

PROYECTO FINAL DE CARRERA INSTITUTO TECNOLÓGICO DE BUENOS AIRES



Análisis de complejos proteicos desordenados utilizando dinámica molecular como complemento al RMN: identificación de residuos claves y comportamiento dinámico

Autores:

- Maria Clara BASTIEN, Legajo 59171
- Franco Giuliano TAVOLARO, Legajo 58052

Tutores:

- Dra. María Laura FERNANDEZ
- Dra. Cristina MARINO BUSLJE

Trabajo Final presentado para la obtención del título de Bioingeniería

Buenos aires, Segundo Cuatrimestre 2023

Agradecimientos

En primer lugar, queremos agradecer a Laura y Cristina que nos apoyaron y tuvieron paciencia en este último paso de nuestro camino. Fueron dos personas fundamentales para llevar a cabo este proyecto, no dejaron de apostar en nosotros y en nuestro potencial. Sin dudas nos llevamos mucho conocimiento y experiencia de ustedes.

Gracias especiales a nuestras familias, que no solo nos acompañaron en estos largos años, sino que siempre estuvieron ahí para nuestros momentos difíciles dándonos las palabras justas para seguir adelante, y en los momentos buenos para festejar y brindar.

Gracias a nuestros amigos y amigas del ITBA, que fueron la excusa más grande para seguir yendo a la facultad y nos motivaron a siempre dar un poco más de nosotros. De más está decir que se convirtieron en amigos de la vida y esta carrera sin ustedes no hubiese sido posible.

También, agradecerle a nuestros amigos del colegio, del barrio y de la vida. Sin entender tanto, siempre estuvieron escuchándonos atentamente en cada pena y victoria, apoyándonos cuando no podíamos estar presentes y ofreciendo su distracción cuando la necesitábamos.

Este proyecto no solo es el resultado de nuestro trabajo, sino también del apoyo, la confianza y el empuje de cada uno de ustedes. ¡Esto también es de ustedes!

Parte de este trabajo ha sido presentado como ***Identifying key residues by MD in disordered protein fuzzy complexes. Beyond NMR.*** en XIII Argentine Congress of Bioinformatics and Computational Biology (XIII CAB2C), XIII International Conference of the Iberoamerican Society of Bioinformatics (XIII SoIBio), 2 de Nov 2023, Rosario, Argentina y en IV International Conference on Bioinformatics, Simulation and Modeling, Talca, Chile, 7 al 10 de Nov 2023.

Índice

1. Glosario	5
2. Resumen	6
3. Abstract	7
4. Introducción	8
5. Objetivos	15
6. Materiales y Métodos	16
6.1. Selección de proteínas:	16
6.2. Simulaciones por Dinámica Molecular:	27
6.3. Análisis:	28
6.4. Gráficos y Análisis de datos: Python, AlphaFold y Chimera	30
7. Resultados y discusión	31
7.1. Análisis de interacciones en los complejos, comparación entre DM y RMN.	31
7.1.1. 2GS0	31
7.1.2. 2M0G	36
7.1.3. 2L14	40
7.1.4. 5URN	44
7.1.5. 2ASQ	48
7.1.6. 2LPB	52
7.1.7. 2LY4	56
7.1.8. 1MV0	60
7.1.9. 2PX9	64
7.2. Otros análisis propuestos	67
7.2.1. Gyrate	67
7.2.2. Prueba de concepto	68
8. Conclusiones	69
9. Referencias	71
10. Anexo	75
10.1. Anexo 1: Parametros RING	75
10.2. Anexo 2: Matrices de Contacto	76
10.3. Anexo 3: Heatmap de Comparación	83
10.4. Anexo 4: Gráficos Gyrate	92

1. Glosario

- **DM** - Dinámica Molecular
- **RMN** - Resonancia Magnética Nuclear
- **IDP** - Proteínas Intrínsecamente Desordenadas (por sus siglas en inglés de *Intrinsically Disordered Proteins*)
- **IDR** - Regiones Intrínsecamente Desordenadas (por sus siglas en inglés de *Intrinsically Disordered Regions*)
- **RMSF** - Raíz de la Fluctuación Cuadrática Media (por sus siglas en inglés de *Root Mean Square Fluctuation*)
- **RMSD** - Raíz de la Desviación Cuadrática Media (por sus siglas en inglés de *Root Mean Square Deviation*)
- **CATH** - del inglés *Class, Architecture, Topology/fold, Homologus superfamily*
- **TEM** - Microscopía Electrónica de Transmisión (por sus siglas en inglés de *Transmission Electron Microscopy*)
- **ELA** - Esclerosis Lateral Amiotrofica
- **FTD** - Demencia Frontotemporal (por sus siglas en ingles de *Fronto Temporal Dementia*)
- **PDB** - del inglés *Protein Data Bank*
- **RING** - del inglés *Residue Interaction Network Generator*
- **GROMACS** - del inglés *Groningen Machine for Chemical Simulations*

2. Resumen

La espectroscopía de Resonancia Magnética Nuclear (RMN) ha sido durante mucho tiempo una herramienta valiosa para caracterizar las conformaciones moleculares de proteínas intrínsecamente desordenadas (IDPs), sin embargo, a menudo proporciona información limitada sobre la conformación estructural de sus complejos. Estas proteínas están asociadas con varios procesos biológicos clave, por lo tanto, su mal funcionamiento podría desencadenar enfermedades como cáncer, Parkinson, Alzheimer, etc. En este estudio, abordamos el desafío de aumentar la información conformacional a partir de datos de RMN para caracterizar los complejos de IDPs, empleando simulaciones de dinámica molecular (DM). La técnica de RMN resulta en un conjunto de estructuras, mientras que la MD proporciona información sobre el comportamiento de los complejos a lo largo del tiempo, mejorando nuestra comprensión de su dinámica estructural.

Nuestra investigación involucró 9 complejos proteína-proteína, donde para cada uno de ellos, se realizaron 2 simulaciones de DM utilizando dos conformeros de RMN diferentes como puntos de partida. En una segunda etapa, los resultados fueron analizados revelando que cada complejo muestra interacciones de residuos con probabilidades variables junto con una amplia gama de estados conformacionales. Destacablemente, identificamos interacciones centrales dentro de estos complejos, arrojando luz sobre los factores clave que influyen en su estabilidad y, finalmente, en su funcionalidad. Además, identificamos 3 complejos en los que las proteínas se separaron lo suficiente como para perder las interacciones centrales. Se necesitarán más datos para determinar si la función asociada requiere que estos complejos sean transitorios o lábiles.

Nuestros hallazgos demuestran que las simulaciones de dinámica molecular complementan los datos de RMN al proporcionar una visión más completa de las estructuras complejas. El enfoque de dinámica molecular perfecciona la información obtenida de la RMN, proporcionando un recuento más preciso de los contactos centrales en una interacción proteína-proteína a lo largo del tiempo. Además, nuestro estudio destaca el papel crucial de residuos específicos en el mantenimiento de la estabilidad y la integridad de las interacciones complejas. En general, la combinación de RMN y dinámica molecular demuestra ser una estrategia poderosa para desentrañar las complejidades de los complejos biomoleculares. En conjunto, este estudio permite una comprensión más profunda del comportamiento y la función de los complejos de IDP y, eventualmente, esta información podría utilizarse para explorar nuevas estrategias de tratamiento de enfermedades.

3. Abstract

Nuclear Magnetic Resonance (NMR) spectroscopy has long been a valuable tool for characterizing intrinsically disordered proteins (IDPs) molecular conformations, yet it often provides limited information about structural conformation of their complexes. These proteins are associated with several key biological processes, therefore their malfunction could trigger diseases such as cáncer, Parkinson, Alzheimer, etc. In this study, we address the challenge of increasing the conformational information from NMR data to characterize IDPs complexes, by employing molecular dynamics simulations (MD). The NMR technique results in an ensemble of structures, whereas MD provides information on the behavior of the complexes over time, improving our understanding of their structural dynamics.

Our investigation involved 9 protein-protein complexes, for each one, 2 MD simulations were performed using two different NMR conformers (complexes) as starting points. In a second stage, the results were analyzed revealing that each complex exhibits residue interactions with varying probabilities together with a wide range of conformational states. Notably, we identified core interactions within these complexes, shedding light on the key factors that influence their stability and eventually, their functionality. Furthermore, we identified 3 complexes where the proteins separated enough to lose the core interactions, more data will be needed to determine if the function associated requires these complexes to be transient or labile.

Our findings demonstrate that molecular dynamics simulations complement NMR data by providing a more comprehensive view of complex structures. The molecular dynamics approach refines the information obtained from NMR, yielding a more accurate count of core contacts within a protein-protein interaction over time. Moreover, our study highlights the pivotal role of specific residues in maintaining the stability and integrity of complex interactions. Overall, the combination of NMR and molecular dynamics proves to be a powerful strategy for unraveling the intricacies of biomolecular complexes. Taken together, this study allows for a deeper understanding of the behavior and function of IDP complexes and eventually, this information could be used to explore new disease treatment strategies.

4. Introducción

Las proteínas son polímeros de aminoácidos; en la naturaleza hay 20 aminoácidos distintos con dos partes, una constante y otra variable. La parte constante es la columna vertebral o backbone que es igual para todos los aminoácidos y se compone de un grupo amino, un grupo carboxilo, un carbono denominado alfa y un hidrógeno. La parte variable permite diferenciar estructural y fisicoquímicamente a los aminoácidos entre sí y se denomina cadena lateral. En el ejemplo de la Figura 1 se representa el aminoácido alanina cuya cadena lateral es un metilo.

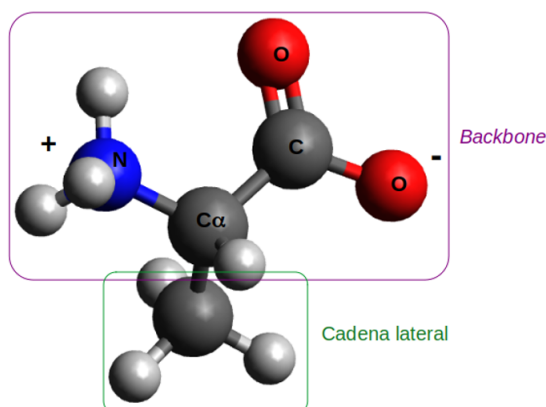


Figura 1: Estructura de un aminoácido (alanina).

La unión entre aminoácidos es lineal, se produce exclusivamente mediante el *backbone* y se da entre el grupo carboxilo del aminoácido “n” y el grupo amino del aminoácido “n+1”. Éstos son enlaces covalentes del tipo amida y se los conoce como uniones peptídicas, por ello se nombra a las proteínas como polipéptidos.

La composición de aminoácidos y la interacción entre ellos determinarán todos los niveles estructurales de las proteínas, ya sea la estructura primaria, secundaria, terciaria y/o cuaternaria. Se le llama estructura primaria a la secuencia ordenada y continua de aminoácidos. La estructura secundaria está dada por el plegamiento de estructuras locales sencillas, mediado casi exclusivamente por la formación de puentes de hidrógeno entre átomos del *backbone*; las estructuras secundarias más comunes son las hélices alfa y las hojas beta (ejemplo Figura 2a). La estructura terciaria de una proteína se define por la disposición tridimensional global siendo determinada por la organización espacial de los segmentos que presentan estructura secundaria, así como por aquellos segmentos que carecen de estructura definida (ejemplo Figura 2b). La estructura terciaria es la configuración tridimensional funcional de muchas proteínas. Sin embargo, existen otras proteínas donde la estructura funcional es el resultado de la interacción de 2 o más cadenas polipeptídicas entre sí. La estructura cuaternaria es el arreglo tridimensional resultante de la interacción y ensamblaje de múltiples cadenas



Figura 2: Diferentes niveles de representación (*cartoon*) del heterodímero 2DYP[1]. a) Estructura secundaria: hojas beta en violeta, hélices alfa en naranja b) Estructura terciaria por cadena: cadena A en celeste y cadena B en amarillo. Estructura cuaternaria: interacción entre ambas cadenas para configurar la forma tridimensional funcional del complejo.

La Figura 3 muestra una clasificación jerárquica de los dominios de proteínas (según CATH[2]). Al agrupar los dominios de las proteínas estructuralmente, se observa que existen dominios compuestos casi exclusivamente por hélices alfa, otras cuya estructura es solo hojas beta u otras donde la estructura final es la combinación hélices alfa u hojas beta en distintas proporciones.

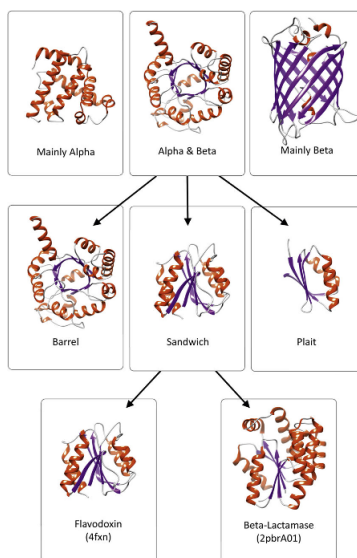


Figura 3: Clasificación jerárquica según CATH, panel superior: clase (según el contenido de estructura secundaria), panel medio: arquitectura (organización espacial gruesa de estructuras secundarias). Panel inferior: topología o plegamiento (organización espacial semejante de estructuras secundarias)

Las proteínas son las macromoléculas más versátiles en los sistemas vivos y cumplen funciones cruciales en casi todos los procesos biológicos. Numerosas funciones fisiológicas dependen de la correcta interacción entre proteínas, entre proteínas y ácidos nucleicos, proteínas y lípidos, etc. Está demostrado que la alteración de dichas interacciones puede inducir la aparición de procesos patológicos[3]. Las interacciones entre macromoléculas subyacen a todo proceso biológico y están reguladas por factores intrínsecos de las proteínas interactuantes, así como por factores extrínsecos como condiciones celulares y ambientales en las que se encuentran (estrés, pH, presión osmótica, fuerza iónica, señales químicas, etc.).

Existe una estrecha relación entre la secuencia, la estructura y la función de la proteína; se define la estructura nativa como la estructura tridimensional en condiciones fisiológicas. Sin embargo, las proteínas en

su estado nativo, no son elementos rígidos, sino un ensamble dinámico de estructuras coexistentes, llamadas conformeros. La importancia de la dinámica de una proteína en relación a sus funciones biológicas es cada vez más evidente.

Si las proteínas presentan muy pequeñas diferencias en su estructura tridimensional entre conformeros se denominan rígidas, si presentan marcadas diferencias se denominan móviles [4]. Hay un grupo de proteínas que carecen de estructura secundaria definida y son extremadamente móviles, a éstas se las denomina proteínas intrínsecamente desordenadas (IDP por *Intrinsically disordered proteins* en inglés); otro grupo de proteínas son aquellas que tienen algunas regiones desordenadas (IDR por *Intrinsically disordered regions* en inglés), tal como se observa en la Figura 4.

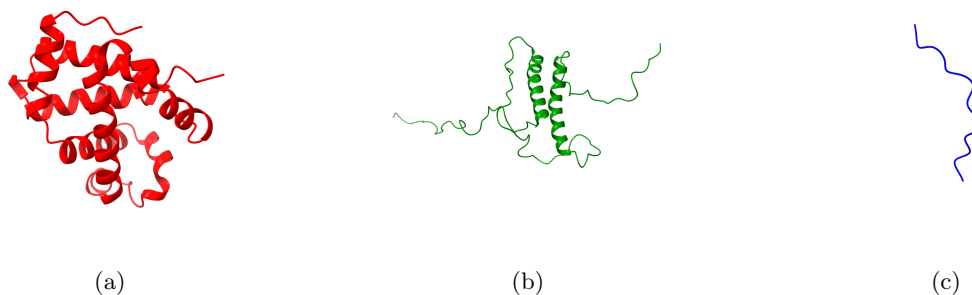


Figura 4: Grados de organización de las proteínas: a- proteína ordenada (Hemoglobina), b- proteína con IDRs (Splicing Factor 1) c- IDP (Myc proto-oncogene)

Se estima que alrededor del 30 % al 40 % de las proteínas en el proteoma humano pueden contener regiones desordenadas. Estas proteínas llevan a cabo infinidad de funciones celulares fisiológicas, pero a la vez están involucradas en numerosos procesos patológicos [3], lo que las convierte en buenas dianas para el desarrollo de drogas[5]. Las proteínas con algún grado de desorden se han encontrado involucradas en enfermedades neurodegenerativas como por ejemplo Parkinson, Alzheimer, Esclerosis Lateral Amiotrófica (ELA), Demencia Frontotemporal (FTD) [6] y cancer[7], por citar algunos ejemplos.

Para conocer la estructura tridimensional de las proteínas encontramos varios métodos experimentales, los más utilizados son: cristalografía de rayos X, criomicroscopía electrónica y espectroscopía de resonancia magnética nuclear.

Por medio de la cristalografía de rayos X es posible obtener la estructura tridimensional de una molécula o un conjunto de moléculas en estado cristalino. Esta técnica permite determinar la posición de cada uno de los átomos del cristal como coordenadas espaciales. Para obtener la estructura de biomoléculas es necesario partir de una muestra altamente purificada en solución. Luego se induce la formación de cristales bajo condiciones específicas que en un paso posterior serán expuestos a un haz de rayos X. Las nubes electrónicas de los átomos de las moléculas que componen el cristal dispersan los rayos X generando un patrón de difracción, éste es registrado mediante un detector de alta sensibilidad. A partir de este patrón de difracción es posible reconstruir la posición espacial de los átomos estudiados[8]. Si bien esta técnica es la más utilizada, la molécula a determinar no se encuentra en un ambiente fisiológico, ya que se está en un estado de agregación particular (cristal) que no es natural.

Frente a esta limitación se recurre a otras técnicas, por ejemplo: a la criomicroscopía electrónica, que es un procedimiento derivado de la microscopía electrónica de transmisión (TEM por *Transmission Electron Microscopy* en inglés). Aquí se emite un haz de electrones para interactuar con la muestra, la interferencia entre electrones dispersos y no dispersos da como resultado el llamado contraste de fase y la formación de imágenes. Debido a que los microscopios electrónicos requieren alto vacío, las muestras hidratadas no pueden examinarse con este método a temperatura ambiente, para ello se implementa la congelación criogénica de

las muestras. Esta técnica no requiere cristalización y puede ser aplicada en proteínas o complejos proteicos de gran tamaño. Dado que las imágenes obtenidas representan proyecciones bidimensionales (2-D) de un objeto tridimensional (3-D), se necesitan muchas imágenes de una proteína en diferentes orientaciones para reconstruir computacionalmente su estructura 3-D[9].

La espectroscopía de resonancia magnética nuclear (RMN) es utilizada en proteínas purificadas de tamaño pequeño en solución. Mediante un imán se aplica un campo magnético fuerte, lo que genera un alineamiento espacial de los átomos que tengan spin mayor a 0 (ej. el ^1H , ^{13}C , ^{15}N , etc.). El alineamiento se produce en dos direcciones: una paralela al campo (menor energía) y otra antiparalela (mayor energía). Mediante un pulso de energía electromagnética ortogonal al campo anterior, es posible desestabilizar el equilibrio de alineamiento donde los átomos vuelven a su estado original emitiendo ondas de radio que son detectadas por el equipo de RMN. Cada átomo emite una onda diferente que es dependiente del entorno químico (átomos vecinos). Al ser ^1H el átomo con momento magnético más fuerte, es el espectro obtenido del mismo el más utilizado para este método. De la comparación con ondas patrón de núcleos en distintos entornos se puede descifrar la estructura de la proteína a partir de su espectro RMN. La señal es detectada por una antena y se digitaliza para el procesamiento y visualización[10]. Como se mencionó, si bien la RMN es una técnica poderosa, puede tener limitaciones en tamaño de proteínas y complejidad estructural en comparación con otras técnicas, como la cristalografía de rayos X. Sin embargo, la RMN tiene la ventaja de trabajar en condiciones más cercanas a las fisiológicas y proporcionar información dinámica sobre las biomoléculas.

El resultado final de estas técnicas es un archivo de texto con las coordenadas X,Y y Z de cada átomo que en general se guarda con un formato específico y de uso generalizado (pdb por Protein Data Bank). Cada molécula resuelta se deposita como un archivo en el banco de datos de estructura de proteínas (Protein Data Bank)[11]. Es una base de datos de acceso libre donde se hallan los datos estructurales de macromoléculas biológicas (mayoritariamente proteínas) y ligandos resueltos mediante las técnicas experimentales mencionadas previamente. Esta fuente sirve como punto de partida para muchas investigaciones y estudios en el campo de la biomedicina estructural.

Dinámica Molecular (DM), es una herramienta de simulación computacional que permite estudiar el movimiento y la interacción entre átomos y moléculas. Este método fue desarrollado en principio para el campo de la física a fines de 1950, hoy en día es utilizado en muchos más campos, dentro de ellos se destacan la Física, Química, Biología Estructural y el modelado de materiales. Como se mencionó previamente las proteínas no son estructuras rígidas sino que son dinámicas e interactúan con su entorno, presentando cambios conformacionales y movimientos que tienen estrecha relación con la funcionalidad de las mismas, el proceso en el que se ven involucradas, o que estén en condiciones tanto fisiológicas como patológicas. Estas simulaciones utilizan un modelo dinámico permitiendo el desarrollo de estudios específicos donde se proporciona una trayectoria con información detallada de los movimientos de la proteína, y las coordenadas atómicas a lo largo de un determinado tiempo. La DM proporciona una alternativa a experimentos que por costos o tiempos no pueden ser llevados a cabo, o también ser un gran complemento a otros estudios. Los resultados obtenidos de la simulación son a nivel nanoscópico y muchas funciones de las proteínas dependen de estas interacciones que los métodos experimentales no logran capturar.

La DM simula un periodo de tiempo determinado permitiendo que los átomos y moléculas interactúen entre sí. El algoritmo de DM esta formado por los siguientes pasos:

1. **Ingreso de Condiciones Iniciales** - Las condiciones iniciales están dadas por las posiciones de los átomos (r) a partir del archivo de configuración inicial y por las velocidades de estos (v) asignadas mediante el algoritmo de velocidad de Maxwell-Boltzmann generado por números aleatorios
2. **Calculo de Fuerzas** - La fuerza sobre cualquier átomo se calcula con la siguiente ecuación:

$$F_i = -\frac{\partial V}{\partial r_i}, i = 1, \dots, N \quad (1)$$

La sumatoria de las fuerzas de unión que pueden ser entre 2, 3 o 4 átomos y las fuerzas entre pares

de átomos no unidos (van der Waals y electrostáticas), a las que se les agregan fuerzas restrictivas y/o externas

3. **Actualización de la configuración** - El movimiento de los átomos se simula resolviendo numéricamente las ecuaciones de movimiento de Newton:

$$\frac{d^2 r_i}{dt^2} = -\frac{F_i}{m_i}, i = 1, \dots, N \quad (2)$$

4. **Salida** - Se escriben las posiciones, velocidades, energías, temperaturas, presión, etc.

Los pasos 2, 3 y 4 listados previamente se repiten hasta completar el tiempo de simulación establecido.

Adicionalmente se calcula la energía cinética y la temperatura del sistema con las ecuaciones enumeradas a continuación. Donde T es la temperatura absoluta, k es la constante de Boltzmann y N_{df} es el número de grados de libertad.

$$E_{kin} = \frac{1}{2} \sum_{i=1}^N m_i v_i^2, i = 1, \dots, N \quad (3)$$

$$\frac{1}{2} N_{df} k T = E_{kin} \quad (4)$$

Se trabaja con un sistema a presión y temperatura constante. La implementación del acoplamiento de temperatura se calcula sobre la base de una modificación del algoritmo de Berendsen, este simula un baño térmico externo con una temperatura determinada mediante un acoplamiento débil con cinética de primer orden. Frente a una variación de temperatura se corrige según la ecuación:

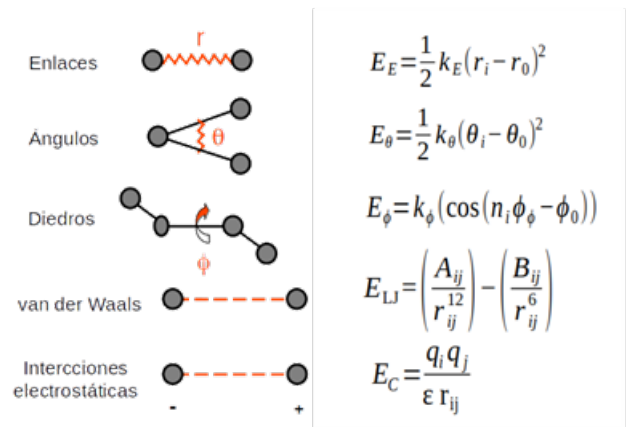
$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (5)$$

En el caso de la presión el algoritmo de Berendsen reescala las coordenadas y los vectores del caja en cada paso, o cada cierta cantidad de pasos determinados con una matriz permitiendo una relajación cinética de primer orden de la presión hacia una presión de referencia dada según la ecuación:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \quad (6)$$

Por último para calcular la energía potencial del sistema se calcula a partir de las interacciones de unión (hasta 4 átomos) y de no unión (más de 4 átomos). En la Figura 5 se muestra un esquema básico del cálculo de energía potencial. Los parámetros en las ecuaciones vienen dados por datos experimentales o de cálculos cuánticos [40].

$$E_p = E_U + E_{NU}$$

$$E_p = E_E + E_\theta + E_\phi + E_{LJ} + E_C$$


Enlaces		$E_E = \frac{1}{2} k_E (r_i - r_0)^2$
Ángulos		$E_\theta = \frac{1}{2} k_\theta (\theta_i - \theta_0)^2$
Diedros		$E_\phi = k_\phi (\cos(n_i \phi_i - \phi_0))$
van der Waals		$E_{LJ} = \left(\frac{A_{ij}}{r_{ij}^{12}} \right) - \left(\frac{B_{ij}}{r_{ij}^6} \right)$
Interacciones electrostáticas		$E_C = \frac{q_i q_j}{\epsilon r_{ij}}$

Figura 5: Cálculo de la Energía Potencial (E_p). La energía potencial se calcula como la sumatoria de las energías de unión (E_U , interacciones hasta 4 átomos) y las energías de no unión (E_{NU} , interacciones mayores a 4 átomos). Donde r_i es la distancia entre dos átomos unidos covalentemente, θ_i es el ángulo entre 3 átomos y ϕ_i es el ángulo diedro entre 4 átomos; r_0 , θ_0 y ϕ_0 son los valores teóricos de equilibrio; k_E , k_θ y k_ϕ son constantes de fuerza de las interacciones. La energía de las interacciones debidas a los enlaces y los ángulos se calculan mediante potenciales armónicos, mientras que las debidas a interacciones de torsión se describen por series de Fourier. La energía potencial debida a interacciones de van der Waals depende de la distancia (r_{ij}) entre dos átomos i y j de los parámetros A_{ij} y B_{ij} y se describe mediante potenciales de Lennard-Jones. La energía debida a las interacciones electrostáticas depende del dieléctrico (ϵ), de la distancia (r_{ij}) entre dos átomos i y j con las cargas q_i y q_j y se describe mediante la ley de Coulomb.

La estructura inicial se obtiene a partir de un archivo de coordenadas (para este trabajo tomamos archivos pdb de complejos resueltos por RMN). Cada complejo proteico se simula en una caja a la cual se le agregan moléculas de agua y una concentración fisiológica de iones. Una vez obtenido el complejo solvatado, se procede a minimizar la estructura mediante un algoritmo de minimización (Ejemplo *steepest descent*); para ello se permite sólo el movimiento de las cadenas laterales dejando que el sistema relaje. A partir de la ultima estructura, se realizan dos simulaciones cortas incorporando primero temperatura (ensamble NVT) y luego temperatura y presión (ensamble NPT), para ello se asigna a cada partícula del sistema velocidades al azar a partir de una distribución de Maxwell-Boltzmann. La última estructura es la configuración de inicio para la simulación de producción en un ensamble NPT. Una vez finalizado el tiempo de simulación se pueden analizar propiedades y comportamiento del sistema.

El análisis de los resultados de las simulaciones de DM es el punto de partida para la comprensión de la estabilidad de cada complejo proteico.

Para el análisis de las interacciones presentes en estos complejos, ya sea a partir de las estructuras de RMN o de las obtenidas por DM se utiliza la plataforma RING (del inglés *Residue Interaction Network Generator*) [12] que proporciona las distancias entre residuos en interacciones proteína-proteína mediante una estrategia basada en la teoría de grafos (Figura 6). Considerando que son las interacciones entre aminoácidos las que contribuyen a la estabilidad y la función de las proteínas, RING se convierte en una herramienta sumamente importante para la investigación en el campo de la biología molecular y la bioinformática. Genera redes probabilísticas y mapas de contacto dependientes de la conformación proteica permitiendo comparar y analizar interacciones proteína-proteína.

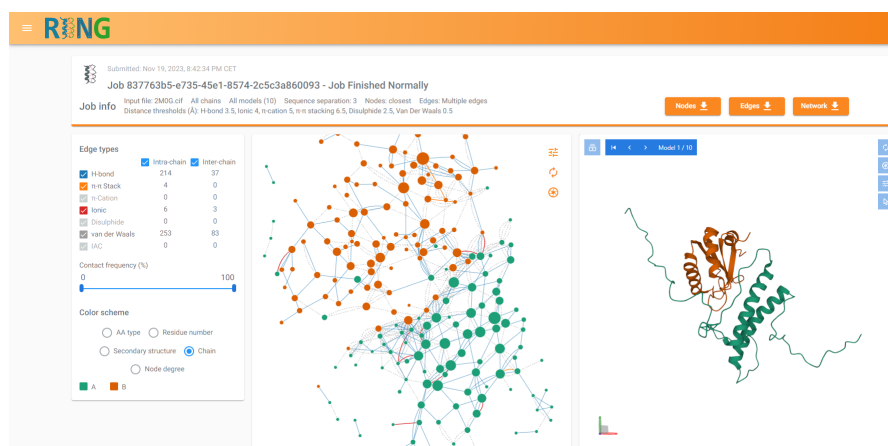


Figura 6: Capturas de pantalla de la salida de la herramienta RING a partir de la estructura de un complejo proteico.

Los complejos objeto de estudio en este proyecto, fueron seleccionados a partir de la base de datos FuzDB[13] (Figura 7). En ésta se recopilan complejos entre proteínas desordenadas determinados experimentalmente (RMN), cuyas interfaces se consideran “difusas” (*fuzzy*) ya que los contactos entre cadenas son dependientes del conformero particular. El RMN otorga un ensamble de complejos que están en equilibrio dinámico en solución, es decir, representa los posibles conformeros que este complejo tendrá en el medio celular.

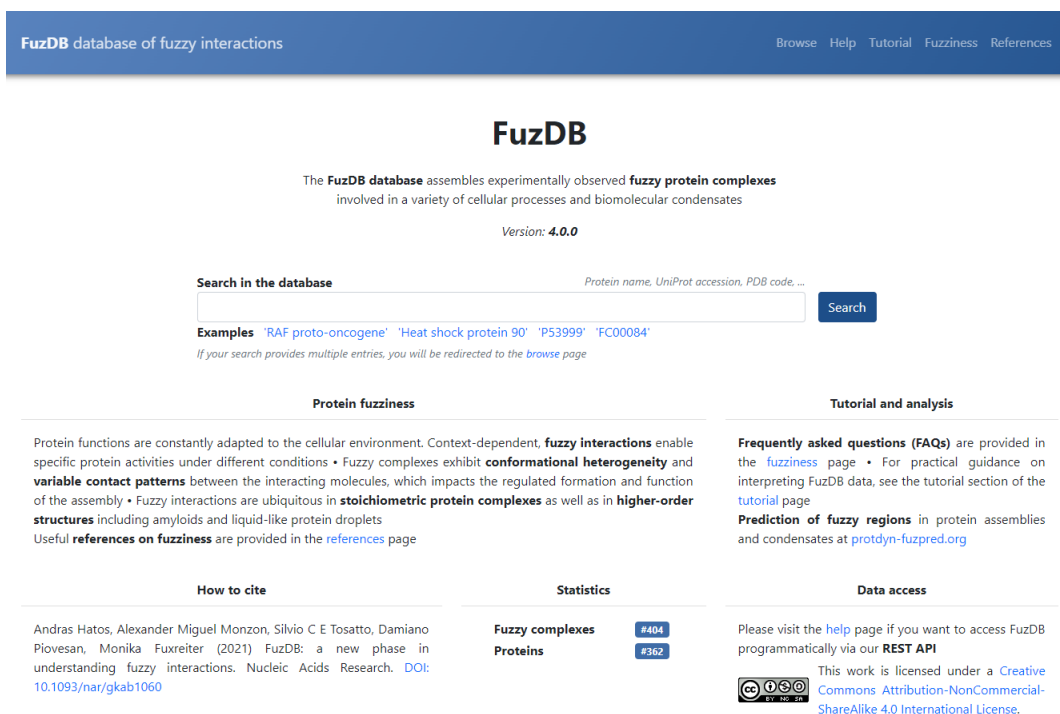


Figura 7: Captura de pantalla de la herramienta FuzDB.

La selección de los complejos estudiados, fue apoyándose en la plataforma MobiDB[14] que ofrece una herramienta de referencia para el estudio de la flexibilidad conformacional en proteínas y su relación con la función biológica. La base de datos MobiDB contiene información sobre regiones de proteínas que muestran diferentes niveles de flexibilidad conformacional. También proporciona información detallada sobre la

estructura y función de las proteínas, así como sobre sus interacciones con otras moléculas biológicas (Figura 8).

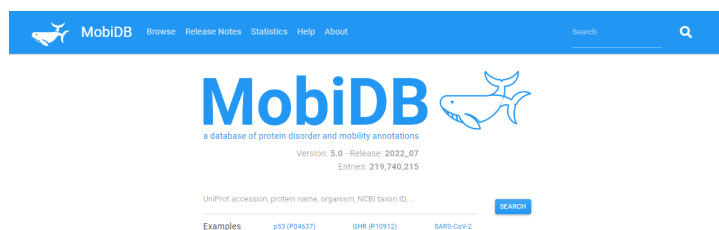


Figura 8: Captura de pantalla de la base de datos MobiDB.

5. Objetivos

Existe una gran cantidad de recursos computacionales para la interpretación y análisis de las conformaciones e interacciones de las IDPs. Encontramos bases de datos, programas de simulación, herramientas de visualización y modelado, así como servidores web con recursos que permiten aplicar algoritmos específicos. La combinación de estas herramientas nos permitió generar resultados que ayudan a una mayor comprensión sobre la estructura y dinámica de las IDPs.

La finalidad del presente trabajo es, partiendo del hecho de que las proteínas y sus complejos son sistemas en constante movimiento, es de suma importancia recorrer sus posibles conformaciones a lo largo del tiempo. De esta manera, tener mayor conocimiento de las superficies de interacción.

Mediante la comparación de los resultados obtenidos a partir de herramientas informáticas como la DM, con el método de resolución experimental RMN, se comprobará si la DM es capaz de explorar todo el espacio conformacional de cada proteína y su complejo, la estabilidad de sus interacciones a lo largo del tiempo, el descubrimiento de nuevos conformeros y contactos y la prevalencia de un conformero con respecto a otro. Esto se aplica principalmente a las IDPs por su naturaleza desordenada y pocas restricciones de movimiento.

De esta forma, planteamos la hipótesis que la dinámica molecular completa la información experimental obtenida por RMN y jerarquiza las interacciones más importantes y conformeros más frecuentes, dado que el RMN da igual importancia a todos los contactos y conformeros. Este conocimiento al que no alcanza el RMN permite identificar enlaces y conformaciones claves que podrían contribuir a detectar dianas terapéuticas para el tratamiento de enfermedades, desestabilizando o estabilizando conformaciones del complejo mediante la intervención en los enlaces de los aminoácidos clave.

Para el desarrollo del presente trabajo, se seleccionaron complejos proteicos constituidos por IDPs. Se comparan los resultados obtenidos a partir de la DM y otras herramientas informáticas, con los datos recopilados del análisis de complejos determinados por RMN y los reportados en la literatura. De esta manera, una vez obtenidos los resultados de la comparación, se podrá desarrollar una conclusión sobre el potencial del análisis de las DM sobre los complejos proteicos compuestos por IDPs.

6. Materiales y Métodos

6.1. Selección de proteínas:

Se seleccionaron 9 complejos de proteínas de la base de datos FuzzDB siguiendo los criterios de importancia de las proteínas y longitud de las cadenas para poder correr las DM con los recursos disponibles.

Además para seleccionar las estructuras se utilizó la base de datos PDB (<https://www.rcsb.org/>) y MobiDB (<https://mobidb.bio.unipd.it/>). Como se mencionó anteriormente, esta última contiene información sobre las zonas desordenadas de cada proteína, esto nos permitió escoger aquellos complejos donde haya una cadena con una amplia región desordenada. Como referencia se tomaron aquellas proteínas cuya porción a analizar se encuentre inmersa en una region con desorden ± 10 aminoácidos consecutivos. En la Figura 9 se muestra un ejemplo de captura de pantalla de la proteína p53 (que esta en nuestro estudio), donde se analizó el fragmento de aminoácidos 20 a 73 que está rodeado por una región de 10 aminoácidos desordenados a cada lado.

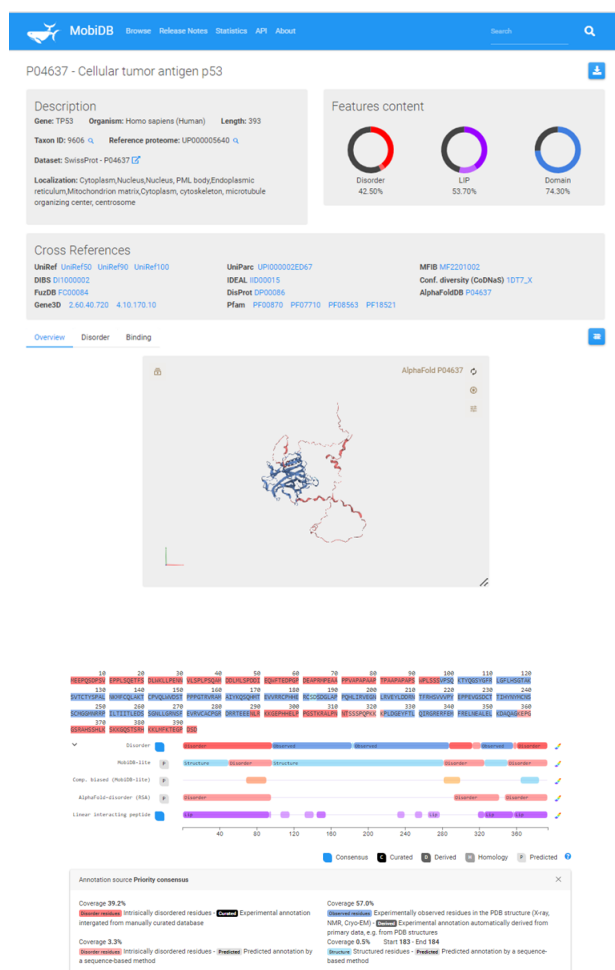


Figura 9: Captura de pantalla de la pagina web de MobiDB con los resultados correspondientes a la proteoma p53, el fragmento a analizar es del aminoácido 20 al 73. Se muestran en azul y rojo los segmentos ordenados y desordenados respectivamente.

Las interacciones entre residuos de cada cadena se obtuvieron del servidor RING. El set de complejos utilizado fue el siguiente:

2GS0 (Figura 10): Dímero conformado por la subunidad TFB1 del Factor de transcripción H de la polimerasa II (TFIIH) (UniProt ID: P32776) y el dominio de transactivación (TAD) de la p53 (UniProt ID: P04637). Método de obtención por NMR, 20 conformeros.[16]

La región TAD de la proteína p53 interactúa con el dominio de pleckstrina de la subunidad p62, homóloga a la TFB1. La proteína p53 actúa como supresor de tumores; induce la detención del crecimiento o la apoptosis según las circunstancias fisiológicas y el tipo de célula. Además participa en la regulación del ciclo celular como un transactivador que regula negativamente la división celular al controlar un conjunto de genes necesarios para este proceso. La interacción entre la subunidad TFB1 y el dominio p53 está directamente relacionada con la capacidad del dominio para activar el inicio de la transcripción y la elongación[16][17][18]. Se observa en la Figura 10 los dominios a analizar (a) y las proteínas enteras (b-c).

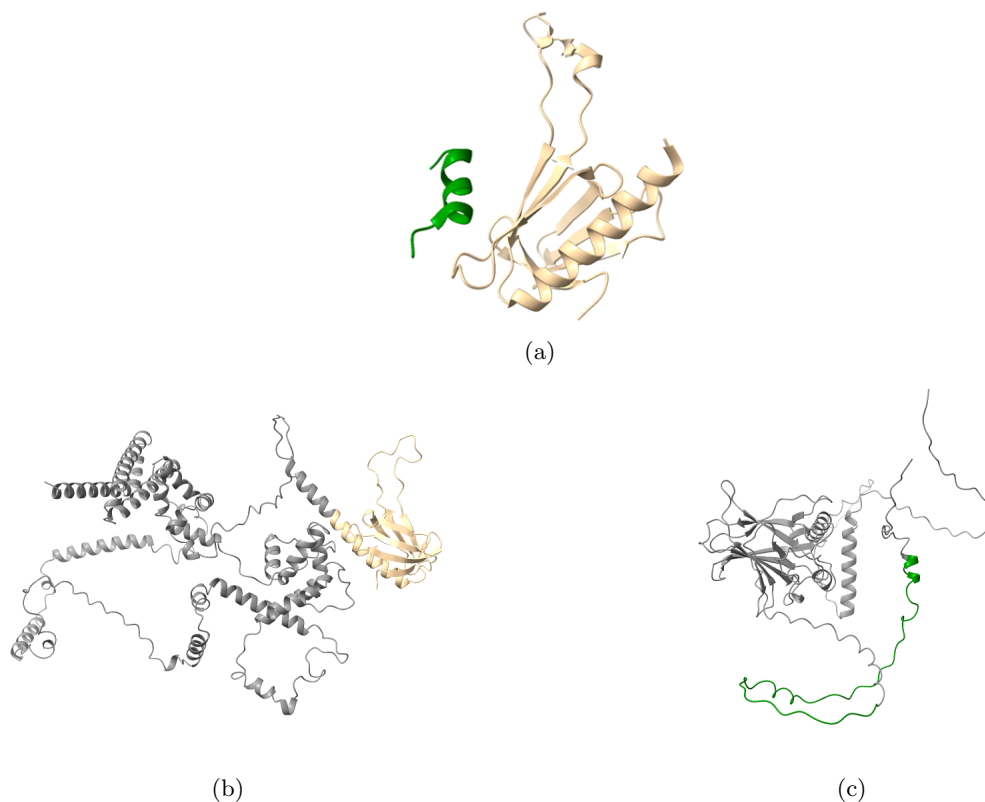


Figura 10: (a) Amarillo la cadena A (TFB1), verde cadena B (TAD). (b-c) Modelos de las proteínas completas predichas con AlphaFold2 (AF)[50], coloreadas según el segmento representado en el complejo y con el mismo código de color.

2M0G (Figura 11): dímero conformado por dominio hélice-horquilla en la región N terminal del splicing factor 1 (SF1) y el motivo homólogo de U2AF del C terminal de la proteína U2AF65. Método de obtención por NMR, 10 conformeros[19].

Estas subunidades juegan un papel fundamental en el splicing del pre-ARNm. El dominio hélice-horquilla de SF1 desempeña un papel esencial en el reconocimiento cooperativo del sitio de empalme 3' al ayudar a mantener una disposición especial de las proteínas SF1 y U2AF65 junto con el ARN[19][20][21]. Se observa en la Figura 11 los dominios a analizar (a) y las proteínas enteras (b-c).

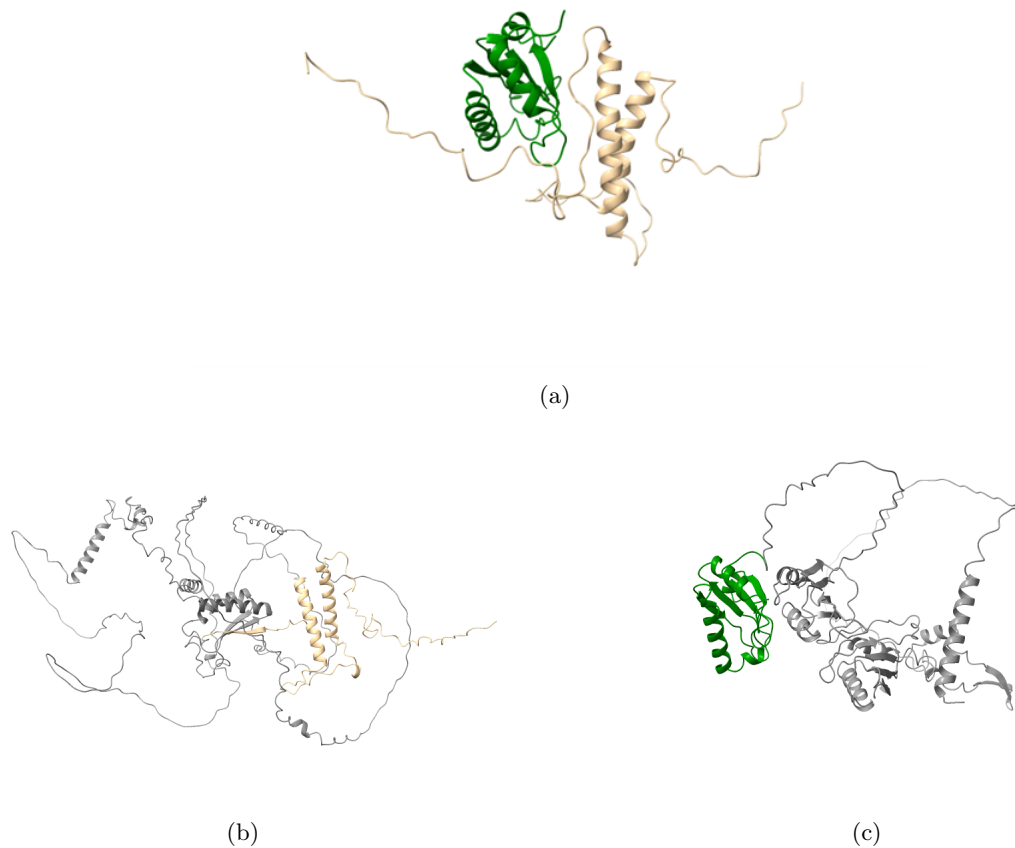


Figura 11: (a) En amarillo la cadena A (Splicing Factor 1) y en verde se observa la cadena B (motivo homólogo U2AF). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

2L14 (Figura 12): dímero conformado por el dominio de transactivación de p53 y el receptor nuclear coactivador del binding domain de la proteína CREB. Método de obtención por NMR, 20 conformeros[22].

Como se mencionó anteriormente, el supresor de tumores p53 actúa en una gran cadena de transducción de señales que median la respuesta celular al estrés, llevando a la detención del ciclo celular, la senescencia o la apoptosis. Debido a su papel en la determinación del destino celular, p53 está estrechamente controlado por numerosas proteínas reguladoras[17][?][23]. Se observa en la Figura 12 los dominios a analizar (a) y las proteínas enteras (b-c).

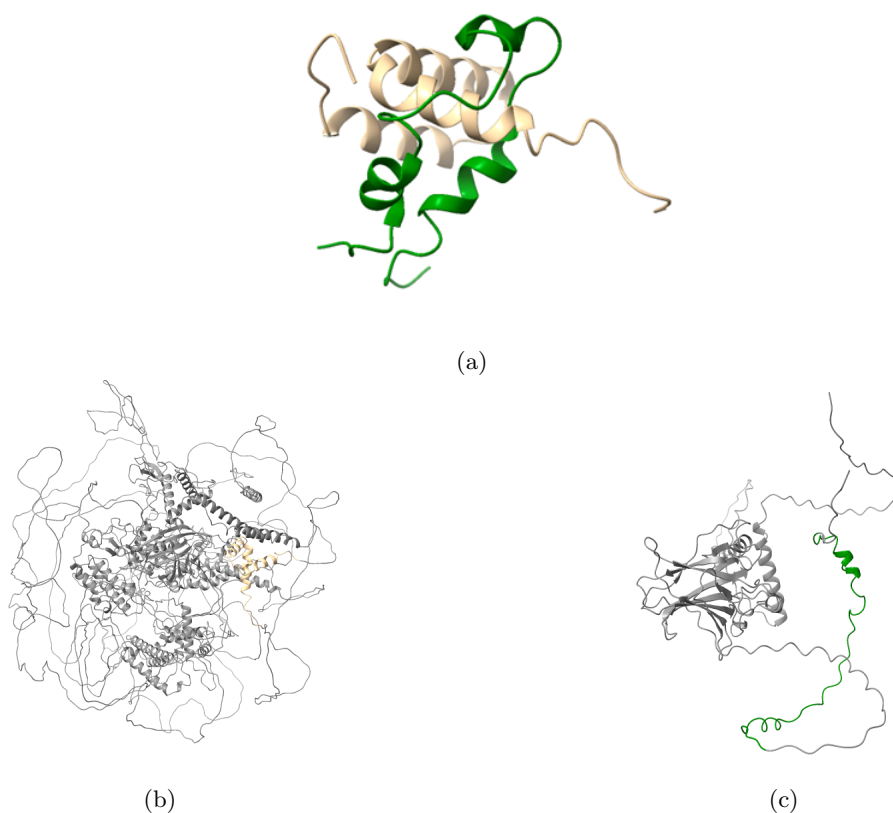


Figura 12: (a) En amarillo la cadena A (binding domain de la proteína CREB) y en verde se observa la cadena B (dominio de transactivación p53). (b-c) Modelos de las proteínas completas coloreadas según el segmento representado en el complejo y con el mismo código de color.

2LPB (Figura 13): dímero conformado por el activador de transcripción ácido Gcn4 y el dominio de activación de la subunidad mediadora Gal11/Med15. Método de obtención por NMR, 13 conformeros[24].

Ambas subunidades se encuentran relacionadas en la transcripción de genes dependientes del ARN polimerasa II. El dominio de unión al activador Gal11 presenta una estructura de cuatro hélices con una pequeña cavidad hidrofóbica poco profunda en su núcleo. Cuando se encuentran el complejo unido, ocho residuos de Gcn4 adquieren una conformación helicoidal, lo que permite que tres residuos aromáticos se introduzcan en la cavidad de Gal11[24][25][26]. Se observa en la Figura 13 los dominios a analizar (a) y las proteínas enteras (b-c).

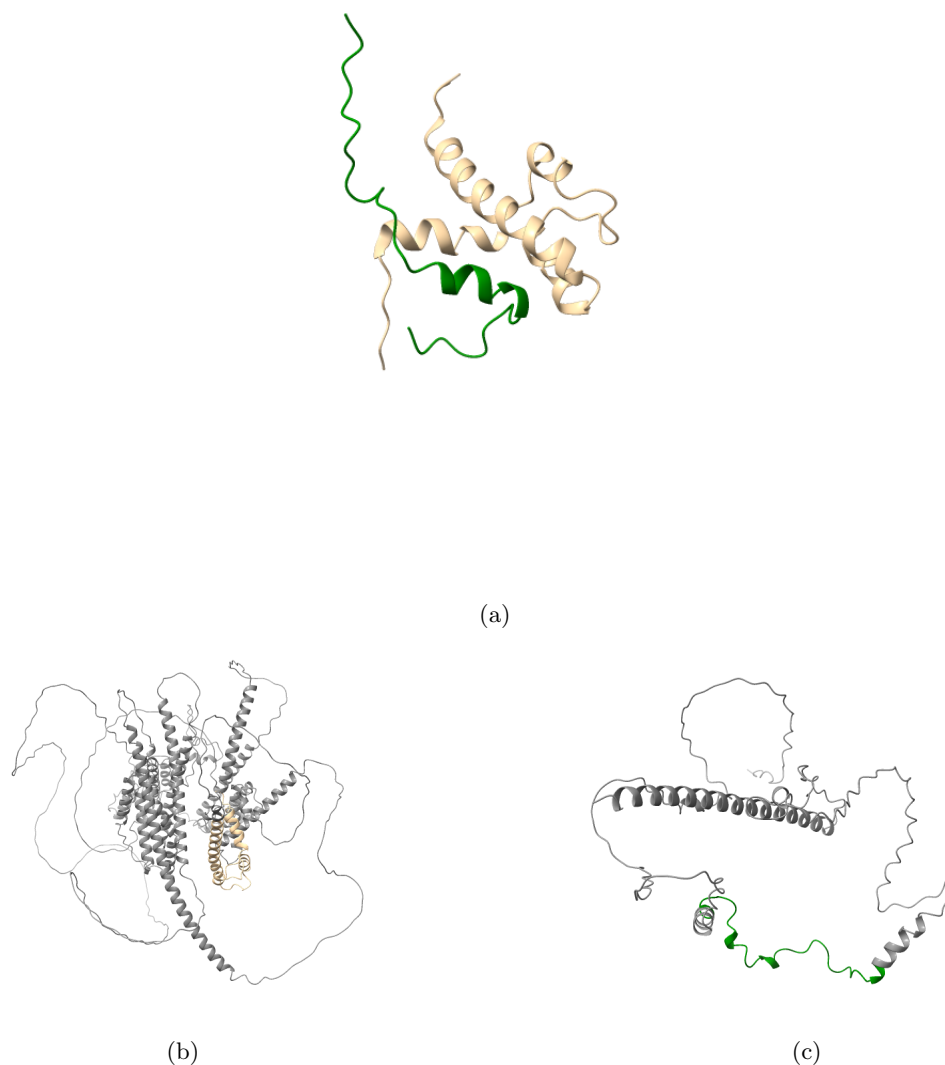


Figura 13: En amarillo la cadena A (dominio de activación de la subunidad mediadora Gal11/Med15) y en verde se observa la cadena B (activador de transcripción ácido Gcn4). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

5URN (Figura 14): Dímero conformado por el dominio 1 del factor de transcripción p65 y la subunidad TFB1 del factor IIIH de transcripción (homólogo al motivo pleckstrin p62 de humanos) y reparación general del ADN. Método de obtención por NMR, 20 conformeros[27].

La interacción entre las subunidades está directamente relacionada con una serie de eventos de transducción de señales iniciados por una amplia gama de estímulos desencadenados por procesos biológicos como la inflamación, la inmunidad, la diferenciación, el crecimiento celular, la tumorigénesis y la apoptosis. Tres residuos hidrofóbicos dentro del dominio 1 de p65 desempeñan roles significativos en la formación de la interfaz de unión con la subunidad TFB1 del factor IIIH, y estos residuos son esenciales para la capacidad de p65 de activar la transcripción[18][27][28]. Se observa en la Figura 14 los dominios a analizar (a) y las proteínas enteras (b-c).

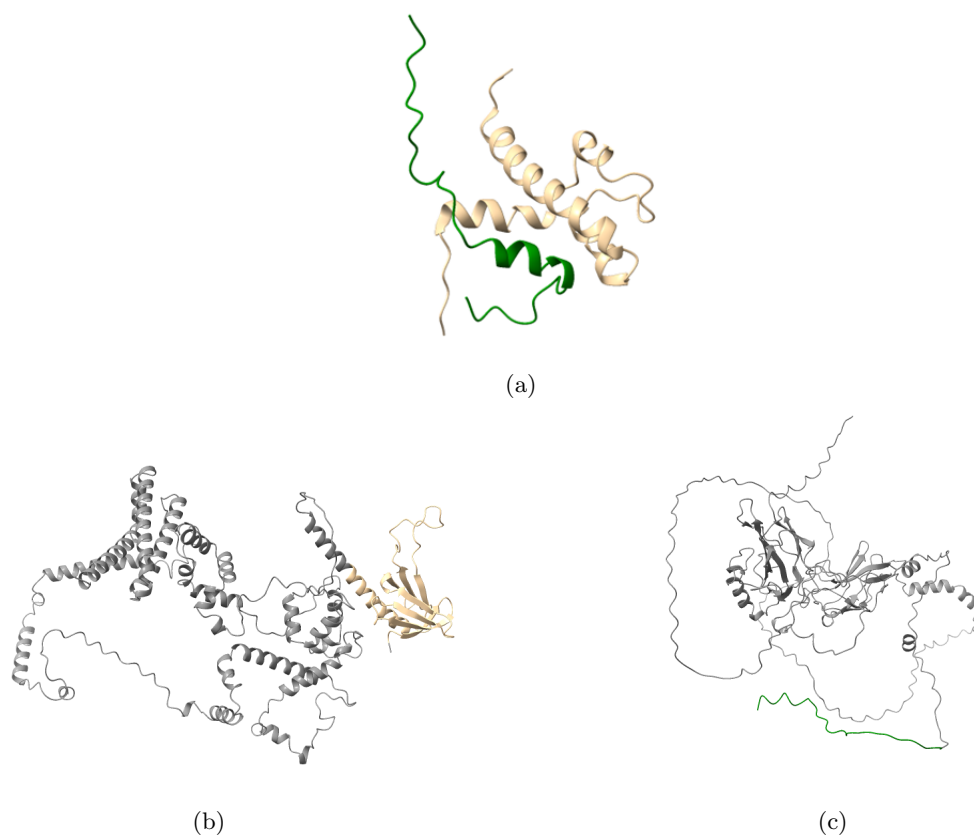


Figura 14: En amarillo la cadena A (subunidad TFB1 del factor IIH) y en verde se observa la cadena B (factor de transcripción p65). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

2LY4 (Figura 15): Dímero conformado por el dominio de transactivación p53 y HMGB1 (High Mobility group box 1). Método de obtención por NMR, 10 conformeros[29].

El grupo HMGB1 facilita la adherencia del ADN a factores de transcripción incluido el factor de supresión de tumor p53 al actuar como una chaperona[17][29][30]. Se observa en la Figura 15 los dominios a analizar (a) y las proteínas enteras (b-c).

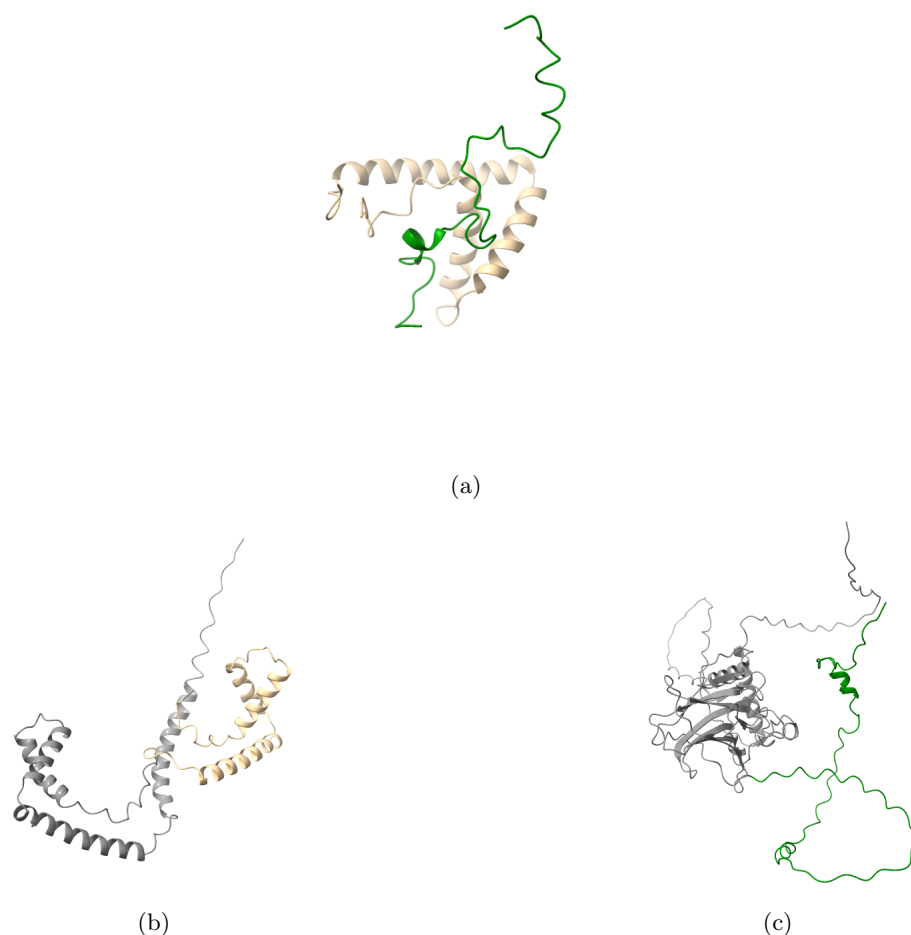


Figura 15: En amarillo la cadena A (HMGB1) y en verde se observa la cadena B (dominio de transactivación p53). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

1MV0 (Figura 16): Dímero conformado por el N terminal de la proteína c-Myc (MB1) y la proteína Bin1. Método de obtención por NMR, 20 conformeros.

c-Myc es una fosfoproteína nuclear que funciona como regulador de la transcripción de genes, que desencadenan el ciclo celular, la apoptosis y el bloqueo de la diferenciación. El extremo N de la oncoproteína c-Myc interactúa con Bin1, una proteína núcleo citoplasmática expresada de forma ubicua con características de supresor de tumores. La interacción c-Myc/Bin1 tiene un significado regulador potencial, ya que la transformación y la apoptosis mediadas por c-Myc pueden modularse mediante la expresión de Bin. La interacción c-Myc/Bin1 depende de la secuencia conservada de Myc Box 1 (MB1) de c-Myc[31][32][33]. Se observa en la Figura 16 los dominios a analizar (a) y las proteínas enteras (b-c).

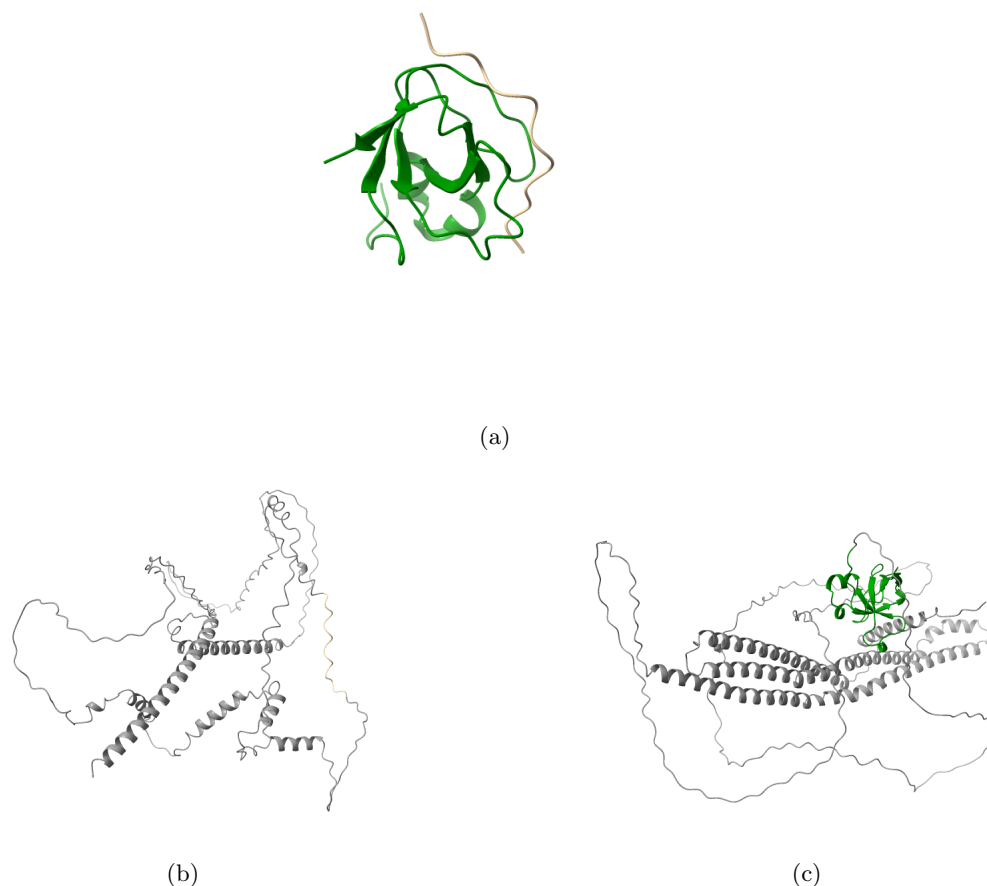


Figura 16: En amarillo la cadena A (proteína c-Myc) y en verde se observa la cadena B (proteína Bin1). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

2PX9 (Figura 17): Dímero conformado por las enzimas SUMO E1 y E2. Método de obtención por NMR, 20 conformeros[34].

La proteína SUMO (del inglés small ubiquitin like modifier) es un tipo de ubiquitina modificadora, realiza funciones esenciales en la mayoría de los organismos. La SUMOilación está involucrada en una gran variedad de procesos celulares fundamentales, incluida la replicación del ADN, la transcripción, la regulación del ciclo celular, la reparación del daño del ADN, la organización de la cromatina, la biogénesis de los ribosomas, el empalme del pre-ARNm, el tráfico nuclear, la transducción de señales y la degradación de proteínas. Un paso esencial en la vía de SUMOilación corresponde a la interacción entre la enzima activadora E1 y la enzima conjugadora E2 permitiendo que la proteína SUMO se logre conjugar con su proteína target[34][35][36]. Se observa en la Figura 17 los dominios a analizar (a) y las proteínas enteras (b-c).

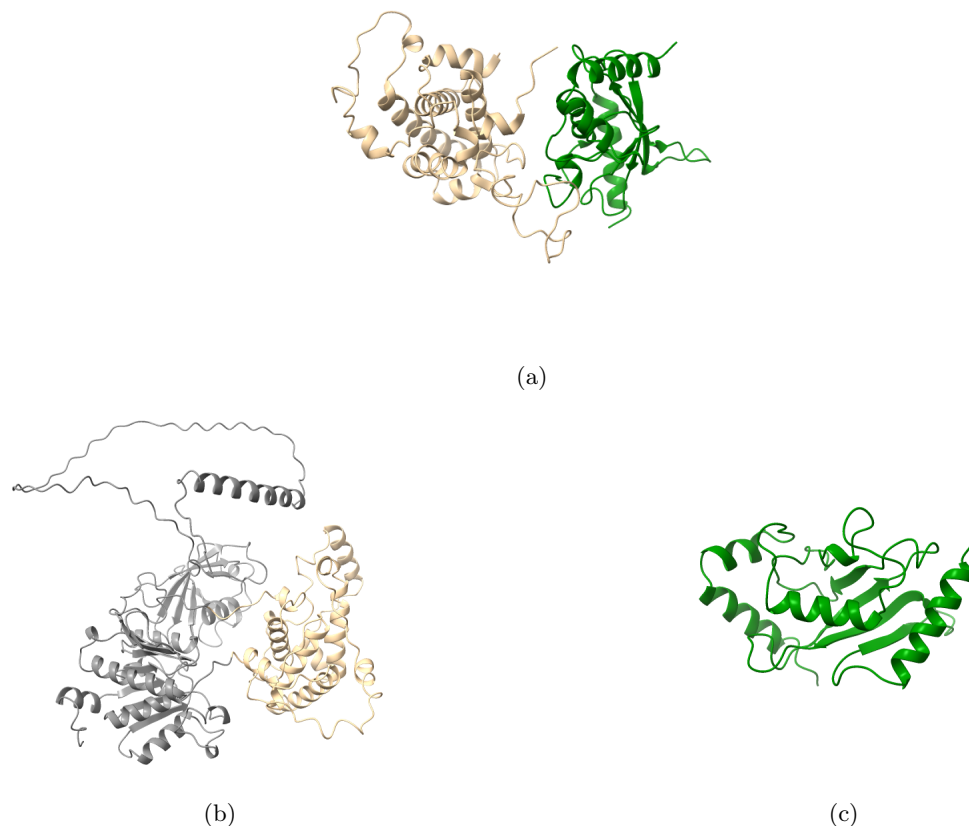


Figura 17: En amarillo la cadena A (Enzima E2) y en verde se observa la cadena B (Enzima E1). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

2ASQ (Figura 18): dímero conformado por SUMO-1 y SUMO binding motif (SBM) con la siguiente secuencia KVDVIDLTIESSSDEEDPPAKR de la proteína PIASX. Método de obtención por NMR, 10 conformeros[37].

Dentro de las proteínas SUMO encontramos 4 parálogos, distinguiendo a la -1 como aquella con mayor diferencia en estructura. Se ha identificado una superficie específica de SUMO como la única región que media la inhibición transcripcional dependiente de SUMO-1. Esta superficie es idéntica a la región que interacciona con la SBM.(22). Debido a que la inhibición transcripcional es una función común de la sumoilación, se sugiere que la interacción SUMO-SBM es responsable del reclutamiento de factores celulares que median las funciones inhibitoras transcripcionales de SUMO[37][38][39]. Se observa en la Figura 18 los dominios a analizar (a) y las proteínas enteras (b-c).

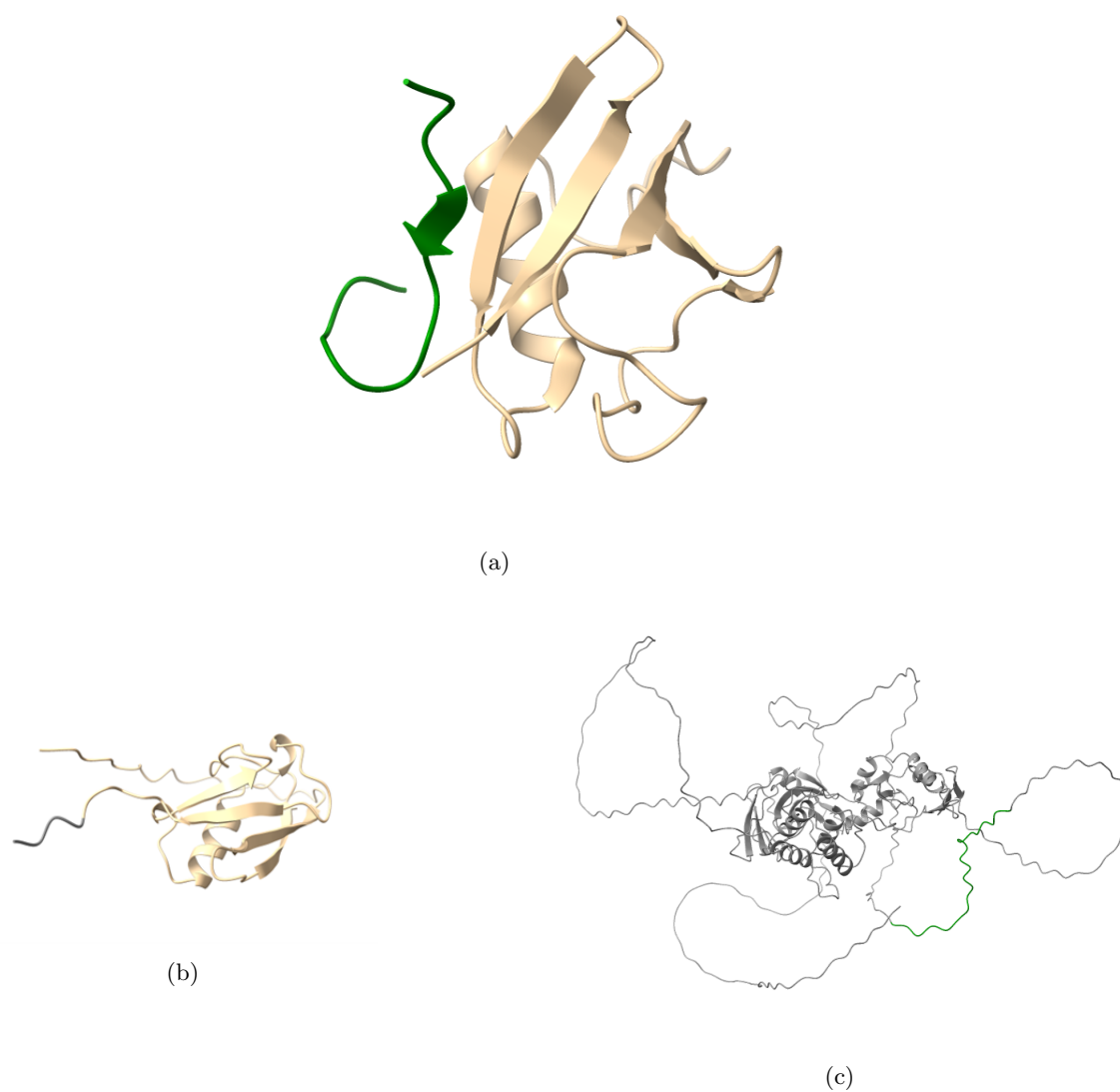


Figura 18: En amarillo la cadena A (SUMO-1) y en verde se observa la cadena B (SBM). (b-c) Modelos de las proteínas completas (AF) coloreadas según el segmento representado en el complejo y con el mismo código de color.

A continuación en el Cuadro 1, se muestra un resumen de los complejos mencionados previamente.

PDB	N°conf.	Representación	Cadenas y rango	Longitud PDB	Uniprot
2GS0	20		TFB1 levadura A 1-115	115	P3D776
			P53 humana B 45-58	13	P04637
2M0G	10		Splicing Factor 1 humana A 1-145	145	Q15637
			U2AF2 humana B 372-475	104	P26368
2L14	20		CBP raton A 2059-2117	59	P45481
			P53 humana B 13-61	49	P04637
2LPB	13		MED15 levadura A 158-238	81	P19659
			GCN4 levadura B 101-134	34	P03069
5URN	20		TFB1 levadura A 2-115	115	P32776
			TF65 humana B 521-551	33	Q04206
2LY4	10		HMGB1 humana A 2-84	83	P09429
			P53 humana B 14-60	46	P04637
1MV0	20		MYC humana A 55-68	14	P01106
			BIN1 humana B 513-593	81	O00499
2PX9	20		CAE2 humana A 166-382	217	Q9UBT2
			UBT9 humana B 1-158	158	P63279
2ASQ	10		SUMO1 humana A 1-97	97	P63165
			PIASx humana B 466-488	25	O75928

Cuadro 1: Tabla a modo resumen de los complejos estudiados, con la longitud de cada cadena y su código Uniprot [15]. Beige cadenas A, verde, cadenas B.

Una vez seleccionado el set de proteínas se utilizó la herramienta ChimeraX (visualizador de moléculas) [51] para hallar el conformero más lejano, y de esta forma poder realizar dos dinámicas por cada complejo partiendo de dos condiciones espaciales iniciales distintas.

- **DM 1** - utilizando como punto de partida el conformero 1
- **DM 2** - utilizando como punto de partida el conformero más lejano

Para medir la similitud entre las estructuras de los conformeros, estos deben superponerse en el espacio y su diferencia se mide a través de la desviación media cuadrática (RMSD por Root Mean Square Deviation en inglés). Se mide la distancia promedio entre los átomos de una estructura y los átomos correspondientes en la otra estructura después de una superposición óptima. El valor resultante del RMSD se expresa en Å (angstroms) y cuanto menor sea el valor del RMSD, mayor será la similitud estructural entre las proteínas comparadas. Valores altos de RMSD denotan proteínas con grandes diferencias entre sus conformaciones.

Este confórmero lejano se identifica utilizando el comando “MatchMaker” de Chimera donde como primer paso se realiza un alineamiento de secuencias en pares y luego realiza un ajuste de la alineación comparando un átomo por cada residuo. Luego, se calcula el RMSD de cada confórmero contra el primero. A partir de esta diferencia se seleccionó el confórmero con valor más alto de RMSD para la segunda dinámica.

6.2. Simulaciones por Dinámica Molecular:

Para realizar las simulaciones por DM de 1000 ns, se ha utilizado la herramienta de simulación GROMACS [41]. A continuación se encuentra un listado de los pasos a seguir para a partir de una estructura estática lograr simular la estructura en movimiento.

1. Obtener el archivo de coordenadas de la proteína o complejo proteico con el que se va a trabajar (formato PDB)
2. Trabajar a partir de la unidad biológica.
3. Verificar si hay residuos perdidos, evaluar su importancia para reconstruirlos
4. Identificar la presencia y la importancia de iones y ligandos. Decidir si se van a incluir los ligandos.
5. Agregar hidrógenos si no estuvieran en la estructura original.
6. La estructura proteica se modela dentro de una caja. A ésta se le agregan moléculas de agua explícita y iones. La cantidad de iones es aquella que permite neutralizar el sistema y alcanzar una concentración fisiológica para realizar el experimento. La caja de simulación debe ser un poliedro que llene el espacio para permitir condiciones periódicas de contorno. Los distintos tipos de cajas son el cubo, el dodecaedro rómbico y el octaedro truncado. En este trabajo se utilizó un dodecaedro rómbico porque requiere menos moléculas de agua que una caja cúbica.
7. Realizar la minimización de energía. Durante este proceso se relaja la estructura proveniente del pdb. Dentro de los métodos de minimización encontramos a “steepest descent” y “conjugate gradient” como los más utilizados. En ambos métodos se modifican las coordenadas de los átomos en función de la dirección del gradiente negativo. La diferencia entre ellos es que en el primero no se tiene en cuenta los pasos anteriores, esto hace que cerca de los mínimos locales su convergencia sea lenta. El segundo converge de forma más rápida al tener en cuenta la información de los pasos anteriores.
8. Equilibrar la proteína: se agrega la temperatura y la presión en dos pasos consecutivos.
9. Se realiza la simulación de producción.

Para este trabajo se utilizó Gromacs[41], ya que es uno de los programas de simulación computacional más utilizados. Permite el uso de campos de fuerza como GROMOS, OPLS, AMBER, CHARMM, etc. El programa GROMACS es altamente adaptable y flexible, ya que permite al usuario agregar campos de fuerza y especificar funciones. GROMACS realiza dinámicas de equilibrio, de no equilibrio, y permite calcular la energía libre en ciertos sistemas. Se encuentra disponible al dominio público y se distribuye junto con su código fuente y documentación bajo la Licencia Pública General de GNU. Un grupo de desarrolladores de las Universidades de Groningen, Uppsala y Estocolmo, así como el Instituto Max Planck para la Investigación de Polímeros en Maguncia se encargan de su mantenimiento.

Se utiliza el campo de fuerzas de todos los átomos CHARMM36m (v. jul 2021)[42] con el modelo de aguas el agua TIP3P[43] acorde al campo de fuerzas elegido. Se utiliza una caja del tipo dodecaedro rómbico con una distancia entre el soluto y la caja de al menos 2 nm, según la movilidad del complejo. Se utiliza una concentración fisiológica de KCl de 0.15 M. Para el proceso de minimización se utiliza el algoritmo steepest descent. Se realiza una primera equilibración mediante un ensamble NVT o canónico de 100 ps, luego una segunda equilibración con un ensamble NPT de 200 ps. Para la trayectoria de producción se utiliza un ensamble NPT durante 1 microsegundo. Los siguientes parámetros se utilizan para las dos equilibraciones y la producción (excepto los de presión que solo se utilizan para la segunda equilibración y la producción):

algoritmo de integración leap frog [44], paso de tiempo 2 fs, acoplamiento de temperatura de Berendsen modificado (velocity rescaling)[45], temperatura de referencia de 300 K y un acoplamiento de presión de relajación exponencial (C-rescale)[46] con una presión de referencia de 1 bar con una compresibilidad de 4.5×10^{-5} . Se utilizó un radio de corte de 1.2 nm para interacciones coulómicas y de Lennard Jones. Se aplica el método de Particle Mesh Ewald (PME)[47] para las interacciones electrostáticas de largo alcance. Los primeros 50 ns de la trayectoria de producción se consideran equilibración y no se incorporan en el análisis. Se utiliza el paquete GROMACS v. 2021.5[48]

6.3. Análisis:

Una vez realizadas las simulaciones se llevaron a cabo dos tipos de análisis; por cadena y por complejo entero, para cada dinámica.

Utilizando el programa GROMACS se obtuvo **por cadena** la siguiente información:

- RMSD del backbone: el comando “gm x rms” compara cada estructura de la trayectoria con la inicial al calcular la desviación cuadrática media (Root-Mean-Square Deviation)
- RMSF: el comando “gm x rmsf” calcula la fluctuación cuadrática media al cuantificar las desviaciones promedio de las posiciones del átomo central de cada residuo de la proteína en comparación con una estructura de referencia
- Gyrate: el comando “gm x gyrate” calcula el radio de giro de una molécula. Es la distancia cuadrática media de cada átomo de la proteína a su centroide
- End to End: el comando “gm x pairdist” entre el primer y último residuo de la cadena permite cuantificar la separación espacial entre las terminales de las cadenas permitiendo evaluar la estabilidad estructural o cambios conformacionales a lo largo de la dinámica

Utilizando el programa GROMACS, Python y RING se obtuvo **por complejo** la siguiente información:

- Matriz RMSD: el comando “gm x rms” obtiene una matriz de comparación entre una trayectoria consigo misma, estas matrices se generaron desde el nanosegundo 50, instante de tiempo donde se puede considerar que los sistemas se encuentran equilibrados
- Cluster: el comando “gm x cluster” permite agrupar estructuras partiendo de las matrices de RMSD. El agrupamiento se realiza mediante el método de gromos, este tiene en cuenta un valor de punto de corte (cut off) para dejar de agrupar conformaciones. El valor de cut off se eligió según la población de cada cluster y buscando que solo se forme un cluster del tamaño 1. El resultado es un archivo de PDB con un representante de cada cluster que se lo denomina modelo. Además el comando permite obtener información sobre los conformeros en el tiempo realizando gráficas con la población de clusters
- Matriz de Contacto: mediante el uso del algoritmo de RING para procesar un archivo de PDB se puede obtener la lista de contactos intra e inter cadena del complejo por modelo (archivo .txt, Anexo X). El algoritmo identifica cada interacción basándose en distancias de corte (distance thresholds). Se utilizó el corte relajado, y manualmente se ajustó el distance threshold para las interacciones Van der Waals a 1.46 Å, los demás valores para cada interacción se encuentran listados en el Anexo 2. Una vez obtenidas las listas mediante el uso de un script de python se procesan los datos para obtener la frecuencia de aparición de cada contacto por modelo, estas mismas son las que conforman la matriz de contacto inter cadena
- Interacciones Core: se definen como aquellas interacciones que se mantienen a lo largo de la trayectoria por modelo (frecuencia de aparición es mayor al 75 %) y coinciden con los contactos vistos por RMN
- Contactos para clusters mas poblados: se seleccionan los clusters mas poblados de cada dinámica (representando al menos el 75 % del tiempo de simulación) y se buscan todos los contactos presentes en estos clusters

- Heatmap de Comparación: se busca comparar punto a punto las matrices de contacto entre el RMN y cada dinámica creando dos mapas de calor respectivamente para cada dinámica (Figura 19). Se calcula 1 menos la diferencia absoluta entre las frecuencias de las matrices a comparar, tomando como resultado 1 cuando ambos puntos tienen el mismo valor y 0 cuando tienen valor nulo. De esta forma, también se obtiene un índice de similitud como la suma de los valores obtenidos de la comparación, dividido por la dimensión de la matriz.

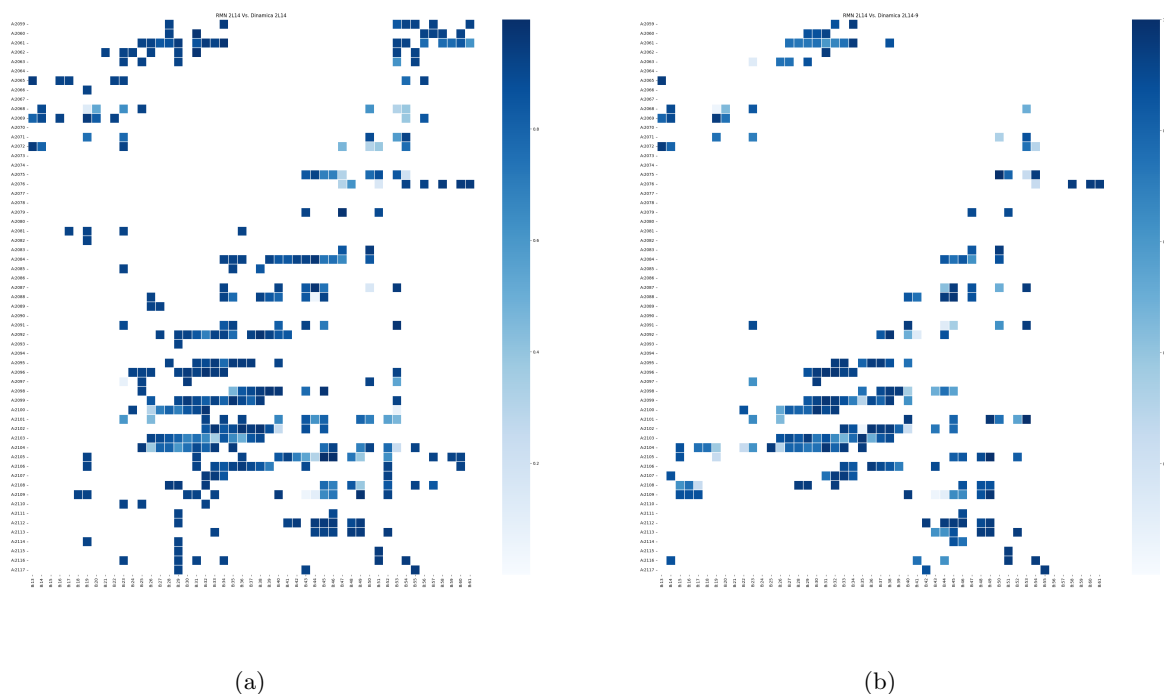


Figura 19: Los píxeles más oscuros representan una mayor similitud de frecuencia en esos puntos entre las matrices, mientras que los más claros representan mayor diferencia de frecuencia. (a) Heatmap de comparación entre la matriz de contactos del RMN y la matriz de contactos de la DM 1 2L14. (b) Heatmap de comparación entre la matriz de contactos del RMN y la matriz de contactos de la DM 2 2L14-9.

- Matriz de comparación: se busca mediante el uso de un código de color (Figura 20) ilustrar qué interacciones se mantienen coincidentes al comparar las matrices de contacto de cada dinámica por separado con la de RMN. Los píxeles verdes representan que ambas matrices tuvieron un contacto entre los respectivos residuos, los píxeles amarillos representan un contacto para una de las dos matrices, los píxeles azules representan contactos para la otra, y los blancos representan ningún contacto de ambas matrices

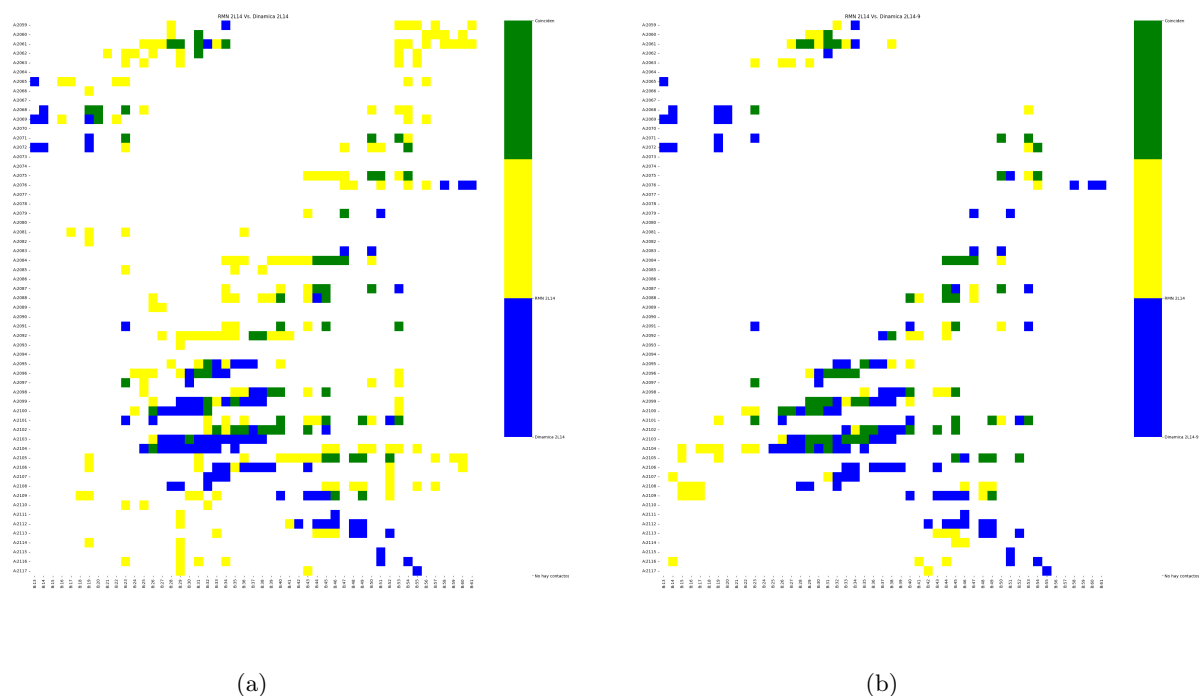


Figura 20: Los píxeles verdes representan contacto en ambos métodos, los amarillos en el método de RMN, los azules en el método de dinámica, y los blancos ningún contacto para ambos métodos. (a) Matriz de comparación entre la matriz de contactos del RMN y la matriz de contactos de la DM 1 2L14. (b) Matriz de comparación entre la matriz de contactos del RMN y la matriz de contactos de la DM 2 2L14-9.

- Distancia entre interacciones: utilizando el comando “gmx pairdist” se mide la distancia mínima sobre cada cadena lateral sin hidrógenos. Obtener la distancia mínima a lo largo de la trayectoria permite calcular el porcentaje total en el que la distancia permanece por debajo a 0,7nm o 7Å (threshold máximo utilizado por RING)

6.4. Gráficos y Análisis de datos: Python, AlphaFold y Chimera

Los gráficos y el manejo de datos presentados en este trabajo fueron realizados mediante Python[49] (librerías pandas, numpy, matplotlib.pyplot y seaborn) desde el sistema Google Collaborate, AlphaFold[50], para la obtención de las proteínas completas presentes en cada complejo, y ChimeraX-1.4[51] para iOS.

7. Resultados y discusión

7.1. Análisis de interacciones en los complejos, comparación entre DM y RMN.

7.1.1. 2GS0

A continuación se muestran los resultados de las DM realizadas partiendo del conformero 1 (2GS0) y el conformero más lejano 15 (2GS0-15):

En primer lugar hemos comparado las matrices de contacto del RMN, de la DM 1 y de la DM 2 (Figura 21 a, b y c respectivamente) y la matriz de comparación para cada dinámica con el RMN (Figura 22, a y b). Cada píxel es un contacto y la frecuencia del mismo se ve en la intensidad de color. Un contacto puede tener intensidad baja en la matriz si se da en un número reducido de complejos. La comparación de las matrices RMN vs DM1 y RMN vs DM2, muestra que son similares (score 0.79 y 0.71 respectivamente). Esto demuestra que en este caso, la DM reproduce los conformeros que aparecen en el RMN (por eso el score de similitud es alto) y naturalmente, aparecen nuevos, lo que se puede visualizar en el mapa de calor de la figura 22.

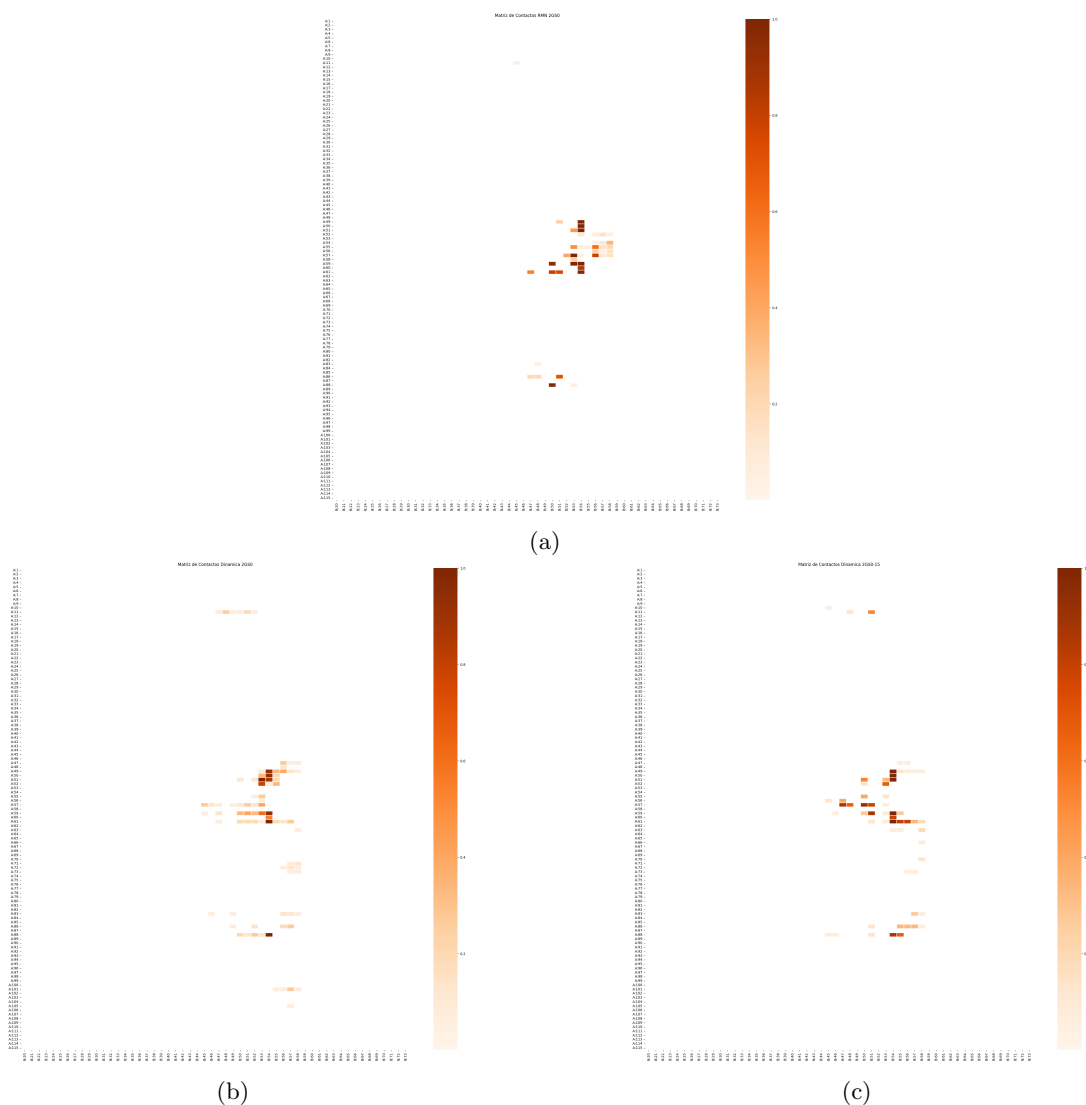


Figura 21: Matriz de contacto del: a) RMN, b) DM1, c) DM2

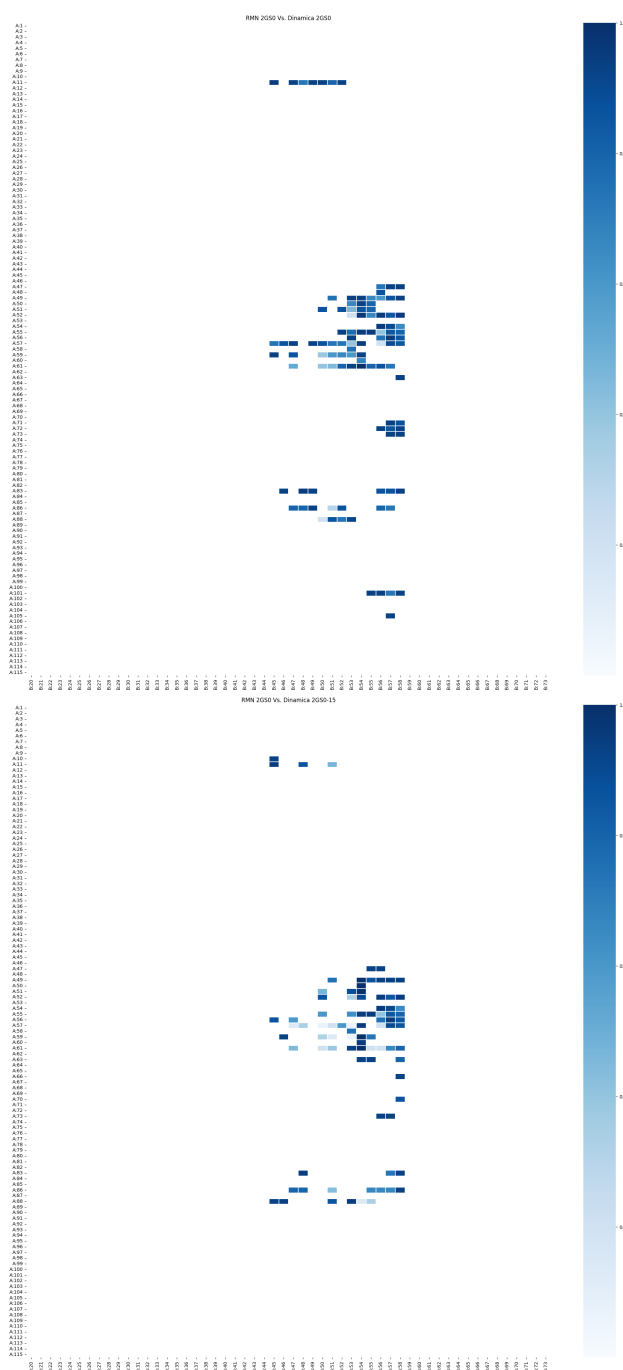
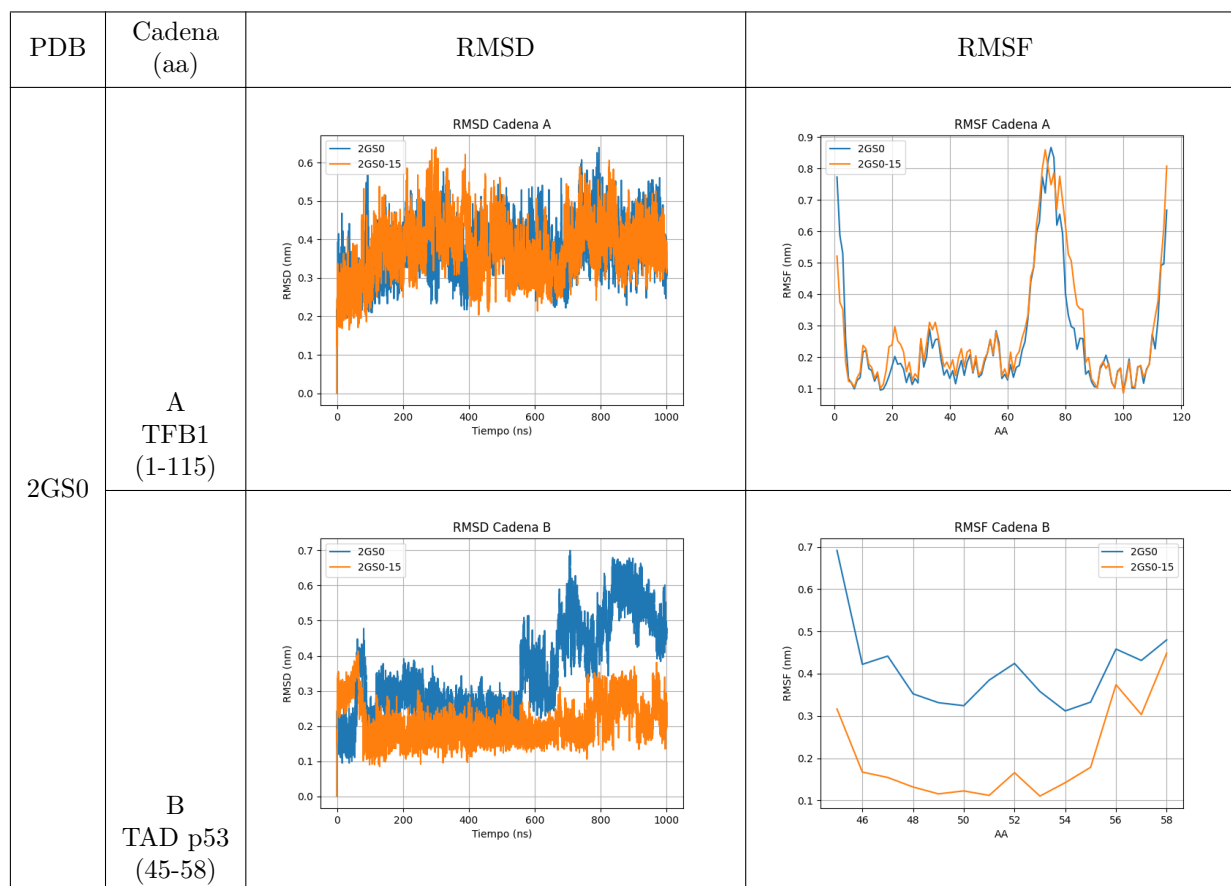


Figura 22: Heatmap de comparación. Panel superior: RMN - DM1, panel inferior: RMN - DM2. En azul oscuro se ven las posiciones donde ambas matrices son similares (no es la frecuencia del contacto, sino que el contacto en ambas matrices tiene la misma probabilidad)

En segundo lugar se analizaron las dinámicas moleculares. Se observa en el Cuadro 2 tanto el RMSD como el RMSF de cada cadena para las DM realizadas. En ambos casos vemos que las dinámicas se comportan de forma esperada con altos grados de oscilación, al ser proteínas con regiones desordenadas. La cadena A tiene un comportamiento similar en ambas DMs a lo largo del tiempo. En cambio, las DMs de la cadena B a partir de los 600 ns se separan. Este comportamiento probablemente hace alusión a que partiendo de una condición espacial inicial diferente, la proteína tiene la posibilidad de adoptar una nueva conformación y oscila alrededor

de 2 conformaciones ligeramente diferentes. Por otro lado, vemos que el RMSF de ambas dinámicas, aunque parten de distinto lugar, tienen el mismo perfil, coincidiendo en la movilidad de las regiones.



Cuadro 2: Representaciones gráficas del RMSD y RMSF para las DM 2GS0 y 2GS0-15. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

De la distribución de los conformeros en el tiempo por cluster para cada DM (Figura 23), se desprende que en ambas dinámicas el cluster 1 es el que tiene el 75 % de los conformeros muestreados. En el Cuadro 3 se puede ver el número de conformeros por cluster.

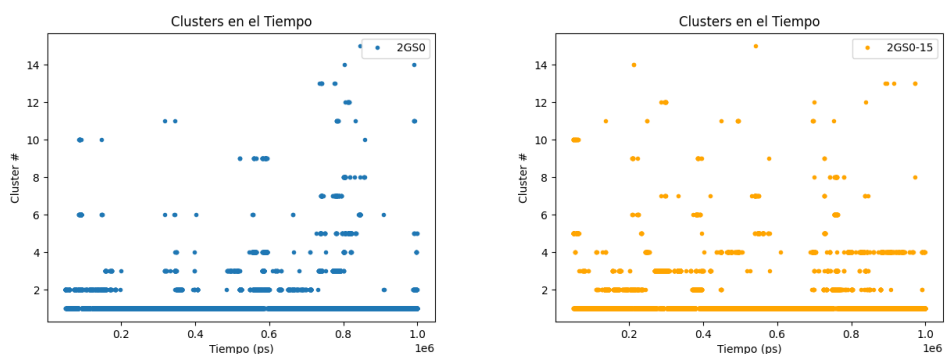
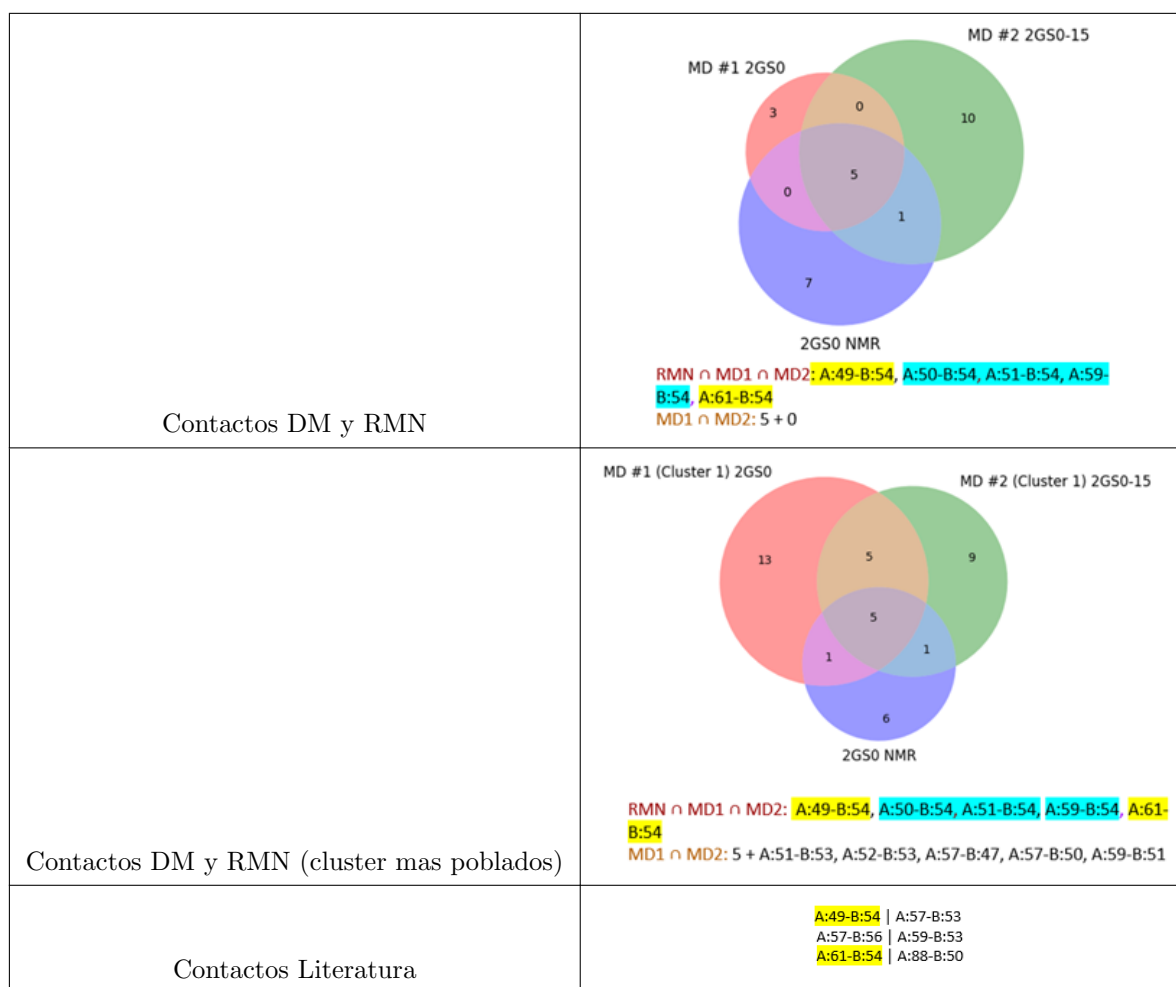


Figura 23: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2GS0 y 2GS0-15.

2GS0			2GS0-15		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	16378	86 %	1	16016	84 %
2	1459	8 %	2	1165	6 %
3	486	3 %	3	845	4 %
4	192	1 %	4	542	3 %
5	156	1 %	5	190	1 %
6	117	1 %	6	93	0 %
7	67	0 %	7	42	0 %
8	55	0 %	8	26	0 %
9	41	0 %	9	24	0 %
10	16	0 %	10	22	0 %
11	15	0 %	11	15	0 %
12	9	0 %	12	13	0 %
13	7	0 %	13	5	0 %
14	2	0 %	14	2	0 %
15	1	0 %	15	1	0 %
Total	19001			19001	

Cuadro 3: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

En el Cuadro 4 se puede observar los diagramas de Venn con las interacciones de cada dinámica, del RMN y la coincidencia entre ellas. Las compartidas por ambas dinámicas y el RMN son 5, éstas coinciden con las interacciones reportadas analizando sólo el cluster más poblado de cada DM. Además dos de ellas (contacto hidrofóbico: A:49-B:54, Pi-Catión: A:61-B:54) se encuentran también en la literatura, siendo la más fuerte del tipo pi-catión. Estas interacciones son tomadas como “core”, es decir las fundamentales para mantener el complejo. Sin embargo, analizando sólo el cluster 1 de las dos dinámicas, se observan otros 5 contactos que sin duda son relevantes dado que se mantienen a lo largo del tiempo (Figura 24).



Cuadro 4: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters más poblados, es decir los que contienen al 75 % de los complejos muestreados a lo largo del tiempo (centro). A la derecha se encuentran los contactos reportados como relevantes en la literatura. Los contactos resaltados en amarillo son coinciden entre las 3 técnicas y en la literatura, los resaltados en azul, coinciden entre las dos DMs.

A partir de la bibliografía consultada sobre el complejo 2GS0 [16], se calcularon las distancias de las posibles 6 interacciones detalladas en la misma (Cuadro 4). A modo de resumen en el Cuadro 5 se detalla el porcentaje total en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como para la DM 2. Se observa que la distancia entre A:61-B:54 y entre A:49-B:54 se encuentra por debajo de 0,7 nm más del 75 % del tiempo de simulación en ambos casos, a diferencia de los otros 4 contactos reportados en la literatura.

A partir de nuestro análisis, consideramos que alguno de los contactos reportados por la literatura pueden tener una importancia inferior, dado que aunque están presentes en algún conformero, no se mantienen a lo largo del tiempo. Sin embargo, es evidente la importancia de los 10 contactos (los dos que coinciden y los 8 propuestos) para la correcta ubicación de la cadena B, dado que la fijan en toda su longitud. Se puede hipotetizar que la interfaz del complejo mantiene la conformación en hélice de esa región de la cadena B al darle menos grados de libertad por las interacciones (Figura 24).

2GS0 — 2GS0-15		Cadena A				
		Glu 49	Lys 57	Met 59	Arg 61	Met 88
Cadena B	Ile 50	-	-	-	-	48 % — 6 %
	Trp 53	-	44 % — 12 %	52 % — 39 %	-	-
	Phe 54	99 % — 100 %	-	-	99 % — 100 %	-
	Glu 56	-	0 % — 1 %	-	-	-

Cuadro 5: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM. Se toman los aminoácidos que interaccionan según la literatura para el complejo 2GS0

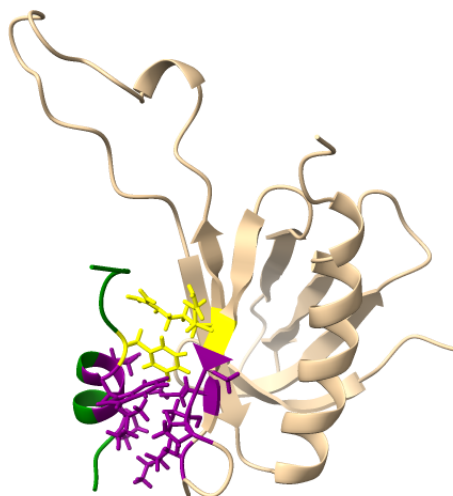


Figura 24: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos “core” (en contacto en todas las DMs, RMN y la literatura). En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.2. 2M0G

Se obtuvieron los siguientes resultados de la comparación de las matrices de contacto de las DM1 y DM2 completas partiendo del conformero 1 (2M0G) y el conformero más lejano 6 (2M0G-6) con el RMN.

La comparación entre las matrices de contacto del RMN y las DMs dan un score de 0.75 y 0.71 respectivamente para los conformeros más alejados del RMN. Al igual que con el complejo anterior, las matrices de contacto tienen un grado de similitud elevado (Anexo 3). Otra vez, denotando que las DMs tienen conformeros similares estableciendo aproximadamente los mismos contactos con la misma probabilidad.

Se observa en el Cuadro 6 tanto el RMSD como el RMSF de cada cadena para las DM realizadas. El RMSF es muy similar en ambas dinámicas para las dos cadenas, mostrando que en las dos, las zonas más móviles han sido los mismos segmentos. Por el otro, el RMSD de la cadena A tiene un comportamiento similar en ambas dinámicas, mientras que la cadena B, dependiendo de la conformación de entrada, tiene una oscilación alrededor de distintas conformaciones. Una explicación posible es que hay al menos dos conformeros más estables diferentes, alrededor del cual oscilan los complejos.

PDB	Cadena (aa)	RMSD	RMSF
2M0G	A SF1 (1-145)		
	B U2AF2C (372-475)		

Cuadro 6: Representaciones gráficas del RMSD y RMSF para las DM 2M0G y 2M0G-6. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Además observamos en la Figura 25 la distribución de los conformeros en el tiempo por cluster para cada DM. Detalladamente en el Cuadro 7 observamos que para la DM 1 los clusters 1 y 2 representan el 75 % de los complejos mientras que para la DM 2 este porcentaje se encuentra entre los clusters 1, 2 y 3.

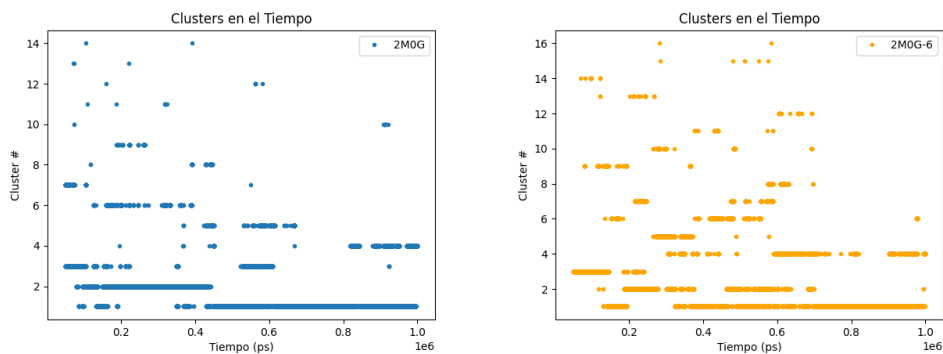
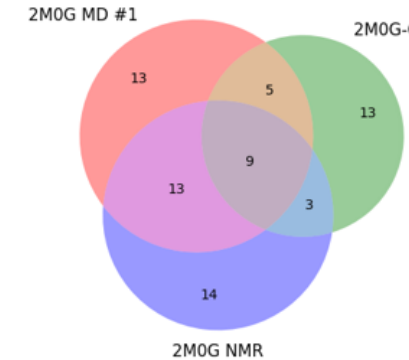
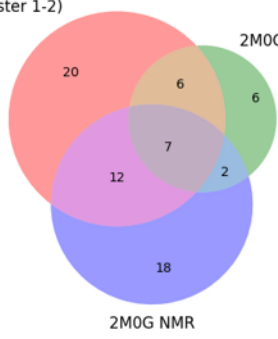


Figura 25: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2M0G y 2M0G-6.

2M0G			2M0G-6		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	9592	50 %	1	9686	51 %
2	5213	27 %	2	2821	15 %
3	1825	10 %	3	1887	10 %
4	1026	5 %	4	1811	10 %
5	535	3 %	5	998	5 %
6	513	3 %	6	773	4 %
7	161	1 %	7	464	2 %
8	55	0 %	8	259	1 %
9	52	0 %	9	120	1 %
10	10	0 %	10	80	0 %
11	7	0 %	11	29	0 %
12	7	0 %	12	27	0 %
13	3	0 %	13	23	0 %
14	2	0 %	14	15	0 %
			15	6	0 %
			16	2	0 %
Total	19001			19001	

Cuadro 7: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Se puede observar en el Cuadro 8 los diagramas de Venn con las interacciones de cada dinámica, del RMN y la coincidencia entre ellas. Las interacciones compartidas entre ambas dinámicas y el RMN son 9, dentro de ellas las más fuertes son de tipo Pi-cación (A:21-B:454) y contacto iónico (A:49-B:462). Por otro lado, observando los diagramas de Venn para los clusters más poblados de cada DM vemos que las interacciones coincidentes se reducen a 7, probablemente las otras dos estén presentes en clusters con baja representación. Dentro de estas se encuentran las mencionadas anteriormente (A:21-B:454 y A:49-B:462). Además vemos que las interacciones A:53-B:460 y A:49-B:462 se encuentran en ambos diagramas de Venn y en la literatura. Analizando los clusters más representativos de cada dinámica, aparece otra interacción reportada en la literatura (A:56-B:458) y que está presente en ambas dinámicas pero no en el RMN.

Contactos DM y RMN	 2M0G MD #1 2M0G-6 MD #2 2M0G NMR RMN $\cap$ MD1 $\cap$ MD2: A:21-B:394, A:22-B:401, A:21-B:454, A:22-B:452, A:22-B:405, A:22-B:453, A:22-B:459, A:49-B:462, A:53-B:460 MD1 $\cap$ MD2: 9 + A:21-B:455, A:22-B:383, A:22-B:402, A:22-B:454, A:25-B:453
Contactos DM y RMN (cluster mas poblados)	 2M0G MD #1 (Cluster 1-2) 2M0G-6 MD #2 (Cluster 1-2-3) 2M0G NMR RMN $\cap$ MD1 $\cap$ MD2: A:21-B:454, A:22-B:405, A:22-B:452, A:22-B:453, A:22-B:459, A:49-B:462, A:53-B:460 MD1 $\cap$ MD2: 7 + A:22-B:383, A:22-B:454, A:25-B:453, A:56-B:458, A:60-B:453, A:64-B:453
Contactos Literatura	A:39-B:381 A:39-B:458 A:39-B:460 A:40-B:381 A:40-B:458 A:40-B:460 A:49-B:462 A:53-B:381 A:53-B:458 A:53-B:460 A:56-B:381 A:56-B:458 A:56-B:460

Cuadro 8: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters más poblados (centro). A la derecha se encuentran los contactos reportados en la literatura. Los contactos resaltados en amarillo son coinciden entre las 3 técnicas y en la literatura, los resaltados en azul, coinciden entre las dos DMs y verde, coincide entre la DM de los clusters más representativos y la literatura.

A partir de la bibliografía consultada sobre el complejo 2M0G [19], se calcularon las distancias de las posibles interacciones detalladas en la misma. A modo de resumen en el Cuadro 9 se detallan el porcentajes total en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como para la DM 2. Se observa que la distancia entre A:40-B:381, A:40-B:460, A:53-B:381 y A:53-B:458 se encuentra por debajo de 0,7 nm más del 75 % del tiempo solo para una DM. En este caso se observa que la distancia entre A:56-B:460, se encuentra por debajo de 0,7 nm más del 75 % del tiempo de simulación formando un contacto que no hemos detectado con el análisis realizado de las DMs

2M0G — 2M0G-6		Cadena A				
		Val 39	Ile 40	Glu 49	Ile 53	Leu 56
Cadena B	Met 381	0 — 0	99 % — 1 %	-	99 % — 59 %	1 % — 0
	Val 458	0 — 0	0 — 0	-	62 % — 91 %	97 % — 97 %
	Val 460	30 % — 0	100 % — 1 %	-	100 % — 98 %	99 % — 86 %
	Lys 462	-	-	99 % — 98 %	-	-

Cuadro 9: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM. Se toman los aminoácidos que interaccionan según la literatura para el complejo 2M0G

En la Figura 26, se representan los aminoácidos en contacto “core” en amarillo (intersección entre las 2 DMs y el RMN). Observamos al igual que en el ejemplo anterior, que hay otro set de contactos que permanecen durante el tiempo en los clusters más representativos que parecen ser relevantes al aumentar la interfaz de interacción entre las cadenas, de esta manera estabilizando el complejo. Es relevante destacar que los nuevos contactos encontrados por este análisis, mantienen una región muy desestructurada en una conformación fija, cerca de la región globular. Esta región desestructurada, podría tomar configuraciones muy diversas, alejándose de la interfaz de interacción de no existir estos contactos.

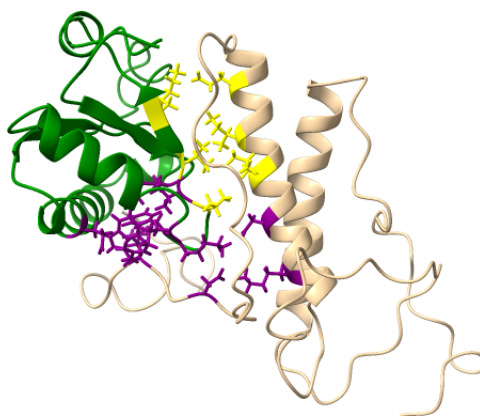


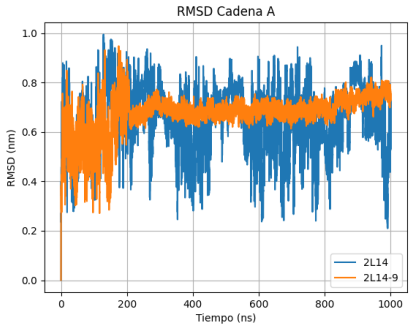
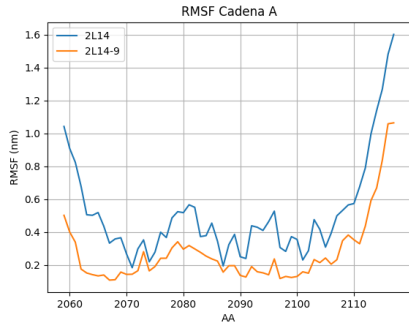
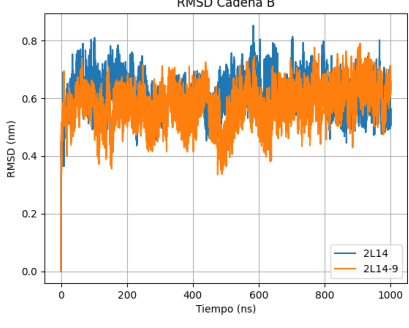
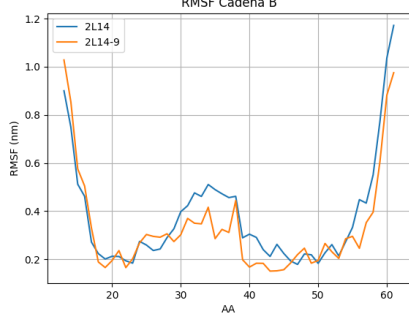
Figura 26: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos “core” (en contacto en todas las DMs, RMN y la literatura). En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.3. 2L14

Se obtuvieron los siguientes resultados de las dinámicas realizadas partiendo del conformero 1 (2L14) y el conformero más lejano 9 (2L14-9):

La comparación entre las matrices de contacto del RMN y las DMs dan un score de 0.76 y 0.74 respectivamente para los conformeros más alejados del RMN. Al igual que con el complejo anterior, las matrices de contacto tienen un grado de similitud elevado (Anexo 3). Otra vez, denotando que las DMs tienen conformeros similares estableciendo aproximadamente los mismos contactos con la misma probabilidad.

Se observa en el Cuadro 10 tanto el RMSD como el RMSF de cada cadena para las DMs realizadas, en ambos casos vemos que las dinámicas se comportan de la forma esperada con altos grados de oscilación al ser proteínas con regiones desordenadas. Por el otro lado las diferencias de RMSD en la cadena A hacen alusión a que partiendo de una condición espacial inicial diferente se obtiene un comportamiento distinto del complejo, con oscilaciones mayores.

PDB	Cadena (aa)	RMSD	RMSF
2L14	A CBP (2059-2117)		
	B TAD p53 (13-61)		

Cuadro 10: Representaciones gráficas del RMSD y RMSF para las DM 2L14y 2L14-9. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Además observamos en la Figura 27 y en el Cuadro 11 la distribución de los cónfórmeros en el tiempo por cluster para cada DM. Observamos que para la DM 1 los clusters 1 y 2 representan el 75 % de la dinámica mientras que para la DM 2 este porcentaje se encuentra en el cluster 1.

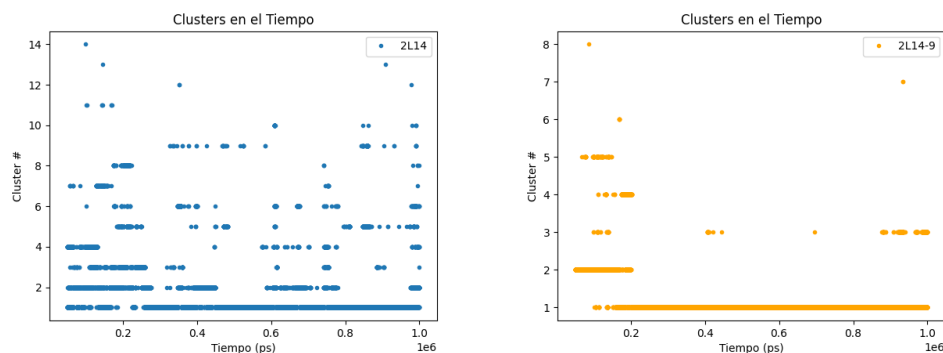
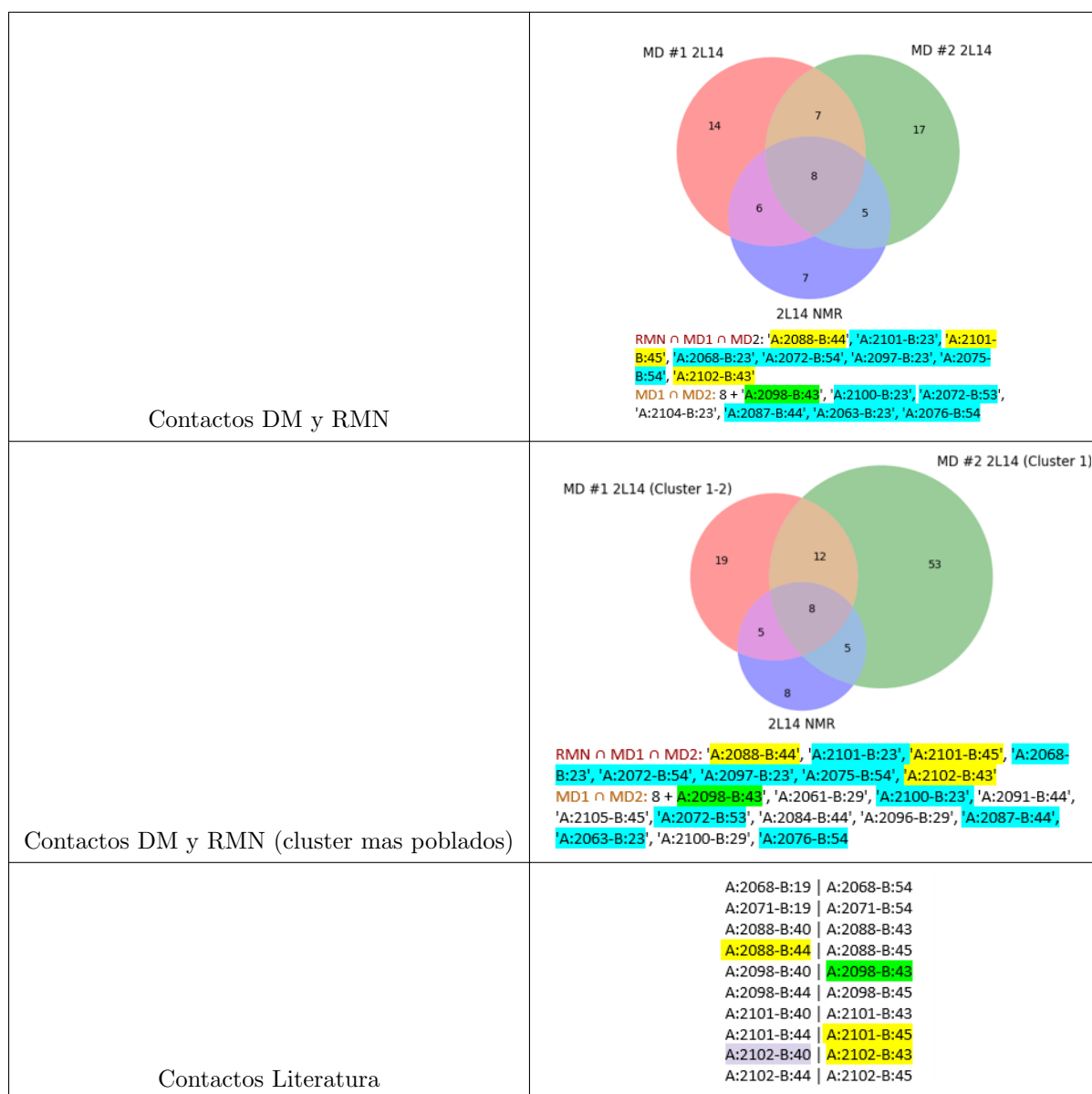


Figura 27: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2L14y 2L14-9.

2L14			2L14-9		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	12724	67 %	1	16128	85 %
2	3294	17 %	2	2260	12 %
3	1098	6 %	3	263	1 %
4	636	3 %	4	208	1 %
5	460	2 %	5	134	1 %
6	370	2 %	6	5	0 %
7	176	1 %	7	2	0 %
8	115	1 %	8	1	0 %
9	88	0 %			
10	26	0 %			
11	8	0 %			
12	3	0 %			
13	2	0 %			
14	1	0 %			
Total	19001			19001	

Cuadro 11: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Se puede observar en el Cuadro 12 los diagramas de Venn con las interacciones de cada dinámica, del RMN y la coincidencia entre ellas. Las interacciones compartidas entre ambas dinámicas y el RMN son 8, estas se mantienen tomando los clusters más poblados y 3 de ellas se reportan en la literatura (A:2088-B:44, A:2101-B:45, A:2102-B:43). Estas son las interacciones que llamamos “core”, fundamentales para la unión del complejo. Otra interacción, reportada (A:2098-B:43) sólo se observa en las DM al analizar los clusters más representativos.



Cuadro 12: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters más poblados, es decir los que contienen al 75 % de los complejos muestreados a lo largo del tiempo (centro). A la derecha se encuentran los contactos reportados como relevantes en la literatura. Se sigue el mismo esquema de colores para resaltar los contactos.

A partir de la bibliografía consultada sobre el complejo 2L14[22], se calcularon las distancias de las posibles interacciones detalladas en la misma. A modo de resumen en la Cuadro 13 se detallan el porcentajes total en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como para la DM 2. Se observa que la distancia entre A:2068-B:19, A:2017-B:19, A:2088-B:40, A:2098-B:40, A:2098-B:44, A:2101-B:43, A:2101-B:44, A:2101-B:45 y A:2102-B:45 se encuentra por debajo de 0,7 nm más del 75 % del tiempo solo para una de las simulaciones.

Por otro lado, se observa que la distancia entre A:2102-B:40, se encuentra por debajo de 0,7 nm más del 75 % del tiempo de simulación y no la hemos detectado en nuestro análisis.

2L14 — 2L14-9		Cadena A					
		Leu 2068	Leu 2071	Leu 2088	Met 2098	Phe 2101	Ile 2102
Cadena B	Phe 19	100 % — 5 %	99 % — 0	-	-	-	-
	Met 40	-	-	8 % — 76 %	28 % — 100 %	20 % — 0	93 % — 76 %
	Leu 43	-	-	51 % — 17 %	93 % — 98 %	92 % — 0	92 % — 100 %
	Met 44	-	-	100 % — 100 %	57 % — 100 %	84 % — 9 %	6 % — 28 %
	Leu 45	-	-	2 % — 26 %	0 — 100 %	67 % — 100 %	0 — 100 %
	Phe 54	0 — 0	16 % — 10 %	-	-	-	-

Cuadro 13: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM del complejo 2L14.

La Figura 28 muestra los contactos en la estructura 3D del complejo. Los core y los relevantes según nuestro análisis. Otra vez observamos que los contactos core son relevantes para la unión del complejo, pero los que se desprenden de nuestro análisis establecen una red de contactos densa entre las dos cadenas en una gran área de superficie. Se puede hipotetizar que la unión entre estas cadenas es fuerte y estable en el tiempo y que todos esos contactos son los responsables de mantener al complejo en contacto cercano con una gran superficie de interacción.

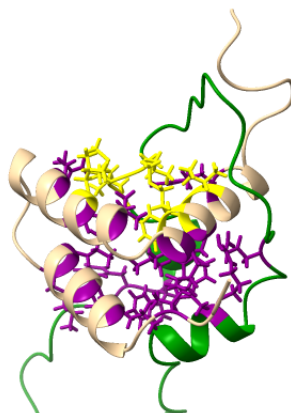


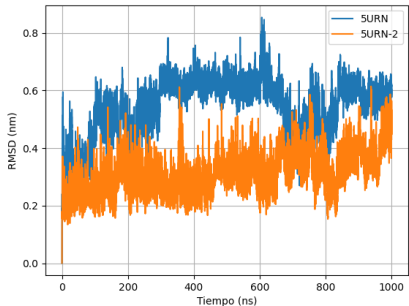
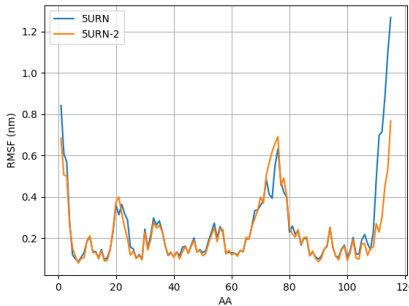
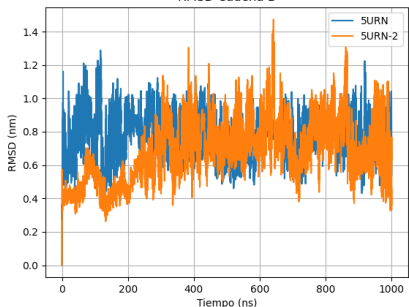
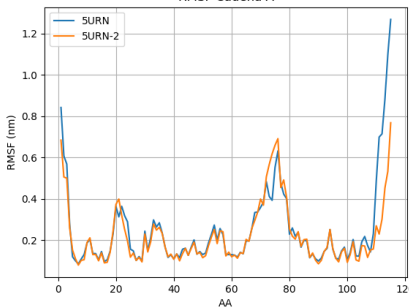
Figura 28: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos en contacto en todas las DMs, RMN y la literatura. En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.4. 5URN

Se obtuvieron los siguientes resultados de las dinámicas realizadas partiendo del conformero 1 (5URN) y el conformero más lejano 2 (5URN-2):

La comparación entre las matrices de contacto del RMN con las dos dinámicas de los conformeros más lejanos dieron un score de 0.83 y 0.84 respectivamente (Anexo 3). Demostrando también en este caso, que las DMs reproducen las conformaciones de los complejos similares al RMN.

Se observa en el Cuadro 14 tanto el RMSD como el RMSF de cada cadena para las DMs realizadas. En ambos casos vemos que las dinámicas se comportan de forma esperada con altos grados de oscilación, al ser proteínas con regiones desordenadas. Por el otro lado, las diferencias de RMSD en la cadena A, hacen alusión a que partiendo de una condición espacial inicial diferente el complejo quede oscilando alrededor de conformaciones ligeramente diferentes.

PDB	Cadena (aa)	RMSD	RMSF
5URN	A TFB1 (2-115)		
	B p65 (521-551)		

Cuadro 14: Representaciones gráficas del RMSD y RMSF para las DM 5URN y 5URN-2.

Observamos en la Figura 29 y en el Cuadro 15 que para ambas DMs los clusters 1 y 2 representan el 75 % de la dinámica.

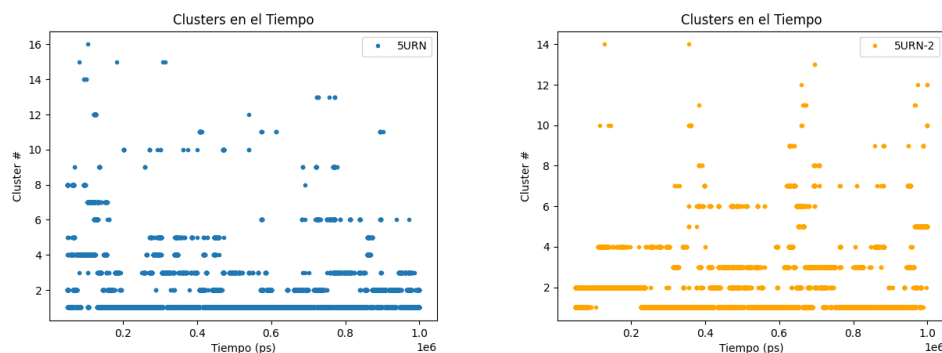
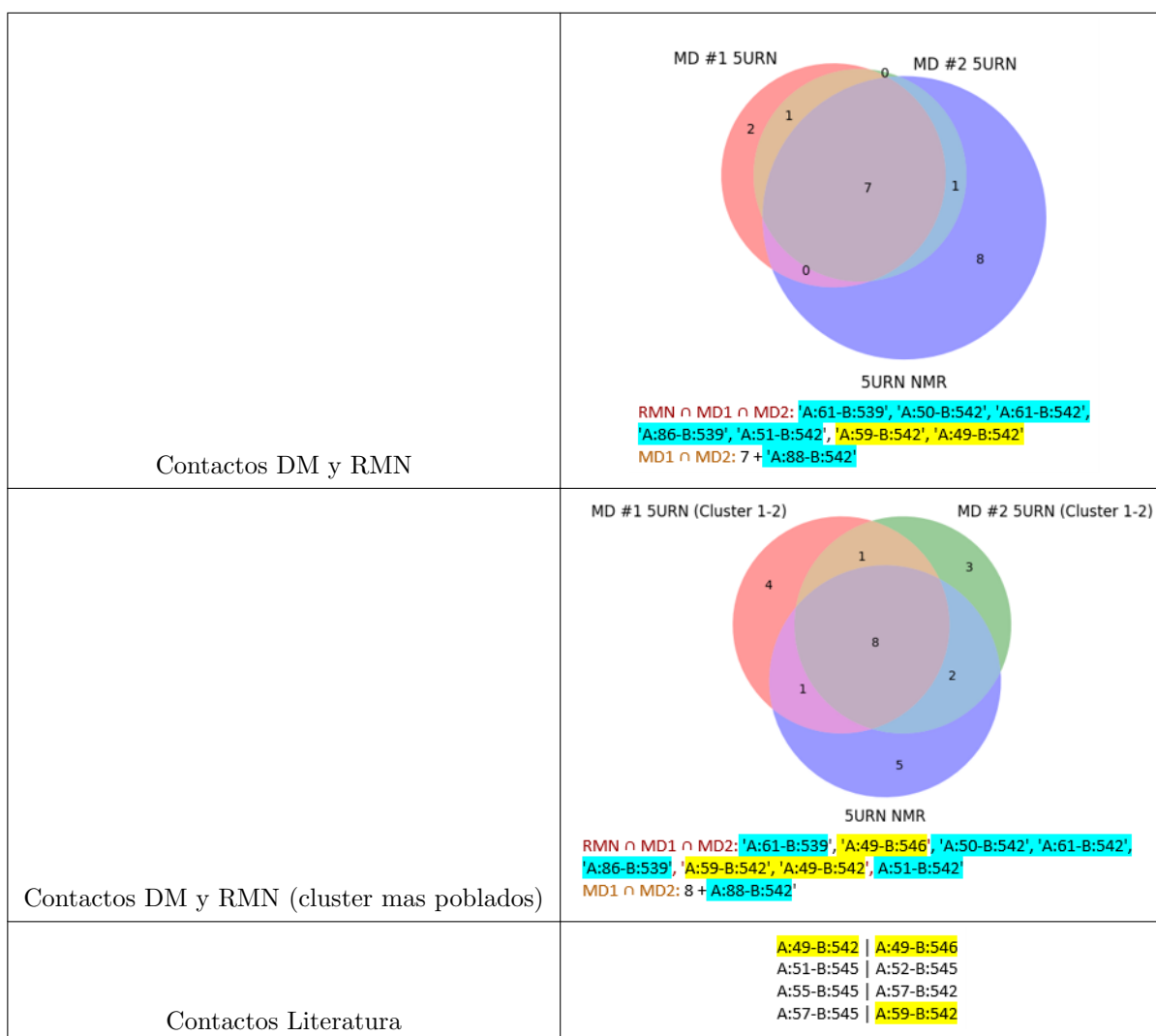


Figura 29: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 5URN y 5URN-2. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

5URN			5URN-2		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	12550	66 %	1	10729	56 %
2	2780	15 %	2	4342	23 %
3	1373	7 %	3	1952	10 %
4	1262	7 %	4	868	5 %
5	316	2 %	5	534	3 %
6	276	1 %	6	288	2 %
7	181	1 %	7	201	1 %
8	81	0 %	8	27	0 %
9	68	0 %	9	25	0 %
10	38	0 %	10	20	0 %
11	37	0 %	11	7	0 %
12	22	0 %	12	4	0 %
13	8	0 %	13	2	0 %
14	4	0 %	14	2	0 %
15	4	0 %			
16	1	0 %			
Total	19001			19001	

Cuadro 15: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Se puede observar en el Cuadro 16 los diagramas de Venn con las interacciones de cada dinámica, del RMN y la coincidencia entre ellas. Las interacciones que coinciden entre ambas dinámicas y el RMN son 7, de las cuales 2 se reportan en la literatura (puente de hidrógeno A:59-B:542 y contacto hidrofóbico A:49-B:542). Tomando sólo los clusters más poblados observamos que el número de interacciones comunes entre las DM y el RMN asciende a 8, y encontramos que a las interacciones que coinciden con la literatura ahora se les suma la A:49-B:546, contacto hidrofóbico. Dentro de las interacciones comunes encontradas en ambos diagramas de Venn vemos la interacción A:61-B:542 que es del tipo pi catión.



Cuadro 16: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters más poblados, es decir los que contienen al 75 % de los complejos muestreados a lo largo del tiempo (centro). A la derecha se encuentran los contactos reportados como relevantes en la literatura. Se sigue el mismo esquema de colores para resaltar los contactos.

A partir de la bibliografía consultada sobre el complejo 5URN [27], se calcularon las distancias de las posibles interacciones detalladas en la misma. A modo de resumen en el Cuadro 17 se detallan el porcentajes total en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como para la DM 2. Se observa que la distancia entre A:49-B-546 se encuentra por debajo de 0,7 nm más del 70 % del tiempo solo para una de las simulaciones.

5URN — 5URN-2		Cadena A					
		Gln 49	Thr 51	Pro 52	Ser 55	Lys 57	Met 59
Cadena B	Phe 542	100 % — 96 %	-	-	-	39 % — 18 %	100 % — 100 %
	Leu 545	-	4 % — 51 %	18 % — 67 %	9 % — 41 %	3 % — 28 %	-
	Leu 546	83 % — 46 %	-	-	-	-	-

Cuadro 17: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM. Se toman los aminoácidos que interaccionan según la literatura para el complejo 5URN

Representando en la estructura 3D los contactos core y los encontrados en los clusters más representativos, otra vez se puede observar que este segundo set de contactos dá mayor superficie de interacción y estabilidad al complejo a lo largo del tiempo (Figura 30).

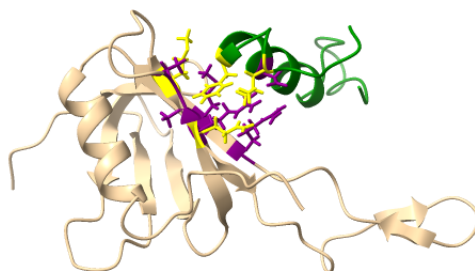


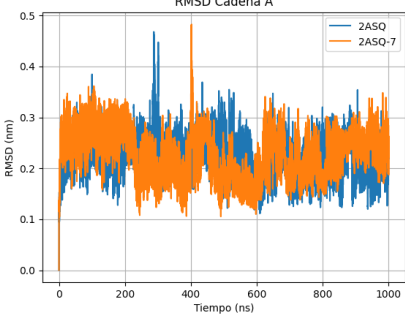
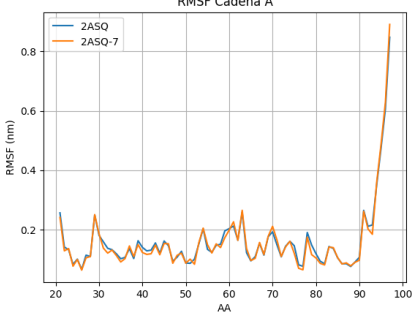
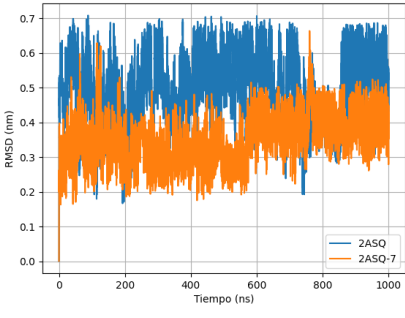
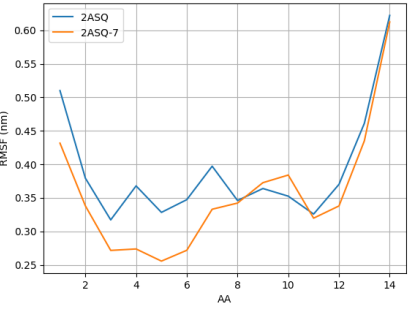
Figura 30: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos en contacto en todas las DMs, RMN y la literatura. En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.5. 2ASQ

Se obtuvieron los siguientes resultados de las dinámicas realizadas partiendo del cónfórmero 1 (2ASQ) y el cónfórmero más lejano 7 (2ASQ-7):

En este caso vemos un resultado diferente a los casos anteriores. La comparación entre la matriz de contacto entre una de las dinámicas moleculares y el RMN da un score de 0.53, demostrando que no son tan similares. La otra DM tiene una matriz de contacto más similar al RMN, con un score de 0.79. (Anexo 3)

Se observa en la Cuadro 18 tanto el RMSD como el RMSF de cada cadena para las DM realizadas, en ambos casos vemos que las dinámicas se comportan de la forma esperada con un alto grado de oscilación al ser proteínas con regiones desordenadas.

PDB	Cadena (aa)	RMSD	RMSF
2ASQ	A SUMO-1 (21-97)		
	B SBM PIASx (466-479)		

Cuadro 18: Representaciones gráficas del RMSD y RMSF para las DM 2ASQ y 2ASQ-7. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Además, observamos en la Figura 31 y el Cuadro 19 la distribución de los confórmeros en el tiempo por cluster para cada DM. En el caso de la DM 1 encontramos 6 clusters de los cuales sólo en el primero los complejos se encuentran unidos. Sorprendentemente, el 98 % de las representaciones se encuentran en este cluster. Esto explica que las matrices de contacto entre la DM 1 y el RMN tengan un score bajo en la comparación. En ellas se muestran la probabilidad de que un par de residuos esté en contacto, según la cantidad de modelos en los que aparece. En este caso son 6 (un representante por cada cluster, de los cuales en 5 las cadenas están separadas). Esto da una matriz de contactos que no coincide con el RMN, 10 modelos todos unidos. Por otro lado en la DM 2 obtenemos 7 clusters de los cuales en todos se mantienen en contacto las cadenas. De igual forma el cluster 1 obtiene el 95 % de las representaciones totales.

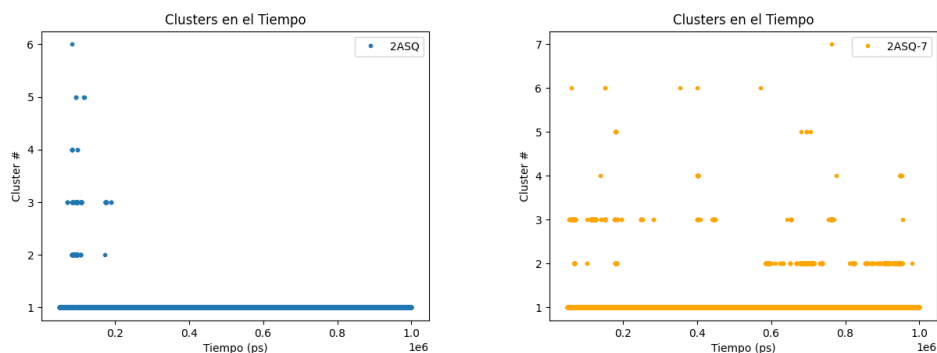


Figura 31: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2ASQ y 2ASQ-7.

2ASQ			2ASQ-7		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	18686	98 %	1	18144	95 %
2	183	1 %	2	572	3 %
3	116	1 %	3	233	1 %
4	9	0 %	4	33	0 %
5	6	0 %	5	12	0 %
6	1	0 %	6	6	0 %
			7	1	0 %
Total	19001			19001	

Cuadro 19: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Luego de realizado el análisis por cadena, se prosiguió analizando el complejo en su totalidad. Se encontró que en la DM 1 las cadenas se separan en el espacio a lo largo del tiempo de simulación. Como se puede observar en la Figura 32, la distancia entre ambas cadenas es tal que no existe interacción entre las mismas.

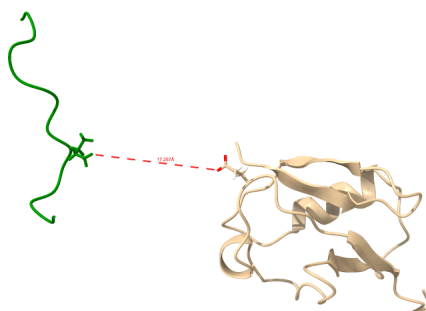
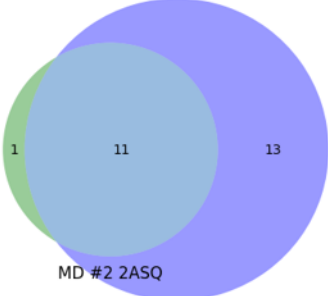


Figura 32: En amarillo la cadena A (Enzima E2), en verde se observa la cadena B (Enzima E1) y en la línea punteada en rojo marca la distancia en armnstrongs (Å) de los aminoacidos mas cercanos entre las cadenas.

Como mencionamos, el cluster 1 contiene al 98 % de los frames de la DM, diciendo de cierta manera que el 98 % del tiempo de dinámica el complejo está unido, por lo que el análisis de la DM en su totalidad nos conduciría a un resultado erróneo.

En el Cuadro 20 observamos que en la DM 1 completa hay una sola interacción entre las dos cadenas (concordando con lo anterior) y ninguna coincidencia con el RMN. La DM 2 comparte varias interacciones con el RMN. Considerando las coincidencias entre técnicas como un indicador de que un contacto es “core”, esto daría como resultado que no hay ninguna interacción de este tipo. Por lo tanto el complejo no tiene las interacciones necesarias para mantener las cadenas unidas.

Sin embargo, al analizar los clusters más representativos (recordar que un sólo cluster en la DM 1 contiene el 95 % de las estructuras) se encuentran interacciones comunes entre ambas DMs y el RMN. De esta manera hay 16 interacciones entre las 2 DMs y el RMN y otra entre las 2 dinámicas del cluster representativo. Comparando estas interacciones relevantes con las reportadas en la literatura, coinciden 4: un puente de hidrógeno (A:35-B:2), tres contactos hidrofóbicos (A:35-B:4, A:36-B:5, A:47-B:7). La interacción A:38-B:7 (contacto hidrofóbico), reportada en la literatura se observa en las DM pero no al analizar el RMN.

Contactos DM y RMN	 MD #2 2ASQ 2ASQ NMR RMN $\cap$ MD1 $\cap$ MD2: 0 MD1 $\cap$ MD2: 0 (MD1 showed structures with separated chains)								
Contactos DM y RMN (cluster mas poblados)	 MD #1 2ASQ (Cluster 1) MD #2 2ASQ (Cluster 1) 2ASQ NMR RMN $\cap$ MD1 $\cap$ MD2: A:54-B:5 'A:50-B:7', 'A:35-B:3', 'A:34-B:5', 'A:35-B:4', 'A:37-B:7', 'A:36-B:4', 'A:35-B:5', 'A:35-B:2' 'A:36-B:5', 'A:37-B:6', 'A:37-B:5', 'A:23-B:4', 'A:54-B:5', 'A:37-B:4', 'A:36-B:7', 'A:47-B:7' MD1 $\cap$ MD2: 16 + A:38-B:7								
Contactos Literatura	<table border="0"> <tr> <td>A:33-B:2</td> <td>A:34-B:2</td> </tr> <tr> <td>A:35-B:2</td> <td>A:34-B:3</td> </tr> <tr> <td>A:35-B:4</td> <td>A:36-B:5</td> </tr> <tr> <td>A:38-B:7</td> <td>A:47-B:7</td> </tr> </table>	A:33-B:2	A:34-B:2	A:35-B:2	A:34-B:3	A:35-B:4	A:36-B:5	A:38-B:7	A:47-B:7
A:33-B:2	A:34-B:2								
A:35-B:2	A:34-B:3								
A:35-B:4	A:36-B:5								
A:38-B:7	A:47-B:7								

Cuadro 20: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters más poblados (centro). A la derecha se encuentran los contactos reportados en la literatura.

Este análisis revela que los 24 contactos presentes en el ensamble de complejos de RMN no tienen todos la misma importancia. Además, tomar las DMs sin un clustering y sin verificar que el análisis se está haciendo

con el ensamble de complejos más representativos, tampoco permite identificar los contactos relevantes. La Figura 33 muestra los contactos “core”, resultado de la intersección entre contactos de las DM analizando los complejos del cluster más representativo, y el segundo set de contactos relevantes. Se observa que la interfaz del complejo está estabilizada por una amplia red de interacciones a lo largo de toda su superficie. Este es un ejemplo interesante que resalta la utilidad del análisis propuesto en este trabajo.



Figura 33: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos en contacto en todas las DMs, RMN y la literatura. En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.6. 2LPB

A continuación se analizaron las dinámicas realizadas partiendo del conformero 1 (2LPB) y el conformero más lejano 6 (2LPB-6). El comportamiento de este complejo es un interesante caso de estudio, que se desarrollará a continuación.

La comparación de las matrices de contacto entre el RMN y la DM del conformero 1 y la del 2 dió un score de 0.84 y 0.85 respectivamente (Anexo 3). En este caso, las matrices de contacto son muy similares. Otra vez, mostrando que en la DM está representado el conjunto de conformeros que se obtienen por RMN.

En el Cuadro 21 se muestra que el RMSD de cada cadena para las DMs realizadas se comportan de forma esperada, con altos grados de oscilación al ser proteínas con regiones desordenadas. A la vez se observa que hay ligeras diferencias en el comportamiento de la cadena B entre las dos corridas dependiendo del estado inicial de la dinámica. Al igual que en otros casos, probablemente debido a conformaciones ligeramente distintas alrededor de las cuales los complejos permanecen oscilando. El RMSF de la cadena A tiene exactamente el mismo perfil a diferencia de la cadena B, donde aparenta haber regiones de distinta movilidad entre las dinámicas, indica la posibilidad de que una región quede menos flexible en un complejo que en otros.

PDB	Cadena (aa)	RMSD	RMSF
2LPB	A Gal11 (158-238)		
	B Gcn4 (101-134)		

Cuadro 21: Representaciones gráficas del RMSD y RMSF para las DM 2LPB y 2LPB-6. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Además, observamos en la Figura 34 y el Cuadro 22 que los clusters 1 y 2 representan el 75 % de los frames de la DM. Lo que significa que los complejos están en esas conformaciones más del 75 % del tiempo de simulación.

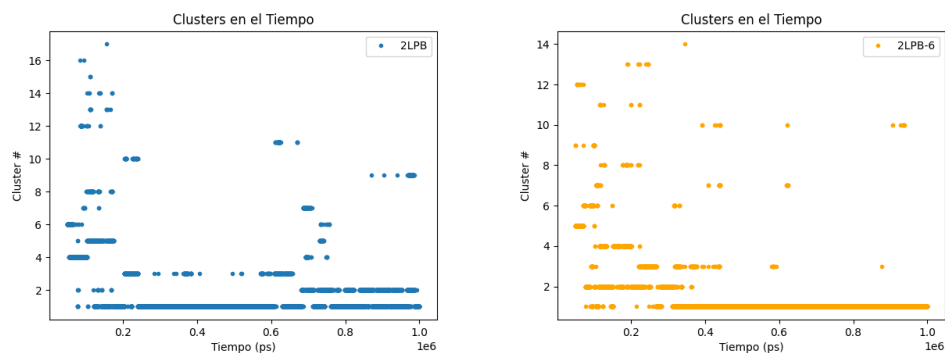
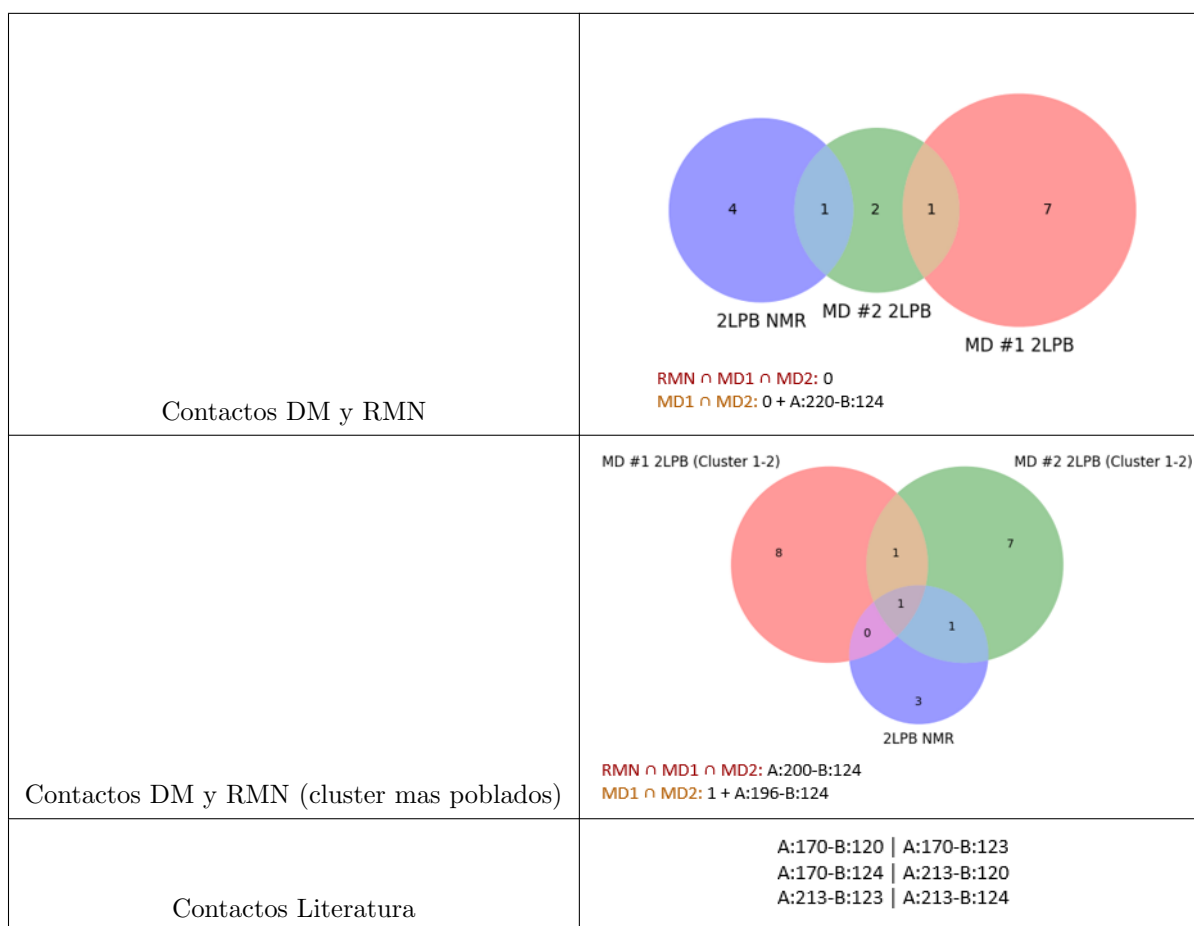


Figura 34: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2LPB y 2LPB-6.

2LPB			2LPB-6		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	12696	67 %	1	13656	72 %
2	2310	12 %	2	2486	13 %
3	1342	7 %	3	954	5 %
4	820	4 %	4	848	4 %
5	764	4 %	5	430	2 %
6	228	1 %	6	276	1 %
7	196	1 %	7	130	1 %
8	144	1 %	8	58	0 %
9	136	1 %	9	54	0 %
10	112	1 %	10	44	0 %
11	98	1 %	11	26	0 %
12	94	0 %	12	20	0 %
13	30	0 %	13	18	0 %
14	20	0 %	14	1	0 %
15	6	0 %			
16	4	0 %			
17	1	0 %			
Total	19001			19001	

Cuadro 22: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

El Cuadro 23 muestra los diagramas de Venn con las interacciones entre residuos de cada DM completa, del NMR y la coincidencia entre ellos. Si bien los tres conjuntos de datos presentan interacciones (el RMN tiene 5 interacciones, la DM 1 8 interacciones y la DM 2 4 interacciones), no hay coincidencias entre ellos. La DM 2 comparte 1 contacto con el RMN y 1 con la DM 1, a su vez, la DM 1 no comparte contactos con el RMN. Contrariamente, tomando los clusters más poblados las 3 técnicas comparten 1 contacto (contacto hidrofóbico A:200-B:124) y hay otro contacto representado en la mayor parte de las DMs (A:196-B:124). Ninguno de los contactos presentes en las DM coinciden con los reportados en la literatura.



Cuadro 23: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters mas poblados (centro). A la derecha se encuentran los contactos reportados en la literatura.

A partir de la bibliografía consultada sobre el complejo 2LPB [24], se calcularon las distancias de las posibles interacciones detalladas en la misma. A modo de resumen en el Cuadro 24 se detallan el porcentajes total en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como para la DM 2. Se observa que la distancia entre A:213-B:120, y A:213-B:124 se encuentra por debajo de 0,7 nm más del 75 % del tiempo sólo para una de las DM. Por otro lado, no se observa ninguna interacción con distancia por debajo de 0,7 nm más del 75 % del tiempo de simulación para los otros 4 contactos reportados en la literatura.

2LPB — 2LPB-6		Cadena A	
		Val 170	Met 213
Cadena B	Trp 120	0 — 0	99 % — 1 %
	Leu 123	0 — 0	0 — 0
	Phe 124	30 % — 0	100 % — 1 %

Cuadro 24: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM. Se toman los aminoácidos que interaccionan según la literatura para el complejo 2LPB

Este resultado sugiere que los contactos reportados en la literatura y el RMN no son relevantes cuando se considera el comportamiento de los complejos a lo largo del tiempo. Otra observación es que sólo 2 contactos son los fundamentales para mantener el complejo: A:200-B:124 y A:196-B:124, representados en la Figura 35. Harán falta otros experimentos para comprobar la estabilidad de este complejo en un entorno natural.

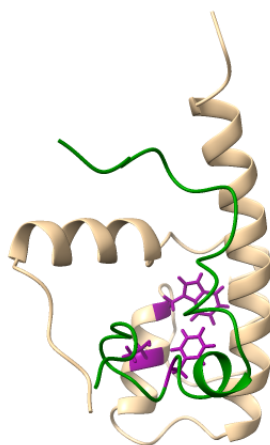


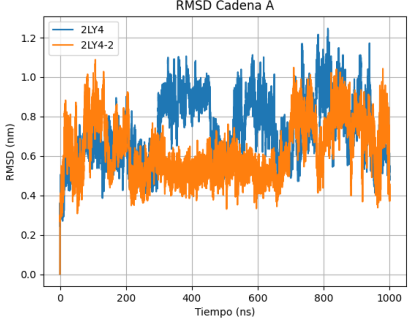
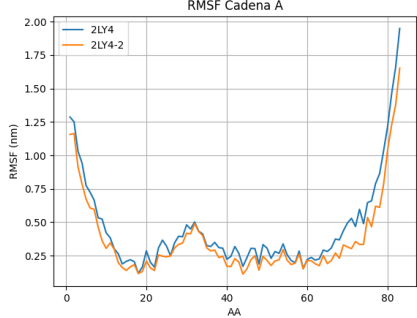
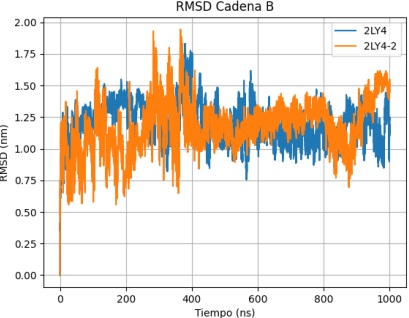
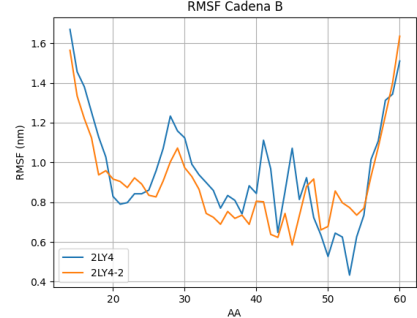
Figura 35: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos en contacto en todas las DMs, RMN y la literatura. En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.7. 2LY4

El comportamiento del complejo 2LY4 es particular, diferente de los anteriores, haciendo el mismo un buen caso de estudio que permite comprobar los alcances del análisis propuesto. Nuevamente la comparación de las matrices de contacto entre ambas DMs y el RMN dan un alto score de similitud 0.82 y 0.82 (Anexo 3)

Esto refleja que el ensamble de complejos obtenidos por las DMs y los modelos de RMN tienen una distribución similar de conformaciones. Recordar que el score de similitud compara si la probabilidad de estar en contacto de cada par de residuos es similar, por ejemplo, si un contacto es poco probable en el RMN (presente en pocos modelos), tiene que ser poco probable también el la DM.

Se observa en el Cuadro 25 que tanto el RMSD como el RMSF de cada cadena para las DM realizadas se comportan de la forma esperada, con altos grados de oscilación al ser proteínas con regiones desordenadas. Se observan leves diferencias en los RMSD en ambas cadenas durante las dos DMs mientras que el RMSF sigue el mismo perfil. Indicando que partiendo de una condición espacial inicial diferente se obtiene el mismo comportamiento del complejo.

PDB	Cadena (aa)	RMSD	RMSF
2LY4	A HMGB1 (2-84)		
	B TAD p53 (14-60)		

Cuadro 25: Representaciones gráficas del RMSD y RMSF para las DM 2LY4 y 2LY4-2. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Observamos en la Figura 36 y el Cuadro 26 la distribución de los conformeros en el tiempo por cluster para cada DM. Observamos que para la DM 1 los clusters 1, 2, 3 y 4 representan el 75 % de la dinámica mientras que para la DM 2 los clusters 1, 2, 3 representan el 75 % de la dinámica. En este caso resalta que no es muy predominante un cluster respecto de otros.

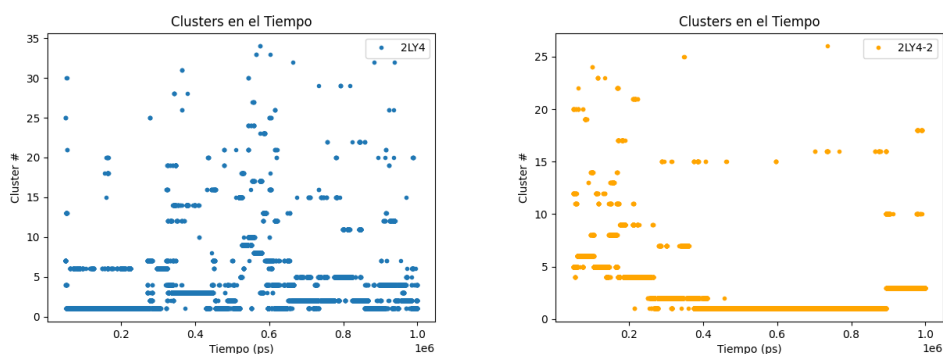
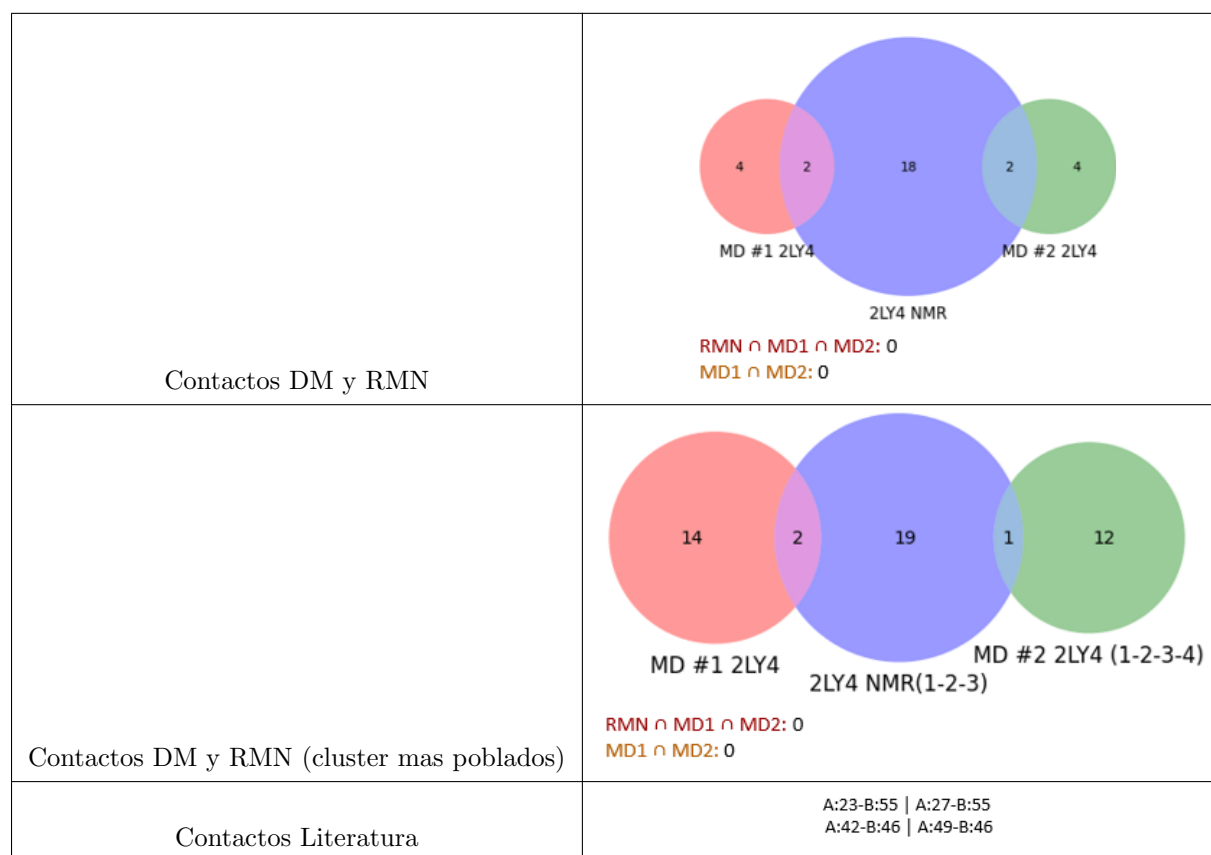


Figura 36: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2LY4 y 2LY4-2.

2LY4			2LY4-2		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	7181	38 %	1	10212	54 %
2	3459	18 %	2	1978	10 %
3	2066	11 %	3	1971	10 %
4	1556	8 %	4	1573	8 %
5	1218	6 %	5	923	5 %
6	903	5 %	6	748	4 %
7	467	2 %	7	402	2 %
8	323	2 %	8	347	2 %
9	314	2 %	9	257	1 %
10	240	1 %	10	131	1 %
11	196	1 %	11	93	0 %
12	192	1 %	12	54	0 %
13	168	1 %	13	53	0 %
14	131	1 %	14	46	0 %
15	131	1 %	15	42	0 %
16	119	1 %	16	36	0 %
17	93	0 %	17	31	0 %
18	47	0 %	18	25	0 %
19	35	0 %	19	20	0 %
20	29	0 %	20	19	0 %
21	23	0 %	21	17	0 %
22	19	0 %	22	11	0 %
23	18	0 %	23	7	0 %
24	16	0 %	24	2	0 %
25	14	0 %	25	2	0 %
26	9	0 %	26	1	0 %
27	6	0 %			
28	6	0 %			
29	6	0 %			
30	5	0 %			
31	3	0 %			
32	3	0 %			
33	3	0 %			
34	2	0 %			
Total	19001			19001	

Cuadro 26: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

En el Cuadro 27 se muestra el diagrama de Venn con las interacciones de cada dinámica completa, del RMN y la coincidencia entre ellas. Se puede apreciar que no hay contactos comunes. El ensamble de complejos resuelto por RMN presenta un conjunto de 22 contactos y ambas DMs 6 contactos cada una, pero no los mismos. Las 2 interacciones comunes entre la DM 2 y el RMN son del tipo contacto hidrofóbico, mientras que las 2 interacciones comunes entre la DM 1 y el RMN son un contacto hidrofóbico y un pi catión (A:9 arginina - B:53 doble anillo aromático triptófano). Las intersecciones entre la DM 1 y el RMN tomando los clusters más poblados coincide con los resultados mencionados anteriormente. En el caso de la intersección de la DM 2, con sus clusters más poblados y el RMN se observa uno solo de los contactos hidrofóbicos observados anteriormente.

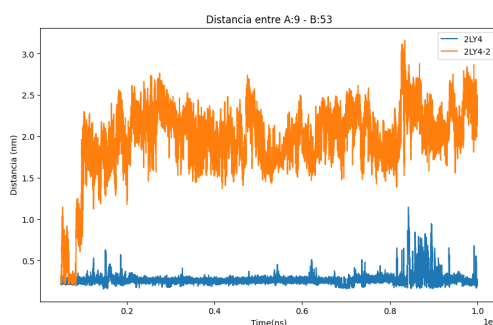


Cuadro 27: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters mas poblados (centro). A la derecha se encuentran los contactos reportados en la literatura. No se reportan interacciones coincidentes entre las dinámicas y RMN en ambos casos.

Se gráfico la distancia (Figura 37) entre ambos aminoácidos (A:9 arginina - B:53 doble anillo aromático triptófano) ya que al ser una interacción que podría ser del tipo core se buscaba demostrar por que no aparecía en la DM 2, y efectivamente en la DM 2 los átomos que forman esta interacción se separan.



(a)



(b)

Figura 37: a) Interacción Pi catión en rojo entre A:9 arginina - B:53 doble anillo aromático triptófano. b) Distancia entre aminoácidos A:9 y B:53 para DM 1 (azul) y DM 2 (naranja).

A partir de la bibliografía consultada sobre el complejo 2LY4 [29], se calcularon las distancias de las posibles interacciones detalladas en la misma. A modo de resumen en el Cuadro 28 se detallan los porcentajes totales en el que la distancia entre los aminoácidos permanece por debajo de 0,7 nm tanto para la DM 1 como

para la DM 2. Se observa que para ninguno de los casos la distancia entre aminoácidos es significativamente menor a 0,7 nm.

2LY4 — 2LY4-2		Cadena A			
		Arg 23	Lys 27	Lys 42	Lys 49
Cadena B	Ser 46	-	-	0 — 0	0 — 3 %
	Thr 55	20 % — 0	10 % — 0	-	-

Cuadro 28: Porcentaje de tiempo donde la distancia entre aminoácidos de ambas cadenas se encuentra $< 0,7$ nm para cada DM. Se toman los aminoácidos que interaccionan según la literatura para el complejo 2LY4

Según el análisis llevado a cabo, este complejo no presenta contactos relevantes, lo que indicaría su poca estabilidad o que este complejo es transiente en estado fisiológico. Si bien en el RMN y las dinámicas de los clusters más representativos muestran residuos en contacto, en ningún caso son los mismos, lo que permite hipotetizar que la interfaz de interacción de este complejo es muy lábil. No hay ningún par de residuos en contacto con probabilidad de estar en todos los modelos a lo largo del tiempo. Este es un caso excelente para demostrar la capacidad de este análisis para sugerir que este complejo será inestable o efímero.

7.1.8. 1MV0

Se obtuvieron los siguientes resultados de las dinámicas realizadas partiendo del confórmero 1 (1MV0) y el confórmero más lejano 8 (1MV0-8):

La comparación de las matrices de contacto entre el RMN y la DM del confórmero 1 y la del 2 dió un score de 0.53 y 0.53 respectivamente (Anexo 3). En este caso, las matrices de contacto no son similares.

Se observa en el Cuadro 29 tanto el RMSD como el RMSF de cada cadena para las DM realizadas. En ambos casos vemos que las dinámicas se comportan de forma esperada con altos grados de oscilación, al ser proteínas con regiones desordenadas. El comportamiento de ambas cadenas en ambas dinámicas es casi idéntico, tanto en el RMSD como en el perfil de RMSF.

PDB	Cadena (aa)	RMSD	RMSF
1MV0	A MB1 c-Myc (55-68)		
	B Bin1 (513-593)		

Cuadro 29: Representaciones gráficas del RMSD y RMSF para las DM 1MV0 y 1MV0-8. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.

Además, observamos en la Figura 38 y el Cuadro 30 la distribución de los conformeros en el tiempo por cluster para cada DM. En el caso de la DM 1 encontramos 7 clusters de los cuales solo en el primero los complejos se encuentran unidos, además se observa que el 93 % de las representaciones se encuentran en ese cluster. En la DM 2 sucede lo mismo, dentro de los 13 clusters solo el primero es en el que las cadena se encuentran unidas y en este caso el 70 % de las representaciones se encuentran en este cluster

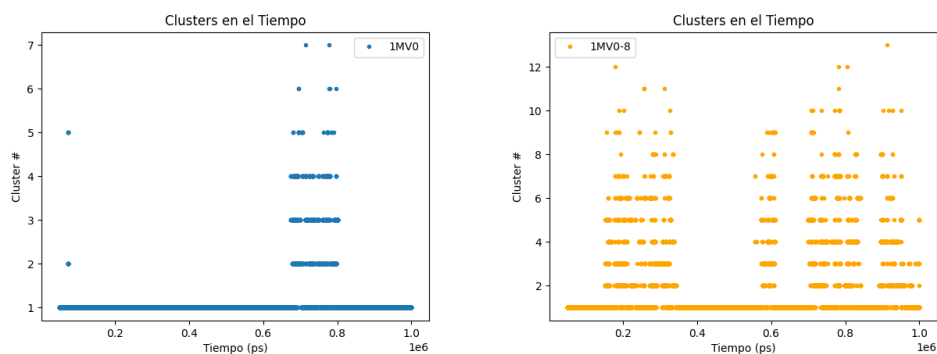


Figura 38: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 1MV0 y 1MV0-8.

1MV0			1MV0-8		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	17642	93 %	1	13387	70 %
2	680	4 %	2	2267	12 %
3	436	2 %	3	1199	6 %
4	203	1 %	4	1071	6 %
5	33	0 %	5	424	2 %
6	5	0 %	6	330	2 %
7	2	0 %	7	206	1 %
			8	53	0 %
			9	35	0 %
			10	20	0 %
			11	5	0 %
			12	3	0 %
			13	1	0 %
Total	19001			19001	

Cuadro 30: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Luego de realizado el análisis por cadena, se prosiguió analizando el complejo en su totalidad. Se encontró que para ambas dinámicas las cadenas se separan en el espacio a lo largo del tiempo de simulación. Como se puede observar en la Figura 39, la distancia entre ambas cadenas es tal que no existe interacción entre las mismas.

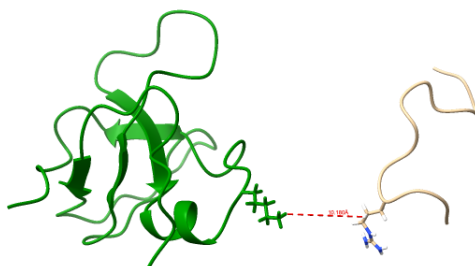


Figura 39: En amarillo la cadena A (proteína c-Myc), en verde se observa la cadena B (proteína Bin1) y en la línea punteada en rojo marca la distancia en armstrongs (Å) de los aminoácidos mas cercanos entre las cadenas.

Se observa en el Cuadro 31 diagramas de Venn con las interacciones de cada dinámica, del RMN y la coincidencia entre ellas. Tomando todos los clusters vemos que no se encuentran interacciones en las DMs aunque en el RMN si se reportan 22 interacciones. Mientras que tomando solo el cluster 1 de cada DM se encuentran 19 contactos en la DM 1 y 15 en la DM 2. Si bien no hay una intersección entre el RMN y ambas DM, se observan coincidencias entre cada una y el RMN por separado. Entre la DM 1 y el RMN se reportan 3 contactos hidrofóbicos (A:59-B:474, A:59-B:477, A:59-B:451), y entre la DM 2 y el RMN 2 puentes de hidrógeno (A:65-B:448, A:65-B:446), el primer puente de hidrógeno se encuentra reportado en la literatura.

De estos datos concluiríamos que no hay contactos “core” en este complejo. Sin embargo hay 4 contactos fundamentales que probablemente contengan las cadenas unidas en el complejo, que son los contactos que se conservan a lo largo del tiempo. Es decir los que se mantienen en las dos DM analizando los clusters más representativos y a su vez donde las cadenas se encuentran unidas; A:55-B:448, A:55-B:451, A:55-B:441 y A:55-B:472 (Figura 40). En este caso consideramos que el RMN al dar sólo 10 modelos, no permite conocer el comportamiento dinámico de la mayoría de los complejos a lo largo del tiempo. Decimos que los contactos del RMN no son los más duraderos y aunque haya contactos relevantes son diferentes.

Contactos DM y RMN	 1MV0 NMR RMN $\cap$ MD1 $\cap$ MD2: 0 MD1 $\cap$ MD2: 0 (Both MD showed structures with separated chains)
Contactos DM y RMN (cluster mas poblados)	 MD #1 1MV0 (Cluster 1) MD #2 1MV0 (Cluster 1) 1MV0 NMR RMN $\cap$ MD1 $\cap$ MD2: 0 MD1 $\cap$ MD2: 0 + 'A:55-B:448', 'A:55-B:451', 'A:55-B:441', 'A:55-B:472'
Contactos Literatura	A:62-B:420 A:62-B:448 A:62-B:451 A:62-B:472 A:65-B:448 A:65-B:418 A:65-B:451 A:65-B:477

Cuadro 31: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters mas poblados (centro). A la derecha se encuentran los contactos reportados en la literatura.



Figura 40: Representación en ribbon del complejo. En amarillo están representados en sticks los residuos en contacto en todas las DMs, RMN y la literatura. En violeta los contactos relevantes (DM1 y DM2 de los clusters más poblados)

7.1.9. 2PX9

Se obtuvieron los siguientes resultados de las dinámicas realizadas partiendo del conformero 1 (2PX9) y el conformero más lejano 19 (2PX9-19):

La comparación de las matrices de contacto entre el RMN y las DM dió un score de 0.8 para ambos casos (Anexo 3). Otra vez hay similitud entre las matrices de contacto del RMN y las dinámicas.

Se observa en el Cuadro 32 tanto el RMSD como el RMSF de cada cadena para las DM realizadas, en ambos casos vemos que las dinámicas se comportan de forma esperada con altos grados de oscilación al ser proteínas con regiones desordenadas. Tanto el RMSD como el RMSF tienen perfiles casi idénticos a lo largo de la DM.

PDB	Cadena (aa)	RMSD	RMSF
2PX9	A SAE2 (166-382)		
	B SUMO UBC9 (2-158)		

Cuadro 32: Representaciones gráficas del RMSD y RMSF para las DM 2PX9 y 2PX9-19. “En azul se muestran los resultados de la DM 1 y en naranja, los de la DM 2.”

Observamos en la Figura 41 y el Cuadro 33 la distribución de los conformeros en el tiempo por cluster para cada DM. Para ambas dinámicas vemos que el número total de clusters es alrededor de 30, siendo el primer cluster el más poblado con el 45 % de las representaciones totales para cada dinámica. Este resultado nos muestra que hay una gran diversidad de conformaciones, que da un número elevado de cluster y ninguno contiene la mayoría de los conformeros del complejo.

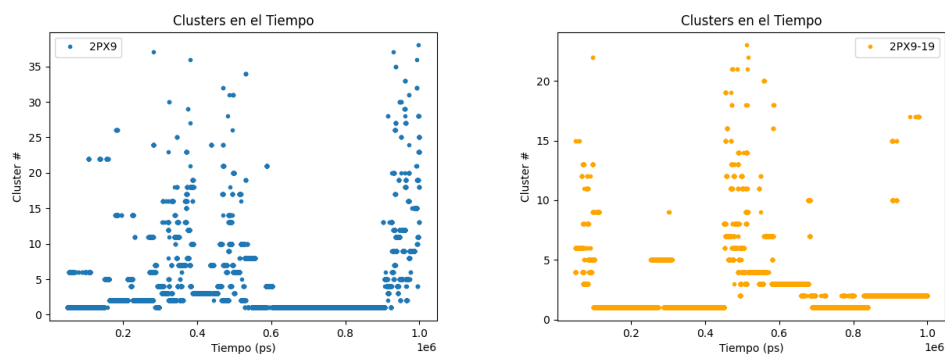


Figura 41: Distribución de conformeros a lo largo del tiempo de simulación para cada cluster de la dinámica 2PX9 y 2PX9-19.

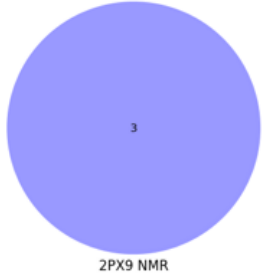
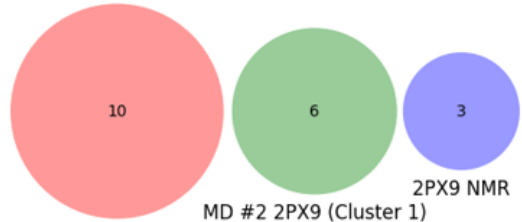
2PX9			2PX9-19		
Cluster	# Estructuras	Porcentaje	Cluster	# Estructuras	Porcentaje
1	8686	46 %	1	8898	47 %
2	2453	13 %	2	3691	19 %
3	1567	8 %	3	2103	11 %
4	1076	6 %	4	1234	6 %
5	807	4 %	5	1113	6 %
6	758	4 %	6	577	3 %
7	630	3 %	7	559	3 %
8	533	3 %	8	192	1 %
9	408	2 %	9	149	1 %
10	357	2 %	10	134	1 %
11	294	2 %	11	108	1 %
12	234	1 %	12	57	0 %
13	187	1 %	13	57	0 %
14	172	1 %	14	33	0 %
15	141	1 %	15	29	0 %
16	113	1 %	16	16	0 %
17	107	1 %	17	12	0 %
18	87	0 %	18	11	0 %
19	77	0 %	19	10	0 %
20	51	0 %	20	7	0 %
21	45	0 %	21	7	0 %
22	41	0 %	22	3	0 %
23	28	0 %	23	1	0 %
24	28	0 %			
25	27	0 %			
26	20	0 %			
27	17	0 %			
28	11	0 %			
29	8	0 %			
30	8	0 %			
31	7	0 %			
32	6	0 %			
33	5	0 %			
34	4	0 %			
35	3	0 %			
36	2	0 %			
37	2	0 %			
38	1	0 %			
Total	19001			19001	

Cuadro 33: Cantidad de estructuras asociadas a cada cluster para la DM 1 y DM 2.

Al analizar el Cuadro 34 vemos que los contactos observados tomando el RMN y las dos dinámicas completas dan como resultado 3 contactos para el RMN y ninguno para las DMs. Al analizar el cluster más poblado se observan algunos contactos en las DMs, pero hay que recordar que en cada dinámica el cluster más poblado no es representativo del comportamiento del complejo a lo largo del tiempo. Ninguno contiene un grupo representativo de estructuras (mayor al 75 %).

Frente a estos resultados se decidió analizar las interacciones solo para los cluster más poblados y a su vez donde las cadenas se encuentran unidas, de esta forma se evita perder contactos al solo trabajar con

representaciones del complejo donde las cadenas interactúan.

Contactos DM y RMN	 2PX9 NMR RMN $\cap$ MD1 $\cap$ MD2: 0 MD1 $\cap$ MD2: 0 (Both MD showed structures with separated chains)
Contactos DM y RMN (cluster mas poblados)	 MD #1 2PX9 (Cluster 1) MD #2 2PX9 (Cluster 1) 2PX9 NMR RMN $\cap$ MD1 $\cap$ MD2: 0 MD1 $\cap$ MD2: 0
Contactos Literatura	weak interactions

Cuadro 34: Diagramas de Venn con las interacciones de todos los clusters (izquierda) y tomando solo los clusters mas poblados (centro). A la derecha se encuentran los contactos reportados en la literatura. No se observan interacciones coincidentes en ninguno de los casos analizados.

Este resultado coincide con lo reportado en la literatura ya que son interacciones débiles y un complejo lábil [34]. Este es un resultado relevante porque al analizar su comportamiento dinámico vemos que el complejo permanecerá unido un tiempo muy corto, alineándonos con lo encontrado en la literatura que menciona que el complejo contiene interacciones débiles y transitorias. En este caso el RMN al no tener la característica temporal nos habría dado un resultado que no permite tener idea de este tipo de comportamiento característico. Por consiguiente, no se identifican interacciones de importancia para reportar.

7.2. Otros análisis propuestos

7.2.1. Gyrator

Con el objetivo de profundizar en el análisis y abrir la posibilidad de descubrir hallazgos significativos sobre las proteínas donde no se reportan interacciones coincidentes, se realizó un examen detallado de los gráficos de Gyrator con el fin de identificar algún patrón que diferencie estos complejos de los que sí se reportan interacciones coincidentes (Anexo 4). Sin embargo, no se observaron patrones consistentes que permitan llegar a conclusiones definitivas. Este hallazgo resalta la complejidad inherente al estudio de los complejos de proteínas desordenadas y la necesidad de emplear enfoques multifacéticos para comprender mejor su comportamiento y propiedades.

7.2.2. Prueba de concepto

Como prueba complementaria a lo expuesto anteriormente, se realizó una tercera dinámica molecular de los complejos 2L14 y 2M0G. Partiendo desde el primer modelo de RMN de los respectivos complejos, se realizó la simulación manteniendo todos los parámetros utilizados exceptuando la temperatura, la cual se aumentó de 300 K (23°C) a 333 K (60°C) para observar el comportamiento de las mismas en este estado.

2M0G: 2M0G: Para el caso del complejo 2M0G los resultados obtenidos fueron una conservación de 5 de los 7 contactos “core” que encontramos tanto en el RMN como en las dinámicas a 300 K. De estos contactos permanecieron el Pi-cation, entre los aminoácidos A:21-B:454, y el contacto iónico más doble puente de hidrógeno entre los aminoácidos A:49-B:462 que además esta reportado en la literatura.

2L14: Por otro lado, para el complejo 2L14 obtuvimos resultados diferentes. Se encontraron 2 contactos core. entre el RMN, la DM 1 y la DM 2 a 333K.

De todas maneras es un análisis que no se ha desarrollado en profundidad. El resultado de que hayan contactos conservados al someter a las proteínas a una temperatura tan extrema y alejada de las condiciones celulares, podría permitir concluir que ese complejo es muy estable.

8. Conclusiones

A través de la aplicación de dinámicas moleculares en los complejos proteicos formados por proteínas desordenadas, se ha descubierto que las simulaciones reflejan una alta flexibilidad y dinamismo, que concuerda con la naturaleza de las proteínas que poseen regiones desordenadas. Se han observado varios comportamientos distintos para los complejos:

- Complejos donde las cadenas interactúan entre sí durante toda la dinámica, y se encuentran contactos entre estas que coinciden con los reportados en la literatura y en el RMN
- Complejos donde las cadenas interactúan entre sí durante toda la dinámica, pero no se encuentran interacciones coincidentes con el RMN ni la literatura
- Complejos donde las cadenas interactúan entre sí pero de forma transitoria a lo largo del tiempo y no presentan interacciones permanentes al ser estructuras altamente dinámicas

Se ha observado que la dinámica molecular revela contactos “core” y otros contactos fundamentales para la cohesión del complejo cuando es posible. Estos contactos pueden ser distintos a los observados en el RMN ya que este proporciona un número limitado de complejos y cada uno de ellos es equiprobable, es decir tienen el mismo peso sin noción de si son representativo del conjunto de conformaciones a lo largo del tiempo. El análisis de las dinámicas completas también padece un problema similar al RMN. Dado que se considera todos los clusters de complejos, sin incluir la importante noción de que algunos son muy representativos del mismo y por el contrario otros contienen conformaciones muy improbables, que aparecen un periodo ínfimo del tiempo explorado en la dinámica.

La implementación de la clusterización y consideración del comportamiento de los clusters más representativos a lo largo del tiempo, ha incluido una dimensión extra al análisis ya que permite hipotetizar que hay un conjunto menor de estructuras que es representativo de lo que ocurrirá en el contexto celular donde las proteínas están en continuo movimiento en un ambiente líquido.

El análisis ha permitido también observar separaciones y comportamientos que el RMN no logra identificar, debido a las limitaciones temporales de esta técnica experimental. De esta forma los resultados de RMN se complementan por las simulaciones de dinámica molecular.

Por otro lado, el enfoque de la evaluación de los clusters más poblados en las dinámicas ha proporcionado un panorama detallado de las interacciones intermoleculares que se pierden al tomar todos los clusters en casos de complejos altamente dinámicos. Esto subraya la importancia de los distintos abordajes a la hora de analizar las dinámicas y de cada complejo en particular.

La consistencia entre las simulaciones de DM y los datos de RMN, junto con las confirmaciones de las interacciones reportadas en la literatura, refuerza la validez de nuestros métodos y subraya el valor de la dinámica molecular como un complemento indispensable en la investigación de las interacciones entre proteínas desordenadas. Este enfoque multidimensional no solo amplía nuestro conocimiento sobre la función y estructura de este tipo particular de complejos proteicos sino que también es crucial para el diseño de futuras intervenciones terapéuticas y para la comprensión detallada de los mecanismos moleculares que rigen la biología de las proteínas.

Por tanto, los resultados de este estudio enfatizan el valor de un enfoque integrador en la investigación científica. La implementación de métodos complementarios, como la dinámica molecular y el RMN, amplía significativamente el espectro de datos disponibles, facilitando así un entendimiento más integral y holístico de las interacciones intermoleculares y la conformación de los complejos proteicos. La sinergia entre estas técnicas ha demostrado ser esencial para superar las limitaciones inherentes a cada método por separado.

A pesar de los avances significativos presentados, este trabajo no está exento de limitaciones. La generalización de los resultados puede estar condicionada por el alcance específico de los complejos analizados, y se reconoce la necesidad de extender la investigación a una gama más amplia de complejos para validar aún más la aplicabilidad de estos enfoques combinados. Se ha hecho evidente la necesidad de incluir en los análisis futuros un espectro más amplio de conformaciones, y dinámicas realizadas partiendo de otras condiciones iniciales para obtener resultados extendidos que puedan llevar a una comprensión más profunda de la dinámica proteica y del complejo estudiado. En futuras investigaciones, sería provechoso complementar los resultados con el análisis de las interacciones y su rol en la funcionalidad específica del complejo. Con esto en mente, el presente estudio no solo contribuye a la base de conocimientos existente sino que también establece un precedente para la investigación futura, subrayando la importancia de la colaboración interdisciplinaria en el avance de la ciencia molecular y el desarrollo de posibles dianas terapéuticas.

9. Referencias

- [1] Goddard-Borger ED, Carapito R, Jeltsch JM, Phalip V, Stick RV, Varrot A. -L-Arabinofuranosylated pyrrolidines as arabinanase inhibitors. *Chem Commun (Camb)*. 2011 Sep 14;47(34):9684-6. doi: 10.1039/c1cc13675e. Epub 2011 Jul 20. PMID: 21773614.
- [2] Sillitoe I, Dawson N, Thornton J, Orengo C. The history of the CATH structural classification of protein domains. *Biochimie*. 2015 Dec;119:209-17. doi: 10.1016/j.biochi.2015.08.004. Epub 2015 Aug 4. PMID: 26253692; PMCID: PMC4678953.
- [3] Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC. Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev*. 2014 Jul 9;114(13):6844-79. doi: 10.1021/cr400713r. Epub 2014 May 15. PMID: 24830552; PMCID: PMC4100540.
- [4] Kufareva I, Abagyan R. Methods of protein structure comparison. *Methods Mol Biol*. 2012;857:231-57. doi: 10.1007/978-1-61779-588-6_10. PMID : 22323224; PMCID : PMC4321859.
- [5] Choudhary S, Lopus M, Hosur RV. Targeting disorders in unstructured and structured proteins in various diseases. *Biophys Chem*. 2022 Feb;281:106742. doi: 10.1016/j.bpc.2021.106742. Epub 2021 Dec 11. PMID: 34922214.
- [6] Krishnarajuna B, Ivanova MI, Ramamoorthy A. Aggregation and the Intrinsic Structural Disorder of Di-peptide Repeat Peptides of C9orf72-Related Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Characterized by NMR. *J Phys Chem B*. 2021 Nov 18;125(45):12446-12456. doi: 10.1021/acs.jpcc.1c08149. Epub 2021 Nov 9. PMID: 34751579.
- [7] Zou H, Pan T, Gao Y, Chen R, Li S, Guo J, Tian Z, Xu G, Xu J, Ma Y, Li Y. Pan-cancer assessment of mutational landscape in intrinsically disordered hotspots reveals potential driver genes. *Nucleic Acids Res*. 2022 May 20;50(9):e49. doi: 10.1093/nar/gkac028. PMID: 35061901; PMCID: PMC9122534.
- [8] Smyth MS, Martin JH. x ray crystallography. *Mol Pathol*. 2000 Feb;53(1):8-14. doi: 10.1136/mp.53.1.8. PMID: 10884915; PMCID: PMC1186895.
- [9] Quentin D, Raunser S. Electron cryomicroscopy as a powerful tool in biomedical research. *J Mol Med (Berl)*. 2018 Jun;96(6):483-493. doi: 10.1007/s00109-018-1640-y. Epub 2018 May 5. PMID: 29730699; PMCID: PMC5988769.
- [10] Yunfei Hu, Kai Cheng, Lichun He, Xu Zhang, Bin Jiang, Ling Jiang, Conggang Li, Guan Wang, Yunhuang Yang, and Maili Liu NMR-Based Methods for Protein Analysis *Analytical Chemistry* 2021 93 (4), 1866-1879 DOI: 10.1021/acs.analchem.0c03830
- [11] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank (2000) *Nucleic Acids Research* 28: 235-242 <https://doi.org/10.1093/nar/28.1.235>.
- [12] Clementel D, Del Conte A, Monzon AM, Camagni GF, Minervini G, Piovesan D, Tosatto SCE. RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res*. 2022 50(W1):W651-W656. from <https://ring.biocomputingup.it/submit>
- [13] Andras Hatos, Alexander Miguel Monzon, Silvio C E Tosatto, Damiano Piovesan, Monika Fuxreiter (2021) FuzDB: a new phase in understanding fuzzy interactions. *Nucleic Acids Research*. DOI: 10.1093/nar/gkab1060
- [14] Piovesan D, Del Conte A, Clementel D, Monzon AM, Bevilacqua M, Aspromonte MC, Iserte JA, Orti FE, Marino-Buslje C, Tosatto SCE. MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res*. 2023 Jan 6;51(D1):D438-D444. doi: 10.1093/nar/gkac1065. PMID: 36416266; PMCID: PMC9825420.
- [15] The UniProt Consortium , UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, Volume 51, Issue D1, 6 January 2023, Pages D523–D531, <https://doi.org/10.1093/nar/gkac1052>

- [16] Di Lello P, Jenkins LMM, Jones TN, Nguyen BD, Hara T, Yamaguchi H, Dikeakos JD, Appella E, Legault P, Omichinski JG. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIF and the activation domain of p53. *Mol Cell*. 2006 Jun 23;22(6):731-740. doi: 10.1016/j.molcel.2006.05.007. PMID: 16793543.
- [17] UniProt Consortium. (2023). UniProt entry P04637 [Tumor protein p53]. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P04637/entry>
- [18] UniProt Consortium. (2023). UniProt entry P32776. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P32776/entry>
- [19] Zhang Y, Madl T, Bagdiul I, Kern T, Kang HS, Zou P, Mäusbacher N, Sieber SA, Krämer A, Sattler M. Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3'-splice site recognition. *Nucleic Acids Res*. 2013 Jan;41(2):1343-54. doi: 10.1093/nar/gks1097. Epub 2012 Nov 21. PMID: 23175611; PMCID: PMC3553976.
- [20] UniProt Consortium. (2023). UniProt entry Q15637. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/Q15637/entry>
- [21] UniProt Consortium. (2023). UniProt entry P26368. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P26368/entry>
- [22] Lee CW, Martinez-Yamout MA, Dyson HJ, Wright PE. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry*. 2010 Nov 23;49(46):9964-71. doi: 10.1021/bi1012996. Epub 2010 Oct 29. PMID: 20961098; PMCID: PMC2982890.
- [23] UniProt Consortium. (2023). UniProt entry P45481. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P45481/entry>
- [24] Brzovic PS, Heikaus CC, Kisselev L, Vernon R, Herbig E, Pacheco D, Warfield L, Littlefield P, Baker D, Klevit RE, Hahn S. The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol Cell*. 2011 Dec 23;44(6):942-53. doi: 10.1016/j.molcel.2011.11.008. PMID: 22195967; PMCID: PMC3246216.
- [25] UniProt Consortium. (2023). UniProt entry P19659. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P19659/entry>
- [26] UniProt Consortium. (2023). UniProt entry P03069. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P03069/entry>
- [27] Lecoq L, Raiola L, Chabot PR, Cyr N, Arseneault G, Legault P, Omichinski JG. Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF- κ B and transcription regulatory factors. *Nucleic Acids Res*. 2017 May 19;45(9):5564-5576. doi: 10.1093/nar/gkx146. PMID: 28334776; PMCID: PMC5435986.
- [28] UniProt Consortium. (2023). UniProt entry Q04206. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/Q04206/entry>
- [29] Rowell JP, Simpson KL, Stott K, Watson M, Thomas JO. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. *Structure*. 2012 Dec 5;20(12):2014-24. doi: 10.1016/j.str.2012.09.004. Epub 2012 Oct 11. PMID: 23063560.
- [30] UniProt Consortium. (2023). UniProt entry P09429. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P09429/entry>
- [31] Pineda-Lucena A, Ho CS, Mao DY, Sheng Y, Laister RC, Muhandiram R, Lu Y, Seet BT, Katz S, Szyperki T, Penn LZ, Arrowsmith CH. A structure-based model of the c-Myc/Bin1 protein interaction shows alternative splicing of Bin1 and c-Myc phosphorylation are key binding determinants. *J Mol Biol*. 2005 Aug 5;351(1):182-94. doi: 10.1016/j.jmb.2005.05.046. PMID: 15992821.

- [32] UniProt Consortium. (2023). UniProt entry P01106. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P01106/entry>
- [33] UniProt Consortium. (2023). UniProt entry O0049. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/O0049/entry>
- [34] Wang J, Hu W, Cai S, Lee B, Song J, Chen Y. The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. *Mol Cell*. 2007 Jul 20;27(2):228-237. doi: 10.1016/j.molcel.2007.05.023. PMID: 17643372; PMCID: PMC1978417.
- [35] UniProt Consortium. (2023). UniProt entry Q9UBT2. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/Q9UBT2/entry>
- [36] UniProt Consortium. (2023). UniProt entry P63279. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P63279/entry>
- [37] Song J, Zhang Z, Hu W, Chen Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J Biol Chem*. 2005 Dec 2;280(48):40122-9. doi: 10.1074/jbc.M507059200. Epub 2005 Oct 3. PMID: 16204249.
- [38] UniProt Consortium. (2023). UniProt entry P63165 [Protein Name]. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/P63165/entry>
- [39] UniProt Consortium. (2023). UniProt entry O75928. Retrieved November 17, 2023, from <https://www.uniprot.org/uniprotkb/O75928/entry>
- [40] M. Abraham, B. Hess, D. van der Spoel, E. Lindahl, “GROMACS - Groningen Machine for Chemical Simulations: User Manual 2023.3”, 2023.
- [41] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-18. from <https://www.gromacs.org/>
- [42] Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M, de Groot, B.L., Grubmuller, H., and MacKerell, A.D., Jr., CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins,” *Nature Methods*, 14:71-73, 2016, PMC5199616.https://mackerell1.umaryland.edu/charmm_ff.shtml#gromacs.
- [43] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, 79 926–935 (1983).
- [44] R.W. Hockney, S.P. Goel, and J. Eastwood, “Quiet High Resolution Computer Models of a Plasma,” *J. Comp. Phys.*, 14 148–158 (1974)
- [45] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, 126 014101 (2007)
- [46] H.J.C. Berendsen, J.P.M. Postma, A. DiNola, and J.R. Haak, “Molecular dynamics with coupling to an external bath,” *J. Chem. Phys.*, 81 3684–3690 (1984)
- [47] Essman, U. Perera, L. Berkowitz, M. L. Darden, H. L. T. Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995, 103, 85778592.
- [48] MJ Abraham, T Murtola, R Schulz, S Páll, JC Smith, B Hess, E Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, Volumes 1–2, pp:19-25
- [49] Python Software Foundation. (2023). Python Language Reference, version 3.10. [Software]. Available from <https://www.python.org>

- [50] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Židek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, Sameer Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, *Nucleic Acids Research*, Volume 50, Issue D1, 7 January 2022, Pages D439–D444, <https://doi.org/10.1093/nar/gkab1061>
- [51] Goddard, T. D., Huang, C. C., & Ferrin, T. E. (2023). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. [Software]. Available from <https://www.rbvi.ucsf.edu/chimerax/>

10. Anexo

10.1. Anexo 1: Parametros RING

Los parámetros establecidos como distance threshold para cada tipo de interacción son los siguientes:

- Hydrogen Bonds: 5.5A
- Ionic Bonds: 5A
- Pi-Cation bonds: 7A
- Van der Walls: 1,46A
- Pi-Pi Stack Bond: 7A
- Disulphide bond: 3A

10.2. Anexo 2: Matrices de Contacto

2M0G

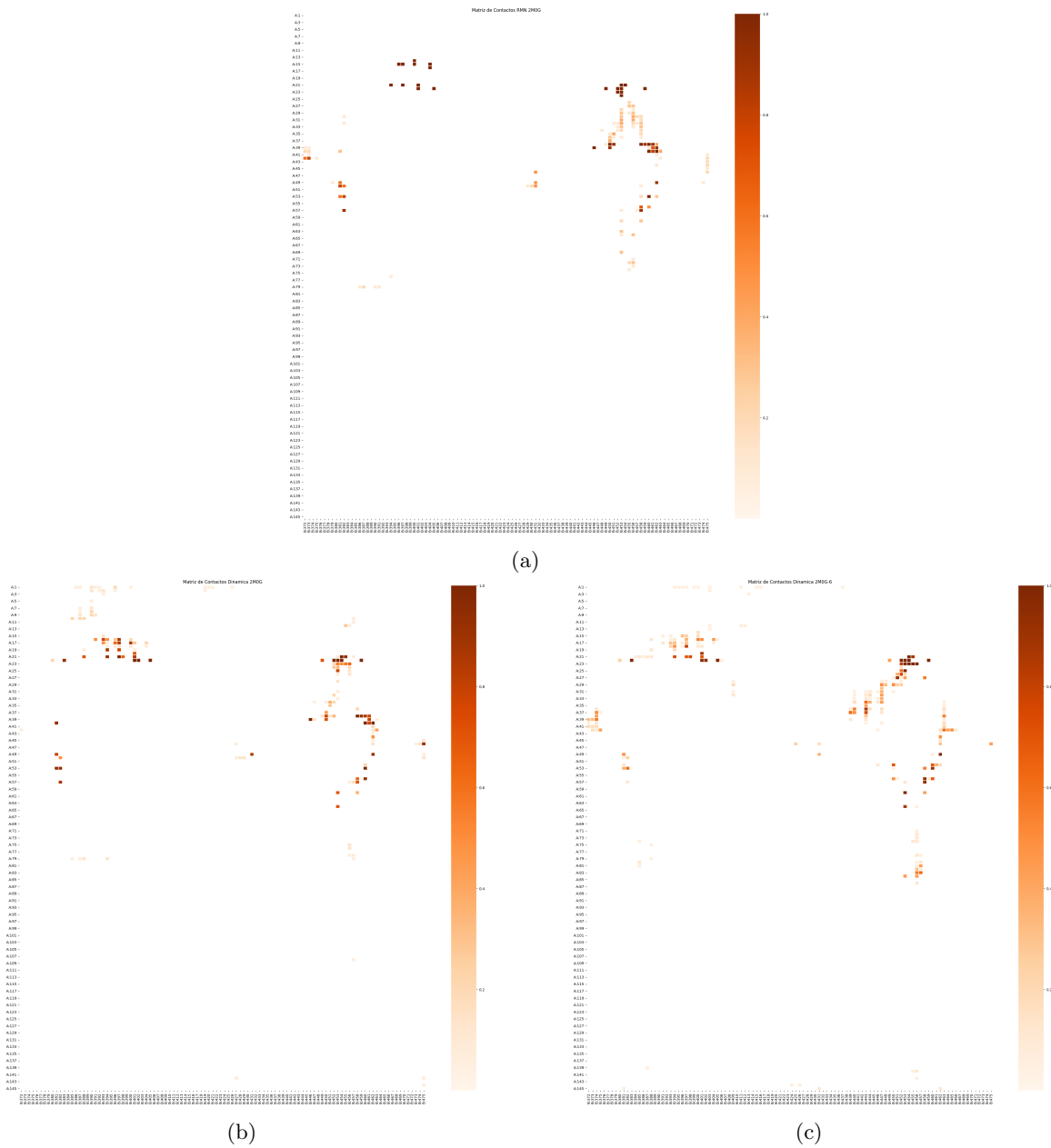


Figura 42: Matriz de contacto del: a) RMN, b) DM1, c) DM2

5URN

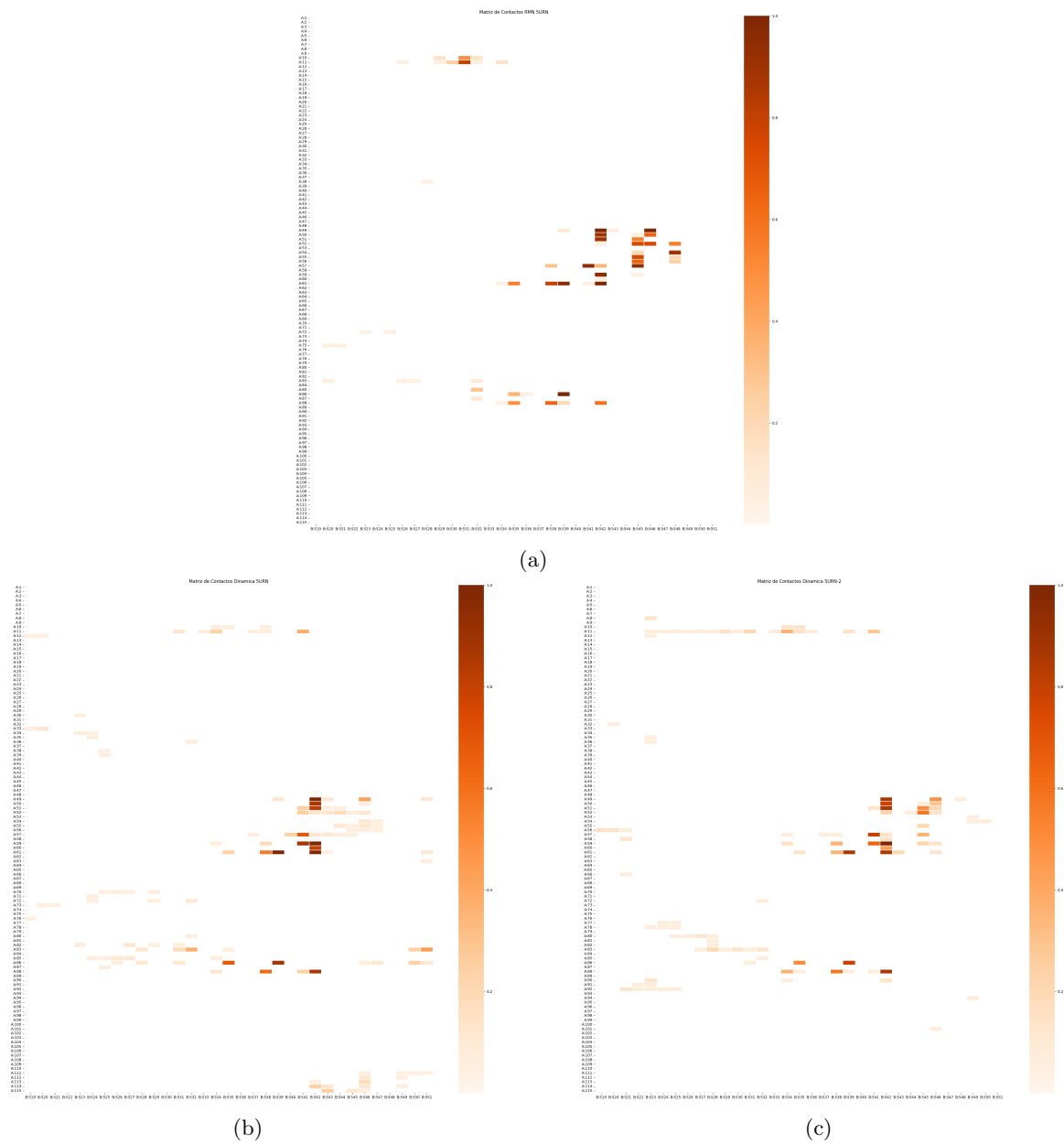


Figura 44: Matriz de contacto del: a) RMN, b) DM1, c) DM2

2ASQ

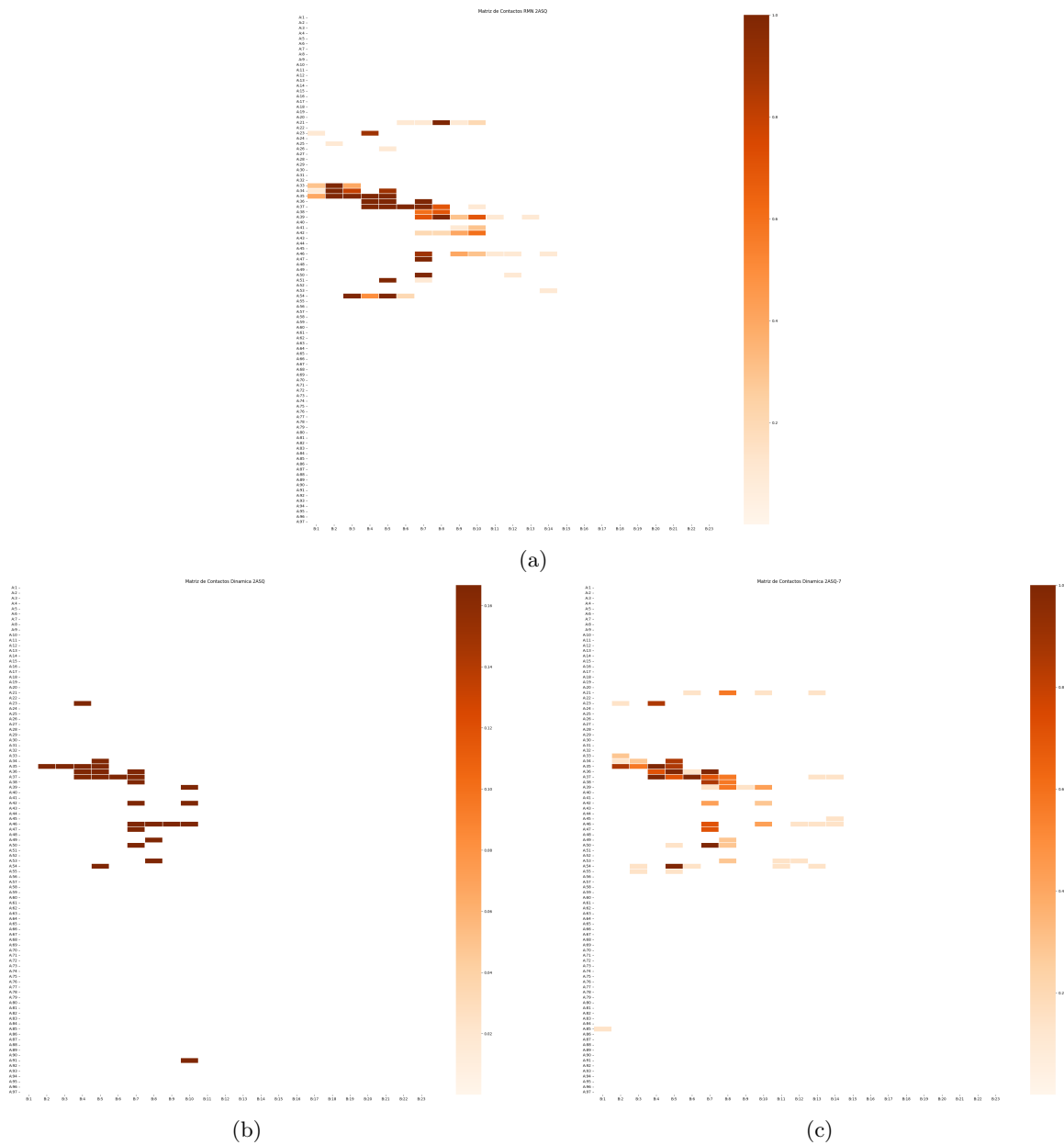


Figura 45: Matriz de contacto del: a) RMN, b) DM1, c) DM2

2LPB

Figura 46: Matriz de contacto del: a) RMN, b) DM1, c) DM2

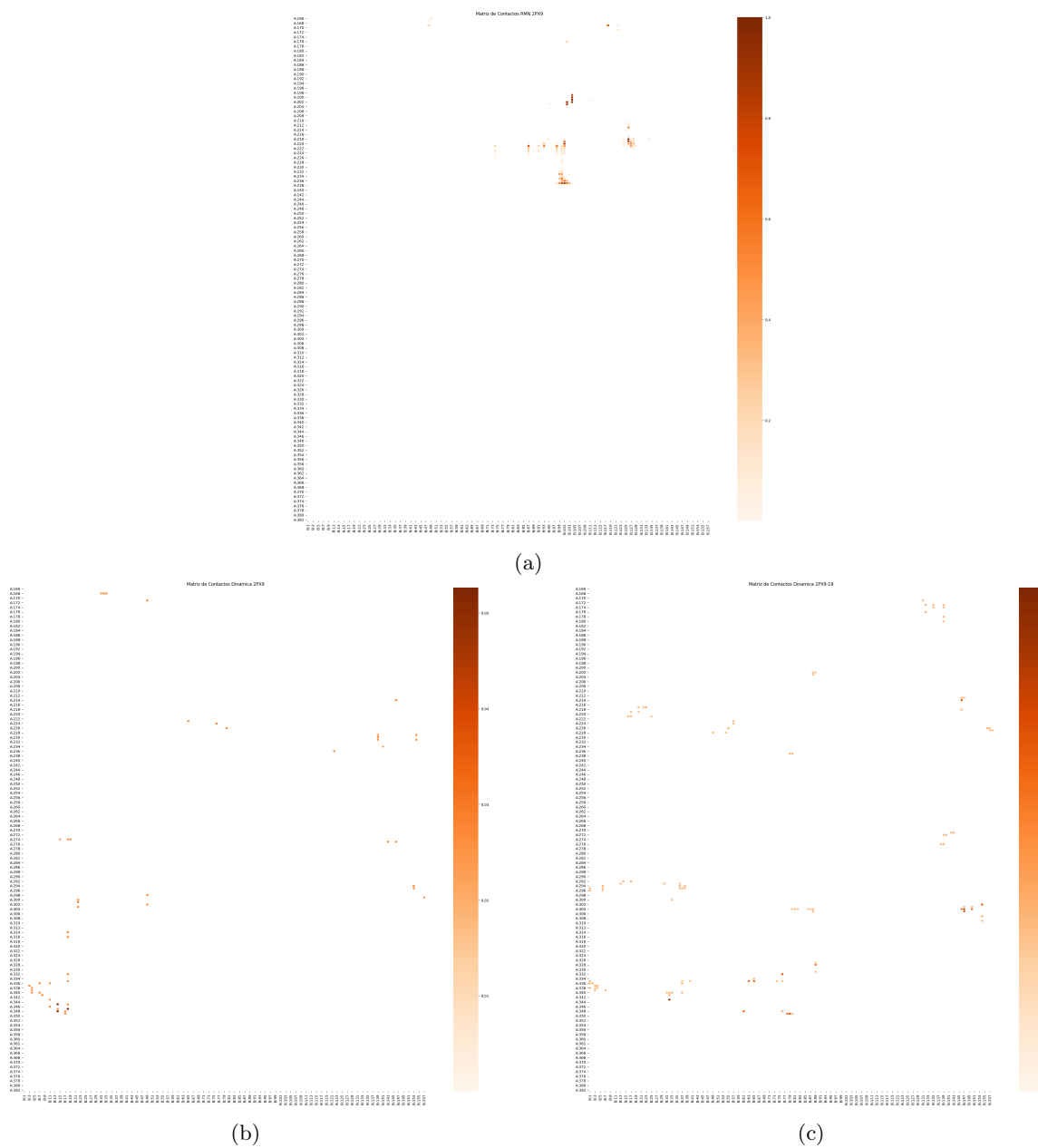
2LY4

Figura 47: Matriz de contacto del: a) RMN, b) DM1, c) DM2

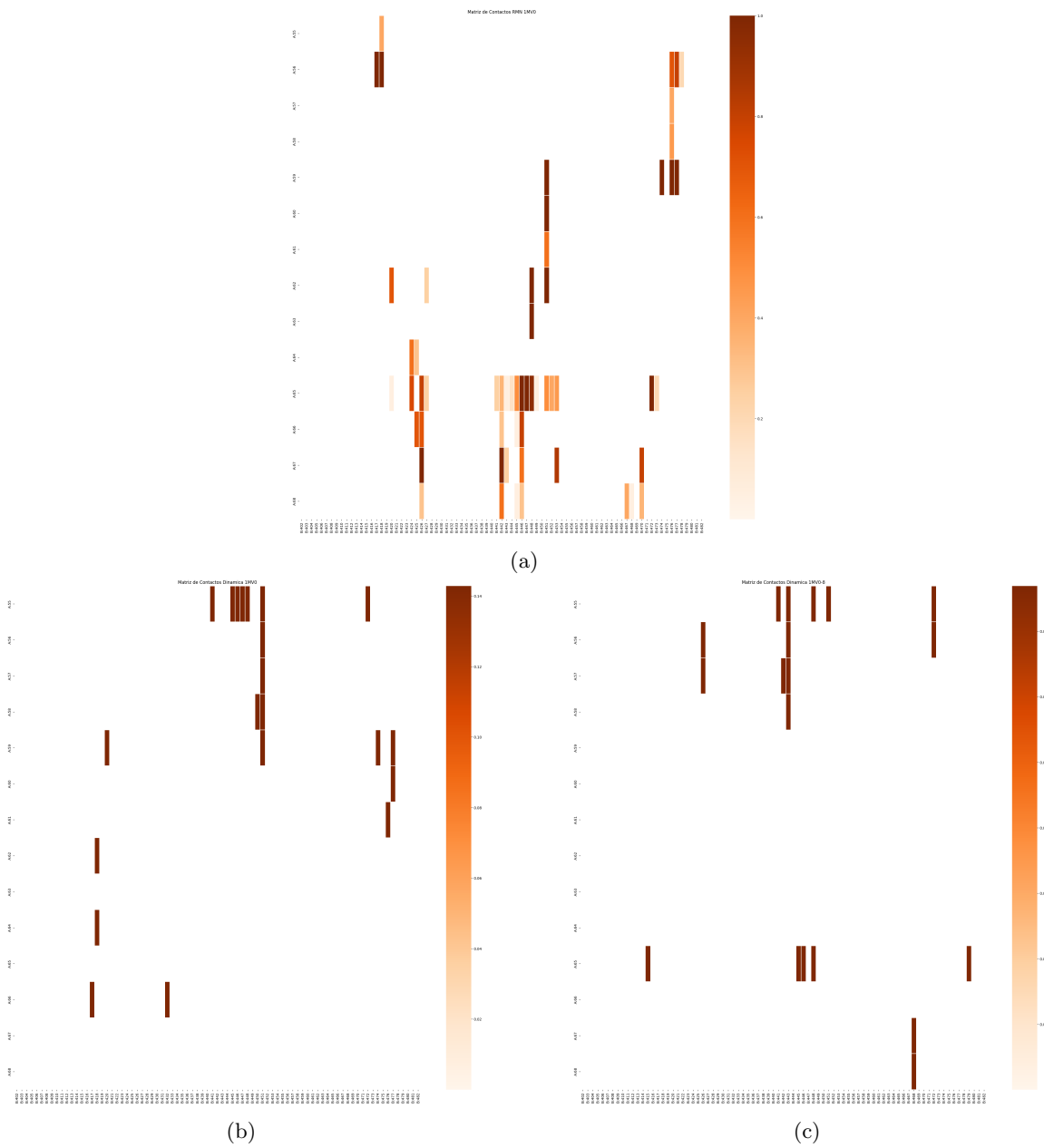
1MV0

Figura 48: Matriz de contacto del: a) RMN, b) DM1, c) DM2

Mapa de Contactos (Omniscia - 27/03)

Mapa de Contactos (Omniscia - 27/03)

Matriz de Contactos Dinamica 2020-19

10.3. Anexo 3: Heatmap de Comparación

2M0G

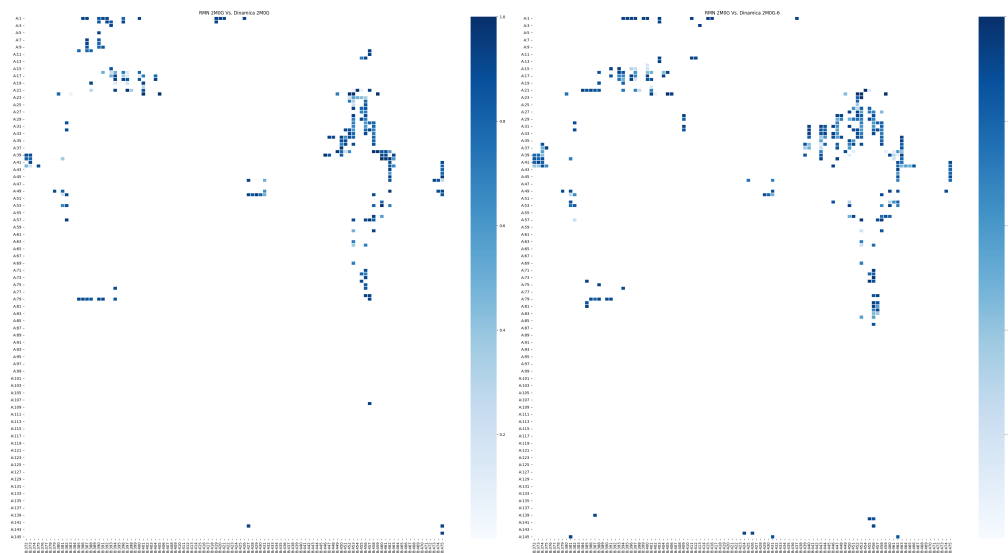


Figura 50: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

2L14

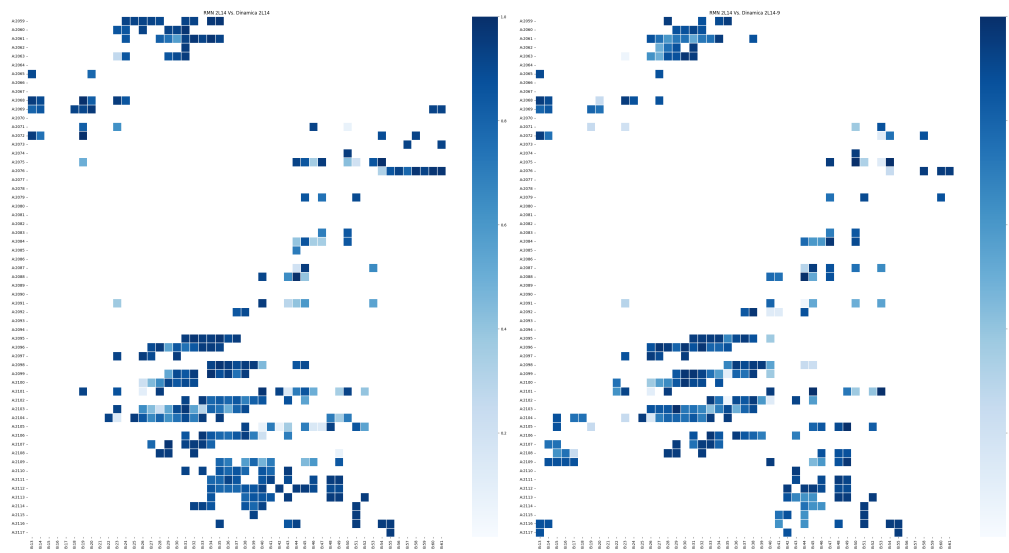


Figura 51: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

5URN

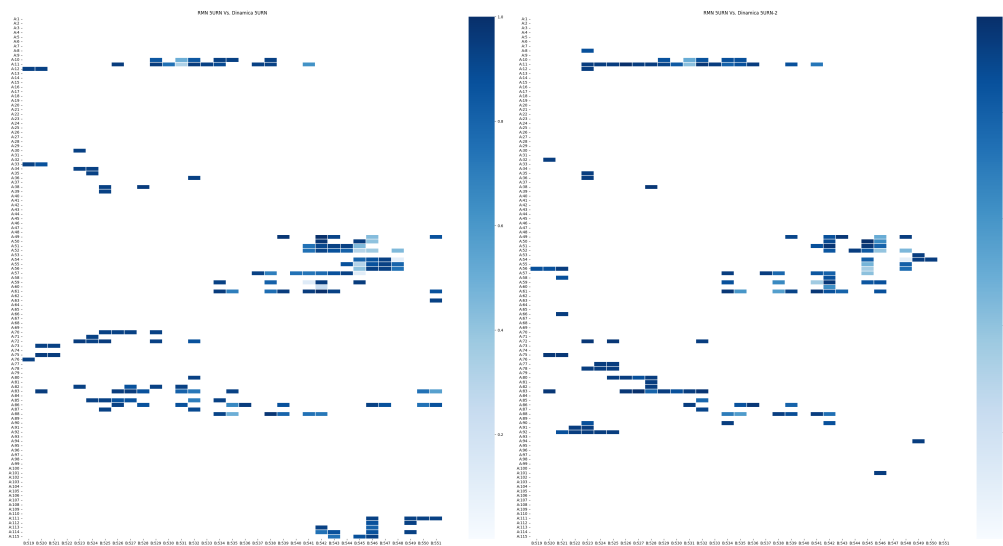


Figura 52: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

2ASQ

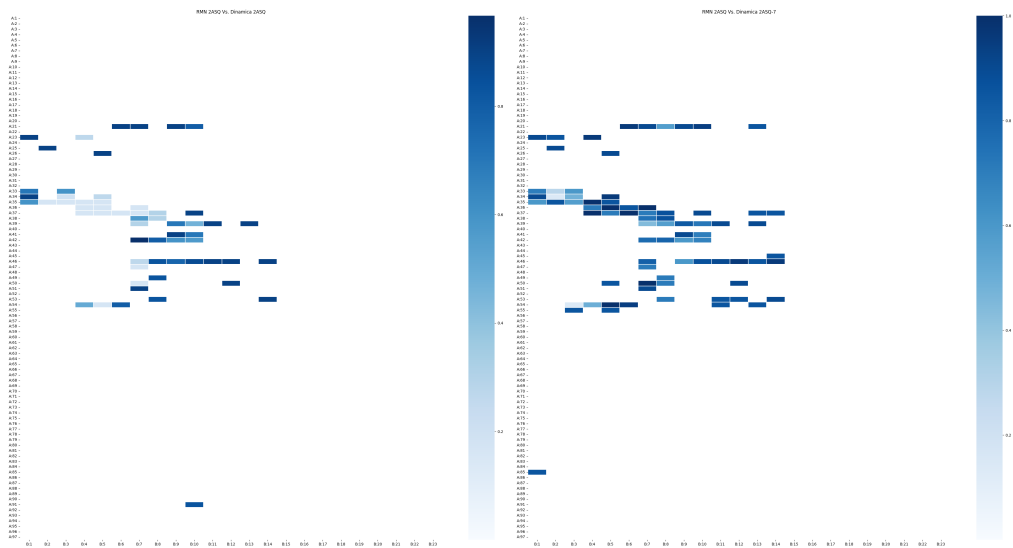


Figura 53: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

2LPB

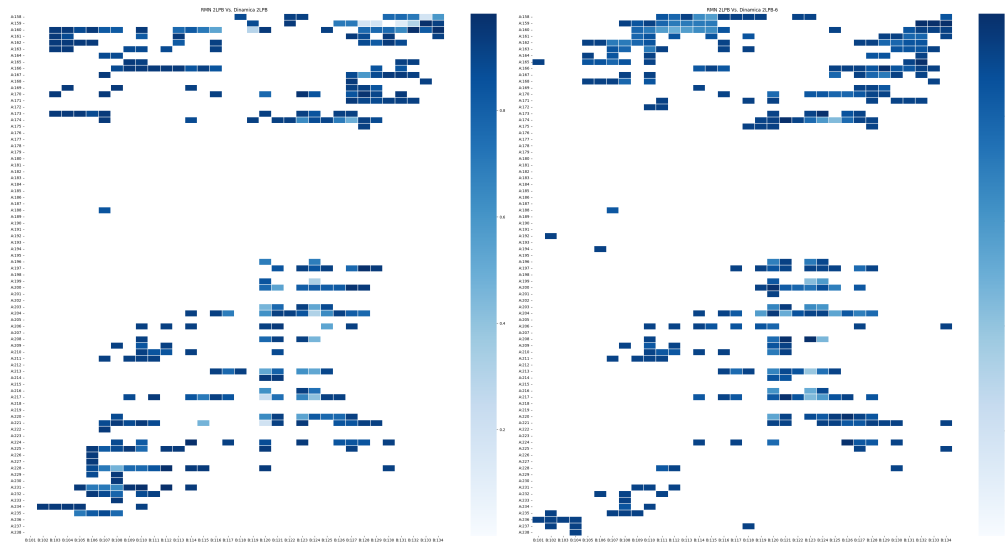


Figura 54: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

2LY4

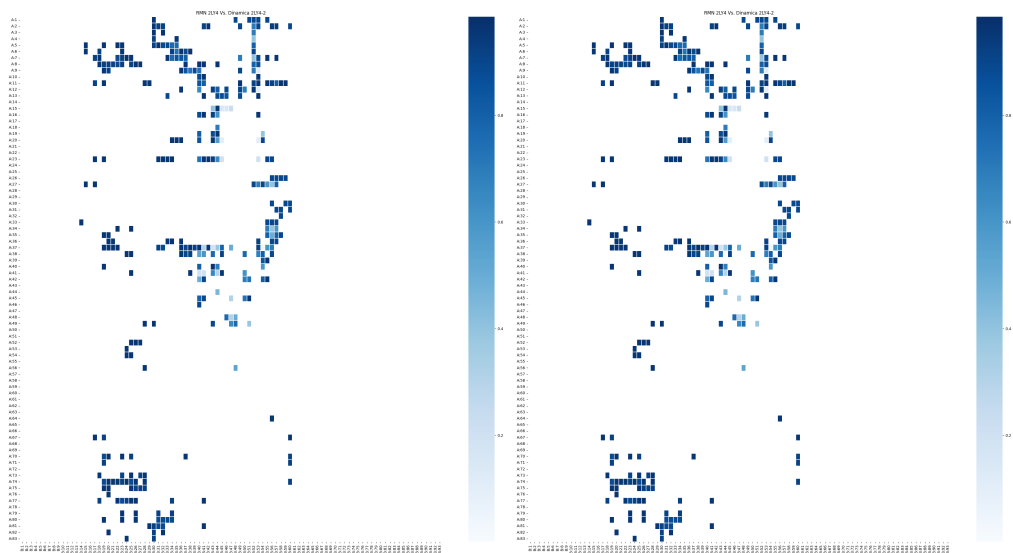


Figura 55: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

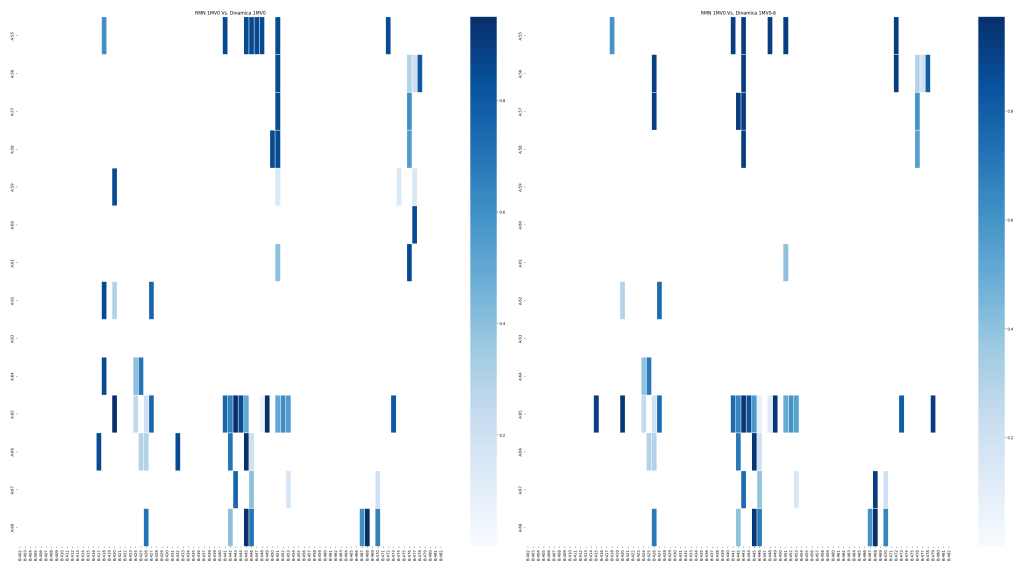
1MV0

Figura 56: Heatmap de comparación entre a) RMN - DM1 b) RMN - DM2. .

Figure 1 displays two heatmaps showing the distribution of 1000 randomised trials across 1000 randomised trials. The left heatmap is titled "1000 (2015 to 2020) (Dynamics 2015)" and the right heatmap is titled "1000 (2015 to 2020) (Dynamics 2015)". Both heatmaps show a color scale from 0.00 to 0.10, with a vertical color bar on the left of each heatmap indicating the scale. The heatmaps show a distribution of trials across 1000 randomised trials, with a color scale from 0.00 to 0.10. The left heatmap shows a distribution of trials across 1000 randomised trials, with a color scale from 0.00 to 0.10. The right heatmap shows a distribution of trials across 1000 randomised trials, with a color scale from 0.00 to 0.10.

91

10.4. Anexo 4: Gráficos Gyrate

2GS0

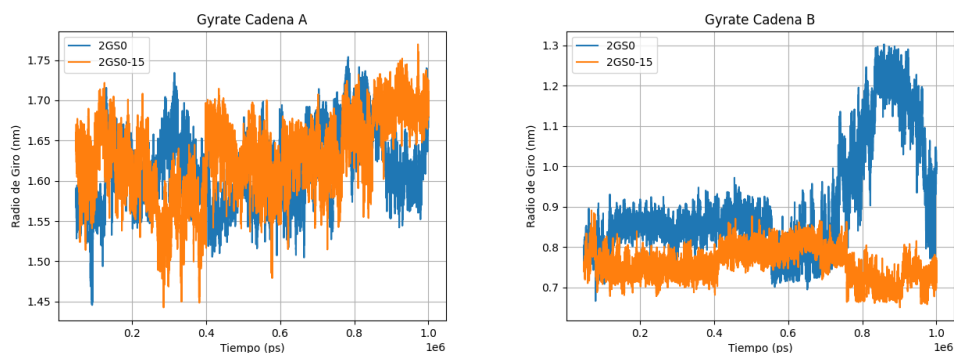


Figura 58: Cadena A y B del complejo 2GS0, en azul DM 1 y en naranja DM 2.

2M0G

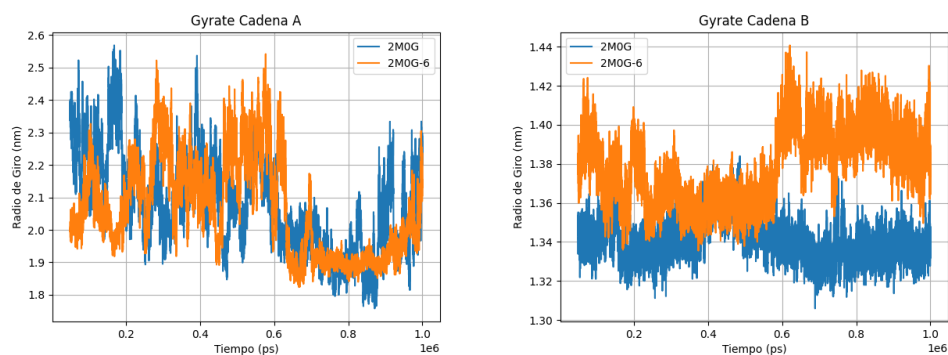


Figura 59: Cadena A y B del complejo 2M0G, en azul DM 1 y en naranja DM 2.

2L14

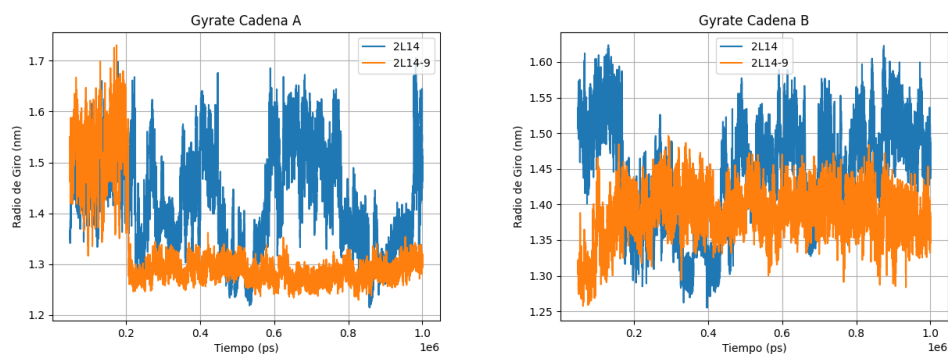


Figura 60: Cadena A y B del complejo 2L14, en azul DM 1 y en naranja DM 2.

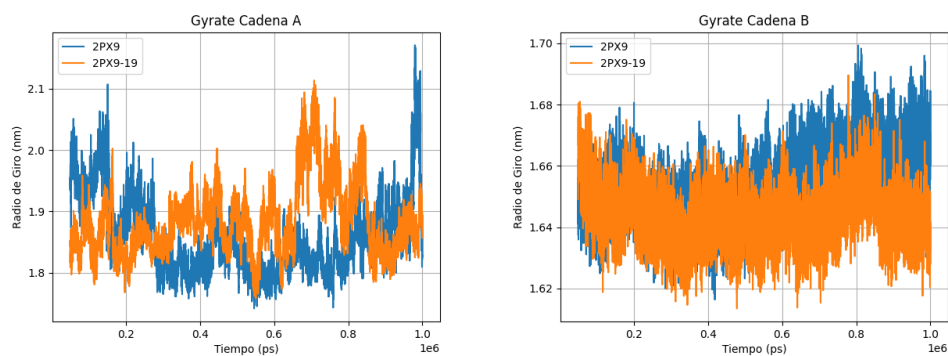
2PX9

Figura 61: Cadena A y B del complejo 2PX9, en azul DM 1 y en naranja DM 2.

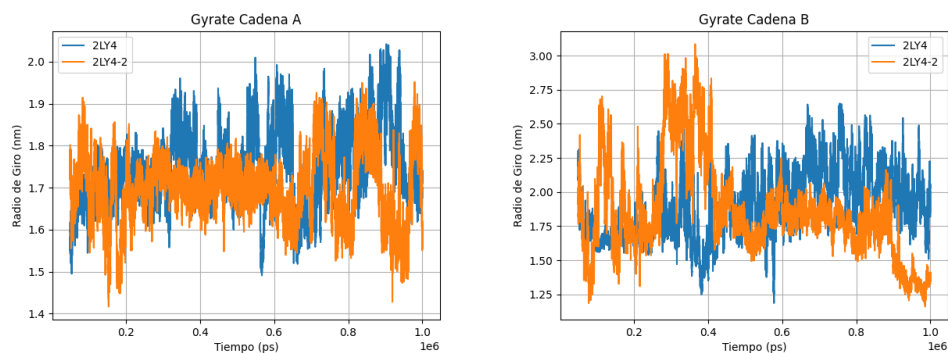
2LY4

Figura 62: Cadena A y B del complejo 2LY4, en azul DM 1 y en naranja DM 2.

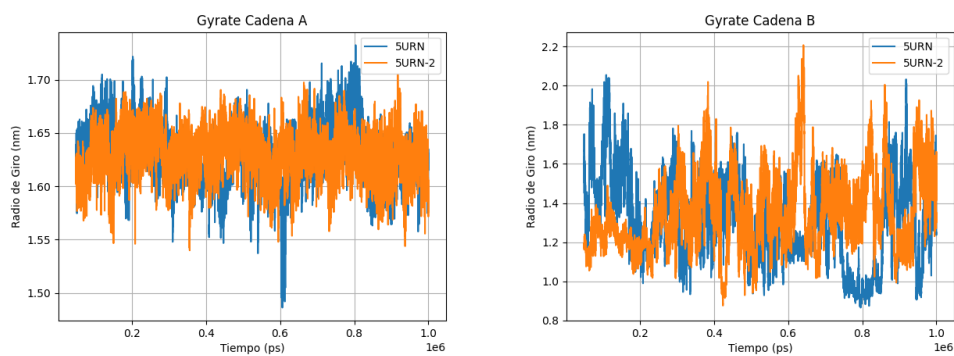
5URN

Figura 63: Cadena A y B del complejo 5URN, en azul DM 1 y en naranja DM 2.

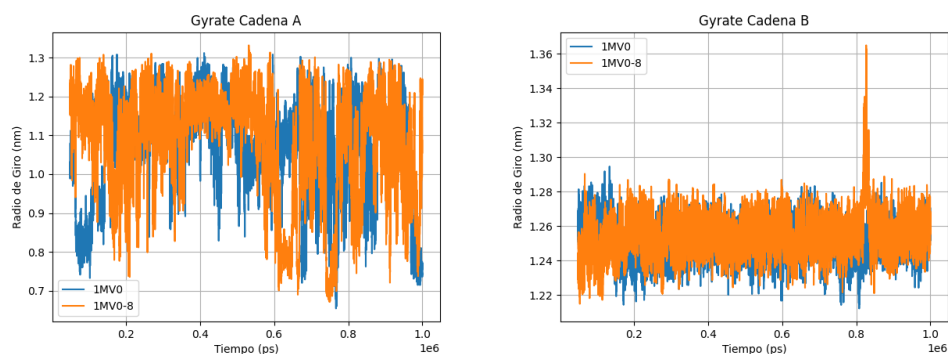
1MV0

Figura 64: Cadena A y B del complejo 1MV0, en azul DM 1 y en naranja DM 2.

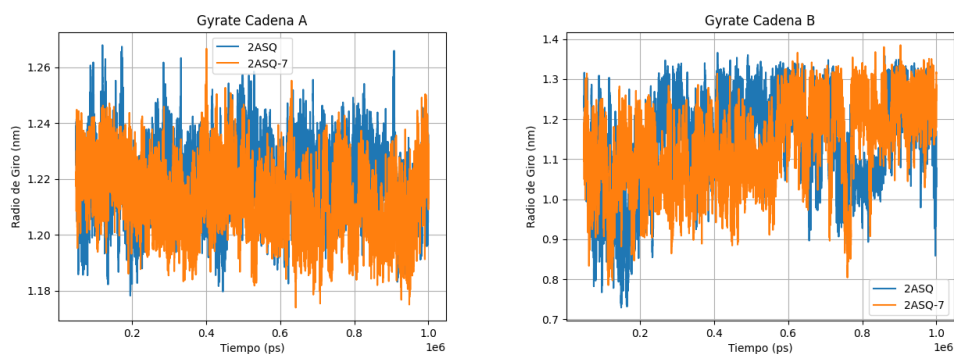
2ASQ

Figura 65: Cadena A y B del complejo 2ASQ, en azul DM 1 y en naranja DM 2.

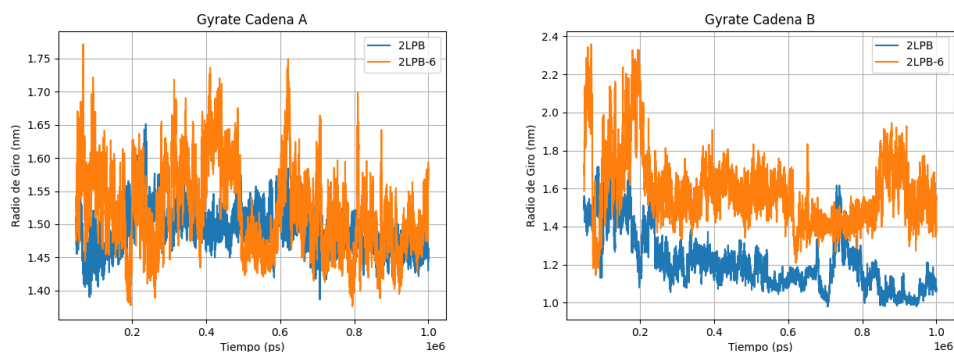
2LPB

Figura 66: Cadena A y B del complejo 2LPB, en azul DM 1 y en naranja DM 2.