

# Química Computacional

## Taller 3

Gian Pietro Miscione y Sebastián Franco Ulloa

Módulo de Mecánica Molecular

Mayo 9, 2017

QUIM-3516

Este taller tiene como objetivo familiarizar al estudiante con la técnica computacional de dinámicas moleculares (MD). Para esto haremos una simulación del complejo entre una ADN girasa y su ligando nativo presente en uno de los cristales ya reportados. Esto se hará usando GROMACS, un programa gratuito para hacer este tipo de simulaciones. El orden de trabajo se ilustra en la figura 1. Los resultados de estas simulaciones se visualizarán con el programa, también gratuito, VMD (*Visual Molecular Dynamics*).

### 1. Parametrización del ligando

La simulación la haremos con el campo de fuerza AMBER99SB-ILDN, un campo de fuerza típico en la simulación de proteínas. Aunque GROMACS puede asignar los *atom types* necesarios a los átomos de una proteína, debemos crear *atom types* para nuestro ligando.

- Descargue del repositorio del taller (<https://github.com/cebasfu93/QuimicaComputacional/tree/master/Taller3>) el archivo *4plb\_wild\_fixed.pdb*. Este archivo es muy parecido al que se descargó en el Taller 2 (código PDB: 4PLB).
- Abra el archivo descargado en Maestro y elimine la proteína. Salve un archivo *31N.pdb* donde sólo esté el ligando nativo del cristal. Para esto: *Click Derecho en la entrada* → *Export* → *Structures...* En *Files* seleccione “Export all entries to the same file”. En *Files of type* seleccione “By Extension (\*.\*)”. Guárdelo bajo el nombre de *31N.pdb*.
- Verifique el paso anterior abriendo en Maestro el archivo *31N.pdb*. Asegúrese de que sólo se haya guardado el ligando.
- Ahora necesitamos calcular las cargas de los átomos usando un método QM. Para esto primero necesitamos un archivo *.com*. Para esto, abra con *gaussview* el archivo que acaba de salvar.
- Verifique que la geometría esté bien (e.g. que los hidrógenos estén bien colocados).
- Salve el compuesto en formato *.com*.
- Abra el archivo *.com* en un editor de texto (e.g. *gedit*) y edítelo para que el encabezado se vea de la siguiente manera. OJO: Las líneas en blanco se deben dejar a propósito. También, borre la conectividad en caso de haber sido escrita y asegúrese de dejar una línea en blanco al final del documento.

```
%chk=31N.chk
%mem=5000MB
%nproc=8
#P HF/6-31G* SCF=Tight Pop=(MK) IOP(6/33=2)
```

```
31N_SCF
```

```
1 1
```

```
C 42.51000 32.42800 28.66000
```

- Lo anterior en un método compatible con los campos de fuerza AMBER para calcular cargas parciales de átomos.
- Para correr este cálculo necesitamos una máquina que tenga instalado *Gaussian*, el programa por excelencia para hacer cálculos QM.
- Conéctese a la dirección “estudiantes@157.253.60.12”
- Cree una carpeta con su nombre usando el comando *mkdir*.
- Usando el comando *scp* lleve el archivo *31N.com* a su carpeta.
- Vaya a la carpeta que creó y ejecute el siguiente comando:

```
g09 < 31N.com > 31N.log
```

- El archivo *31N.log* tiene, en algún lugar, las cargas parciales de los átomos de nuestro ligando. Para extraerlas usaremos herramientas de AMBER (la empresa, no el campo de fuerza). Estas herramientas están instaladas en otra máquina.
- Conéctese usando el comando *ssh* al usuario “guest@157.253.72.13”
- Cree una carpeta con su nombre (si todavía no la hay) y lleve el archivo *31N.log* a esa carpeta. De ahora en adelante trabajaremos en esta carpeta.
- Para extraer las cargas y la geometría necesaria usamos la herramienta *antechamber* que se puede descargar de manera gratuita de la página de AMBER:

```
antechamber -i res.log -fi gout -o res.mol2 -fo mol2 -c resp -nc # -m #
```

- Las opciones *-nc* y *-m* piden la carga y la multiplicidad del sistema respectivamente, en este caso ambas son 1.
- El resultado de este cálculo es un archivo *31N.mol2* que contiene las coordenadas y las cargas parciales de todos los átomos.
- Por defecto, *antechamber* nombra al residuo “MOL”. Es preferible ponerle a nuestro ligando un nombre que lo distinga. Para esto, usamos el comando *sed*, el cual reemplaza un patrón de caracteres por otro:

```
sed s'/MOL/31N/' 31N.mol2 > 31N_2.mol2
```

- El resultado de este comando es un archivo *31N\_2.mol2* que dice “31N” donde antes decía “MOL”. Revise este archivo con un editor de texto y arréglole. ¿Qué toca arreglar?
- Cambie el archivo de nombre de “31N\_2.mol2” a “31N.mol2”
- Ahora debemos calcular los demás parámetros. Para esto primero revisamos si sí se pueden calcular:

```
parmchk -i 31N.mol2 -f mol2 -o 31N.frcmod -a y
```

- El resultado del cálculo anterior es un archivo *31N.frcmod* que muestra todos los parámetros que puede asignar. Si no logra asignar algún parámetro lo va a poner en 0,000 y al lado derecho dirá “ATTN: Needs Revision”.

- Ya que sabemos que sí se pueden calcular, vamos a usar *tleap* para asignarlos y ponerlos en los archivos correctos. Para esto vaya al repositorio del curso y descargue los archivos “*leapLig.in*” y “*leapComplex.in*”.
- Los archivos *\*.in* contienen la información necesaria para que *tleap* haga lo que queremos que haga. Entienda cada línea del archivo *leapLig.in* y ejecútelo de la siguiente manera:

```
tleap -sf leapLig.in
```

- El archivo *leapLig.in* lo que hace es importar los parámetros y asignarlos a una geometría. El resultado de ese cálculo es un archivo *leap.log* que tiene la información de lo que hizo *tleap* y otros 3 archivos: *31N.prmtop* que tiene la geometría del ligando en formato AMBER (el programa, no la empresa), *31N.inpcrd* que tiene la conectividad del ligando en formato AMBER y *31N.lib* que tiene los parámetros del ligando en formato AMBER.
- Ahora debemos unir los parámetros del receptor y del ligando. Para esto vamos a usar otra vez *tleap*, esta vez con el archivo *leapComplex.in*.
- Además necesita descargar del repositorio del taller el archivo *complex.pdb*.
- El archivo *complex.pdb* es el mismo archivo *4plb\_wild\_fixed.pdb* pero alineando el compuesto de *31N.mol2* con el ligando nativo y eliminando el ligando nativo. Esto se puede hacer con la herramienta *Superposition* de Maestro. Esto es necesario ya que en los pasos anteriores generamos *atom types* a átomos con nombres específicos que salen del archivo *31N.mol2*. Si se usara *4plb\_wild\_fixed.pdb* se obtendría un error diciendo que no se encontraron *atom types* para todos los átomos. Por tiempo, no vamos a generar el archivo *complex.pdb* en el taller.
- Además del ligando, existe otro átomo para el cual no se tienen parámetros en el campo de fuerza predeterminado de AMBER99SB-ILDN: el magnesio que tiene la ADN girasa como cofactor. Estos parámetros se pueden encontrar en internet. Una copia de los archivos con los parámetros se encuentran en el repositorio del taller: “MG.frmod” y “MG.prep”
- Ahora puede abrir el archivo *leapComplex.in* y entender qué hace cada línea.
- Corra el archivo similar a como se hizo la vez pasada:

```
tleap -sf leapComplex.in
```

- Este cálculo debe arrojar dos archivos: *sys.prmtop* y *sys.inpcrd*. El primero tiene la coordenadas del sistema en formato AMBER y el segundo tiene la topología (conectividad y parámetros) de todos los átomos también en formato de AMBER.
- Ya casi estamos listos para empezar la simulación, solo falta convertir los archivos de coordenadas y topología a un formato reconocible por GROMACS. Para esto se usa un *script* disponible en internet llamado *acpye* el cual ya está en la máquina en la que se está trabajando. Para hacer la conversión solo se debe correr el siguiente comando:

```
acpye -p sys.prmtop -x sys.inpcrd -r
```

- El resultado del cálculo anterior son dos archivos con la misma información anterior pero en formato de GROMACS: *sys\_GMX.gro* (conectividad) y *sys\_GMX.top* (topología).

## 2. Creación de la caja

A partir de ahora trabajaremos con GROMACS pero en la misma carpeta que antes. Una vez con la topología y las coordenadas se puede proceder a definir la región del espacio que se simulará. Note que entre más grande sea esta región más se le va a exigir a los procesadores durante la simulación pues habrán más moléculas de disolvente.

- Para definir la región en el espacio a simular (la caja), se usa el siguiente comando

```
gmx_mpi editconf -f sys_.gro -o sys_BOX.gro -d 1.5
```

- La opción `-f` establece el archivo de coordenadas con el que se trabajará. La opción `-o` especifica el nombre del archivo con la información de la caja. La opción `-d` forza a los átomos del sistema a encontrarse al menos  $1,5nm$  ( $15\text{\AA}$ ) de la caja. Por defecto se usa una caja ortorómbica (ángulos iguales, lados distintos) pero se puede especificar otro tipo de caja.
- Al final del archivo `*.gro` que se generó se deberían encontrar las dimensiones de la caja.

## 3. Adición de agua

Ahora podemos agregar moléculas de agua al espacio en el que no se encuentre la proteína. Las moléculas de agua se agregan dependiendo del modelo de solvatación que se desee usar.

- Descargar del repositorio del taller los archivos `tip3p.itp` y `ions.itp`.
- Ejecutar el siguiente comando para solvatar el sistema:

```
gmx_mpi solvate -cp sys_BOX.gro -cs spc216.gro -p sys_GMX.top -o sys_SOLV.gro
```

- La opción `-cp` especifica el archivo de coordenadas a usar (el que se generó en el paso anterior). La opción `-cs` especifica el modelo de solvatación a utilizar, “`spc216.gro`” significa el modelo *TIP3P*. La opción `-p` especifica el archivo de topología del sistema y `-o` el nombre del archivo de salida. Los archivos de coordenadas y de topología ahora deberían tener las coordenadas y conectividad de todas las moléculas de agua que se agregaron.
- Note que al haber agregado átomos (oxígenos e hidrógenos), tenemos una topología incompatible con el archivo de coordenadas pues la topología no tiene parámetros para estos nuevos átomos. Para hacerlos compatibles debemos abrir el archivo de topología con `vi` y agregar los parámetros pertinentes. Dos modificaciones deben ser hechas:
  1. Al comienzo del archivo hay una sección llamada `[ atomtypes ]`. Al final de la sección se deben agregar las cargas y parámetros de van der Waals para los nuevos `atom types` (OW y HW). Agregue las siguientes líneas:

```
OW OW 16.0000 -0.8476 A 3.15061e-01 6.36272e-01
HW HW 1.0000 0.4238 A 0.00000e-01 0.00000e-01
```

2. Al final del archivo hay una sección llamada `[ system ]`. Antes de que esta sección inicie se debe agregar una línea que llame al archivo con los parámetros de nuestro modelo de solvatación. Escriba entonces la siguiente línea:

```
#include “./tip3p.itp”
```

## 4. Adición de iones

Aunque el sistema ya está solvatado, aún debemos agregar iones para simular correctamente las condiciones fisiológicas y neutralizar el sistema, el cual, como ya debe saber en este momento, tiene una carga de casi  $-60e$ . Es importante mencionar que a partir de este momento son necesarios dos pasos para hacer prácticamente cualquier cosa: Unir las coordenadas, la topología y las especificaciones en un único archivo, y luego ejecutar ese archivo.

- Primero uniremos las coordenadas, topología y parámetros del cálculo con el comando *grompp*:

```
gmx_mpi grompp -f minim.mdp -c sys_SOLV.gro -p sys_GMX.top -o sys_ION.tpr
```

- El archivo *\*.mdp* es un archivo de texto que contiene todas las especificaciones técnicas de lo que se va a hacer, en este caso, como solo se van a agregar iones, no se va a simular entonces no es importante. Ese archivo se puede descargar del repositorio del taller. Al igual que los demás comandos de GROMACS, éste requiere de un archivo de coordenadas y de topología.
- El resultado de este cálculo es un archivo *\*.tpr* fácil de manipular pero que no se puede abrir como un archivo de texto (es un binario).
- Para agregar los átomos solo hace falta ejecutar el archivo que acabamos de generar. Para esto se debe correr el siguiente comando:

```
gmx_mpi genion -s sys_ION.tpr -o sys_ION.gro -p sys_GMX.top -pname So -neutral
```

- Después de la ejecución preguntará qué moléculas se desean reemplazar por iones. Selecciona el solvente.
- El resultado de este cálculo es un archivo *\*.gro* en el que algunas moléculas de agua fueron reemplazadas por átomos de tipo “So”. “So” quiere decir sodio, la razón por la que no se usa “Na” es porque puede haber un *atom type* (algún nitrógeno) que ya tenga este nombre (puede ocurrir lo mismo con “So” pero es menos probable, de todas maneras siempre es mejor asegurarse).
- Al igual que en la sección anterior, ahora es necesario agregar los parámetros de los átomos que se agregaron al sistema, nuevamente se deben hacer dos cosas:

1. En la sección de *[ atomtypes ]* agregar la siguiente línea.

So So 23.0000 1.0000 A 3.32840e-01 1.15897e-02
--

2. Antes de la sección *[ system ]* agregar la siguiente línea.

#include “./ions.itp”
-----------------------

## 5. Minimización

Ahora sí tenemos el sistema que deseamos simular. Lo primero que debemos hacer es una minimización, para asegurarnos de que la simulación no empiece en un punto energéticamente desfavorable.

- Nuevamente usamos *grompp* para pegar los archivos pertinentes de la siguiente manera:

```
gmx_mpi grompp -f minim-solvated.mdp -c sys_ION.gro -p sys_GMX.top -o sys_MIN.tpr
```

- Esta vez el archivo *\*.mdp* sí es importante pues nos especifica aspectos como el número máximo de iteraciones, el método con el que se minimizará el sistema, etc.

- Ahora usamos el archivo *\*.tpr* que acabamos de generar para minimizar el sistema:

```
gmx_mpi mdrun -v -gpu_id 0 -deffnm sys_MIN
```

- La herramienta *mdrun* es la encargada de hacer minimizaciones, calentamientos, equilibraciones y producciones. La opción *gpu\_id* especifica la *gpu* en la que se correrá el cálculo. Para asegurarse de que nadie más esté usando esa *gpu* puede correr el comando *nvidia-smi* y ver el porcentaje de *gpu* usado actualmente. Si está en uso, puede cambiar el número 0 por 1, 2 o 3 (la máquina de trabajo solo tiene 4 *gpu*'s). La opción *-deffnm* busca un archivo *\*.tpr* con el nombre que se dió y todos los archivos de *output* se escriben con ese nombre pero diferentes extensiones.
- El resultado de este cálculo es un archivo *\*.trr* que es un formato de trayectoria, tiene todas las geometrías que se evaluaron durante la minimización (es un binario). También se obtiene un archivo *\*.gro* que tiene el último marco de la minimización, es decir, el sistema ya minimizado. Se obtiene un archivo *\*.log* con la información de lo que hizo GROMACS durante el cálculo y un archivo *\*.edr* que tiene información como energías y propiedades fisicoquímicas durante la minimización.
- Si se desea extraer información sobre la energía, se debe correr el siguiente comando:

```
gmx_mpi energy -s sys_MIN.tpr -f sys_MIN.edr -o sys_MIN.xvg
```

Este comando necesita de un binario *\*.tpr* y un archivo de energías *\*.edr*. El resultado del cálculo es un archivo de texto de tipo *\*.xvg* donde cada columna corresponde a una de las propiedades que se le pidió al comando imprimir tras su ejecución. Estos datos se pueden graficar con programas como *xmgrace* y *gnuplot*, o se puede escribir un *script* que lo haga en el lenguaje de programación de su elección.

- Se recomienda imprimir y graficar la energía potencial del sistema para verificar que efectivamente la energía convergió a un mínimo.

## 6. Generación de restricciones

Una buena práctica durante el calentamiento y la equilibración es fijar en el espacio algunos átomos. Esto lo hacemos para no permitir que el sistema cambie demasiado durante estas etapas y poder extraer información útil durante la producción. La idea detrás de una restricción es que se le debe decir a GROMACS que un cierto conjunto de átomos les aplique una constante de fuerza considerablemente alta, para que no se muevan demasiado.

- Primero creamos un grupo de átomos. Para esto usamos el siguiente comando:

```
gmx_mpi make_ndx -f sys_MIN.gro -o index.ndx
```

- Al ejecutar el comando, se va a solicitar al usuario que define el grupo de interés. Suponiendo que el “backbone” y el “DNA” corresponden a los grupos “4” y “11”, respectivamente, para crear un grupo con ambos conjuntos de átomos se necesita indicar “4 || 11”. Cree un grupo que contenga el “backbone”, “DNA” y “31N”. Cree otro grupo con todo nuestro sistema de interés: “Protein”, “DNA”, “31N” y “MG” y otro con la parte del sistema que no es de nuestro interés: “Water” y “So”. Asegúrese que el número de átomos en los últimos dos grupos sumen los átomos de todo el sistema.
- El resultado de este cálculo es un archivo de índices de tipo *\*.ndx*. Este archivo es de texto y está dividido en diferentes secciones. Cada sección corresponde a un grupo de átomos, entre ellos, los que definimos.
- Ahora, para crear las restricciones sobre el primer grupo que definimos se usa el siguiente comando:

```
gmx_mpi genrestr -f sys_MIN.gro -n index.ndx -o restraints.itp
```

- Al ejecutar este comando se va a solicitar al usuario que especifique el grupo al que se le desean generar las restricciones. Seleccione el primero que definimos.
- El resultado de este cálculo es un archivo de texto *\*.itp* que contiene una lista con los átomos a restringir y tres columnas indicando las constantes de fuerza en cada dirección para cada átomo. Note que estas son restricciones flexibles pues no se fija la posición de los átomos, solo se les pone una fuerza muy grande en todas las direcciones.
- Para hacer vigentes las restricciones que acabamos de crear es necesario agregarlas a la topología. Para esto, abra con *vi* el archivo de topología y antes de la sección *[ system ]*, agregue la siguiente línea:

```
#include "restraints.itp"
```

## 7. Calentamiento

Ya que tenemos el sistema minimizado y parte de él restringido, podemos empezar a calentar el sistema.

- Descargue del repositorio del curso el archivo *nvt.mdp*, el cual contienen las especificaciones técnicas del calentamiento (termostato a usar, tiempo de calentamiento, constante de acople con el termostato, etc.).
- Para asegurarnos de que el calentamiento ocurra lo más uniformemente posible y que la energía que se escriba sea representativa de nuestro sistema de interés, debemos dividir nuestro sistema en dos: interés y no-interés.
- Abra con *vi* el archivo *nvt.mdp* y busque las opciones *energygrps* y *tc-grps*. Cambie los nombres que están después de estas palabras claves a los nombres de los 2 últimos grupos que se definieron en la sección anterior.
- Ahora unimos todos los archivos pertinentes para hacer el calentamiento:

```
gmx_mpi grompp -f nvt.mdp -c sys_MIN.gro -p sys_GMX.top -n index.ndx -o sys_NVT.tpr
```

- Al final de este cálculo se debe obtener un único archivo *\*.tpr*.
- Solo falta ejecutar el archivo que acabamos de generar:

```
gmx_mpi mdrun -v -gpu_id 0 -deffnm sys_NVT
```

- El resultado de este cálculo, al igual que en la minimización son archivos *\*.trr*, *\*.gro*, *\*.edr* y *\*.log*. En este caso, el primero de los archivos tiene la trayectoria de los átomos durante el calentamiento y el segundo archivo tiene la geometría del último paso del calentamiento. Además, esta vez se debió escribir un archivo de tipo *\*.cpt*, el cual tiene la misma función que los *checkpoints* de GAUSSIAN. Almacena toda la información del cálculo periódicamente para que, si cae por alguna razón, se pueda retomar el cálculo donde se había quedado.

## 8. Equilibración

Ya que el sistema está calentado, podríamos empezar a tomar datos a partir de este momento, peor debido a que el sistema puede seguir buscando una conformación estable a esa temperatura, es necesario darle un poco de tiempo para que se equilibre.

- Descargue del repositorio del taller el archivo *npt.mdp* y hágale las mismas modificaciones que se le hizo a *nvt.mdp* en la sección anterior.

- Nuevamente empezamos por unir todos los archivos pertinentes:

```
gmx_mpi grompp -f npt.mdp -c sys_NVT.gro -t sys_NVT.cpt -p sys_GMX.top -n index.ndx -o sys_NPT.tpr
```

- Y procedemos a correrlo:

```
gmx_mpi mdrun -v -gpu id 0 -deffnm sys_NPT
```

- Los archivos que se deben obtener tras este cálculo deben ser los mismos formatos a los que se obtuvieron en la sección anterior.

## 9. Producción

Al fin vamos a llevar a cabo la producción de la dinámica molecular, de donde realmente sacaremos la información de nuestro sistema. Lo primero que se debe hacer es QUITAR LAS RESTRICCIONES!!! Para esto elimine la línea que se escribió en la topología que las incluía.

- Descargue del repositorio del taller el archivo *production.mdp* y haga las mismas modificaciones que se les hizo a los archivos *\*.mdp* en las secciones anteriores.
- Unimos todos los archivos de interés:

```
gmx_mpi grompp -f production.mdp -c sys_NPT.gro -t sys_NPT.cpt -p sys_GMX.top -n index.ndx -o sys_MD.tpr
```

- Ya que este es el paso más demorado, no tiene sentido esperar a que termine frente al computador. En cambio, es preferible dejar el cálculo corriendo en el fondo de la máquina. Para esto se usa el siguiente comando:

```
nohup gmx_mpi mdrun -v -gpu_id 0 -deffnm sys_MD &
```

- El símbolo *et* (&) al final del comando hace que se corra en el fondo de la máquina. El comando *nohup* hace que todo lo que debió decirnos el comando en la terminal, se escriba en un archivo (*nohup.out*).
- Los resultados de esta simulación son los siguientes archivos: *\*.xtc*, un archivo con la trayectoria de la simulación, equivalente al *\*.trr* de las secciones pasadas. *.gro* un archivo con las coordenadas del último marco de la simulación. *.edr*, *\*.log* y *\*.cpt*, archivos con la información de la energía, avances de GROMACS y *checkpoints*, respectivamente.

## 10. Visualización

Una vez terminamos la simulación de dinámica molecular, estamos listos para observar la trayectoria de la producción y analizar lo que más se pueda de ella. Desafortunadamente, por la forma en la que opera GROMACS, los archivos de trayectorias no muestran lo que queremos ver, se ve muy mal debido a la forma en la que trata las condiciones de frontera periódicas. Antes de la visualización debemos arreglar los archivos de trayectoria.

- Supongamos que se va a ver la trayectoria de la producción (aunque esto aplica para cualquier trayectoria en cualquier formato de GROMACS (*\*.xtc* o *\*.trr*):

```
gmx_mpi trjconv -f sys_MD.xtc -o sys_MD.FIX.xtc -s sys_MD.tpr -pbc mol -ur compact
```

- Al ejecutar el comando, la terminal va a solicitar al usuario el grupo cuya trayectoria se desea arreglar. Lo mejor es seleccionar todo el sistema (grupo 0).



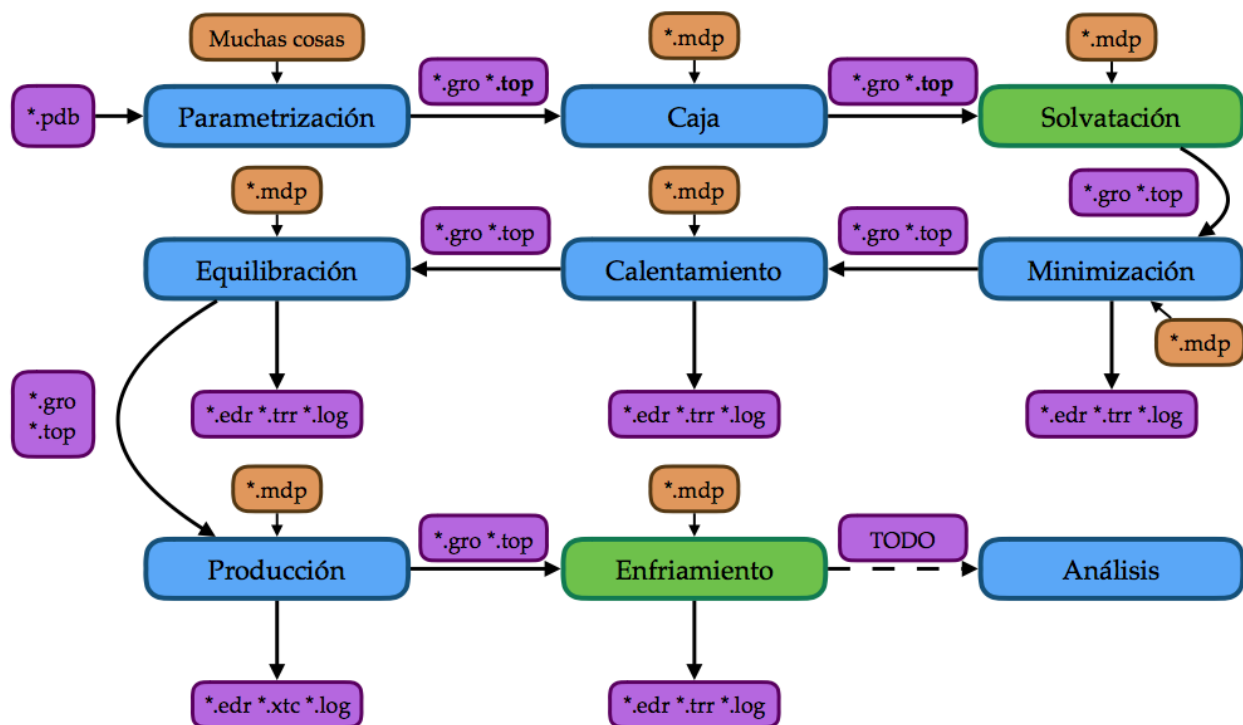


Figura 1: Orden lógico de una simulación de dinámica molecular.

- El resultado del cálculo anterior es un único archivo llamado *sys\_MD\_FIX.xtc* que tiene la misma información que la trayectoria original pero ahora sí se puede analizar más fácilmente.
- Para abrir una trayectoria debe copiar la carpeta en la que se trabajó a un computador que tenga instalado *VMD* y que tenga suficiente memoria RAM como para abrir las trayectorias (cuánta depende de la extensión de la trayectoria).
- Abra *VMD*.
- Abra la última estructura (*\*.gro*) del paso anterior a la dinámica que se desea estudiar. Por ejemplo, si desea ver la producción (*sys\_MD\_FIX.xtc*), entonces debe abrir el archivo *sys\_NPT.gro*, el cual tiene la geometría del último paso de la equilibración. Para abrir esta geometría:

*File* → *New Molecule...* → *Browse* → *\*Seleccione el archivo .gro\** → *Load*

- Ya debería poder ver todo el sistema en la interfaz gráfica de *VMD*.
- Para abrir la trayectoria:

*File* → *New Molecule...* → *Load files for* → *\*Nombre del archivo .gro que se acaba de importar\** → *Filename* → *\*Seleccione el archivo .xtc\** → *Load*

- Debería poder ver cómo se carga, un marco a la vez, la trayectoria de interés.

## 11. Análisis de trayectorias

Esto es otra historia completa.