



UNIVERSIDAD  
DE MURCIA

UNIVERSIDAD DE MURCIA - FACULTAD DE BIOLOGÍA

---

## Modelado molecular del tripeptido: **ALA-ARG-ASP**

---

ASIGNATURA: (6654) MODELADO MOLECULAR - MÁSTER DE BIOINFORMÁTICA

2023-2024

*Autor:*

Francisco Javier Redondo Aniorte

*orcid :*

0000-0002-0794-5408

*Coordinador del curso:*

Adolfo Bastida

14 de marzo de 2024

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Metodología</b>	<b>5</b>
2.1. Construcción . . . . .	5
2.1.1. Topología . . . . .	5
2.1.2. Condiciones periódicas de contorno . . . . .	5
2.2. Equilibrado (caso de uso 298K) . . . . .	6
2.2.1. Temperatura . . . . .	6
2.2.2. Cargas . . . . .	7
2.2.3. Posiciones finales y velocidades . . . . .	8
2.3. Ejecución (caso de uso 298K) . . . . .	9
2.3.1. Fichero de entrada de ejecución . . . . .	9
2.3.2. Trayectoria de la simulación . . . . .	10
2.3.3. Visualización de la simulación . . . . .	10
2.4. Análisis (caso de uso 298K) . . . . .	11
2.4.1. Radio de giro . . . . .	11
2.4.2. Distancias de enlace . . . . .	11
2.4.3. Ángulos de enlace . . . . .	12
2.4.4. Ángulos dihedros . . . . .	12
2.4.5. Puentes de hidrógeno . . . . .	13
2.4.6. Temperatura . . . . .	14
2.4.7. Velocidad . . . . .	14
2.4.8. Energía total . . . . .	15
2.4.9. Energía cinética . . . . .	15
<b>3. Resultados</b>	<b>16</b>
3.1. Realizar simulaciones a 298K y 400K. . . . .	16
3.2. Evolución de parámetros con el tiempo . . . . .	16
3.2.1. Energía total y cinética . . . . .	16
3.2.2. Temperatura del sistema . . . . .	17
3.2.3. Dos distancias de enlace seleccionadas . . . . .	17

---

3.2.4. Dos ángulos . . . . .	18
3.2.5. Dos dihedros . . . . .	19
3.2.6. Radio de giro . . . . .	21
3.2.7. Velocidades de 5 átomos diferentes . . . . .	21
3.3. Evaluar los valores medios de la energía cinética del sistema a ambas temperaturas y compararlas con los resultados esperados desde el punto de vista de la termodinámica estadística . . . . .	22
3.4. Extra: Simulación de 500 ps a 298K con exportación cada 10 fs . . . . .	23
3.4.1. Histograma para la temperatura. . . . .	24
3.4.2. Histogramas para tres componentes de la velocidad de un átomo seleccionado. . . . .	24
3.4.3. Crear el diagrama de Ramachandran para los tres residuos. . . . .	25
3.4.4. Identificar las conformaciones de probabilidad más alta. . . . .	25

## 1. Introducción

En este trabajo se va a realizar una simulación molecular de un tripéptido, en un entorno, con condiciones constantes de número de partículas (N) y volumen (V). Además, en el equilibrio, el sistema permanece a temperatura constante, y se puede considerar que está en contacto con un baño térmico. A estas condiciones se le denominan Colectivo (o ensemble) NVT.

Se enumeran a continuación las tareas y objetivos a alcanzar:

1. Realizar simulaciones a 298 K y 400 K.
2. Mostrar cómo evolucionan con el tiempo los siguientes parámetros: - Energía total y cinética; - Temperatura del sistema; - Dos distancias de enlace seleccionadas; - Dos ángulos; - Dos dihedros; - Radio de giro; - Velocidades de 5 átomos diferentes. Comparar los resultados para cada una de las temperaturas.
3. Evaluar los valores medios de la energía cinética del sistema a ambas temperaturas y compararlas con los resultados esperados desde el punto de vista de la termodinámica estadística.
4. Ejecutar una simulación de 500 ps a 298 K y exportar los resultados cada 10 fs.
  - a) Obtener el histograma para la temperatura.
  - b) Obtener los histogramas para las tres componentes de la velocidad de un átomo seleccionado.
  - c) Crear el diagrama de Ramachandran para los tres residuos.
  - d) Identificar las conformaciones de probabilidad más alta.

Inicialmente se parte de un fichero [ard.pdb](#) con la secuencia de aminoácidos (ALA-ARG-ASP), flanqueada por dos extremos terminales:

- ACE (CH<sub>3</sub>CO-)
- NME (-NHCH<sub>3</sub>)

La secuencia queda de esta manera:

ACE-ALA-ARG-ASP-NME

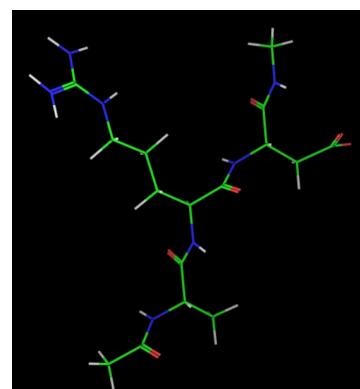


Figura 1: Péptido ACE-ALA-ARG-ASP-NME

Es interesante hacer constar que, a pesar de que el péptido, globalmente, es neutro, no lo son dos de sus aminoácidos en condiciones fisiológicas. Concretamente, la arginina (ARG) es positiva, mientras que el ácido aspártico (ASP) tiene carga neta negativa (Figura 2). Esto tendrá que tenerse en cuenta en la fase de neutralización de carga.

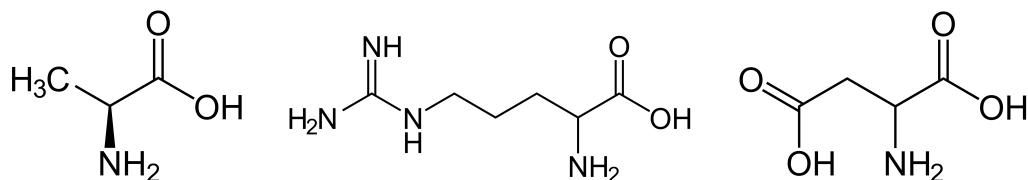


Figura 2: Aminoácidos del tripéptido: ALA, ARG, ASP

Los ficheros de la práctica se encuentran en el cluster de servidores de la UMU, en la ruta siguiente: *alumno24@dayhoff.inf.um.es* :~ /mm/ard. No obstante, por mayor comodidad se ha subido el contenido al repositorio de Github: <https://github.com/fjredondo/ard-tripeptide> [5] <sup>1</sup>

---

<sup>1</sup>Por limitaciones de espacio, no se ha subido al repositorio de Github [5] ficheros de más de 40 MB. Debido a ello, dichos ficheros han de ser consultados en: *alumno24@dayhoff.inf.um.es* :~ /mm/ard

## 2. Metodología

Se han utilizado las siguientes aplicaciones: pymol [4], gmx [2] y gnuplot [3]

### 2.1. Construcción

Inicialmente se parte de un fichero [ard.pdb](#) con la secuencia de aminoácidos (ALA-ARG-ASP), flanqueada por dos extremos terminales: ACE (CH<sub>3</sub>CO-) y NME (-NHCH<sub>3</sub>)

#### 2.1.1. Topología

Se lanza el siguiente comando para obtener las topologías, con las opciones: campo de fuerzas “charmm27.ff/forcefield.itp” y modelo de agua “TIP3P”

```
$ gmx pdb2gmx -f ard.pdb -o ard.gro -p ard.top -ter
```

Se obtienen los ficheros de [ard.top](#) (topología molecular, [.top](#)) y [ard.gro](#) (estructura molecular, [.gro](#)). Se comprueba, así mismo, que el contenido de los ficheros es consecuente con la secuencia peptídica.

#### 2.1.2. Condiciones periódicas de contorno

Con la siguiente sentencia, se crea un caja de 27 nm<sup>3</sup> que contiene el péptido y las nuevas coordenadas se almacenan en el fichero: [ard-box.gro](#).

```
$ gmx editconf -f ard.gro -o ard-box.gro -bt cubic -box 3.0 3.0 3.0
```

Hasta este punto, los ficheros contienen las coordenadas de 58 átomos. A continuación, se solvata el péptido con moléculas de agua.

```
$ gmx solvate -cp ard-box.gro -cs -o ard-box-solv.gro -p ard.top
# Output configuration contains 2638 atoms in 865 residues
# Volume : 27 (nm^3)
# Density : 979.957 (g/l)
# Number of SOL molecules: 860
$ gmx editconf -f ard-box-solv.gro -o ard-box-solv.pdb
```

El resultado (2638 átomos) se almacena en el fichero: [ard-box-solv.gro](#). Por último, para que pueda ser dibujado se convierte a un [.pdb](#), como se muestra en la Figura 3.

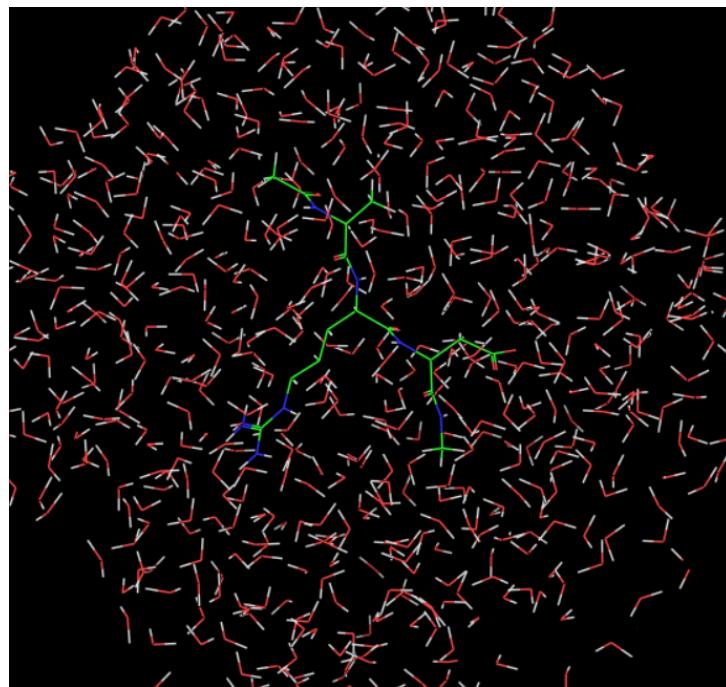


Figura 3: Peptido solvatado ([ard-box-solv.pdb](#))

## 2.2. Equilibrado (caso de uso 298K)

En esta fase comienza el equilibrado de temperaturas y, en ella, se generarán archivos de salida con velocidades y energías dependientes de dicha temperatura. Por ello, se va a exponer, como primer caso de uso, el equilibrado a 298 K. No obstante, los resultados para 400 K, pueden reproducirse, de igual manera, siguiendo los mismos pasos. Las particularidades en relación a la temperatura de 400 K, se expondrán en la Sección [3.1](#).

### 2.2.1. Temperatura

Los parámetros de frecuencia de actualización de datos de coordenadas, velocidades y energías, y, la configuración de temperatura, se encuentra en el siguiente fichero [.mdp](#): [equiNVT.mdp](#).

Tabla 1: Fichero de parámetros de frecuencia de actualización de datos: `equiNVTmdp`

```
; RUN CONTROL PARAMETERS
dt      = 0.0005 ; 0.5 fs
nsteps  = 800000 ; 400 ps
nstcomm = 4000 ; number of steps for
                  center of mass motion removal
; OUTPUT CONTROL OPTIONS
nstxout = 4000 ; Save coordinates every 2 ps
nstvout = 4000 ; Save velocities every 2 ps
nstlog  = 4000 ; Update log every 2 ps
nstenergy = 4000 ; Save energies every 2 ps
; OPTIONS FOR ELECTROSTATICS AND VDW
; Temperature coupling
ref_t   = 298 ; Avobe fluid phase trasition
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel = yes ; Should be the first equilibration
gen_temp = 298.0 ; Temperature to generate
                  corresponding Maxwell distribution
```

Se explica a continuación el detalle de la simulación:

Se consideran intervalos temporales de 0.5 fs con un total de 800000 pasos, lo que equivale a:  $800000(steps) \times 0.5(fs) \times 10^{-3}(\frac{ps}{fs}) = 400(ps)$ . El registro de datos se planifica cada 2 ps, o lo que es lo mismo, cada 4000 pasos, durante los 400 ps de la simulación.

Esta configuración se va a utilizar para la obtención de un fichero binario `.tpr`, el cual puede servir de entrada para la ejecución mediante el comando "gmx mdrun".

```
$ gmx grompp -f equiNVTmdp -c ard-box-solv.gro -p ard.top -o ard-a.tpr
```

### 2.2.2. Cargas

Sobre el fichero binario `.tpr`, obtenido en el equilibrado de la temperatura, se aplica una neutralización de las cargas de los péptidos. En este caso, se van a neutralizar las cargas de los aminoácidos no neutros, como son, la arginina y el ácido aspártico.

```
$ gmx genion -s ard-a.tpr -p ard.top -o ard.gro -nn 1 -np 1
# Replacing 2 solute molecules in topology file (ard.top)
# by 1 NA and 1 CL ions.
```

Se ha neutralizado la carga global del sistema añadiendo un ión  $Cl^-$  y un  $Na^+$  (permutando con dos moléculas de solvente elegidas al azar).

### 2.2.3. Posiciones finales y velocidades

Una vez hecho esto, se recalculan las posiciones finales y velocidades, teniendo en cuenta los iones añadidos. Por último, se obtiene el fichero .pdb, para su visualización.

```
$ gmx grompp -f equiNVTmdp -c ard.gro -p ard.top -o ard.tpr
$ gmx editconf -f ard.gro -o ard.pdb
```

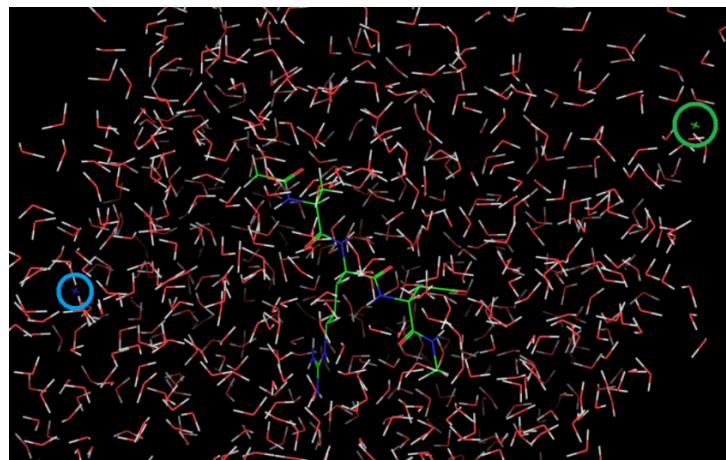


Figura 4: Iones  $Cl^-$  y un  $Na^+$  añadidos en fichero ard.pdb

El fichero ard.pdb muestra en las últimas posiciones los dos iones añadidos:

ATOM	2632	HW2	SOL	863	18.610	28.530	21.940	1.00	0.00
ATOM	2633	NA	NA	864	27.280	23.160	25.040	1.00	0.00
ATOM	2634	CL	CL	865	25.940	6.340	29.260	1.00	0.00

En este momento, se está en condiciones de generar un fichero .g96. El siguiente comando se encuentra en el script `submit-eck.sh`

```
$ gmx mdrun -deffnm ard -c ard.g96 -nt 1
```

El script anterior se lanza sobre el gestor de trabajos Slurm [1], mediante el comando sbatch. Con squeue es posible monitorizar el estado del trabajo hasta su finalización. Por último, la salida de log se envía a través de los ficheros siguientes: `slurm-35234.out` y `ard.log`. Este último, contiene el detalle de las coordenadas, las velocidades y las energías (KJ/mol). Al final del fichero se muestra una estadística con las medias.

El fichero con las posiciones espaciales y las velocidades según cada eje (x,y,z), de los 2634 átomos, se ha almacenado en: `ard.g96`. Este fichero será utilizado en la fase de ejecución.

### 2.3. Ejecución (caso de uso 298K)

En esta fase entra la topología **.top**: ([ard.top](#)), el fichero con las posiciones y las velocidades **.g96**: [ard.g96](#) y, por último, el archivo de parámetros **.mdp**: [runNVT.mdp](#).

Tabla 2: Fichero de parámetros de frecuencia de actualización de datos: [runNVT.mdp](#)

```
; RUN CONTROL PARAMETERS
dt          = 0.0005 ; 0.5 fs
nsteps      = 4000   ; 2 ps
nstcomm     = 1000   ; number of steps for
                      center of mass motion removal
; OUTPUT CONTROL OPTIONS
nstxout     = 2       ; Save coordinates every 1 fs
nstvout     = 2       ; Save coordinates every 1 fs
nstlog      = 2       ; Save coordinates every 1 fs
nstenergy   = 2       ; Save coordinates every 1 fs
; OPTIONS FOR ELECTROSTATICS AND VDW
; Temperature coupling
ref_t       = 298    ; Avobe fluid phase trasition
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel     = no      ; Should be the first equilibration
```

Se explica a continuación el detalle de la simulación:

Se consideran intervalos temporales de 0.5 fs con un total de 4000 pasos, lo que equivale a:  $4000(steps) \times 0.5(fs) \times 10^{-3}(\frac{ps}{fs}) = 2(ps)$ . El registro de datos se planifica cada 1 fs, o lo que es lo mismo cada 2 pasos, durante los 2 ps de la simulación.

#### 2.3.1. Fichero de entrada de ejecución

Se lanza el siguiente comando:

```
$ gmx grompp -f runNVT.mdp -c ard.g96 -p ard.top -o ard.tpr
```

En este punto se obtiene el fichero binario **.tpr**, entrada para la ejecución mediante el comando "gmx mdrun".

### 2.3.2. Trayectoria de la simulación

Las coordenadas y velocidades se almacenan en un fichero `.trr`, salida de la ejecución mediante el comando ”`gmx mdrun`”.

```
$ gmx mdrun -deffnm ard -c ard.g96 -nt 1
```

Al igual que se hizo en la fase de equilibrado se lanza un trabajo sobre Slurm [1]. El script en este caso es el siguiente: `submit-eck.sh`. La salida de log se envía a través de los ficheros siguientes: `slurm-35236.out` y `ard.log`.

### 2.3.3. Visualización de la simulación

La visualización se realiza transformando los ficheros en formato binario `.tpr` y `.trr` en ficheros de formato `.pdb`

```
$ gmx traj -f ard.trr -s ard.tpr -oxt cartoon.pdb
```

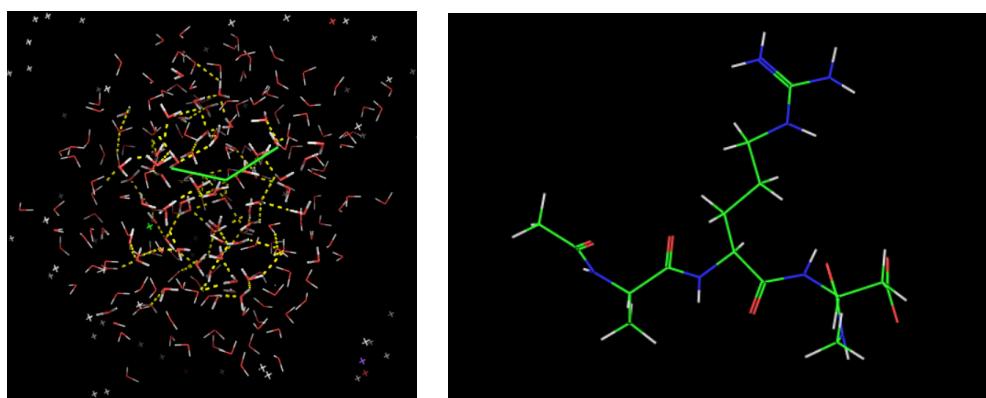


Figura 5: Simulaciones del movimiento del péptido con y sin agua

Por limitaciones de espacio, no se ha subido al repositorio de Github ficheros de más de 40 MB, debido a ello, los ficheros “`ard.trr`”, “`cartoon.pdb`”, han de ser consultados desde el servidor `alumno24@dayhoff.inf.um.es` en la ruta: `/mm/ard`

## 2.4. Análisis (caso de uso 298K)

Se va a realizar el análisis sobre el péptido considerando el caso de uso de temperatura a 298K.

La siguiente figura muestra las posiciones de los átomos que van a ser analizadas en los siguientes apartados.

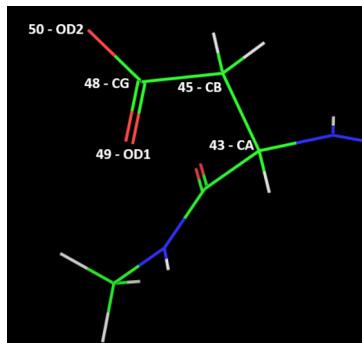


Figura 6: Péptido con posiciones a analizar

### 2.4.1. Radio de giro

Se trata de realizar un cálculo del radio de giro de la molécula, a partir de la raíz cuadrada de la masa ponderada, cada 1 fs en el intervalo de 2 ps. La fórmula es la siguiente:

$$Rg(x) = \sqrt{\left(\sum_i w_i (R_i(y)^2 + R_i(z)^2)\right) / \left(\sum_i w_i\right)}$$

Se utiliza para ello la función [gmx gyrate](#).

```
$ gmx gyrate -f ard.trr -s ard.tpr -xvg none
# Selected 1: 'Protein'
> gnuplot
gnuplot > plot 'gyrate.xvg'
```

El siguiente fichero almacena 2001 medidas del radio de giro: [gyrate.xvg](#). Se visualiza en la Figura 17.

### 2.4.2. Distancias de enlace

Se crea un fichero [.ndx](#) con la información de los átomos cuyo enlace se va a analizar: [distances-43-45.ndx](#)

```
[CA-CB-asp-4]
43 45
```

Se utiliza para ello la función [gmx distance](#).

```
$ gmx distance -f ard.trr -s ard.tpr -n distances_43_45.ndx -oall -xvg none
# CA-CB-asp-4:
# Number of samples: 2001
```

```
# Average distance: 0.15490 nm
# Standard deviation: 0.00343 nm
```

Como se puede apreciar, es posible visualizar en la salida el resultado medio y la desviación estándar de la distancia. Estos resultados se utilizarán en el apartado 3.2.3

El siguiente fichero almacena 2001 medidas de distancia: [dist-CA-CB-asp-4.xvg](#). Se visualiza en la Figura 11.

#### 2.4.3. Ángulos de enlace

Se crea un fichero [.ndx](#) con la información de los átomos cuyos ángulos de enlace se va a analizar: [angles-43-45-48.ndx](#)

```
[CA-CB-CG-asp-4]
43 45 48
```

Se utiliza para ello la función [gmx angle](#).

```
$ gmx angle -f ard.trr -n distances_43_45_48.ndx -ov -xvg none
# Group 0 ( CA-CB-CG-asp-4) has 3 elements
# Found points in the range from 102 to 121 (max 180)
# < angle > = 111.256
# < angle^2 > = 12388.8
# Std. Dev. = 3.29757
```

Como se puede apreciar, es posible visualizar en la salida el resultado medio y la desviación estándar de la distancia. Estos resultados se utilizarán en el apartado 3.2.4

El siguiente fichero almacena los ángulos y su probabilidad de aparición: [angdist-43-45-48.xvg](#). Se visualiza en la siguiente Figura 13.

#### 2.4.4. Ángulos dihedros

Se crea un fichero [.ndx](#) con la información de los átomos cuyos ángulos de enlace se va a analizar: [dihedral-49.ndx](#)

```
[CA-CB-CG-OD1-asp-4]
43 45 48 49
```

Se utiliza para ello la función [gmx angle](#).

```
$ gmx angle -f ard.trr -type dihedral -n dihedral_49.ndx -oc -xvg none
# Group 0 (CA-CB-CG-OD1-asp-4) has 4 elements
# Found points in the range from 238 to 301 (max 360)
# < angle > = 96.167
# < angle^2 > = 9400.55
# Std. Dev. = 12.3476
```

Como se puede apreciar, es posible visualizar en la salida el rango, el ángulo medio y la desviación estándar de la distancia. Estos resultados se utilizarán en el apartado 3.2.5

El siguiente fichero almacena los ángulos y su probabilidad de aparición: [angdist-dihedral-49.xvg](#). Se visualiza en la siguiente Figura 15.

También es posible obtener los ángulos *phi* y *psi* por separado para cada aminoácido.

Se utiliza para ello la función [gmx rama](#).

```
$ gmx rama -f ard.trr -s ard.tpr -xvg none
# Found 3 phi-psi combinations
$ grep ASP-4 rama.xvg | awk '{print $1}' | cat -n > phi-asp-4.dat
$ grep ASP-4 rama.xvg | awk '{print $2}' | cat -n > psi-asp-4.dat
$ join phi-asp-4.dat psi-asp-4.dat > asp-4.dat
```

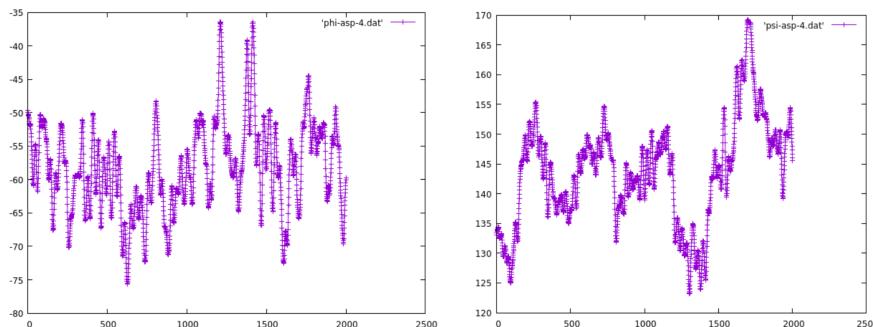


Figura 7: Ángulos *phi* ([phi.dat](#)) y *psi* ([psi.dat](#)) de ASP a 298K

#### 2.4.5. Puentes de hidrógeno

Se utiliza para ello la función [gmx hbond](#).

```
$ gmx hbond -f ard.trr -s ard.tpr -a 120 -r 0.18 -noda -num hbond_num.xvg
-xvg none -nonitacc -hx hbond_hx.xvg
# Checking for overlap in atoms between Protein and SOL
# Calculating hydrogen bonds between Protein (58 atoms) and SOL #(2574 atoms)
# Found 865 donors and 864 acceptors
# Last frame      2000 time      2.000
# Average number of hbonds per timeframe 8.172 out of 373680 possible
```

El siguiente fichero almacena el número de puentes de hidrógeno en función del tiempo: [hbond-num.xvg](#).

#### 2.4.6. Temperatura

Se utiliza para ello la función `gmx traj`.

```
$ gmx traj -f ard.trr -s ard.tpr -xvg none -ot
# Selected 0: 'System'
> gnuplot
gnuplot> width = 2
gnuplot> hist(x,width)=width*floor(x/width)+width/2.0
gnuplot> set boxwidth width*0.9
gnuplot> set style fill solid 0.5
# Histograma de la temperatura
gnuplot> plot 'temp.xvg' u (hist($2,width)):(1.0) smooth freq w
          boxes lc rgb'green' notitle
```

El siguiente fichero `.xvg` almacena la temperatura cada 1fs durante 2ps: `temp.xvg`.

#### 2.4.7. Velocidad

Se utiliza para ello la función `gmx traj`.

```
$ gmx traj -f ard.trr -s ard.tpr -xvg none -ov
# Selected 0: 'System'
> gnuplot
gnuplot> n=100
gnuplot> max=3. #max value
gnuplot> min=-3. #min value
gnuplot> width=(max-min)/n #interval width
gnuplot> hist(x,width)=width*floor(x/width)+width/2.0
gnuplot> set boxwidth width*0.9
gnuplot> set style fill solid 0.5 # fill style
# Histograma de la coordenada x de la velocidad
gnuplot> plot 'veloc.xvg' u (hist($2,width)):(1.0) smooth freq w
          boxes lc rgb'purple' notitle
# Histograma de la coordenada y de la velocidad
gnuplot> plot 'veloc.xvg' u (hist($3,width)):(1.0) smooth freq w
          boxes lc rgb'orange' notitle
# Histograma de la coordenada z de la velocidad
gnuplot> plot 'veloc.xvg' u (hist($4,width)):(1.0) smooth freq w
          boxes lc rgb'blue' notitle
```

El siguiente fichero `.xvg` almacena las 3 coordenadas de la velocidad cada 1fs durante 2ps: `veloc.xvg`.

#### 2.4.8. Energía total

Se utiliza para ello la función [gmx energy](#).

```
$ gmx energy -f ard.edr -s ard.tpr -xvg none
# Selected 15: 'Total-Energy'
# Energy      Average   Err.Est.      RMSD  Tot-Drift
# -----
# Total Energy -27239.8        69     181.174   -127.941 (kJ/mol)
```

El siguiente fichero [.xvg](#) almacena la Energía Total cada 1fs durante 2ps: [energy-total.xvg](#).

#### 2.4.9. Energía cinética

Se utiliza para ello la función [gmx energy](#).

```
$ gmx energy -f ard.edr -s ard.tpr -xvg none
# Selected 14: 'Kinetic-En.'
# Statistics over 4001 steps [ 0.0000 through 2.0000 ps ], 1 data sets
# All statistics are over 2001 points (frames)
# Energy      Average   Err.Est.      RMSD  Tot-Drift
# -----
# Kinetic En.  9745.28       14     135.3    -68.0321 (kJ/mol)
```

El siguiente fichero [.xvg](#) almacena la Energía Cinética cada 1fs durante 2ps: [energy-kinetic.xvg](#).

### 3. Resultados

En este apartado se van a mostrar los resultados de aplicar las técnicas de modelado molecular detalladas en el apartado 2.

#### 3.1. Realizar simulaciones a 298K y 400K.

#### 3.2. Evolución de parámetros con el tiempo

Cada uno de los parámetros analizados en este apartado van a ser comparados a dos temperaturas 298K y 400K.

##### 3.2.1. Energía total y cinética

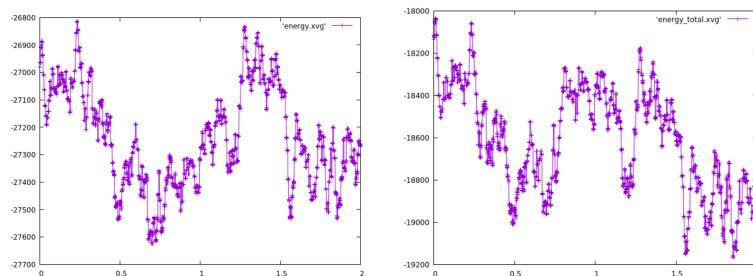


Figura 8: Energía total del sistema a 298K (8: [energy-total-298.xvg](#)) y 400K ([energy-total-400.xvg](#))

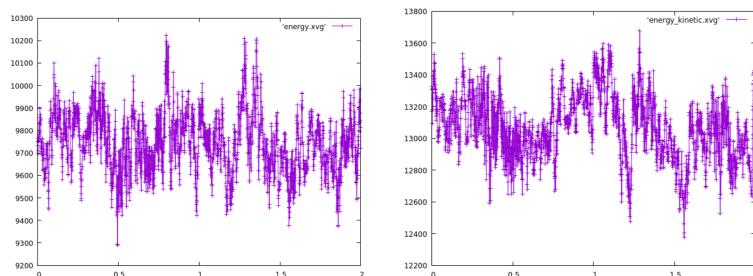


Figura 9: Energía cinética del sistema a 298K ([energy-kinetic-298.xvg](#)) y 400K ([energy-kinetic-400.xvg](#))

En las condiciones de la simulación Colectivo (o ensemble) NVT, la Energía total media no es constante, sin embargo, sí lo es la Energía cinética media, la cual está relacionada con la temperatura de equilibrio.

La Energía cinética de la Figura 9 está expresada en  $\frac{KJ}{mol}$ . Se puede comprobar que es más alta y tiene una dispersión mayor, en la gráfica de la derecha, correspondiente a 400K.

### 3.2.2. Temperatura del sistema

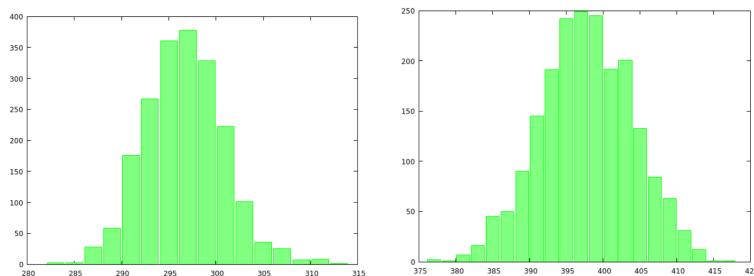


Figura 10: Temperatura del sistema a 298K ([temp-298.xvg](#)) y 400K ([temp-400.xvg](#))

Los histogramas de la Figura 10, son consecuentes con la termodinámica estadística. Tienen una forma más esbelta, o con más átomos concentrados alrededor de la temperatura de equilibrio, cuando ésta es más baja (caso 298K, a la izquierda) y, una forma más aplanada o, lo que es lo mismo, con menos átomos en la zona coincidente con la temperatura de equilibrio, cuando ésta es más alta (caso 400K, a la derecha). También, en este último caso, los átomos abarcan un rango mayor de temperaturas.

### 3.2.3. Dos distancias de enlace seleccionadas

Se ha realizado la comparativa de los enlaces CA-CB y CG-OD2 en el aminoácido ASP que se muestran en la Figura 6.

	298 K	400 K
Average distance (nm)	0.15490	0.15530
Standard deviation (nm)	0.00343	0.00366

Tabla 3: Comparativa para CA-CB (ASP) a 298K y 400K

	298 K	400 K
Average distance (nm)	0.12584	0.12588
Standard deviation (nm)	0.00197	0.00203

Tabla 4: Comparativa para CG-OD2 (ASP) a 298K y 400K

Se puede observar cómo, en ambos casos, aumenta con la temperatura, tanto la desviación estándar, como la distancia media de enlace, lo cual es consecuente con un aumento del movimiento de vibración en los enlaces.

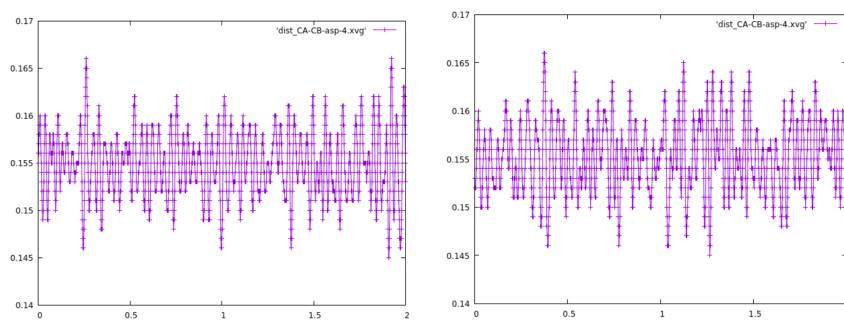


Figura 11: Dist. enlaces (nm) CA-CB en ASP a 298K ([dist-CA-CB-298.xvg](#)) y 400K ([dist-CA-CB-400.xvg](#))

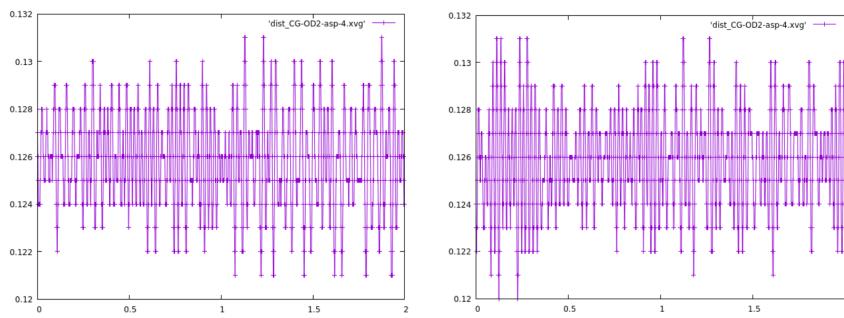


Figura 12: Dist. enlaces (nm) CG-OD2 en ASP a 298K ([dist-CG-OD2-298.xvg](#)) y 400K ([dist-CG-OD2-400.xvg](#))

### 3.2.4. Dos ángulos

Se ha realizado la comparativa de los ángulos en CA-CB-CG y CB-CG-OD2 en el aminoácido ASP que se muestran en la Figura 6.

CA-CB-CG-asp-4	298 K	400 K
Range	[102-121]	[100-122]
angle	111.256	111.006
$angle^2$	12388.8	12337.5
Std. Dev.	3.29757	3.89793

Tabla 5: Comparativa para CA-CB-CG (ASP) a 298K y 400K

CB-CG-OD2-asp-4	298 K	400 K
Range	[110-127]	[108-129]
angle	118.468	118.013
$angle^2$	14040.3	13937.3
Std. Dev.	2.3461	3.1994

Tabla 6: Comparativa para CB-CG-OD2 (ASP) a 298K y 400K

Se puede observar cómo, en ambos casos, aumenta con la temperatura, tanto la desviación estándar, como el rango de ángulos alcanzado en los movimientos de flexión.

### 3 Resultados

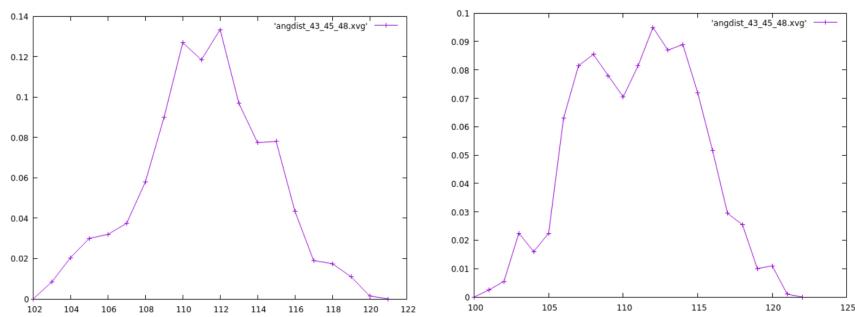


Figura 13: Diagrama de densidad de ángulos de CA-CB-CG en ASP a 298K ([angdist-43-45-48-298.xvg](#)) y 400K ([angdist-43-45-48-400.xvg](#))

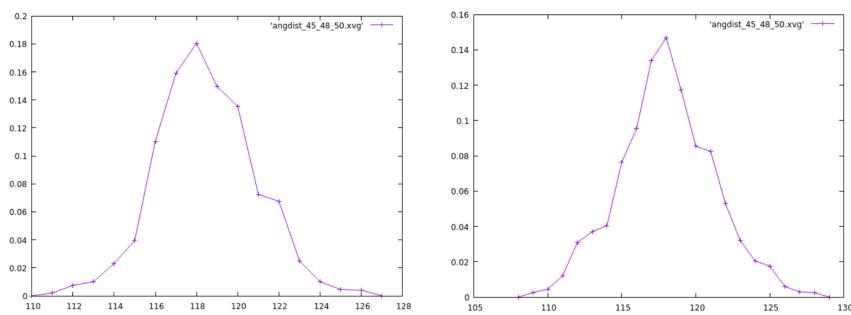


Figura 14: Diagrama de densidad de ángulos de CB-CG-OD2 en ASP a 298K ([angdist-45-48-50-298.xvg](#)) y 400K ([angdist-45-48-50-400.xvg](#))

En las gráficas de las Figuras 13 y 14, se puede ver que la distribución de densidad de ángulos se aproxima a una campana de gauss. El pico mayor de frecuencia coincide con la energía potencial más baja, es decir, el ángulo entre los tres átomos más estable o de menor energía.

Se puede observar cómo las gráficas relativas a 400K (a la derecha en ambas Figuras), son más aplanadas y anchas.

#### 3.2.5. Dos dihedros

Se ha realizado la comparativa de los ángulos dihedros en CA-CB-CG-OD1 y CA-CB-CG-OD2 en el aminoácido ASP que se muestran en la Figura 6.

CA-CB-CG-OD1-asp-4	298 K	400 K
Range	[238-301]	[214-305]
angle	96.167	85.4003
$angle^2$	9400.55	7755.69
Std. Dev.	12.3476	21.5051

Tabla 7: Comparativa para CA-CB-CG-OD1 (ASP) a 298K y 400K

### 3 Resultados

CA-CB-CG-OD2-asp-4	298 K	400 K
Range	[59-115]	[35-124]
angle	-87.0568	-93.723
$angle^2$	7744.45	9237.34
Std. Dev.	12.8673	21.2916

Tabla 8: Comparativa para CA-CB-CG-OD2 (ASP) a 298K y 400K

Se puede observar cómo, en ambos casos, aumenta con la temperatura, tanto la desviación estándar, como el rango de ángulos alcanzado en los movimientos de torsión.

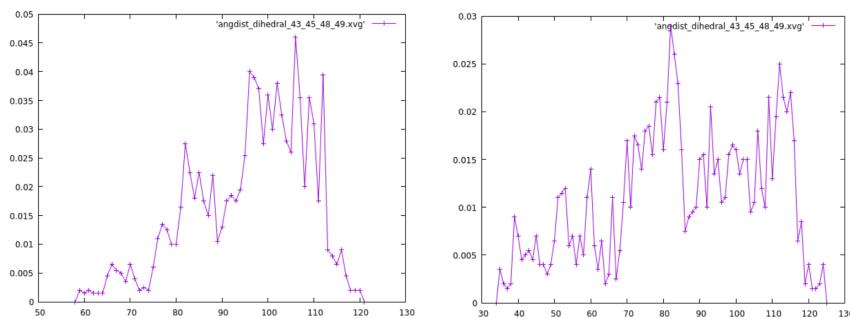


Figura 15: Diagrama de densidad del dihedro CA-CB-CG-OD1 en ASP a 298K ([dihedral-49-298.xvg](#)) y 400K ([dihedral-49-400.xvg](#))

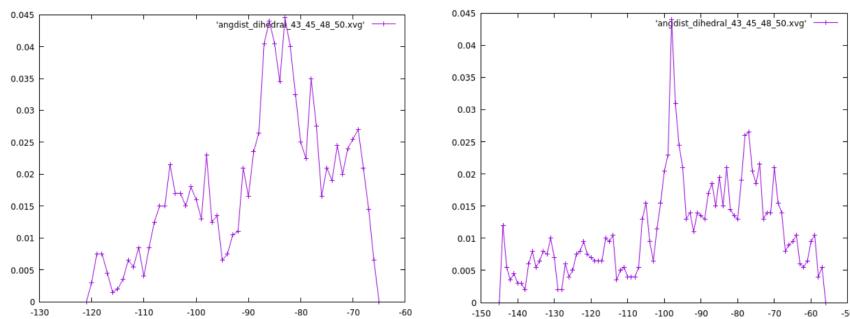


Figura 16: Diagrama de densidad del dihedro CA-CB-CG-OD2 en ASP a 298K ([dihederal-50-298.xvg](#)) y 400K ([dihederal-50-400.xvg](#))

En las gráficas de las Figuras 15 y 16, se puede ver que la distribución de densidad de ángulos de torsión no se aproxima a una campana de gauss, ya que en este caso se muestra la suma de dos ángulos. Sin embargo, se puede observar cómo las gráficas relativas a 400K (a la derecha en ambas Figuras), son más aplazadas y anchas. Esto se debe al aumento de los ángulos de torsión.

### 3 Resultados

#### 3.2.6. Radio de giro

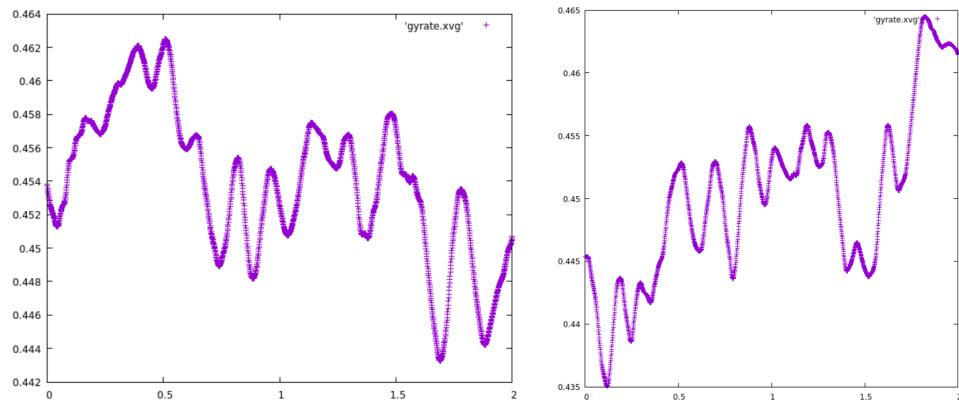


Figura 17: Radio de giro del tripéptido a 298K ([gyrate-298.xvg](#)) y 400K ([gyrate-400.xvg](#))

En la Figura 17 se realiza el cálculo del radio de giro de la molécula, a partir de la raíz cuadrada de la masa ponderada, cada 1 fs (desde el instante 0fs) en el intervalo de 2 ps. Se muestran 2001 medidas del radio de giro. Como se puede observar, a 400K el rango de valores para el radio de giro es mayor: [(0, 435) – (0, 465)]

#### 3.2.7. Velocidades de 5 átomos diferentes

Los átomos seleccionados son: CA, CB, CG, OD1 y OD2, en el aminoácido ASP (la Figura 6 muestra la posición de dichos átomos).

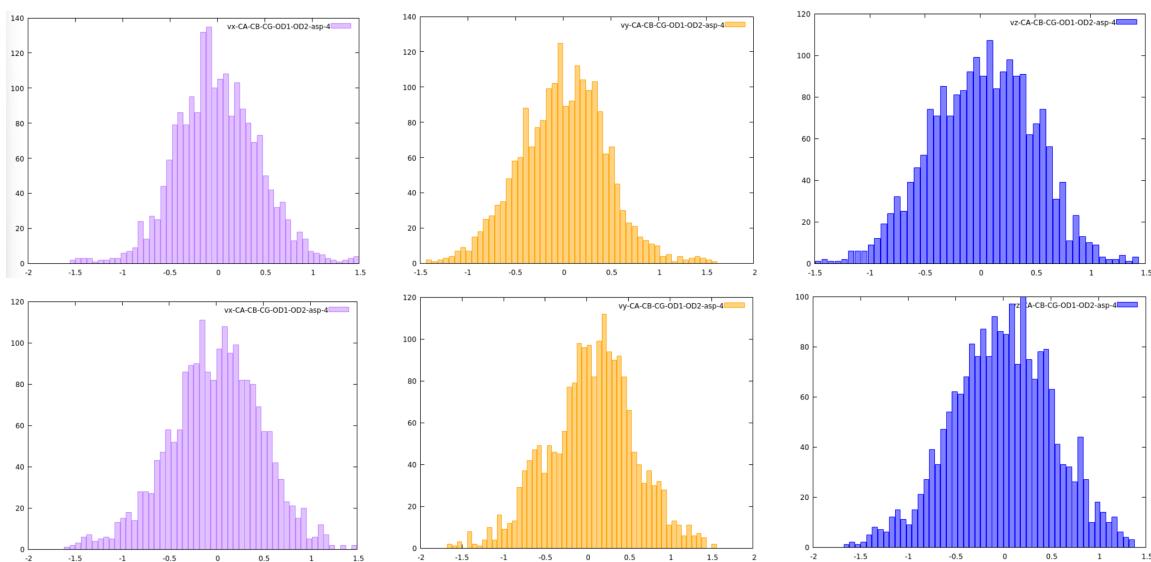


Figura 18: Histogramas de cada coordenada de la velocidad a 298K (fila 1) y 400K (fila 2)

### 3.3. Evaluar los valores medios de la energía cinética del sistema a ambas temperaturas y compararlas con los resultados esperados desde el punto de vista de la termodinámica estadística

Energía Cinética (KJ/mol)	298 K	400 K
Average	9745.28	13066.4
Err.Est.	14	50
RMSD	135.3	204.006
Tot-Drift	-68.0321	-183.223

Tabla 9: Energía cinética del sistema a 298K y 400K

Como la energía cinética es cuadrática con las componentes de la velocidad, al separar estas tres componentes, cada una contribuye con  $(1/2)k_B T$  a la energía cinética promedio en equilibrio térmico. Por lo tanto la energía cinética de la partícula es  $(3/2)k_B T$ , donde  $k_B$  es la constante de Boltzmann.

La constante de Boltzmann es:  $k = 1,380649 \times 10^{-23} J/K$ . Por otro lado, el número de Avogadro es:  $N_A = 6,02214076 \times 10^{23} mol^{-1}$ . Finalmente, la constante de Boltzmann puede expresarse de la siguiente manera:  $k_B = k \times N_A = 8,314462 \times 10^{-3} (\frac{KJ}{mol*K})$

Como se indica en el fichero [ard.g96](#) hay un total de 2634 átomos.

$$Ec_{298K} = 2634 \times \frac{3}{2}k_B(\frac{KJ}{mol*K}) \times 298(K) = 9789,42(\frac{KJ}{mol})$$

$$Ec_{400K} = 2634 \times \frac{3}{2}k_B(\frac{KJ}{mol*K}) \times 400(K) = 13140,16(\frac{KJ}{mol})$$

Los resultados según la termodinámica estadística son muy cercanos a los datos obtenidos en la Tabla 9

### 3.4. Extra: Simulación de 500 ps a 298K con exportación cada 10 fs

En la fase de ejecución entra la topología `.top`: ([ard.top](#)), el fichero con las posiciones y las velocidades `.g96`: [ard.g96](#) y, por último, el archivo de parámetros `.mdp`: [runNVT.mdp](#).

Tabla 10: Fichero de parámetros de frecuencia de actualización de datos: [runNVT.mdp](#)

```
; RUN CONTROL PARAMETERS
dt          = 0.0005 ; 0.5 fs
nsteps      = 1000000 ; 500 ps
nstcomm     = 1000    ; number of steps for
                      center of mass motion removal
; OUTPUT CONTROL OPTIONS
nstxout     = 20     ; Save coordinates every 10 fs
nstvout     = 20     ; Save coordinates every 10 fs
nstlog      = 20     ; Save coordinates every 10 fs
nstenergy   = 20     ; Save coordinates every 10 fs
; OPTIONS FOR ELECTROSTATICS AND VDW
; Temperature coupling
ref_t       = 298   ; Avobe fluid phase trasition
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel     = no     ; Should be the first equilibration
```

Se explica a continuación el detalle de la simulación:

Se consideran intervalos temporales de 0.5 fs con un total de 4000 pasos, lo que equivale a:  $1000000(steps) \times 0.5(fs) \times 10^{-3}(\frac{ps}{fs}) = 500(ps)$ . El registro de datos se planifica cada 10 fs, o lo que es lo mismo cada 20 pasos, durante los 500 ps de la simulación. Se realizarán, por tanto, 50000 anotaciones o puntos de registro.

### 3.4.1. Histograma para la temperatura.

El fichero siguiente permite obtener las gráficas de la Figura 19: [temp.xvg](#).

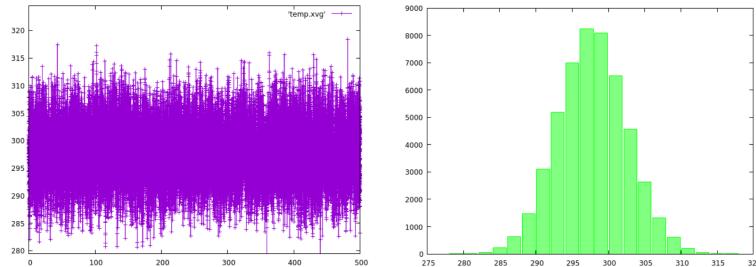


Figura 19: Diagrama e histograma de temperatura (298K)

En la Figura 19, se muestra a la izquierda en el eje de ordenadas la Temperatura absoluta (K) y en el eje de absisas el tiempo (ns). Como se puede ver la temperatura media es de 298K. A la derecha se muestra un histograma que representa en el eje de abcisas la Temperatura absoluta (K) y en el eje de ordenadas el número de átomos. Como se puede ver es una campana de Gauss centrada en 298K.

### 3.4.2. Histogramas para tres componentes de la velocidad de un átomo seleccionado.

El fichero siguiente permite obtener las gráficas de la Figura 20: [veloc-43.xvg](#).

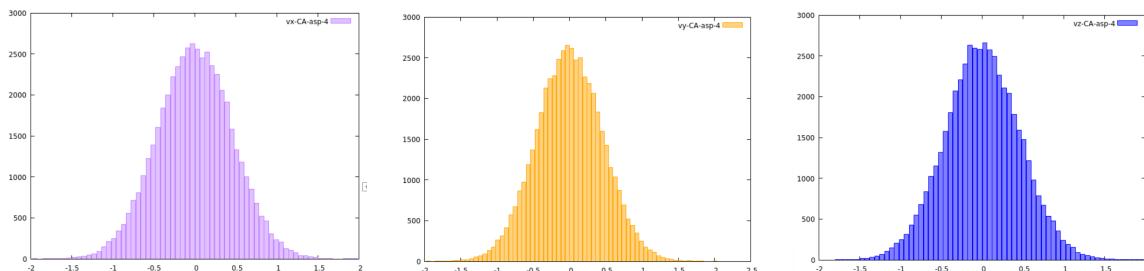


Figura 20: Histogramas de cada coordenada de la velocidad del átomo CA-asp4

Físicamente el módulo de la velocidad de una molécula  $v$ , es igual a la raíz de la suma de los cuadrados de las velocidades coordenadas de la partícula  $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$ , y como cada una de ellas siguen distribuciones gaussianas entonces  $v$  debe seguir una distribución de Maxwell-Boltzmann:  $f(v) = 4\pi \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2kT}}$ , donde  $m$  es la masa de la partícula y  $kT$  viene dada por el producto de la constante de Boltzmann y la temperatura absoluta [6].

### 3.4.3. Crear el diagrama de Ramachandran para los tres residuos.

Los ficheros siguientes permiten obtener las gráficas de la Figura 21: `ala-2.dat`, `arg-3.dat`, `asp-4.dat` y `rama.xvg`. Todos ellos almacenan los ángulos  $\phi$  y  $\psi$

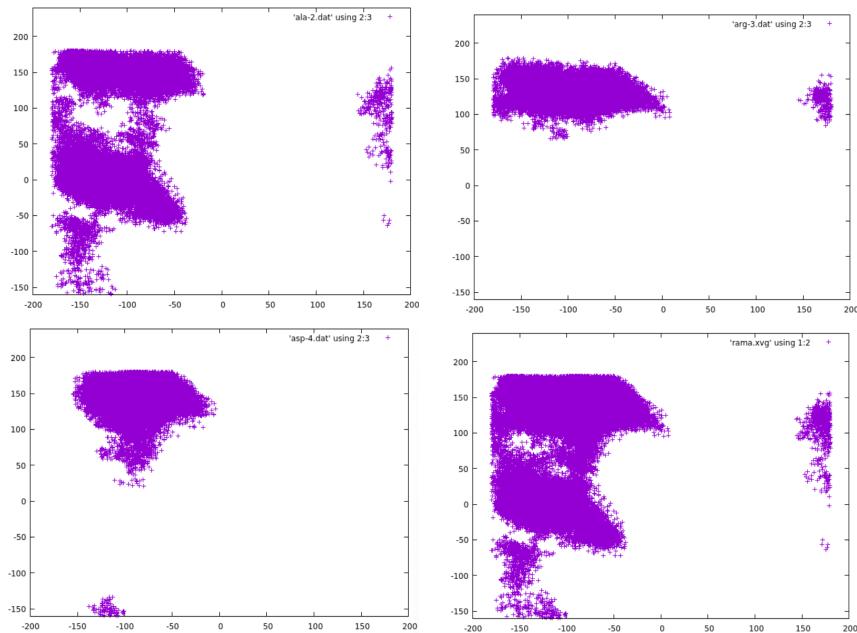


Figura 21: Ángulos  $\phi$  (abcisa) y  $\psi$  (ordenada) para ALA, ARG, ASP y todo el tripéptido a 298K

### 3.4.4. Identificar las conformaciones de probabilidad más alta.

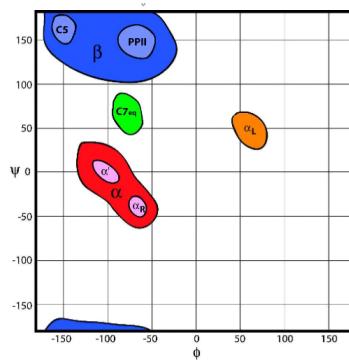


Figura 22: Conformaciones de referencia

En la Figura 21 se pueden ver los ángulos más frecuentes para cada uno de los aminoácidos y para el tripéptido (cuadrante abajo derecha), mientras que la Figura 22, muestra un diagrama de referencia de las conformaciones más habituales.

Se observan tres zonas en la Figura 21 que se pueden relacionar con alguna de las zonas de referencia de la Figura 22:

- La franja vertical que comprende el rango de ángulos  $\phi[(0) - (150)]$  no tiene ninguna ocurrencia. Los movimientos de torsión de los dihedros tienen prohibida esa zona. Corresponde con la zona  $\alpha_L$  (naranja) de las conformaciones de referencia.
- La superficie flanqueada por los ángulos  $\phi[(-180) - (0)];\psi[(180) - (100)]$  es compartida por todos los aminoácidos. Corresponde con la zona  $\beta$  (azul) de las conformaciones de referencia.
- La superficie flanqueada por los ángulos  $\phi[(-180) - (-20)];\psi[(70) - (-50)]$  es exclusiva de la alanina. Corresponde con la zona C7 (verde) y  $\alpha$  (roja) de las conformaciones de referencia.

## Referencias

- [1] Andy B. Yoo, Morris A. Jette y Mark Grondona. «SLURM: Simple Linux Utility for Resource Management». En: *Job Scheduling Strategies for Parallel Processing*. Ed. por Dror Feitelson, Larry Rudolph y Uwe Schwiegelshohn. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, págs. 44-60. ISBN: 978-3-540-39727-4.
- [2] Mark James Abraham et al. *GROMACS User Manual*. version 2016.4. Disponible en [www.gromacs.org](http://www.gromacs.org). Uppsala University. Sweden, jun. de 2017.
- [3] Thomas Williams, Colin Kelley y many others. *Gnuplot 5.2: an interactive plotting program*. <http://gnuplot.sourceforge.net/>. Jun. de 2018.
- [4] LLC Schrödinger y Warren DeLano. *PyMOL*. Ver. 2.4.0. 20 de mayo de 2020. URL: <http://www.pymol.org/pymol>.
- [5] F. J. Redondo Aniorte. *Bioinformatics - Tripeptide molecular modelling: ALA-ARG-ASP*. 2024. URL: <https://github.com/fjredondo/ard-tripeptide> (visitado 2024).
- [6] Wikipedia. *Distribución de Maxwell-Boltzmann — Wikipedia, La enciclopedia libre*. [Internet; descargado 7-febrero-2024]. 2024. URL: <https://w.wiki/9Tui>.

## Listado de comandos y tipos de fichero

- .g96** Puede ser un archivo de configuración inicial/final de GROMOS-96, o un archivo de trayectoria de coordenadas, o una combinación de ambos. [8](#), [9](#), [23](#)
- .gro** Archivo de estructura molecular que contiene una estructura molecular en formato Gromos87. Es generado a partir de un .pdb por el programa "gmx pdb2gmx". Las columnas contienen la siguiente información (de izquierda a derecha): - número de residuo (5 posiciones, entero) - nombre del residuo (5 caracteres) - nombre del átomo (5 caracteres) - número de átomo (5 posiciones, entero) - posición (en nm, x y z en 3 columnas, cada una de 8 posiciones con 3 decimales) - velocidad (en nm/ps, x y z en 3 columnas, cada una de 8 posiciones con 4 decimales). [5](#)
- .mdp** El archivo de parámetros de dinámica molecular (mdp) contiene toda la información sobre la simulación de dinámica molecular en sí, por ejemplo, time-step, number of steps, temperature, pressure etc. [6](#), [9](#), [23](#)
- .ndx** El archivo de índice GROMACS contiene algunos conjuntos de átomos definidos por el usuario. [11](#), [12](#)
- .pdb** Archivos de estructura molecular en el formato de archivo del banco de datos de proteínas. Las coordenadas están en Amstrongs. [5](#), [8](#), [10](#)
- .top** Archivo de topología molecular obtenido a partir de un archivo de estructura "pdb". Este archivo de topología contiene una descripción completa de todas las interacciones en su péptido. Es generado a partir de un .pdb por el programa "gmx pdb2gmx". [5](#), [9](#), [23](#)
- .tpr** Archivo de entrada de ejecución binaria portátil. Este archivo contiene la estructura inicial de su simulación, la topología molecular y todos los parámetros de simulación. Puede ser obtenido al combinar archivos de estructura molecular (.gro), de topología (.top), de parámetros (.mdp) y, opcionalmente, de índices (.ndx). Es la entrada del programa "gmx mdrun". [7](#), [9](#), [10](#)
- .trr** Los archivos con la extensión trr contienen la trayectoria de una simulación. Tiene un formato portable binario. Incluye todas las coordenadas, velocidades, fuerzas y energías que se indicaron en el fichero .mdp. Es la salida del programa "gmx mdrun". [10](#)
- .xvg** Produce una salida PostScript(tm), que es muy adecuada para su inclusión, por ejemplo.

Documentos LaTeX, pero también para otros procesadores de texto. [14](#), [15](#)

**gmx angle** Calcula la distribución de ángulos para varios ángulos o diedros. Con la opción -ov, se puede trazar el ángulo promedio de un grupo de ángulos en función del tiempo. [12](#)

**gmx distance** calcula distancias entre pares de posiciones en función del tiempo. Cada selección especifica un conjunto independiente de distancias para calcular. Cada selección debe constar de pares de posiciones y las distancias se calculan entre las posiciones 1-2, 3-4, etc. [11](#)

**gmx energy** Extrae componentes energéticos y los almacena en un archivo. [15](#)

**gmx gyrate** Calcula el radio de giro de una molécula y los radios de giro alrededor de los ejes x, y y z, en función del tiempo. Los átomos están explícitamente ponderados en masa. [11](#)

**gmx hbond** Utiliza la definición geométrica de los puentes de hidrógeno para definirlos en toda la estructura. El parámetro -num permite obtener una gráfica del número de puentes de hidrógeno en función del tiempo en el fichero de salida.. [13](#)

**gmx rama** Selecciona las combinaciones diédricas phi/psi de su archivo de topología y las calcula en función del tiempo. [13](#)

**gmx traj** Permite obtener medidas de temperatura (-ot), velocidad (-ov) dependiendo del parámetro solicitado. [14](#)