

Primer reporte:
Docking y dinámica molecular de la proteína Bcl-xL anclada a
membrana con el ligando Genistein

Luis Garreta

5 de septiembre de 2022

Índice

1. Preparación de las estructuras	1
1.1. Proteína Bcl-xL completa	1
1.2. Ligando Genistein	3
1.3. Proteína Bcl-xL anclada a membrana	3
2. Equilibración del sistema.	4
3. Acoplamiento Proteína-Ligando (<i>Docking</i>)	6
4. Selección de las poses de acomplamiento para análisis.	6
5. Simulación con Dinámica Molecular de 200ns para los 6 sistemas	8
5.1. Conteo de enlaces de hidrógeno	10
5.2. Radio de Giro (RG)	11
5.3. Root Mean Squared Deviation (RMSD).	12
5.4. Root Mean Square Fluctuation (RMSF)	13
5.5. Energías de interacción de enlaces no-covalentes (<i>non-bonded energies</i>).	14
5.5.1. Otras energías de interacción	15
5.6. Resumen	16
6. Disponibilidad	16

1. Preparación de las estructuras

1.1. Proteína Bcl-xL completa

Construimos un modelo de la estructura 3D de la proteína Bcl-xL anclada a membrana basados en los trabajos de [Tyagi et al., 2021] y [Ryzhov et al., 2020] en donde se combina dos proteína Bcl-xL: una que carece de la hélice C-terminal (PDB: 1lxl) y otra que solo tiene la hélice C-terminal (PDB 6f46). De la primera, que la vamos a llamar como *principal*, se tomó los residuos del 1 al 222 y de la segunda, que la vamos a llamar de aquí en adelante como *hélice*, se tomó la hélice α 8, residues 209-231 [Tyagi et al., 2021] (Figura 1).

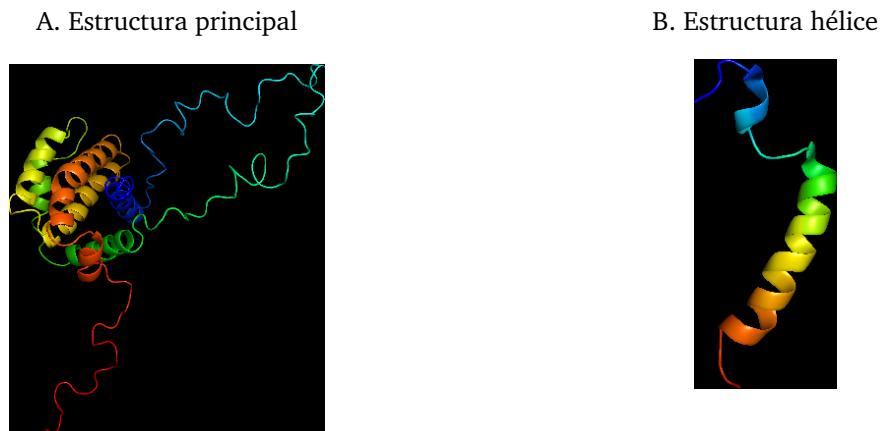


Figura 1: Estructuras de proteínas usadas para construir la proteína Bcl-xL anclada a membrana. (A) Bcl-xL que carece de la hélice C-terminal. (B) Bcl-xL con solo la hélice C-terminal.

El modelo lo construimos con la ayuda del software MODELLER para modelado comparativo [XXX]. Para esto, tomamos la secuencia completa de aminoácidos de la proteína *principal* Bcl-xL en formato fasta y construimos un alineamiento tomando como plantillas tanto a la estructura de la proteína *principal* (aminoácidos del 1 al 206) como a la estructura de la proteína *hélice* (aminoácidos 207 al 231) previamente alineada a lo largo del eje Z negativo. El archivo de alineamiento que construimos se ilustra en la Figura 2.

```
>P1;1lx1_sequence
sequence ::::::::::::
MSQSNRELVVDLFSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMTPSAINGNPSWHILADSPAVNGATG
HSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVNVNELFRDGVNWGRG
VAFFSFGGALCVESVDKEMQVLVSRIAAMATYLNDHLEPWIQENGWDTFVELYGNNAAAESRKQGERF
NRWFLTGMTVAGVVLLGSLFS*
```

```
>P1;1lx1_template
structureX:bclxl_MAIN.pdb:1:A:206:A:Bcl-xl:Lacking Helix 8::
MSQSNRELVVDLFSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMTPSAINGNPSWHILADSPAVNGATG
HSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVNVNELFRDGVNWGRG
VAFFSFGGALCVESVDKEMQVLVSRIAAMATYLNDHLEPWIQENGWDTFVELYGNNAAAESRKG————
*—————*
```

```
>P1;6f46_template
structureX:bclxl_TMD_ZZ.pdb:207:A:231:A:::
—————
```

```
NRWFLTGMTVAGVVLLGSLFS*
```

Figura 2: Archivo de alineamiento usado para la construcción de la estructura Bcl-xL completa.

Como resultado obtuvimos 10 modelos de los cuales tomamos como nuestra estructura Bcl-xL completa al mejor de ellos (Figura 3) de acuerdo a la función objetivo del software MODELLER.

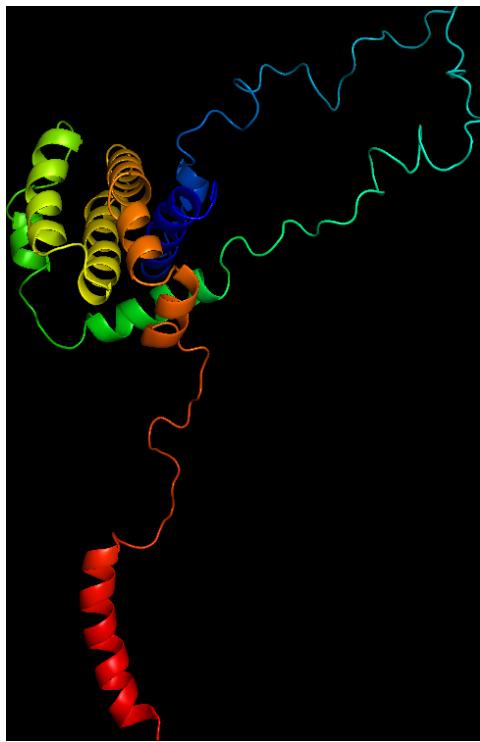


Figura 3: Mejor modelo 3D para la estructura Bcl-xL completa obtenido mediante MODELLER.

Finalmente, orientamos toda la estructura con respecto a la normal de la membrana (Eje Z por definición) usando como referencia las coordenadas pre-orientadas de la estructura *hélice* obtenidas de la base de datos de proteínas de membrana (OPM) (<https://opm.phar.umich.edu/>).

1.2. Ligando Genistein

La estructura química del compuesto Genistein la obtuvimos del banco de datos de proteínas Brookhaven PDB (PDB ID: GEN), de donde descargamos las coordenadas optimizadas en la versión ideal, en formato SDF (Figura 4).

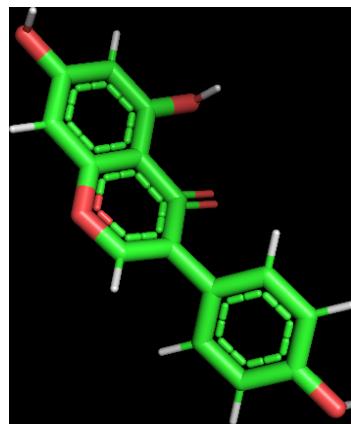


Figura 4: Estructura 3D ligando Genistein.

1.3. Proteína Bcl-xL anclada a membrana

La construcción del sistema inicial de proteína Bcl-xL anclada a membrana se realizó utilizando el módulo constructor de membranas del servidor CHARMM-GUI [Jo et al., 2007]. La proteína se ancló a una membrana tipo bicapa lipídica compuesta 100 % por palmitoyloleoylphosphati-dylcholine (POPC) con 412 moléculas en la capa

superior y 411 en la capa inferior. Todas las cadenas laterales de residuos ácidos y de histidina de la proteína fueron protonadas. El sistema se solvató usando un modelo de agua TIP3P con 100580 moléculas y se neutralizó con iones de KCl (197 K⁺, 105 Cl⁻), para un total de 415756 átomos y un tamaño de celda de simulación de 168.2 Å × 168.2 Å × 155.9 Å. Una imagen del sistema construido se muestra en la Figura 5.

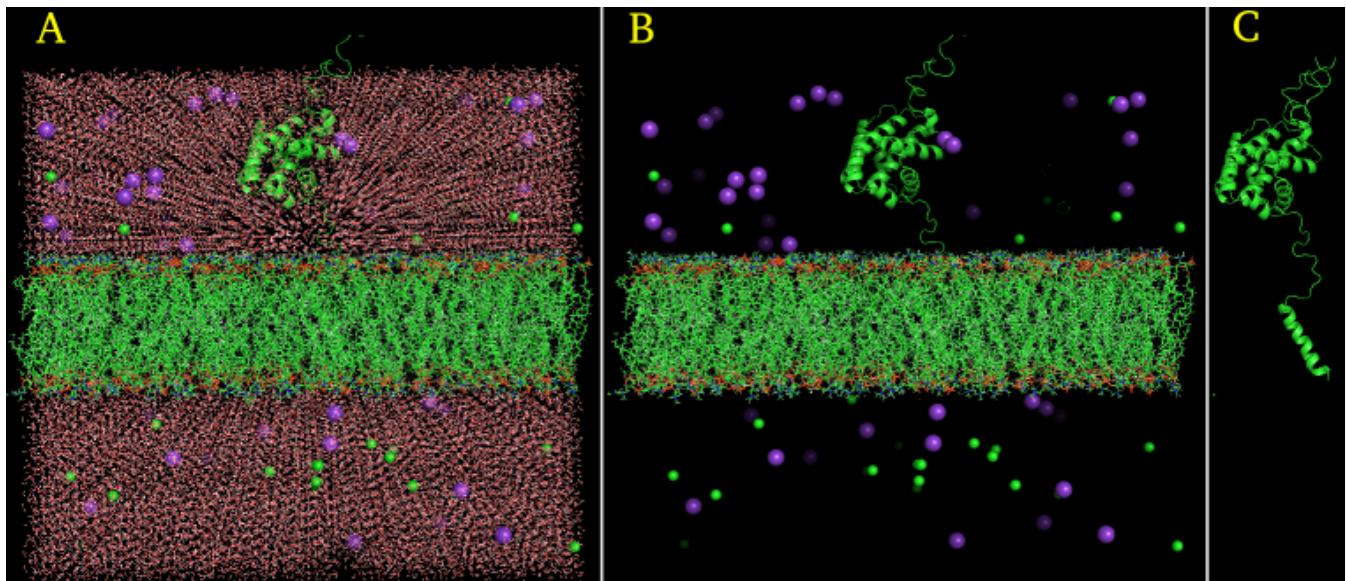


Figura 5: Sistema inicial de la proteína Bcl-xL anclada a membrana. (A) Proteína, bicapa lipídica, solvente y contraíones. (B) Sistema sin solvente. (C) Solo proteína.

2. Equilibración del sistema.

Al anterior sistema se le realizó primero un proceso de minimización de 10000 pasos, luego un proceso de simulación de equilibración de DM de 1ns, y finalmente una simulación de producción de DM de 20ns con un paso de tiempo de 2fs.

Para la equilibración se siguió el protocolo por defecto de CHARMM-GUI que realiza una equilibración gradual al sistema en seis etapas a través de la aplicación de diferentes restricciones aplicadas tanto a la proteína, al agua, a los contraíones, y a las moléculas de lípidos [Jo et al., 2007]. Para las primeras dos etapas (250ps) se usó una dinámica NVT (volumen y temperatura constante), mientras que para las restantes (750ps) se usó una dinámica NPAT (presión, área, y temperatura constante) a 303.15 K. La Figura 6 muestra el comportamiento a través del tiempo de las principales variables que describen la termodinámica del sistema durante el equilibrio.

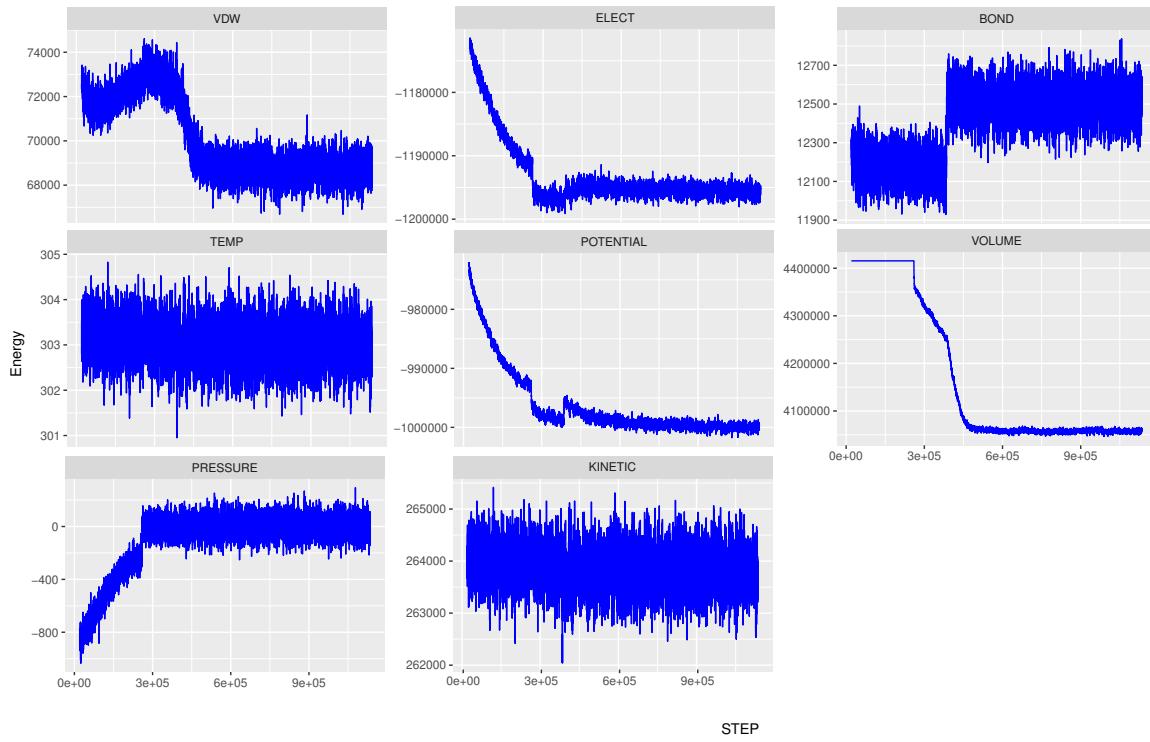


Figura 6: Comportamiento de las variables termodinámicas del sistema durante el equilibrio.

Después de relajar el sistema en la equilibración, ejecutamos una simulación de producción de DM de 10ns para evaluar la estabilidad estructural de la proteína en el tiempo. La Figura X muestra los perfiles de RMSD (equilibración y producción) del esqueleto de la proteína (*backbone*) para los átomos C α con referencia a la estructura inicial.

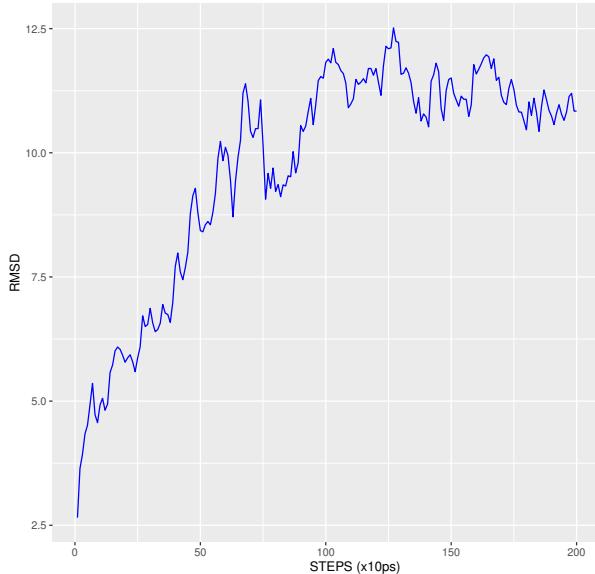


Figura 7: Perfil RMSD para átomos C α -atoms del esqueleto de la proteína. Primeros 10ns corresponden a la simulación de equilibración y los segundos 10ns corresponden a la simulación de producción.

3. Acoplamiento Proteína-Ligando (*Docking*)

Realizamos un acoplamiento o *docking* del ligando Genistein con la proteína Bcl-xL completa utilizando tres herramientas: autodock4 [Morris et al., 2009], autodock vina [Trott & Olson, 2010] y swissdock [Grosdidier et al., 2011]. Las dos primeras las ejecutamos mediante scripts de línea de comandos, mientras que la tercera utilizamos el servidor web de SwissDock (<http://www.swissdock.ch/>).

Para realizar el *docking* con las tres herramientas, tomamos como receptor la estructura de la proteína Bcl-xL completa (Figura 3), obtenida del último marco de los 10ns de la simulación de producción. El receptor contiene solo a la estructura de la proteína, sin membrana, ni el solvente, ni los contraiones. Mientras que el ligando lo tomamos de la estructura optimizada del compuesto de Genistein, versión ideal, descargada desde el PDB (Figura 4) (<https://www.rcsb.org/ligand/GEN>).

El tipo de acoplamiento que realizamos fue rígido y a ciegas o *blind docking*. Rígido ya que sólo se permite la flexibilidad del ligando y no del receptor, y a ciegas ya que se explora toda la superficie de la proteína sin especificar un sitio o región determinado. Sin embargo, para realizar una exploración más reducida, usamos una caja de cuadrícula cúbica centrada alrededor de la cabeza de la proteína (Figura 8).

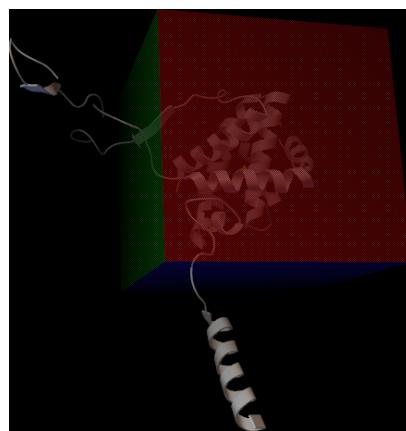


Figura 8: Caja de cuadrícula cúbica de acoplamiento centrada alrededor de la cabeza de la proteína.

4. Selección de las poses de acoplamiento para análisis.

Obtuvimos en total 326 resultados de posibles posiciones o *poses* del ligando en la proteína receptora: 50 poses de Autodock4, 20 poses de Vina, y 256 poses de SwissDock. Para reducir este número de poses para posteriores análisis, realizamos un agrupamiento jerárquico por cada conjunto de resultados y de cada grupo seleccionamos 10 poses para un total de 30 (Figura 9).

Realizamos dos procesos de selección, en el primero, la selección de las poses la hicimos de acuerdo a su posición buscando poses que estén ubicadas en diferentes sitios del receptor, en vez de poses con alta afinidad de unión (*affinity*) que pueden estar muy cercanas entre ellas. Además, las poses que produce cada herramienta ya vienen de una preselección interna en que tienen en cuenta esa afinidad de unión.

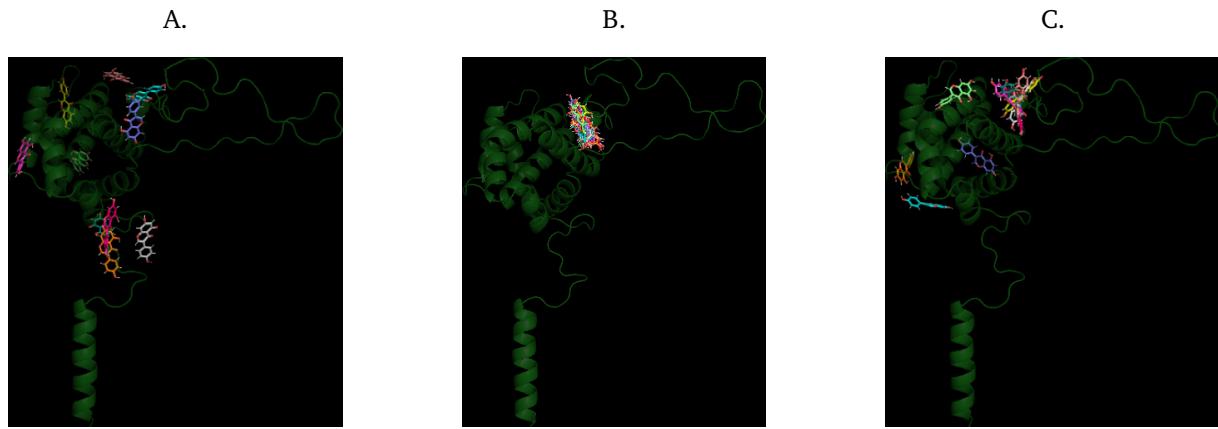


Figura 9: Poses de docking preseleccionadas de cada herramienta. (A) 10 Poses de Autodock4. (B) 10 Poses de Vina. (C) 10 Poses de SwissDock.

Para el segundo proceso de selección, todas las 30 poses resultantes de las tres herramientas de docking se agruparon jerárquicamente de acuerdo a su posición en diez grupos y se seleccionó de cada uno de ellos a la pose que visualmente estaba más cerca de las hélices que de los dominios de bucle, los cuales son generalmente muy flexibles. Después, con cada pose se construyó un sistema proteína-ligando, sin membrana ni solvente, y se le ejecutó una simulación de DM de 1ns con el objetivo de calcular la energía de enlace libre o FBE (*free binding energy*) para cada sistema.

La FBE la calculamos utilizando un enfoque de re-evaluación o *rescoring* que ofrecen las herramientas de Autodock4 y Vina—cada una de acuerdo a propia función de puntaje—para reevaluar una pose específica de un ligando en un receptor y estimar la energía de enlace libre. Para esto, obtuvimos todos los *frames* o cuadros de cada trayectoria resultante de la dinámica y reevaluamos cada uno de ellos tanto con la función de puntaje de Autodock4 como la de Vina. El promedio de estos cálculos de FBE por *frame* es el que tomamos como la FBE del sistema.

Teniendo en cuenta lo anterior, al final seleccionamos seis poses de docking que las nombramos como *ad22*, *av01*, *sw06*, *sw14*, *sw22*, y *sw25*, donde los prefijos *ad*, *av*, y *sw* se refieren a Autodock4, Vina, y SwissDock, respectivamente, que son las herramientas que generaron las respectivas poses. Las poses y su ubicación final en la proteína receptora se puede observar en la Figura X.

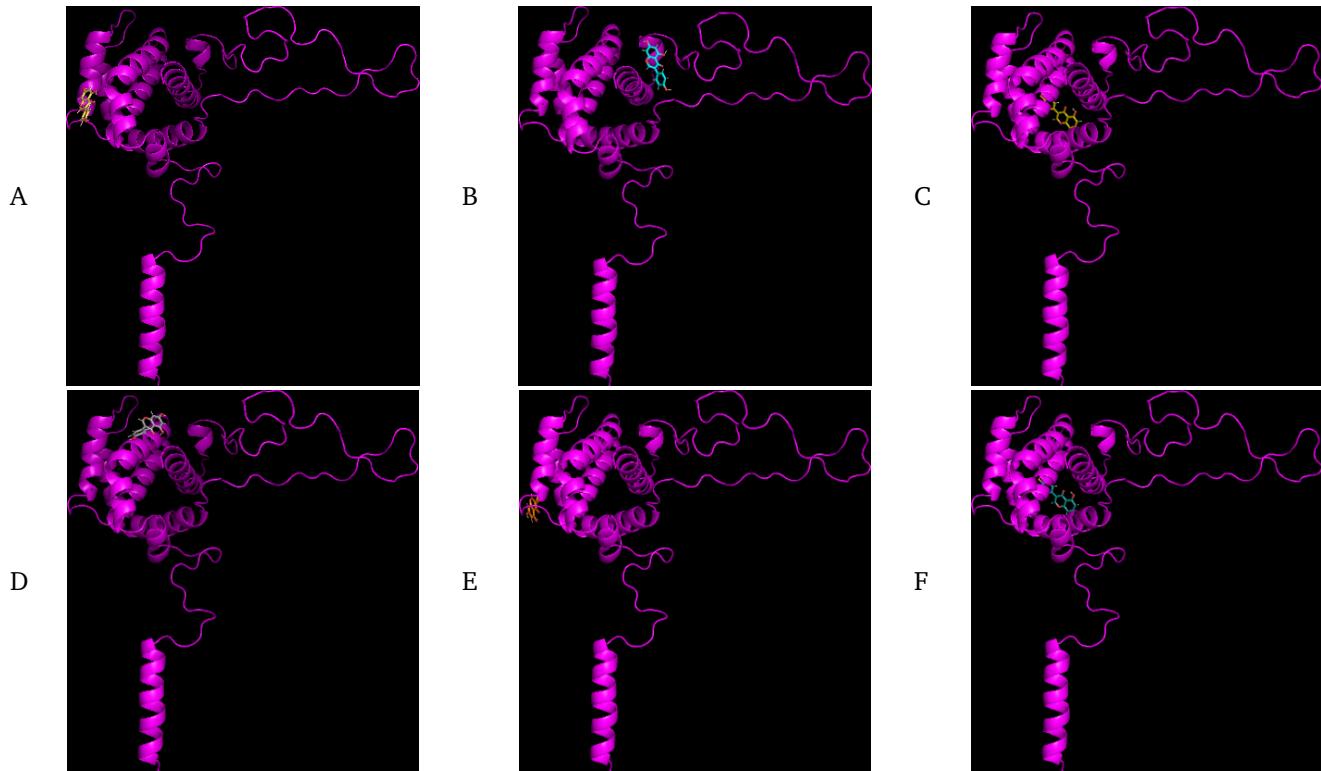


Figura 10: Poses de docking seleccionadas para análisis con dinámica molecular. (A) Pose ad22. (B) Pose av01. (C) Pose sw06. (D) Pose sw14 (E) Pose sw22. (F) Pose sw25.

5. Simulación con Dinámica Molecular de 200ns para los 6 sistemas

Con cada una de las poses construimos un complejo ligando-proteína anclado a membrana como el descrito en la sección 1.3, pero ahora adicionado una de las seis poses del ligando seleccionadas en el proceso anterior. A estos seis sistemas le realizamos una simulación de dinámica molecular de 200ns con un paso de tiempo de 2fs.

Para tener una idea de los sistemas que estamos simulando en la Figura 11 mostramos una instantánea o snapshot de uno de los sistemas ligando-proteína anclado a membrana, donde el solvente y los contraíones se han escondido solo para observar mejor el sistema. Mientras que en la Figura 11 presentamos algunas instantáneas representativas de la trayectoria seguida por cada complejo, con etiquetas en la parte superior que muestran el tiempo de la simulación en que se tomó la instantánea, y en la parte central está el nombre de la pose del ligando. La proteína se muestra en color azul plateado mientras que el ligando aparece de color verde claro.

De aquí en adelante cada uno de los seis sistemas los nombraremos de acuerdo al identificador de la pose del ligando, es decir, los resultados se presentan para los sistemas ad22, av01, sw06, sw14, sw22, y sw25.

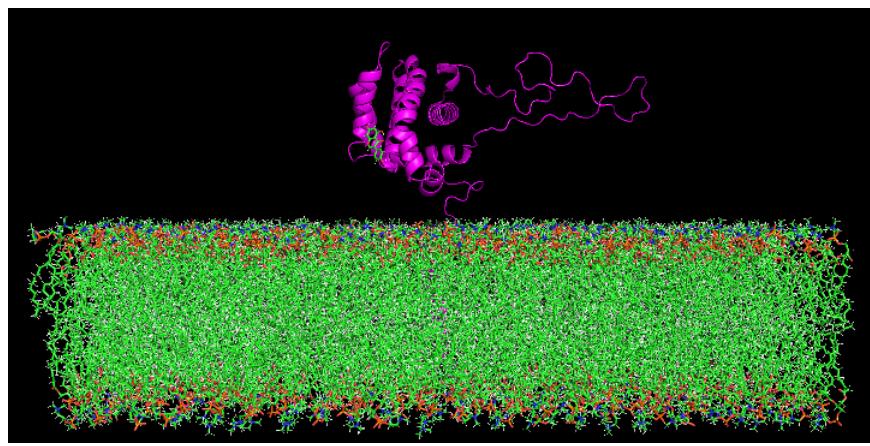


Figura 11: Uno de los sistemas iniciales proteína-ligando anclado a membrana.

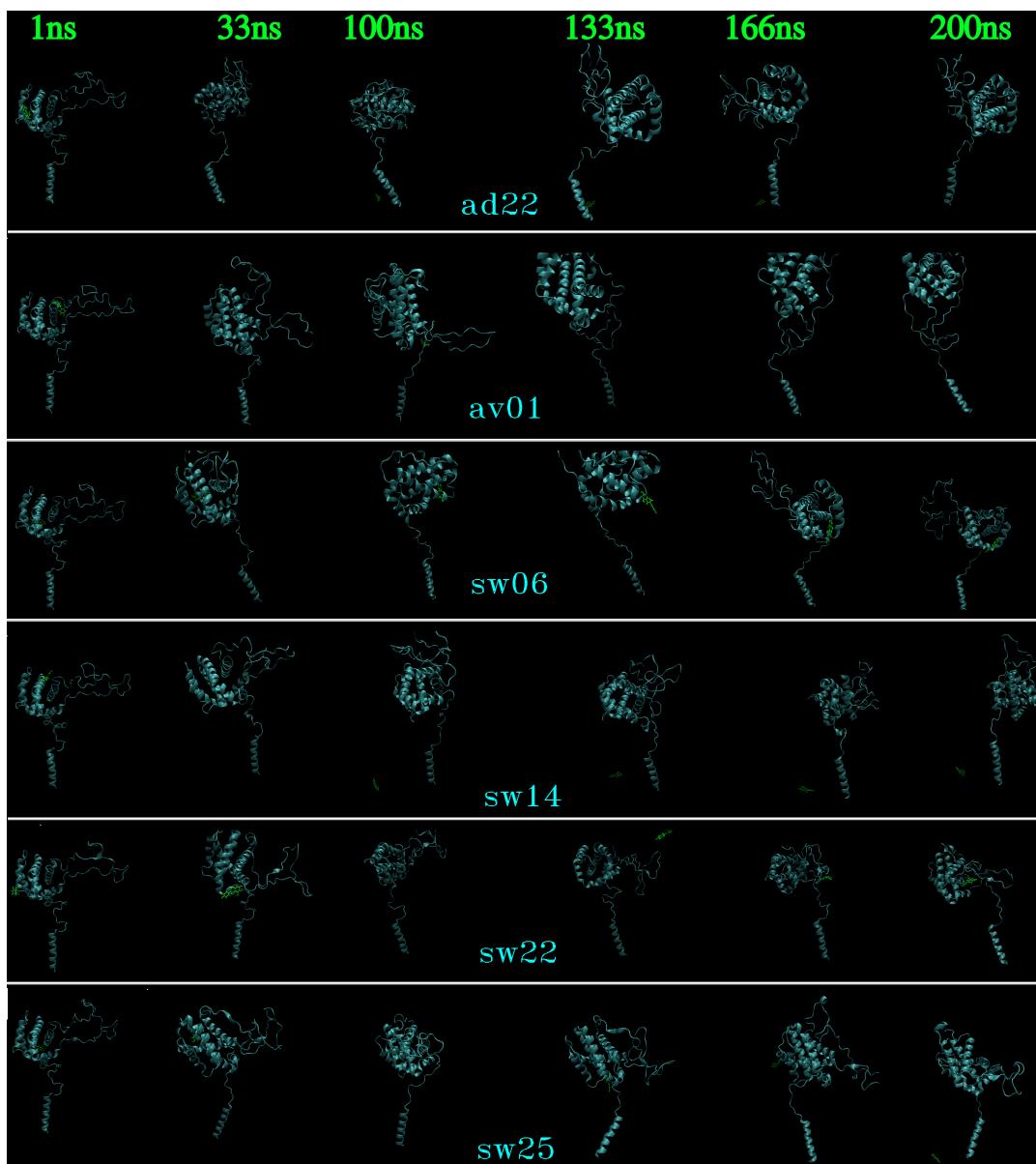


Figura 12: Imágenes trayectoria de la simulación de 200ns de los seis complejos ligando-proteína.

5.1. Conteo de enlaces de hidrógeno

La Figura 13 muestra el conteo de enlaces de hidrógeno para los seis sistemas. Observamos que solo el sistema *sw06* mantiene algunos enlaces hasta el final de la simulación. Mientras que el sistema *sw22* pierde totalmente los enlaces hacia la mitad de la simulación para luego volver a mantenerlos hasta el final. En los otros sistemas vemos que a partir de un tiempo específico, ya no se generan ningún enlace, lo que puede significar que el ligando ya no está unido a la proteína.

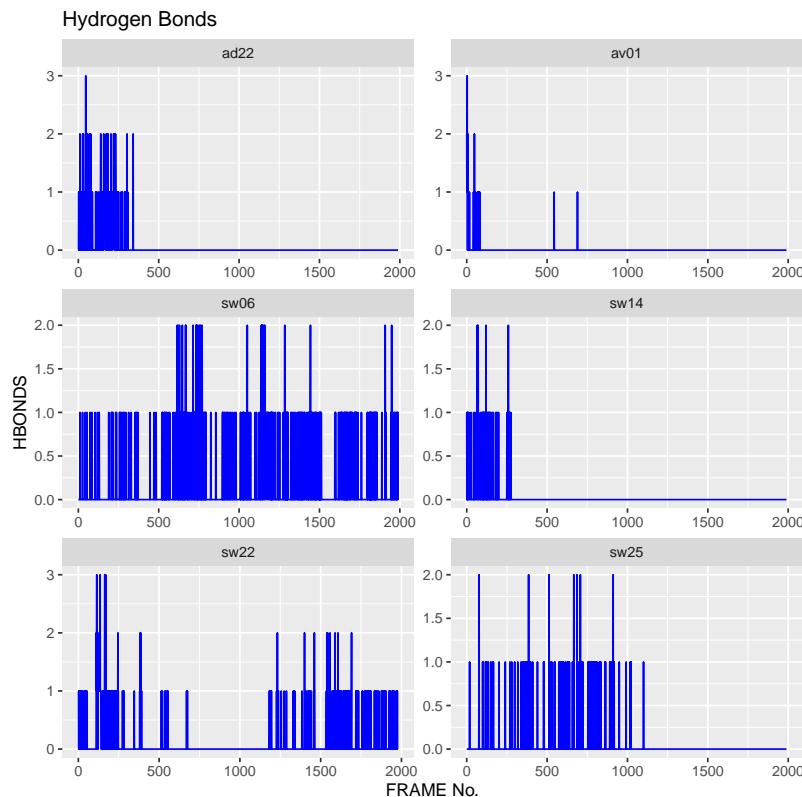


Figura 13: Conteo de enlaces de hidrógeno para los seis sistemas. El conteo se realizó sobre los enlaces de hidrógeno que se forman entre los átomos del ligando y los átomos de la proteína.

5.2. Radio de Giro (RG)

La Figura 14 muestra el radio de giro calculado para dos selecciones: para el complejo proteína-ligando (línea de color rojo) y para la proteína sola (línea de color azul). Observamos que en todos los sistemas el valor de RG en las dos selecciones es fluctuante. Sin embargo, vemos que en la gráfica de RG para el sistema *sw06* las dos selecciones tienen fluctuaciones iguales y por tanto el suponemos que el ligando permanece muy unido a la proteína.

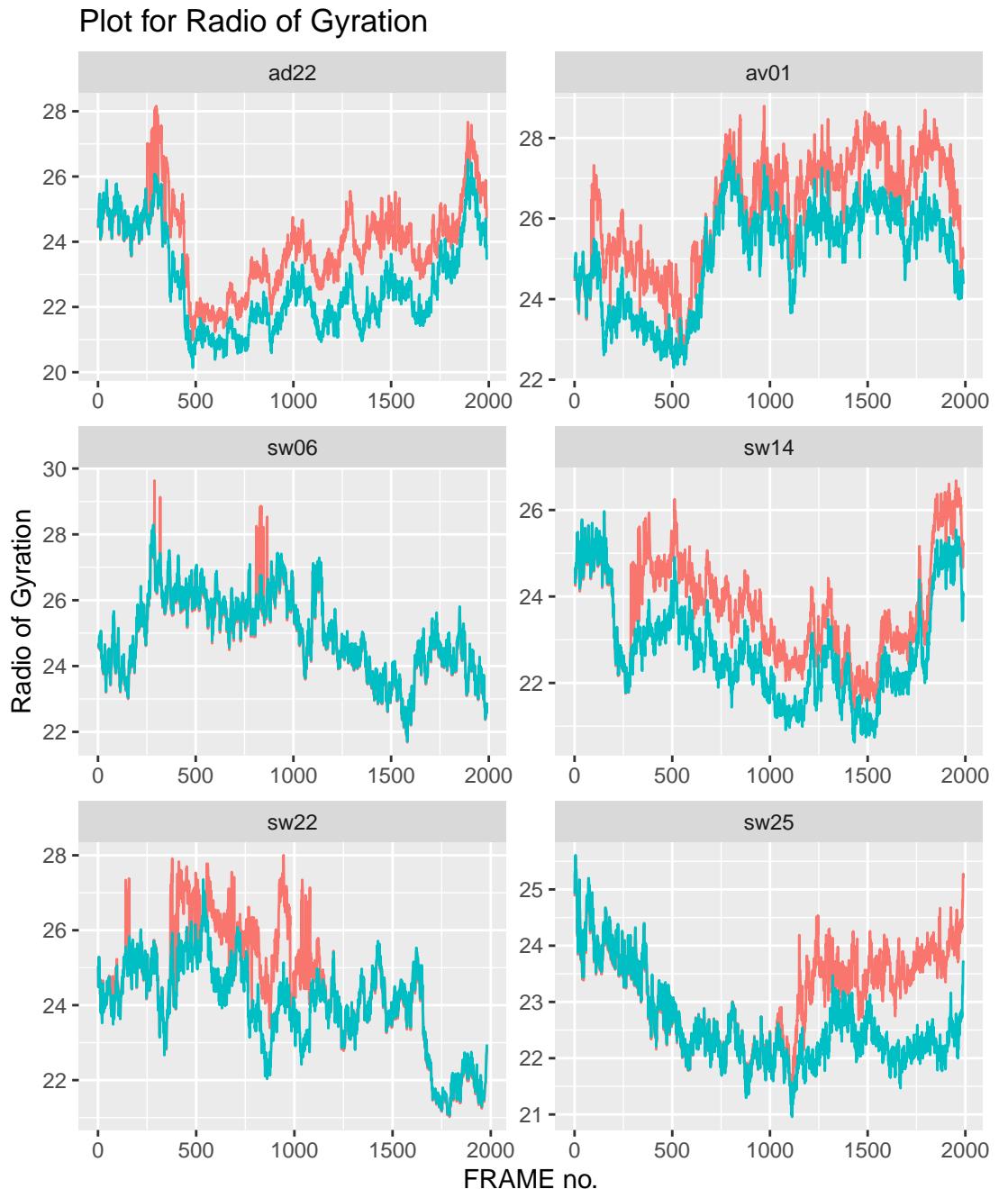


Figura 14: Radio de Giro para proteína-ligando en los seis sistemas. La línea roja corresponde al RG para el complejo proteína-ligando, mientras que la línea azul corresponde al cálculo de RG para la proteína sola.

5.3. Root Mean Squared Deviation (RMSD).

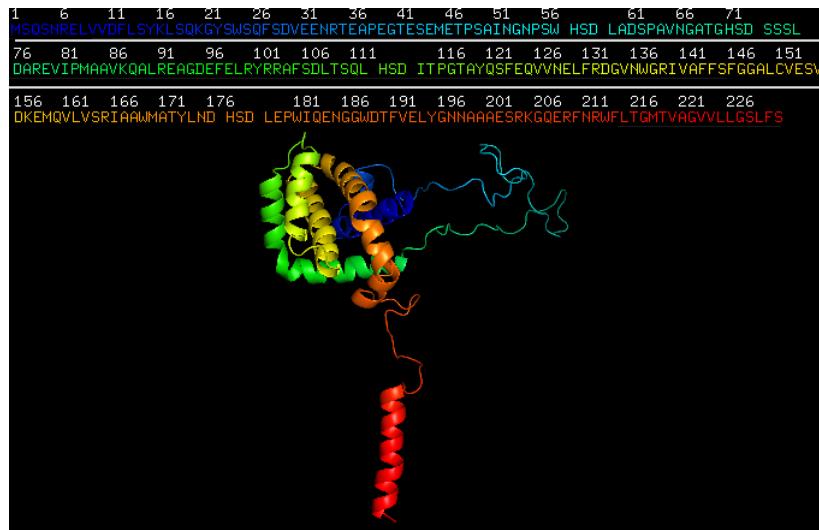
En la Figura 15 se presentan los valores de RMSD para los seis sistemas, calculados entre las estructuras simuladas a lo largo del tiempo y la estructura inicial, tanto para la proteína (línea azul) como para el ligando (línea roja). Observamos que todos los sistemas empiezan a estabilizarse aproximadamente a partir la mitad de la simulación (aprox. cuadro 1000). Sin embargo, en el sistema *sw06*, donde el ligando se mantiene unido a la proteína en el tiempo, vemos que después de un creciente RMSD, la proteína hace un cambio brusco y empieza a estabilizarse, diferente a los otros sistemas. En cuanto a la estabilidad del ligando, en general todavía se presentan grandes fluctuaciones y desviaciones, a excepción de los sistemas *sw22* y *sw14*, donde al final de la simulación las fluctuaciones disminuyen.



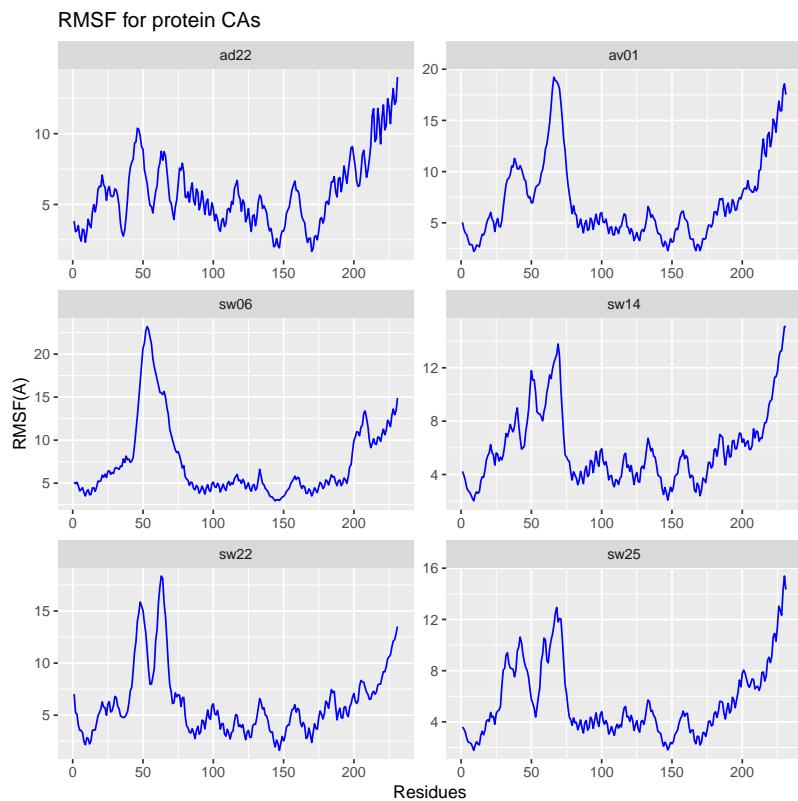
Figura 15: RMSD de la proteína y del ligando para los seis sistemas. El RMSD se calcula con referencia a la estructura inicial en la simulación, tanto de la proteína como del ligando. Cada paso de tiempo (Frame No.) corresponde a 2fs. En la parte superior de cada cuadro aparece el nombre del sistema.

5.4. Root Mean Square Fluctuation (RMSF)

La Figura 16 presenta los valores de RMSF para los seis sistemas junto con la proteína y residuos numerados. Los residuos que se tomaron para el cálculo fueron los carbonos alfa de la proteína. Aunque todos los sistemas en general presentan una misma tendencia, vamos a analizar la gráfica del sistema *sw06*. En esta gráfica podemos observar cuatro picos de importancia. El primero alrededor del residuo 50, que corresponde al loop flexible más grande (color cyan en la proteína). El segundo, es alrededor del residuo 205, que pertenece al loop que conecta la hélice C-terminal (color rojo), anclada a la membrana. También se observa una serie de fluctuaciones menores entre los residuos ~147 y ~153, que pertenecen a la hélice 5 de la proteína (color amarillo).



A.



B.

Figura 16: RMSFs para los residuos de la proteína. (A) Estructura de la proteína con el número del residuo y su color representativo dentro de la estructura de la proteína. (B) Valores de RMSF de cada residuo.

5.5. Energías de interacción de enlaces no-covalentes (*non-bonded energies*).

La Figura 17 muestra los gráficos de energías de interacción de enlaces no-covalentes para los seis sistemas. No observamos un patrón de interés y tampoco podemos establecer cual tipo de energía energía no-covalente es responsable de las interacciones con el ligando.

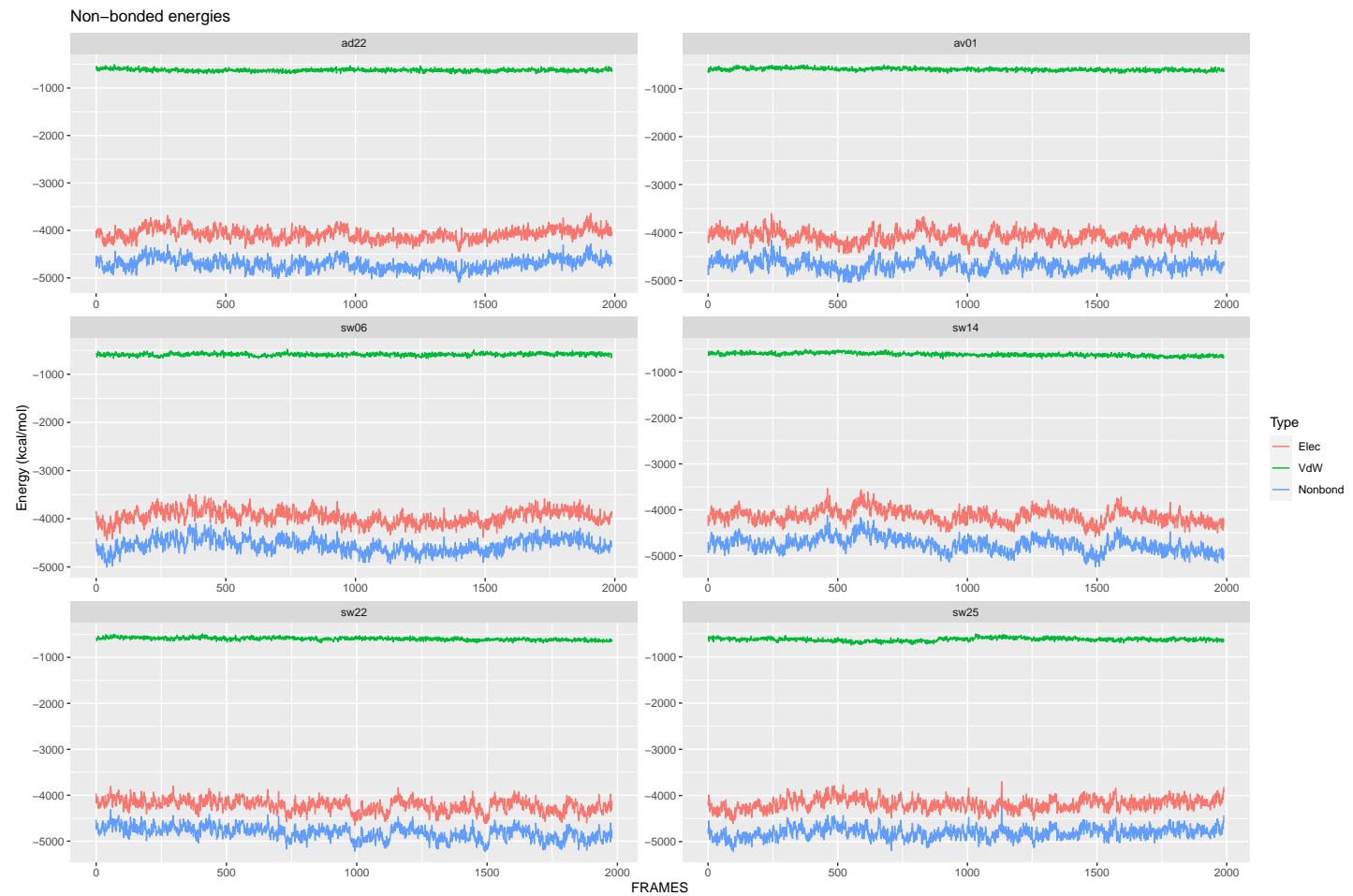


Figura 17: Energías de interacción de enlaces no-covalentes entre los átomos de las seis poses del ligando y los de la proteína receptora.

5.5.1. Otras energías de interacción

Además de las energías anteriores, también calculamos otras energías de interacción entre los residuos del ligando y los de la proteína. La Figura X muestra estas energías para el caso del sistema *sw06* ya que para los otros sistemas las interacciones o no existen o son muy débiles.

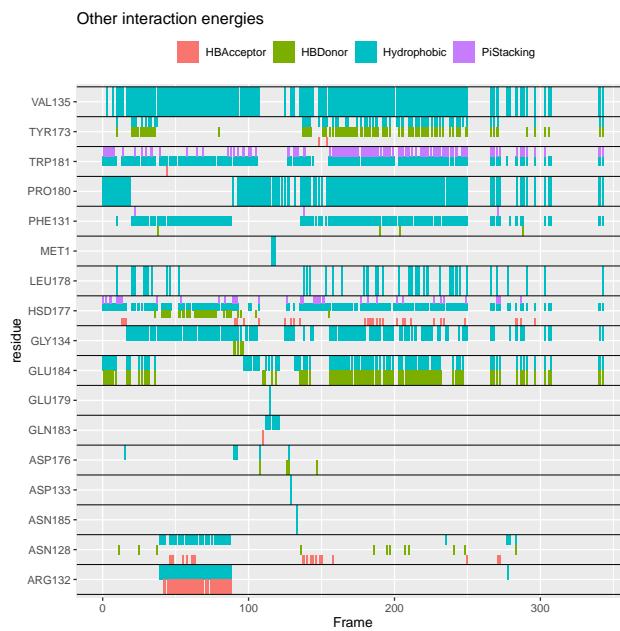


Figura 18: Otras energías de interacción de enlaces no-covalentes entre los átomos de las seis poses del ligando y los de la proteína receptora.

5.6. Resumen

La Figura X muestra los principales mediciones descritas anteriormente para el ligando sw06, que es el único sistema que se mantiene unido a la proteína al final de los 200ns de simulación.

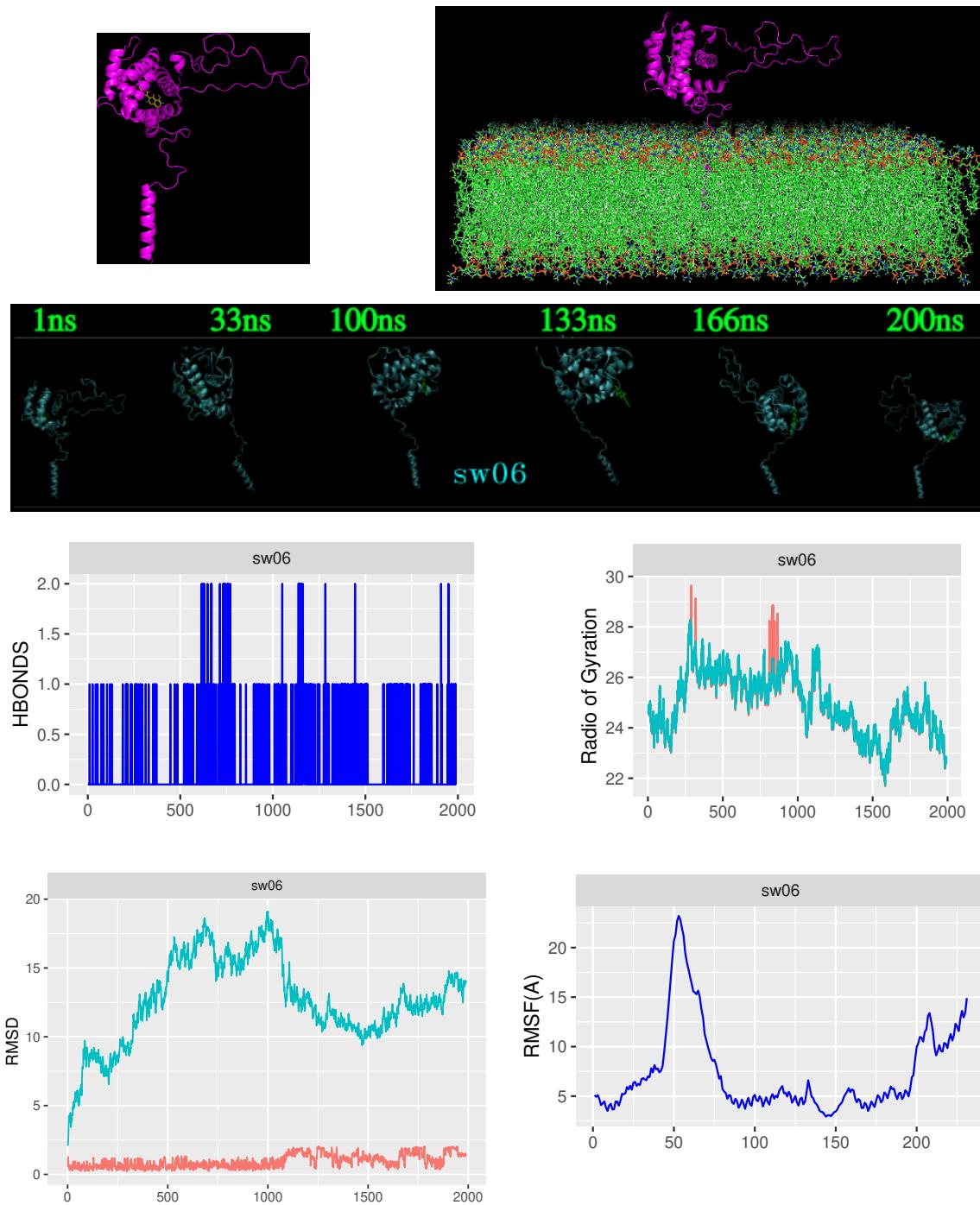


Figura 19: Principales gráficos de medidas realizadas sobre el sistema proteína-ligando anclado a membrana sw06.

6. Disponibilidad

- Los videos de las trayectorias se puede descargar desde el repositorio: <https://github.com/bioinfolg/mdsims.git>

References

- [Grosdidier et al., 2011] Grosdidier, A., Zoete, V., & Michelin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(SUPPL. 2), 270–277.
- [Jo et al., 2007] Jo, S., Kim, T., & Im, W. (2007). Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE*, 2(9), 1–10.
- [Morris et al., 2009] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785–2791.
- [Ryzhov et al., 2020] Ryzhov, P., Tian, Y., Yao, Y., Bobkov, A. A., Im, W., & Marassi, F. M. (2020). Conformational States of the Cytoprotective Protein Bcl-xL. *Biophysical journal*, 119(7), 1324–1334.
- [Trott & Olson, 2010] Trott, O. & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455–461.
- [Tyagi et al., 2021] Tyagi, V., Vasquez-Montes, V., Freites, J. A., Kyrychenko, A., Tobias, D. J., & Ladokhin, A. S. (2021). Effects of Cardiolipin on the Conformational Dynamics of Membrane-Anchored Bcl-xL.