

Simulación con Dinámica Molecular: Minimización de energía proteína en agua

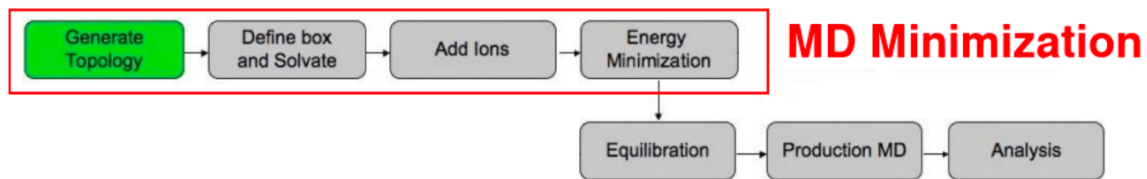
Taller

Selectiva 1

Linea de énfasis Bioinformática

Sergio Andrés Castañeda Garzón

Proceso:



1. Descarga del archivo PDB.

Se realiza la descarga del archivo 1UBQ (Structure of ubiquitin refined at 1.8 Å resolution)

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

Structure Summary 3D View Annotations Experiment Sequence

Biological Assembly 1

1UBQ

STRUCTURE OF UBIQUITIN REFINED AT 1.8 ANGSTROMS RESOLUTION

DOI: 10.2210/pdb1UBQ/pdb BMRB: 5387

Classification: CHROMOSOMAL PROTEIN

Organism(s): Homo sapiens

Mutation(s): No

Deposited: 1987-01-02 Released: 1987-04-16

Deposition Author(s): Vijay-Kumar, S., Bugg, C.E., Cook, W.J.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.80 Å

R-Value Observed: 0.176

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore	11	11
Ramachandran outliers	10	10
Sidechain outliers	8.8%	8.8%
RSRZ outliers	5.3%	5.3%

This is version 1.2 of the entry. See complete history.

Literature

Download Primary Citation

Structure of ubiquitin refined at 1.8 Å resolution.

Vijay-Kumar, S., Bugg, C.E., Cook, W.J., (1987) J Mol Biol 194: 531-544

PubMed: 3041007 Search on PubMed

DOI: 10.1016/0022-2836(87)90679-6

Primary Citation of Related Structures: 1UBQ

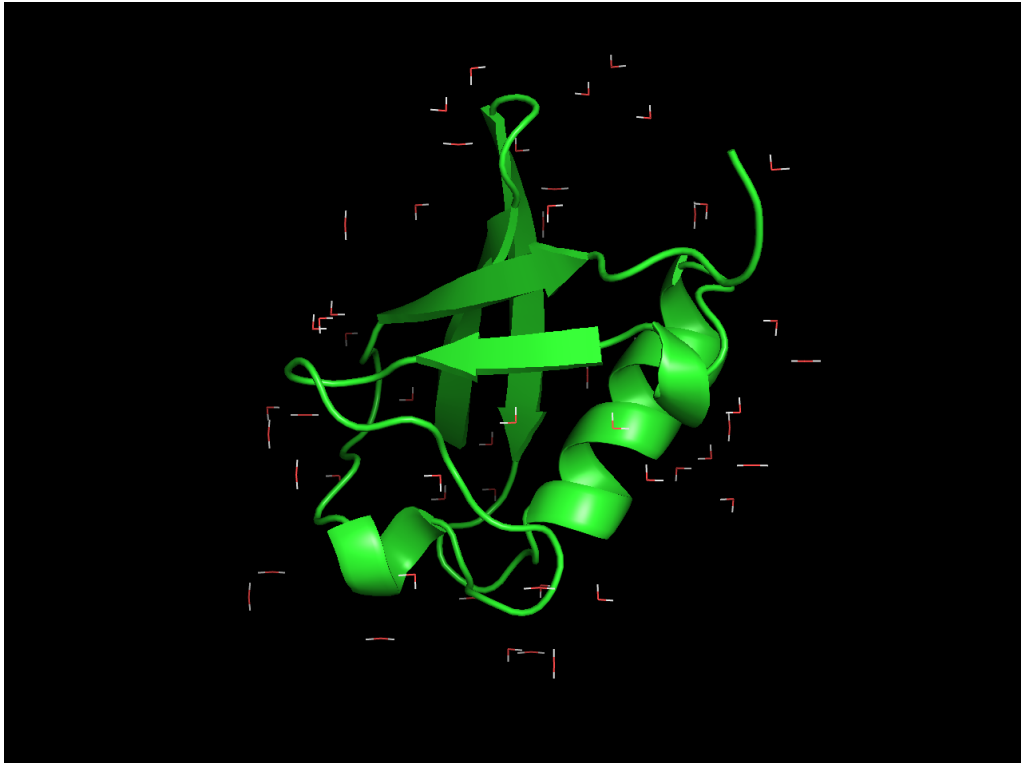
PubMed Abstract:

The crystal structure of human erythrocytic ubiquitin has been refined at 1.8 Å resolution using a restrained least-squares procedure. The crystallographic R-factor for the final model is 0.176. Bond lengths and bond angles in the molecule have root- ...

Fuente: <https://www.rcsb.org/structure/1UBQ>

2. Estructura inicial formato PDB

Se realiza la visualización del archivo descargado utilizando la herramienta pymol obteniendo la siguiente imagen (se adjunta archivo).



Archivo PDB descargado y visualizado a través de pymol. Fuente: Herramienta pymol

3. Crear archivos de topología y coordenadas de GROMACS.

Para este paso se utiliza la herramienta GROMACS. Esta herramienta necesita trabajar los archivos .pdb en su propios formatos. Los archivos .pdb son compatibles con GROMACS, pero deben convertirse a un archivo de coordenadas .gro. El primer paso en la simulación de MD con GROMACS es crear archivos de topología (.top) y coordenadas (.gro) compatibles con GROMACS.

Para dicho propósito se ejecuta el siguiente script (utilizando la función pdb2gmx).

```
> gmx pdb2gmx -f /Users/sergio.castaneda/Desktop/1ubq.pdb -o protein.gro -p topol.top -i posre.itp -ignh -v
```

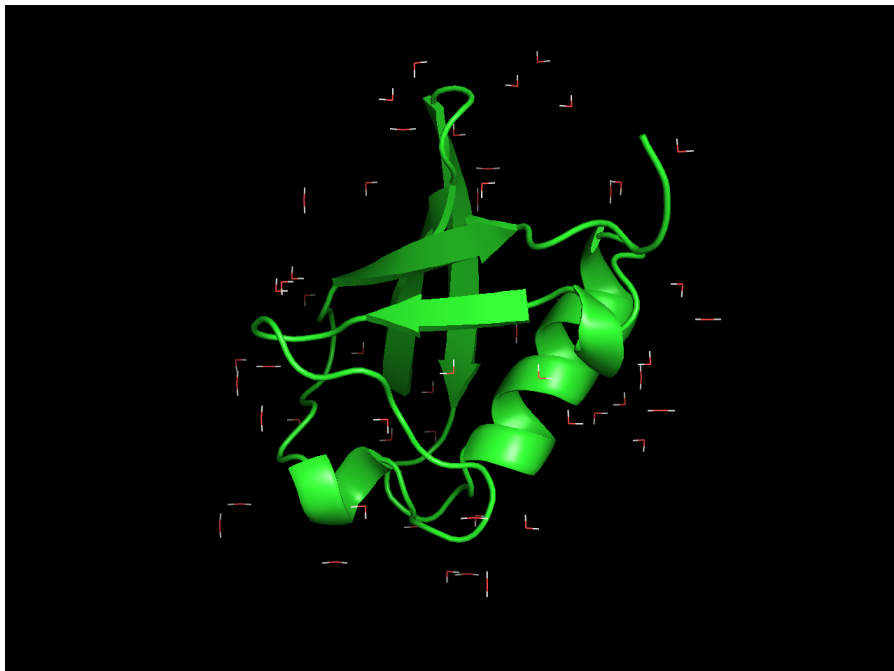
Con este comando se obtienen los 3 archivos correspondientes a estructura y topología que serán utilizados posteriormente (.top, .gro y .itp) (Se adjuntan archivos)

4. Estructura inicial en formato GROMACS

Para visualizar la estructura de la proteína en formato gromacs (protein.gro) con la herramienta pymol se utiliza el siguiente script.

```
> pymol protein.gro
```

Se obtiene la siguiente imagen, correspondiente a la imagen vista anteriormente pero con el formato soportado por GROMACS (se adjunta archivo).



Archivo formato gromacs (.gro) visualizado a través de pymol. Fuente: Herramienta pymol

5. Preparación de una caja de agua que encierre la proteína

La estructura de la proteína (sistema) se debe encerrar dentro de una caja, para limitar la simulación a este espacio, y la caja se debe llenar con un solvente (ej. agua). En los siguientes pasos, se definirá un tamaño de caja para el sistema, se centrará la proteína en la caja, se solvatará y, finalmente, se agregará restricciones al sistema. Los archivos usados para este paso son:

- protein.gro # archivo de coordenadas en formato gromacs.
- topol.top # archivo de topología en formato gromacs
- posre.itp # archivo de restricción de posición

Para crear una caja rectangular alrededor de la proteína se utiliza la función editconf de GROMACS, utilizando el siguiente script.

```
> gmx editconf -f protein.gro -o protein_box.gro -bt triclinic -d 1.2 -c
```

Los parámetros utilizados en este comando corresponden a:

'-bt triclinic' crear una caja rectangular.

'-d 1.2' crea un búfer de 1.2 nm entre el exterior de la proteína y el borde de la caja

'-c' centra la proteína en la caja y coloca la esquina de la caja en {0, 0, 0} en espacio cartesiano.

'-o protein_box.gro' es el archivo de salida.

6. Adición del solvente a la caja del sistema

El anterior comando crea la caja pero aún no se ha llenado de solvente. Con el siguiente comando de GROMACS llenará la caja con moléculas de solvente. Para ello se utiliza la función solvate de GROMACS.

```
> gmx solvate -cp protein_box.gro -cs spc216.gro -o protein_sol.gro -p topol.top
```

'-cp' especifica el archivo de coordenadas del sistema que se va a solvatar

'-cs spc216.gro' le dice a genbox que use un archivo de coordenadas SPC (un modelo de agua simple de tres puntos) para completar la caja.

'-o protein_sol.gro' es el archivo de coordenadas

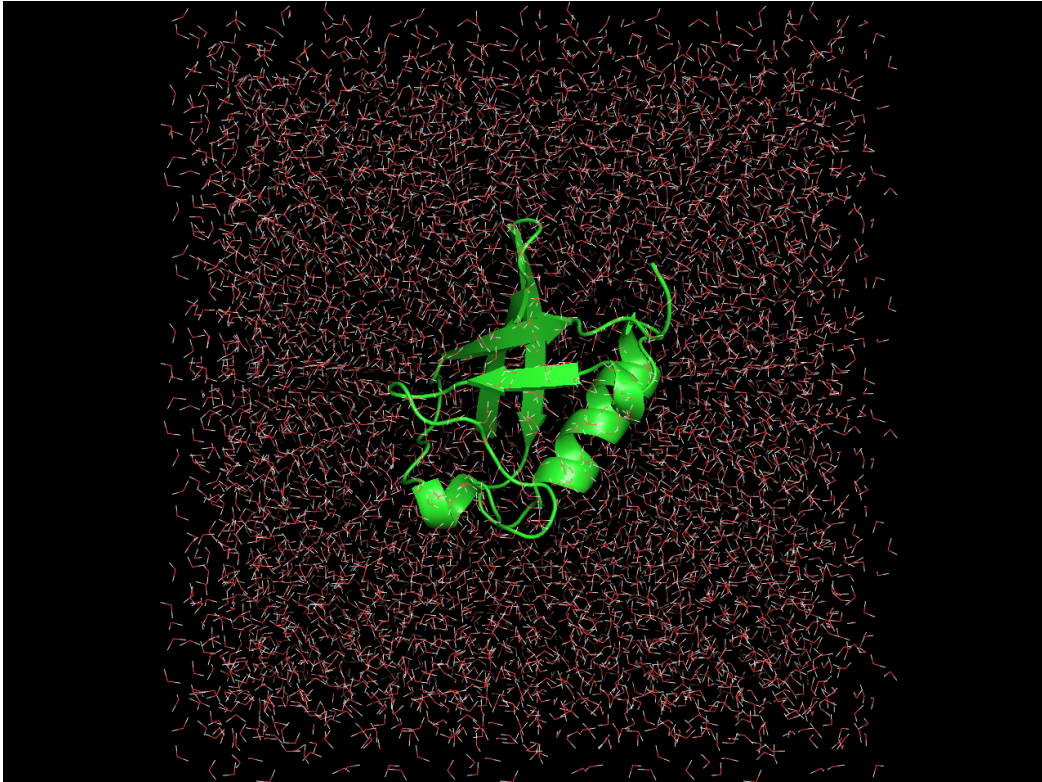
'-p topol.top' es el archivo de topología que también se proporciona en la línea de comandos.

7. Proteína solvatada encerrada en caja

Para visualizar el proceso realizado anteriormente en donde se obtuvo el archivo protein_sol.gro, se ejecuta el comando

```
> pymol protein_sol.gro
```

Obteniendo la siguiente imagen (se adjunta archivo):



Proteína solvatada encerrada en caja rectangular. Fuente: Herramienta pymol

8. Adición de iones al sistema para neutralizar la carga neta

El último paso antes de la simulación es agregar suficientes iones al sistema para neutralizar la carga neta o, alternativamente, agregar suficientes iones para neutralizar la carga neta y alcanzar cierta concentración fisiológica.

La herramienta de GROMACS con la función 'genion' busca en el archivo de coordenadas y reemplaza aleatoriamente las moléculas de agua con iones. Sin embargo, para que funcione, se requiere un archivo de entrada preprocesado (con extensión .tpr) que contenga toda la información de los archivos de coordenadas y topología. Para generar dicho archivo se utilizará la función 'grompp'. Además del archivo de topología y coordenadas, 'grompp' también requiere que proporcione un archivo de parámetros MD (.mdp) como entrada. Un ejemplo de archivo .mdp, llamado "genion.mdp" con todos los parámetros predeterminados fue descargado del repositorio de github <https://github.com/luigarreta/taller-simulacionDM.git>.

Se realiza la clonación del archivo de github con el siguiente comando:

```
> git clone https://github.com/luigarreta/taller-simulacionDM.git
```

Posteriormente se obtiene el archivo de entrada .tpr por medio del siguiente comando:

```
> gmx grompp -f /Users/sergio.castaneda/taller-simulacionDM/genion.mdp -c protein_sol.gro -p topol.top -o genion_input.tpr
```

La carga neta del sistema de ubiquitina ya es 0, así que en lugar de neutralizar el sistema, se agrega suficiente NaCl para alcanzar una concentración de sal de 100 mM. Para tal fin se ejecuta el siguiente comando:

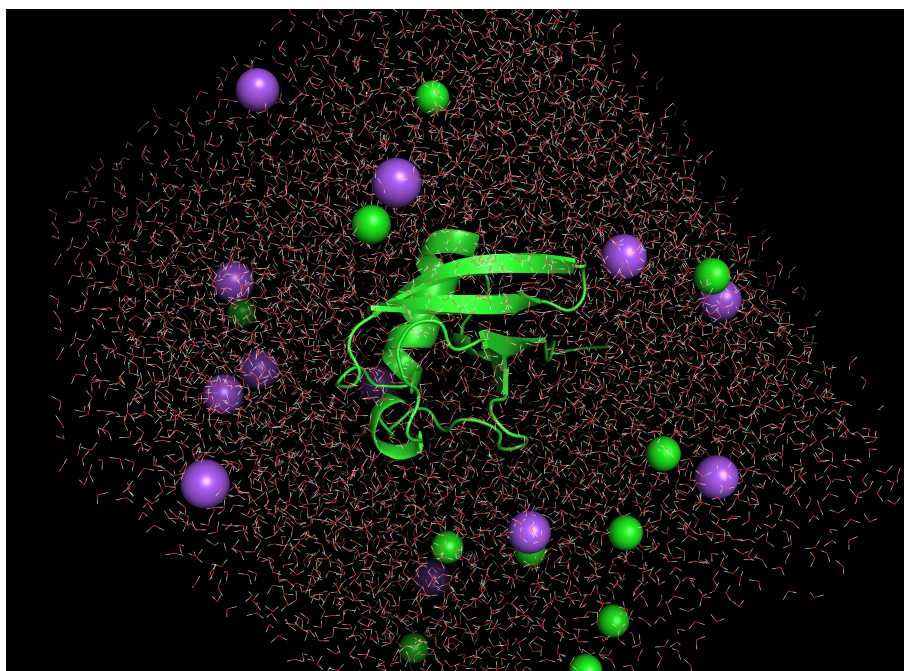
```
> gmx genion -s genion_input.tpr -p topol.top -o protein_sol_nacl.gro -pname NA -pq 1 -nname CL -nq -1 -conc 0.1 -neutral
```

“-s genion_input.tpr”, es el archivo de entrada creado con el comando anterior.

“-o protein_sol_nacl.gro” es el nuevo archivo que se crea y que contiene el nuevo sistema adicionado los iones.

9. Sistema adicionado con iones

Se realiza la visualización de la estructura de la proteína solvatada y con iones en formato gromacs (protein_sol_nacl.gro) con la herramienta pymol (se adjunta archivo).



Proteína solvatada y con iones en formato GROMACS. Fuente: Herramienta pymol

10. Minimización de energía

Antes de ejecutar la dinámica real, deberá realizar una minimización de energía. El propósito de una minimización es relajar la geometría molecular del sistema para deshacerse de cualquier choque

atómico u otras irregularidades que puedan existir. Los archivos que necesita para iniciar este paso son:

"protein_sol_nacl.gro" # archivo de coordenadas en formato gromacs
"topol.top" # archivo de topología en formato gromacs
"posre.itp" # archivo de restricción de posición
"ubq_min.mdp" # archivo de parámetros de minimización de energía

El archivo 'ubq_min.mdp' contiene los parámetros de ejecución para la minimización. Se busca una copia de este archivo en el repositorio descargado en el paso anterior (<https://github.com/luisgarreta/taller-simulacionDM.git>). En resumen, este archivo de parámetros exige una minimización de la energía de descenso (steepest descent) que no exceda los 2000 pasos. La minimización se considerará exitosa si la fuerza máxima sobre cualquier átomo es inferior a 1000 kJ/mol/nm, momento en el que se detendrá la minimización.

El principal motor de integración de GROMACS es una función llamada 'mdrun'. Como entrada, requiere un archivo de entrada de ejecución preprocesado (.tpr) que contiene las coordenadas, la topología y los parámetros del sistema para la minimización en sí. Como antes, 'grompp' se usa para generar ese archivo.

```
> gmx grompp -f /Users/sergio.castaneda/taller-simulacionDM/ubq_min.mdp -c  
protein_sol_nacl.gro -p topol.top -o input_min.tpr
```

Al generar el archivo .tpr se realiza la minimización utilizando el comando:

```
> gmx mdrun -s input_min.tpr -deffnm ubiquitin_min -v
```

Al usar la opción '-v', se puede ver el progreso de la minimización en la pantalla. Usando el método de descenso (steepest descent), los sistemas pequeños pueden equilibrarse después de solo unos pocos cientos de pasos; los sistemas más grandes pueden requerir varios miles de pasos. Depende principalmente de la complejidad del sistema y la calidad de la estructura original inicial.

Una vez que se complete la minimización (que debería tomar alrededor de 500 pasos), se generan varios archivos, todos con el prefijo 'ubiquitin_min':

ubiquitin_min.gro # coordenadas del sistema en el paso final
ubiquitin_min.trr # archivo de trayectoria
ubiquitin_min.log # archivo de registro
ubiquitin_min.edr # archivo de energía

El primer archivo, 'ubiquitin_min.gro', contiene las coordenadas completamente minimizadas del sistema. El segundo archivo, 'ubiquitin_min.trr' contiene una trayectoria de minimización. La frecuencia con la que se escriben los fotogramas en la trayectoria se especifica en el archivo .mdp

que se pasó a grompp. El archivo de registro contiene información sobre los parámetros de ejecución utilizados para la minimización, así como varias energías del sistema durante la minimización. Por último, ubiquitin_min.edr es un archivo binario de energía que puede leerse con la herramienta GROMACS 'energy', lo que proporciona muchas piezas de información valiosa. En este caso, se utilizó para comprobar la energía potencial del sistema en función del paso de minimización. Para tal fin se utilizó el siguiente comando:

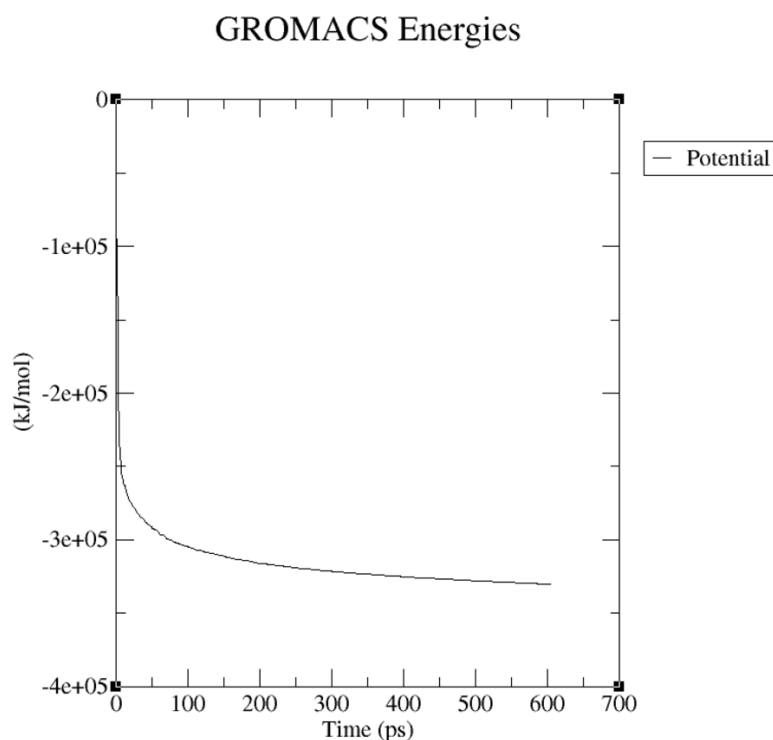
```
> gmx energy -f ubiquitin_min.edr -o energy.xvg
```

11. Curva de minimización de energía

El archivo "energy.xvg" es un archivo de datos simple de dos columnas que se puede trazar en un programa como Grace (xmgrace) o abrir el archivo en un editor de texto y tomar los datos de las dos columnas y graficarlos en excel. Se utilizó la herramienta xmgrace desde la línea de comandos. Se activa inicialmente Xquartz para una correcta visualización.

```
> xmgrace energy.xvg
```

Se obtiene una imagen como la siguiente donde se observa el proceso de minimización de energía exitoso.



Curva de minimización de energía. Fuente: Herramienta xmgrace