34, es necesario conocer de manera correcta la forma en que los átomos se encuentran enlazados o interactuando entre ellos. De esta forma, bajo la nomenclatura del campo de fuerza CHARMM, un archivo de topología (RTF o top) define para cada residuo los grupos de átomos, tipos de átomos, enlaces y carga de ellos. La estructura del archivo de topología es bastante evidente, por lo que solo se mostrará el archivo correspondiente a una molécula de agua de tipo TIP3.

```
* Residue topology file for TIP3 water
*
31 1

MASS      4 HT      1.00800 H ! TIPS3P WATER HYDROGEN
MASS     75 OT     15.99940 O ! TIPS3P WATER OXYGEN

RESI TIP3      0.000 ! tip3p, generate using noangle nodihedral
GROUP
ATOM OH2  OT    -0.834
ATOM H1   HT     0.417
ATOM H2   HT     0.417
BOND OH2 H1 OH2 H2 H1 H2    ! the last bond is needed for shake
ANGLE H1 OH2 H2              ! required
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE

END
```

Un archivo de parámetros (PAR) permite determinar la energía asociada a una estructura a través de la definición de las constantes de fuerza de enlaces, ángulos y torsiones (dihedros) y parámetros de van der Waals. En CHARMM existen parámetros estándar para ácidos nucleicos, lípidos, carbohidratos y agua. Un ejemplo de un archivo de parámetros para simular el agua se muestra a continuación

```
* parameter file needed to simulate TIP3 water
*

BONDS
!atom type Kb          b0
HT  HT      0.000      1.5139 ! ALLOW WAT
                        ! FROM TIPS3P GEOMETRY (FOR SHAKE/W PARA
OT  HT     450.000      0.9572 ! ALLOW  WAT
                        ! FROM TIPS3P GEO

ANGLES
!atom types      Ktheta  Theta0  Kub      SO
HT  OT  HT      55.000   104.5200 ! ALLOW WAT
                        ! TIP3P GEOMETRY, ADM JR.

DIHEDRALS

IMPROPER

NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
          cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5

!atom ignored      epsilon  Rmin/2 ignored  eps,1-4  Rmin/2,1
OT      0.000000 -0.152100   1.768200 ! ALLOW  WAT
                        !TIP3P OXYGEN PARAMETERS, adm jr., NBFIX obsolete
HT      0.000000 -0.046000   0.224500 ! ALLOW WAT
                        !TIP3P HYDROGEN PARAMETERS, adm jr., NBFIX obsolete
Da      0.000000 -0.488067   1.51093
END
```

Cabe destacar que existe una convención en el nombre de las moléculas, los cuales deben coincidir con el archivo de topología correspondiente. Errores en la forma de nombrar las moléculas o átomos

resultarán en resultados incorrectos, por lo cual ambos archivos de topología y parámetros son entregados en conjunto y no es buena idea mezclar versiones ellos, sin embargo, la estructura de estos archivos permite la incorporación de nuevas topologías y parámetros.

Si observa con detención el archivo observará que para simular el agua no son necesarios parámetros para dihedros o impropios, ya que solo hay 3 átomos en este residuo. Algunos parámetros incluyen también otras correcciones, como por ejemplo parámetros CMAP en enlaces peptídicos.

4.6. Implementación de un programa de Dinámica Molecular

A partir de las diferentes secciones descritas podemos generalizar un esquema de un programa para realizar una simulación de dinámica molecular:

Algorithm 2 Dinámica Molecular

M pasos de integración

t tiempo origen

dt variación de tiempo

- 1: inicializar (estructura, velocidades)
 - 2: **for** $t = 0 \rightarrow M; t+dt$ **do**
 - 3: Calcular energía y gradiente (aplicación de función de energía potencial)
 - 4: Calcular nuevas posiciones (integrar)
 - 5: Aplicar constricciones a enlaces (shake)
 - 6: Describir las propiedades (q , T , E)
 - 7: Escalar velocidades
 - 8: Guardar información
 - 9: **end for**
-

De este programa ya hemos analizado los aspectos teóricos correspondiente a las líneas 3,4,5 y 8 los cuales son de fácil implementación computacional. A continuación ahondaremos en otros aspectos, que son indispensables para la ejecución de este algoritmo.

1: Inicialización

Un punto importante es la definición de las **coordenadas iniciales**, las cuales habitualmente se obtienen desde una simulación previa o a partir de una configuración tipo grilla cristalina o “lattice” para líquidos o de estructuras de rayos x para proteínas.

En el primer caso es útil partir con una configuración “fcc” que represente tanto la densidad adecuada y que posea una orientación aleatoria de los momentos dipolares. Cabe resaltar que esta configuración debe ser equilibrada hasta alcanzar la pérdida de la configuración inicial (melting). De manera similar, la configuración obtenida por rayos x debe ser sometida a un proceso de minimización energética, esto permite eliminar fuerzas de gran magnitud actuando sobre los átomos del sistema, lo cual emerge de la repulsión de van der Waals o por geometrías mal determinadas durante el proceso de obtención cristalográfica. Además, es posible agregar constricciones sobre todos o algunos átomos del sistema con el fin de prevenir cambios bruscos en la configuración inicial debido a estas fuerzas, las cuales disminuyen durante los procesos de minimización y equilibración.

En este programa las velocidades y posiciones de cada partícula son calculadas en cada paso, sin

embargo se requiere solucionar en el primer paso la siguiente ecuación

$$r_i(\delta t) = r_i(0) + v_i(0)\delta t + \frac{1}{2}F_i(0)\delta t^2/m \quad (35)$$

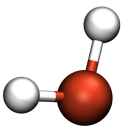
Para resolver la ecuación 35 es necesario obtener las **velocidades iniciales** $\mathbf{v}(0)$ del sistema, las cuales se obtienen a partir de un caso especial que relaciona la ecuación de Boltzmann para energías traslacionales ($E_i = \frac{1}{2}m_i v_i^2$)

$$P(v)dv = F(v)dv = \frac{v^2 e^{-mv^2/2kT} dv}{\int_0^\infty v^2 e^{-mv^2/2kT} dv} \quad (36)$$

donde $P(v)dv$ es la probabilidad de encontrar una molécula con velocidad v y dv , $F(v)dv$ la fracción de moléculas con velocidad v y dv , con moléculas de masa m y temperatura T . Cabe destacar que la velocidad corresponde a $\sqrt{v_x^2 + v_y^2 + v_z^2}$.

Donde al desarrollar la integral del denominador, la cual actúa como un normalizador, obtenemos la llamada ecuación de Maxwell-Boltzmann de las velocidades,

$$F(|\vec{v}|) = 4\pi v^2 f(\vec{v}) = 4\pi \left[\frac{m}{2\pi kT} \right]^{3/2} v^2 e^{-mv^2/2kT} \quad (37)$$



distribución de Maxwell-Boltzmann de las velocidades. Nos indica que al aumentar la temperatura (o disminuyendo la masa) el sistema tiende a una distribución más ancha de las velocidades, junto con un cambio hacia mayores velocidades de las partículas. Cabe destacar que esta distribución tiene una inclinación ("skewed distribution"), por lo cual su máximo está localizado en

$$\bar{v} = \sqrt{\frac{2kT}{m}}$$

la velocidad más probable no es lo mismo que la velocidad media

$$\langle v \rangle \equiv \int_0^\infty dv v f(|\vec{v}|) = \sqrt{\frac{8}{\pi}} \sqrt{\frac{kT}{m}}$$

y la raíz de la velocidad media (r.m.s.),

$$\bar{v}_{rms} \equiv \sqrt{\langle v^2 \rangle} = \sqrt{\frac{3kT}{m}}$$

Otro punto importante de observar es que los componentes de las velocidades tienen una distribución del tipo Gaussiana, por ejemplo para la componente x se cumple que

$$P(v_x)dv_x = F(v_x)dv_x = \left[\frac{m}{2\pi kT} \right]^{1/2} v^2 e^{-mv^2/2kT} dv_x$$

a partir de lo cual es posible calcular la velocidad promedio como

$$\langle v \rangle = \int_0^\infty P(v)dv = \left(\frac{8kT}{\pi m} \right)^{1/2} \quad (38)$$

$$\langle v^2 \rangle = \int_0^\infty P(v^2)dv = \frac{3kT}{m} \quad (39)$$

Recordando que para un sistema de N partículas sin constricciones internas la energía cinética del sistema corresponde a $K = \frac{1}{2} \sum_{i=1}^N m_i v_i^2$, la energía cinética del sistema entonces queda definida como

$$\langle K \rangle = \frac{3}{2} kT$$

cuando definimos la temperatura de un sistema. Si consideramos el principio de la equipartición de la energía, donde cada grado de libertad que aparece al cuadrado en el Hamiltoniano del sistema aporta con $\frac{kT}{2}$, para un sistema con $3N$ grados de libertad con N_c constricciones internas, el número total de grados de libertad correspondiente para el sistema es $N_f = 3N - N_c$, lo cual define la temperatura para un sistema con restricciones en sus grados de libertad como

$$\langle K \rangle = \frac{N_f kT}{2} \quad (40)$$

De esta forma, la ecuación de Maxwell-Boltzmann no tan solo nos permite inicializar las velocidades del sistema, sino que también nos permite calcular la temperatura de un sistema a partir de las energías cinéticas instantáneas o **Temperatura cinética instantánea** del sistema para cada partícula

$$T_{\text{instantanea}} = \frac{2K}{N_f k} = \frac{1}{N_f k} \sum_{i=1}^N m_i v_i^2 \quad (41)$$

Mientras que la Temperatura termodinámica, de naturaleza macroscópica, corresponde al promedio de la Temperatura cinética instantánea $\langle T_{\text{instantanea}} \rangle$.

Cabe recordar que las ecuaciones del movimiento que usamos son inestables e inducen al caos. Debido a esto, más que reproducir el sistema a tiempo real lo que se desea es sobretodo *muestrear* el espacio de fase. En este contexto, la hipótesis fundamental es que el promedio de una propiedad O a lo largo del tiempo de nuestra simulación corresponde a la propiedad promedio

$$\langle \hat{O} \rangle = \frac{1}{t_{\text{tot}}} \int_0^{t_{\text{tot}}} O(t) dt = \frac{1}{N} \sum_i O(t_i)$$

volveremos a ello más adelante.



Una simulación larga o muchas cortas. Las llamadas trayectorias de dinámica molecular se propagan hasta que las propiedades, por ejemplo la temperatura, se calculen hasta que estas sean estables. Un solo valor instantáneo no es **nunca** representativo, sólo el promedio de las propiedades. De esta forma, toda simulación necesita de los procesos de calentamiento y equilibración antes de calcular cualquier propiedad en equilibrio, estando ya en el ergódico. Una recomendación es realizar múltiples simulaciones de corta duración, cambiando las condiciones iniciales, que una sola simulación larga, ya que esto le permite explorar con mayor seguridad los distintos caminos disponibles en el espacio de fase. ver Karplus et al. Prot. Sci. 7, 649 (1998)

En términos prácticos, las velocidades iniciales se obtienen a partir de la generación de una distribución de Maxwell-Boltzmann de velocidades siguiendo el siguiente procedimiento

1. A partir de un generador de números aleatorios se debe generar una distribución normal de media cero y una varianza unitaria. Este set corresponde a $3N$ números que conforman esta a esta distribución).

2. Multiplicar por $\sqrt{\frac{kT}{m_i}}$ para obtener la distribución de velocidades correspondientes con la temperatura dada.
3. Corregir para que la suma de los momentos sea cero.

Cabe destacar que existen más de un número de distribuciones de las velocidades que cumplen las condiciones dadas, ya que estas dependen de la generación de números aleatorios, sin embargo, este “ensamble” de velocidades iniciales satisface la relación $\langle K \rangle = \frac{N_f kT}{2}$.

Por ejemplo, la tabla 6 muestra las coordenadas y velocidades de un sistema de 3 partículas obtenidos utilizando dinámica molecular.

Cuadro 6: Coordenadas en el plano xy de tres átomos de Argón en sus primeros tres tiempos de integración. el potencial utilizado corresponde a un potencial $4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$. Las velocidades están en unidades de \AA ps^{-1} . La masa del Argón es $39,95 \text{ g/mol}$, las unidades de fuerza corresponden a 10 J/mol \AA

	Partícula 1		Partícula 2		Partícula 3	
	x	y	x	x	x	y
Paso 0						
fuerza	55.47	55.47	-68.3	12.8	12.8	-68.3
Posicion	0	0	4	0	0	4
Velocidades	0	0	0	1	1	0
Paso 1						
fuerzas	55.56	55.56	-68.8	13.23	13.23	-68.8
posiciones	0.00174	0.00174	3.9979	0.0504	0.0504	3.9979
velocidades	0.0347		-0.0427	1.008	1.008	-0.0427
Paso 2						
fuerzas	54.3	54.3	-68.3	13.98	13.98	-68.3
posiciones	0.00695	0.00695	3.9914	0.1016	0.1016	3.9914
velocidades	0.0695	0.0695	-0.0057	0.0163	0.0163	-0.0057
Paso 3						
fuerzas	5134	5134	-66.5	15.14	15.14	-66.5
posiciones	0.0156	0.0156	3.9807	0.154	0.154	3.9807
velocidades	0.1383	0.1383	-0.172	1.0333	1.0333	-0.172

A Desarrolle las ecuaciones del movimiento para comprobar los datos de la tabla 6 y extiendala hasta el cuarto paso. Si encuentra que faltan datos complételos y compruebe el resultado propuesto.

2: Integración de las ecuaciones del movimiento

Una simulación de dinámica molecular permite observar la evolución temporal de un sistema a través de la integración de las ecuaciones del movimiento, habitualmente utilizando entre $10^3 - 10^6$ pasos de integración (“timesteps”), que en “tiempo real” corresponden a rangos temporales que van desde los femtosegundos a los nanosegundos (en contados casos en el rango de los microsegundos).

Considerando que la constante de auto difusión experimental a $273K$ y $0,1MPa$ para el Argón en estado líquido es $0,157 \text{ cm}^2 \text{ s}^{-1}$ y para el Agua en las mismas condiciones es de $0,276 \text{ cm}^2 \text{ s}^{-1}$, una simulación molecular permitirá observar con claridad el movimiento de estas partículas si la evolución o trayectoria de simulación alcanza el orden de los $0,5 - 1,0 \text{ ps}$ (picosegundos). Por otro

Cuadro 7: Escalas de movimiento				
tipo	scala	orden de tiempo	amplitud	descripción
corto	femto, pico	$10^{-15} - 10^{-12} s$	0.001 - 0.1 Å	<ul style="list-style-type: none"> ■ vibración de enlace ■ ángulo ■ movimiento dihedro
mediano	pico, nano	$10^{-12} - 10^{-9} s$	0.1 - 10 Å	<ul style="list-style-type: none"> ■ movimiento de cadenas laterales ■ movimiento de loops ■ movimientos colectivos
largos	nano, micro	$10^{-9} - 10^{-6} s$	1 - 100 Å	<ul style="list-style-type: none"> ■ plagamiento de péptidos ■ transición helice-coil
muy largo	micro, segundo	$10^{-6} - 10^{-1} s$	10 - 100 Å	plegamiento proteico

lado, la escala de los movimientos moleculares de sistemas de interés biológicos van más allá de las escalas de tiempo (ver tabla 7), por lo cual es necesario realizar simulaciones de mayor extensión, que permitan explorar con mayor probabilidad las propiedades de interés, por ejemplo, para ver el movimiento de una hélice de una proteína sería necesario realizar a lo menos 10^{-6} de pasos de integración, para observar este movimiento, si es que ocurre.

B Discuta la frase “para ver el movimiento de una hélice de una proteína sería necesario realizar a lo menos 10^{-6} de pasos de integración, para observar este movimiento, si es que ocurre”.

Algoritmos

A lo largo de la historia se han desarrollado diferentes métodos que permiten integrar las ecuaciones del movimiento de manera eficientes, en términos del uso de recursos computacionales, certera y precisa.

El poder de computo actual nos permite alcanzar cada vez mayores tiempos y escalas de simulación, sin embargo, esto habitualmente va en el tamaño y complejidad de los problemas en estudio, tanto en cantidad de átomos, metodologías empleadas y el tiempo para lograr la convergencia de los observables. Es por esto que se requiere una optimizar no tan solo los algoritmos para que sean eficientes y rápidos, sino que hay que buscar diferentes trucos que permitan avanzar en los estudios con los recursos de hardware disponibles.

Una de las primeras herramientas es modificar el valor de de tiempo asignado al paso de integración dt o timestep, sin embargo, un cambio en el valor de dt influye directamente en la precisión del cálculo de las posiciones de las partículas y, subsecuentemente, introduce a errores en los cálculos de las fuerzas y energías.

El valor del paso de integración limita la extensión de la trayectoria de dinámica molecular. Por ejemplo, al cambiar este paso 1 a $2fs$ el tiempo de simulación se duplica, reduciendo a la mitad el tiempo de computo. En cada paso de integración la operación computacional más “lenta” es el ciclo que recorrer todos los pares de partículas para calcular la energía potencial y las fuerzas del sistema (pairwise interaction). Cabe resaltar que el gasto computacional escala a N^2 , siendo N el número de partículas.

De esta forma, en un ensamble canónico la estabilidad numérica de los algoritmos se ve reflejada directamente en la conservación de la energía. Para obtenerla uno debe elegir un tiempo de integración que sea a lo menos un orden de magnitud menor que la magnitud del movimiento más rápido presente en el sistema, lo que en términos físicos esta relacionado con los periodo vibracionales presentes en el sistema y reflejados en la función de la energía potencial.

En la práctica, el valor del paso de integración esta limitado por la conservación de la energía. Habitualmente se optimiza realizando pequeñas simulaciones en un ensamble NVE, donde la energía del sistema en un comienzo fluctúa por imprecisiones numéricas, para luego converger a un rango de menor fluctuación, de al rededor de $1/5000$ partes de energía total del sistema cada 20 pasos de simulación.

Observables del sistema

Cabe recordar que en dinámica molecular las ecuaciones del movimiento son integradas avanzando en el tiempo ($t + dt$). Para caracterizar apropiadamente el sistema, este debe cumplir con las restricciones termodinámicas del ensamble elegido, lo cual se traduce en que el sistema debe alcanzar el equilibrio térmico antes de poder calcular las propiedades de interés.

Por ejemplo, para establecer si un sistema compuesto de gases nobles se encuentra en equilibrio es posible describir la energía cinética E_K , temperatura T y el desplazamiento promedio de las partículas (mean square displacement, MSD). Los cuales al ocupar unidades reducidas ($m = 1$) quedan descritas como

$$E_K = \frac{1}{2} \sum_i m_i |\mathbf{v}_i|^2 = \frac{1}{2} \sum_i |\mathbf{v}_i|^2 \quad (42)$$

$$\frac{3}{2}(N-1)k_B T = \frac{1}{2} \sum_i |\mathbf{v}_i|^2 = E_K \implies T = \frac{2}{3k_B(N-1)} E_K. \quad (43)$$

$$MSD = \frac{1}{N} \sum_i |\Delta \mathbf{r}_i|^2 = \frac{1}{N} \sum_i |\mathbf{r}_t - \mathbf{r}_0|^2. \quad (44)$$

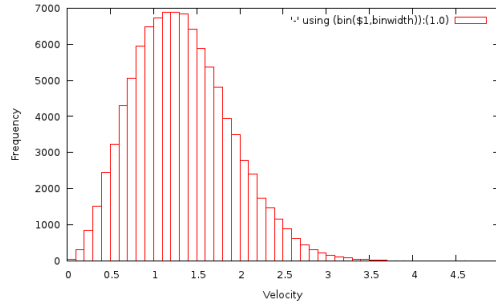


Figura 18: Distribución de la energía cinética en un sistema en equilibrio.

Por otro lado, cuando se estudian propiedades estructurales utilizando dinámica molecular es habitual utilizar criterios basados en energías, presión, volumen y temperatura del sistema para decidir si la simulación se encuentra en equilibrio. Por ejemplo, en el equilibrio termodinámico las velocidades de los átomos del sistema deben seguir una distribución del tipo Maxwell-Boltzmann (figura 18).

En mecánica estadística, el valor promedio de un observable A esta dado por el promedio de observaciones del conjunto de conformaciones

$$\langle A \rangle_{\text{conjunto}} = \int \int d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) \quad (45)$$

donde $A(\mathbf{p}^N, \mathbf{r}^N)$ es el observable de interés en términos de los momentos y posiciones del sistema. La densidad de probabilidad del sistema esta dada por

$$\rho(\mathbf{p}^N, \mathbf{r}^N) = \frac{1}{Q} \exp[-\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)/k_B T] \quad (46)$$

con \mathcal{H} el Hamiltoniano del sistema, T la temperatura, k_B la constante de Boltzmann y Q la función de partición definida como

$$Q = \int \int d\mathbf{p}^N d\mathbf{r}^N \exp[-\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)/k_B T] \quad (47)$$

Esta integral es, generalmente, difícil o imposible de calcular ya que deben conocerse todos los estados posibles del sistema. En dinámica molecular los puntos en el colectivo son calculados de manera secuencial en el tiempo, luego, para calcular el promedio del colectivo la simulación debe pasar por todos los estados posibles a partir de las restricciones termodinámicas propias o impuestas al sistema.

Otra forma de calcular una propiedad es determinar el promedio temporal de del observable A , el cual se puede expresar como

$$\langle A \rangle_{\text{tiempo}} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\mathbf{p}^N, \mathbf{r}^N) dt \quad (48)$$

que para términos prácticos puede describirse como

$$\langle A \rangle_{\text{tiempo}} \cong \frac{1}{M} \sum_{t=1}^M A(\mathbf{p}^N, \mathbf{r}^N) \quad (49)$$

donde M es el número de configuraciones exploradas. De esta manera se calculan promedios temporales, pero los observables experimentales corresponden a promedios del colectivo. Para resolver este problema se hace uso de un axioma fundamental de la mecánica estadística, la *hipotesis del ergódico*, la cual establece que el promedio temporal es igual al promedio de colectivo

$$\langle A \rangle_{\text{tiempo}} = \langle A \rangle_{\text{colectivo}} \quad (50)$$

La idea básica de este axioma es que si el sistema evoluciona en un tiempo infinito, este explorará eventualmente todos los estados posibles. El objetivo de una simulación molecular es generar una cantidad suficiente de conformaciones representativas que permitan cumplir esta igualdad, calculando a partir de ellas las propiedades estructurales, dinámicas y termodinámicas utilizando un tiempo de cálculo razonable con los recursos computacionales disponibles. Esto se traduce en recolectar en un tiempo finito la mayor cantidad de muestras del espacio de fase (momentos y coordenadas).

Esto significa que debemos analizar si nuestra simulación alcanzó o no un equilibrio del observable estudiado antes de confiar en el valor del promedio temporal. La función de autocorrelación temporal del observable A , $\langle a(t_0)a(t_0 + t) \rangle$ permite relacionar valores que se encuentran separados por un tiempo t . Esta función es independiente del punto de inicio y puede escribirse $\langle a(0)a(t) \rangle$. Esta función decae desde un valor $\langle a(0)a(0) \rangle = \langle a^2 \rangle$ hasta el infinito como

$$\lim_{t \rightarrow \infty} \langle a(0)a(t) \rangle \equiv \lim_{t \rightarrow \infty} \langle a(0) \rangle \langle a(t) \rangle = 0 \quad (51)$$

mientras las variables $a(0)$ y $a(t)$ no se encuentren correlacionadas. El decaimiento ocurre en un tiempo característico τ_a definido como

$$\tau_a = \int_0^{\infty} dt \langle a(0)a(t) \rangle / \langle a^2 \rangle \quad (52)$$

que puede ser aproximada como $\langle a(0)a(t) \rangle \propto \exp(-t/\tau_a)$. En la práctica se traduce en generar la curva $y = \exp(-x/a_0)$, donde y es la función de autocorrelación, x el tiempo y a_0 es el tiempo de autocorrelación τ .

C A partir de las líneas de código entregadas programe la función de autocorrelación.

```

funcion Autocorrelacion {x} {
! x representa una serie temporal
    mean = Mean(x)
    autocorrelation = x.length/2
    for (t = 0; t < autocorrelation.length; t++) {
        n = 0; // Numerador
        d = 0; // Denominador
        for (i = 0; i < x.length; i++) {
            xim = x[i] - mean
            n += xim * (x[(i + t) \% x.length] - mean)
            d += xim * xim;
        }
        autocorrelation[t] = n / d
    }
    return autocorrelation
}

```

D Genere series que estén correlacionadas de forma lineal ($a_{i+d} = f(a_i)$) y no correlacionadas entre ellas, calcule con su función el tiempo de autocorrelación y describa la relación de este con el valor d .

Finalmente, el correcto análisis de los observables nos permite la cantidad de pasos de integración a realizar, cantidad que depende tanto de la escala de tiempo en que ocurren los fenómenos a observar y el tiempo necesario para que el valor estimado de un observable o propiedad haya convergido.

En resumen y en general, podemos decir que dinámica molecular permite la medición de propiedades macroscópicas a través de la determinación del promedio de una propiedad obtenida de nuestra simulación. La hipótesis ergódica establece que el tiempo en que un sistema reside en un estado A es proporcional al volumen del espacio de fase en que A se encuentra en ese valor. Esto aplica en sistemas que se encuentran en equilibrio durante un largo periodo de tiempo, lo cual tiene como resultado que el promedio y el promedio del ensamble de una variable son iguales. Si promediamos por un tiempo largo de tiempo, es decir una gran cantidad de microestados, y nuestro modelo de interacción es correcto – nuestro campo de fuerza – podemos utilizar estos modelos para predecir propiedades de sistemas reales que se encuentran en equilibrio. Aun más, podemos utilizar estas simulaciones para obtener información estadísticas de propiedades de un sistema compuesto por múltiples partículas y de la forma en como fluctúan estas bajo distintas condiciones de simulación.

E Discuta como la integración de las ecuaciones del movimiento para obtener las configuraciones de un sistema puede generar errores en la determinación estadística de un observable.

7: Escalar Velocidades

En las secciones anteriores se nombraron diferentes termostatos que pueden ser utilizados para ajustar la temperatura del sistema, sin embargo, la implementación más simple corresponde al rescalamiento de velocidades para mantener la energía cinética constante (simulación isocinética). Para esto se debe medir la temperatura instantánea T_i en un tiempo i del sistema luego de actualizar las velocidades, de tal manera que si escalamos estas velocidades por una constante λ debe cumplirse que

$$\mathbf{v}_{objetivo} = \lambda \mathbf{v}_i = \sqrt{(T_{objetivo}/T_i)} \mathbf{v}_{actual} \quad (53)$$

Este tipo de termostato es útil para calentar o en las etapas de inicialización de una simulación. Cabe destacar que se puede realizar el rescalamiento de las velocidades en cada paso de integración o solo en alguno de ellos.

- F** Desarrolle el pseudocódigo de un programa que permita integrar las ecuaciones de movimiento y controlar la temperatura a través del rescalamiento de velocidades y el termostato de Berendsen (ecuación 32).
- G** Discuta el por qué el rescalamiento de velocidades no es apropiado para realizar simulaciones en el ensamble canónico y si el termostato de Berendsen.

8: Guardar la información

Un punto importante es guardar los datos generados por una simulación para realizar el procesamiento de estos y obtener la información.

Para guardar las coordenadas podemos utilizar diferentes tipos de formato de archivo. En particular un formato general es el formato *xyz* el cual es delimitado por espacios; coordenadas en angstrom) y tiene la siguiente estructura

```

línea 1: <numero de átomos, N>
línea 2: línea de comentario
línea 3-N+2: elemento atómico, coordenada x, y & z

```

A su vez, este formato posibilita guardar una trayectoria, para ello solamente se debe incorporar una nueva configuración tras otra

```

3
0
C 0.000000    0.000000    0.000000
O 0.000000    0.000000    1.159076
O 0.000000    0.000000   -1.159076
3
1
C 0.000000    0.000000    0.000000
O 0.000000    0.000000    1.200000
O 0.000000    0.000000   -1.159076

```

De esta forma, para generar un archivo *xyz* solo se requiere una estructura como la presentada a continuación

```

for (t=0; t<n; t++)
  print $nAtom
  print $t
  for (n=0; n<nAtom; n++)
    print "Ar $rx[n][t] $ry[n][t] $rz[n][t]"
  }
}

```



Formato xyz. Este formato es reconocido por VMD. A su vez, puede ser utilizado para almacenar cualquier información que pueda ser proyectada en el espacio cartesiano, como la velocidad, fuerzas sobre átomos, potenciales, entre otras. <http://www.ks.uiuc.edu/Research/vmd/plugins/molfile/xyzplugin.html>

Ejemplos

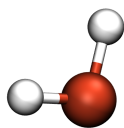
5. Argón: Un sistema tipo Lennar-Jones

Este es el sistema de menor complejidad que podemos estudiar. Al ser el Argón un gas noble, es posible modelar el comportamiento de N partículas solo en base a interacciones no enlazantes por medio de un Potencial de pares de tipo Lennard-Jones, como el descrito por Rahman en 1964 y posteriormente por Levesque, Verlet y K urkijarvi (Phys. Rev. A 7, 1690 (1973)). Ellos describieron con  xito el comportamiento de un l quido LJ cercano a su punto tripe utilizando un sistema compuesto por 864  tomos de Arg n en un ensamble NVE a $94,4K$, densidad $\rho = 1,374 \text{kd}/m^3$, un potencial LJ de $\frac{\epsilon}{k_b} = 119,8K$ y $\sigma = 3,405$. Considere una masa molar $M = 39,948$.

En este pr ctico Ud. describir  el sistema en su punto tripe. Como condici n inicial ud. dispone de un archivo PDB `argon864.pdb` que contiene las posiciones de 864  tomos en una estructura ordenada (fcc), los cuales est n en una caja c bica de lado $L = 34,739 \text{  }$.

Al igual Rahman, ud. deber  utilizar  la funci n de distribuci n radial, la distribuci n de la velocidades del sistema y el movimiento de las part culas para corroborar los resultados de din mica molecular.

A Reproduzca la investigaci n realizada por Rahman en el a o 1964



Correlation in the motions of Liquid Arg n. Rahman 1964. Este trabajo se realiz  utilizando el computador CDC-3600, el demor  45s para calcular cada paso de integraci n. Alcanzando un total de 2000 pasos. Aneesur Rahman es considerado el padre de la din mica molecular. Physical Review, vol. 136, Issue 2A, pp. 405-411

El programa de din mica Molecular

En las secciones anteriores se han visto los aspectos de mayor relevancia de un programa de din mica molecular, con los cuales Ud. deber  generar un programa que permita caracterizar utilizando din mica molecular el sistema de Arg n anteriormente descrito.

B Implemente un programa cuyos requisitos son:

- Leer un archivo en formato pdb o xyz.
- Describir el sistema utilizando un potencial de Lennard-Jones.
- Integrar el movimiento en el espacio Cartesiano xyz utilizando el algoritmo de Verlet.
- Incluir condiciones peri dicas de borde.
- Utilizar la convenci n de la m nima imagen.
- Entregar un archivo de salida en formato xyz con las coordenadas at micas.
- Entregar un archivo con las velocidades at micas.
- Explorar el espacio de fase en los ensambles can nico y microcan nico.
- Utilizar el m todo de rescalamiento de velocidades y otro tipo de termostato.

Como ayuda, a continuaci n encontrar  diferentes c digos desarrollados en el lenguaje DaV-Code, desde el cual Ud. puede extraer las estructuras b sicas para generar su propio sistema de din mica molecular.

```

Function IniMomentaAndVelocities(n,Temp)
! Reduced Units
!-----

! Generate random momenta
do i = 1,n
! assign x component of momenta with random numbers between 0 and 1
    px0(i) = random01()
end do

! calculate total momenta
do i = 1,n
    pxsum += px0(i)
end do
!divide by number of particles
    pxsum /= n
! Set total momentum to zero
do i = 1,n
    px0(i) -= pxsum
end do
do i = 1,n
    kineticTemp += px0(i)*px0(i)
end do
! reduced units , i.e., m = k = 1,
! m is the atomic mass of particles, and k is Boltzmann's constant.
! Review dimensions?

scale = sqrt(3*n*Temp/kineticTemp)
do i = 1,n
    px0(i) *= scale
end do
end

```

```

Function Forces(x,fx,n,Epot)
! Reduced Units
!-----
!
epsilon = 1.0
sigma = 1.25
sigma2 = sigma * sigma
! Explain these lines...
Rcut = 2.0
Rcut2 = Rcut * Rcut

do i = 1,n
    fx(i) = 0.0
end do
Epot = 0.0
do i = 1,n-1
    do j = i+1,n
        dx = x(i) - x(j)
        r2 = dx*dx
        if (r2 < Rcut2) then
            fr2 = sigma2 / r2
            fr6 = fr2 * fr2 * fr2
            fpr = 48.d0 * epsilon * fr6*(fr6-0.5)/r2
            fxi = fpr * dx

            fx(i) = fx(i) + fxi
            fx(j) = fx(j) - fxi

            Epot = Epot + 4.0 * epsilon * fr6 * (fr6 - 1.0)
        end if
    end do
end do
end

```

```

Function Verlet(amass,x,vx,ax,n,dt,Ekin,Epot)
!-----
dt2 = 0.5 * dt
do i = 1,n
    vx(i) = vx(i) + ax(i) * dt
    x(i) = x(i) + vx(i) * dt
end do

```

```

Forces(x,ax,n,Epot)

Ekin = 0.0
do i = 1,n
    ax(i) = ax(i) / amass
    vx(i) = vx(i) + ax(i) * dt2
    Ekin = Ekin + vx(iatm)**2
end do
Ekin = 0.5 * amass * Ekin
end

```

```

Function BoundaryConditions(x,n,dcell)
!-----
do i = 1,n
    if (x(i) < 0) then
        x(i) = x(i) + dcell
    else if (x(i) > dcell) then
        x(i) = x(i) - dcell
    end if
end do
end

```

La ciencia

C Además, debe discutir o informar lo siguiente:

- Demostrar que la velocidad de las partículas evoluciona a una distribución de Maxwell-Boltzmann independiente de las condiciones iniciales. (Ayuda: utilice histogramas)
- Ajustar el tiempo de simulación de acuerdo a la evolución del sistema hacia una distribución de Maxwell-Boltzmann.
- Demostrar si existe o no una dependencia entre el valor de dt (tiempo de integración, timestep), la energía cinética, potencial y sus fluctuaciones. (Ayuda: Equipartición de la energía).
- Realizar un experimento que le permita definir un valor adecuado de tiempo de integración.
- Caracterizar las fluctuaciones en la temperatura y su relación con el tiempo de integrador y tamaño del sistema.
- Caracterizar el efecto de la temperatura sobre la estructura del Argón.

6. Agua: un modelo de mayor complejidad

En esta unidad utilizaremos VMD y NAMD para realizar una simulación de dinámica molecular de una caja de agua, la cual servirá en el resto de este práctico para explorar de las capacidades de VMD y NAMD para estudiar las propiedades estáticas y dinámicas que permitan describir distintos sistemas.

A primera vista una molécula de agua parece ser simple: dos átomos de hidrógenos unidos a un átomo de oxígeno. Cada átomo de hidrogeno tiene carga $+1$ y el átomo de oxígeno $-2e$. La diferencia de electronegatividad de los átomos presentes generan dipolos al interior de la molécula, modificando tanto la distribución de cargas como la geometría de la molécula. Para una molécula aislada la longitud del enlace O-H es de 0.9584 \AA y el ángulo H-O-H es de 104.45° , mientras que en estado gaseoso son 0.95718 \AA y 104.474° . En estado líquido, utilizando estudios de difracción de neutrones, se ha reportado una distancia de enlace de 0.95718 \AA y 104.474° , respectivamente.

El modelo de agua TIP3P (transferable intermolecular potential with 3 points), implementado en el campo de fuerza de CHARMM, representa el agua a partir de las propiedades de 3 sitios con

cargas, dispuestos a una distancia rígida y parámetros de LJ para cada átomo (figura 19). Este es similar al modelo TIP3P desarrollado por Jorgensen y colaboradores, ⁴ el cual esta compuesto de un átomo de oxígeno unido a dos hidrógenos en una molécula que tiene sus grados de libertad fijos.

Este modelo de agua puede ser representado por un potencial de la siguiente forma

$$E_{ab} = \sum_i^n \sum_j^n \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^6}$$

donde $k_C = 332,1 \text{ kcal/mol}$ es una constante dieléctrica y los parámetros de LJ correspondiente son

Cuadro 8: default

r_{OH}	$\angle HOH$	σ	ϵ (kJ/mol)	$q(O)$ (e)	$q(H)$ (e)	$q(M)$ (e)	r_{OM}
0.9572	104.52	3.15066	0.63627	-0.834	0.417	0	0

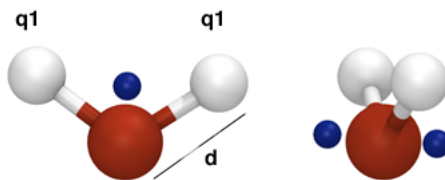


Figura 19: Modelos de Agua. *Izq.* Modelo de tres sitios. Los átomos de oxígeno e hidrógeno se encuentran unidos a través de enlaces flexibles o rígidos. Las cargas positivas se localizan en el núcleo de los átomos de hidrógeno, mientras que las cargas negativas se localizan en la bisectriz del ángulo H-O-H (esfera azul). *Der.* Modelo de cuatro sitios, donde las cargas negativas se encuentran localizadas sobre y bajo el núcleo del átomo de oxígeno (partícula M en la tabla 6) .

6.1. Coordenadas Iniciales del sistema

- 1 En VMD inicie la consola Tk. Para esto seleccione el menú **Extensions** \rightarrow **Tk Console**. Ingrese al directorio de trabajo. Si utiliza MS Windos, use cremillas para navegar a la dirección, i.e., `cd "C:\Documents and Settings"`, se recomienda colocar el directorio de trabajo en la raíz del sistema `cd "C:"` o en un directorio libre de espacios.
- 2 Para crear la caja utilizamos el plugin **solvate** de VMD ⁵. El plugin se inicia escribiendo en la consola el comando

```
<Main>0% package require solvate
1.5
```

Si se ha iniciado correctamente, aparecerá la versión del mismo en pantalla, en nuestro caso 1.5. Este plugin nos permite agregar un solvente a cualquier soluto.

⁴Jorgensen et al. 1983, J. Chem. Phys. 79: 926; Neria et al 1996, J. Chem. Phys. 105: 1902

⁵<http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/>,

3 Los distintos parámetros e instrucciones para su uso se visualizan ejecutando el comando `solvate`.

```
<Main>1% solvate
Info) Usage: solvate <psffile> <pdbfile> <option1?> <option2?>...
Info) Usage: solvate <option1?> <option2?>... to just create a water box
Info) Options:
Info) -o <output prefix> (data will be written to output.psf/output.pdb)
Info) -s <segid prefix> (should be either one or two letters; default WT)
Info) -b <boundary> (minimum distance between water and solute, default 2.4)
Info) -minmax xmin ymin zmin xmax ymax zmax
Info) -rotate (rotate molecule to minimize water volume)
Info) -rotsel <selection> (selection of atoms to check for rotation)
Info) -rotinc <increment> (degree increment for rotation)
Info) -t <pad in all directions> (override with any of the following)
Info) -x <pad negative x>
Info) -y <pad negative y>
Info) -z <pad negative z>
Info) +x <pad positive x>
Info) +y <pad positive y>
Info) +z <pad positive z>
Info) The following options allow the use of solvent other than water:
Info) -spsf <solventpsf> (PSF file for nonstandard solvent)
Info) -spdb <solventpdb> (PDB file for nonstandard solvent)
Info) -stop <solventtop> (Topology file for nonstandard solvent)
Info) -ws <size> (Box length for nonstandard solvent)
Info) -ks <keyatom> (Atom occuring once per residue for nonstandard solvent)
```

4 Escriba el comando

```
<Main>2% solvate -o BOX -minmax {-20 -10 -10} {20 10 10}
```

el que generará una caja de agua, en ausencia de un soluto, de dimensiones 40x20x20 Å. Si el plugin se ha ejecutado correctamente debería encontrar dos archivos en el directorio de trabajo: `BOX.psf` y `BOX.pdb`. El archivo `BOX.pdb` es el archivo que contiene las posiciones en el espacio cartesiano, mientras el archivo `BOX.psf` contiene la conectividad del sistema.

La caja de solvente se genera a partir de la duplicación de sistemas de menor tamaño, cortando cada dimensión para alcanzar el tamaño deseado. Es necesario verificar si el tamaño alcanzado es el adecuado, como también si la densidad del sistema corresponde con las descritas de manera experimental. VMD cuenta con una función de medición por el comando `measure`, el cuál permite cuantificar, entre otros, largo de enlaces, ángulos, dihedros, centro y tamaño del sistema, radio de giro, etc. A continuación utilizaremos esta herramienta para medir el tamaño del sistema.

5 Seleccione utilizando la columna “T” en la ventana `Main` el sistema a medir, si desea esconder un sistema lo puede hacer seleccionando la columna “D”.

6 Ejecute el comando `measure minmax` sobre una selección de átomos

```
<Main>3% set Aguas [atomselect top all]
<Main>4% set medidas [measure minmax $Aguas]
{-20.4420 -10.6625 -10.5232} {20.5997 10.6650 10.3884}
```

el comando `measure minmax` entrega una lista, donde cada elemento son las coordenadas cartesianas mínimas y máximas entre las partículas presentes en el sistema.

- A** Calcule la densidad del sistema.
- B** Describa utilizando las herramientas programadas por Ud. y las de VMD el sistema obtenido.
- 7** Lea detenidamente el manual de usuario de VMD, en él encontrará mayor información de distintos comandos que no han sido mencionados en este documento y de como utilizar este programa para distintas tareas a través de sus módulos y comandos.⁶
- C** Describa y utilice al menos 2 herramientas que no se hayan descrito en este práctico que permitan describir el sistema en estudio. Ayuda: por ejemplo cantidad de puentes de hidrógeno.
- D** Compare sus resultados con los datos disponibles de las siguientes y otras investigaciones.
- Reviews in Computational Chemistry 13 183 (1999).
 - W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926-935 (1983).
 - W. L. Jorgensen and J. D. Madura. Temperature and size dependence for Monte Carlo simulations of TIP4P water. Mol. Phys. 56, 1381-1392 (1985).
 - W. L. Jorgensen and C. Jenson. Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density. J. Comput. Chem. 19, 1179-1186 (1998).
 - M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J. Chem. Phys. 112, 8910-8922 (2000).

6.2. Uso básico de NAMD

6.2.1. Archivo de configuración básico

En el directorio `practico-difusion/DM` se encuentran los archivos necesarios para realizar las simulaciones de dinámica molecular programa NAMD que aparecen en este práctico, tanto en un ensamble canónico (NVT) como un ensamble microcanónico (NVE). A continuación serán detallados los parámetros más importantes.

Los archivos de configuración para NAMD tienen una estructura básica⁷. En este ejemplo visualizaremos los archivos `practico-difusion/DM/ejemplo_nve.namd` y `practico-difusion/DM/ejemplo_nvt.namd` utilizado para generar las trayectorias de las cajas de agua que Ud. acaba de analizar.

- 1** En un editor de texto (por ejemplo notepad en Windows o TextEdit en Linux) abra el archivo `practico-difusion/DM/ejemplo_nve.namd`.
- La primera sección de este archivo referencia los archivos de estructura, conectividad y parámetros (en nuestro caso del campo de fuerza charmm) a utilizar.

```
# input
coordinates      BOX.pdb
structure        BOX.psf
parameters       par_all27_prot_lipid_na.inp
paratypecharmm   on
```

⁶www.ks.uiuc.edu/Research/vmd/current/ug/.

⁷Para mayor información se recomienda leer el manual del usuario y los tutoriales disponibles en

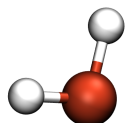
- Para controlar el nombre de los archivos de salida hemos utilizado una variable usando el comando Tcl “set output”, para que en este caso los archivos de salida tengan el nombre base BOX. Por otro lado, el parámetro “dcdfreq” permite controlar cada cuantos pasos de integración guardaremos las configuraciones del sistema que en conjunto generan una trayectoria.

```
# output
set output          BOX
outputname          $output
dcdfile             ${output}.dcd
xstFile             ${output}.xst
dcdfreq             50
xstFreq             50

outputEnergies 1    ;cada cuanto tiempos de integración
outputPressure 1    ; almaceno las energías y presión del sistema.
```

- En esta simulación utilizamos las capacidades de NAMD para realizar múltiples pasos de integración para el cálculo de las fuerzas de enlace e interacciones de corto y largo alcance.

```
# Multiple timestepping
firsttimestep       0
timestep            1      ;# femtosegundo
stepspcycle         20     ;
nonbondedFreq       2      ;#calc. efectivo timestep x nbf
fullElectFrequency  4      ;#calc. efectivo timestep x fEF
;# calc. efectivo timestep x fullElectFrequency
```



Grados de libertad e Integrador. En un sistema LJ no existen grados de libertad asociados interacciones intramoleculares, las ecuaciones del movimiento pueden integrarse utilizando mayores tiempos de integración, por ejemplo cada $\Delta t = 10$ fs. ¿Por qué debe ser menor cuando existen interacciones intramoleculares?

- Para reducir el tiempo de cálculo y controlar el volumen de la celda durante la simulación utilizaremos condiciones periódicas de borde (sec. 3.2.4). Para este caso, el tamaño de la celda unitaria viene dada por el tamaño de la caja de agua creada en la unidad 4, la cual se encuentra centrada en el origen. El comando `wrapAll` especifica como se almacenarán las coordenadas de los átomos que atraviesan los límites de la caja periódica.

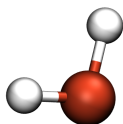
```
# PBC
cellBasisVector1    40.00 0.00 0.00
cellBasisVector2    0.00 20.00 0.00
cellBasisVector3    0.00 0.00 20.00
cellOrigin          0.00 0.00 0.00
wrapAll             on
dcdUnitCell         yes
```

- Durante una simulación en un conjunto NVE se busca mantener la Energía del sistema constante, para lo cual es necesario asignar la energía inicial cinética del sistema correspondiente a una temperatura específica.

```
# Temperature control
set temperature      298
reinitvels           $temperature ;# reiniciar las temperaturas
```

- Finalmente, si es necesario se puede realizar una minimización de la energía potencial del sistema, por ejemplo utilizando el método de gradientes conjugados, para luego, realizar una simulación de dinámica molecular utilizando el algoritmo de Verlet por el tiempo deseado, por ejemplo 1 ns.

```
minimize 1000          ;#gradienates conjugados
run          1000000    ;# timesteps de dinámica molecular
                        ;# equivalente a numsteps
```



Dinámica de Langevin. Utilizar dinámica de Langevin es una manera de controlar la energía cinética del sistema, lo que permite controlar la temperatura o presión del sistema. El método hace uso de la ecuación de Langevin para partícula única:

$$m_i \frac{dx_i(t)}{dt^2} = F_i\{x_i(t)\} - \gamma_i \frac{dx_i(t)}{dt} m_i + R_i(t) \quad (54)$$

El primer termino de la derecha representa la fuerza de la partícula, mientras que el segundo representa un factor de amortiguación de fricción aplicado a la partícula con un factor de fricción igual a γ_i . El tercer termino corresponde a fuerzas aleatorias que pueden ser aplicadas sobre la partícula como resultado de la interacción con el solvente. Estos términos son usados para mantener la energía cinética del sistema para estabilizar la temperatura a un valor constante.

6.2.2. Control de presión y temperatura

- Por otro lado, en un ambiente NVT se busca mantener la temperatura durante la simulación, para lo cual es posible utilizar un termostato adecuado.

Por ejemplo, es posible utilizar el método de rescalamiento de las velocidades cada cierta cantidad de pasos de integración.

```
rescaleFreq 20          ;# numero de pasos de integración entre rescalamiento
rescaleTemp 290         ;# temperatura final
```

o a través de la incorporación de un termostato de Langevin.

```
# Temperature control
set temperature      298
temperature          $temperature ;# initial temperature

# Langevin Dynamics
langevin             on ;# langevin dynamics
langevinDamping      1 ;# damping coefficient (gamma) of 1/ps
langevinTemp         $temperature ;# temperatura del baño
langevinHydrogen     no ;# no efecta hidrógenos,solo a los átomos pesados
```

- Para realizar simulaciones en ambiente isobárico-isotérmico (NPT) se puede utilizar el método de piston de Nosé-Hoover Langevin piston, en un ambiente semi-isotrópico o anisotrópico según sea el caso.

```
# Constant Pressure Control
langevinPiston       on
langevinPistonTarget 1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod 100.    ;# oscillation period around 100 fs
langevinPistonDecay  50.     ;# oscillation decay time of 50 fs
langevinPistonTemp   $temperature ;# coupled to heat bath

# Volume control
useGroupPressure     yes ;# needed for rigid bonds
```

```

useFlexibleCell      no  ;# no for water box, yes for membrane
useConstantArea     no  ;#Permitir área x-y constante,
                      ;# mientras varia la dimensión z
                      ;# no for water box, maybe for membrane

```

6.2.3. Disminuyendo el tiempo de cálculo

- Podemos hacer uso de constricciones a los diferentes grados de libertad, por ejemplo, restringiendo el largo de los enlaces a su valor de equilibrio, por otro lado, el uso de un integrador más extenso nos permite aumentar el tiempo de la simulación disminuyendo el tiempo de calculo necesario.

```

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1    ;# nonbonded forces every step
fullElectFrequency  2    ;# PME only every other step

```

- Adicionalmente, el cálculo de las propiedades electrostáticas es de alta demanda computacional, NAMD permite disminuir el tiempo de cálculo utilizando el método de *Particle mesh Ewald* (PME) el cuál es de mayor precisión y de menor costo que utilizar cutoff de mayor tamaño, sin embargo, solo es aplicable a sistemas periódicos.

```

#PME (for full-system periodic electrostatics)
;#PME (Particle Mesh Ewald) interacciones
;#electrostáticas bajo condiciones periódicas de borde;
PME                yes
# let NAMD determine grid
PMEGridSpacing     1.0

```

Estos no son todos los parámetros configurables en una simulación con el programa NAMD, el resto de ellos se encuentran detallados en el manual de usuario del programa, por ejemplo el control de las funciones para interacciones no enlazantes.

```

exclude scaled1-4    ;# Las interacciones 1-3 no se calculan
1-4scaling 1.0       ;# Las interacciones 1-4 no se escalan
switching on         ;# atenuación de interacciones Van Der Waals
switchdist 10        ;# y Electrostáticas a cierta distancia del cutoff.
cutoff 12.0          ;# distancia de exclusion de calculo de interacciones
pairlistdist 14.0     ;# distancia total para calcular todo tipo de interacciones

```

Archivo	utilidad	comando
pdb	archivo de coordenadas	coordinates
psf	archivo de conectividad	structure
par	archivo de parámetros	parameters
top	archivo de topología	Utilizado en psfgen
coor	archivo de coordenadas finales	outputname
dcd	archivo de trayectoria	dcdfile
out	archivo de registro de ejecución y energías	salida.log

Cuadro 9: Resumen de los archivos de entrada y salidas de NAMD

Ejecución del programa

- 2 Para ejecutar el programa debe abrir un ventana de terminal. En windows ejecute el programa “cmd”. En ambiente Unix abra un consola de Terminal. Luego acceda al directorio donde se encuentran los archivos necesarios.

- Windows:

```
cd CARPETA_TRABAJO
C:\CARPETA_DONDE_ESTA_NAMD\bin\namd2.exe archivo.namd >salida.log
```

- Unix/Linux:

```
cd CARPETA_TRABAJO
/CARPETA_DONDE_ESTA_NAMD/namd2 archivo.namd >salida.log
```



Variables de ambiente. Es recomendable colocar la ubicación del programa NAMD en las variables de ambiente del sistema

- Windows: path = %PATH%;C:\CARPETA_DONDE_ESTA_NAMD
- Unix/Linux: export PATH = \$PATH:/CARPETA_DONDE_ESTA_NAMD

- 3 Una vez terminada la simulación se obtendrá un archivo de registro **salida.log** el cuál contiene los parámetros utilizados y valores como la energía potencial, temperatura, volumen, etc. Tras una correcta ejecución del programa ud. debe obtener líneas finales como las siguientes

=====

```
WallClock: 613.973755  CPUTime: 613.973755  Memory: 2428.781250 MB
Program finished.
```

- 4 El resultado final es un archivo **log** que contiene las energías, temperatura, presión y volumen del sistema. Un archivo **dcd** con las coordenadas de cada marco de simulación seleccionado para ser almacenado utilizando los valores de **dcdFreq**. Cabe destacar que la trayectoria representa un valor temporal de **[timestep * numstep]** femtosegundos (fs) y que cada cuadro representa las coordenadas almacenadas cada **[timestep * dcdFreq]** fs. Otros archivos son generados tras la correcta ejecución del programa, los cuales se encuentran detallados en la tabla 9, que resume los archivos de entrada y salidas necesarios para realizar una simulación.

Análisis del archivo de salida

En esta sección, para ilustrar el análisis, estudiaremos la evolución temporal de la energía potencial y temperatura del sistema durante la trayectoria de ejemplo. Para esto utilizaremos el archivo de salida **solvate.log** obtenido en la sección 6.2.2, el cual será analizado utilizando las herramientas incorporadas en VMD.

- 5 En VMD seleccione el menú **Extensions → Analysis → NAMD Plot**. Al hacerlo se abrirá la ventana NAMD Plot
- 6 Seleccione el menú **File → Select NAMD log File**, se abrirá una ventana que le permitirá navegar hasta el archivo **solvate.log**
- 7 Para seleccionar que parámetro graficar complete la casilla correspondiente, en nuestro caso seleccione **TOTAL**
- 8 Seleccione el menú **File → Plot Selected Data**, se abrirá una ventana con un gráfico de la evolución temporal del parámetro seleccionado.

- 9 Los datos pueden ser extraídos para ser utilizados en otras aplicaciones, para ello en la ventana Multiplot seleccione el menú File → Export to ..., a modo de ejemplo, se muestran los datos en formato ASCII que se encuentran en el archivo especificado

```
0 205899.3369
1 22423.9753
2 1887.7740
3 -1395.1914
```

- E Realice un gráfico de la evolución de la energía cinética y de la temperatura del sistema.
- F ¿Cuales son los valores promedios para los distintos parámetros? es apropiado considerar los valores desde el comienzo de la simulación? Determine el periodo de tiempo en que el sistema ha encontrado el equilibrio termodinámico.

6.3. Agua líquida

En 1971 Rahman y Stillinger en su investigación “Molecular Dynamics Study of Liquid Water” realizaron la primera simulación molecular “realística” del agua. Ellos utilizaron un sistema compuesto por 216 moléculas de agua rígidas a una densidad de 1gcm^{-3} . Integrando las ecuaciones de movimiento y utilizando un potencial de pares para representar la interacción entre múltiples partículas, lograron contrastar el comportamiento cinético a una temperatura de $307,45\text{K}$ con diferentes datos experimentales.

En este práctico Ud. describirá un sistema de aguas en estado líquido. Al igual Rahman, ud. deberá utilizar la función de distribución radial, la distribución de la velocidades del sistema y el movimiento de las partículas para corroborar los resultados de dinámica molecular.

- G Reproduzca la investigación realizada por Rahman en el año 1970. Para ello utilice NAMD y VMD y el modelo de agua TIP3.

utilizando un ensamble NVE

- Demuestre que la velocidad de las partículas evoluciona a una distribución de Maxwell-Boltzmann independiente de las condiciones iniciales.
- Estudie el efecto de la temperatura en la estructura del agua y la constante de difusión.
- Corroborar que una simulación de largo duración converge al resultado de un conjunto de simulaciones cortas.
- Estudie el efecto del tamaño de la caja periódica sobre diferentes observables.

utilizando un ensamble NVT

- Demuestre que la velocidad de las partículas evoluciona a una distribución de Maxwell-Boltzmann utilizando rescalamiento de velocidades y el termostato de Langevin.
- Explique el por que el método de rescalamiento de velocidades no permite muestrear en el ensamble microcanónico, mientras que el termostato de Langevin si lo permite.
- Estudie el efecto del termostato utilizado en la estructura del agua y la constante de difusión.
- Estudie el efecto de utilizar un valor de `timestep 2.0 ;# 2fs/step` y `rigidBonds all`

Conclusiones Finales

Este práctico introduce al uso básico de VMD en NAMD para realizar y analizar simulaciones moleculares. Se ha recorrido el proceso de realizar una simulación de dinámica molecular desde la observación de fenómenos estáticos a fenómenos dinámicos que son comparables con datos experimentales. Herramientas de programación combinadas con el conocimiento físico y químico permiten entender fenómenos de importancia biológica, actuando como un microscopio molecular.