

# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE CIENCIAS**

### **POTENCIAL DE FUERZA MEDIA ASOCIADO A LA INTERACCIÓN ENTRE LA PROTEÍNA RIBOSÓMICA S8 Y SU APTÁMERO DE ARN**

#### **TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE FÍSICO**

#### **PROYECTO DE INVESTIGACIÓN**

**JUAN JOSÉ CEPEDA ESPÍN**  
juan.cepeda@epn.edu.ec

**DIRECTOR: MARCO VINICIO BAYAS REA, PH.D.**  
marco.bayas@epn.edu.ec

**Quito, junio 2023**

## DECLARACIÓN

Yo, Juan José Cepeda Espín, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

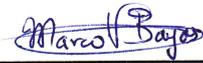


---

**JUAN JOSÉ CEPEDA ESPÍN**

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Juan José Cepeda Espín, bajo mi supervisión.



---

**Marco Bayas Rea, Ph.D.**

DIRECTOR

## **DEDICATORIA**

*Para el ser más maravilloso que tuve la oportunidad de compartir en este pequeño trayecto llamado vida. Mamá Rosa este trabajo es para ti. TE AMO.*

## **AGRADECIMIENTOS**

A Dios por ser esa fuente inagotable de luz.

A mis padres Ramiro y Juanita por la eterna comprensión mostrada y su infinito apoyo hacia todas las acciones emprendidas por mí. Sin ustedes me sería imposible haber cumplido este sueño.

A mis hermanos Santiago y Luis su ímpetu y semblante para lograr las metas me contagiaron a seguir creyendo en mí. Su ejemplo me es imposible de igualar.

A mis tíos, primos y familia en general que con sus grandes gestos de generosidad me brindaron respaldado.

A mis amigos por compartir la alegría hacia los logros académicos y personales. Además, les agradezco por su sincera amistad y por estar presente en los momentos más difíciles.

Un valioso agradecimiento para Tavy, sus ojos de admiración hacia mi trabajo contribuyeron a que yo pudiera concluirlo.

Finalmente, quiero agradecer a mi tutor Marco Bayas, por despertar en mí la fascinación hacia este universo denominado Biofísica. Poder describir fenómenos biológicos mediante la física causó en mi un gran augurio que me es difícil describirlo. También le agradezco por toda su paciencia así como su bondad para compartir conocimiento.

# Índice de Contenido

<b>Índice de Figuras</b>	<b>ix</b>
<b>Índice de Tablas</b>	<b>xii</b>
<b>Resumen</b>	<b>xiii</b>
<b>Abstract</b>	<b>xiv</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Aptámeros de ARN . . . . .	2
<b>2 Marco Teórico</b>	<b>5</b>
2.1 Bases Biofísicas . . . . .	5
2.2 Simulaciones de Dinámica Molecular . . . . .	5
2.2.1 Integración Numérica . . . . .	6
2.2.2 Energía en la Dinámica Molecular . . . . .	7
2.2.3 Control de Temperatura y Presión . . . . .	9
2.3 Dinámica Molecular Dirigida (SMD) . . . . .	10
2.4 Cálculo de la Energía Libre . . . . .	11
2.4.1 Muestreo “Umbrella” . . . . .	13

<b>3</b>	<b>Metodología</b>	<b>16</b>
3.1	NAMD: NAnoscale Molecular Dynamics . . . . .	16
3.2	Ensamble NPT del Complejo Molecular . . . . .	17
3.2.1	Minimización de la Energía del Sistema . . . . .	20
3.2.2	Calentamiento del Sistema . . . . .	21
3.2.3	Equilibración . . . . .	21
3.3	Simulaciones SMD . . . . .	22
3.4	Cálculo del Potencial de Fuerza Media PMF . . . . .	23
3.4.1	Método de Histogramas Ponderados . . . . .	24
3.5	Características Estructurales del Complejo . . . . .	26
3.5.1	Caracterización de las trayectorias de DM . . . . .	27
<b>4</b>	<b>Resultados</b>	<b>28</b>
4.1	Minimización y Calentamiento . . . . .	28
4.2	Equilibración . . . . .	29
4.3	Simulación SMD . . . . .	31
4.4	Potencial de Fuerza Media PMF . . . . .	39
<b>5</b>	<b>Conclusiones</b>	<b>48</b>
	<b>Referencias</b>	<b>58</b>
	<b>Anexos</b>	<b>60</b>
<b>A</b>	<b>Preparación del Sistema</b>	<b>60</b>
A.1	Script para separar las moléculas del complejo 4PDB . . . . .	60
A.2	Script para la generación de los archivos PSF y PDB del complejo . . . . .	61
A.3	Script para alinear el complejo en la dirección z . . . . .	62

A.4	Script de Solvatación del Complejo . . . . .	63
<b>B</b>	<b>Minimización de la Energía</b>	<b>64</b>
<b>C</b>	<b>Calentamiento del Sistema</b>	<b>66</b>
<b>D</b>	<b>Equilibración del Sistema</b>	<b>68</b>
<b>E</b>	<b>Simulación SMD</b>	<b>71</b>
<b>F</b>	<b>Equilibración del sistema para cada ventana</b>	<b>74</b>
F.1	Archivo de configuración . . . . .	74
F.2	Archivo Colvars . . . . .	76

# Índice de Figuras

1.1	Complejo aptámero-proteína. Código 4PDB. Figura generada con VMD [23].	4
2.1	Simulación SMD a velocidad constante para el caso unidimensional. El átomo “dummy” se encuentra de color salmón, mientras que el átomo SMD está de color cyan oscuro. A medida que el átomo dummy se mueve a velocidad constante el átomo SMD experimenta una fuerza que depende la distancia entre estos átomos. . . . .	11
3.1	Diagrama de flujo del proceso simulado para la obtención del PMF concerniente a la disociación del complejo proteína ribosomal S8 y su aptámero de ARN. . . . .	18
3.2	Complejo solvatado y neutralizado. La proteína ribosómica S8 se muestra como una superficie roja mientras que el aptámero de ARN se esquematiza de color azul. El complejo se encuentra inmerso en una caja con moléculas de agua. Los iones de sodio y cloro se muestran como esferas de color amarillo y celeste respectivamente. Figura generada con VMD. . . . .	20
3.3	Ubicación de los los átomos $C_{\alpha}$ de residuos 35 54 113 130 en el plano XY. En color cyan se resaltan los átomos SMD seleccionados, mientras que el punto cuadrangular de color amarillo representa el centro de masa del “backbone” de la proteína S8 dentro del plano XY. . . . .	22

4.1	Evolución temporal del RMSD y RG para la proteína ribosómica S8 y su aptámero de ARN obtenido durante el calentamiento a 310 K. . . . .	29
4.2	Evolución temporal de la desviación cuadrática media <i>rmsd</i> y radio de giro <i>RG</i> obtenidos para la proteína S8 y su aptámero de ARN durante la equilibración del sistema. . . . .	30
4.3	Evolución temporal de la coordenada de reacción $\xi$ (verde) y del ángulo $\theta$ (negro) durante la equilibración del sistema. El ángulo $\theta$ corresponde al ángulo formado por el eje $z$ y el vector que une los centros de masa entre las biomoléculas (proteína S8 y aptámero de ARN). . . . .	31
4.4	Evolución temporal de la energía del sistema durante la equilibración a 310K efectuada durante $10^6$ pasos de 1 <i>fs</i> . La línea punteada corresponde al valor medio de la energía del sistema durante los últimos 0.2 <i>ns</i> ( $\bar{E} = -111054.41kcal/mol$ ). . . . .	32
4.5	Snapshots de la simulación SMD efectuada para disociar el complejo aptámero de ARN-proteína S8 dentro de una caja de agua. En color cyan junto con $\uparrow$ se resaltan los átomos SMD que fueron halados en la dirección $+z$ durante la simulación, mientras que en la representación CPK se muestran las moléculas de agua confinadas entre las biomoléculas señaladas. . . . .	33
4.6	Parámetros obtenidos durante la simulación SMD del sistema a 310K. Evolución temporal de la distancia de los centros de masa a lo largo del eje $z$ (verde), radio de giro <i>RG</i> (naranja), distancia <i>rmsd</i> (azul) y ángulo $\theta$ que forma el vector que une los centros de masa con el eje $z$ (negro). El radio de giro <i>RG</i> y el <i>rmsd</i> fueron calculados para el backbone la proteína ribosomal S8. . . . .	35
4.7	Fuerza media de los átomos SMD en función de la coordenada de reacción $\xi$	36

4.8	Distancia entre los centros de masa de los residuos a) <i>CYT17</i> – <i>ALA114</i> y b) <i>ADE24</i> – <i>SER132</i> . . . . .	37
4.9	Distancia entre los centros de masa de los residuos a) <i>URA5</i> – <i>GLN8</i> y b) <i>ADE4</i> – <i>THR27</i> . . . . .	38
4.10	Evolución temporal de la coordenada de reacción $\xi$ durante la equilibración del sistema en las sendas ventanas. Los límites de cada ventana se los representa con líneas entrecortadas y el punto medio de cada ventana se lo resaltó con una línea continua de color amarillo. . . . .	40
4.11	Evolución temporal del <i>RMSD</i> calculada para el backbone de la proteína ribosomal S8 durante la equilibración del sistema molecular para cada ventana. . . . .	42
4.12	Evolución temporal del <i>RG</i> calculada para el backbone de la proteína ribosomal S8 durante la equilibración del sistema molecular para cada ventana. . . . .	43
4.13	Evolución temporal del ángulo $\theta$ formado por el eje z y el vector que une los centros de masa entre las biomoléculas (proteína y ARN) durante la equilibración de cada ventana. . . . .	44
4.14	Histogramas de la coordenada de reacción para cada ventana. El tiempo de equilibración empleado fue de 200 <i>ps</i> . descartando los primeros 20 <i>ps</i> . . . . .	45
4.15	Perfil de energía libre (PMF) obtenido durante el proceso de disociación forzado del complejo 4PDB inmerso en una caja de agua. El PMF encontrado se lo expresa en función de la coordenada de reacción $\xi$ . . . . .	47

# Índice de Tablas

4.1	Tiempo asociado a la ruptura de enlaces moleculares entre los grupos fosfato y amino del ARN y la proteína S8. Junto a los tiempos de ruptura se presenta el valor de la coordenada de reacción $\xi$ y la fuerza media de los átomos SMD para cada instante de ruptura. . . . .	39
4.2	Promedios estadísticos del radio de giro $RG$ y del ángulo $\theta$ junto con su desviación cuadrática media $\sigma$ encontrados durante la equilibración de cada ventana. Los promedios junto con $\sigma$ se encontraron descartando los primeros 20000 pasos. . . . .	41
4.3	Porcentaje de solapamiento entre ventanas aledañas con respecto al muestreo de $\xi$ . . . . .	46

## Resumen

El proceso de disociación del complejo formado por la proteína ribosómica S8 y su aptámero de ARN se estudió por medio de simulaciones de Dinámica Molecular. El muestreo de las configuraciones accedidas por el complejo durante el proceso de disociación forzada permitió encontrar el potencial de fuerza media PMF. En primer lugar, el complejo se colocó en una caja con moléculas de agua lo suficientemente grande para simular la disociación. Una vez que el sistema fue equilibrado, se efectuó la disociación al aplicar una fuerza externa con velocidad constante de  $0.25 \text{ (\AA/ps)}$  en la dirección  $z$ . Esta fuerza fue aplicada sobre los  $C_\alpha$  de los aminoácidos *THR35*, *LYS54*, *LYS113* y *SER130* de la proteína. La aplicación de la fuerza externa indujo la disociación del complejo en un tiempo de  $160 \text{ ps}$ . Para obtener el PMF, se definió la coordenada de reacción  $\xi$  como la distancia que une los centros de masa del ARN y de la proteína a lo largo del eje  $z$ . El muestreo de  $\xi$  se lo realizó utilizando 12 subintervalos (ventanas) comprendidos entre  $22.4$  y  $36.8 \text{ \AA}$ . El PMF, calculado mediante la ponderación de histogramas WHAM, reveló que el costo energético requerido para disociar el complejo fue de  $21.3 \text{ kcal/mol}$ . Finalmente, la existencia de regiones con valores constantes del PMF revelaron la presencia de contactos críticos en la formación del complejo. Los contactos críticos identificados fueron aquellos formados por los residuos de la proteína *ALA114*, *SER132*, *GLN80* & *THR27* con sus contrapartes en el ARN, *CYT17*, *ADE24*, *URA5* & *ADE4*, respectivamente.

## Abstract

The dissociation process of the complex between S8 ribosomal protein and its RNA aptamer was studied with Molecular Dynamics simulations. By sampling the configurations accessed by the complex during the forced dissociation, the corresponding potential of mean force (PMF) was determined. First, the complex was placed in a water box large enough to simulate the dissociation. After the system reached equilibrium, the dissociation was induced by applying an external force with constant speed of  $0.25 \text{ (\AA/ps)}$  in the z direction. Such force was applied on the  $C_\alpha$  of *THR35*, *LYS54*, *LYS113* and *SER130* aminoacids of the S8 ribosomal protein. The application of the external force induced the dissociation of the complex in  $160 \text{ ps}$ . In order to obtain the PMF, the reaction coordinate  $\xi$  was defined as the distance between the RNA and protein centers of mass along the z-axis.  $\xi$  was sampled on 12 intervals (windows) between  $22.4$  and  $36.8 \text{ \AA}$ . The PMF calculated by the weighted histogram analysis method WHAM, revealed that the energy cost required to dissociate the complex was  $21.3 \text{ kcal/mol}$ . Finally, the existence of regions of constant PMF revealed the presence of contacts critical for the formation of the complex. The identified critical contacts are formed by the protein residues *ALA114*, *SER132*, *GLN80* & *THR27* with their RNA counterpart *CYT17*, *ADE24*, *URA5* & *ADE4*, respectively.

# Capítulo 1

## Introducción

Los aptámeros son oligonucleótidos de cadena simple de ADN o ARN que presentan una estructura tridimensional definida la cual les permite enlazarse a ciertos blancos moleculares [1]. En 1990 tres grupos independientes de investigación mostraron la posibilidad de aislar moléculas de ARN y ADN capaces de reconocer distintos blancos por medio del proceso conocido como SELEX (Systematic Evolution of Ligands by EXponential Enrichment) [1–3]. Este método usa la química combinatoria para seleccionar ácidos nucleicos dentro de una librería de  $10^{15}$  componentes que puedan reconocer y unirse a moléculas específicas. Las cadenas de ácidos nucleicos derivados de la metodología SELEX se denominaron “aptámeros”.

A partir de la invención de la técnica SELEX [1, 3], se han producido aptámeros para distintos blancos entre los que se distinguen moléculas pequeñas, proteínas, virus, bacterias, células e incluso tejidos celulares [4, 5]. Los aptámeros se unen selectivamente a sus blancos con constantes de disociación en el orden de los picomoles ( $pM$ ) hasta los nanomoles ( $nM$ ) [5]. El reconocimiento de un blanco por parte de su aptámero se da primordialmente por interacciones electrostáticas y enlaces hidrógeno, los cuales surgen por diferentes con-

tactos a lo largo de las superficies interactuantes [6].

El carácter específico de las interacciones asociadas a los aptámeros ha incentivado el desarrollo de herramientas moleculares útiles en el diseño de biosensores [7] o en biomedicina [8]. Existen estudios para el desarrollo de aplicaciones concernientes al área de antiinfección e inmunoterapia [9], o la acción de aptámeros como agentes de difusión de biomarcadores en células y tejidos [10]. Además, los aptámeros no requieren de ensayos con animales o una respuesta inmune para su reproducción; de hecho, el protocolo *in vitro* SELEX muestra ventajas con respecto a la generación de anticuerpos al no emplear sustancias tóxicas que afectarían al animal [11].

## **1.1 Aptámeros de ARN**

Los aptámeros de ARN son oligonucleótidos de cadena simple que no suelen formar cadenas dobles extensas. La especificidad de sus estructuras les permiten distinguir varios blancos moleculares como las proteínas. Técnicas experimentales como la cristalografía de Rayos X y la resonancia magnética nuclear (NMR) suministran información adicional sobre las regiones o dominios de unión entre aminoácidos y nucleótidos dentro de los complejos ARN-proteína. El conjunto de datos generados por las técnicas experimentales mencionadas permiten determinar el tipo de interacción intermolecular conjuntamente de los aminoácidos involucrados en los dominios RBP (RNA Binding Protein) [12, 13].

La generación de aptámeros a través de la técnica SELEX ha permitido la identificación de mecanismos de unión y funciones específicas del ARN, las cuales requieren que un conjunto de ácidos nucleicos posean estructuras definidas [1, 3]. Las estructuras de los aptámeros de ARN muestran que ciertas secuencias de ácidos nucleicos se enlazan selectivamente a re-

giones dentro de las proteínas siendo capaces de reconocer lugares cargados positivamente, incluso con ausencia de ligandos de ácidos nucleicos [14].

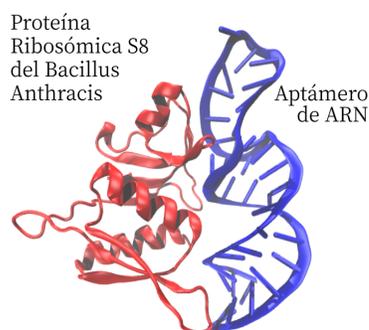
Los complejos proteína-ARN se caracterizan por poseer contactos, principalmente entre los aminoácidos con carga positiva y los ácidos nucleicos con carga negativa [15]. Estas interacciones moleculares incluyen enlaces hidrógeno, interacciones iónicas, de Van der Waals e interacciones de apilamiento  $\pi\pi$  [14, 16]. La especificidad de unión entre aptámeros de ARN y proteínas se debe primordialmente a enlaces hidrógeno y puentes salinos. Los componentes de la proteína que presentan mayor afinidad son los residuos con carga, tanto positiva (Arginina y Lisina), como negativa (Asp y Glu)[17].

Estudios sobre el reconocimiento de proteínas por parte de aptámeros de ARN se han centrado principalmente en los ribosomas de la bacteria *Escherichia Coli* [17]. Las proteínas ribosomales dentro de esta bacteria poseen un rol importante en la formación de sitios funcionales y en los estados iniciales del ensamblaje ribosomal [17]. La proteína ribosómica S8 de la bacteria *E. Coli* tiene un papel crucial en la formación de la subunidad ribosomal 30S, ya que interacciona con el grupo fosfato de la hélice H21 del ARN ribosomal 16S. Por lo tanto, el estudio del complejo (proteína S8 - aptámero de ARN) puede brindar información adicional sobre la organización y funcionamiento de los ribosomas dentro de la bacteria *E. Coli*. Puesto que la proteína S8 es esencial para la síntesis de proteínas, el estudio del complejo puede ser de utilidad en el diseño de nuevos fármacos para tratar infecciones ocasionadas por esta bacteria.

Una forma de caracterizar el mecanismo de interacción en un complejo molecular es la determinación de su potencial de fuerza media (Potential of Mean Force, PMF). El PMF de un sistema molecular es el potencial asociado al promedio de las fuerzas entre los gru-

pos moleculares de interés [18, 19]. Adicionalmente, el PMF cuantifica la energía libre de un sistema en función de cierta coordenada, denominada coordenada de reacción [20] y puede ser obtenido mediante técnicas especiales de muestreo como: fuerza de sesgo adaptativo (Adaptative Biasing Force, ABF) [21] o la igualdad de Jarzynski [20]. El PMF de un sistema puede obtenerse mediante la aplicación de dichas técnicas a las trayectorias de dinámica molecular correspondientes.

En este trabajo se utilizará como modelo de interacción aptámero-receptor aquel correspondiente al archivo pdb 4PDB (Fig. 1.1). El complejo escogido corresponde al aptámero de la proteína ribosómica S8, sobre el cual existe información estructural [22] al ser un aptámero representativo de ARN. Partiendo de la estructura tridimensional del complejo, se determinará el potencial de fuerza media para esta interacción mediante el análisis de trayectorias de dinámica molecular.



**Figura 1.1:** Complejo aptámero-proteína. Código 4PDB. Figura generada con VMD [23].

## Capítulo 2

### Marco Teórico

#### 2.1 Bases Biofísicas

La afinidad entre grupos moleculares dentro de sistemas biológicos se debe principalmente a interacciones no covalentes asociadas a las fuerzas de atracción electrostáticas entre iones y grupos moleculares. A pesar de que los enlaces no covalentes son débiles, el efecto acumulativo de los mismos brinda cierta estabilidad a los complejos moleculares [24–26]. Las interacciones no covalentes predominantes dentro de los sistemas biológicos se deben primordialmente a enlaces iónicos, enlaces de hidrógeno e interacciones de Van der Waals. Los detalles de estas interacciones a nivel atómico pueden estudiarse con ayuda de las estructuras de los complejos correspondientes, determinadas principalmente con cristalografía de rayos X y con simulaciones de dinámica molecular de estas estructuras [27].

#### 2.2 Simulaciones de Dinámica Molecular

Estas simulaciones describen la dinámica de un sistema de partículas que interactúan entre sí mediante la resolución de sus ecuaciones de movimiento. Las interacciones entre

partículas son definidas a través de los llamados campos de fuerza, los cuales son determinados con base en datos experimentales [28, 29]. Las simulaciones de dinámica molecular permiten el análisis del comportamiento promedio de un sistema con muchas partículas, el cual, se desarrolla bajo la suposición de que la hipótesis ergódica se cumpla [30, 31]. Estas simulaciones usualmente toman pasos de 1 femtosegundo [28, 32]. Particularmente, las simulaciones de DM permiten estudiar diversas funciones biológicas de macromoléculas aportando con información que sería difícil de alcanzar de forma experimental [27, 33]. Así, existen notables aplicaciones de la Dinámica Molecular como el análisis referente a cambios conformacionales [34], diferencias en energía libre [20] y el estudio con respecto a la estabilidad estructural y dinámica de diversos complejos moleculares [29].

### 2.2.1 Integración Numérica

Las trayectorias de un conjunto de partículas en DM están determinadas por la integración numérica de las ecuaciones de movimiento [29, 35]. Al conocer las condiciones iniciales  $\vec{r}_i(0)$  y  $\dot{\vec{r}}_i(0)$  del sistema de estudio, se puede determinar las posiciones  $\vec{r}_i(t)$  y velocidades  $\dot{\vec{r}}_i(t)$  para un tiempo  $t$ . Dado el potencial de interacción  $U$ , las ecuaciones de movimiento de un sistema de partículas son:

$$-\nabla_{\vec{r}_i} U(\vec{r}_i) = m_i \ddot{\vec{r}}_i \quad (2.1)$$

La integración numérica de la ecuación (2.1) se puede efectuar por medio del algoritmo de Verlet [30, 36, 37], el cual, discretiza las posiciones y velocidades de los átomos mediante:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + (\Delta t)^2 \vec{a}_i(t) \quad (2.2)$$

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2 \Delta t} \quad (2.3)$$

Donde,  $\Delta t$  es el intervalo de tiempo empleado para integrar las ecuaciones (2.2, 2.3). En el algoritmo de Verlet las posiciones y velocidades de las partículas se actualizan en diferentes tiempos, lo cual conduce a la acumulación de errores en simulaciones largas. Además, este algoritmo no es de utilidad para una simulación que se inicializa únicamente con las posiciones de las partículas. Alternativamente, se puede utilizar la version velocidad del algoritmo de verlet. El software NAMD emplea la version velocidad de este algoritmo donde la posición y velocidad de cada partícula se actualizan mediante [38]:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)(\Delta t)^2 \quad (2.4)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{1}{2}\left[\vec{a}_i(t + \Delta t) + \vec{a}_i(t)\right] \quad (2.5)$$

Matemáticamente, los dos algoritmos son equivalentes con la ventaja de que en este último las posiciones y velocidades se calculan para el mismo instante de tiempo. De esta manera, teniendo en cuenta las condiciones iniciales del sistema junto con las interacciones interatómicas y la aplicación iterativa de las ecuaciones (2.4) & (2.5) la evolución de un sistema biomolecular puede ser simulado.

### 2.2.2 Energía en la Dinámica Molecular

La energía potencial  $U$  en DM está representada a través de campos de fuerza en los cuales se expresa de manera explícita la interacción entre los átomos del sistema [29, 39, 40]. Así, la energía potencial, se expresa como una función  $U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$  que describe los enlaces, ángulos e interacciones electrostáticas y de Van der Waals de un átomo con respecto a los demás.

$$U(\vec{r}_1, \dots, \vec{r}_N) = \underbrace{\sum_{enlace} k_i^{enlace} (r_i - r_0)^2}_{U_{enlace}} + \quad (2.6)$$

$$\underbrace{\sum_{angulos} k_i^{angulo} (\theta_i - \theta_0)^2}_{U_{angulos}} + \quad (2.7)$$

$$\underbrace{\sum_{torsion} k_i^{angulo} [1 + \cos(n_i \phi_i + \delta_i)] + \sum_{torsion} k_i^{angulo} (\phi_i - \delta_i)^2}_{U_{torsion}} + \quad (2.8)$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{U_{vdw}} + \quad (2.9)$$

$$\underbrace{\sum_i \sum_{j \neq i} \frac{q_i q_j}{\varepsilon r_{ij}}}_{U_{Coulomb}} \quad (2.10)$$

El término  $U_{enlace}$ , corresponde a la interacción armónica entre pares de átomos  $(i, j)$  con enlace covalente. Donde  $r_i - r_0$  es la distancia que un átomo se ha desplazado con respecto al equilibrio. El segundo término  $U_{angulos}$ , está asociado a los ángulos de enlace entre tres átomos.  $U_{angulos}$  describe el potencial vibracional angular en la tripleta de átomos  $(i, j, k)$ , donde  $\theta$  es el ángulo entre los vectores  $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$  y  $\vec{r}_{kj} = \vec{r}_j - \vec{r}_k$ ,  $\theta_0$  es el ángulo de equilibrio y  $k_i^{angulo}$  es la constante angular. El tercer y cuarto término están asociados a los ángulos de torsión. El primero de ellos corresponde al potencial de torsión, este describe el comportamiento angular armónico entre los planos formados por los tres primeros átomos de los enlaces consecutivos dentro de la cuádrupla  $(i, j, k, l)$ . El ángulo entre los planos  $(i, j, k)$  y  $(j, k, l)$  se lo denota con la letra  $\phi$ ,  $\delta$  es el ángulo de desfase y  $n$  es un número entero e indica la periodicidad. Mientras que el segundo término relacionado con el potencial de torsión, corresponde a un potencial sinusoidal para ángulos diedros impropios

donde  $\delta$  actúa como ángulo equilibrio. Las dos últimas contribuciones  $U_{vdW}$  y  $U_{Coulomb}$  son las contribuciones no enlazantes, las cuales están asociadas a la interacción entre pares mediante interacciones electrostáticas por medio del potencial Coulombiano y del potencial de Lennard-Jones. El potencial de Lennard-Jones se lo expresa en función del radio de van der Waals  $\sigma$  y de la profundidad del potencial  $\varepsilon$  [5, 28].

### 2.2.3 Control de Temperatura y Presión

En las simulaciones de Dinámica Molecular, una vez que se ha definido la energía de interacción, es importante considerar el tipo de ensamble estadístico que se va a muestrear conjuntamente del control que se ejercerá sobre ciertas variables termodinámicas. El control de los parámetros macroscópicos dentro de una simulación DM define el tipo de ensamble.

**Temperatura Constante.** Los métodos para controlar la temperatura de una simulación DM llevan el nombre de “termostatos”. En DM varios termostatos típicos han sido implementados como el de Berendsen [41], reescalamiento de velocidad [42] el termostato de Nosé–Hoover [43, 44] y la dinámica de Langevin [45]. Todos estos tienen en común en que la velocidad de las partículas se ajustan tal que la temperatura del sistema permanezca constante. En este trabajo se empleó la dinámica de Langevin implementada por NAMD [28]. La dinámica de Langevin reproduce la interacción del sistema con un reservorio de calor a cierta temperatura al agregar dos términos adicionales a la ecuación de movimiento (Ec. 2.1). Los términos adicionales corresponden a la fuerza de fricción ejercida por el medio y a una fuerza aleatoria.

**Presión Constante.** Los métodos para controlar la presión de una simulación DM llevan el nombre de “barostatos”, entre los cuales está el de Parrinello-Rahman [46], y el barostato

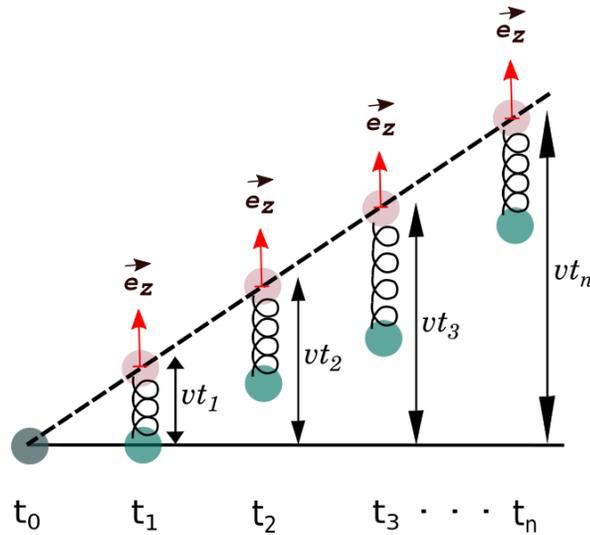
de Nosé-Hoover [47] . Los barostatos mantienen la presión del sistema en cierto valor al reescalar las posiciones de las partículas y el volumen de la celda simulada. En este trabajo se empleó la metodología híbrida denominada “Nosé-Hoover Langevin piston” para el control de presión y temperatura en el ensamble *NPT* la cual esta implementada por NAMD [28].

### 2.3 Dinámica Molecular Dirigida (SMD)

La Dinámica Molecular Dirigida (Steered Molecular Dynamics, SMD) es un subgrupo de simulaciones DM que permiten aplicar fuerzas a átomos específicos dentro de un sistema [48]. Mediante la aplicación de estas fuerzas se logra que un átomo permanezca en un lugar o, a su vez que este se mueva. Las simulaciones “SMD” permiten estudiar diferentes propiedades de enlace entre biomoléculas ante una manipulación externa [48]. Así, “SMD” figura como una herramienta útil para explorar eventos de unión/desunión asociados a las interacciones moleculares.

La fuerza externa aplicada permite acelerar procesos de transición entre estados. Esta fuerza puede ser aplicada de dos maneras, la primera es halando o empujando un grupo de átomos a velocidad constante y la segunda es cuando la fuerza empleada es constante. Para las simulaciones SMD a velocidad constante, es necesario aplicar la fuerza externa a ciertos átomos (átomos SMD) los cuales se encuentran unidos a otro grupo de átomos (átomos dummy) mediante un resorte virtual (Fig. 2.1). En este caso, los átomos SMD serán los  $C_\alpha$  de los residuos que están en contacto directo con el ARN.

En las simulaciones a velocidad constante, los átomos SMD perciben una fuerza que depende de las distancias entre un tiempo inicial  $t_0$  y un tiempo  $t$ . La fuerza que experimenta



**Figura 2.1:** Simulación SMD a velocidad constante para el caso unidimensional. El átomo “dummy” se encuentra de color salmón, mientras que el átomo SMD está de color cyan oscuro. A medida que el átomo dummy se mueve a velocidad constante el átomo SMD experimenta una fuerza que depende la distancia entre estos átomos.

el átomo SMD con respecto al dummy se determina mediante:

$$\vec{F} = -\nabla U \quad (2.11)$$

$$U = \frac{1}{2}k[v t - (\vec{r} - \vec{r}_0) \cdot \vec{n}]^2 \quad (2.12)$$

Donde,  $\vec{F}$  representa la fuerza externa,  $U$  es la energía potencial,  $k$  es la constante elástica,  $v$  velocidad de empuje,  $t$  tiempo,  $\vec{r}_0$  posición inicial del átomo SMD,  $\vec{r}$  posición del átomo SMD al tiempo  $t$  y  $\vec{n}$  es la dirección de empuje.

## 2.4 Cálculo de la Energía Libre

El cambio en energía libre entre estados que se encuentran en equilibrio se fundamenta en la igualdad de Jarzynski [49]. La igualdad de Jarzynski (Ec. 2.13) relaciona el cambio en energía libre  $\Delta F$  entre dos estados de equilibrio, con el trabajo  $W$  realizado en el sistema

durante un proceso de no equilibrio [49, 50].

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F} \quad (2.13)$$

Donde  $\beta = 1/k_B T$  con  $T$  temperatura del sistema y  $k_B$  constante de Boltzmann. La igualdad de Jarzynski está definida dentro de un ensamble canónico, sin embargo esta igualdad puede ser extendida para un ensamble NPT, en el cual la energía libre asociada a dicho ensamble es la energía libre de Gibbs  $G$  [51].

La igualdad de Jarzynski se la extiende a simulaciones de dinámica molecular al establecer de forma explícita un parámetro  $\lambda$  el cual relaciona la topografía molecular entre los estados inicial y final de un sistema [52]. Generalmente,  $\lambda$  se lo considera como un parámetro generalizado que permite cuantificar el progreso de un sistema a lo largo de un trayecto entre estos estados [52]. Asumiendo que la energía  $E$  del sistema depende de  $\lambda$  tal que  $E(\lambda)$  transita suavemente de  $E_0$  a  $E_1$ , la energía libre  $A(\lambda)$  se define como:

$$A(\lambda) = -\frac{1}{\beta} \ln Q \quad (2.14)$$

Teniendo en cuenta todos los grados de libertad respecto a  $\lambda$ , la distribución de probabilidad se expresa como [53]:

$$Q(\lambda) = \frac{\int \delta[\lambda(r) - \lambda] \exp\{-\beta E\} \mathbf{d}^N r}{\int \exp\{-\beta E\} \mathbf{d}^N r} \quad (2.15)$$

$Q(\lambda)d\lambda$  se interpreta como la probabilidad de encontrar al sistema en un pequeño intervalo alrededor de  $\lambda$ . En simulaciones de dinámica molecular, la integral en el espacio de fases de la Ec. (2.15) es difícil de lograrla. No obstante, si el sistema es ergódico, y al tener en

cuenta la ocurrencia de  $\lambda$  dentro de un intervalo ( $\rho(\lambda)$ );  $Q(\lambda)$  es igual a [53]:

$$P(\lambda) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \rho[\lambda(t')] dt' \quad (2.16)$$

En otras palabras, el promedio en el ensamble  $Q(\lambda)$  es igual al promedio temporal  $P(\lambda)$ .

### 2.4.1 Muestreo “Umbrella”

El muestreo “Umbrella” o Umbrella Sampling (US), es una técnica computacional asociada al muestreo conformacional de sistemas moleculares empleada para el cálculo de la energía libre en procesos de asociación-disociación. Durante estos procesos, diferentes trayectorias pueden explorarse. Así, si se establece un “potencial artificial” que guíe al sistema biomolecular por ciertos trayectos el correspondiente cambio en energía libre puede ser obtenido. La técnica US añade este potencial, conocido también como potencial de sesgo, a la energía potencial del sistema [54]. El potencial de sesgo se sitúa en múltiples regiones a lo largo del camino de disociación para, posteriormente realizar un muestreo conformacional en intervalos consecutivos conocidos como “ventanas” [53]. Precisamente, el camino de disociación define la coordenada de reacción  $\xi$ , siendo  $\xi$  un parámetro continuo que distingue cambios conformacionales dentro del sistema molecular.

Generalmente, el potencial de sesgo empleado en el muestreo US es un potencial armónico de la forma:

$$w_i = k_i(\xi - \xi_i)^2 \quad (2.17)$$

Donde el subíndice  $i$  indica la ventana en la que se aplicará potencial de sesgo,  $\xi_i$  es la posición referencial dentro de la ventana  $i$ , y  $k_i$  es la constante elástica empleada para mantener el sistema de estudio dentro de cada ventana. Para cada ventana el potencial es escogido de

forma adecuada, tal que, este fuerce al sistema a encontrarse en ciertas posiciones deseadas con respecto a  $\xi$  [54]. El potencial  $w_i$  se añade a la energía potencial del sistema mediante [53]:

$$U'(r) = U(r) + w_i(\xi) \quad (2.18)$$

La distribución de probabilidad no sesgada del sistema se calcula al integrar sobre todo el espacio teniendo en cuenta los  $N$  grados de libertad con respecto a  $\xi$  [53].

$$P_i(\xi) = \frac{\int \exp\{-\beta U(r)\} \delta[\xi(r) - \xi] \mathbf{d}^N r}{\int \exp\{-\beta [U(r)]\} \mathbf{d}^N r} \quad (2.19)$$

Las simulaciones DM proveen una distribución de probabilidad conocida también como distribución sesgada a lo largo de la coordenada de reacción [53].

$$P'_i(\xi) = \frac{\int \exp\{-\beta [U(r) + w_i(\xi)]\} \delta[\xi(r) - \xi] \mathbf{d}^N r}{\int \exp\{-\beta [U(r) + w_i(\xi)]\} \mathbf{d}^N r} \quad (2.20)$$

La distribución de probabilidad no sesgada se relaciona con la distribución de probabilidad sesgada por medio de [53]:

$$P_i(\xi) = P'_i(\xi) \exp\{\beta w_i(\xi)\} \langle \exp\{-\beta w_i(\xi)\} \rangle \quad (2.21)$$

A partir de la Ec. (2.21), dado que se tiene una función de distribución respecto a  $\xi$ , la energía libre puede ser evaluada mediante la ec. (2.14) [53]. La energía libre  $A_i(\xi)$  obtenida, también se la denomina como el Potencial de Fuerza Media “Potential of Mean Force” (PMF) [54].

$$\begin{aligned} A_i(\xi) &= -\frac{1}{\beta} \ln P_i(\xi) - w_i(\xi) - \frac{1}{\beta} \ln \langle \exp\{-\beta w_i(\xi)\} \rangle \\ &= -\frac{1}{\beta} \ln P'_i(\xi) - w_i(\xi) + G_i \end{aligned} \quad (2.22)$$

Si se tiene en cuenta dos estados (etiquetados como 1 y 0), la diferencia en energía libre entre estos es  $\Delta A(\xi) = A_1(\xi) - A_0(\xi)$ , la cual, al considerar la relación encontrada en la Ec. (2.22) se escribe como:

$$\Delta A(\xi) = -\frac{1}{\beta} \ln \left( \frac{P'_1}{P'_0} \right) - w_1(\xi) + w_0(\xi) - \frac{1}{\beta} \ln \left( \frac{\langle \exp\{-\beta w_1(\xi)\} \rangle}{\langle \exp\{-\beta w_0(\xi)\} \rangle} \right) \quad (2.23)$$

El término  $G_i$  de la Ec. (2.22) debe ser calculado. Para ello es necesario tener en cuenta que las curvas de energía libre  $A_i(\xi)$  asociadas a cada ventana necesitan ser combinadas.  $G_i$  está asociado con el potencial de sesgo, y es el encargado de conectar las curvas de energía libre  $A_i(\xi)$  de cada ventana. Así, teniendo en cuenta la distribución de probabilidad no sesgada  $P(\xi)$ ,  $G_i$  puede ser calculado mediante [53]:

$$\exp\{-\beta G_i\} = \langle \exp\{-\beta w_i(\xi)\} \rangle \quad (2.24)$$

$$= \int P(\xi) \exp\{-\beta w_i(\xi)\} d\xi \quad (2.25)$$

La Ec. (2.24) se considera como un caso límite del teorema de Jarzynski [53]. Por medio de la metodología de perturbación de energía libre FEP (Free Energy Perturbation) [55] se extiende la igualdad de Jarzynski; tal que, el promedio exponencial del potencial  $w$  es semejante al promedio exponencial del trabajo  $W$ .

## Capítulo 3

### Metodología

#### 3.1 NAMD: NANoscale Molecular Dynamics

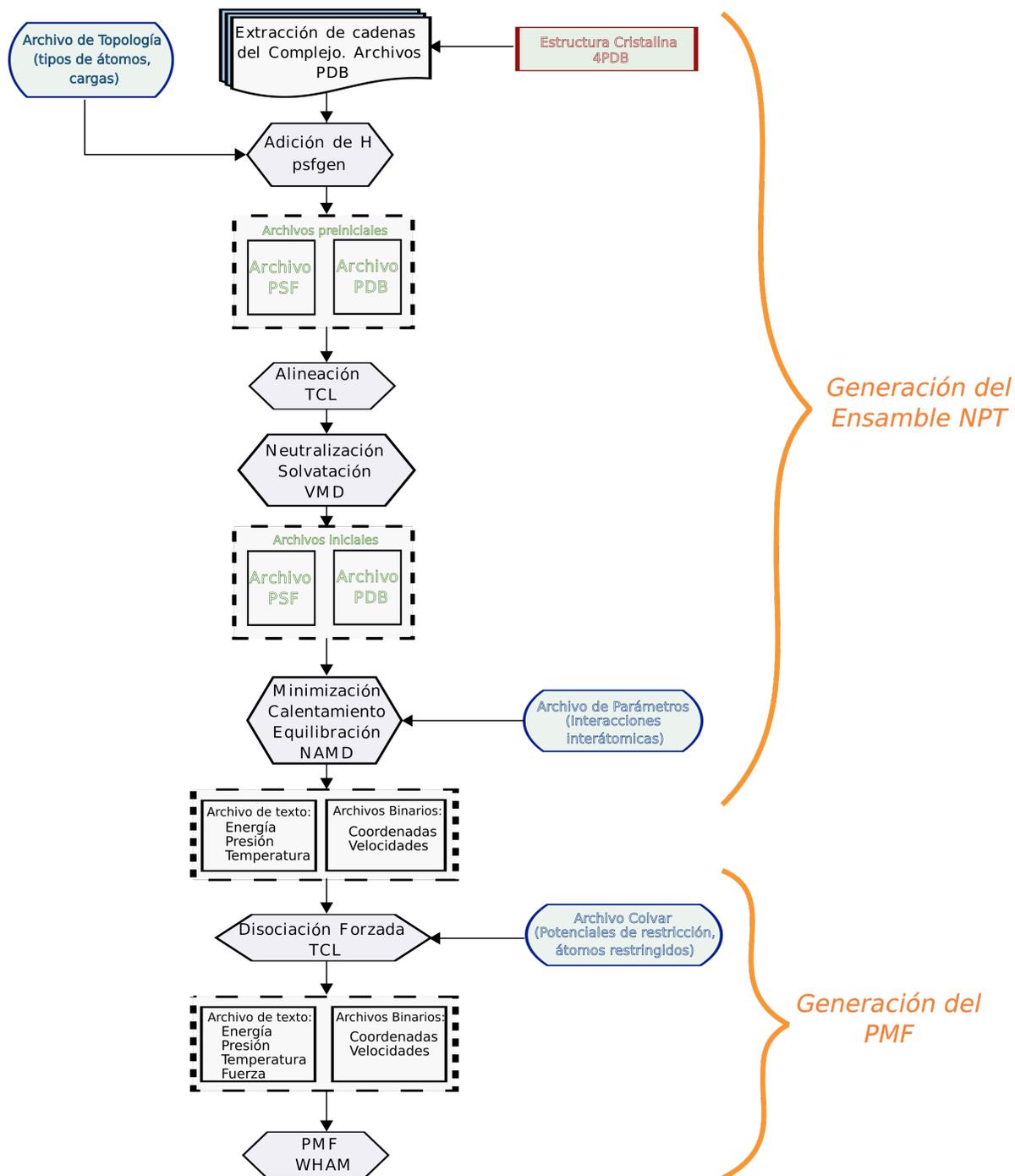
NAMD (NANoscale Molecular Dynamics) [56] es un software de acceso libre que permite realizar simulaciones MD de alto rendimiento para distintos sistemas biomoleculares [29]. Está escrito en C++ y se basa en el sistema de programación en paralelo Charm++, el cual permite balancear las tareas realizadas entre procesadores [29, 56]. Este software trabaja con campos de fuerza de simulación biomolecular como AMBER o CHARMM. En este trabajo se utilizó el campo de fuerza CHARMM 27 [57] junto con la versión NAMD 2.15alpha2 para Linux x86-64 multicore. NAMD junto con el programa de visualización y modelamiento molecular VMD (Visual Molecular Dynamics) [23], ejecutan varias actividades como la configuración de simulaciones, análisis de trayectoria o la generación de archivos de coordenadas. Para ejecutar una simulación de Dinámica Molecular, el programa NAMD requiere los siguientes archivos:

- Archivo de coordenadas en formato “pdb” (protein data bank) [58]. Este archivo contiene las coordenadas de todos los átomos del complejo molecular.

- Archivo PSF (protein structure file). Archivo que contiene información estructural del complejo molecular. En este archivo se detallan las características sobre los tipos de átomos, enlaces, ángulos, ángulos dihedros, torsiones impropias y términos cruzados.
- Archivo de parámetros de campo de fuerza. Este archivo contiene información concerniente a las fuerzas que experimentan los átomos del sistema de estudio. En el presente trabajo se empleó el campo de fuerza CHARMM27 [57].
- Archivo de topología de campo de fuerza. En este archivo se define las características usadas por el archivo de campo de fuerza. Dentro de este archivo se define la manera en que los átomos se encuentran conectados junto con la carga y el tipo de enlace que hay entre ellos. Además, este archivo contiene coordenadas internas las cuales permiten la asignación de coordenadas a los átomos de hidrógeno y otros átomos faltantes en el archivo de coordenadas.
- Archivo de configuración. Este archivo contiene todas las especificaciones referente a las opciones que NAMD debe emplear durante la ejecución de la simulación. En este archivo se escribe de forma explícita el archivo de parámetros, el control de temperatura, las condiciones de contorno entre otras restricciones.

## 3.2 Ensamble NPT del Complejo Molecular

La figura 3.1 muestra esquemáticamente el procedimiento para obtener el ensamble NPT del complejo. La estructura cristalina del complejo estudiado se la obtuvo del Protein Data Bank con código pdb: 4PDB. La obtención del ensamble NPT consta de 4 etapas.



**Figura 3.1:** Diagrama de flujo del proceso simulado para la obtención del PMF concerniente a la disociación del complejo proteína ribosomal S8 y su aptámero de ARN.

### **Etapa I. Redefinición de los elementos del complejo**

A partir del archivo 4PDB se identificó los segmentos correspondientes a la proteína y a la cadena de ARN. La separación y escritura de las moléculas en archivos *pdb* se las efectuó mediante la ejecución del script en tcl *extract* (Anexo A.1). De esta manera se generaron los archivos *arn.pdb* y *protein.pdb*.

### **Etapa II. Generación de archivos *psf* para las biomoléculas**

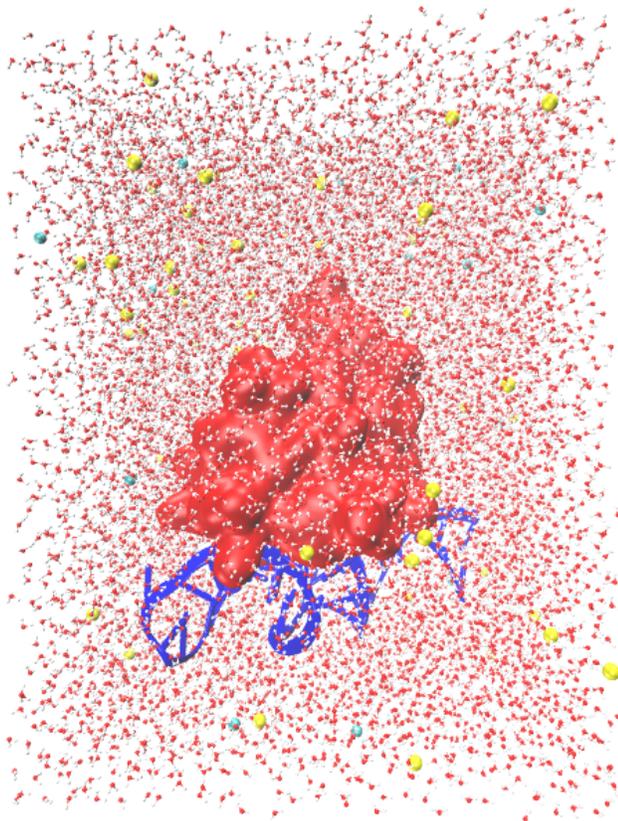
Con los archivos de coordenadas independientes de la proteína y del ARN, se procedió a generar los archivos PSF y PDB del complejo. Los archivos PSF y PDB cumplen con la nomenclatura y especificaciones del campo de fuerzas CHARMM 27. La generación de los nuevos archivos PSF y PDB se logró al ejecutar el script *generator* (Anexo A.2).

### **Etapa III. Solvatación del complejo**

Con la finalidad de simular un medio fisiológico, se colocó el complejo dentro de una caja con moléculas de agua lo suficientemente grande tal que, durante el proceso de disociación, las biomoléculas permanezcan dentro de dicha caja. Mediante el paquete *solvate* del programa VMD, se situó al complejo en una caja con moléculas de agua, donde los átomos más apartados del complejo distan en 10 Å en las direcciones  $x$  y; mientras que en la dirección  $+z$  se estipuló un margen de 25 Å. Previamente a la solvatación, el complejo fue alineado con respecto al eje  $z$ . Para esto, el vector que une los centros de masa de la proteína y del ARN fue orientado en la dirección  $+z$ . En los anexos A.3, A.4 se encuentran los “scripts” empleados para alinear y solvatar al complejo respectivamente. Como resultado se obtuvieron los archivos *aptamer\_alg\_wb.psf* y *aptamer\_alg\_wb.pdb*.

#### **Etapas IV. Neutralización del sistema**

Mediante el paquete *autoionize* de VMD, se añadieron iones de  $Na^+$  y  $Cl^-$  al complejo solvatado de modo que la carga final fue cero. Los iones fueron añadidos con una concentración de 100 *mM*. El sistema resultante se muestra en la Figura 3.2.



**Figura 3.2:** Complejo solvatado y neutralizado. La proteína ribosómica S8 se muestra como una superficie roja mientras que el aptámero de ARN se esquematiza de color azul. El complejo se encuentra inmerso en una caja con moléculas de agua. Los iones de sodio y cloro se muestran como esferas de color amarillo y celeste respectivamente. Figura generada con VMD.

#### **3.2.1 Minimización de la Energía del Sistema**

El método de minimización empleado por NAMD es el denominado método del gradiente conjugado [28]. El proceso de minimización de la energía para el sistema de estudio se

realizó en 1000 pasos. Las coordenadas de los átomos del sistema se salvaron en un archivo binario cada 100 pasos, mientras que la energía se salvó cada 50 pasos. La restricción del sistema consistió en fijar los fósforos con residuos 1, 10 y 19 (Anexo B).

### 3.2.2 Calentamiento del Sistema

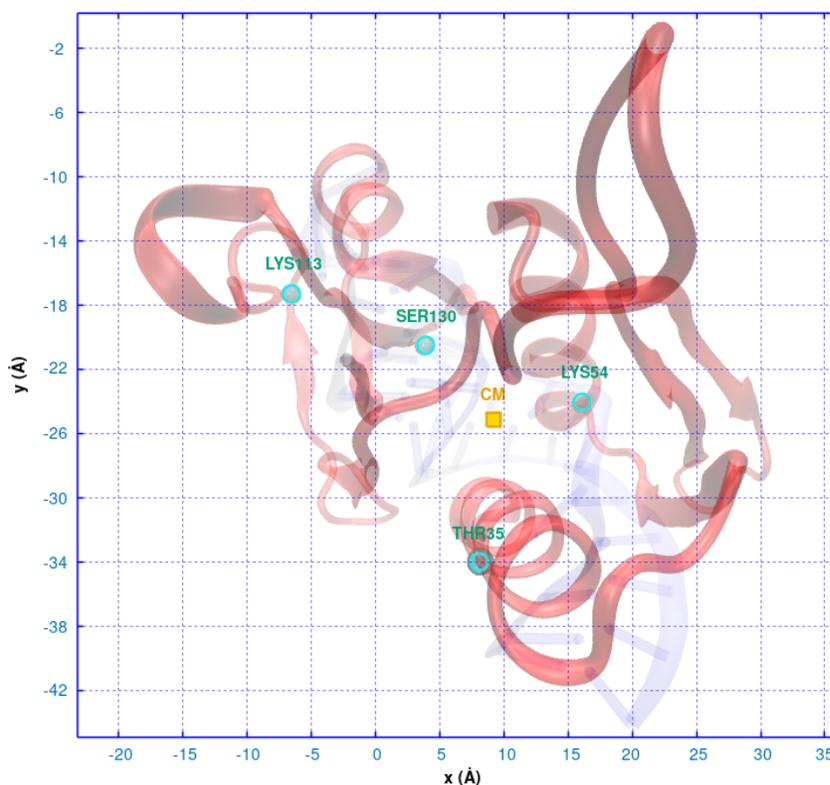
Posterior a la minimización, se procedió a calentar el sistema a una temperatura de 310  $K$  manteniendo las mismas restricciones de la minimización. El sistema se calentó de 0 a 310  $K$  en 20000 pasos de 1  $fs$ , de modo que las temperaturas se reasignaron en incrementos de 0.3878 $K$  cada 25 pasos. NAMD reasigna la temperatura del sistema al incrementar de forma gradual la velocidad de los átomos [28]. Entre cada incremento, el sistema se mantuvo a temperatura constante al emplear la Dinámica de Langevin, como se explicó en la Sección 2.2.3. El archivo de configuración utilizado para calentar el sistema se muestra en el Anexo C.

### 3.2.3 Equilibración

El sistema resultante del calentamiento se lo equilibró por 1  $ns$ . Se aplicó una restricción armónica a los átomos de fósforo de los residuos 1, 10 y 19 de la molécula de ARN. La constante armónica empleada fue de 5  $kcal/mol/\text{Å}^2$ . Se almacenaron las coordenadas del sistema cada 1000 pasos equivalente a conocer las posiciones de cada uno de los átomos del sistema cada 1  $ps$ . La presión del sistema se mantuvo constante a 1  $atm$ , empleando un coeficiente de amortiguamiento  $\gamma = 1ps^{-1}$ . Todos estos detalles mencionados se encuentran en el Anexo D.

### 3.3 Simulaciones SMD

Una vez que el sistema se ha equilibrado a 310 K, se simuló el proceso de disociación al halar los átomos  $C_\alpha$  de los aminoácidos *THR35*, *LYS54*, *LYS113* y *SER130* con velocidad constante de  $0.25 \text{ \AA}/ps$ . Tres de los átomos escogidos (*LYS54*, *LYS113* y *SER130*) se encuentran en la zona de enlace del complejo; mientras que el átomo restante (*THR35*), se lo escogió debido a que el lazo libre en el que se encuentra presentó oposición a la disociación. La figura 3.3 muestra la distribución geométrica de los átomos SMD en el plano XY con respecto a la estructura del complejo. Se puede observar que los átomos SMD ( $C_\alpha$  35 54 & 130) se encuentran distribuidos geoméricamente alrededor del centro de masa de la proteína.



**Figura 3.3:** Ubicación de los los átomos  $C_\alpha$  de residuos 35 54 113 130 en el plano XY. En color cian se resaltan los átomos SMD seleccionados, mientras que el punto cuadrangular de color amarillo representa el centro de masa del “backbone” de la proteína S8 dentro del plano XY.

La simulación SMD empleada para disociar el complejo se ejecutó durante 200000 pasos de 1  $fs$ . Durante la simulación SMD se aplicó una fuerza elástica con constante  $k = 5 \text{ kcal/mol}/\text{Å}^2$  manteniendo fijo al fósforo del ARN con residuo 10. Las coordenadas de los átomos constituyentes del sistema se almacenaron cada 100 pasos equivalente a conocer la estructura del sistema cada 100  $fs$ . El archivo de configuración empleado para la simulación SMD se encuentra en el Anexo E.

### 3.4 Cálculo del Potencial de Fuerza Media PMF

El PMF referente a la disociación del complejo involucra la definición de la coordenada de reacción  $\xi$ . En base a la información obtenida de la simulación SMD, se definió la coordenada de reacción  $\xi$  como la distancia que une los centros de masa del ARN y la proteína a lo largo del eje  $z$ . Una vez definida la coordenada de reacción, se establecieron intervalos equidistantes dentro de  $\xi$ . Estos intervalos, conocidos como “ventanas” [53], corresponden a regiones en las que el sistema se equilibrará independientemente. Para cada ventana se estableció un ancho de 1.2 Å comprendidas desde 22.4 Å hasta 36.8 Å. Se tomó como nodo inicial al lugar dentro de la coordenada de reacción que diste 0.2 Å de la configuración inicial, mientras que el nodo final se estableció como la región en la cual el complejo se ha disociado. De esta manera, al tomar 12 ventanas con un ancho de 1.2 Å se recorrió todo el dominio explorado por  $\xi$ .

Seguidamente, se buscó estructuras cuyo valor de  $\xi$  sea el punto medio de cada ventana. Cada estructura se la obtuvo a partir de la simulación SMD, en donde se seleccionaron los archivos referentes a las coordenadas del sistema junto con la dimensión de cada celda. Las estructuras seleccionadas para cada ventana fueron equilibradas durante  $2 \times 10^5$  pasos de 1  $fs$ , con restricciones idénticas a las empleadas durante la equilibración (Sección 3.2.3).

Adicionalmente, en cada ventana se estableció una constante elástica de  $2.5 \text{ kcal/mol}/\text{\AA}^2$  (Anexo F.2) referente al potencial sesgado (Ec. 2.17). Las coordenadas de los sendos sistemas se almacenaron cada 100 pasos de  $1 \text{ fs}$ . El script empleado para la equilibración de cada ventana se encuentra en el Anexo F.1. Una vez que se ha equilibrado el sistema en cada una de las ventanas, se procedió a calcular el PMF a través del análisis de los valores de  $\xi$  (ver Fig. 3.1). A partir de las trayectorias de las 12 ventanas, conjuntamente de la ponderación de histogramas “WHAM” (Weighted Histogram Analysis Method) [59] se obtuvo el perfil de energía libre.

### 3.4.1 Método de Histogramas Ponderados

El análisis de los resultados del muestreo “Umbrella” se efectúa al considerar el aporte ponderado de los histogramas obtenidos para cada ventana. El método de histogramas ponderados expresa la probabilidad de distribución global no sesgada como la superposición de la distribución de cada ventana [53, 59].

$$P^{(*)}(\xi_k) = \sum_i^S a_i(\xi_k) P_i^{(*)}(\xi) \quad (3.1)$$

$$= \sum_i^S a_i(\xi_k) \exp\{\beta w_i(\xi_k)\} \langle \exp\{-\beta w_i\} \rangle P_i'^{(*)}(\xi_k) \quad (3.2)$$

$$P_k^* = \sum_i^S a_{ik} c_{ik} f_i P_{ik}'^{(*)} \quad (3.3)$$

Donde:

\* : indica que la cantidad es estimada.

$S$  : número de ventanas o simulaciones.

$\xi_k$  : posición del bin  $k$  a lo largo  $\xi$ .

$a_{ik}$  : pesos estadísticos que minimizan el error de la distribución dentro de la ventana

$i$  para el bin  $k$ .

La ecuación (3.3) se la obtiene empleando la relación entre las distribuciones sesgadas y no sesgadas (Ec. 2.21) dentro de la distribución global (Ec. 3.1). A continuación se presentan un conjunto de ecuaciones denominadas como “ecuaciones *wham*” [19], las cuales permiten estimar los factores  $G_i$  de la expresión (2.22). Así, bajo la condición  $\sum a_{ik} = 1$  y considerando los factores  $n_{ik}$  como los puntos de datos asociados a la ventana  $i$ , la primera ecuación se expresa como [60]:

$$P_k^* = \frac{\sum_{i=1}^S n_{ik}}{\sum_{j=1}^S M_j \exp\{-\beta w_{jk} + \beta G_j\}} \quad (3.4)$$

Donde  $M_j$  es el número total de puntos de datos muestreados para cada ventana [59, 60].

La segunda ecuación proviene del Teorema de Jarzynski (Ec. 2.24).

$$\begin{aligned} \exp\{-\beta G_i\} &= \langle \exp\{-\beta w_{ik}\} \rangle \\ G_i &= -\frac{1}{\beta} \ln \left| \int P(\xi) \exp\{-\beta w_i(\xi)\} d\xi \right| \end{aligned} \quad (3.5)$$

De forma práctica una distribución de probabilidad exacta es difícil de lograrla. Por lo cual, es necesario discretizar  $P(\xi)$ . La discretización se realiza al definir intervalos discretos donde se agrupe  $\xi$ . Establecido el número de intervalos o bins, la última integral se transforma en la denominada segunda ecuación *wham* [60].

$$G_i = -\frac{1}{\beta} \ln \left| \sum_k P_k^* \exp\{-\beta w_{ik}\} \right| \quad (3.6)$$

Las ecuaciones *wham* son un conjunto de ecuaciones implícitas cuya resolución se da de

forma iterativa [53, 60]. Usualmente se inicia con un intento de  $G_i$ , luego se usa la primera ecuación (Ec. 3.4) para obtener la distribución  $P_k^*$ . Después, se emplea  $P_k^*$  en la segunda ecuación (Ec. 3.6) y así obtener una mejor estimación de  $G_i$ . Este proceso se efectúa reiteradamente hasta su convergencia. Finalmente el PMF es obtenido a través de la relación entre energía libre y la distribución de probabilidad (Ec. 2.14):

$$A(\xi_k) = -k_B T \ln(P_k^*) \quad (3.7)$$

La implementación de la distribución global descrita conjuntamente de la obtención del PMF se la efectuó con el programa *wham*<sup>1</sup> desarrollado por A. Grossfield.

Al combinar el muestreo “Umbrella” con la metodología WHAM se realiza un muestreo de estados de alta energía del sistema, lo que permite efectuar una construcción precisa del perfil de energía libre [53]. Al usar el esquema de ponderación de histogramas es necesario definir una coordenada de reacción específica, y, evaluar el método por vías de transición confiables como el análisis de parámetros (RMSD, RG). Cabe resaltar que una adecuada selección y análisis de  $\xi$  permite estudiar cambios conformacionales del sistema de estudio.

### 3.5 Características Estructurales del Complejo

La evaluación de las características del complejo se las efectuó mediante el análisis de  $\xi$ , el ángulo  $\theta$  formado por el vector que une los centros de masa del complejo molecular con el eje z, la distancia RMSD y el radio de Giro. La evolución temporal de estas cantidades permitieron describir cuantitativamente el complejo de estudio. Los parámetros RMSD y RG describen la estabilidad como la compacidad del ARN y de la proteína S8.

<sup>1</sup>Grossfield, A, “WHAM: an implementation of the weighted histogram analysis method”, <http://membrane.urmc.rochester.edu/content/wham/>, version 2.0.10

### 3.5.1 Caracterización de las trayectorias de DM

#### Raíz de la Desviación Cuadrática Media RMSD

La raíz de la desviación cuadrática media (Root Mean Square Deviation RMSD), es una magnitud que evalúa la distancia promedio entre los átomos de un sistema. El RMSD calcula la desviación estructural de un sistema con respecto a una estructura referencial. Esta cantidad se expresa como:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i(t) - \vec{r}_i(t_0))^2} \quad (3.8)$$

Donde,  $N$  es el número de átomos del sistema,  $\vec{r}_i(t)$  posición del átomo  $i$  a un tiempo  $t$  y  $\vec{r}_i(t_0)$  es la posición del átomo  $i$  al tiempo referencial  $t_0$  [28].

#### Radio de Giro RG

El radio de giro es una medida que describe la distribución de los átomos de un sistema con respecto a su centro de masa. Se define como:

$$RG = \sqrt{\frac{\sum_{i=1}^N m_i (\vec{r}_i - \vec{r}_{CM})^2}{\sum_{i=1}^N m_i}} \quad (3.9)$$

Donde  $N$  es el número total de átomos,  $m_i$  y  $\vec{r}_i$  son la masa y posición del átomo  $i$ ,  $\vec{r}_{CM}$  posición del centro de masa del sistema analizado. El radio de giro usualmente es empleado como un estimador de la compacidad de un sistema molecular [61].

## Capítulo 4

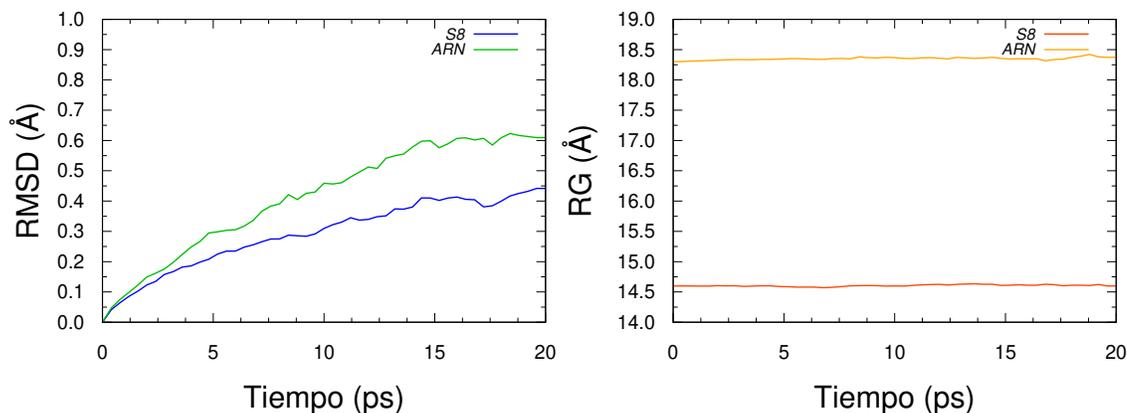
### Resultados

El sistema molecular obtenido al final de la preparación (Sección 3.2) contó con un total de 40260 átomos de los cuales 12288 corresponden al solvente (agua). Las moléculas de agua que bordean al complejo forman un paralelogramo de dimensiones (61.65, 73.51, 95.71) Å. El sistema tuvo una carga neta de  $3.649 \times 10^{-6} e$ .

#### 4.1 Minimización y Calentamiento

Durante la minimización, el sistema alcanzó la configuración con mínima energía en  $-161797 \text{ kcal/mol}$ ; mientras que al final del calentamiento su energía fue de  $-110120.24 \text{ kcal/mol}$ . El comportamiento del sistema fue analizado a través de la evolución temporal del RMSD y del RG. El RMSD calculado para la etapa de calentamiento del sistema, fue obtenido con respecto al estado final de la minimización; es decir, se tomó como estructuras referenciales aquellas alcanzadas al final de la minimización. La figura 4.1 muestra la evolución temporal del RMSD y del RG obtenidas para la proteína S8 y el ARN respectivamente durante el calentamiento del sistema. En ambos casos el RMSD tiende a crecer rápidamente durante la etapa de inicial del calentamiento. El comportamiento mostrado del RMSD, se debe a

las nuevas posiciones que adquiere la proteína juntamente del ARN a lo largo de la simulación. Por su parte el radio de giro, tanto para la proteína cuanto para el ARN, alcanzó un valor cuasi constante en la etapa de calentamiento. Así, RG para la proteína se estabilizó en 14.6 Å mientras que para el ARN el RG se estabilizó en 18.35 Å (Fig. 4.1).

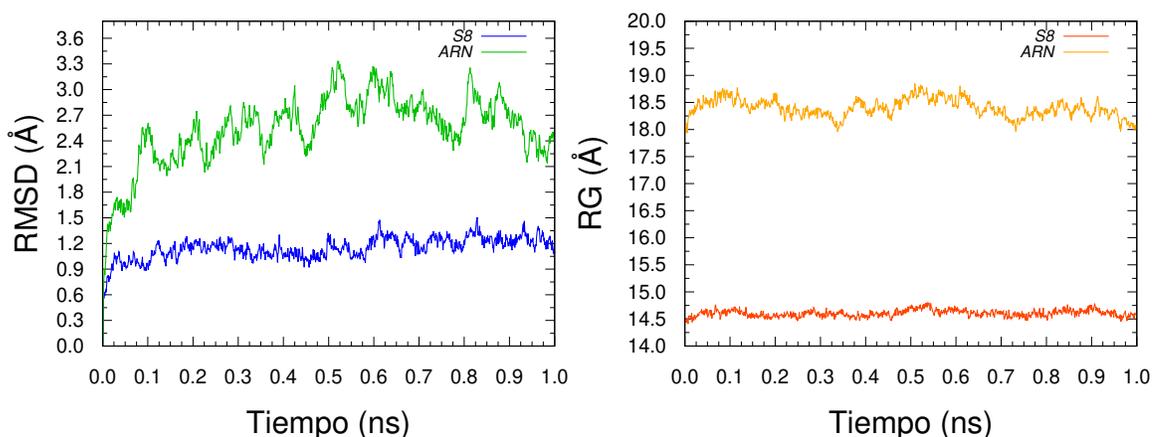


**Figura 4.1:** Evolución temporal del RMSD y RG para la proteína ribosómica S8 y su aptámero de ARN obtenido durante el calentamiento a 310 K.

## 4.2 Equilibración

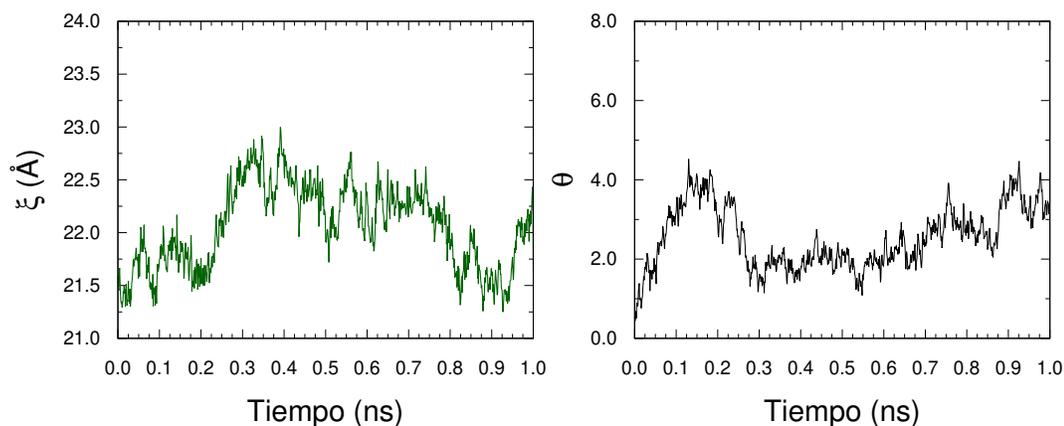
Con el sistema equilibrado a 310 K y 1 atm de presión, se procedió a analizar el comportamiento del RMSD para el “backbone” de la proteína S8 y los grupos fosfato del ARN. Junto con el RMSD se analizó la evolución temporal del radio de giro para los dos grupos moleculares mencionados. La evolución temporal del RMSD y el radio de giro (RG) para la proteína S8 y el ARN se presentan en la figura 4.2. El RMSD obtenido durante la equilibración, fue calculado con respecto al estado final del sistema en la etapa de calentamiento. Se observa que el RMSD de la parte proteica del sistema alcanzó un valor estable alrededor de 1.25 Å, es decir, a partir de los 0.8 ns. Previo a que el RMSD se estabilice, este mostró un comportamiento creciente, lo cual refleja que el sistema molecular exploró nuevos estados

durante la equilibración. Por otro lado, teniendo en cuenta el mismo intervalo temporal en el que el sistema se ha estabilizado ( $0.8 - 1 \text{ ns}$ ), el RG de la proteína S8 osciló alrededor de  $14.62 \text{ \AA}$ . En cambio, a partir de los  $0.8 \text{ ns}$ , el RMSD del ARN se estabilizó alrededor de  $2.64 \text{ \AA}$  (Fig. 4.2). El radio de giro del ARN creció durante los primeros  $0.05 \text{ ns}$ , a partir de este instante el RG osciló alrededor de  $18.5 \text{ \AA}$ .



**Figura 4.2:** Evolución temporal de la desviación cuadrática media *rmsd* y radio de giro *RG* obtenidos para la proteína S8 y su aptámero de ARN durante la equilibración del sistema.

El comportamiento en función del tiempo de la coordenada de reacción  $\xi$  y el ángulo  $\theta$  se muestra en la figura 4.3. Al observar la evolución de  $\xi$  durante la equilibración, se puede inferir que el sistema molecular alcanzó cierto “estado estable” cuando  $\xi = 22.4 \text{ \AA}$ , región en la cual el RMSD de la proteína se estabilizó. Este estado, asociado a un microestado del sistema, se lo tomó como punto de partida para disociar el complejo. En el intervalo temporal en el que la proteína del sistema se estabilizó ( $0.8 - 1 \text{ ns}$ ), el ángulo  $\theta$  fluctuó alrededor de  $3.17^\circ$ . Además, al analizar el comportamiento de  $\theta$  en función del tiempo se observa que el complejo molecular rotó con respecto al eje  $z$ . Se observó un incremento máximo de  $4.5^\circ$  aproximadamente, pese a que inicialmente el complejo fue alineado a lo largo del eje  $z$ .

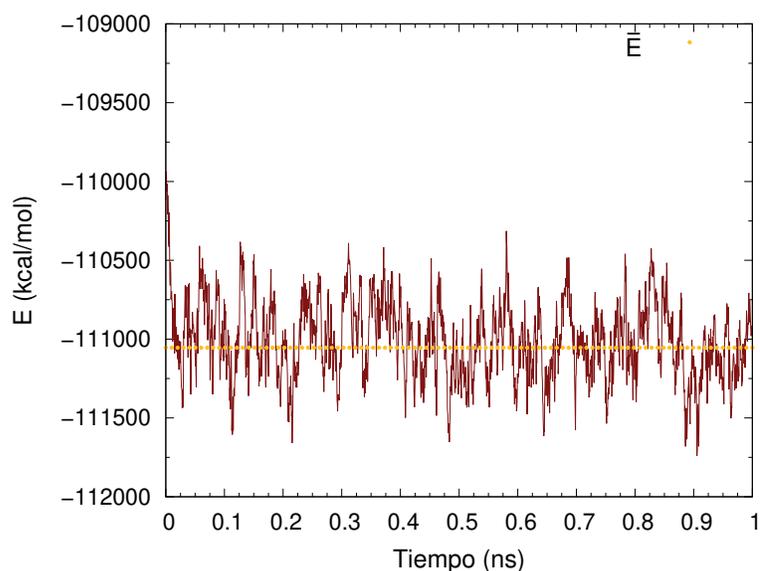


**Figura 4.3:** Evolución temporal de la coordenada de reacción  $\xi$  (verde) y del ángulo  $\theta$  (negro) durante la equilibración del sistema. El ángulo  $\theta$  corresponde al ángulo formado por el eje  $z$  y el vector que une los centros de masa entre las biomoléculas (proteína S8 y aptámero de ARN).

Adicionalmente, se estudió la evolución de la energía del sistema durante el proceso de equilibración (Fig. 4.4). En la figura 4.4 se ha esquematizado el valor medio de la energía referente a los últimos 0.2 *ns*, intervalo en el cual el RMSD de la proteína se estabilizó. Así, se muestra que el sistema alcanzó un valor de  $-111054.41 \text{ kcal/mol}$  en un tiempo de 1 *ns* equivalente a  $10^6$  pasos de 1 *fs*. El comportamiento oscilatorio alrededor de  $-111054.41 \text{ kcal/mol}$  corrobora que el sistema se ha equilibrado, ya que como se observa en la Fig. 4.4, no existe un crecimiento paulatino de la energía durante la equilibración.

### 4.3 Simulación SMD

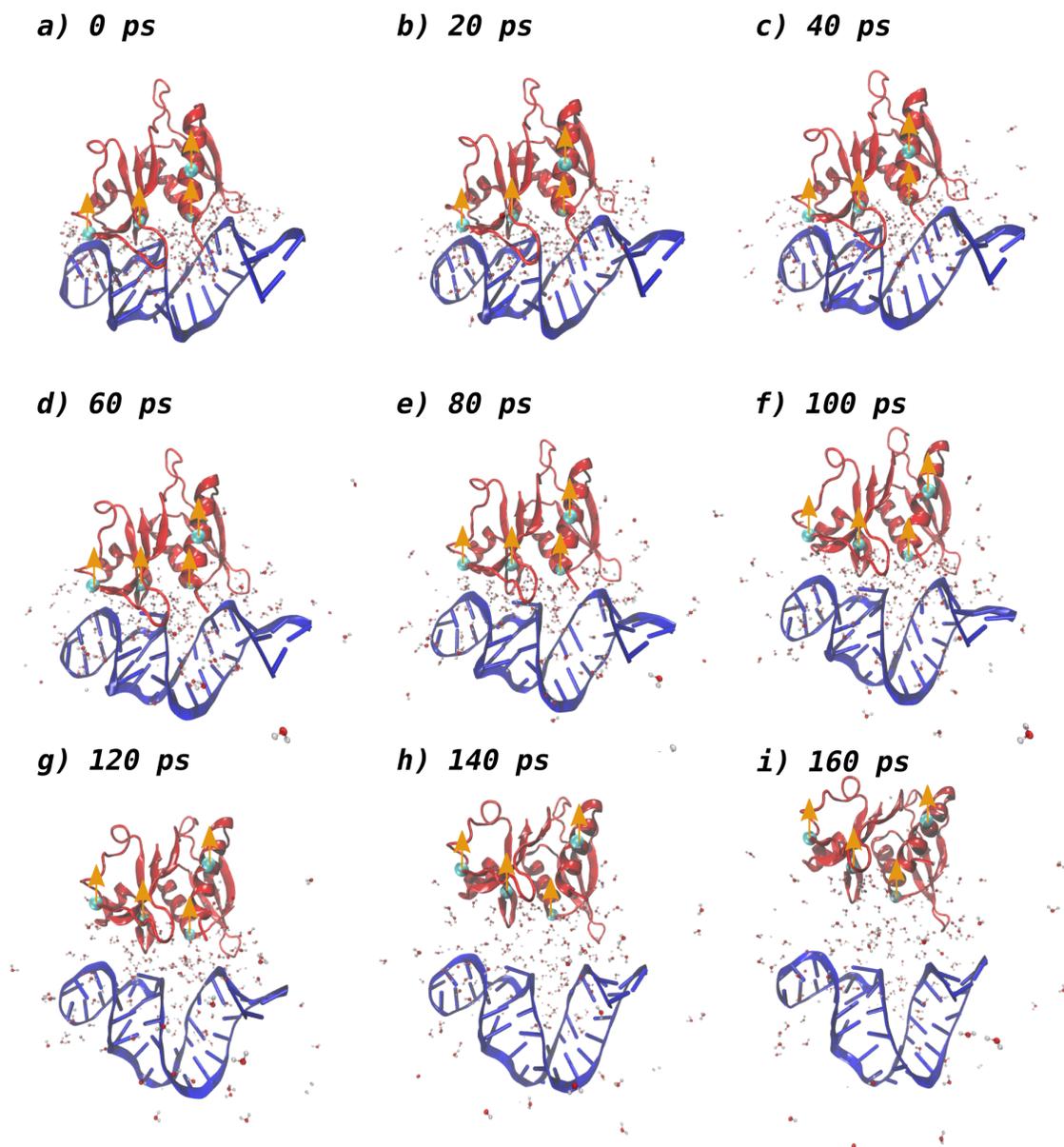
Con el sistema equilibrado, se procedió a disociar el complejo molecular como se detalló en la sección 3.3. La fuerza disociatoria empleada fue aplicada durante 160 *ps*. Este tiempo de simulación fue el suficiente para disociar el complejo molecular en las condiciones establecidas. Los “snapshots” de la simulación SMD realizada se presenta en la figura 4.5. En la figura 4.5 se aprecia como la proteína ribosomal S8 varía espacialmente mientras



**Figura 4.4:** Evolución temporal de la energía del sistema durante la equilibración a  $310K$  efectuada durante  $10^6$  pasos de  $1 fs$ . La línea punteada corresponde al valor medio de la energía del sistema durante los últimos  $0.2 ns$  ( $\bar{E} = -111054.41 kcal/mol$ ).

que el aptámero de ARN tiende a acomodarse, tal que muestra cierta oposición al proceso disociatorio. Este hecho se lo asocia a la interacción entre los grupos fosfato del ARN y amino de la proteína. Visualmente, se nota que en  $t = 160 ps$  tanto la proteína como el aptámero de ARN no interactúan y el espacio entre estas biomoléculas se llena con moléculas de agua.

Durante el proceso de disociación forzada se analizó el comportamiento de los mismos parámetros estudiados a lo largo de la equilibración del sistema, i.e,  $\xi$ , RG, RMSD y  $\theta$ . El comportamiento de estas variables se muestra en la figura 4.6. Se observa que  $\xi$  toma valores entre  $22.4$  y  $39.6 \text{ \AA}$ . A partir de los  $30 ps$  aproximadamente el complejo empieza a separarse, ya que, en el intervalo de  $0 - 30 ps$  no se evidencia un cambio notorio en la pendiente de  $\xi vs. t$  (Fig. 4.6). Alcanzado los  $30 ps$  de la simulación, se observa que la pendiente cambia paulatinamente. El cambio observado en la pendiente de  $\xi vs. t$  está relacionado con la ruptura de enlaces y por ende con el proceso de disociación, ya que



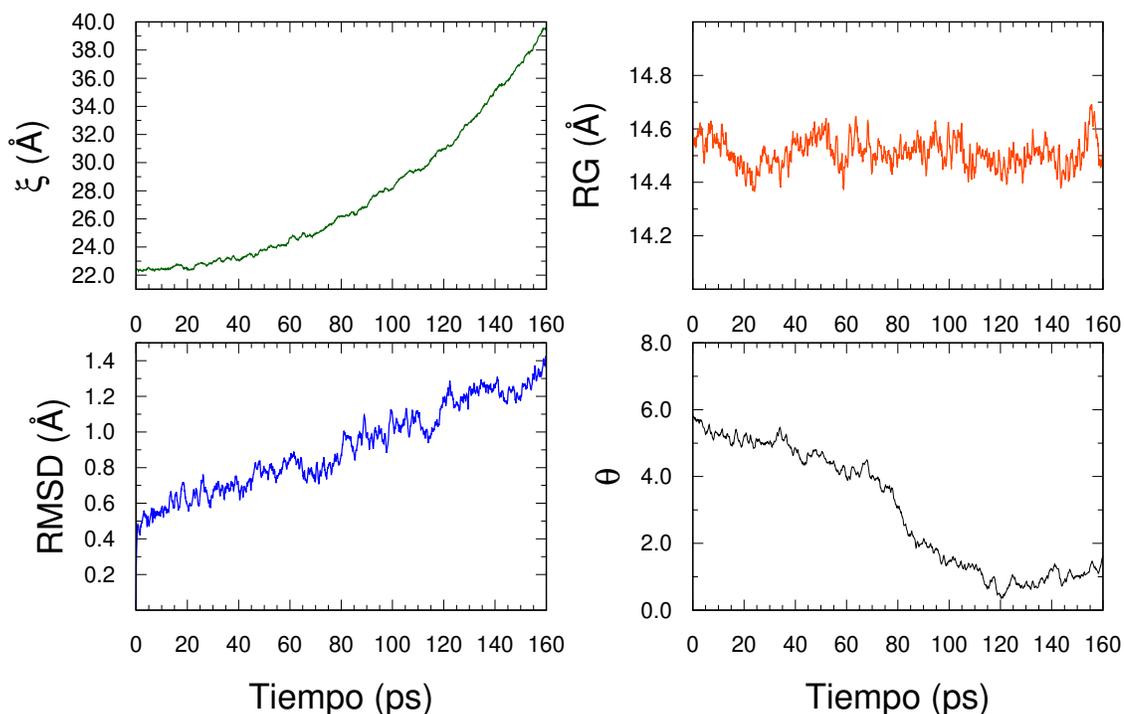
**Figura 4.5:** Snapshots de la simulación SMD efectuada para disociar el complejo aptámero de ARN-proteína S8 dentro de una caja de agua. En color cian junto con  $\uparrow$  se resaltan los átomos SMD que fueron halados en la dirección  $+z$  durante la simulación, mientras que en la representación CPK se muestran las moléculas de agua confinadas entre las biomoléculas señaladas.

variaciones grandes de  $\xi$  representan cambios conformacionales del complejo.

Al analizar la evolución temporal del RMSD del backbone de la proteína S8, se observa que el RMSD crece durante la disociación tomándole alrededor de 120 *ps* estabilizarse. El valor medio alcanzado entre 120 – 160 *ps* fue de 1.237 Å. Teniendo en cuenta el intervalo temporal en el que el RMSD se estabilizó, el RG osciló alrededor de 14.50 Å. En cuanto a la evolución temporal de  $\theta$ , se observa que este decrece durante la disociación; y además, se evidenció una notable disminución a partir del instante  $t = 30$  *ps*. En este mismo instante se observó un cambio en la pendiente de  $\xi$ , por lo que se infiere que la disociación como tal ocurre alrededor de dicho instante. Adicionalmente, a partir de  $t = 120$  *ps*,  $\theta$  se estabilizó en torno a 1.05°. Por otro lado, al analizar la evolución del radio de giro del backbone de la proteína, en líneas generales se infiere que no existe desnaturalización de la proteína ribosomal, ya que durante los 160 *ps* de la simulación SMD el RG mostró un comportamiento estable.

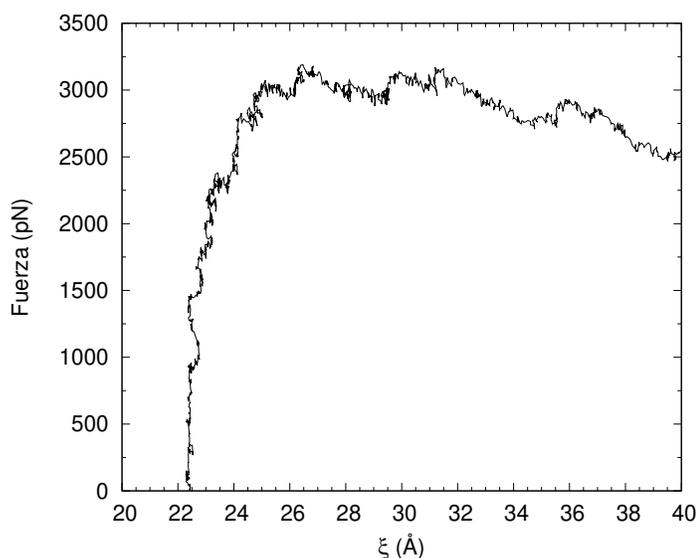
En la figura 4.7 se muestra la fuerza media de los átomos SMD con respecto a la coordenada de reacción  $\xi$ . La fuerza media alcanzó un valor máximo igual a 3189.65 *pN* en  $\xi = 26.47$  Å concerniente al paso temporal 86.4 *ps*. En general, se observa que la fuerza media de los átomos SMD tiende a crecer de forma rápida durante un primer intervalo (22.44 – 26.47 Å); cuando dicha fuerza alcanza su valor máximo esta tiende a afianzarse alrededor de 3000 *pN*. Posteriormente, en  $\xi = 32$  Å la fuerza media de los átomos SMD empieza a disminuir. El comportamiento observado de la fuerza media de los átomos SMD con respecto a  $\xi$  está en relación con la naturaleza de la misma, ya que esta es una fuerza que hala los átomos a velocidad constante. Entonces, el hecho de que la fuerza crezca, se estabilice y finalmente disminuya se debe a que en un principio existe una mayor cantidad de enlaces moleculares presentes en la interacción del complejo, después cuando la fuerza se estabiliza

algunos de los enlaces se han roto para finalmente disminuir ya que la mayoría de enlaces se han disipado.



**Figura 4.6:** Parámetros obtenidos durante la simulación SMD del sistema a 310K. Evolución temporal de la distancia de los centros de masa a lo largo del eje z (verde), radio de giro RG (naranja), distancia rmsd (azul) y ángulo  $\theta$  que forma el vector que une los centros de masa con el eje z (negro). El radio de giro RG y el rmsd fueron calculados para el backbone la proteína ribosomal S8.

Durante la simulación SMD, se evidenció la interacción y posterior ruptura entre grupos moleculares del ARN y de la proteína ribosomal S8. El primer contacto molecular que se disoció fue entre *CYT17* con *ALA114*. La distancia entre los centros de masa de estos residuos presentó un crecimiento no lineal a partir del instante  $t = 47.8ps$ . En la figura 4.8 se muestra la distancia entre los centros de masa de los residuos *CYT17* – *ALA114*. Estos mostraron una interacción directa ya que se observó el enlace molecular entre los grupos fosfato y amino respectivamente. Por su parte, el aminoácido concerniente al residuo 113 se reorientó para interactuar con *CYT17*. La interacción entre *CYT17* con *LYS113* se dió una vez que

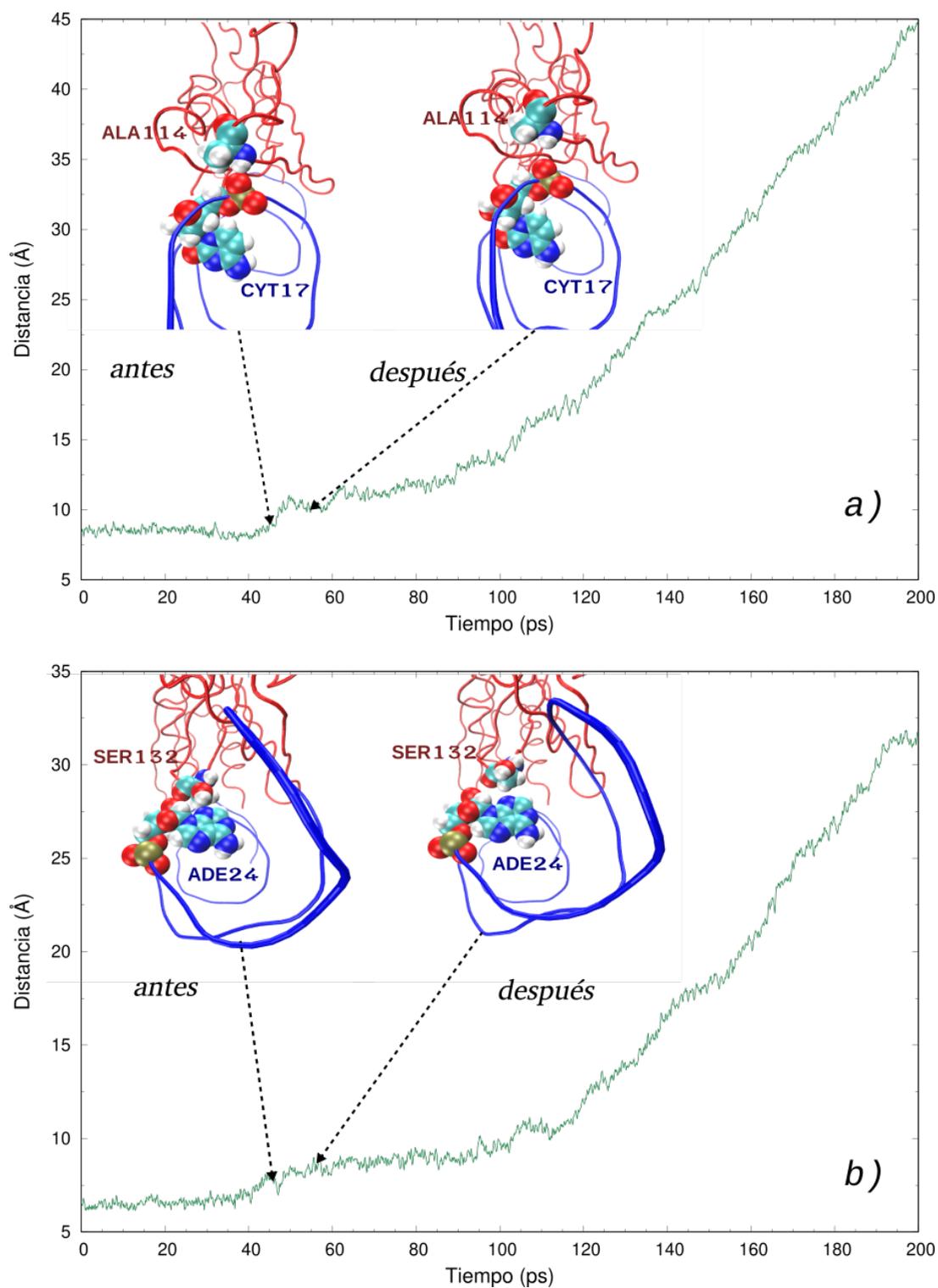


**Figura 4.7:** Fuerza media de los átomos SMD en función de la coordenada de reacción  $\xi$

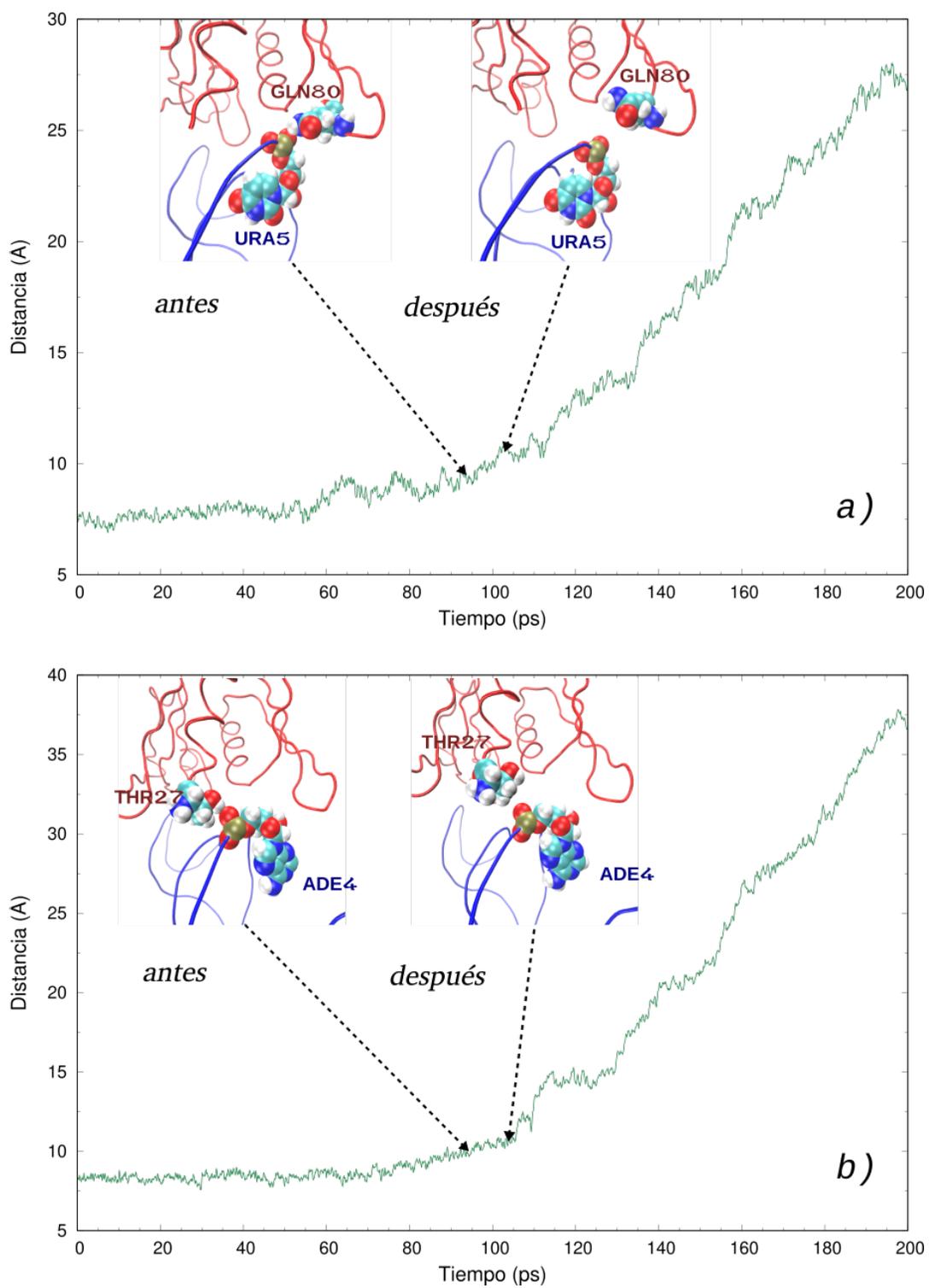
*ALA114* se separó de dicha base nitrogenada. El par *CYT17* – *LYS113* interactuaron hasta  $t = 85$  ps. Previamente, los residuos concernientes al par *ADE24* – *SER132* se apartaron. El enlace entre estos tuvo una duración de 52.8 ps. Precisamente, a partir de este instante la distancia entre los centros de masa de los residuos *ADE24* – *SER132* creció paulatinamente (Fig. 4.8).

La distancia entre los centros de masa de los residuos *URA5* – *GLN80* mostró un cambio de comportamiento al tiempo  $t = 100$  ps. En este instante se evidenció la ruptura entre estos residuos (Fig. 4.9). Por su parte, el residuo *ADE4* del ARN junto con *THR27* de la proteína estuvieron enlazados a lo largo de 102 ps. Al observar la evolución temporal de la distancia entre los centros de masa de estos últimos residuos (Fig. 4.9), se evidencia un cambio de pendiente en dicho instante.

Los pares moleculares, el tiempo, la región dentro de  $\xi$  y la fuerza media asociada la ruptura de los enlaces mencionados se presenta en la tabla 4.1.



**Figura 4.8:** Distancia entre los centros de masa de los residuos a) *CYT17* – *ALA114* y b) *ADE24* – *SER132*.



**Figura 4.9:** Distancia entre los centros de masa de los residuos a) *URA5* – *GLN8* y b) *ADE4* – *THR27*.

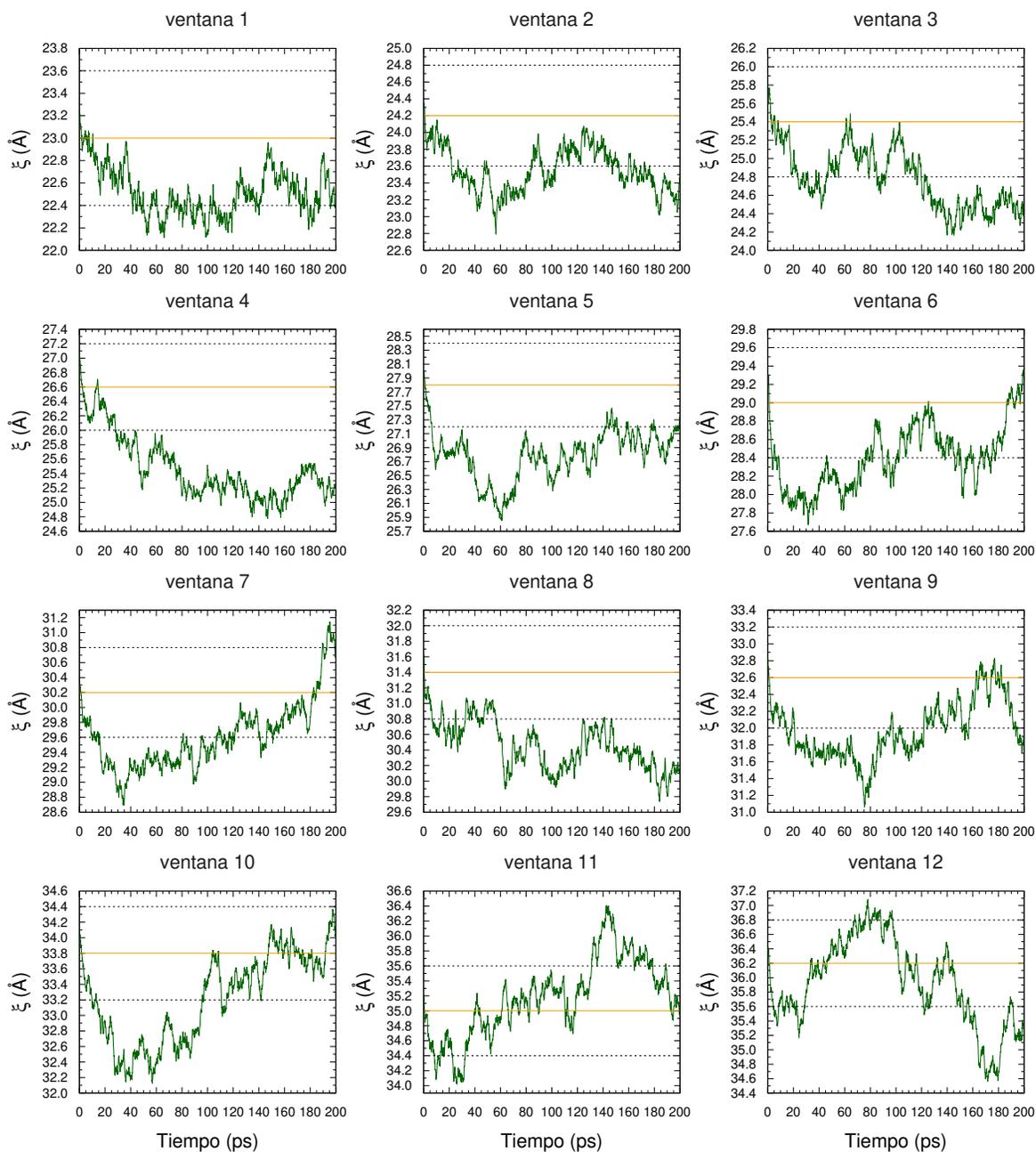
Contactos	Tiempo (ps)	$\xi$ (Å)	F (pN)
CYT17-ALA114	47.8	23.552	2296.00
ADE24-SER132	52.8	24.119	2426.52
CYT17-LYS113	85.0	26.437	3138.09
URA5-GLN80	100.0	28.125	2978.35
ADE4-THR27	102.0	28.644	3021.29

**Tabla 4.1:** Tiempo asociado a la ruptura de enlaces moleculares entre los grupos fosfato y amino del ARN y la proteína S8. Junto a los tiempos de ruptura se presenta el valor de la coordenada de reacción  $\xi$  y la fuerza media de los átomos SMD para cada instante de ruptura.

## 4.4 Potencial de Fuerza Media PMF

El comportamiento de  $\xi$  durante la equilibración de cada ventana se muestra en la figura 4.10. Se observa que  $\xi$  exploró con mayor frecuencia las regiones aledañas al límite inferior de cada una de las ventanas. Concretamente,  $\xi$  exploró regiones concerniente a la ventana anterior. La tendencia de  $\xi$  en transitar lugares inferiores al punto medio de cada ventana contribuyó a la obtención de un PMF “suave”. Finalmente cabe analizar el comportamiento del radio de giro  $RG$  y del ángulo  $\theta$  que forma el eje  $z$  con el vector que une los centros de masa de los constituyentes moleculares (ARN y proteína S8).

La evolución temporal de los parámetros RMSD,  $RG$  y  $\theta$  se muestra en las figuras 4.11, 4.10, 4.12 y 4.13 respectivamente. Para el cálculo del RMSD se tomó como estructura referencial a la proteína situada en el punto medio de cada ventana. El RMSD del backbone de la proteína ribosomal S8 encontrado para cada ventana exhibe un breve incremento durante los primeros pasos de la simulación. Una vez que se ha superado la etapa rápida de crecimiento, el RMSD se estabilizó en cada ventana. De hecho, el perfil de energía libre (Fig. 4.15), se lo obtuvo suprimiendo los primeros 20000 pasos (20 ps). Durante ese tiempo el sistema aun no se ha equilibrado y el RMSD de la proteína aun no se ha estabilizado.

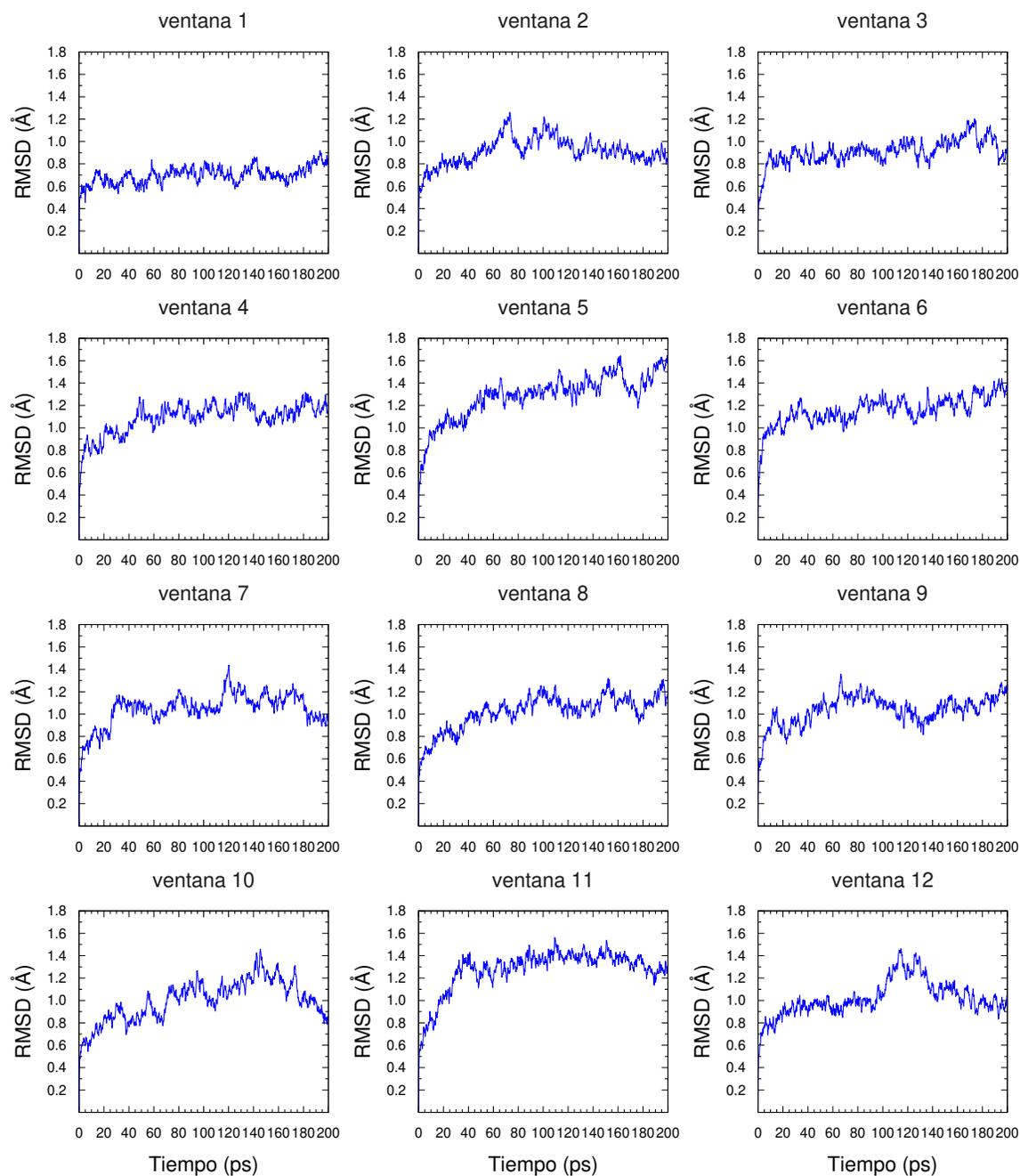


**Figura 4.10:** Evolución temporal de la coordenada de reacción  $\xi$  durante la equilibración del sistema en las sendas ventanas. Los límites de cada ventana se los representa con líneas entrecortadas y el punto medio de cada ventana se lo resaltó con una línea continua de color amarillo.

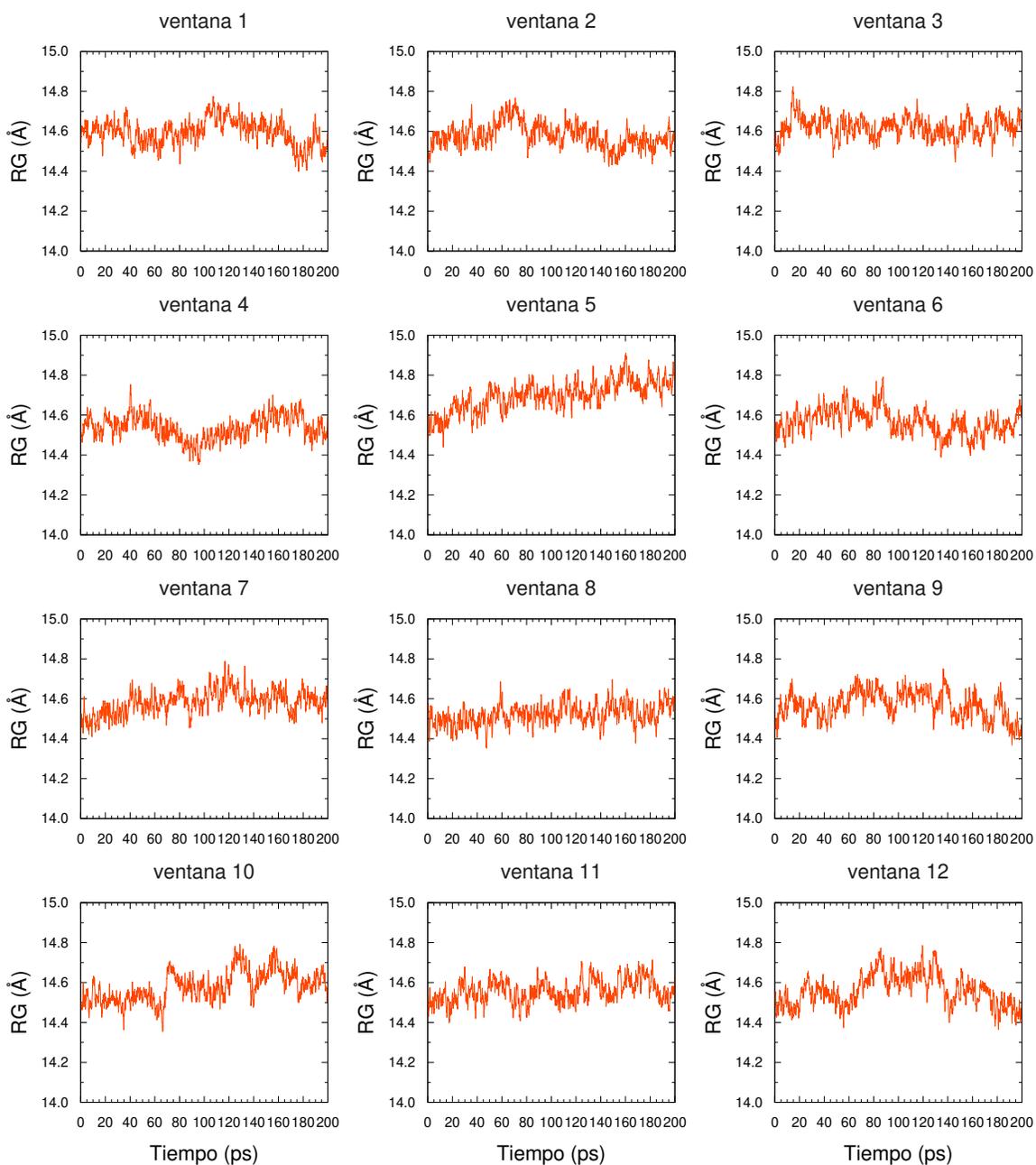
De manera general, se observa que el radio de giro de la parte proteica del sistema (Fig. 4.12) tiende a cierto valor constante en cada ventana, dichos valores se los muestra en la tabla 4.2. En cambio, al analizar la evolución temporal del ángulo  $\theta$  (Fig. 4.13), se observa que la proteína sufre una ligera rotación no mayor a  $2.7^\circ$  con excepción de las dos primeras ventanas (Tabla 4.2). No obstante, en todas las ventanas  $\theta$  no presentó grandes desviaciones durante la equilibración. Las características que se han descrito con respecto a RMSD,  $\xi$ ,  $RG$  y  $\theta$  corroboraron que la proteína S8 no se desnaturalizó durante la equilibración del sistema en cada ventana.

Ventana	$RG(\text{Å})$	$\sigma$	$\theta$	$\sigma$
1	14.594	0.066	3.806	0.393
2	14.581	0.061	3.750	0.542
3	14.619	0.048	1.499	0.442
4	14.538	0.062	0.711	0.321
5	14.711	0.064	2.102	0.387
6	14.569	0.065	2.669	0.396
7	14.592	0.055	1.808	0.399
8	14.525	0.052	1.483	0.855
9	14.570	0.069	1.995	0.405
10	14.586	0.075	1.364	0.563
11	14.564	0.056	1.827	0.519
12	14.568	0.078	1.550	0.573

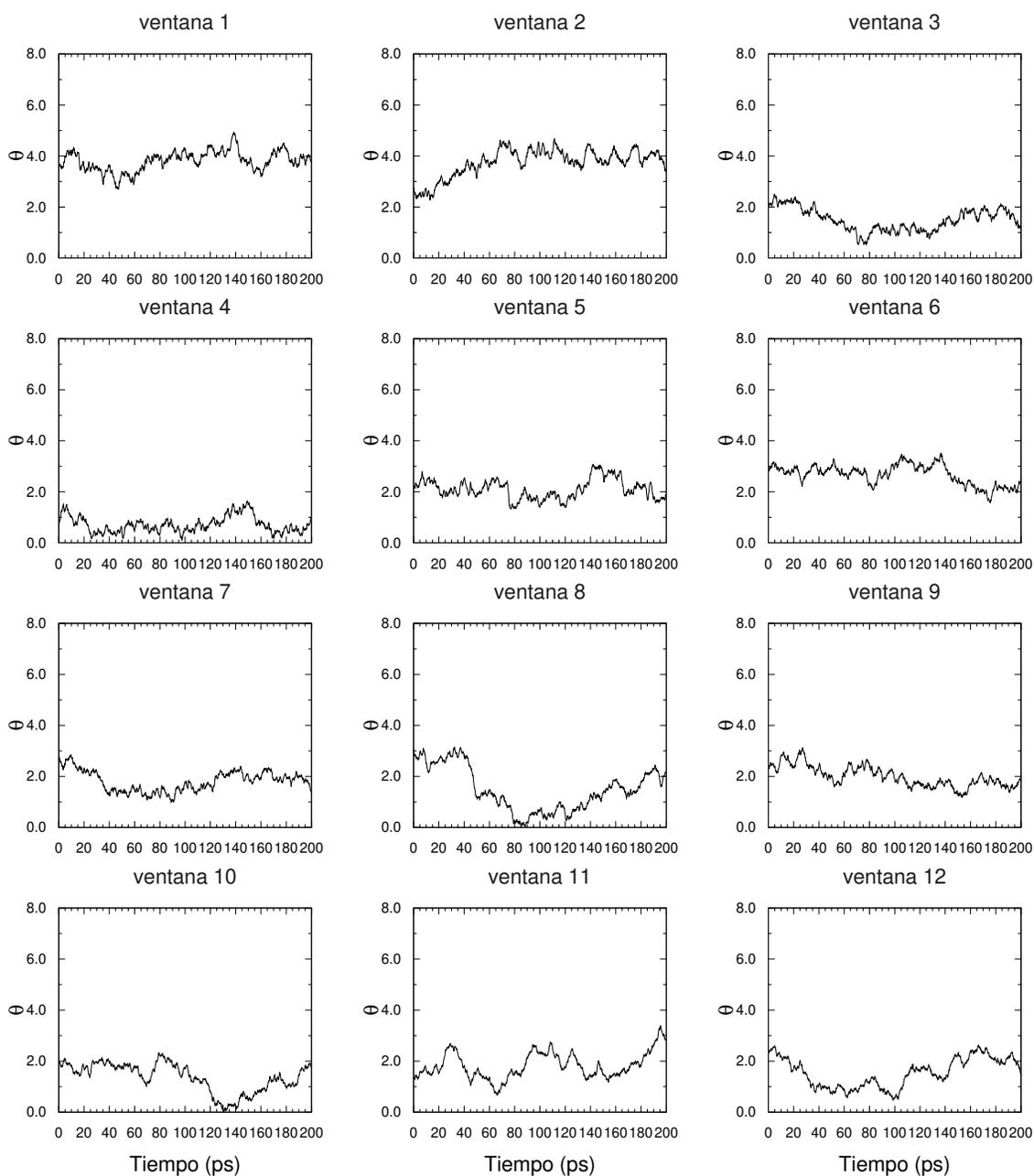
**Tabla 4.2:** Promedios estadísticos del radio de giro  $RG$  y del ángulo  $\theta$  junto con su desviación cuadrática media  $\sigma$  encontrados durante la equilibración de cada ventana. Los promedios junto con  $\sigma$  se encontraron descartando los primeros 20000 pasos.



**Figura 4.11:** Evolución temporal del *RMSD* calculada para el backbone de la proteína ribosomal S8 durante la equilibración del sistema molecular para cada ventana.

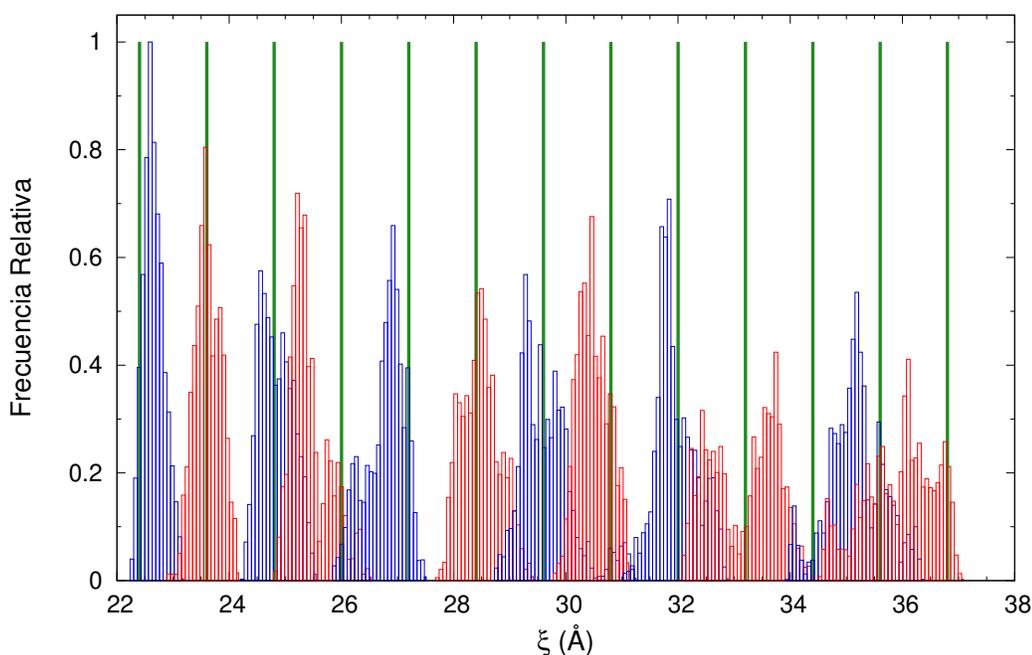


**Figura 4.12:** Evolución temporal del  $RG$  calculada para el backbone de la proteína ribosomal S8 durante la equilibración del sistema molecular para cada ventana.



**Figura 4.13:** Evolución temporal del ángulo  $\theta$  formado por el eje z y el vector que une los centros de masa entre las biomoléculas (proteína y ARN) durante la equilibración de cada ventana.

Los resultados del muestreo de la coordenada de reacción  $\xi$  se lo presenta en la Figura 4.14. Se observa que existieron valores adoptados frecuentemente por el complejo, como es el caso de las dos primeras ventanas. Esto se debe a que el sistema en las dos primeras ventanas se encuentra cerca a su configuración de equilibrio. Al tener en cuenta el muestreo de  $\xi$  en cada ventana (Fig. 4.14), vemos que en la vecindad  $\xi = 24 \text{ \AA}$  y  $\xi = 28 \text{ \AA}$  existen lugares que  $\xi$  no ha explorado; por lo que, dentro del perfil del PMF (Fig. 4.15) se aprecian puntos que sobresalen del resto de la curva.



**Figura 4.14:** Histogramas de la coordenada de reacción para cada ventana. El tiempo de equilibrio empleado fue de 200 *ps.* descartando los primeros 20 *ps*

Los porcentajes de solapamiento se presenta en la tabla 4.3, en ella se corrobora el hecho de que ciertas ventanas no presentan solapamiento como es el caso de las ventanas 2 & 3 junto con las ventanas 5 & 6. Existe también ventanas aledañas que tienen un porcentaje alto de solapamiento (ventanas 11 & 12) como otras que poseen un solapamiento bajo (ventanas 8 & 9). La literatura recomienda que el porcentaje de solapamiento entre ventanas vecinas se encuentre entre el 30% y 40%. El solapamiento requerido es importante para disponer

de un muestreo adecuado de las configuraciones. Sin embargo, para alcanzar el valor recomendado de solapamiento se requiere un mayor tiempo de equilibración del sistema.

Ventanas Aledañas	Porcentaje
1-2	1.46 %
2-3	0.00 %
3-4	33.25 %
4-5	8.70 %
5-6	0.00 %
6-7	10.61 %
7-8	18.39 %
8-9	0.50 %
9-10	26.29 %
10-11	5.18 %
11-12	47.76 %
<b>Promedio</b>	<b>13.83%</b>

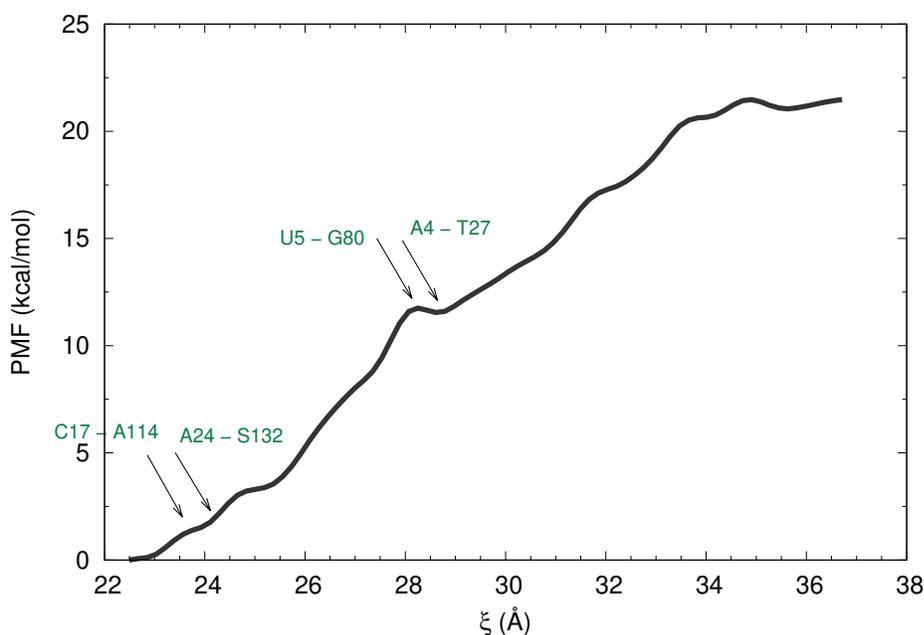
**Tabla 4.3:** Porcentaje de solapamiento entre ventanas aledañas con respecto al muestreo de  $\xi$ .

El PMF obtenido (Fig. 4.15) se lo representó como función de la distancia de los centros de masa en el eje  $+z$ . El PMF, expresado en función de  $\xi$ , muestra que la energía requerida para disociar el complejo es de  $21.3 \text{ kcal/mol}$  ( $\sim 34k_B T$  a  $310 \text{ K}$ ). El perfil de energía libre encontrado crece de forma irregular hasta los  $21.3 \text{ kcal/mol}$ , valor en el cual el complejo se ha disociado. Para poner en perspectiva esta magnitud, se puede considerar la hidrólisis de 1 mol de ATP que, a temperatura ambiente libera alrededor de  $7.3 \text{ kcal/mol}$  de energía libre [62]. Al comparar la energía libre requerida para la disociación del complejo con valores referenciales de importancia biológica se constata que la energía libre encontrada es aproximadamente 3 veces la energía liberada durante la hidrólisis del ATP.

En la Fig. 4.15 se observa la existencia de mesetas, regiones con PMF constante, comprendidas entre  $23.5 < \xi < 24.2 \text{ \AA}$  y  $28.3 < \xi < 28.9 \text{ \AA}$ . Justamente, en  $\xi = 23.55$  y  $24.11 \text{ \AA}$  los residuos *CYT17* – *ALA114* junto con *ADE24* – *SER132* se disociaron. La ruptura de enlaces entre estos dos pares de residuos (Fig. 4.8) está relacionada con la primera meseta

dentro del perfil del PMF.

Para  $\xi = 28.13$  y  $28.64$  Å los enlaces U5 - G80 y A4 - T27 se desligaron respectivamente (Fig. 4.9). La ruptura de los enlaces entre estos dos pares de residuos ocurrió cuasi-simultáneamente (Tabla 4.1). En consecuencia, el perfil del PMF presentó otra meseta asociada a estas rupturas. Al relacionar los valores adquiridos por  $\xi$  en las mesetas dentro del PMF, se observa que la fuerza media de los átomos SMD encuentra mínimos locales (Fig. 4.7). Adicionalmente, si tenemos en cuenta la distribución de la coordenada de reacción en cada ventana (Fig. 4.14), vemos que cerca a  $\xi = 24$  Å y  $\xi = 27.5$  Å existen lugares que  $\xi$  no ha explorado. Sin embargo, este mal muestreo entre las ventanas 2-3 y 5-6 se contrarrestó al emplear una media móvil con 4 puntos a los datos del PMF. Con el tratamiento de los datos mencionado, se suavizó el perfil del PMF el cual se muestra en la figura 4.15.



**Figura 4.15:** Perfil de energía libre (PMF) obtenido durante el proceso de disociación forzado del complejo 4PDB inmerso en una caja de agua. El PMF encontrado se lo expresa en función de la coordenada de reacción  $\xi$ .

## Capítulo 5

### Conclusiones

El cambio de energía libre asociado al proceso de disociación del complejo aptámero de ARN proteína ribosomal S8 fue caracterizado por medio del potencial de fuerza media. El perfil del PMF creció hasta el valor energético límite de  $21.3 \text{ kcal/mol}$ . Dentro del perfil se observaron fases transitorias en las que el PMF permaneció relativamente constante. Las mesetas encontradas están asociadas a la ruptura de contactos dentro del complejo molecular. Estas mesetas mostraron la existencia de contactos energéticamente más altos en comparación con los demás. El perfil encontrado mostró características similares a aquel obtenido por Yang et al., donde estudiaron el PMF concerniente a la disociación entre la trombina y su aptámero de ADN [63].

La energía libre requerida para la disociación es comparable con valores energéticos de otros procesos bioquímicos. En condiciones ambientales, en promedio cada molécula forma 3.5 enlaces hidrógeno con otras moléculas de agua en estado líquido [64]. Teniendo en cuenta que la energía asociada a un enlace hidrógeno se encuentra entre  $3 \sim 10 \text{ kcal/mol}$  [65, 66], se concluye que el número total de enlaces, asociados a la ruptura de contactos, fue de 15 enlaces aproximadamente. Similarmente, al comparar la energía de disociación con

la hidrólisis del ATP, se encontró que  $\Delta H(dis) \approx 3\Delta H(ATP)$ . Por lo que se infiere que existe suficiente energía libre para romper varios enlaces químicos entre ellos, enlaces hidrógeno.

Las regiones cuasi constantes en el perfil del PMF están asociadas a la ruptura de enlaces entre los grupos moleculares del complejo estudiado. En  $\xi = 23.55$  y  $24.11$  Å los residuos *CYT17* y *ADE24* dejaron de interactuar con los aminoácidos *ALA114* y *SER132* respectivamente. La ruptura de estos enlaces se refleja en la presencia de la meseta en  $23.5 < \xi < 24.2$  correspondiente a  $1.5$  kcal/mol. El PMF presentó otra meseta comprendida entre  $28.3 < \xi < 28.9$ , en donde se constató que los residuos *URA5* *GLN80* conjuntamente de *ADE4* *THR27* se apartaron. Estos últimos residuos se disociaron con una energía de  $11.6$  kcal/mol. En las regiones concernientes a la ruptura de enlaces la fuerza media disociatoria alcanzó mínimos locales.

El radio de giro y el RMSD concernientes al backbone de la proteína ribosomal S8 no exhibieron cambios considerables durante el proceso disociatorio. El RMSD se estabilizó a partir de  $120$  ps alcanzando el valor medio de  $1.237$  Å mientras que el RG fluctuó alrededor de  $14.5$  Å. Sin embargo, se pudo constatar que el complejo rotó ligeramente con respecto al eje z. El ángulo que forma el vector que une los centros de masa del complejo sufrió un incremento de  $4^\circ$  durante la equilibración del sistema. La rotación del complejo se dió a pesar que la fuerza externa simulada fue aplicada en la dirección  $+z$  por lo cual se resalta la fuerza de enlace en las regiones de unión del complejo.

La fuerza media de los átomos SMD con respecto a  $\xi$  reportó un pico máximo cuando  $\xi = 26.47$  Å para posteriormente disminuir en  $\xi = 32$  Å. Durante este intervalo la fuerza media de los átomos SMD se afianzó alrededor de  $3000$  pN, con la peculiaridad de poseer

un comportamiento lineal. Así, se infirió que la disociación del complejo ocurrió durante este intervalo. Además, cuando  $\xi = 32 \text{ \AA}$  se constató que el espacio comprendido entre la proteína S8 y el ARN se llenó con moléculas de agua. Es preciso señalar que la fuerza disociatoria aplicada permitió separar al complejo en  $\xi = 39.6 \text{ \AA}$ .

El estudio realizado ha permitido asociar eventos moleculares específicos con las características del PMF para la disociación del complejo entre la proteína ribosomal S8 y su aptámero de ARN. Particularmente, las mesetas en el PMF están ligadas a la ruptura de contactos críticos específicos. Los contactos identificados son aquellos formados por los residuos de la proteína *ALA114*, *SER132*, *GLN80* & *THR27* con sus contrapartes en el ARN *CYT17*, *ADE24*, *URA5* & *ADE4*, respectivamente. Finalmente, es importante disponer de una descripción física de los mecanismos de unión entre aptámeros y sus blancos moleculares puesto que, permite categorizar la importancia de los diferentes contactos moleculares presentes en la interfase.

## Referencias

- [1] C. Tuerk and L. Gold, “Systematic evolution of ligands by exponential enrichment: Rna ligands to bacteriophage t4 dna polymerase.,” *Science (New York, N.Y.)*, vol. 249, pp. 505–510, 4968 1990.
- [2] D. L. Robertson and G. F. Joyce, “Selection in vitro of an rna enzyme that specifically cleaves single-stranded dna,” *Nature*, vol. 344, pp. 467–468, 6265 1990.
- [3] A. D. Ellington and J. W. Szostak, “In vitro selection of rna molecules that bind specific ligands,” *Nature*, vol. 346, pp. 818–822, 6287 1990.
- [4] A. D. Keefe, S. Pai, and A. Ellington, “Aptamers as therapeutics,” *Nature Reviews Drug Discovery*, vol. 9, pp. 537–550, 7 2010.
- [5] P. S. Pendergrast and H. N. Marsh, “Nucleic acid aptamers for target validation and therapeutic applications,” *Journal of biomolecular techniques : JBT*, vol. 16, pp. 224–234, 3 2005.
- [6] S. Cai, J. Yan, H. Xiong, Y. Liu, D. Peng, and Z. Liu, “Investigations on the interface of nucleic acid aptamers and binding targets,” *The Analyst*, vol. 143, pp. 5317–5338, 2018.
- [7] I. Jeddi and L. Saiz, “Computational design of single-stranded dna hairpin aptamers immobilized on a biosensor substrate,” *Scientific Reports*, vol. 11, p. 10 984, 1 2021.

- [8] B. Guan and X. Zhang, “Aptamers as versatile ligands for biomedical and pharmaceutical applications,” *International Journal of Nanomedicine*, vol. 15, pp. 1059–1071, 2020.
- [9] J. F. Lee, G. M. Stovall, and A. D. Ellington, “Aptamer therapeutics advance,” *Curr Opin Chem Biol*, vol. 10, pp. 282–289, 3 2006.
- [10] C. S. M. Ferreira, M. C. Cheung, S. Missailidis, S. Bisland, and J. Gariépy, “Phototoxic aptamers selectively enter and kill epithelial cancer cells,” *Nucleic Acids Research*, vol. 37, no. 3, pp. 866–876, Dec. 2008.
- [11] V. Thiviyanathan and D. G. Gorenstein, “Aptamers and the next generation of diagnostic reagents,” *Proteomics. Clinical applications*, vol. 6, pp. 563–573, 11-12 2012.
- [12] K. Han and C. Nepal, “Pri-modeler: Extracting rna structural elements from pdb files of protein-rna complexes,” *FEBS letters*, vol. 581, pp. 1881–90, 9 2007.
- [13] L. Pérez-Cano and J. Fernández-Recio, “Optimal protein-rna area, opr: A propensity-based method to identify rna-binding sites on proteins,” *Proteins*, vol. 78, pp. 25–35, 1 2010.
- [14] N. Bjerregaard, P. A. Andreasen, and D. M. Dupont, “Expected and unexpected features of protein-binding rna aptamers,” *Wiley interdisciplinary reviews. RNA*, vol. 7, pp. 744–757, 6 2016.
- [15] T. Hermann and D. J. Patel, “Adaptive recognition by nucleic acid aptamers,” *Science (New York, N.Y.)*, vol. 287, pp. 820–825, 5454 2000.
- [16] M. Corley, M. C. Burns, and G. W. Yeo, “How rna-binding proteins interact with rna: Molecules and mechanisms,” *Molecular cell*, vol. 78, pp. 9–29, 1 2020.

- [17] A. Gupta and M. Gribskov, “The role of rna sequence and structure in rna–protein interactions,” *Journal of Molecular Biology*, vol. 409, pp. 574–87, 4 2011.
- [18] J. G. Kirkwood, “Statistical mechanics of fluid mixtures,” *The Journal of Chemical Physics*, vol. 3, pp. 300–313, 5 1935.
- [19] B. Roux, “The calculation of the potential of mean force using computer simulations,” *Computer Physics Communications*, vol. 91, no. 1, pp. 275–282, 1995, ISSN: 0010-4655.
- [20] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, “Free energy calculation from steered molecular dynamics simulations using jarzynski’s equality,” *J. Chem. Phys.*, vol. 119, pp. 3559–3566, 6 2003.
- [21] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, “Adaptive biasing force method for scalar and vector free energy calculations,” *J. Chem. Phys.*, vol. 128, no. 44120, 2008.
- [22] M. Davlieva, J. Donarski, J. Wang, Y. Shamoo, and E. P. Nikonowicz, “Structure analysis of free and bound states of an rna aptamer against ribosomal protein s8 from bacillus anthracis.,” *Nucleic Acids Res*, vol. 42, pp. 10 795–10 808, 16 2014.
- [23] W. Humphrey, A. Dalke, and K. Schulten, “Vmd: Visual molecular dynamics,” *Journal of molecular graphics*, vol. 14, pp. 33–28, 1 1996.
- [24] H. Lodish *et al.*, *Biología celular y molecular*, 5th ed. Medica Panamericana, 2006, ISBN: 950-06-1374-3.
- [25] W. Müller-Sterl, *Bioquímica. Fundamentos para Medicina y Ciencias de la Vida*. Revert’e, S. A., 2008, ISBN: 978-84-291-7393-2.
- [26] K. Müller-Dethlefs and P. Hobza, “Noncovalent interactions: A challenge for experiment and theory,” *Chem Rev*, vol. 100, pp. 143–168, 1 2000.

- [27] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature*, vol. 9, no. 9, pp. 646–652, 2002.
- [28] R. Bernardi *et al.*, *NAMD User’s Guide Version 2.14*. Theoretical Biophysics Group, University of Illinois and Beckman Institute, 2020.
- [29] J. C. Phillips *et al.*, “Scalable molecular dynamics with namd,” *Journal of computational chemistry*, vol. 26, pp. 1781–1802, 16 2005.
- [30] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. Academic Press, 2002, ISBN: 0-12-267351-4.
- [31] S.-C. Wang, *Interdisciplinary Computing in Java Programming*, 1st ed. Springer, Boston, MA, 2003, ISBN: 978-1-4020-7513-1.
- [32] S. A. Hollingsworth and R. O. Dror, “Molecular dynamics simulation for all,” *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018.
- [33] M. Karplus and G. A. Petsko, “Molecular dynamics simulations in biology,” *Nature*, vol. 347, pp. 631–639, 1990.
- [34] M. Moradi and E. Tajkhorshid, “Mechanistic picture for conformational transition of a membrane transporter at atomic resolution,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 18 916–18 921, 47 2013.
- [35] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From algorithms to Applications*, 2nd ed. Academic Press, 2002, ISBN: 0-12-267351-4.
- [36] M. P. Johansson, V. R. I. Kaila, and D. Sundholm, *Biomolecular Simulations: Methods and Protocols*, L. Monticelli and E. Salonen, Eds. Humana Press, 2013, ISBN: 978-1-62703-016-8.
- [37] L. Verlet, “Computer “experiments” on classical fluids thermodynamical properties of lennard-jones molecules,” *Physical Review*, vol. 159, pp. 98–103, 1 1967.

- [38] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford University Press, 1987, ISBN: 0-19-855375-7.
- [39] C. Sagui and T. A. Darden, “Molecular dynamics simulations of biomolecules: Long-range electrostatic effects,” *Annual review of biophysics and biomolecular structure*, vol. 28, pp. 155–179, 1999.
- [40] M. P. Allen, “Introduction to molecular dynamics simulation,” in *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, J. von Neumann Institute for Computing, Ed. Jülich: NIC Series, 2004, vol. 23, pp. 1–28, ISBN: 3-00-012641-4.
- [41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *J. Chem. Phys.*, vol. 81, pp. 3684–3690, 8 1984.
- [42] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, vol. 126, 014101 2007.
- [43] S. Nosé, “A unified formulation of the constant temperature molecular dynamics methods,” *J. Chem. Phys.*, vol. 81, pp. 511–519, 1984.
- [44] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Phys. Rev. A*, vol. 31, pp. 1695–1697, 3 1985.
- [45] D. S. Lemons and A. Gythiel, “Paul langevin’s 1908 paper “on the theory of brownian motion” [“sur la théorie du mouvement brownien,” c. r. acad. sci. (paris) 146, 530–533 (1908)],” *Am. J. Phys.*, vol. 65, pp. 1079–1081, 11 1997.
- [46] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, pp. 7182–7190, 12 1981.

- [47] G. J. Martyna, D. J. Tobias, and M. L. Klein, “Constant pressure molecular dynamics algorithms,” *J. Chem. Phys.*, vol. 101, pp. 4177–4189, 5 1994.
- [48] S. Izrailev *et al.*, “Steered molecular dynamics,” in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, P. Deuffhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 39–65, ISBN: 978-3-642-58360-5.
- [49] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, 14 1997.
- [50] C. Jarzynski, “Nonequilibrium work relations: Foundations and applications,” *The European Physical Journal B*, vol. 64, pp. 331–340, 3 2008.
- [51] G. E. Crooks, “Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems,” *Journal of Statistical Physics*, vol. 90, pp. 1481–1487, 5 1998.
- [52] D. L. Beveridge and F. M. DiCapua, “Free energy via molecular simulation: Applications to chemical and biomolecular systems,” *Annual Review of Biophysics and Biophysical Chemistry*, vol. 18, pp. 431–492, 1989.
- [53] J. Kästner, “Umbrella sampling,” *WIREs Comput. Mol. Sci.*, vol. 1, no. 6, pp. 932–942, 2011.
- [54] J. Leszczynski, *Challenges and Advances in Computational Chemistry and Physics. Quantum Modeling of Complex Molecular Systems*. Springer, 2015, vol. 21, ISBN: 978-3-319-21625-6.
- [55] C. J. Cramer, *Essentials of Computational Chemistry. Theories and Models*, 2nd ed. John Wiley & Sons Ltd, 2004, ISBN: 0-470-09181-9.

- [56] L. V. Kalé, A. Bhatele, E. J. Bohm, and J. C. Phillips, “Namd (nanoscale molecular dynamics),” in *Encyclopedia of Parallel Computing*, D. Padua, Ed. Boston, MA: Springer US, 2011, pp. 1249–1254, ISBN: 978-0-387-09766-4.
- [57] A. D. MacKerell *et al.*, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *The Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, 18 1998.
- [58] H. Berman, K. Henrick, and H. Nakamura, “Announcing the worldwide protein data bank,” *Nature Structural & Molecular Biology*, vol. 10, pp. 980–980, 12 2003.
- [59] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, “The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method,” *J Comput Chem*, vol. 13, pp. 1011–1021, 8 1992.
- [60] K. Karis, “The free energy of vesicular transmembrane lipid transfer studied with molecular dynamics simulations,” Master Thesis, Department of Theoretical Physics, School of Engineering Sciences. KTH Engineering Sciences, 2014.
- [61] M. I. Lobanov, N. S. Bogatyreva, and O. V. Galzitskaia, “Radius of gyration is indicator of compactness of protein structure,” *Mol Biol (Mosk)*, vol. 42, pp. 701–6, 4 2008.
- [62] N. V. Bhagavan and C.-E. Ha, *Essentials of Medical Biochemistry*, 2nd ed. Academic Press, 2015, ISBN: 0124166873.
- [63] C. Yang, E. Kim, and Y. Pak, “Potential of mean force simulation by pulling a dna aptamer in complex with thrombin,” *Bulletin of the Korean Chemical Society*, vol. 3, pp. 3597–3600, 11 2012.
- [64] P. Ball, “Water as an active constituent in cell biology,” *Chemical reviews*, vol. 108, pp. 74–108, 2008.

- [65] B. Alberts *et al.*, *Molecular Biology of the Cell*, 6th ed. Garland Science, Taylor & Francis Group, LLC, 2015.
- [66] G. A. Jeffrey, *An introduction to hydrogen bonding*. Oxford University Press, 1997.

# **Anexos**

# Anexo A

## Preparación del Sistema

### A.1 Script para separar las moléculas del complejo 4PDB

```
#!/bin/sh
# the next line restarts using tcsh |
exec tcsh "$0" "$@"

# Main Program
set file [lindex $argv 0]
set infile [open "$file.pdb"]
set outfile1 [open protein.pdb w]
set outfile2 [open arn.pdb w]
puts $outfile1 "REMARK pdb file for the Protein part of the
               complex in the pdb file $file (chain A)"
puts $outfile2 "REMARK pdb file for the ARN part of the
               complex in the pdb file $file (chain I)"

set text {}
while {$text != "ATOM"} {
    set pos [tell $infile]
    gets $infile line
    set text [string range $line 0 3]
}
seek $infile $pos start
gets $infile line
while {[string last "ATOM" $line] != -1} {
    puts $outfile1 $line
    gets $infile line
}
puts $outfile1 "END"
gets $infile line
while {[string last "ATOM" $line] != -1} {
    puts $outfile2 $line
    gets $infile line
}
puts $outfile2 "END"
close $outfile1
close $outfile2
```

## A.2 Script para la generación de los archivos PSF y PDB del complejo

```
#!/bin/csh

# Run the psfgen program, taking everything until "ENDMOL"
# as input.

psfgen << ENDMOL

# Read in the topology definitions for the residues we
# will create.
# This must match the parameter file used for the simulation
# as well.

topology top_all127_prot_na.rtf

pdbalias residue G GUA
pdbalias residue A ADE
pdbalias residue C CYT
pdbalias residue U URA

pdbalias residue HIS HSE
pdbalias atom ILE CDI CD

# Building segments of residues
segment arn {
    first none
    last none
    pdb arn.pdb
}

coordpdb arn.pdb arn

segment pro {
    pdb protein.pdb
}

coordpdb protein.pdb pro

guesscoord

writepsf aptamer.psf

# This ends the matching coordinate pdb file. The psf and pdb files
# are a matched set with identical atom ordering as needed by NAMD.

writepdb aptamer.pdb

ENDMOL
```

## A.3 Script para alinear el complejo en la dirección z

Este script, escrito en TCL, alinea el vector que une los centros de masa de las partes constituyentes del complejo con respecto al eje  $+z$ .

```
#####
#Procedure to find the angle between two vectors
proc agl {v1 v2} {
  if {[llength $v1] != [llength $v2]} {
    error "Vectors must have same dimension"
  }
  set pi [expr atan(1.0)*4]
  #set pi [expr acos(-1.0)]

  set mod1 [veclength $v1]
  set mod2 [veclength $v2]
  set dotprod [vecdot $v1 $v2]

  set frac [expr $dotprod/($mod1*$mod2)]
  set angle [expr acos($frac)]
  set angle [expr $angle*180/$pi]

  return $angle
}
#####
set molname aptamer

mol new ${molname}.psf
mol addfile ${molname}.pdb

set everyone [atomselect top all]
set prot [atomselect top "protein"]
set arn [atomselect top "nucleic"]

set com_prot [measure center $prot weight mass]
set com_arn [measure center $arn weight mass]
set vec [vecsub $com_prot $com_arn]

#
#set Mvec [transvecinv $vec]

#$everyone move $Mvec
#set $Mvec [transaxis y -90]
#$everyone move $Mvec

$everyone move [transvecinv $vec]
$everyone move [transaxis y -90]

$everyone writepsf ${molname}_alg.psf
$everyone writpdb ${molname}_alg.pdb

mol delete top

### Load the system aligned to z-axis
mol new ${molname}_alg.psf
mol addfile ${molname}_alg.pdb

set prot [atomselect top "protein"]
set arn [atomselect top "nucleic"]
```

```
set com_prot [measure center $prot weight mass]
set com_arn [measure center $arn weight mass]

set zvec {0 0 1}
set Mvec [vecsub $com_prot $com_arn]

set angulo [agl $zvec $Mvec]

puts "angle between vectors is: $angulo"
```

## A.4 Script de Solvatación del Complejo

Este script, escrito en lenguaje TCL, permite situar al complejo en una caja con agua.

```
set molname "aptamer_alg"
mol new ${molname}.psf
mol addfile ${molname}.pdb

package require solvate

solvate ${molname}.psf ${molname}.pdb -t 10 +z 25 -o ${molname}_wb
```

## Anexo B

# Minimización de la Energía

```
#####
## MINIMIZACION ##
#####

# Minimization of aptmaer complex, R COM vector between protein
# and RNA is aligned to z-axis
# The molecular complex is neutralized in a water box
# 3 Phosphures atoms of RNA were fixed (res 1 , 10 y 19)

#####
## ADJUSTABLE PARAMETERS ##
#####

set inputname          neuSystem

structure              ./input/$inputname.psf
coordinates            ./input/$inputname.pdb

set temperature        0
set outputname        ${inputname}_min
firsttimestep         0

#####
## SIMULATION PARAMETERS ##
#####

paraTypeCharmm        on
parameters            ../common/par_all127_prot_na.prm

## NOTE: don't use temperature for initial velocities when
##          .vel file is specified!
## Consider that in minimization this parameter isn't used
temperature           $temperature

# Force-Field Parameters
exclude               scaled1-4
l-4scaling            1.0
cutoff                11.0
switching             on
switchdist            9.0      ;# cutoff -2
```

```

pairlistdist      13.0      ;# cutoff +2
margin            1.0

#Integrator Parameters
timestep          1.0      ;#1 fs / step
rigidBonds        all
nonbondedFreq     1
fullElectFrequency 5
stepspercycle     10

#Periodic Boundary Conditions
cellBasisVector1  61.7      0.0      0.0
cellBasisVector2  0.0      73.6      0.0
cellBasisVector3  0.0      0.0      95.7
cellOrigin        9.029     -25.442   -10.964

wrapAll           on

#Electrostatic Force Evaluation
#PME (for full system periodic electrostatics)
PME               yes
PMEGridSizeX     64
PMEGridSizeY     75
PMEGridSizeZ     100

#Restrictions
fixedAtoms        on
fixedAtomsFile    ./input/aux_atoms.pdb
fixedAtomsCol     B

#Output
outputName        $outputname
restartfreq       100
dcdfreq          100
outputEnergies    50

#####
## EXTRA PARAMETERS                                     ##
#####

#####
## EXECUTION SCRIPT                                     ##
#####

#Minimization
minimize          1000

```

## Anexo C

### Calentamiento del Sistema

```
#####
## HEATING ##
#####

# Heating of aptmaer complex whose R COM between protein and RNA
# is aligned to z-axis
# Complex is neutralizaed in a water box.
# 3 Phosphorus of RNA were fixed (res 1 , 10 y 19)

#####
## ADJUSTABLE PARAMETERS ##
#####

set inputname      neuSystem

structure          ./input/$inputname.psf
coordinates        ./input/min_fr9.pdb
bincoordinates     ../1-5-minimizacion/${inputname}_min.coor
set temperature    0
set outputname     ${inputname}_heat
firsttimestep      0

#####
## SIMULATION PARAMETERS ##
#####

paraTypeCharmm    on
parameters        ../common/par_all127_prot_na.prm

## NOTE: don't use temperature for initial velocities when
##         .vel file is specified!
## Consider that in minimization this parameter isn't used

# Force-Field Parameters
exclude           scaled1-4
l-4scaling        1.0
cutoff            11.0
switching         on
switchdist        9.0      ;# cutoff -2
pairlistdist      13.0     ;# cutoff +2
```

```

margin                1.0

#Integrator Parameters
timestep              1.0      ;#1 fs / step
rigidBonds            all
nonbondedFreq         1
fullElectFrequency    5
stepspercycle         10

#Periodic Boundary Conditions
extendedSystem        ../1-5-minimizacion/${inputname}_min.xsc
wrapAll               on

#Electrostatic Force Evaluation
#PME (for full system periodic electrostatics)
PME                   yes
PMEGridSizeX          64
PMEGridSizeY          75
PMEGridSizeZ          100

#Restrictions
fixedAtoms            on
fixedAtomsFile        ./input/aux_atoms.pdb
fixedAtomsCol         B

#Output
outputName            $outputname
restartfreq           400 ;# 400 steps
dedfreq               400
outputEnergies        100

#####
## EXTRA PARAMETERS                                     ##
#####

#####
## EXECUTION SCRIPT                                     ##
#####

#Heating Protocol
temperature            $temperature
reassignFreq          25
reassignIncr          0.3875
reassignHold          310

#Execution
run                   20000

```

## Anexo D

### Equilibración del Sistema

```
#####
## EQUILIBRATION ##
#####

# Equilibration of aptamer complex whose R COM between
# protein and RNA is aligned to z-axis
# Complex is neutralized in a water box.
# 3 phosphorus RNA atoms were restricted a an elastic
# constant of 5 kcal/mol
# Eq. time 1ns

#####
## ADJUSTABLE PARAMETERS ##
#####

set inputname          neuSystem

structure              ./input/$inputname.psf
coordinates            ./input/heat_fr49.pdb
bincoordinates        ../1-6-calentamiento/${inputname}_heat.coor
binvelocities         ../1-6-calentamiento/${inputname}_heat.vel

set temperature        310 ;# for Langevin Control
set outputname         ${inputname}_eq
firsttimestep         0

#####
## SIMULATION PARAMETERS ##
#####

paraTypeCharmm        on
parameters            ../common/par_all27_prot_na.prm

## NOTE: don't use temperature for initial velocities when
##          .vel file is specified!
## Consider that in minimization this parameter isn't used

# Force-Field Parameters
exclude              scaled1-4
1-4scaling           1.0
```

```

cutoff          11.0
switching       on
switchdist      9.0      ;# cutoff -2
pairlistdist    13.0     ;# cutoff +2
margin          1.0

#Integrator Parameters
timestep        1.0      ;#1 fs / step
rigidBonds      all
nonbondedFreq   1
fullElectFrequency 5
stepspercycle   10

#Periodic Boundary Conditions
extendedSystem  ../1-6-calentamiento/${inputname}_heat.xsc
wrapAll         on

#Electrostatic Force Evaluation
#PME (for full system periodic electrostatics)
PME             yes
PMEGridSizeX   64
PMEGridSizeY   75
PMEGridSizeZ   100

##Restrictions
#fixedAtoms     on
#fixedAtomsFile ../input/fix_atoms.pdb
#fixedAtomsCol  B

constraints     on
consexp         2
consref         ../input/constrainatomsK5.pdb
conskfile       ../input/constrainatomsK5.pdb
conskcol        O

#Additional constraint for movement in z direction
selectConstraints on
selectConstrZ    off
selectConstrX    on
selectConstrY    on

#Control for constant temperature
langevin        on
langevinDamping 1 ;# gamma coefficient 1/ps
langevinTemp     $temperature
langevinHydrogen off

#Control for constant pressure
useGroupPressure yes ;# needed for rigid bonds
useFlexibleCell  no
useConstantArea  no

langevinPiston    on
langevinPistonTarget 1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod 200
langevinPistonDecay 50
langevinPistonTemp $temperature

#Output
outputName       $outputname
restartfreq      1000
dcdfreq          1000

```

```
velDCDfile      $outputname.veldcd
velDCDfreq      1000
outputEnergies  500
```

```
#####
## EXTRA PARAMETERS                                ##
#####
```

```
#####
## EXECUTION SCRIPT                                ##
#####
```

```
run              1000000    ;# 1ns
```

## Anexo E

### Simulación SMD

```
#####
## SMD ##
#####

## SMD on protein. 1 RNA atom is fixed ##
## constant k=5 ##
## smd atoms are CA of residues 35 54 113 130 ##
## force with constant velocity of 0.00025 ##

#####
## ADJUSTABLE PARAMETERS ##
#####

set inputname          neuSystem

structure               ./input/${inputname}.psf
coordinates             ./input/eq_fr999.pdb

binCoordinates         ./input/${inputname}_eq.coor
binVelocities          ./input/${inputname}_eq.vel
extendedSystem         ./input/${inputname}_eq.xsc

set temperature        310
set outputname         smdrefi_feri_ch

proc get_first_ts { xscfile } {
    set fd [open $xscfile r]
    gets $fd
    gets $fd
    gets $fd line
    set ts [lindex $line 0]
    close $fd
    return $ts
}

set firsttime [get_first_ts ./input/${inputname}_eq.xsc]

firsttimestep          0

#####
```

```

## SIMULATION PARAMETERS ##
#####

# Input
paraTypeCharmm      on
parameters          ../common/par_all127_prot_na.prm

##temperature      $temperature
##don't use temperature for initial velocities when .vel file is specified!

# Force-Field Parameters
exclude             scaled1-4
l-4scaling          1.0
cutoff              11.0
switching           on
switchdist         9.0      ;# cutoff -2
pairlistdist       13.0     ;# cutoff +2
margin              1.0

# Integrator Parameters
timestep            1.0      ;# 1 fs / step
rigidBonds          all
nonbondedFreq      1
fullElectFrequency  5
stepspercycle      10

# Constant Temperature Control
langevin            on
langevinDamping     1 ;# gamma coefficient 5/ps
langevinTemp        $temperature
langevinHydrogen    off

# Periodic Boundary Conditions
# It's not necessary to set cellBasisVectors1,2,3 and cellOrigin
# because this information are in .restart and .xsc files in Adjustable Parameters

wrapAll             on

#Electrostatic Force Evaluation
#PME (for full system periodic electrostatics)
PME                 yes
PMEGridSizeX       64
PMEGridSizeY       75
PMEGridSizeZ       100

#Control for constant pressure (variable volume)
useGroupPressure   yes ;# needed for rigid bonds
useFlexibleCell    no ;# no for water box
useConstantArea    no ;# no for waterer box

langevinPiston     on
langevinPistonTarget 1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod 200
langevinPistonDecay 50
langevinPistonTemp  $temperature

```



## Anexo F

# Equilibración del sistema para cada ventana

### F.1 Archivo de configuración

```
#####
##      UMBRELLA SAMPLING                      ##
#####

## US of COM between ARN & PROTEIN, window 0      ##
##                                                  ##
##                                                  ##

#####
## ADJUSTABLE PARAMETERS                        ##
#####

set inputname          aptamer_alg_wb_n

structure              ../../input/$inputname.psf
coordinates            ../../input/eq_fr99.pdb

binCoordinates        ../../input/rename/COMs-${num}.restart.coor
extendedSystem        ../../input/rename/COMs-${num}.restart.xsc

set temperature        310
set outputname        win${num}

firsttimestep         0

#####
## SIMULATION PARAMETERS                        ##
#####

# Input
```

```

paraTypeCharmm          on
parameters              ../../common/par_all27_prot_na.prm
temperature             $temperature

# Force-Field Parameters
exclude                 scaled1-4
l-4scaling              1.0
cutoff                  11.0
switching               on
switchdist              9.0      ;# cutoff -2
pairlistdist           13.0     ;# cutoff +2
margin                  1.0

# Integrator Parameters
timestep                1.0      ;# 1 fs / step
rigidBonds              all
nonbondedFreq           1
fullElectFrequency      5
stepspercycle           10

# Constant Temperature Control
langevin                on
langevinDamping          1 ;# gamma coefficient 5/ps
langevinTemp             $temperature
langevinHydrogen         off

# Periodic Boundary Conditions
# It's not necessary to set cellBasisVectors1,2,3 and cellOrigin
# because this information are in .restart and .xsc files in Adjustable Parameters

wrapAll                  on

#Electrostatic Force Evaluation
#PME (for full system periodic electrostatics)
PME                      yes
PMEGridSizeX             64
PMEGridSizeY             75
PMEGridSizeZ             108

#Control for constant pressure (variable volume)
useGroupPressure         yes ;# needed for rigid bonds
useFlexibleCell          no ;# no for water box
useConstantArea          no ;# no for water box

langevinPiston           on
langevinPistonTarget     1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod     200
langevinPistonDecay      50
langevinPistonTemp       $temperature

#Output
outputName               $outputname

restartfreq              200
dcdfreq                  100
xstFreq                  100

outputEnergies           200

```

```

outputPressure      200

#####
## EXTRA PARAMETERS                                     ##
#####

colvars             on
colvarsConfig       US-win${num}.in

constraints         on
consexp             2
consref             ../../input/rest01.ref
conskfile           ../../input/rest01.ref
conskcol           O

selectConstraints   on
selectConstrZ       off
selectConstrX       on
selectConstrY       on

#####
## EXECUTION SCRIPT                                     ##
#####

run                 105000

```

## F.2 Archivo Colvars

```

colvarsTrajFrequency 20
##colvarsRestartFrequency 20000

colvar {
  name Translocation

  width 0.1

  lowerboundary 20.2
  upperboundary 37.3

  lowerWall 20.2
  upperWall 37.3

  lowerwallconstant 100.0
  upperwallconstant 100.0

  distanceZ {
    main {
      atomnumbers {
1224 1238 1252 1264 1283 1293 1305 1322 1341 1355 1379 1398 1422 1436 1446 1460 1477
1493 1517 1534 1549 1571 1590 1605 1623 1635 1645 1656 1678 1697 1719 1741 1756 1775
1785 1800 1819 1838 1860 1884 1899 1906 1926 1945 1969 1981 1997 2012 2033 2052 2067
2079 2093 2115 2132 2139 2158 2177 2201 2220 2240 2259 2281 2302 2309 2319 2333 2347
2362 2386 2402 2421 2435 2442 2461 2483 2507 2526 2537 2561 2573 2580 2599 2623 2639
2660 2670 2692 2702 2714 2729 2747 2759 2783 2799 2818 2832 2839 2858 2865 2884 2894
2913 2929 2940 2954 2965 2987 2994 3010 3027 3041 3053 3075 3087 3097 3121 3138 3157
3174 3188 3195 3202 3217 3233 3249 3259 3280 3296

```

```
    }  
  }  
  ref {  
    atomnumbers {  
1   35  69 103 136 66 200 231 261 292 325 359 389 423 456 486 517 548 578 608  
639 673 707 741 774 804 837 867 898 931 965 999 1033 1064 1097 1127 1158 1189  
    }  
  }  
  axis ( 0.0, 0.0, 1.0 )  
}  
  
harmonic {  
  name Z  
  colvars Translocation  
  centers CENTER  
  forceConstant 0.025;# 2.5 * 0.1^2 = 0.025  
}
```