



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Maestría y Doctorado en Ciencias Bioquímicas**

ANÁLISIS ESTRUCTURAL DE LA PROTEÍNA AKR1B1 PARA LA  
IDENTIFICACIÓN DE DOMINIOS DE INTERACCIÓN PROTEÍNA-PROTEÍNA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestro en Ciencias

PRESENTA:

Q.F.B. FRANCISCO JAVIER GÓMEZ ZALDÍVAR

TUTOR

Dr. MARCO ANTONIO VELASCO VELÁZQUEZ  
[Facultad de Medicina](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. ALEJANDRO SOSA PEINADO  
[Facultad de Medicina](#)

Dr. ROGELIO RODRÍGUEZ SOTRES  
[Facultad de Química](#)

Ciudad de México. Mayo, 2022



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **AGRADECIMIENTOS**

A la Universidad Nacional Autónoma de México que, a través de su Programa de Maestría y Doctorado en Ciencias Bioquímicas, me brindó la oportunidad de continuar con mi formación académica y profesional.

A mi tutor, el Dr. Marco Antonio Velasco Velázquez, por permitirme formar parte de su grupo de investigación, así como por su guía y compromiso para el desarrollo del presente trabajo de investigación.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para realizar mis estudios de maestría (número de apoyo 759554).

A los proyectos UNAM-PAPIIT IV200121, UNAM-PAPIIT IN206622 y CONACyT A1-S-18285 por el apoyo financiero para la realización de este trabajo de investigación.

Al Departamento de Supercómputo (Miztli-UNAM) de la Dirección General de Cómputo y de Tecnologías de la Información y Computación por la asignación de los recursos para realizar los experimentos computacionales.

A los miembros del comité tutor, Dres. Alejandro Sosa Peinado y Rogelio Rodríguez Sotres por sus valiosas recomendaciones y aportaciones para el desarrollo de este trabajo de investigación.

A los miembros del comité sinodal, Dres. Adela Rodríguez Romero, Marcelino Arciniega Castro, Gabriel del Río Guerra, José Salud Rodríguez Zavala e Ismael Bustos Jaimes, por sus observaciones que aportaron y enriquecieron esta tesis.

Al Dr. Luis Córdova Bahena por todas sus enseñanzas, consejos y apoyo brindado a lo largo del desarrollo del presente trabajo de investigación.

A todos los miembros del laboratorio de Farmacología Molecular de la Facultad de Medicina por toda su retroalimentación, que me ayudó a potenciar mi aprendizaje.

## ÍNDICE

<b>ÍNDICE DE FIGURAS</b> .....	<b>V</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>VII</b>
<b>LISTA DE ABREVIATURAS</b> .....	<b>VIII</b>
<b>RESUMEN</b> .....	<b>11</b>
<b>CAPÍTULO I. INTRODUCCIÓN</b> .....	<b>13</b>
<b>CAPÍTULO II. ANTECEDENTES</b> .....	<b>16</b>
2.1 <i>Protein Data Bank</i> .....	16
2.2 Interacción de proteínas y otras moléculas.....	17
2.2.1 Interacciones proteína-proteína (IPP).....	18
2.3 Bases de datos sobre IPP.....	20
2.4 Métodos computacionales para la predicción de IPP.....	21
2.4.1 Acoplamiento molecular <i>in silico</i> .....	22
2.4.2 Simulaciones de dinámica molecular (SDM) .....	38
2.5 AKR1B1 como blanco terapéutico .....	53
2.5.1 Superfamilia de las aldo-ceto reductasas (AKR's).....	53
2.5.2 Aldosa Reductasa (AKR1B1): estructura y función.....	54
2.5.3 AKR1B1 como blanco farmacológico en diversas patologías.....	56
2.5.4 Inhibición farmacológica de AKR1B1.....	61
2.5.5 IPP de AKR1B1 .....	63
<b>CAPÍTULO III. PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>65</b>
<b>CAPÍTULO IV. OBJETIVOS</b> .....	<b>65</b>
<b>CAPÍTULO V. METODOLOGÍA</b> .....	<b>67</b>
5.1 Características de hardware y software utilizados.....	67
5.2 Caracterización del dominio de IPP de AKR1B1.....	69
5.2.1 Selección de la estructura cristalográfica de AKR1B1 para ensayos <i>in silico</i> .....	69
5.2.2 Acoplamientos moleculares <i>in silico</i> .....	69
5.2.3 Simulaciones de dinámica molecular.....	71
5.2.4 Alineamiento múltiple de secuencias .....	74

<b>CAPÍTULO VI. RESULTADOS Y DISCUSIÓN.....</b>	<b>75</b>
6.1 Comparación de estructuras de AKR1B1 humanas disponibles en el <i>Protein Data Bank</i> .....	75
6.2 Ensayos de validación de servidores de acoplamiento proteína-proteína. ..	78
6.3 Acoplamientos moleculares AKR1B1 y DAD .....	80
6.4 Simulaciones de dinámica molecular de complejos estructurales AKR1B1-DAD .....	82
6.5 Mutagénesis de alanina <i>in silico</i> .....	94
6.6 Alineamiento múltiple de secuencias .....	98
<b>CAPÍTULO VII. CONCLUSIONES .....</b>	<b>107</b>
<b>CAPÍTULO VIII. PERSPECTIVAS .....</b>	<b>107</b>
<b>CAPÍTULO IX. REFERENCIAS .....</b>	<b>109</b>

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Ilustración esquemática de los criterios de evaluación CAPRI.....	34
<b>Figura 2.</b> Principales términos representados en los campos de fuerza. ....	40
<b>Figura 3.</b> Representación esquemática de las PBC .....	46
<b>Figura 4.</b> Criterio de convención de mínima imagen. ....	46
<b>Figura 5.</b> Etapas de una simulación de dinámica molecular.....	47
<b>Figura 6.</b> Estructura tridimensional de AKR1B1 (código PDB:6F7R) .....	55
<b>Figura 7.</b> Mecanismo de reacción secuencial y ordenado para la reducción de aldehídos por medio de AKR1B1. ....	55
<b>Figura 8.</b> Principales reacciones que cataliza AKR1B1 .....	56
<b>Figura 9.</b> Estructura tridimensional de AKR1B1 co-cristalizada con NADPH y el inhibidor epalrestat (código PDB: 4JIR) .....	62
<b>Figura 10.</b> Alineamiento estructural AKR1B1 y HDAC3-DAD.....	64
<b>Figura 11.</b> Flujo de trabajo que representa de manera general la metodología seguida en el presente proyecto. ....	68
<b>Figura 12.</b> Conformaciones de inicio de AKR1B1 y DAD para SDM. ....	72
<b>Figura 13.</b> Comparación de RMSD-C $\alpha$ y agrupamiento jerárquico entre todas las estructuras de AKR1B1 humanas disponibles en el PDB. ....	75
<b>Figura 14.</b> Comparación de RMSD del esqueleto proteico entre las trayectorias de dinámica molecular de AKR1B1-NADPH, AKR1B1 apoenzima y RMSF de C $\alpha$ . .	76
<b>Figura 15.</b> Ensayos de validación de servidores de acoplamiento molecular.....	78
<b>Figura 16.</b> Acoplamientos moleculares AKR1B1-DAD .....	80
<b>Figura 17.</b> Sitio de interacción de AKR1B1 y DAD.. ....	82
<b>Figura 18.</b> Distancia entre centros de masas de AKR1B1 y DAD durante SDM de 500 ns. ....	83

<b>Figura 19.</b> RMSD en función del tiempo del esqueleto proteico de AKR1B1 en los diferentes complejos estructurales AKR1B1-DAD y radio de giro de AKR1B1. ....	84
<b>Figura 20.</b> RMSF de los C $\alpha$ de AKR1B1 en los tres complejos estructurales AKR1B1-DAD durante la SDM. ....	85
<b>Figura 21.</b> RMSD del esqueleto proteico de DAD de la trayectoria con respecto a la conformación inicial y comparación de RMSD con agrupamiento jerárquico entre las conformaciones adoptadas por DAD en SDM de 500 ns. ....	86
<b>Figura 22.</b> RMSF de los C $\alpha$ de DAD en los tres complejos estructurales durante la SDM. ....	88
<b>Figura 23.</b> Porcentaje de tiempo en el que los residuos de DAD presentan helicidad .....	89
<b>Figura 24.</b> Parámetro de orden generalizado S <sup>2</sup> del enlace N-H en los residuos de DAD de los tres complejos estructurales durante la SDM. ....	90
<b>Figura 25.</b> Contribución energética por residuo de AKR1B1 en el complejo estructural WT .....	92
<b>Figura 26.</b> Interacciones intermoleculares del complejo WT. ....	93
<b>Figura 27.</b> Análisis comparativo del RMSD del esqueleto proteico de AKR1B1 en el complejo estructural AKR1B1-DAD WT y de complejos mutantes. ....	94
<b>Figura 28.</b> Análisis de interacciones intermoleculares de complejos AKR1B1-DAD WT y mutantes. ....	96
<b>Figura 29.</b> Alineamiento múltiple de secuencias de AKR1B1. ....	98-99
<b>Figura 30.</b> Alineamiento múltiple de secuencias de los aminoácidos de las hélices $\alpha$ 8 y H2 de AKR1B1. ....	100-101
<b>Figura 31.</b> Alineamiento múltiple de secuencias de DAD. ....	102-103

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Bases de datos sobre IPP.....	21
<b>Tabla 2.</b> Resumen de los criterios que se establecen en CAPRI para clasificar los complejos predichos por acoplamiento molecular.....	34
<b>Tabla 3.</b> Tipos de ensamblajes usados en dinámica molecular.....	45
<b>Tabla 4.</b> Expresión de mRNA de AKR1B1 en diferentes tipos de tumores y neplasias hematológicas contenidas en la base de datos <i>Oncomine</i> .....	58-59
<b>Tabla 5.</b> Expresión y actividad específica de AKR1B1 en áreas no tumorales y tumorales de tejidos cancerosos humanos. ....	59
<b>Tabla 6.</b> Valores de IC50 de inhibidores farmacológicos de AKR1B1. ....	62
<b>Tabla 7.</b> Clasificación de la calidad de la predicción de los modelos de acoplamiento con la puntuación DockQ. ....	70
<b>Tabla 8.</b> Comparación de RMSD y energía libre de unión entre los complejos estructurales AKR1B1-DAD que se utilizaron en la SDM.....	83
<b>Tabla 9.</b> Energía libre de unión de los complejos estructurales AKR1B1-DAD en SDM de 500 ns.....	91
<b>Tabla 10.</b> Energía libre de unión de residuos de DAD que más favorecieron y desfavorecieron la unión con AKR1B1.....	92-93
<b>Tabla 11.</b> Energía libre de unión mediante cálculos MM/PBSA para el complejo tipo salvaje (WT) y complejos generados a través de mutaciones en residuos de interfaz energéticamente importantes de AKR1B1.....	95



## LISTA DE ABREVIATURAS

<b>Å</b>	Angstrom
<b>ΔG</b>	Energía libre de Gibbs
<b>μM</b>	Micromolar
<b>AIR</b>	Restricciones de Interacción Ambigua ( <i>Ambiguous Interaction Restrictions</i> por sus siglas en inglés)
<b>ALN</b>	Alineamiento
<b>Cα</b>	Carbono alfa
<b>CAPRI</b>	<i>Critical Assessment of Predicted Interactions</i> (por sus siglas en inglés)
<b>CASF</b>	<i>Comparative Assessment of Scoring Functions</i> (por sus siglas en inglés)
<b>CASP</b>	<i>Critical Assessment of Structure Prediction</i> (por sus siglas en inglés)
<b>Cl-</b>	Ión cloruro
<b>co-IP</b>	Co-inmunoprecipitación
<b>COO-</b>	Carboxilato
<b>cryoEM</b>	Criomicroscopía electrónica ( <i>cryogenic electron microscopy</i> por sus siglas en inglés)
<b>DAD</b>	Dominio de activación de desacetilasa ( <i>Deacetylase Activation Domain</i> por sus siglas en inglés)
<b>DNA</b>	Ácido desoxirribonucleico
<b>EMT</b>	Transición epitelio mesénquima ( <i>Epithelial-Mesenchymal Transition</i> por sus siglas en inglés)
<b>FEP</b>	Perturbación de la energía libre ( <i>Free Energy Perturbation</i> por sus siglas en inglés)
<b>FFT</b>	Transformada rápida de Fourier ( <i>Fast Fourier Transform</i> por sus siglas en inglés)
<b>FRET</b>	Transferencia de energía por resonancia de fluorescencia ( <i>Fluorescence Resonance Energy Transfer</i> por sus siglas en inglés)
<b>fs</b>	Femtosegundo

<b>GSH</b>	Glutación reducido
<b>GS-DHN</b>	3-glutationil-1,4-dihidroxinonano
<b>GS-HNE</b>	3-glutationil-4-hidroxinonanal
<b>GSSG</b>	Glutación oxidado
<b>HDAC3</b>	Histona desacetilasa 3
<b>HK</b>	Hexocinasa
<b>HNE</b>	4-hidroxinonenal
<b>IC50</b>	Concentración inhibitoria media
<b>IPP</b>	Interacción proteína-proteína
<b>ITC</b>	Calorimetría de titulación isotérmica ( <i>Isothermal Titration Calorimetry</i> por sus siglas en inglés)
<b>K+</b>	Ión potasio
<b>kDa</b>	Kilodalton
<b>kJ</b>	Kilojoules
<b>LIE</b>	Energía de interacción lineal ( <i>Linear Interaction Energy</i> por sus siglas en inglés)
<b>MM-GBSA</b>	Mecánica molecular del área de superficie generalizada de Born ( <i>Molecular Mechanics Generalized Born Surface Area</i> por sus siglas en inglés).
<b>MM-PBSA</b>	Mecánica molecular del área de superficie de Poisson-Boltzmann ( <i>Molecular Mechanics of Poisson-Boltzmann Surface Area</i> por sus siglas en inglés)
<b>MUC</b>	Modo de unión consenso
<b>NADPH</b>	Nicotinamida-adenina dinucleótido fosfato reducido
<b>NFκB</b>	Factor nuclear kappa B
<b>-NH3+</b>	Grupo amino
<b>nM</b>	Nanomolar
<b>nm</b>	Nanómetro
<b>NMR</b>	Resonancia magnética nuclear ( <i>Nuclear Magnetic Resonance</i> por sus siglas en inglés)

<b>ns</b>	Nanosegundo
<b>PBC</b>	Condiciones periódicas de la frontera ( <i>Periodic Boundary Conditions</i> por sus siglas en inglés)
<b>PGF2<math>\alpha</math></b>	Prostaglandina F2 $\alpha$
<b>PGH2</b>	Prostaglandina H2
<b>Rg</b>	Radio de giro
<b>RMSD</b>	Raíz de la desviación cuadrática media ( <i>Root Mean Squared Deviation</i> por sus siglas en inglés)
<b>RMSF</b>	Raíz de la fluctuación cuadrática media ( <i>Root Mean Squared Fluctuation</i> por sus siglas en inglés)
<b>RNA</b>	Ácido ribonucleico
<b>ROS</b>	Especies reactivas de oxígeno ( <i>Reactive Oxygen Species</i> por sus siglas en inglés)
<b>SASA</b>	Área de superficie accesible al solvente ( <i>Solvent Accessible Surface Area</i> por sus siglas en inglés)
<b>SDM</b>	Simulaciones de Dinámica Molecular
<b>SMRT/NCOR1</b>	Correpresor de receptores nucleares 1
<b>SPR</b>	Resonancia de plasmón superficial ( <i>Surface Plasmon Resonance</i> por sus siglas en inglés)
<b>TI</b>	Integración termodinámica ( <i>Thermodynamic Integration</i> por sus siglas en inglés)
<b>WT</b>	Tipo silvestre ( <i>Wild-type</i> por sus siglas en inglés)

## RESUMEN

En años recientes, el avance en el desarrollo de nuevas herramientas bioinformáticas y quimioinformáticas han permitido comprender más a fondo las bases moleculares de las enfermedades humanas y otros fenómenos biológicos, así mismo han posibilitado el desarrollo óptimo de nuevos agentes terapéuticos para el tratamiento de estas mismas. Muchos de los procesos biológicos importantes como la transducción de señales y la formación de complejos proteicos se logran llevar a cabo gracias a las interacciones entre proteínas. Por lo tanto, su identificación y caracterización son de importancia en la bioquímica y biología celular.

En este contexto, el presente trabajo de investigación se planteó con el objetivo de identificar dominios de interacción proteína-proteína (IPP) de la enzima aldo-ceto reductasa humana AKR1B1 empleando métodos computacionales como el acoplamiento molecular *in silico* y las simulaciones de dinámica molecular (SDM).

AKR1B1 es una potencial diana terapéutica debido a que su expresión está relacionada con diversos tipos de cáncer y su interacción con el dominio ***Deacetylase Activation Domain*** (DAD) provoca hipertrofia cardíaca grave. La inhibición de la actividad enzimática ha sido una estrategia muy utilizada para desarrollar nuevos tratamientos. Por el contrario, el desarrollo de inhibidores de la IPP ha sido una estrategia poco explotada, debido a la falta de información de complejos estructurales proteína-proteína de esta enzima, lo cual limita el desarrollo óptimo de nuevos agentes terapéuticos.

En la literatura, se ha reportado que AKR1B1 tiene IPP con el dominio DAD. Aunque se ha determinado experimentalmente que el residuo L289 de AKR1B1 es esencial para mantener esta interacción, se desconoce el modo de asociación entre estas proteínas, así como la identidad de otros residuos que participan en la interacción

Con la finalidad de predecir y caracterizar el modo de unión entre AKR1B1 y DAD, en primera instancia se realizó una búsqueda de las estructuras de AKR1B1 humanas que han sido dilucidadas y que se encuentran publicadas en la base de

datos *Protein Data Bank*. Posteriormente, se compararon todas entre sí mediante el cálculo de *Root-Mean-Square Deviation* (RMSD). Dado que no se observaron diferencias estructurales importantes, se decidió trabajar con la estructura de mejor calidad y resolución (6F7R). Por otro lado, para predecir el modo de unión entre AKR1B1 y DAD se efectuaron acoplamiento moleculares *in silico* empleando tres servidores que han tenido éxito en CAPRI (**Critical Assessment of PR**ediction of **I**nteractions). Antes de hacer los acoplamiento, dichos servidores se sometieron a ensayos de validación con la finalidad de determinar si son capaces de predecir la conformación de un complejo proteico cuya estructura cristalográfica es conocida (complejo DAD-HDAC3; PDB:4A69). Se obtuvo que las predicciones mejor calificadas por los tres servidores fueron de alta calidad, siguiendo los criterios que se establecen en CAPRI. Por lo tanto, se determinó que estos servidores son adecuados y confiables, así que se llevaron a cabo los acoplamiento de AKR1B1 y DAD utilizando estos servidores y se estableció un modo de unión consenso.

Finalmente, se efectuaron SDM de 500 ns de diferentes conformaciones en las que se consideró el modo de unión consenso predicho por el acoplamiento molecular y otras dos conformaciones, en donde DAD se une a un sitio lejano de la L289, con la finalidad de evaluar el modo de unión más estable. Para esto se hicieron análisis de RMSD, RMSF, distancia entre centros de masas de AKRB1-DAD, helicidad y parámetro de orden generalizado para DAD.

También se realizó un cálculo teórico de la energía libre de unión mediante el enfoque MM-PBSA, obteniendo que el modo de unión más estable fue el predicho con acoplamiento molecular. Asimismo, se determinó que los residuos más favorables energéticamente para la interacción son E193, Q197, S282, T289, L289 y S290. Estos residuos fueron mutados por alaninas (mutaciones simples) y se sometieron a SDM para observar el efecto de las mutaciones sobre la formación del complejo estructural AKR1B1-DAD.

Se espera que la información generada en el presente trabajo pueda ser la base para un futuro desarrollo de inhibidores de la IPP de esta enzima.

## **CAPÍTULO I. INTRODUCCIÓN**

La aldo-ceto reductasa humana AKR1B1, también conocida como aldosa reductasa, es una enzima multifuncional que desempeña funciones involucradas en biosíntesis, metabolismo intermediario y la detoxificación, actuando sobre aldehídos con relevancia biológica como: esteroides, productos de la peroxidación lipídica y productos finales de la glicosilación<sup>1</sup>. Además, cataliza la conversión de glucosa en sorbitol de manera dependiente de NADPH, siendo este el paso limitante de la vía del poliol. La importancia biológica de esta proteína se debe a que se le ha relacionado con diversas patologías diabéticas, inflamatorias, transición epitelio mesénquima (EMT) y se encuentra sobre-expresada en diversos tipos de cáncer<sup>2,3,4</sup>.

Diversos estudios han determinado que la expresión de esta enzima está fuertemente relacionada con el fenotipo agresivo en células de cáncer de mama triple negativo y de pulmón<sup>5,6</sup>. En dichos estudios se observó que el silenciamiento de la expresión de AKR1B1 conlleva a la disminución significativa en la migración celular, tasa de crecimiento y formación de colonias. Estos resultados indican que AKR1B1 desempeña un papel en la EMT del cáncer.

En la literatura, se han mencionado los posibles mecanismos con los que se relaciona la presencia de AKR1B1 con la EMT en distintos tipos de cáncer. Estos mecanismos presentan diferencias dependiendo del tipo de cáncer que se está estudiando, sin embargo, todos implican alguno de los siguientes enfoques, ya sea mediante la actividad enzimática de AKR1B1 o su interacción con otras proteínas<sup>7</sup>.

En el trabajo de Schwab, A y colaboradores<sup>5</sup> muestran el vínculo que existe entre la actividad de AKR1B1 en la vía del poliol, la diferenciación y la agresividad en células madre cancerosas. Asimismo, Wu y colaboradores<sup>6</sup> informaron que AKR1B1 puede inducir EMT a través de la síntesis de PGF2 $\alpha$  y la activación del factor nuclear NF $\kappa$ B en células de cáncer de mama triple negativo.

Por otra parte, en el estudio de Jia-Xing Zhao y colaboradores<sup>8</sup> se determinó que la sobre-expresión de AKR1B1 en hepatocarcinoma desencadena la vía de señalización de AKT/mTOR a través de la interacción con el dominio cinasa de

AKT1 y como consecuencia se observó un aumento en la tumorigenicidad en las células HepG2. Finalmente, Ming-Bing Xiao y colaboradores<sup>9</sup> reportaron que la interacción de AKR1B1 con el receptor  $\beta$ 2-adrenérgico promueve la proliferación y dificulta la apoptosis a través de la vía ERK1/2 en la línea celular de cáncer de páncreas BXP-3.

De igual importancia, Thiagarajan y colaboradores<sup>10</sup> reportaron que AKR1B1 compete con la desacetilasa de histona 3 (HDAC3) para unirse al dominio DAD (*Deacetylase Activation Domain*) del correpresor de receptores nucleares (SMRT/NCOR1). Las principales consecuencias de esta interacción incluyen la disociación del complejo HDAC3-DAD, conduciendo a la degradación de HDAC3. La reducción de los niveles de HDAC3 se ve acompañada de un aumento en la actividad de PPAR $\gamma$ , lo que provoca a una hipertrofia cardíaca grave debido a la acumulación de cantidades patológicas de lípidos en el corazón de ratones.

La interacción AKR1B1-DAD se demostró experimentalmente mediante ensayos de co-inmunoprecipitación (co-IP), además los autores identificaron mediante ensayos de mutagénesis sitio dirigida que la L289 de AKR1B1 es importante para que se mantenga la formación del complejo estructural. Sin embargo, aún no se ha caracterizado de manera más detallada la interacción AKR1B1-DAD.

Dadas estas razones, la inhibición de la actividad enzimática de AKR1B1 ha sido utilizada como la estrategia principal en la búsqueda de nuevos agentes terapéuticos y algunos inhibidores han mostrado moderado éxito. Por el contrario, en la actualidad no existen reportes sobre el desarrollo de inhibidores de la interacción proteína-proteína (IPP) de AKR1B1, debido a la falta de información de complejos estructurales proteína-proteína de esta enzima.

Las interacciones proteína-proteína juegan un papel esencial en muchos procesos biológicos, por lo que su estudio y entendimiento, permitirá conocer las funciones de las proteínas, las vías y mecanismos por las que se establecen distintas enfermedades y otros fenómenos fisiológicos. Esta valiosa información conducirá a enfoques mejorados para el descubrimiento de fármacos. Por esta razón, en las últimas décadas se han desarrollado diversas técnicas experimentales para

identificar IPP y con ello los estudios sobre redes de proteínas han aumentado significativamente. Sin embargo, esta gran cantidad de resultados suelen tener desacuerdos entre sí y a veces son contradictorios, además existe una limitación en la comprensión de las superficies de interacción de manera detallada, por lo que dificulta el desarrollo óptimo de nuevos agentes terapéuticos<sup>11,12</sup>.

En años recientes, los métodos computacionales como el acoplamiento molecular y las simulaciones de dinámica molecular se han aplicado con éxito para predecir y caracterizar de manera detallada las IPP. Estos métodos permiten un análisis detallado de los aspectos estructurales y funcionales de las IPP, proporcionando información crucial sobre los mecanismos de interacción. La evaluación crítica de la predicción de interacciones (CAPRI: *Critical Assessment of PRediction of Interactions*) es una iniciativa comunitaria establecida en 2001 que ha impulsado el progreso en los métodos de acoplamiento proteína-proteína, ya que reta a los diferentes algoritmos de acoplamiento para predecir a ciegas la estructura de complejos proteína-proteína. De este modo se ha incentivado a los creadores de estos algoritmos a mejorar sus métodos de predicción y de este modo obtener predicciones más confiables<sup>13-15</sup>.

El objetivo del presente trabajo de investigación consistió en caracterizar de manera detallada la interacción AKR1B1-DAD mediante el uso de métodos computacionales para identificar dominios o residuos que sean importantes para la IPP y, por lo tanto, se espera que la información generada en este trabajo pueda ser la base para el futuro desarrollo óptimo de inhibidores de la IPP de AKR1B1.



## **CAPÍTULO II. ANTECEDENTES**

### **2.1 Protein Data Bank**

El *Protein Data Bank*, es una base de datos que contiene información sobre las estructuras tridimensionales de macromoléculas biológicas como proteínas, ácidos nucleicos y sus complejos, que fueron determinadas experimentalmente por diversos grupos de investigación. Actualmente, el *Protein Data Bank* almacena datos de modelos estructurales a nivel atómico derivados de tres métodos: cristalografía de rayos X, espectroscopia de resonancia magnética nuclear (NMR) y crio-microscopía electrónica (cryoEM)<sup>16</sup>.

El *Protein Data Bank* fue creado en 1971 por el *Brookhaven National Laboratory*. En 2003, se originó el wwPDB, que incluye la colaboración entre RSCB-PDB, PDBe (Europa), PDBj (Japón), y BioMagResBank BMRB (USA), cuyo objetivo fue operar un sistema de software global, unificado, público y gratuito<sup>16</sup>.

Los archivos PDB contienen la información completa de los modelos estructurales y se identifican por un código de cuatro caracteres, por ejemplo (6F7R). De manera más específica, los archivos PDB contienen datos de las coordenadas atómicas, secuencia y composición de los biopolímeros y ligandos que se encuentra unidos, etcétera. Además, se puede conocer los indicadores de calidad global importantes para la estructura comparándose contra las estructuras previamente depositadas.

Entre los indicadores de calidad para las estructuras cristalográficas se encuentran los valores de R y “*Rfree*”. Este último es una medida del ajuste del modelo a un pequeño subconjunto de los datos experimentales que no se utilizaron en el refinamiento del modelo. El siguiente indicador es el “*Clashscore*” que indica el número de pares de átomos que se encuentran inusualmente cercanos. Otro de los indicadores son los “*Ramachandran outliers*” o valores atípicos de Ramachandran. Estos valores representan el porcentaje de los residuos que presentan una combinación de los ángulos diedros  $\phi$  y  $\psi$  no adecuada con respecto al número total de residuos presentes en la estructura. Asimismo, se evalúan los “*Sidechain outliers*” o valores atípicos de la cadena lateral, que se

determinan como el porcentaje de residuos con una conformación de la cadena lateral inusual, con respecto al número total de los residuos presentes en la estructura. El último indicador es el *real space R-value outliers* normalizado (RSRZ) y se calcula como el porcentaje de valores atípicos de RSRZ ( $>2$ ), con respecto al número total de residuos para los que se calculó RSRZ. Este valor se calcula solo para aminoácidos y nucleótidos estándar en cadenas de proteínas, DNA y RNA<sup>16,17</sup>.

La resolución a la que se obtiene la estructura también es un dato importante, ya que indica el nivel de detalle presente en el patrón de difracción y el nivel de detalle que se verá cuando se calcule el mapa de densidad de electrones. En la cristalografía de rayos X, las estructuras de alta resolución ( $\leq 1.0 \text{ \AA}$ ) están muy ordenadas, puesto que las proteínas presentes en el cristal se encuentran alineadas de forma idéntica, por lo que la dispersión de los rayos X se realiza de manera similar. Por lo tanto, el patrón de difracción muestra los detalles finos del cristal, por el contrario, las estructuras de baja resolución ( $\geq 3.0 \text{ \AA}$ ) únicamente muestran los contornos básicos de la cadena proteica, así que los datos de la estructura atómica se tienen que inferir<sup>17</sup>.

## **2.2 Interacción de proteínas y otras moléculas**

La mayoría de los procesos biológicos operan gracias a las funciones de las proteínas. Debido a su importancia, se han dilucidado extensas redes de funciones para descifrar los posibles mecanismos como interacción y asociación entre proteínas y otro tipo de moléculas que gobiernan estos fenómenos biológicos. Algunos ejemplos de estas interacciones son proteína-DNA, que se presentan en la regulación de la expresión génica, interacciones proteína-RNA en la regulación de la transcripción y procesamiento del RNA e interacciones proteína-proteína en la transducción de señales<sup>18</sup>.

El estudio de estas redes de interacción ha proporcionado información valiosa para comprender de manera integral la biología a nivel molecular.

### 2.2.1 Interacciones proteína-proteína (IPP)

Nos enfocaremos especialmente en las IPP, debido a que están involucradas en el adecuado funcionamiento de la mayoría de los procesos fisiológicos en las células. Sin embargo, pueden ser responsables del desarrollo de procesos patológicos tales como Alzheimer, enfermedades priónicas, cáncer, entre otras.

La arquitectura básica de la red de interacción proteína-proteína es similar en todas las células. Las proteínas están acopladas directa o indirectamente en una gran red de interacción proteína-proteína. Se ha reportado previamente que algunas proteínas actúan como proteínas centrales, ya que están altamente conectadas a otras. Asimismo, se ha observado que el silenciamiento de la expresión de una proteína que tiene un papel central en muchas redes tiende a ser letal. A este fenómeno se le conoce comúnmente como la regla de la centralidad-letalidad<sup>19</sup>. Sin embargo, también se ha reportado que esta regla no es general para todos los centros, ya que se ha observado que muchos centros se enriquecen en fenotipos condicionales<sup>20</sup>.

En la literatura se han informado diferentes tipos de IPP. Un ejemplo son las interacciones permanentes y transitorias. Las IPP permanentes son fuertes e irreversibles, con una afinidad estimada del orden nanomolar, por otra parte, las IPP transitorias implican interacciones de proteínas que se forman y se rompen fácilmente y tienen afinidades del orden de micromolar<sup>21</sup>.

Las IPP pueden estar reguladas por diversos factores ambientales tales como la temperatura, pH, fuerza iónica, o por mecanismos celulares como modificaciones covalentes, unión con ligandos de manera no covalente, etcétera<sup>22,23</sup>.

Debido a su importancia, se han desarrollado muchos métodos experimentales basados en principios biofísicos, bioquímicos o genéticos, para identificar y caracterizar IPP. Cada enfoque tiene sus fortalezas y limitaciones. A continuación, se mencionarán algunos de estos métodos.

### *2.2.1.1 Fluorescence Resonance Energy Transfer (FRET)*

FRET es un fenómeno útil para el estudio de las interacciones bioquímicas, incluidas las IPP. El procedimiento consiste en que los socios de interacción se encuentran fusionados con un fluoróforo donante y otro aceptor. Por lo tanto, cuando existe un acercamiento (1-10 nm) entre los socios de interacción que permita una superposición entre el espectro de emisión del donante y espectro de excitación del aceptor, hay una transferencia de energía del fluoróforo donante al aceptor, como consecuencia existe un apagado en la fluorescencia del donante y un incremento en la fluorescencia del receptor<sup>24</sup>.

Entre las principales ventajas de este método es que se puede usar para visualizar y cuantificar IPP en tiempo real en células vivas y con alta resolución espaciotemporal. Sin embargo, sus principales desventajas son los cambios en el entorno local, autofluorescencia, fotoblanqueo y de manera general la baja relación señal-ruido<sup>24</sup>.

### *2.2.1.2 Sistema de doble híbrido*

Este método se desarrolló en levadura y es uno de los más empleados para detectar o confirmar IPP. El principio del doble híbrido se basa en que dos proteínas que se expresan de forma separada se fusionan con el dominio de unión a DNA del factor transcripcional Gal4 y el dominio de activación de Gal4, respectivamente. Cuando existe una interacción entre las proteínas, la funcionalidad de Gal4 se restablece e induce la transcripción de varios genes reporteros. Una de las principales ventajas de este método es que permite detectar interacciones transitorias que a menudo son relevantes en cascadas de señalización. Mientras que su desventaja es que el método genera muchos resultados falsos-positivos<sup>25</sup>.

### *2.2.1.3 Calorimetría de titulación isotérmica (ITC)*

La ITC es una técnica que se ha utilizado ampliamente para estudiar las interacciones proteína-ligando y proteína-proteína. Esta técnica se basa en el hecho de que, al interaccionar dos proteínas o socios de interacción, se producen cambios en los parámetros termodinámicos  $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  y  $K_a$ , los cuales son

cuantificados. De manera general, un experimento de ITC común consiste en colocar una de las proteínas interactuantes en una jeringa de inyección y la otra en una celda calorimétrica, posteriormente se realizan inyecciones paso a paso, finalmente se mide el calor liberado o absorbido en condiciones de equilibrio durante la titulación, y se determinan los parámetros restantes<sup>24,26</sup>.

#### *2.2.1.4 Resonancia de plasmón superficial (SPR)*

SPR es un método óptico que mide el índice de refracción de capas muy delgadas de un material adsorbido en un metal. La técnica consiste incidir energía luminosa en un ángulo definido y parte de esta es capaz de interactuar con los electrones deslocalizados en la película de metal (plasmón), reduciendo así la intensidad de la luz reflejada. La respuesta es proporcional a la masa unida a la superficie, es decir, el número de moléculas unidas multiplicado por su masa molecular. La generalidad se extiende a todo tipo de moléculas, por lo que se pueden detectar y estudiar las IPP. En este caso, una de las proteínas interactuantes se inmoviliza sobre la superficie de una película de oro y el socio de interacción se inyecta en la superficie<sup>25,27</sup>.

La formación del complejo proteína-proteína se puede monitorear como el aumento de la respuesta durante la fase de asociación en el sensograma, que es una representación gráfica de la respuesta dependiente del tiempo. Asimismo, el análisis SPR puede proporcionar información sobre las constantes de tasa de asociación y disociación ( $K_{on}$  y  $K_{off}$ ) para determinar la constante de afinidad en el equilibrio  $K_d$ <sup>27,28</sup>.

En general, una limitante de cualquier método de detección de IPP es que solo se puede confiar cuando se detectó la interacción. Sin embargo, aquellas IPP que no son capaces de detectar, se tiene la incertidumbre de saber si existen.

### **2.3 Bases de datos sobre IPP**

Exististe una gran variedad de bases de datos que contienen información acerca de las interacciones entre las proteínas y que fueron determinadas experimentalmente, las cuales se muestran en la tabla 1.

<b>Base de datos</b>	<b>Numero de interacciones</b>	<b>URL</b>	<b>Referencia</b>
LiveDIP	55733	<a href="http://dip.doe-mpi.ucla.edu">http://dip.doe-mpi.ucla.edu</a>	29
BIND	83517	<a href="http://bind.ca">http://bind.ca</a>	30
STRING	730000	<a href="http://string.embl.de">http://string.embl.de</a>	31
MINT	71854	<a href="http://mint.bio.uniroma2.it/mint">http://mint.bio.uniroma2.it/mint</a>	32
IntAct	68165	<a href="http://www.ebi.ac.uk/intact">http://www.ebi.ac.uk/intact</a>	33
BioGRID	116000	<a href="http://www.thebiogrid.org">http://www.thebiogrid.org</a>	34
HPRD	33710	<a href="http://www.hprd.org">http://www.hprd.org</a>	35

**Tabla 1.** Bases de datos sobre IPP. Modificado de<sup>36</sup>.

Con la ayuda de los métodos experimentales, se han realizado esfuerzos significativos para comprender la naturaleza de las IPP. No obstante, estos requieren de un trabajo intenso y pueden presentar diversos errores sistemáticos. Esto genera que haya una falta de datos continuos, lo que limita nuestra capacidad para comprender las bases moleculares de las enfermedades humanas y otros fenómenos biológicos<sup>37,38</sup>. Sin embargo, en años recientes el desarrollo de enfoques computacionales, así como el experimento CAPRI<sup>13-15</sup>, ha permitido la predicción de IPP de una manera rápida, confiable y superando algunas de las limitaciones de las técnicas experimentales.

#### **2.4 Métodos computacionales para la predicción de IPP**

Las herramientas computacionales cada vez se utilizan más para la predicción de IPP potencialmente relevantes desde el punto de vista biológico. Los enfoques bioinformáticos consideran el término IPP como las asociaciones entre proteínas que incluyen aspectos de relación evolutiva, función y estructura<sup>39,40</sup>.

Se pueden predecir IPP de *novo* mediante técnicas computacionales como el acoplamiento molecular y las simulaciones de dinámica molecular (SDM) para proteínas cuya estructura tridimensional ha sido determinada. Sin embargo, aún no se han dilucidado todas las estructuras proteicas que forman el proteoma humano. Para superar este problema se han desarrollado estrategias para predecir y construir modelos computacionales de las estructuras proteicas.

La evaluación crítica de la predicción de la estructura (CASP: **Critical Assessment of Structure Prediction**), es un experimento comunitario que evalúa el estado del arte en el modelado de la estructura de las proteínas y ha servido durante mucho

tiempo como la evaluación estándar de oro para la precisión de la predicción de la estructura<sup>41,42</sup>. Debido a esto, en los últimos 5 años se ha generado un avance importante en esta área.

La predicción de la estructura basada en plantillas aprovecha las estructuras tridimensionales de proteínas homólogas que se encuentran en el *Protein Data Bank* para generar un modelo tridimensional de la secuencia de aminoácidos problema. Este enfoque es el que genera los resultados más precisos. Sin embargo, una de sus principales limitaciones es cuando el homólogo más cercano no se ha resuelto experimentalmente. Otro enfoque muy utilizado es el modelado sin plantilla o *ab initio*, el cual es el más desafiante, ya que se predice la estructura de la proteína a partir de la secuencia de aminoácidos, tomando en cuenta sus propiedades fisicoquímicas y los términos energéticos que impulsan el plegamiento. Usando este enfoque se han obtenido predicciones de la estructura con una precisión baja para proteínas con más de 130 aminoácidos<sup>42</sup>.

El experimento CASP14 marcó un incremento extraordinario en la precisión de la predicción de las estructuras tridimensionales de las proteínas con la aparición del sistema de aprendizaje profundo AlphaFold2. Este sistema fue desarrollado por DeepMind de Alphabets/Google. En el CASP14, las estructuras predichas por esta inteligencia artificial fueron por mucho las más precisas en comparación con los métodos de la competencia, obteniendo una precisión casi experimental<sup>43</sup>.

Recientemente, AlphaFold ha creado una base de datos de acceso abierto (<https://alphafold.ebi.ac.uk/>), que contiene las predicciones de estructura de las proteínas para el proteoma humano y otras proteínas clave de interés.

A continuación, se abordarán dos técnicas para la predicción y caracterización *in silico* de IPP a partir de estructuras tridimensionales resueltas experimentalmente que se usarán en este trabajo.

#### 2.4.1 Acoplamiento molecular *in silico*

El acoplamiento molecular *in silico*, es un método computacional para predecir el modo de unión de un ligando con un receptor objetivo, donde el receptor suele ser

una proteína o una molécula de ácidos nucleicos como el DNA o RNA, mientras que el ligando es una molécula pequeña (<1 kDa), péptido u otra proteína<sup>44,45</sup>.

Este método implica explorar un espacio de búsqueda definido mediante uno o varios algoritmos de búsqueda, generando varias conformaciones que representan los posibles modos de unión del ligando con el receptor y una función de evaluación que clasifica de manera adecuada las conformaciones candidatas para determinar el mejor modo de unión<sup>44,45</sup>.

#### *2.4.1.1 Algoritmos de búsqueda*

Los algoritmos de búsqueda se definen mediante un conjunto de reglas y parámetros aplicados para predecir las diversas conformaciones. Se pueden clasificar en dos categorías principales: sistemáticos y estocásticos. En el primer caso, se muestrea el espacio de búsqueda en intervalos predefinidos, por lo que el resultado de la búsqueda es determinista y la calidad del resultado depende de la granularidad (tamaño en el que se divide el espacio de búsqueda: posibles orientaciones y conformaciones) del muestreo del espacio de búsqueda. Los algoritmos de búsqueda sistemáticos se utilizan frecuentemente en el acoplamiento rígido proteína-proteína<sup>44</sup>.

Por otra parte, los algoritmos estocásticos realizan iterativamente cambios aleatorios en sus pasos, por lo tanto, el resultado de la búsqueda varía. Es por esto por lo que los algoritmos de búsqueda estocástica son más adecuados para el acoplamiento flexible ligando-proteína<sup>44</sup>.

Los algoritmos de búsqueda también pueden clasificarse de acuerdo con la amplitud con la que exploran el espacio de búsqueda, ya sea de manera local o global. En el primer caso, hay una tendencia a encontrar la energía mínima más cercana a la conformación actual, mientras que los algoritmos de búsqueda globales exploran la energía mínima más favorable dentro del espacio de búsqueda definido<sup>44</sup>.



A continuación, se mencionan algunos métodos que emplean diferentes algoritmos de búsqueda que suelen usarse en programas o servidores de acoplamiento molecular.

#### Complementariedad de forma

Los algoritmos de complementariedad de forma toman en cuenta la superposición geométrica entre dos moléculas. Se efectúan varios alineamientos estructurales entre ligando y receptor. Este enfoque permite identificar los probables sitios de unión de una proteína mediante una búsqueda de superficie macromolecular<sup>45</sup>.

#### Construcción incremental

Este método consiste en dividir el ligando en fragmentos que posteriormente se acoplan de manera separada en el sitio del receptor. Ya que se ha realizado el acoplamiento, los fragmentos se fusionan. Una de las ventajas de la fragmentación es que permite que el algoritmo considere la flexibilidad del ligando<sup>45</sup>.

#### Recocido simulado

El algoritmo de recocido simulado es un algoritmo de optimización heurístico, siguiendo la estrategia de mejora iterativa con el objetivo de acercarse al mínimo energético global. Sin embargo, esta búsqueda suele quedar estancada en un óptimo local, por lo que se acostumbra a hacer el proceso varias veces, partiendo de diferentes configuraciones generadas aleatoriamente. En el algoritmo de recocido simulado, un sistema biomolecular se somete a una simulación con una temperatura alta, posteriormente la temperatura se disminuye de manera gradual hasta que sus átomos se reposicionan para generar una conformación de baja energía. De tal manera que la temperatura funciona como un parámetro de control del algoritmo. Una vez que la energía llega a un equilibrio térmico a una temperatura dada, el sistema se vuelve a enfriar y así sucesivamente. Cada solución generada se compara con la anterior, y si la energía es menor se acepta y el proceso se repite. Los resultados que se obtienen mediante este algoritmo suelen ser muy precisos porque analiza de manera más detallada el estado

conformacional y la flexibilidad del ligando, así como de la proteína en diferentes estados termodinámicos en un intervalo de tiempo<sup>45,46</sup>.

### Programación evolutiva

Los algoritmos de programación evolutiva se inspiran en los principios darwinianos de evolución y selección natural para resolver problemas. La idea principal consiste en simular la evolución de las especies a través de procesos de selección, mutación y reproducción aplicando aproximaciones computacionales llamadas operadores genéticos. Sin embargo, los operadores más usados son los de mutación, que están basados en distribuciones de probabilidad. De manera general estos algoritmos trabajan de la siguiente manera: en primera instancia las soluciones candidatas (individuos o cromosomas) son modificados empleando los operadores genéticos, después se seleccionan las soluciones individuales mejor clasificadas (parentales y mutadas) para la siguiente generación. Este procedimiento se repite, de tal manera que cada vez evoluciona hacia mejores soluciones. El objetivo final es encontrar la solución óptima<sup>45,47</sup>.

De acuerdo con los resultados de las rondas 38-45 en CAPRI<sup>48</sup>, los servidores de acoplamiento molecular proteína-proteína mejor evaluados tienen en común que utilizan los métodos de complementariedad de forma<sup>49,50</sup>. Un algoritmo ampliamente empleado es el de transformada rápida de Fourier (FFT). La mayoría de los algoritmos basados en FFT tiene un enfoque totalmente geométrico, en el cual definen al receptor y ligando en cuadrículas, mediante funciones tridimensionales discretas, posteriormente se evalúa el grado de superposición y penetración de la superficie molecular sobre los desplazamientos de las moléculas en tres dimensiones empleando la FFT para acelerar el muestreo del espacio traslacional. Finalmente, se realiza una exploración de las orientaciones relativas de las moléculas<sup>51,52</sup>.

En virtud de la alta eficiencia numérica del algoritmo FFT, es posible explorar el espacio conformacional de miles de supuestos complejos proteína-proteína, ya que acelera el cálculo de la energía de interacción en cada orientación receptor-ligando que es expresada como la suma de las funciones de correlación. Sin

embargo, la aceleración eficiente se logra solamente en el subespacio tridimensional de traslación o de rotación<sup>53</sup>.

Otro algoritmo comúnmente utilizado es el de *Geometric Hashing*, este algoritmo consiste en establecer puntos críticos de la superficie de la proteína y ligando para definir marcos de coordenadas locales que se almacenan en una tabla de *hash* de búsqueda rápida y eficiente en memoria. En este proceso, no solo ahorra memoria y acelera la búsqueda, sino que también limita la búsqueda a un conjunto finito (pero grande) de todas las transformaciones posibles<sup>54</sup>. *Hash* es una función que transforma cualquier bloque arbitrario de datos (clave) en una nueva serie de caracteres con una longitud fija (valores). Mientras que las tablas de *hash* es un contenedor que asocia los elementos claves con los valores, que son obtenidos aplicando la función *hash*<sup>55</sup>.

En la etapa de reconocimiento, los marcos de coordenadas locales se superponen y las posiciones de los puntos críticos cercanas al receptor se utilizan para buscar marcos candidatos del ligando en la tabla de *hash*. Un alto número de correspondencias con un marco en particular indica una gran similitud local en la forma de las superficies. Sin embargo, se considera que una implementación basada en tablas *hash* es ineficiente debido a la distribución no uniforme de los datos en el espacio *hash*. Esto da como resultado tiempos de búsqueda más largos, ya que las listas de entradas *hash* tienen diferentes longitudes<sup>56</sup>.

#### 2.4.1.2 Funciones de evaluación

Todos los métodos de acoplamiento molecular requieren de al menos una función de evaluación para clasificar los modos de unión candidatos. Las funciones de evaluación son métodos matemáticos para estimar la afinidad de unión entre un ligando y receptor. La energía libre de unión ( $\Delta G$ ) se obtiene con la ecuación de Gibbs-Helmholtz:

$$\Delta G_{unión} = \Delta H_{unión} - T\Delta S_{unión}$$

Donde  $\Delta G$  es la energía libre de unión,  $\Delta H$  es la entalpía de unión,  $T$  es la temperatura en grados Kelvin y  $\Delta S$  es la entropía de unión. Asimismo,  $\Delta G$  se

puede correlacionar con la constante de disociación  $K_d$ , tal como se muestra en la siguiente ecuación:

$$\Delta G_{unión} = -RT \ln K_d$$

donde  $R$  es la constante universal de los gases ideales y  $K_d$  es la constante de disociación<sup>45</sup>. Finalmente, la energía libre de unión también puede definirse como la diferencia de energía libre entre los estados unido y no unido.

$$\Delta G_{unión} = \Delta G_{complejo} - (\Delta G_{ligando} + \Delta G_{receptor})$$

Para evaluar el desempeño de las funciones de evaluación, se ha desarrollado una evaluación comparativa de las funciones de evaluación o **Comparative Assessment of Scoring Functions** (CASF), con el objetivo promover la mejora en las fortalezas y debilidades de las distintas funciones de evaluación. Cada función de evaluación se califica mediante cuatro métricas, que incluyen "poder de puntuación", "poder de clasificación", "poder de acoplamiento" y "poder de detección". Desde la implementación de CASF se ha observado una mejora considerable en las funciones de evaluación<sup>57</sup>.

Las funciones de evaluación se pueden clasificar en tres tipos: empíricas, basadas en campos de fuerza y basadas en el conocimiento.

#### Funciones de evaluación empíricas

Los métodos empíricos evalúan la afinidad de unión de un complejo proteína-ligando, sumando los factores energéticos importantes para la unión tales como enlaces de hidrógeno, efectos hidrofóbicos, choques estéricos, interacciones polares, etcétera. Se estiman los parámetros de la función con base en complejos proteína-ligando con afinidades y estructuras conocidas para optimizar el peso de los factores energéticos mediante un análisis de regresión lineal<sup>58,59</sup>.

En vista de que sus términos de energía son simples, estas funciones de evaluación son adecuadas para predecir la afinidad de unión y la conformación del ligando con un bajo costo computacional<sup>59</sup>. A pesar de la evaluación directa y rápida de la afinidad de unión, estas funciones de evaluación sufren las

limitaciones del término de penalización debido a las malas estructuras y la alta dependencia de la posición del átomo de hidrógeno<sup>60</sup>.

En el trabajo de Xiao Hui Ma y colaboradores<sup>61</sup> se desarrolló una función de evaluación empírica para calcular la energía libre de unión de 20 complejos proteína-proteína disponibles en el *Protein Data Bank*, y obtuvieron que los valores de  $\Delta G$  de unión predichos en 16 de los 20 complejos, tuvieron una diferencia de energía con respecto a los  $\Delta G$  de unión experimentales de  $\pm 1.0$  cal, mientras que los 4 complejos restantes, presentaron una diferencia de  $\Delta G$  de unión en el intervalo de  $\pm 1.5$ - $2.1$  cal con respecto a los valores experimentales. Sin embargo, la función de energía libre de unión que ellos presentan se basa en un tratamiento aproximado en el que las moléculas se tratan como cuerpos rígidos.

En el CASF-2016 participaron 26 funciones de evaluación, de las cuales 15 eran funciones de evaluación empíricas. De manera general, la mayoría de las funciones de evaluación mostraron tasas de éxito mayor al 70%. Sin embargo, la mayoría de las funciones son relativamente exitosas en uno o dos de los aspectos evaluados. En particular, las funciones de evaluación empíricas tuvieron al menos un representante entre las tres mejores funciones evaluadas en cada una de las pruebas. En la prueba de poder de puntuación, la función X-Score fue la segunda mejor calificada, obteniendo un coeficiente de correlación de Pearson de 0.65 en esta prueba. Asimismo, en la prueba de poder de clasificación, las funciones empíricas ChemPLP, X-Score y PLP2 obtuvieron los mejores resultados, presentando coeficientes de correlación de Spearman de aproximadamente 0.7. Por otro lado, en la prueba de poder de acoplamiento, las funciones ChemPLP, GlideScore-SP y ChemScore obtuvieron tasas de éxito de 85%, siendo las más altas en esta prueba. Finalmente, en la prueba de poder de detección, nuevamente las funciones mencionadas anteriormente desempeñaron el mejor papel, ya que obtuvieron los valores de factor de mejora más altos (15, 14, 12), respectivamente<sup>57</sup>.

Sin embargo, también hubo funciones empíricas que estuvieron entre las tres peores en cuanto a su desempeño en las pruebas. En la prueba de poder de

puntuación, la función London-dG MOE fue la tercera peor calificada, obteniendo un coeficiente de correlación de Pearson de 0.45. También en la prueba de poder de clasificación, la función GlideScore tuvo el valor más bajo (0.35) de coeficiente de correlación de Spearman. Asimismo, en la prueba de poder de acoplamiento, la función ASE obtuvo una tasa de éxito de 50%, siendo de las más bajas en esta prueba. Por último, esta misma función fue la peor evaluada en la prueba de poder de detección, presentando un valor de factor de mejora de tres<sup>57</sup>.

#### Funciones de evaluación basadas en campos de fuerza

Las funciones de evaluación basadas en campos de fuerza de mecánica molecular calculan la energía de unión considerando las interacciones de Van der Waals y electrostáticas entre los pares de átomos del complejo proteína-ligando. Sin embargo, estos métodos comúnmente no consideran los componentes entrópicos y el efecto del disolvente o los utilizan de manera reducida. Debido a esto, el rendimiento de las funciones de evaluación basadas en campos de fuerza no es lo suficientemente bueno<sup>58,59</sup>.

La ecuación que involucra los contactos atómicos por pares entre el ligando y la proteína consta de un potencial de Lennard-Jones y un término de Coulomb, tal como se muestra en la siguiente ecuación:

$$\Delta G_{unión} = \sum_{i=1}^{ligando} \sum_{j=1}^{proteína} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

En donde los parámetros A y B se definen para cada par de combinaciones de tipos de átomos diferentes, R es la distancia entre los centros atómicos, q es la carga parcial de cada átomo y  $\epsilon$  es la constante dieléctrica. La elección concreta de cada uno de estos parámetros es relevante, ya que puede tener efectos muy importantes sobre el comportamiento y el rendimiento de la función de evaluación<sup>58</sup>.

En general, las funciones de evaluación basadas en campos de fuerza pueden calcular directamente las interacciones entre los átomos de proteína y ligando. No

obstante, requieren un alto costo computacional<sup>59</sup>. Una de las principales limitaciones de estas funciones de evaluación es que no contemplan los términos entrópicos y de solvatación. En ausencia de los efectos de solvatación, existe un sesgo hacia las interacciones coulombicas. Asimismo, sufren de distancias de corte para el tratamiento de interacciones no enlazadas, lo que complica el tratamiento preciso de los efectos de largo alcance<sup>60</sup>.

En el trabajo de Shuangye Yin y colaboradores<sup>62</sup> desarrollaron una función de evaluación basada en campos de fuerza llamada MedusaScore. Esta función mostro un rendimiento sólido, ya que de un conjunto de 100 complejos proteína-ligando reconoció con éxito la pose cristalográfica de 82. Siendo superior a otras funciones de evaluación conocidas como X-Score, PLP, LUDI y DrugScore, las cuales tuvieron una tasa de éxito del 65, 76, 67 y 71 por ciento, respectivamente. De la misma manera, para la predicción de la afinidad de unión de un conjunto de 1057 complejos se obtuvo un coeficiente de correlación entre MedusaScore y la constante de disociación experimental de 0.60. Esta precisión de la predicción también es significativamente mayor que otras 14 funciones de evaluación que se han probado con el mismo conjunto de datos. La principal limitación de esta función de evaluación se debe a la contribución entrópica, por lo que no se pudo obtener una mayor precisión de la predicción de la afinidad de unión.

Por otra parte, en el CASF-2016 las funciones de evaluación basadas en campos de fuerza tuvieron cuatro representantes y mostraron resultados medios, ya que solo en dos de las pruebas hubo un representante de estas funciones, que estuvieron entre las mejores. En la prueba de poder de clasificación, la función GoldScore fue de las mejor evaluadas, obteniendo un coeficiente de correlación de Spearman de 0.7, empatando con las mejores. Asimismo, la función GBVI-WSA-dG fue de las mejor calificadas en la prueba de evaluación del poder de acoplamiento, teniendo una tasa de éxito del 83%<sup>57</sup>. Por otro lado, solamente en las pruebas de poder de acoplamiento y poder de detección hubo un representante que estuvo entre las tres funciones con peor evaluación, y se trata de una misma función llamada D-Score, la cual obtuvo una tasa de éxito del 25%

en la prueba de poder de acoplamiento y un factor de mejora de tres en la prueba de poder de detección<sup>57</sup>.

Funciones de evaluación basadas en el conocimiento.

Las funciones de evaluación basadas en el conocimiento se basan en el principio estadístico inverso de Boltzmann de contactos intermoleculares identificados de un gran conjunto de complejos proteína-ligando. En este caso se asume que la frecuencia de diferentes pares de átomos en diferentes distancias está relacionada con la interacción de dos átomos y transforma la frecuencia para deducir el potencial de fuerza media (PMF) dependiente de la distancia<sup>58,59</sup>. Las ecuaciones centrales se describen a continuación:

$$\Delta G_{unión} = \sum_{kl} A_{ij}(r) \quad (1)$$

$$A_{ij}(r) = -K_b \ln \left( \frac{\rho_{ij}(r)}{\rho_{ij}^*} \right) \quad (2)$$

Aquí el  $\Delta G$  de unión se define como la suma de todos los pares de átomos de proteína-ligando  $kl$  del potencial de interacción entre los átomos de tipo  $ij$  en el radio  $r$  observado. En la ecuación 2,  $\rho_{ij}(r)$  es la densidad numérica del par de átomos  $ij$  a la distancia  $r$ , y  $\rho_{ij}^*$  es la densidad numérica del mismo par de átomos en un estado de referencia, donde se supone que las interacciones interatómicas son cero. Con este enfoque, se asume que la frecuencia de aparición de un contacto por pares es una medida de su contribución energética a la unión proteína-ligando. Si un contacto por pares específico ocurre con más frecuencia que en el estado de referencia, indica una interacción energéticamente favorable entre el par de átomos dado; si ocurre con menos frecuencia, indica una interacción desfavorable<sup>59,63</sup>.

La mayor ventaja de las funciones de evaluación basadas en el conocimiento es que puede usar más complejos proteína-ligando, con respecto a las funciones empíricas, y también es eficiente en cuanto al consumo de recursos computacionales<sup>63</sup>. Sin embargo, la principal limitante es ubicar el estado de referencia, asimismo las funciones de evaluación basadas en el conocimiento son



adecuadas para la predicción de poses de unión, no obstante, tienen deficiencias en la predicción de las afinidades de unión<sup>60</sup>.

En el CASF-2016 participaron cinco funciones de evaluación basadas en el conocimiento. Con excepción de la función DrugScore, el desempeño de estas funciones no fue adecuado. En tres de las cuatro pruebas hubo al menos un representante que estuvo entre las tres peores funciones calificadas. En la prueba de poder de puntuación, la función PMF fue la de peor desempeño, ya que obtuvo un coeficiente de correlación de Pearson de 0.3, así mismo, esta función fue la segunda con tasa de éxito más baja (40%) en la prueba de poder de acoplamiento. También obtuvo el valor más bajo del factor de mejora (3) en la prueba de poder de detección. Por el contrario, la función DrugScore fue de las únicas funciones que logró tener éxito tanto en la prueba de poder de puntuación (coeficiente de correlación de Pearson= 0.65), así como en la de poder de acoplamiento (tasa de éxito del 80%) y en la prueba de poder de detección (factor de mejora = 10) en el CASF-16<sup>57</sup>.

Sheng-You Huang y Xiaoqin Zou<sup>64</sup> utilizaron un método iterativo para desarrollar una función de evaluación basada en el conocimiento (ITScore-PP) a fin de predecir complejos proteína-proteína. La función se derivó usando las estructuras de un conjunto de 851 complejos proteína-proteína que contienen interfaces biológicas verdaderas. El uso de un método iterativo evita el desafiante problema del estado de referencia al derivar potenciales basados en el conocimiento. La función de evaluación se utilizó para predecir el modo de unión de los ligandos nativos en un conjunto diverso de 91 complejos proteína-proteína y obtuvo una tasa de éxito del 98.9% considerando las diez soluciones mejor calificadas. Asimismo, ITScore-PP se probó para determinar la afinidad de unión de un conjunto de 74 complejos proteína-proteína con afinidades conocidas. Los resultados indicaron un buen coeficiente de correlación (0.71) entre la afinidad experimental y la predicha. También se hicieron acoplamientos no unidos y la tasa de éxito fue del 40.7%, este tipo de acoplamiento es desafiante para las funciones de evaluación y en este aspecto aún tiene deficiencias en la predicción.

Aunque las funciones de evaluación convencionales son las que predominan en el campo y han mostrado ser eficaces, su avance se ha visto opacado por las funciones de evaluación desarrolladas con la ayuda de técnicas de aprendizaje automático, tales como  $\Delta_{\text{vinaRF20}}$ , ya que mostró un rendimiento superior a las funciones de evaluación convencionales en el CASF-2016, obteniendo los mejores resultados en las cuatro pruebas.

#### 2.4.1.3 CAPRI (*Critical Assessment of PRediction of Interaction*)

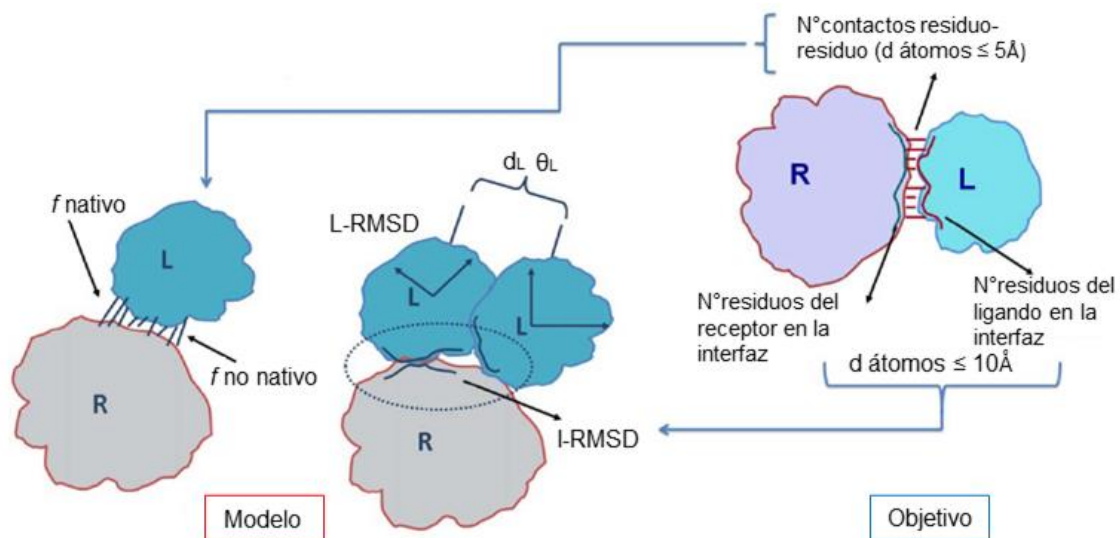
CAPRI es un experimento comunitario implementado en el año 2001 cuyo objetivo es probar el rendimiento de los algoritmos de acoplamiento molecular en predicciones ciegas del modo de asociación entre dos proteínas en función de su estructura tridimensional. Por tal motivo, CAPRI es un factor clave que ha impulsado el progreso de los métodos de acoplamiento molecular proteína-proteína<sup>13-15</sup>.

Aunque los principales objetivos de CAPRI son los complejos proteína-proteína, también se han evaluado complejos proteína-péptido y proteína-DNA para los cuales existe una estructura experimental de los componentes libres<sup>13</sup>.

La evaluación en CAPRI consiste en predecir un complejo cuya estructura tridimensional no ha sido publicada, en donde solo los evaluadores e investigadores que la determinaron la conocen. Se prefiere que las proteínas que forman el complejo tengan una estructura tridimensional conocida en sus formas libres<sup>13,15</sup>.

En el primer día de cada ronda se envían las coordenadas atómicas de los componentes a los diversos predictores registrados y se les asigna un tiempo determinado para enviar sus modelos predichos del complejo en la página web <http://capri.ebi.ac.uk>. Después de la fecha límite, los evaluadores comienzan a evaluar los modelos con respecto a la estructura nativa. Se han establecido principalmente tres métricas para evaluar las propiedades geométricas y biológicas de los modelos (ver figura 1)<sup>13</sup>.

En primer lugar, se determina la fracción de contactos nativos ( $f_{nat}$ ) que son reproducidos en el modelo. Un contacto ligando-receptor se define como cualquier par de átomos de ligando-receptor dentro de una distancia de 5.0 Å. La siguiente métrica es el RMSD del ligando (L-RMSD) y por último el RMSD de los residuos en la interfaz (I-RMSD). Las métricas  $f_{nat}$ , L-RMSD e I-RMSD en conjunto determinan la calidad de un modelo predicho, y con base en estos tres parámetros, los modelos se clasifican en cuatro categorías: calidad alta, calidad media, calidad aceptable e incorrecto, como se muestra en la tabla 2<sup>13,65</sup>.



**Figura 1.** Ilustración esquemática de los criterios de evaluación CAPRI. Donde R y L corresponden al receptor y ligando respectivamente, L-RMSD indica el RMSD del ligando, I-RMSD es el RMSD de la interfaz,  $f_{nativo}$  son los contactos nativos,  $f_{no\ nativos}$  son los contactos no nativos y  $dL$   $\theta_L$  indican el ángulo de mala orientación y el desplazamiento residual del centro de masa del ligando. Modificado de<sup>65</sup>.

Calidad de la predicción	$f_{nat}$	L-RMSD (Å)	I-RMSD (Å)
Alta	$\geq 0.5$	$\leq 1.0$	O $\leq 1.0$
Media	$\geq 0.3$	$1.1 \leq L-RMSD \leq 5.0$	O $1.1 \leq I-RMSD \leq 2.0$
Aceptable	$\geq 0.1$	$5.1 \leq L-RMSD \leq 10.0$	O $2.1 \leq I-RMSD \leq 4.0$
Incorrecto	$< 0.1$	$> 10.0$	Y $> 4.0$

**Tabla 2.** Resumen de los criterios de CAPRI para clasificar los complejos predichos por acoplamiento molecular. Modificado de<sup>65</sup>.

En este trabajo se utilizaron tres servidores de acoplamiento molecular proteína-proteína con diferentes algoritmos de búsqueda y funciones de evaluación que están disponibles en línea de manera gratuita, que han participado en CAPRI y han tenido resultados favorables. El primero de ellos es el servidor PATCHDOCK, el cual estuvo entre los equipos predictores con mejor desempeño en las rondas de CAPRI 13-19, en donde lograron predecir modelos similares al complejo nativo en 6 de 13 objetivos<sup>66</sup>. El siguiente es el servidor ZDOCK, el cual tuvo una demostración exitosa en las rondas de CAPRI 13-19, en general este servidor funcionó bien para tres de los cuatro objetivos en los que tanto el acoplamiento como la puntuación arrojaron al menos una predicción aceptable<sup>67</sup>. Por último, está el servidor HADDOCK, el cual ha mostrado ser de los predictores con mejor desempeño en diversas rondas de CAPRI como en la 20–27, 38-45 y siendo el mejor predictor en la ronda 30 arrojando modelos de calidad muy alta<sup>65,68</sup>.

Las principales características de estos servidores se describen a continuación.

#### 2.4.1.4 PATCHDOCK

PATCHDOCK es un servidor en línea para el acoplamiento molecular proteína-ligando y proteína-proteína. El algoritmo lleva a cabo un acoplamiento rígido basado en la complementariedad de forma local. De manera general PATCHDOCK trabaja de la siguiente manera: en primer lugar, se divide la superficie de las moléculas correspondientes al receptor y ligando en parches cóncavos, convexos y planos de acuerdo con la forma de la superficie. Posteriormente, los parches se filtran, de modo que solo se retienen los parches con residuos “*hot-spots*”. Los *hot-spots* los definen como aquellos residuos que contribuyen con más de 2.0 kcal/mol a la energía libre de unión. Enseguida, los parches se superponen de forma complementaria mediante técnicas de emparejamiento como el *Geometric Hashing* y el *clustering* de poses, de tal manera que los parches cóncavos se combinan con parches convexos y viceversa, mientras que los parches planos se pueden combinar con cualquier tipo de parches, para generar las transformaciones candidatas. Inmediatamente, se examinan las transformaciones candidatas y se descartan aquellas con inserciones inaceptables de los átomos del ligando a los átomos del receptor.

Después los candidatos restantes se clasifican mediante una función de evaluación que considera el ajuste geométrico y la energía de desolvatación atómica. Finalmente, se aplica una agrupación por RMSD para descartar soluciones redundantes<sup>69,70</sup>.

De acuerdo con los autores, la alta eficacia de este algoritmo se debe a la combinación de diversos factores que se enumeran a continuación: 1) enfoca el ajuste inicial de la superficie molecular en parches de superficie localizados, 2) emplea técnicas de detección de patrones espaciales como el *Geometric Hashing* y el *clustering* de poses para generar las transformaciones candidatas, 3) la complementariedad de forma se calcula de manera precisa utilizando la transformada de distancia, 4) hay una detección eficaz de choques estéricos y 5) emplea información biológica enfocándose en parches superficiales ricos en “*hot-spots*”<sup>69,70</sup>.

#### 2.4.1.5 ZDOCK

ZDOCK es un servidor en línea para el acoplamiento molecular proteína-proteína y se basa en el acoplamiento rígido, en donde el muestreo global del espacio traslacional y rotacional se realiza mediante una biblioteca basada en la transformada rápida de Fourier del oeste (FFTW) para calcular la transformada discreta de Fourier en una o más dimensiones y está optimizada para la convolución de alta eficiencia de estructuras moleculares tridimensionales (conv3D)<sup>71</sup>.

La función de evaluación utilizada para calificar las poses de acoplamiento candidatas incluye la complementariedad de forma, un potencial electrostático y un potencial estadístico atómico por pares utilizando las propensiones de contacto de complejos de proteínas transitorias (potencial estadístico basado en el conocimiento)<sup>72</sup>.

#### 2.4.1.6 HADDOCK

*HADDOCK* es un servidor en línea para el acoplamiento molecular proteína-proteína impulsado por alta ambigüedad. Ofrece un enfoque de acoplamiento flexible basado en datos de interacción bioquímica y/o biofísica, tales como datos

de perturbación del cambio químico, que se obtienen de experimentos de titulación por RMN o mutagénesis. Tal información se presenta como restricciones de interacción ambigua (AIR) para impulsar el acoplamiento. El término AIR se refiere a la distancia ambigua entre los residuos que están involucrados en la interacción intermolecular, con un valor máximo de 3.0 Å entre cualquier átomo de un residuo activo de la proteína A y cualquier átomo de residuos tanto activos como pasivos de la proteína B. Los residuos activos y pasivos se definen de acuerdo con el origen de los datos de interacción. En el caso de los datos provenientes de titulación por RMN, los residuos activos son aquellos que sufren una perturbación del cambio químico de manera significativa tras la formación del complejo y además presentan una accesibilidad al disolvente > 50%. Con respecto a los datos de mutagénesis, los residuos activos corresponden a todos los residuos que mediante mutaciones disminuye la formación del complejo y que estén accesibles al disolvente. Por otro lado, los residuos pasivos son aquellos que muestran una perturbación de desplazamiento químico menos importante y/o son vecinos superficiales de los residuos activos y también tienen alta accesibilidad al disolvente<sup>73,74</sup>.

El protocolo de acoplamiento que utiliza *HADDOCK* consta de tres etapas. En la primera etapa, las dos proteínas de entrada se colocan a una distancia de 150 Å entre sí, y cada una se rota de manera aleatoria alrededor de su centro de masa. Posteriormente, se realiza una minimización de energía de cuerpo rígido. En esta etapa se calculan mil conformaciones y luego se afinan las mejores doscientas soluciones. En la segunda etapa se efectúan tres afinamientos con recocido simulado, en donde se permite que las cadenas laterales y el esqueleto de los residuos en la interfaz se muevan. Los residuos a los que se les permite tener movimiento se definen por los residuos activos y pasivos usados en los AIR. La última etapa consiste en un afinamiento con solvente explícito. Después del cálculo, las soluciones se evalúan de acuerdo con la suma de los términos de energía electrostática, de Van der Waals y AIR. Finalmente, las estructuras se agrupan mediante RMSD por pares del esqueleto de los residuos en la interfaz

(iRMSD). Los grupos se forman por al menos dos conformaciones que tienen un iRMSD menor que  $1.0 \text{ \AA}$ <sup>73,74</sup>.

Hasta el momento se han descrito las principales características del acoplamiento molecular *in silico*, así como de tres servidores en línea que realizan acoplamiento molecular. En la siguiente sección se abordará otra técnica para la predicción y caracterización de IPP *in silico*.

#### 2.4.2 Simulación de dinámica molecular (SDM)

La simulación de dinámica molecular es una técnica computacional muy poderosa que permite estudiar una gran variedad de procesos biomoleculares importantes, tales como el plegamiento de proteínas, cambios conformacionales, la unión de ligandos. Así como perturbaciones moleculares debidas a mutaciones, modificaciones postraduccionales, protonación, adición o eliminación de un ligando<sup>75,76</sup>.

Las IPP ocurren principalmente en regiones grandes y poco profundas en la superficie de las proteínas, y las interacciones establecidas a menudo son transitorias, lo que dificulta distinguir las interfaces de unión proteína-proteína. Por lo tanto, es importante considerar la dinámica de las proteínas para determinar qué residuos contribuyen significativamente a la energía libre de unión<sup>77</sup>.

La predicción de residuos cruciales para la formación de los complejos proteína-proteína es una ventaja en el campo del diseño racional de fármacos<sup>77</sup>.

Con las SDM se pueden analizar los aspectos estructurales y funcionales de las IPP, ya que se pueden caracterizar las interfaces proteína-proteína para detectar puntos calientes de unión, también se pueden caracterizar las interacciones importantes entre los socios de unión, la flexibilidad y el muestreo del espacio conformacional de las superficies que interactúan. Por lo tanto, proporciona información relevante sobre los mecanismos de formación de complejos proteína-proteína y apoya al diseño de moduladores de la IPP<sup>78</sup>.

En numerosos estudios han aprovechado el uso sinérgico de las SDM con cálculos de energía libre de unión por los métodos MM-GBSA o MM-PBSA para

predecir la afinidad de unión de los complejos proteína-ligando<sup>79,80</sup> y proteína-proteína<sup>81,82</sup>, y los resultados han mostrado concordancia con los datos experimentales.

Asimismo, se han desarrollado metodologías de mutagénesis de escaneo de alanina *in silico* a partir de trayectorias de dinámica molecular para predecir "puntos críticos" en la interfaz proteína-proteína. La determinación de los puntos críticos se hace mediante la cuantificación de la contribución del aminoácido a la energía libre de unión del complejo proteína-proteína<sup>83</sup>.

Aunque las SDM han mostrado resultados exitosos, es importante ser conscientes de las limitaciones de esta técnica, para hacer un uso razonable de la misma. En primer lugar, las SDM de mecánica molecular no permiten el estudio de procesos químicos, tales como, mecanismos de catálisis. Otra preocupación que surge al utilizar SDM es la validez de los resultados. Un objetivo ideal es reproducir los mismos observables termodinámicos y cinéticos que uno encontraría en los experimentos, sin embargo, esto no se puede saber hasta hacer la validación experimental<sup>77</sup>.

El principal objetivo de la SDM es predecir el movimiento de los átomos de un sistema en función del tiempo, mediante la integración de las ecuaciones de movimiento de Newton para cada átomo. Así que la idea principal de la SDM se basa en que a partir de las posiciones iniciales (coordenadas xyz) de todos los átomos en un sistema, se calcula la fuerza ejercida sobre cada átomo por todos los átomos restantes. Posteriormente, esas fuerzas se utilizan para determinar la velocidad y nueva posición de cada átomo, este cálculo se realiza repetidamente conforme transcurre el tiempo. Al final se genera una trayectoria que describe la configuración del sistema a nivel atómico en cada punto durante el intervalo de tiempo simulado<sup>84</sup>.

#### *2.4.2.1 Campos de fuerza*

Los campos de fuerza son expresiones matemáticas que combinan las funciones de energía potencial de los parámetros estructurales, tales como la deformación del enlace y la geometría del ángulo, torsiones y términos que describen las

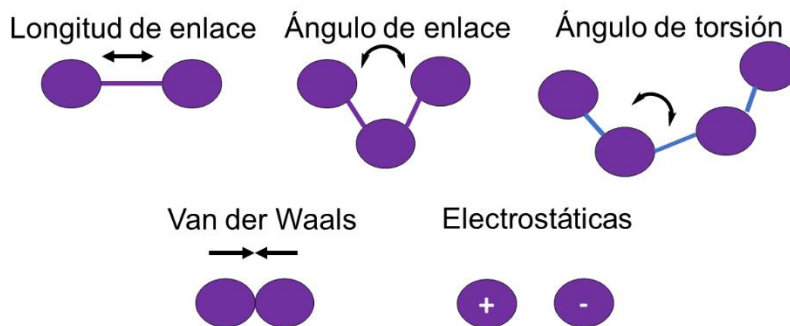


interacciones no enlazantes como las interacciones electrostáticas e interacciones de Van der Waals. Además de un conjunto de parámetros asociados, que describen la dependencia de la energía de un sistema en las coordenadas de sus partículas. Los parámetros se pueden obtener a partir de cálculos de mecánica cuántica, *ab initio* o por medio de datos de experimentos físicos y químicos<sup>85</sup>.

La mecánica molecular solo considera los movimientos nucleares omitiendo los electrones, por lo tanto, las moléculas se definen como un conjunto de átomos que se mantienen unidos por fuerzas elásticas simples (armónicas). Existen diversos campos de fuerza que tienen distintos niveles de complejidad y que están enfocados en modelar diferentes tipos de sistemas. Sin embargo, la expresión de la energía potencial de un sistema molecular que se usa con mayor frecuencia para moléculas orgánicas simples y macromoléculas biológicas es la siguiente:

$$V(r) = \sum_{\text{enlaces}} \frac{k_d}{2} (d - d_0)^2 + \sum_{\text{ángulos}} \frac{k_\theta}{2} (\theta - \theta_0)^2 + \sum_{\text{diédros}} \frac{k_\phi}{2} (1 + \cos(n\phi - \phi_0)) + \sum_{\text{impropios}} \frac{k_\psi}{2} (\psi - \psi_0)^2 + \sum_{LJ} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{electrostáticas}} \frac{q_i q_j}{\varepsilon_D r_{ij}}$$

En donde los primeros cuatro términos se refieren a las contribuciones intramoleculares y los dos últimos términos representan las interacciones repulsivas, de Van der Waals y las interacciones electrostáticas por medio de un potencial de Lennard-Jones y Coulomb, respectivamente<sup>85,86</sup>.



**Figura 2.** Principales términos representados en los campos de fuerza.

### Términos Intramoleculares.

De manera particular, el primer término corresponde al estiramiento de enlace que se representa con un potencial armónico simple que controla la longitud de los enlaces covalentes. En donde  $k_d$  es una constante de fuerza del enlace y  $d_0$  representa la longitud del enlace de referencia. Asimismo, el uso del potencial armónico implica que los enlaces no se pueden romper, por lo tanto, no es posible estudiar procesos químicos. Siendo esta una de las principales limitaciones de la SDM basada en campos de fuerza<sup>85,86</sup>.

El segundo término de la ecuación representa el cambio en la energía potencial debido a la deformación de los ángulos y también suele estar representada por un potencial armónico. El tercer término corresponde a la energía potencial del sistema en función de la rotación alrededor de los ángulos diedros o ángulos de torsión. Estos movimientos de torsión son mucho menos rígidos que los movimientos de estiramiento de enlaces y son imprescindibles para asegurar el grado correcto de rigidez de la molécula y para reproducir los principales cambios conformacionales debidos a las rotaciones sobre los enlaces. La energía de torsión como la que se muestra en la ecuación se representa mediante una función coseno, la  $k_\phi$  determina la altura de la barrera potencial,  $\phi$  representa el ángulo de torsión,  $n$  es el número de mínimos en la función de energía y  $\phi_0$  es el factor de fase, que determina la posición de los mínimos<sup>85,86</sup>.

Finalmente, el término de torsión impropio se utiliza para asegurar la planaridad de los grupos con geometría plana. Por lo que este componente proporciona una función de penalización mediante una contribución positiva a la energía de aquellos movimientos fuera del plano. Donde  $\psi$  es el ángulo impropio correspondiente a la desviación de la planaridad<sup>86</sup>.

### Términos Intermoleculares

Los dos últimos términos en la ecuación se calculan para pares de átomos que se encuentran separados por tres o más enlaces y entre átomos en moléculas diferentes. El primero de estos términos corresponde al potencial de Lennard-

Jones, que se usa frecuentemente para representar las interacciones de Van der Waals. Estas interacciones surgen del equilibrio entre las fuerzas atractivas y repulsivas. La atracción se genera debido a la interacción entre dipolos instantáneos, mientras que el componente repulsivo se debe a la superposición de las nubes electrónicas de los átomos cuya distancia es menor a 0.3 nm<sup>85,86</sup>.

El último término de la ecuación representa las interacciones electrostáticas. El potencial electrostático se representa mediante la suma de los potenciales electrostáticos generados por cargas atómicas parciales utilizando la ley de Coulomb. Existen diversos métodos para determinar las cargas parciales, ya que pueden generarse de datos termodinámicos experimentales y cálculos *ab initio*. Los términos energéticos descritos anteriormente son empleados en muchos campos de fuerza, sin embargo, es posible encontrar términos adicionales como la polarización, potencial de enlace de hidrógeno, entre otros<sup>85,86</sup>.

Existen una gran variedad de campos de fuerza diseñados para estudiar diferentes tipos de sistemas como moléculas orgánicas, proteínas, lípidos, ácidos nucleicos, etcétera. Cada uno tiene sus fortalezas y debilidades, por lo tanto, la elección de cada uno dependerá del problema particular que se quiera estudiar. Entre los campos de fuerza más populares y empleados están TRIPOS<sup>87</sup>, MM2<sup>88</sup>, GAFF<sup>89</sup>, CGenFF (aplicados a moléculas pequeñas)<sup>90</sup>, OPLS<sup>91</sup>, CHARMM<sup>92</sup>, AMBER<sup>93</sup> y GROMOS<sup>94</sup> dirigidos para estudiar macromoléculas. Asimismo, se han desarrollado versiones polarizables de estos últimos cuatro campos de fuerza mencionados. También debe considerarse que estos campos de fuerza evolucionan, ya sea incorporando u optimizando parámetros, por lo tanto, existen diferentes versiones disponibles, por ejemplo, amber99ffsb<sup>95</sup>, CHARMM36m<sup>96</sup>, GROMOS45A3<sup>97</sup>, etcétera.

En el presente trabajo, se han utilizado los campos de fuerza de CHARMM36m y CGenFF, por lo que a continuación se detallan sus funciones.

De manera general, la función de energía potencial de los campos de fuerza de CHARMM<sup>98</sup> para macromoléculas es la siguiente:

$$\begin{aligned}
 V(r) = & \sum_{\text{enlaces}} k_d (d - d_0)^2 + \sum_{\text{ángulos}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{diedros}} k_\phi (1 + \cos(n\phi - \phi_0)) + \\
 & \sum_{\text{impropios}} k_\psi (\psi - \psi_0)^2 + \sum_{\text{Urey-Bradley}} k_{UB} (r_{1,3} - r_{1,3;0})^2 + \sum_{\text{No enlazantes}} \frac{q_i q_j}{\epsilon_D r_{ij}} \\
 & + \epsilon_{ij} \left[ \left( \frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\min,ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{residuos}} V_{CMAP}(\phi, \psi)
 \end{aligned}$$

En la ecuación anterior, la fracción intramolecular de la función de energía potencial incluye los términos para enlaces, ángulos, torsiones, diedros impropios y un término de Urey-Bradley. Este último, es un potencial armónico, en el cual la variación de la energía para la flexión de un ángulo se describe como una función de la distancia entre los átomos en la posición 1-3 en interacciones no enlazantes. Este término proporciona una pequeña mejora en los modos vibracionales de los compuestos. Por otro lado, los términos intermoleculares incluyen las interacciones electrostáticas y de Van der Waals, donde  $R_{\min,ij}$  es el radio en el Lennard-Jones,  $r_{ij}$  es la distancia entre los átomos  $i$  y  $j$  y  $\epsilon_{ij}$  es la profundidad del pozo o *well depth*. El último término incorpora un mapa de energía diedro 2D denominado CMAP, que se utiliza para la corrección torsional del esqueleto en las proteínas y péptidos<sup>98</sup>.

CHARMM36m es un campo de fuerza mejorado para proteínas plegadas e intrínsecamente desordenadas, en donde el potencial CMAP del esqueleto ha sido optimizado, garantizando un ajuste óptimo en las energías de las torsiones del esqueleto de las  $\alpha$ -hélices. Otra mejora involucra la descripción de las interacciones tipo puente salino que se establecen entre los grupos guanidinio y carboxilato empleando un término llamado NBFIX, el cual conduce a interacciones tipo puente salino más realistas. Además, asegura que solo se modifique las interacciones entre arginina-aspartato y arginina-glutamato, entre tanto las interacciones de estos residuos con otros aminoácidos, agua o iones no se ven modificadas<sup>96</sup>.

Por otra parte, el campo de fuerza general de CHARMM, llamado CGenFF, fue desarrollado para simular moléculas orgánicas similares a fármacos, siendo

compatible con los campos de fuerza de CHARMM para macromoléculas, por lo tanto, utiliza una función de energía potencial similar a los otros campos de fuerza de CHARMM. Este campo de fuerzas es extensible a grupos químicos que no se encontraban en el campo de fuerza. En consecuencia, se espera que CGenFF tenga una cobertura más completa de espacio químico. Finalmente, se ha mostrado que este modelo reproduce con precisión datos geométricos, vibracionales y energéticos<sup>90</sup>.

#### 2.4.2.2 Métodos de integración

Las ecuaciones de movimiento de Newton se integran usando algoritmos desarrollados con base en los métodos de diferencias finitas. Este método consiste en sustituir un intervalo de tiempo infinitesimal  $dt$  por un intervalo de tiempo finito  $\Delta t$ , por lo general, este intervalo comprende entre uno y diez femtosegundos ( $10^{-15}$  segundos). De esta manera las ecuaciones de movimiento se resuelven paso a paso, integrándolas en cada intervalo  $\Delta t$ . En cada paso se calculan las fuerzas sobre los átomos, en donde se asume que son constantes en este intervalo de tiempo, y se combinan con las velocidades y posiciones de ese momento para generar un nuevo estado. Una vez sucedido esto se actualizan las fuerzas ejercidas sobre cada átomo, y así sucesivamente hasta generar toda la trayectoria<sup>99</sup>.

Uno de los algoritmos más utilizados es el algoritmo de Verlet<sup>100</sup>, ya que permite evaluar en el mismo instante las posiciones y velocidades de todos los átomos respetando las simetrías fundamentales de las ecuaciones de Newton, asimismo ofrece una buena estabilidad para pasos de tiempo relativamente grandes. Este algoritmo resulta de la combinación de dos series de Taylor para las posiciones del paso nuevo ( $t+\Delta t$ ) y previo ( $t-\Delta t$ ) como se muestra a continuación:

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{1}{2} \frac{d^2x(t)}{dt^2} \Delta t^2 + \frac{1}{3!} \frac{d^3x(t)}{dt^3} \Delta t^3 + \dots$$

$$x(t - \Delta t) = x(t) - \frac{dx(t)}{dt} \Delta t + \frac{1}{2} \frac{d^2x(t)}{dt^2} \Delta t^2 - \frac{1}{3!} \frac{d^3x(t)}{dt^3} \Delta t^3 + \dots$$

Al sumar ambas series se obtiene el algoritmo de Verlet para la posición:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + a_x(t)\Delta t^2$$

Donde “a” corresponde a la aceleración que se obtiene mediante la derivada de la función de energía potencial respecto a las coordenadas. Una característica de este algoritmo es que las velocidades no son necesarias para generar las trayectorias. Sin embargo, si son requeridas para determinar la energía cinética del sistema, por lo tanto, la ecuación empleada para conocer la velocidad es la siguiente:

$$v_x(t) = \frac{x(t + \Delta t) - x(t - \Delta t)}{2\Delta t}$$

#### 2.4.2.3 Ensamblés

Para comenzar una SDM, se requieren datos iniciales como la geometría del sistema, el potencial de interacción y algunas constantes que definen el colectivo de microestados sujetos a una restricción extensiva o también conocido como ensamble<sup>99,101</sup>. Existen diferentes tipos de ensamble que son usados en dinámica molecular, los cuales se listan en la tabla 3.

Ensamble	Grados de libertad	Condición
Microcanónico	N, V, E	Energía constante
Canónico	N, V, T	Temperatura constante
Isobárico-Isoentálpico	N, P, H	Entalpía constante
Isobárico-Isotérmico	N, P, T	Temperatura y presión constante

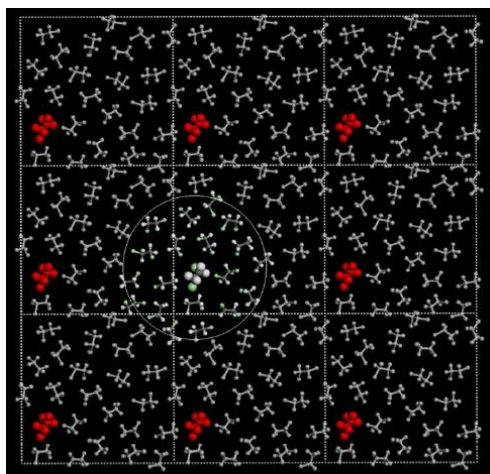
**Tabla 3.** Tipos de ensambles usados en dinámica molecular, donde N es el número de partículas, V es el volumen, E es la energía, T corresponde a la temperatura y H a la entalpía. Recuperado de<sup>101</sup>.

#### 2.4.2.4 Condiciones periódicas de la frontera (periodic boundary conditions, PBC)

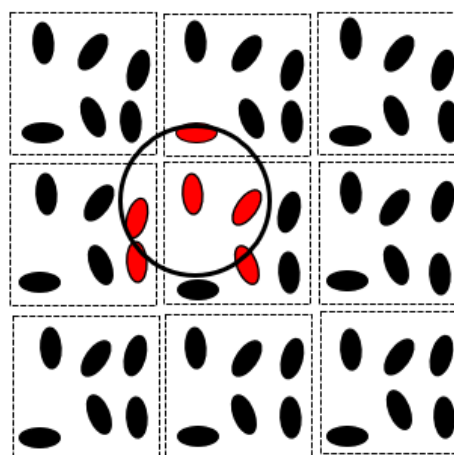
La implementación de condiciones periódicas de la frontera consiste en simular un pequeño número de átomos dentro de una celda principal, que está rodeada por réplicas de esta, de tal manera que se genera un sistema periódico infinito (ver figura 3). La celda principal es aquella que contiene las moléculas del sistema que se van a simular. Existen diferentes tipos de geometrías de celda que suelen

emplearse en SDM, tales como cúbica, octaedro truncado, prisma hexagonal y dodecaedro rómbico principalmente<sup>85,102</sup>.

Solo los átomos que se encuentren en la celda principal se toman en cuenta explícitamente. Sin embargo, cuando un átomo abandona la celda principal por una de sus caras, su imagen de la cara opuesta entra a la celda principal para remplazarlo. Un procedimiento que surge como consecuencia del uso de PBC es el criterio de mínima imagen, el cual considera que un átomo debe interactuar solamente con una de las imágenes de otros átomos, la cual debe ser la más cercana al átomo original, esto puede ilustrarse en la figura 4<sup>85,99</sup>.



**Figura 3.** Representación esquemática de las PBC. Recuperado de<sup>85</sup>.



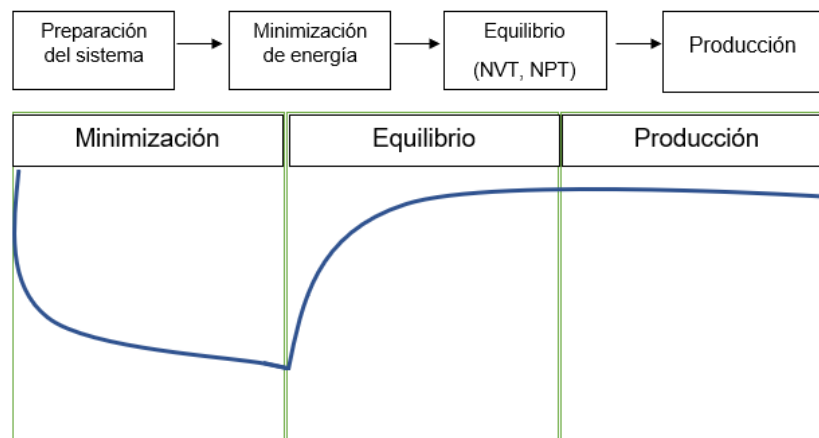
**Figura 4.** Criterio de convención de mínima imagen.

#### 2.4.2.5 Etapas de las SDM

Las diferentes etapas que componen una SDM (ver figura 5), se describen a continuación:

En la preparación del sistema se establece la conformación inicial, se elige el campo de fuerza adecuado para el sistema, y se añaden moléculas de agua y iones. Posteriormente, se realiza un proceso de minimización de la energía, con el objetivo de ajustar la geometría de la estructura, de tal manera, que se encuentre una conformación energéticamente estable. En la etapa de equilibrio, se busca un equilibrio energético, en donde la energía permanece casi constante. Se pueden

emplear diferentes ensambles (ver tabla 3), sin embargo, los más usados son NVT (número constante de partículas “N”, volumen “V” y temperatura “T”) y NPT (número constante de partículas “N”, temperatura “T” y presión “P”), efectuándose dos etapas de equilibrio. Una vez que se ha alcanzado el equilibrio a temperatura y presión deseadas, se realiza la producción de la dinámica, la cual puede ir desde picosegundos hasta segundos, la elección del tiempo de simulación dependerá del tipo de sistema que se está estudiando. Esta última etapa es la que requiere mayor costo computacional, ya que se realiza el cálculo de fuerzas y se genera la trayectoria resultante.



**Figura 5.** Etapas de una simulación de dinámica molecular. Modificado de<sup>101</sup>.

#### 2.4.2.6 Softwares de SDM

Existe una gran variedad de programas para realizar SDM. Entre las opciones más populares y confiables están GROMACS<sup>103,104</sup>, NAMD<sup>105</sup>, AMBER<sup>106,107</sup>, CHARMM<sup>108</sup>, los cuales no deben confundirse con los campos de fuerza. La mayoría de los softwares de dinámica molecular admiten múltiples campos de fuerza, también cada programa contiene herramientas de análisis que permiten calcular una amplia gama de propiedades a partir de las trayectorias. Asimismo, existen programas auxiliares que permiten un análisis más detallado de las propiedades medidas en las trayectorias, tales como VMD<sup>109</sup>, R<sup>110</sup>, MDAnalysis<sup>111,112</sup>, g\_mmpbsa<sup>113</sup>, entre otros.



#### 2.4.2.7 Análisis de trayectorias de SDM

Una vez que se ha ejecutado una SDM, es importante hacer un adecuado análisis de los datos que se pueden extraer de las trayectorias para poder responder las interrogantes que se tienen sobre el sistema en cuestión. A continuación, se describirán algunas métricas de evaluación que permiten determinar el comportamiento de un sistema simulado.

Raíz de la desviación cuadrática media (**Root Mean Squared Deviation**, RMSD)

El RMSD es la medida más utilizada para determinar la similitud estructural en biología. Entre sus principales aplicaciones destaca el monitoreo de cambios estructurales en simulaciones de plegamiento y dinámica de proteínas, evaluar la calidad de esquemas de predicción de estructuras, comparar la diversidad de estructuras de modelos derivadas de experimentos, etcétera<sup>114</sup>.

Esta medida consiste en calcular la distancia media cuadrática mínima de los átomos de una estructura con respecto a otra que tenga los mismos átomos, empleando las coordenadas atómicas. La ecuación con la que se obtiene el valor de RMSD es la siguiente:

$$RMSD (S_i, S_j) = \min \sqrt{\frac{1}{N} \sum_{k=1}^N \left\| \vec{r}_{ik} - \vec{r}_{jk} \right\|^2}$$

Donde N es el número de átomos en la estructura, los subíndices i,j indican estructuras diferentes y r es un vector que representa las coordenadas espaciales de un átomo dado. Se ha establecido que dos estructuras son muy semejantes si su valor de RMSD es menor a 2.0 Å<sup>114,115</sup>.

El RMSD en SDM de proteínas puede determinarse con los átomos del esqueleto, carbonos α, en donde se pueden comparar las conformaciones a los diversos tiempos con respecto a la conformación de inicio, de tal manera que se puede monitorear los cambios que se generaron durante el periodo de simulación.

Otra manera de realizar un análisis de RMSD de una trayectoria es haciendo una comparación de todos contra todos o *Pairwise RMSD*. De esta forma se puede obtener información sobre la similitud mutua, y poder observar los periodos de tiempo en los que existe similitud y variación estructural entre sí.

Raíz de la fluctuación cuadrática media (**Root Mean Squared Fluctuation**, RMSF)

A diferencia de RMSD, RMSF es una medida que permite analizar la flexibilidad de una proteína, ya que representa la fluctuación u oscilación de las posiciones atómicas de sus aminoácidos. Por lo tanto, con esta métrica se puede determinar cuáles son las regiones más móviles de una proteína y como se ve afectada esta movilidad en presencia de un ligando u otra proteína. La ecuación con la que se obtiene el RMSF se muestra a continuación<sup>115</sup>:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{j=1}^T \left( \vec{r}_{ij} - \frac{1}{T} \sum_{k=1}^T \vec{r}_{ik} \right)^2}$$

Donde T es el número total de *frames* o conformaciones, r es un vector de la posición espacial de un átomo i en los *frames* j y k.

Radio de giro ( $R_g$ )

El radio de giro ( $R_g$ ) es un indicador que describe la conformación de equilibrio de un sistema, el cual determina la compacidad de una proteína. El empaquetamiento espacial de los residuos es un factor importante en la estabilidad de las proteínas. La variación del empaquetamiento estará determinada por el tamaño de la proteína, la composición de aminoácidos y el tipo de estructura secundaria que forman<sup>116</sup>.

El  $R_g$  de un grupo de átomos se define como la distancia cuadrática media de cada átomo de la molécula a su centro de masa, tal como se describe en la siguiente ecuación<sup>117</sup>:

$$R_g = \sqrt{\left\{ \sum_{j=1}^N \left[ r_j - \left( \sum_{i=1}^N r_i / N \right) \right]^2 / N \right\}}$$

Donde  $r_i$  y  $r_j$  son los vectores de posición de los átomos  $i$  y  $j$ , y  $N$  es el número de átomos.

### Energía libre de unión

La determinación de la afinidad de unión de los complejos moleculares mediante cálculos de energía libre de unión utilizando enfoques computacionales es de gran importancia, ya que permite conocer o identificar moléculas novedosas que puedan unirse a receptores diana y actuar como fármacos terapéuticos. Asimismo, permite estudiar el reconocimiento molecular, el cual juega un papel importante en las vías de transducción de señales, etcétera. Estos enfoques son el de integración termodinámica (TI), perturbación de la energía libre (FEP), energía de interacción lineal (LIE), mecánica molecular del área de superficie de Poisson-Boltzmann (MM-PBSA) y del área de superficie generalizada de Born (MM-GBSA). Sin embargo, TI y FEP son los más estrictos, pero también los más costosos computacionalmente, por otro lado, LIE y MM-PBSA son computacionalmente más eficientes, ya que determinan la energía libre empleando un conjunto de estructuras en los estados inicial y final<sup>113,118</sup>.

En este trabajo se decidió optar por el enfoque de mecánica molecular del área de superficie de Poisson-Boltzmann (MM-PBSA). Este enfoque es uno de los más populares y utilizados para calcular la energía de unión de complejos biomoleculares, tales como proteína-ligando, proteína-proteína, DNA-proteína, etcétera. MM-PBSA combina conformaciones de SDM con solvente explícito junto con estimadores de energía basados en modelos de solvente implícito. El cambio de la energía libre empleando este enfoque considera tres términos energéticos. El primero corresponde al cambio de la energía potencial en el vacío, el cual considera términos enlazantes como energía de enlace, ángulo y torsión, así como términos no enlazantes como energías de Van der Waals e interacciones

electrostáticas. El siguiente término involucra a la energía de solvatación polar y no polar usando modelos de solvatación implícita. Finalmente, el tercer término considera las contribuciones entrópicas<sup>113</sup>.

Una herramienta que emplea el método de MM-PBSA para la estimación de la energía libre de unión, es el software g\_mmpbsa<sup>113</sup>, el cual es compatible con el paquete de simulación de GROMACS y es de acceso libre.

Este software considera, en términos generales, que la energía libre de unión para un complejo está dada por la siguiente ecuación:

$$\Delta G_{unión} = \Delta G_{complejo} - (\Delta G_{proteína} + \Delta G_{ligando}) \quad (1.0)$$

Por otro lado, el  $\Delta G$  de cada componente de la ecuación 1.0 está dado por:

$$\Delta G_x = \langle E_{MM} \rangle - TS + \langle \Delta G_{solvatación} \rangle \quad (2.0)$$

Donde x es la proteína, ligando o complejo proteína-ligando,  $\langle E_{MM} \rangle$  es la energía potencial promedio de la mecánica molecular en el vacío. El término TS se refiere a la contribución entrópica a la energía libre en el vacío, donde T y S corresponden a la temperatura y la entropía, respectivamente. El último término es la energía libre de solvatación. Con respecto a EMM esta se determina con la siguiente ecuación:

$$E_{MM} = E_{Enlazantes} + (E_{vdw} + E_{elec}) \quad (2.1)$$

Donde  $E_{enlazantes}$  corresponde a las interacciones de enlace, ángulo, diedros e impropios, mientras que  $E_{vdw}$  y  $E_{elec}$  se obtienen empleando una función de potencial de Coulomb y de Lennard-Jones respectivamente.

La energía libre de solvatación se calcula empleando un modelo de solvente implícito y esta se estima con la siguiente ecuación:

$$\Delta G_{solvatación} = \Delta G_{polar} - \Delta G_{no polar} \quad (2.2)$$

Donde  $\Delta G_{polar}$  y  $\Delta G_{no polar}$  corresponden a las contribuciones electrostáticas y no electrostáticas a la energía libre de solvatación.  $\Delta G_{polar}$  representa el término electrostático y se obtiene resolviendo la ecuación de Poisson-Boltzmann (PB)

$$\nabla[\epsilon(r)\nabla\varphi(r)] - \epsilon(r)K(r)^2\sinh[\varphi(r)] + \frac{4\pi\rho^f(r)}{kT} = 0 \quad (2.2.1)$$

Donde  $\varphi(r)$  es el potencial electrostático,  $\epsilon(r)$  es la constante dieléctrica, y  $\rho^f(r)$  es la densidad de carga fija. El término  $K(r)^2$  depende de la fuerza iónica de la solución.

Por otra parte, el término no electrostático  $\Delta G_{no\ polar}$  puede estimarse utilizando varios modelos, sin embargo, en este trabajo se utilizó el modelo de área de superficie accesible al solvente (SASA). Este modelo se basa en el supuesto de que el SASA tiene una dependencia lineal de  $\Delta G_{no\ polar}$ , por lo tanto, este término se estima de la siguiente manera:

$$\Delta G_{no\ polar} = \gamma A + B \quad (2.2.2)$$

Donde  $\gamma$  es un coeficiente relacionado con la tensión superficial, A es el SASA y B es un parámetro de ajuste.

Finalmente, este software también puede realizar la descomposición de la energía de unión para determinar la contribución energética por residuo. En primer lugar, los términos energéticos  $E_{MM}$ ,  $G_{polar}$  y  $G_{nonpolar}$  de los átomos individuales se calculan cuando se encuentran unidos y no unidos y después se determina su contribución a la energía de unión del residuo, como se muestra a continuación:

$$\Delta R_x^{BE} = \sum_{i=0}^n (A_i^{unido} - A_i^{libre}) \quad (3.0)$$

Donde  $A_i$  unido y  $A_i$  libre corresponden a la energía del i-ésimo átomo del residuo x en las formas unidas y no unidas respectivamente, mientras que n es el número total de átomos en el residuo.

La contribución energética de todos los residuos debe ser igual a la energía libre de unión ( $\Delta G_{unión}$ ).

Hasta el momento se ha mencionado la importancia de las proteínas en los procesos biológicos de los seres vivos, así como la relevancia de las IPP y se han descrito algunos métodos experimentales y computacionales para poder

identificarlas. Ahora se trasladará esta información para el estudio de una proteína de interés con relevancia fisiológica y clínica.

## **2.5 AKR1B1 como blanco terapéutico**

### 2.5.1 Superfamilia de las aldo-ceto reductasas (AKR's).

Las aldo-ceto reductasas (AKR's) son una superfamilia de enzimas que se agrupan en quince familias de acuerdo con el porcentaje de identidad entre ellas. Una de sus principales características es que catalizan reacciones de óxido-reducción involucradas en la biosíntesis, metabolismo intermediario y la detoxificación. Los principales sustratos sobre los que actúan son aldehídos con relevancia biológica como: glucosa, esteroides, productos de la peroxidación lipídica y productos finales de la glicosilación<sup>1</sup>.

Dentro de la primera familia de las AKR's, se encuentra la subfamilia de las aldosas reductasas (AKR1B), que son enzimas citoplasmáticas que catalizan la reducción de varios aldehídos. Las proteínas que son miembros de esta subfamilia tienen una alta similitud estructural y un porcentaje de identidad de secuencia mayor al 70%.

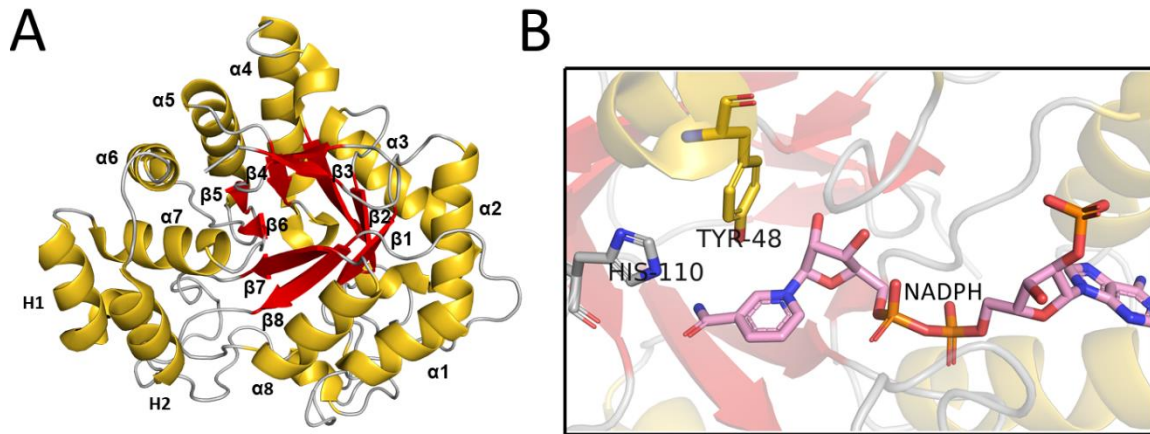
En humanos hay dos miembros de esta subfamilia que han sido muy estudiados y caracterizados, la aldosa reductasa (AKR1B1) y la reductasa del intestino delgado (AKR1B10). Los genes de AKR1B en humanos se encuentran en el cromosoma 7q33-35 y son expresados principalmente en intestino delgado, colon, hígado, timo y glándulas suprarrenales. Por otro lado, hay tres miembros de esta subfamilia que son de origen murino, la aldosa reductasa (AKR1B3), proteína del conducto deferente dependiente de andrógenos "MVDP" (AKR1B7) y factor de crecimiento 1 dependiente de fibroblastos "FR-1" (AKR1B8). Se sabe que la función principal de AKR1B1 y AKR1B10 es la reducción de glucosa a sorbitol y de retinal a retinol respectivamente, mientras que la aldosa reductasa murina cataliza la reacción de conversión de D-treo-aldosa a D-treo-aldono-1,5-lactona. El miembro AKR1B7 reduce principalmente prostaglandinas. Se ha analizado que la diferencia en la especificidad de sustrato entre estas enzimas se debe a una pequeña variabilidad

en los residuos que componen el extremo C-terminal de la estructura tridimensional<sup>119,120,121</sup>.

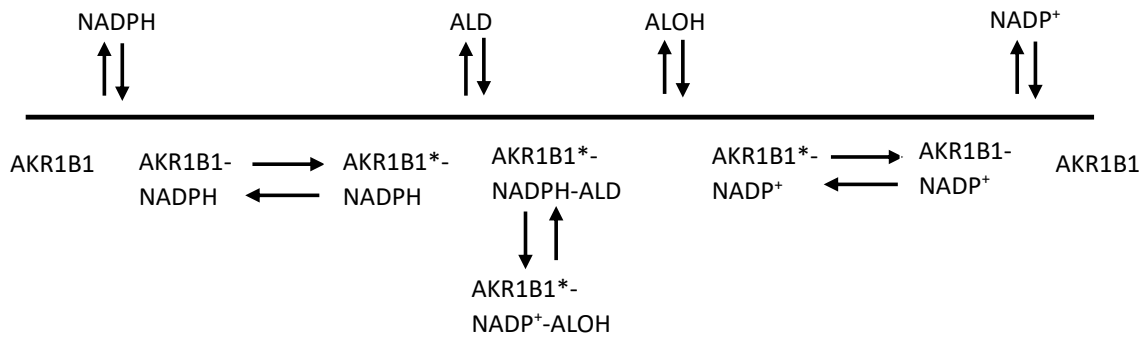
En este trabajo nos enfocaremos en la aldosa reductasa humana AKR1B1, ya que se ha convertido en un blanco terapéutico muy prometedor para el tratamiento de diversas patologías. En el siguiente apartado se abordará más a fondo sobre las características estructurales, fisiológicas y clínicas de esta proteína.

### 2.5.2 Aldosa Reductasa (AKR1B1): estructura y función

La aldosa reductasa humana AKR1B1 contiene 315 residuos de aminoácidos y tiene una masa molecular de 36 kDa. Adopta un plegamiento estructural de tipo barril  $\beta/\alpha$ , el cual está compuesto por ocho láminas betas paralelas y ocho segmentos periféricos con estructura  $\alpha$ -helicoidal adyacentes. El sitio catalítico se encuentra en el núcleo del barril (ver figura 6). Con respecto al mecanismo de catálisis para la reducción de aldehídos, sigue una ruta secuencial y ordenada (ver figura 7), en donde el NADPH se une primero, subsecuentemente lo hace el sustrato. Al darse la unión de la enzima con el NADPH se induce un pequeño cambio conformacional, generando un movimiento para cubrir una fracción de la coenzima, de tal manera que se asegura como si fuese un cinturón de seguridad. El producto se forma en dos pasos. En primer lugar, se realiza la transferencia del hidruro pro-R (proquiralidad R) del NADPH al carbono carbonilo del sustrato. Luego se dona un protón para reducir el grupo carbonilo a un alcohol. Se han hecho numerosos estudios de simulaciones de mecánica cuántica combinadas con mecánica molecular QM/MM, para establecer qué residuo de la enzima actúa como donante de protones, y se ha establecido que la Tyr48 e His110 pueden desempeñar esta función. Sin embargo, algunos autores han concluido que la Tyr48 sea el donador de protones más probable. Una vez que se ha liberado el alcohol se genera un segundo cambio conformacional para liberar la coenzima oxidada<sup>122,123</sup>.



**Figura 6.** Estructura tridimensional de AKR1B1 (código PDB:6F7R). Panel A representación esquemática de diagrama de listón de AKR1B1. El panel B muestra el sitio catalítico de la enzima, destacando los residuos como posibles donadores de H (Tyr48 e His110) y el cofactor NADPH en representación de bastones. Figura generada con PyMOL.

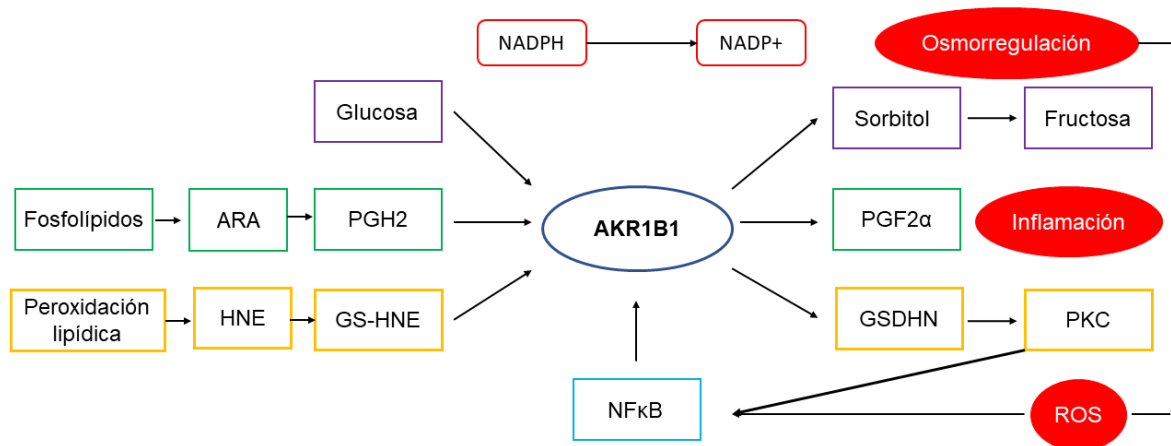


**Figura 7.** Mecanismo de reacción secuencial y ordenado para la reducción de aldehídos por medio de AKR1B1. Nota: AKR1B1\* representa el cambio conformacional después de la unión de NADPH, ALD y ALOH representan a las moléculas de aldehído y alcohol respectivamente. Modificado de<sup>123</sup>.

Las principales funciones de la aldo-ceto reductasa humana AKR1B1 en tejidos normales, consiste en la reducción de glucosa a sorbitol dependiente de NADPH de manera irreversible (ver figura 8). Esta es la primera reacción de la vía de los polioles y constituye el paso limitante de esta vía. En condiciones glucémicas normales solo un pequeño porcentaje de glucosa se metaboliza en esta vía ( $k_{cat}/k_m^{glucosa} AKR1B1 = 6.9 M^{-1}s^{-1}$ )<sup>124</sup>, debido a que la mayoría se fosforila por la hexocinasa “HK” ( $k_{cat}/k_m^{glucosa} HK1 = 1.7 \cdot 10^6 M^{-1}s^{-1}$ ;  $HK2 = 0.9 \cdot 10^6 M^{-1}s^{-1}$ ,  $HK3 = 4.2 \cdot 10^6 M^{-1}s^{-1}$  y  $HK4 = 8.2 \cdot 10^3 M^{-1}s^{-1}$ )<sup>125</sup>, produciendo glucosa-6-fosfato, el cual es



un intermediario importante en la glucólisis y vía de las pentosas fosfato. Asimismo, esta enzima también se encarga de reducir los productos de la peroxidación lipídica que se conjugan con glutatión (GSH), generando especies involucradas en la activación de la vía del factor nuclear NFκB. Finalmente, AKR1B1 participa en el metabolismo de prostaglandinas, catalizando la conversión de PGH2 en PGF2α<sup>126</sup>.



**Figura 8.** Principales reacciones que cataliza AKR1B1. Nota (ARA: ácido araquidónico, PG: prostaglandina, HNE: 4 hidroxinonal, GS-HNE: 3-glutatiónil-4-hidroxinonal, GSDHN: 3-glutatiónil-1,4-dihidroxinonano, PKC: Proteína cinasa C y ROS: especies reactivas de oxígeno). Modificado de<sup>126</sup>.

### 2.5.3 AKR1B1 como blanco farmacológico en diversas patologías

#### 2.5.3.1 Complicaciones Diabéticas

Esta enzima es de importancia clínica debido a que en condiciones de hiperglucemia hay una actividad incrementada de la vía del polirol, lo que genera una mayor formación de sorbitol y consumo de NADPH que conlleva a estrés osmótico y oxidativo, lo cual origina diversas complicaciones diabéticas como retinopatía y nefropatía. La generación de especies reactivas de oxígeno (ROS) mediante la vía del polirol en condiciones de hiperglucemia puede deberse a diversos mecanismos. En primer lugar, hay un alto consumo de NADPH debido a la alta actividad de AKR1B1, el cual es un cofactor necesario para algunas enzimas encargadas de la defensa antioxidante, tales como la glutatión reductasa

que se encarga de catalizar la reducción de glutatión oxidado (GSSG) a glutatión reducido (GSH), que es utilizado por la enzima glutatión peroxidasa para la reducción del peróxido de hidrógeno. En segundo lugar, la conversión de sorbitol en fructosa genera NADH, el cual es un sustrato de la NADH oxidasa que conduce a la formación de anión superóxido, asimismo la fructosa puede aumentar la producción de productos finales de glicación avanzada (AGE), que eventualmente conlleva a la formación de ROS<sup>2,127</sup>.

#### *2.5.3.2 Rol de AKR1B1 en enfermedades inflamatorias*

AKR1B1 juega un papel relevante en diferentes patologías inflamatorias como sepsis, asma, aterosclerosis, entre otras, debido a que produce mediadores inflamatorios al reaccionar con los productos de la peroxidación lipídica, lo cual da lugar a la activación de cascadas de señalización inflamatorias<sup>2,128</sup>. Por ejemplo, se conoce que esta enzima metaboliza los productos de la peroxidación lipídica, tales como el 4-hidroxinonanal (HNE), el cual puede conjugarse con glutatión produciendo 3-glutatiónil-4-hidroxinonanal (GS-HNE), que al reducirse mediante la actividad de AKR1B1 se convierte en 3-glutatiónil-1,4-dihidroxinonano (GS-DHN), lo que a su vez puede promover la activación de la vía de la fosfolipasa C/proteína cinasa C (PLC-PKC), estimulando la producción del factor nuclear NFκB. Este último puede actuar como factor tisular para la expresión de la ciclooxigenasa 2 (COX2), que se encarga de convertir el ácido araquidónico en prostaglandina H2 (PGH2), que posteriormente se convierte en prostaglandina F2α (PGF2α) mediante AKR1B1. El aumento en PGF2α favorece el proceso de inflamación<sup>126,127</sup>.

Finalmente, se ha estudiado que el factor nuclear NFκB es regulador de varios genes involucrados en la inflamación y cáncer.

#### *2.5.3.3 AKR1B1 en cáncer.*

Debido a las diversas funciones de AKR1B1 en diferentes procesos como el metabolismo de la glucosa, productos de peroxidación lipídica y síntesis de prostaglandinas, que conllevan a la formación de ROS y especies inflamatorias, se ha relacionado la expresión de esta enzima con diferentes tipos de cáncer (ver tabla 4)<sup>4</sup>. Se ha reportado que AKR1B1 se sobre-expresa y tiene actividad

incrementada en diferentes tipos de cáncer como cervical, mama, ovario, colon, entre otros (ver tabla 5)<sup>129</sup>. Diversos estudios han determinado que la expresión de esta enzima está fuertemente relacionada con la transición epitelio-mesénquima (EMT) en la que las células cancerosas adquieren un fenotipo invasivo y quimiorresistente<sup>5,6</sup>.

<b>Tipo de cáncer</b>	<b>Fold change(s) vs tejido normal, por estudio</b>	<b>Valor p global</b>
Cerebro (astrocitomas)	2.31, 1.96, 1.96, 1.39, 1.27, 1.22	0.002
Cerebro (glioblastomas)	2.15, 1.31, 1.23, 1.21, -1.64	0.002
Cervical	2.90, 2.12, 2.09, 1.14	1.32E-06
Colon	1.07, -1.50, -1.28	0.118
Esofágico	4.52, 2.99, 1.88, 1.54, 1.29	9.19E-04
Gástrico	1.18, 1.06, 1.04	0.116
Cabeza y cuello	2.61, 1.77, 1.76, 1.41	4.44E-04
Riñón	3.11, 3.00, 2.85, 2.59, 2.42, 2.01, 1.90	3.61E-05
Leucemia (aguda de células B)	5.32, 2.39, 2.31, 1.92	6.95E-36
Leucemia (aguda de células T)	4.52, 1.52, -1.02	1.91E-13
Leucemia (crónica)	1.27, 1.26, -1.48	1.47E-09
Hígado	2.19, 1.28, 1.22, 1.09	0.066
Pulmón (adenocarcinoma)	1.02, -2.43, -1.19, -1.25	0.374
Pulmón (escamoso)	1.42, -1.24, -1.11	0.593
Linfoma	1.86, 1.33, 1.26	0.016
Melanoma	2.43, 1.74, 1.03	0.006
Páncreas	1.74, 1.74, 1.43, 1.41, 1.35	0.069

Vejiga	1.77, 1.26, 1.09	0.007
--------	------------------	-------

**Tabla 4.** Expresión de mRNA de AKR1B1 en diferentes tipos de tumores y neoplasias hematológicas contenidas en la base de datos *Oncomine*. Nota: (el valor de p fue determinado con la prueba estadística de *t-student*, cuyos valores menores a 0.05 se consideran significativos). Modificado de<sup>4</sup>.

<b>Tumor</b>	<b>Inmuno-detección (unidades de densitometría)</b>	<b>Actividad específica (<math>\mu</math>mol NADPH oxidado/h/100 mg proteína)</b>
<i>Área no tumoral</i>		
Mama (5)	~40	12.7 $\pm$ 3.19
Cérvix (4)	~25	13.3 $\pm$ 3.07
Ovario (4)	~35	14.7 $\pm$ 3.61
Colon (4)	~15	10.7 $\pm$ 4.49
<i>Área tumoral</i>		
Mama (5)	~65	26.1 $\pm$ 5.29*
Cérvix (4)	~35	25.4 $\pm$ 2.58*
Ovario (4)	~45	26.5 $\pm$ 2.37*
Colon (4)	~20	32.2 $\pm$ 7.06*

**Tabla 5.** Expresión y actividad específica de AKR1B1 en áreas no tumorales y tumorales de tejidos cancerosos humanos (nota: los valores corresponden a la media  $\pm$  desviación estándar y el número de muestras se da entre paréntesis. Asimismo, el \* indica que la actividad específica en muestras tumorales es significativamente diferente de los valores del área no tumoral (prueba t de *Student*,  $p < 0,05$ )). Reportado por<sup>129</sup>.

En el estudio de Xuebiao Wu y colaboradores<sup>6</sup> mostraron que existe una correlación entre la sobre-expresión de AKR1B1 con el cáncer de mama triple negativo en donde se predice un mal pronóstico. También se determinó que la expresión de AKR1B1 altera la expresión de marcadores de EMT, para esto se analizó la expresión de E-cadherina que es un marcador molecular de tejido epitelial en células MDA-MB231 (células epiteliales de cáncer de mama triple

negativo) y SUM159 (células mesenquimales de cáncer de mama triple negativo) con el gen de AKR1B1 expresado y silenciado.

Los resultados obtenidos mostraron que el silenciamiento de la expresión de AKR1B1 provocó un aumento significativo en la expresión de E-cadherina en ambas líneas celulares. Por el contrario, en las células que expresaban AKR1B1, la E-cadherina se encontró a la baja, lo cual se vio reflejado en una mejora de la migración e invasión de las células de cáncer de mama en algunos ensayos. Estos resultados sugieren que la expresión de AKR1B1 favorece el fenotipo de EMT. Asimismo, propusieron que AKR1B1 regula la EMT a través de un circuito de retroalimentación positiva mediado por Twist2, el cual actúa como un activador transcripcional para inducir directamente la transcripción de AKR1B1. En las células de cáncer de mama triple negativo, la expresión de AKR1B1 se regula por el incremento de Twist2, y como ya se ha mencionado anteriormente, la expresión de AKR1B1 conlleva a la activación del factor nuclear NFκB, que a su vez incrementa la expresión de Twist2, induciendo la EMT.

En otro estudio realizado por Paolo Ceppi y colaboradores<sup>5</sup> también se mostró la correlación entre la expresión de AKR1B1 y el fenotipo EMT en células cancerosas. Analizaron la expresión de marcadores de EMT debido a la expresión de AKR1B1 y lo relacionaron con la capacidad migratoria celular, tasa de crecimiento y formación de colonias, debido a que son características distintivas de la EMT. Para esto analizaron la expresión de E-cadherina y vimetina que es un marcador molecular de tejido mesenquimal en células A549 (células epiteliales de cáncer de pulmón) con el gen de AKR1B1 expresado y silenciado. Los resultados mostraron que en las células que expresan AKR1B1 existe mayor presencia de vimetina, por el contrario, las células que no expresaban esta enzima presentaban mayor expresión de E-cadherina. Asimismo, los ensayos de migración celular indicaron que el silenciamiento de AKR1B1 disminuye la capacidad migratoria, mientras que en los ensayos de tasa de crecimiento y formación de colonias se mostró una reducción significativa en ambas mediciones. Por lo tanto, estos

resultados sugieren que el silenciamiento de la expresión de AKR1B1 tiene un impacto directo sobre el fenotipo de EMT.

Por otra parte, en el estudio de Jia-Xing Zhao y colaboradores<sup>8</sup> se determinó que la sobre-expresión de AKR1B1 en hepatocarcinoma desencadena la vía de señalización de AKT/mTOR a través de la interacción con el dominio cinasa de AKT1 y como consecuencia, se observó un aumento en la formación de lactato y la inflamación hepática para mejorar la hepatocarcinogénesis. Asimismo, mostraron que la sobre-expresión y eliminación de AKR1B1, estabilizaba y desestabilizaba la expresión de AKT1 respectivamente.

También se determinó que la sobre-expresión de AKR1B1 mejoraba de manera significativa la proliferación, la formación de colonias, la migración y la invasión de células de cáncer de hígado SMMC-7721 y HepG2. Por el contrario, al silenciar la expresión del gen de AKR1B1, se presentó una disminución significativa de las mismas características.

Finalmente, Ming Bing Xiao y colaboradores<sup>9</sup> reportaron que la interacción de AKR1B1 con el receptor  $\beta$ 2-adrenérgico promueve la proliferación y dificulta la apoptosis a través de la vía ERK1/2 en la línea celular de cáncer de páncreas BXPC-3. También se ha informado que la expresión de AKR1B1 podría regular la expresión del receptor  $\beta$ 2-adrenérgico en un mecanismo de retroalimentación negativa. Por último, la sobre-expresión de AKR1B1 se asoció con una menor supervivencia en pacientes con cáncer de páncreas.

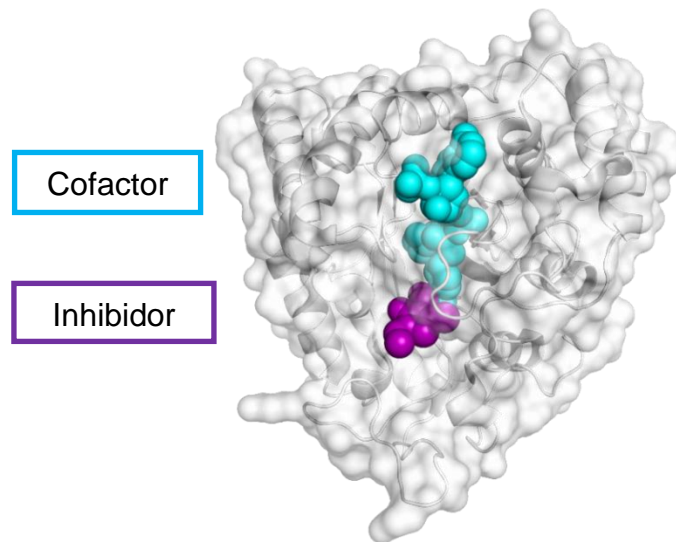
#### 2.5.4 Inhibición farmacológica de AKR1B1

Por su importancia clínica, se han realizado numerosos esfuerzos para desarrollar inhibidores farmacológicos de la actividad de la enzima AKR1B1, tales como el tolrestat, fidarestat, sorbinil, zopolrestat, epalrestat, ponalrestat, entre otros. Los valores de las concentraciones inhibitorias medias (IC50) de estos inhibidores se muestran en la siguiente tabla.

Fármaco	Blanco terapéutico	IC50	Referencia
epalrestat	AKR1B1	2.00 nM	130
tolrestat	AKR1B1	0.31 nM	131
fidarestat	AKR1B1	18.00 nM	132
zopolrestat	AKR1B1	0.31 nM	133
sorbinil	AKR1B1	3.14 $\mu$ M	134

**Tabla 6.** Valores de IC50 de inhibidores farmacológicos de AKR1B1.

El mecanismo de inhibición principal descrito para este grupo de fármacos es que su unión a AKR1B1 involucra dos zonas de contacto en el sitio activo: la primera es una región hidrofílica donde se establecen los puentes de hidrógeno aceptores cerca de la coenzima y la segunda es una zona de contacto hidrofóbica en la hendidura del sitio activo. Tanto tolrestat como zopolrestat se unen a un compartimiento específico en el sitio activo después de que la enzima sufre un pequeño cambio conformacional de un asa (residuos 121-135) y un segmento corto (residuos 298-303) cerca del extremo C-terminal creando un movimiento tipo bisagra para acomodar y secuestrar el inhibidor<sup>135</sup>.



**Figura 9.** Estructura tridimensional de AKR1B1 co-cristalizada con NADPH (color ciano) y el inhibidor epalrestat (color morado). Figura generada con PyMol (Nota: código PDB del cristal: 4JIR).

La inhibición farmacológica ha tenido éxito en el tratamiento de complicaciones diabéticas y en la supresión de trastornos inflamatorios inducidos por estrés oxidativo. Sin embargo, actualmente ningún inhibidor de AKR1B1 se encuentra en evaluación clínica en humanos para el tratamiento del cáncer. Además, ningún tratamiento muestra resultados similares a los que se observan cuando la expresión de AKR1B1 está suprimida. Por lo tanto, es necesario comenzar a centrarse en el desarrollo de inhibidores enfocados en inhibir las IPP de AKR1B1 debido a que también desempeñan un papel importante en el cáncer y otros procesos biológicos<sup>136,137</sup>.

#### 2.5.5 IPP de AKR1B1

Recientemente, Thiagarajan y colaboradores<sup>10</sup>. Reportaron que AKR1B1 compite con la desacetilasa de histona 3 (HDAC3) para unirse al dominio DAD (*Deacetylase Activation Domain*) del correpresor de receptores nucleares (SMRT/NCOR1). Las principales consecuencias de esta interacción incluyen la disociación del complejo HDAC3-DAD, conduciendo a la degradación de HDAC3. La reducción de los niveles de HDAC3 se ve acompañada de un aumento en la actividad de PPAR $\gamma$ , lo que provoca a una hipertrofia cardíaca grave debido a la acumulación de cantidades patológicas de lípidos en el corazón de ratones.

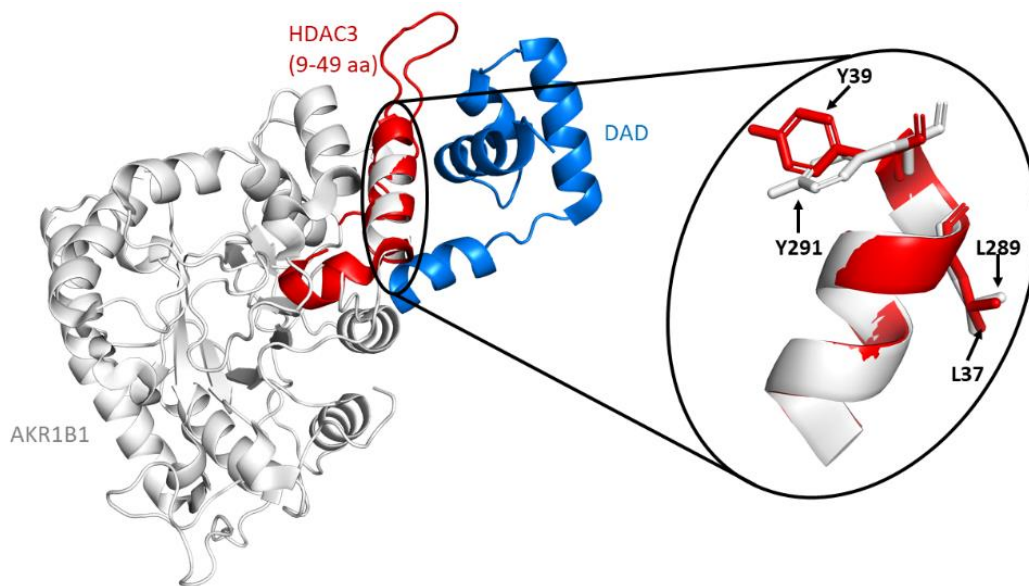
El complejo HDAC3-corepressor reprime varios receptores nucleares, incluidos el receptor de tiroideas, el receptor de ácido retinoico, el receptor de andrógenos, el receptor de glucocorticoides, entre otros. Por lo consiguiente, la degradación de HDAC3, como consecuencia de la interacción AKR1B1-DAD, es esencial para la activación transcripcional de estos receptores nucleares, así como factores de transcripción como el NF $\kappa$ B.

La interacción entre AKR1B1 y DAD, se demostró de manera experimental mediante ensayos de co-IP. En el mismo trabajo se investigó si la interacción AKR1B1-DAD anulaba la actividad enzimática de AKR1B1, midiendo la actividad enzimática en presencia de concentraciones crecientes de una proteína recombinante que corresponde al dominio DAD. Los resultados obtenidos mostraron una disminución moderada (20-30%) en la actividad enzimática, por lo



que estos resultados confirman que la interacción AKR1B1-DAD no se realiza en el sitio catalítico de la enzima.

Posteriormente, mediante métodos computacionales y partiendo de las estructuras cristalizadas de AKR1B1 (PDB: 2DUX) y HDAC3-DAD (PDB: 4A69) los autores identificaron una secuencia en HDAC3 (aminoácidos 28-41) que se asemeja a los residuos 280-293 (H2) de AKR1B1 (ver figura 10). Además, se confirmó de manera experimental la disminución de la interacción de AKR1B1 y DAD cuando se hizo una mutación en la leucina 289 por alanina (L289A). Estos resultados sugieren que la L289 es esencial para la interacción con DAD y confirman que la región de los residuos 280-293 (H2) es importante para mediar la IPP AKR1B1-DAD<sup>10</sup>. Sin embargo, aún no se ha caracterizado de manera más detallada la interacción AKR1B1-DAD.



**Figura 10.** Alineamiento estructural AKR1B1 y HDAC3-DAD. En color blanco se representa AKR1B1, en donde la H2 es estructuralmente semejante a una región de HDAC3 (residuos 30-39) que está coloreada en rojo. El dominio DAD corresponde a la estructura en color azul. La imagen encerrada en el óvalo grande muestra aquellos residuos que son semejantes entre ambas estructuras (L289 y Y291 en AKR1B1 y L37 y Y39 en HDAC3).

### **CAPÍTULO III. PLANTEAMIENTO DEL PROBLEMA**

El interés en la aldo-ceto reductasa humana AKR1B1 se debe principalmente a su relación con diversos tipos de cáncer y otras patologías. En la literatura, se ha destacado la importancia de la interacción de AKR1B1 con otras proteínas, ya que podrían jugar un papel en la EMT de células cancerosas, por lo que la modulación de las IPP de esta enzima se ha convertido en un objetivo terapéutico. No obstante, esas interacciones no se han caracterizado con detalle y en la actualidad no hay reportes de inhibidores de la IPP de AKR1B1. Esto puede deberse a la falta de información estructural sobre los complejos proteína-proteína que forma esta enzima.

Se ha reportado que AKR1B1 interacciona con el dominio DAD del correpresor de receptores nucleares SMRT. Esta IPP es de importancia fisiopatológica, debido a que promueve el desarrollo de hipertrofia cardíaca grave. Aunque se ha determinado experimentalmente que el residuo L289 de AKR1B1 es esencial para mantener esta interacción, se desconoce el modo de asociación entre estas proteínas, así como la identidad de otros residuos que participan en la interacción. Por esta razón, en este trabajo se estudió la interacción AKR1B1-DAD, utilizando herramientas computacionales, con la finalidad de predecir y caracterizar la formación del complejo estructural AKR1B1-DAD. La caracterización de esta interacción y la definición de los residuos claves de AKR1B1 para mediar la IPP permitirá generar hipótesis sobre cómo desarrollar nuevos agentes terapéuticos que interfieran con la formación del complejo.

### **CAPÍTULO IV. OBJETIVOS**

Con base en lo anterior, el objetivo general del presente proyecto consiste en realizar el estudio bioinformático de la IPP entre AKR1B1-DAD, con la finalidad de dilucidar los dominios estructurales en AKR1B1 que participen en la IPP. Por lo tanto, se espera que la información generada en este proyecto contribuya a un futuro desarrollo de inhibidores de la IPP de AKR1B1.

Para lograr este objetivo principal, se plantearon los siguientes objetivos particulares:

- Predecir la región de IPP del complejo AKR1B1-DAD, mediante ensayos de acoplamiento molecular *in silico*.
- Estudiar la región de IPP AKR1B1-DAD, por medio de simulaciones de dinámica molecular y determinar de manera teórica la importancia energética de los residuos para la IPP mediante cálculos MM-PBSA.
- Analizar la estabilidad del complejo estructural AKR1B1-DAD, haciendo mutagénesis de alanina *in silico* de los residuos propuestos como importantes.
- Proponer aquellos residuos potencialmente relevantes para el futuro desarrollo de inhibidores de la IPP en AKR1B1.

## **CAPÍTULO V. METODOLOGÍA**

### **5.1 Características de hardware y software utilizados.**

Se trabajo mediante conexión remota en el equipo de supercomputo “*Miztlil*” de la Universidad Nacional Autónoma de México. Este equipo cuenta con las siguientes características<sup>138</sup>:

#### Procesamiento X86:

- ✓ 332 nodos de cómputo HP Proliant SL230 y SL250.
- ✓ Dos procesadores Intel Xeon por nodo.
- ✓ 5312 núcleos totales de procesamiento.
- ✓ Capacidad de 118 Teraflop/s.

#### Procesamiento GPU

- ✓ 16 GPUs NVIDIA M2090.

#### Memoria RAM

- ✓ 23 terabytes.

#### Almacenamiento:

- ✓ 750 terabytes en disco duro.
- ✓ Diversos sistemas de archivos globales.
- ✓ Almacenamiento principal distribuido basado en SFA12K de Data Direct Networks. y LUSTRE del consorcio OpenSFS.

#### Red:

- ✓ Mellanox Infiniband QDR a 40 Gbps.
- ✓ Red de administración Ethernet a 1 Gbps.
- ✓ Sistema de consolas.

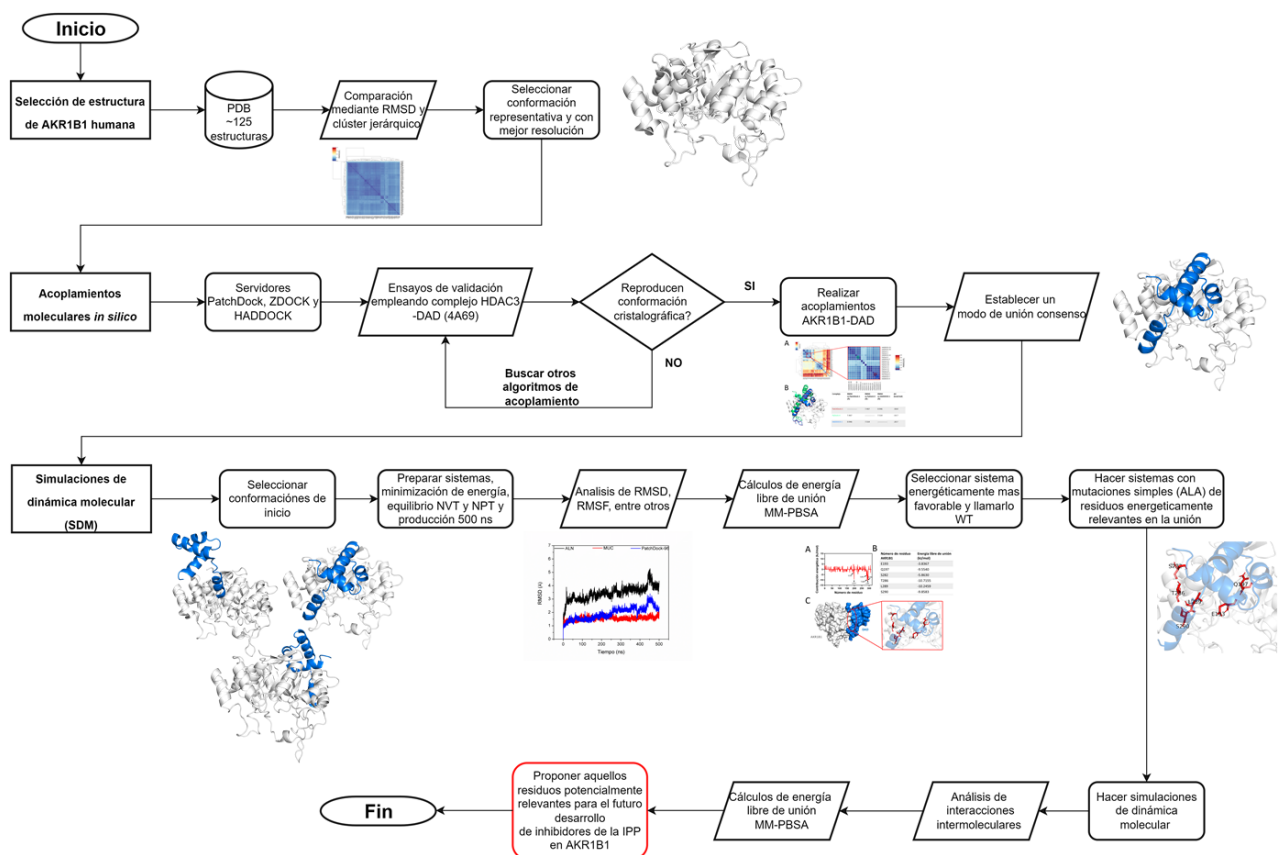
Sistema Operativo:

- ✓ RedHat Enterprise Linux.
- ✓ Scientific Linux RedHat Enterprise.

Los softwares (°) y servidores disponibles en línea utilizados (\*) en este proyecto se listan a continuación.

Base de datos *Protein Data Bank* \*<sup>16</sup>, CHARMM-GUI\*<sup>139</sup>, DockQ°<sup>140</sup>, GROMACS 5.0.4°<sup>141</sup>, g\_mmpbsa°<sup>113</sup>, HADDOCK\*<sup>73,74</sup>, MDanalysis°<sup>111,112</sup>, mTM-align°<sup>142</sup>, Origin(pro)v 8.0°<sup>143</sup>, PATCHDOCK\*<sup>70</sup>, PRODYGY\*<sup>144</sup>, PyMol 2.5°<sup>145</sup>, RStudio°<sup>110</sup>, UCSF Chimera 1.15°<sup>146</sup>, VMD 1.9.3°<sup>109</sup> y ZDOCK\*<sup>71,147</sup>.

En la siguiente figura se representa de manera general el flujo de trabajo seguido en este estudio y a continuación se describirá con más detalle cada proceso



**Figura 11.** Flujo de trabajo que representa de manera general la metodología seguida en el presente proyecto.

## **5.2 Caracterización del dominio de IPP de AKR1B1.**

### 5.2.1 Selección de la estructura cristalográfica de AKR1B1 para ensayos *in silico*

En el *Protein Data Bank* hay depositadas aproximadamente 125 estructuras tridimensionales de AKR1B1 **humanas** co-cristalizadas con su cofactor NADPH y otras moléculas como citrato, fármacos, etcétera, y una de ellas corresponde a la apoenzima (1XGD). Con la finalidad de escoger una única estructura que se pueda emplear para los posteriores análisis, se realizó una comparación de RMSD con el programa mTM-align y posteriormente se utilizó un algoritmo de agrupamiento para determinar si existen diferencias estructurales importantes entre ellas. Dado que el RMSD entre las estructuras fue muy semejante, se decidió elegir aquella estructura (6F7R) con mejores indicadores de calidad del cristal (los cuales se describen en el capítulo de antecedentes; sección: *Protein Data Bank*), y que perteneciera al grupo más representativo. Asimismo, se eligió la estructura de la apoenzima con la finalidad de analizar el comportamiento estructural de la proteína en presencia y ausencia de su cofactor.

Posteriormente, se efectuaron simulaciones de dinámica molecular (la metodología empleada para las SDM se describe en el capítulo Metodología; sección simulaciones de dinámica molecular) de las dos estructuras elegidas para analizar la dinámica estructural de la proteína en sus formas holo y apo en un tiempo determinado y bajo condiciones fisiológicas simuladas.

Al no observar diferencias estructurales considerables, se decidió trabajar con la estructura de AKR1B1 co-cristalizada con su cofactor (PDB:6F7R).

### 5.2.2 Acoplamiento moleculares *in silico*

Para predecir el modo de unión entre AKR1B1 y DAD, se realizaron acoplamiento moleculares *in silico* utilizando tres servidores en línea adecuados para el acoplamiento proteína-proteína de acuerdo con CAPRI: PATCHDOCK, ZDOCK y HADDOCK (las características de estos servidores se describen en el capítulo de antecedentes; sección: acoplamiento moleculares *in silico*). Antes de realizar los acoplamiento de AKR1B1 y DAD, se llevó a cabo una validación de ensayos de acoplamiento, la cual consiste en determinar si un servidor o programa de

acoplamiento es capaz de reproducir la conformación cristalográfica de un ligando con el que la proteína fue cristalizada. Con referencia a lo mencionado, se utilizó la estructura cristalizada del complejo HDAC3-DAD (4A69). A dicha estructura se le eliminaron los ligandos y las moléculas de agua co-cristalizadas y se mantuvo una estructura de las cadenas que eran idénticas (cadena A: HDAC3 y cadena C: DAD; B y D se eliminaron), después cada estructura se guardó en un archivo único, posteriormente se hicieron los acoplamientos usando los tres servidores antes mencionados, siendo HDAC3 la proteína receptora y DAD la proteína ligando.

Finalmente, para evaluar la calidad estructural de los modelos de acoplamiento predichos, se utilizaron los criterios que establece CAPRI (los cuales se describen en el capítulo de antecedentes; sección: 2.4.1.3). La determinación de estos criterios se realizó con el programa DockQ. Este programa ha mostrado reproducir casi por completo la clasificación CAPRI original (90-94%)<sup>140</sup>. Asimismo, implementa una puntuación única en el intervalo de 0-1 que proviene de los criterios Fnat, LRMS e IRMS. La estimación de estos criterios se hace comparando la conformación cristalográfica y las conformaciones generadas por los tres servidores de acoplamiento, con el objetivo de verificar si los tres servidores son capaces de generar soluciones que sean muy semejantes al modo de unión cristalográfico, además de evaluar la calidad de la predicción.

La clasificación de los modelos utilizando la puntuación DockQ se muestran en la siguiente tabla:

<b>Calidad de la predicción</b>	<b>DockQ</b>
Alta	$\geq 0.80$
Media	$0.49 \leq \text{DockQ} < 0.80$
Aceptable	$0.23 \leq \text{DockQ} < 0.49$
Incorrecto	$0.00 \leq \text{DockQ} < 0.23$

**Tabla 7.** Clasificación de la calidad de los modelos de acoplamiento con la puntuación DockQ<sup>140</sup>.

Una vez que se comprobó que los tres servidores reproducen de manera exitosa la conformación del cristal 4A69, se realizó el acoplamiento de AKR1B1 y DAD. La

proteína receptora fue AKR1B1 utilizando la estructura del cristal 6F7R, a dicha estructura se le eliminaron las moléculas de agua co-cristalizadas y el cofactor NADPH se mantuvo. Por otro lado, la proteína que actuó como ligando fue DAD. En el caso de los servidores ZDOCK y HADDOCK se puede especificar si algún residuo es importante para establecer interacciones, y dado que se tiene evidencia experimental de que la L289 en AKR1B1 es importante para establecer IPP, entonces se proporcionó esta información a los servidores

Las treinta mejores soluciones en los tres servidores se compararon mediante una matriz de RMSD y se hizo un análisis de agrupamiento jerárquico, con el objetivo de encontrar un modo de unión consenso. Las soluciones mejor calificadas por los servidores presentaban un modo de unión muy similar, por lo tanto, se calculó la energía de unión con PRODYGY de cada una y la solución con menor energía de unión fue considerada como el modo de unión consenso.

Finalmente, se analizaron las interacciones de manera más detallada empleando PyMol y PRODYGY, a partir de las soluciones generadas por acoplamiento molecular, de tal manera que se logró identificar una región de residuos en AKR1B1 donde se presentan las interacciones con mayor frecuencia.

### 5.2.3 Simulaciones de dinámica molecular

#### 5.2.3.1 SDM de AKR1B1 holoenzima y apoenzima

Antes de comenzar, se removieron los hidrógenos de las estructuras proteicas. La SDM se realizó empleando el software GROMACS y el campo de fuerzas CHARMM36m. El sistema se encerró en una caja cúbica y se solvató con moléculas de agua siguiendo el modelo TIP3P y se adicionaron iones  $K^+$  y  $Cl^-$  a una concentración iónica de 0.15 M para simular condiciones fisiológicas y se utilizaron condiciones periódicas de la frontera. Posteriormente, se hizo un proceso de minimización de la energía en 100000 pasos con el algoritmo del descenso más pronunciado (*steepest descent*). Luego el sistema se equilibró bajo los ensambles NVT y NPT a condiciones de temperatura de 310.15 K y presión de 1 bar en, 250000 pasos (500 ps) usando termostato *v-rescale* y el baróstato de Berendsen, respectivamente. Finalmente, se ejecutó la dinámica durante 500 ns a

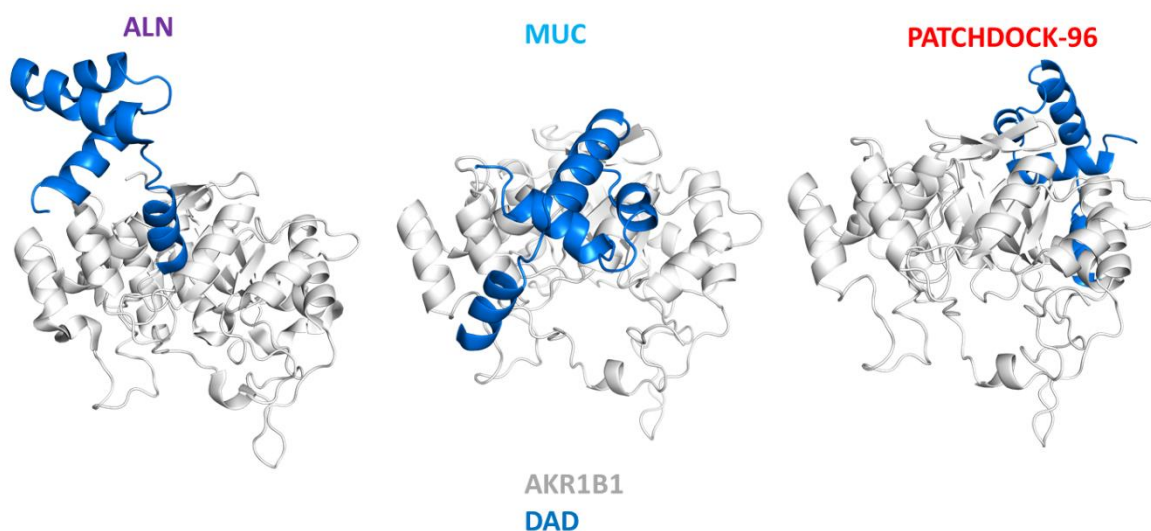


una temperatura de 310.15 K bajo el ensamble NPT. Las ecuaciones de movimiento de Newton se integraron con el algoritmo de salto de rana (*leap-frog*) cada 2 fs.

Las trayectorias se analizaron usando el software MDanalysis para calcular RMSD y RMSF.

### 5.2.3.2 SDM de AKR1B1-DAD

Con el objetivo de caracterizar de manera detallada el modo de unión entre AKR1B1 y DAD, así como determinar aquellos residuos que son más relevantes para la IPP en AKR1B1, se llevaron a cabo SDM del complejo estructural AKR1B1-DAD. Para esto se utilizaron diferentes conformaciones de inicio de AKR1B1-DAD (ver figura 12) provenientes de un alineamiento estructural de 6F7R con 4A69 (ALN), el modo de unión consenso (MUC) predicho por los servidores de acoplamiento molecular y un modo de unión que fue evaluado con baja puntuación por los servidores de acoplamiento (PATCHDOCK-96), en donde DAD se une a una región diferente de la  $\alpha$ -hélice H2 de AKR1B1, con el objetivo de comparar y determinar si el modo de unión consenso predicho por los servidores es el más estable. Además de corroborar que DAD pueda o no unirse de manera estable a otra región de AKR1B1.



**Figura 12.** Conformaciones de inicio de AKR1B1 y DAD para SDM. En color blanco se muestra la estructura de AKR1B1, mientras que en color azul se representa a DAD. ALN corresponde al modo

de unión proveniente de un alineamiento estructural de AKR1B1 y HDAC3-DAD, conservando solamente AKR1B1 y DAD. MUC es el modo de unión consenso predicho por los servidores de acoplamiento. Finalmente, PATCHDOCK-96 fue un modo de unión predicho por el servidor PATCHDOCK y que obtuvo una mala calificación, cabe denotar que DAD se une a una región diferente de la  $\alpha$ -hélice H2 de AKR1B1.

El protocolo de simulación fue el siguiente: en primer lugar, se eliminaron los hidrógenos de las estructuras proteicas, posteriormente se construyó el sistema con la plataforma en línea CHARMM-GUI. Para esto se requirieron las estructuras de AKR1B1 y DAD en formato PDB y con la herramienta *PDB reader* se generaron los archivos de parámetros que son reconocidos por los campos de fuerzas de CHARMM, mientras que para el cofactor se utilizó la estructura en formato sdf y se empleó la herramienta *ligand reader & modeler* para generar los parámetros con el campo de fuerza CGenFF, el cual es compatible con los campos de fuerzas de CHARMM para proteínas. Finalmente, el sistema se generó con la herramienta *Multicomponent Assembler*, en el cual se combinaron las estructuras y parámetros de AKR1B1, DAD y el cofactor.

Para preparar el sistema se generó una caja cúbica, se adicionaron moléculas de solvente TIP3P, se usó una concentración iónica a 0.15 M y se generaron condiciones periódicas de la frontera. La simulación de dinámica molecular se realizó utilizando el software GROMACS y el campo de fuerzas CHARMM36m. La minimización de la energía se llevó a cabo en 100000 pasos con el algoritmo del descenso más pronunciado. Luego el sistema se equilibró bajo los ensambles NVT y NPT a condiciones de temperatura de 310.15 K y presión de 1 bar en, 250000 pasos (500 ps) usando termostato *v-rescale* y el baróstato de Berendsen, respectivamente. Finalmente, se ejecutó la dinámica durante 500 ns a una temperatura de 310.15 K bajo el ensamble NPT. Las ecuaciones de movimiento de Newton se integraron con el algoritmo de salto de rana (*leap-frog*) cada 2 fs.

Las trayectorias se analizaron empleando el software MDAnalysis para calcular RMSD, RMSF y radio de giro. Por otro lado, se calcularon otras métricas como la distancia entre el centro de masas, helicidad y parámetro de orden con

GROMACS. La energía libre de unión se determinó mediante cálculos de MM/PBSA usando g\_mmpbsa. El cálculo de energía de solvatación polar, no polar y de mecánica molecular en el vacío se hizo cada 0.1 nanosegundos. Asimismo, se realizó un análisis de la descomposición energética por residuo para determinar la contribución de energía de cada residuo para establecer la unión.

#### 5.2.3.3 Mutagénesis de alanina *in silico*

Se hicieron seis mutantes simples de aquellos residuos de AKR1B1 que son energéticamente relevantes para la unión con DAD (E193A, Q197A, S282A, T286A, L289A y S290) y una mutante simple T287A, que ya se ha determinado de manera experimental que no altera la unión de AKR1B1-DAD de manera importante. Para llevar esto a cabo, se seleccionó como complejo WT a la conformación que fue más estable en la SDM de los complejos AKR1B1-DAD (modo de unión consenso generado por acoplamiento molecular).

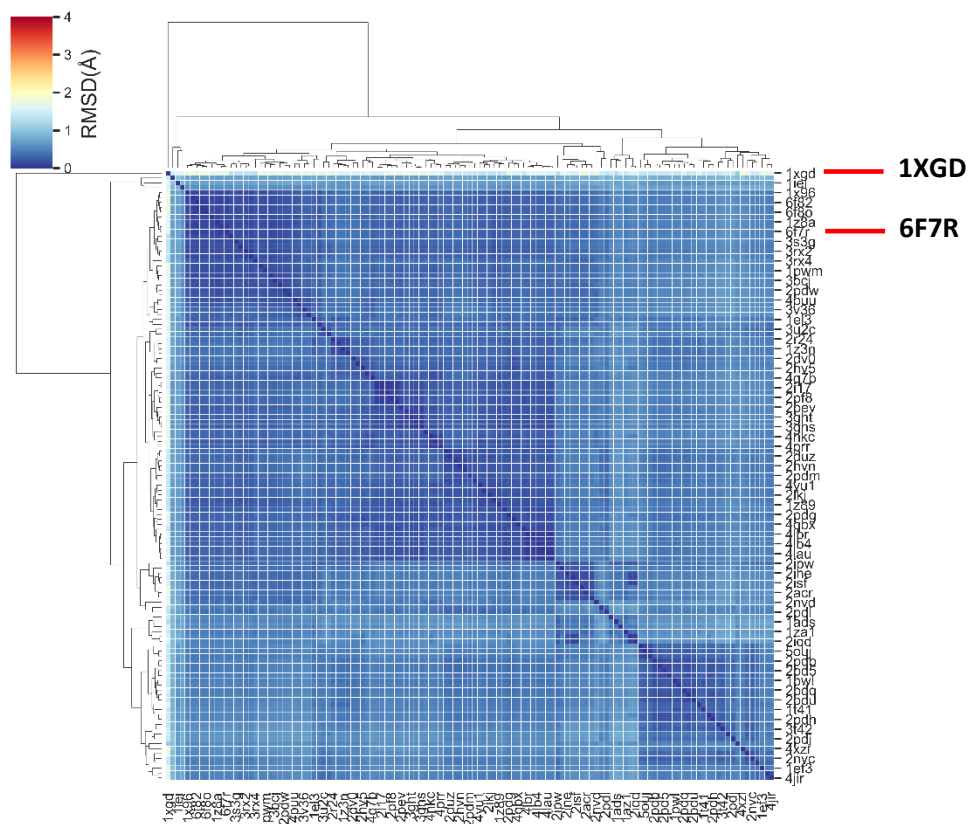
Posteriormente, se produjeron mutaciones simples con ALA de los residuos mencionados anteriormente con PyMol, generando siete nuevos sistemas que fueron sometidos a SDM, después se hicieron cálculos de energía libre de unión MM-PBSA de cada sistema y se realizó un análisis de las interacciones. Esto se hizo con la finalidad de observar cómo se ve afectada la unión con DAD al mutar cada uno de estos residuos y así determinar el papel que pueden desempeñar estos residuos en la interacción proteína-proteína.

#### 5.2.4 Alineamiento múltiple de secuencias

Adicionalmente, se efectuó un alineamiento múltiple de secuencias con el objetivo de observar si los residuos energéticamente relevantes para la unión entre AKR1B1 y DAD se conservan en diferentes especies. Para la comparación se utilizó el programa ClustalΩ y las secuencias de AKR1B1 empleadas son de humano (*Homo sapiens*), ratón (*Mus musculus*), rata (*Rattus norvegicus*), cerdo (*Sus scrofa*), chimpancé (*Pan troglodytes*), murciélago (*Rhinolophus ferrumequinum*), pez cebra (*Danio rerio*) y ave (*Dasyornis broadbenti*). Estas secuencias fueron obtenidas de la base de datos UniProt.

CAPÍTULO VI. RESULTADOS Y DISCUSIÓN

6.1 Comparación de estructuras de AKR1B1 humanas disponibles en el *Protein Data Bank*



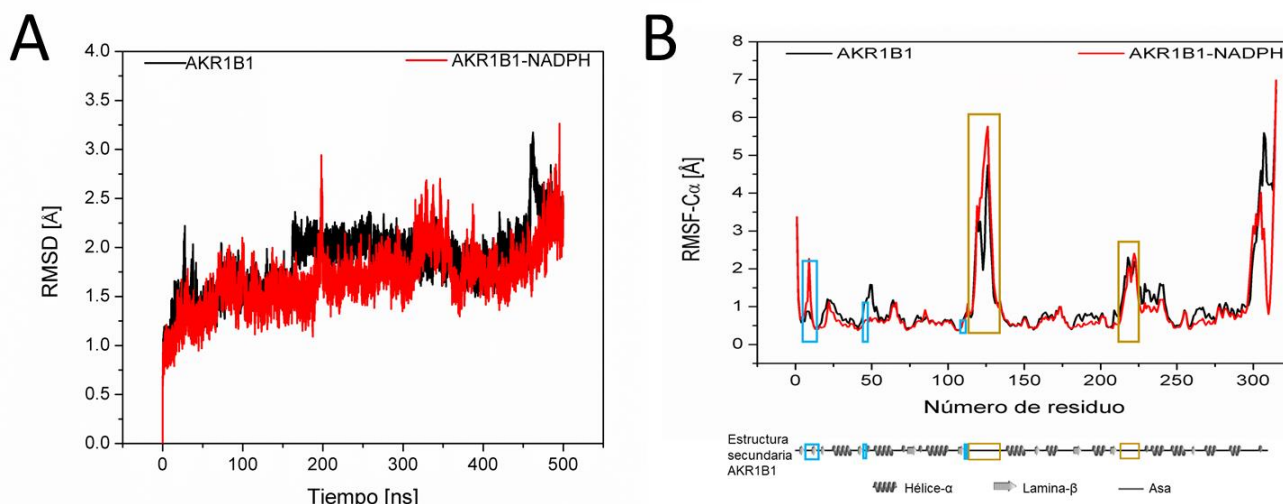
**Figura 13.** Comparación de RMSD-Cα y agrupamiento jerárquico entre todas las estructuras de AKR1B1 humanas disponibles en el *Protein Data Bank*.

Las 125 estructuras cristalográficas de AKR1B1 humanas depositadas en el *Protein Data Bank* se compararon mediante un análisis de RMSD de Cα y un agrupamiento jerárquico con el algoritmo *complete linkage*, con la finalidad de encontrar similitudes o diferencias estructurales que permitieran realizar una selección adecuada de una estructura única que sirva para los posteriores análisis de acoplamiento molecular.

En la figura 13 se observa un mapa de calor que representa un agrupamiento jerárquico con base en los valores de RMSD entre todas las estructuras analizadas. Se puede apreciar que la variación en el RMSD entre todas las

estructuras no es mayor a 2.0 Å, lo cual indica que no existen diferencias estructurales importantes en las estructuras como algún cambio conformacional debido a la presencia de un ligando. Aunque no existan estas diferencias relevantes, el método de agrupamiento generó dos grupos principales. Uno de ellos contiene más del 70% de las estructuras, por lo tanto, una estructura perteneciente a este grupo sería representativa. En consecuencia, se eligió la estructura 6F7R, debido a que era una de las estructuras que presentaba mejores indicadores de calidad de cristal y resolución (0.92 Å), así mismo pertenece al grupo mayoritario.

Adicionalmente, se puede observar en la figura 13 que hay una estructura que es la que presenta mayores diferencias con respecto a las demás, se trata de la proteína en su forma de apoenzima (1XGD). Por lo tanto, se decidió analizar la dinámica estructural de la proteína en ausencia y presencia de su cofactor endógeno NADPH bajo condiciones de temperatura y presión fisiológicas, mediante simulaciones de dinámica molecular durante un periodo de tiempo de 500 ns.



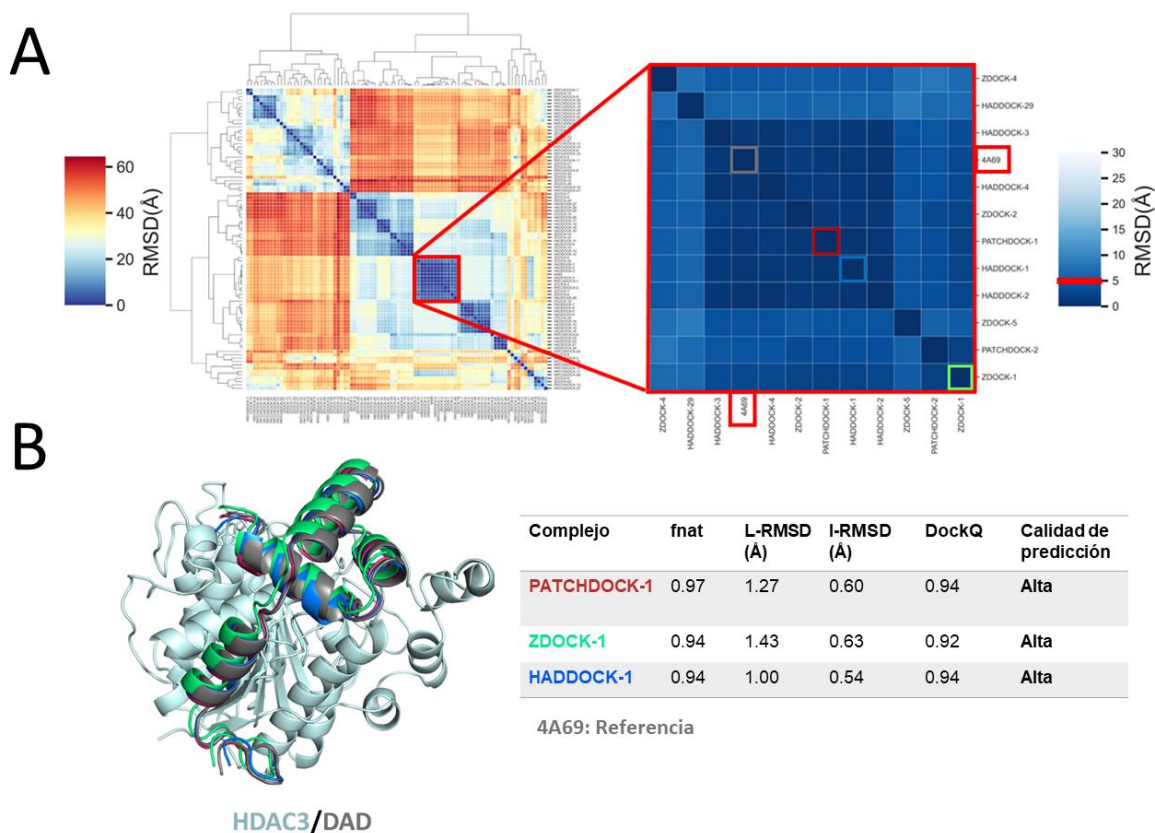
**Figura 14.** En el panel A se presenta un gráfico de la comparación de RMSD del esqueleto proteico entre las trayectorias de dinámica molecular de AKR1B1-NADPH y AKR1B1 apoenzima. Mientras que en panel B, se muestra un gráfico del RMSF de Cα. En los cuadros de color dorado se resaltan las dos regiones de residuos que presentan mayor fluctuación, correspondientes a E120-T140 y L212-L228, mientras que los residuos del sitio de unión a NADPH (10-19 aa), Y48 e

H110 pertenecientes al sitio catalítico se remarcan en cuadros color cian. Finalmente, debajo del número de residuo hay un esquema de la estructura secundaria en la que pertenecen estos residuos.

En la figura 14A se muestra una comparación de RMSD del esqueleto proteico entre ambas trayectorias de dinámica molecular, en donde se observa que no hay diferencias estructurales importantes durante el periodo de simulación, debido a que la diferencia de RMSD entre las conformaciones no supera los 2.0 Å. Igualmente, en la figura 14B se aprecia el análisis de RMSF de C $\alpha$ , donde la fluctuación de los residuos en ambas trayectorias es muy semejante. Las regiones de la proteína que exhibieron mayores fluctuaciones fueron los extremos amino y carboxilo, así como dos asas que abarcan de los residuos E120-T140 y L212-L228 (remarcados en cuadros dorados), estos resultados son coherentes, ya que las asas tienden a ser estructuras muy móviles en las proteínas. De igual manera, se aprecia que los residuos Y48, H110 pertenecientes al sitio catalítico (remarcados en cuadros cian), no presentan diferencias ni fluctuaciones considerables. Sin embargo, aquellos residuos del sitio de unión a NADPH (10-19 aa), presentan mayor fluctuación en AKR1B1-NADPH, lo cual puede deberse a la presencia del NADPH.

Estos resultados sugieren que la presencia del cofactor NADPH no afecta de manera importante la dinámica estructural de la proteína y, por lo tanto, no debería representar inconvenientes trabajar con la proteína en forma de apoenzima u holoenzima. Sin embargo, se decidió trabajar con la estructura 6F7R (holoenzima), debido a que presentó menores diferencias en el RMSD con las otras estructuras de AKR1B1 humanas disponibles en el *Protein Data Bank* (ver figura 13).

6.2 Ensayos de validación de servidores de acoplamiento proteína-proteína.



**Figura 15.** En la figura A se presenta el análisis de agrupamiento de los modos de unión de HDAC3 y DAD predichos por los servidores de acoplamiento molecular PATCHDOCK, ZDOCK y HADDOCK, así como la conformación cristalográfica 4A69. En la parte derecha se hace énfasis en el subgrupo en donde se encuentra 4A69 y las conformaciones semejantes. En la figura B se muestra una comparación estructural entre 4A69 (color gris) y las soluciones mejor calificadas de cada uno de los servidores utilizados. (PATCHDOCK color vino, ZDOCK color verde y HADDOCK color azul). Asimismo, se presenta una tabla con la evaluación de la predicción siguiendo los criterios que se establecen en CAPRI.

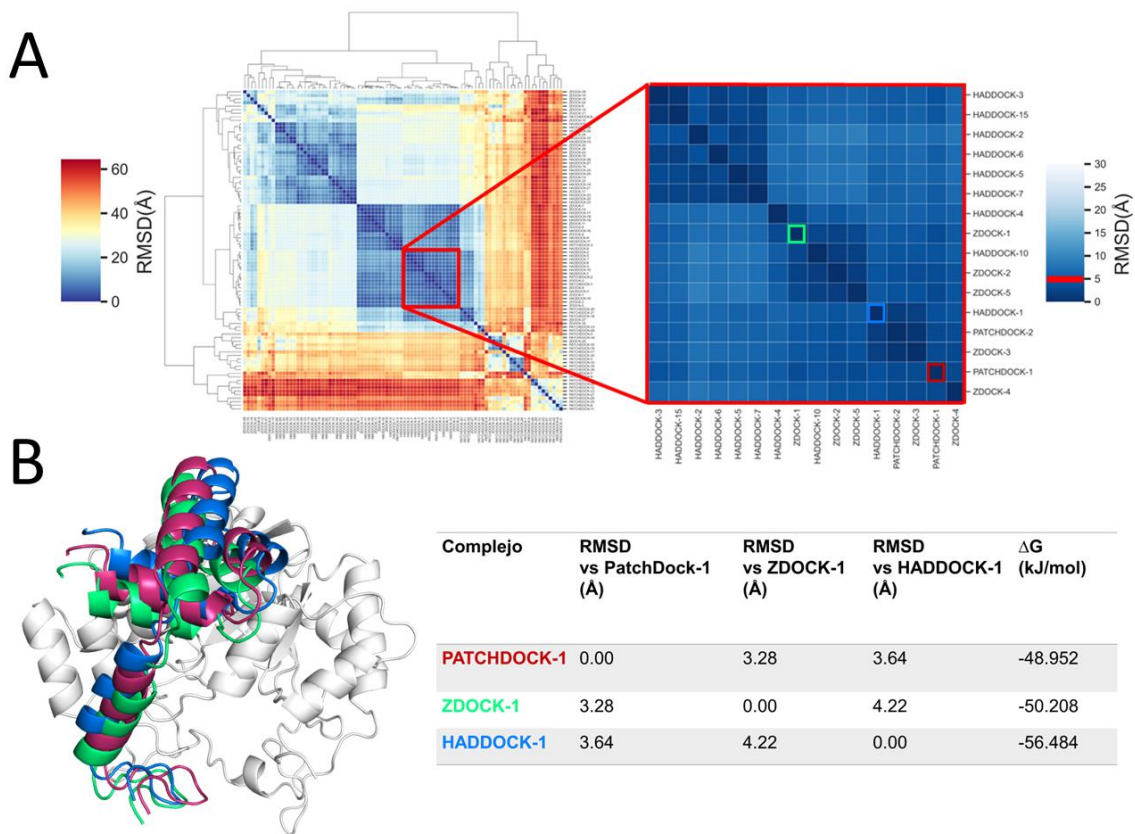
Mediante los ensayos de validación se buscó que los tres servidores de acoplamiento molecular candidatos predijeran el modo de unión del complejo estructural HDAC3-DAD (4A69). Las soluciones mejor calificadas se compararon con la estructura cristalográfica mediante el RMSD del esqueleto proteico y se realizó un agrupamiento jerárquico (ver figura 15A). En el mapa de calor se observan dos grupos principales que presentan diferencias de RMSD muy altas (>40.0 Å), lo cual indica que los modos de unión de DAD entre los grupos se dan

en diferentes regiones de HDAC3. Sin embargo, la finalidad de este ensayo es comparar si hay conformaciones que reproduzcan el modo de unión cristalográfico. El modo de unión cristalográfico se encuentra en el grupo principal inferior derecho. En la figura 15A se hace un énfasis en el subgrupo donde se encuentra este modo de unión y se puede observar que algunas de las soluciones pertenecientes a este subgrupo son PATCHDOCK-1, PATCHDOCK-2, ZDOCK-1, ZDOCK-2, ZDOCK-4, HADDOCK-1, HADDOCK-2, HADDOCK-3 y HADDOCK-4, los cuales son los modos de unión que obtuvieron mejor calificación por los servidores de acoplamiento. Las soluciones mejor calificadas (#1) de cada uno de los servidores se compararon con la referencia cristalográfica siguiendo los criterios que se establecen en CAPRI. Estos criterios son *fnat*, L-RMSD e I-RMSD. En la tabla de la figura 15B, se observa que los tres complejos predichos por los servidores de acoplamiento presentan valores de *fnat*, L-RMSD e I-RMSD que están dentro del intervalo de las predicciones de alta calidad de acuerdo con CAPRI (ver tabla 2). Igualmente, obtuvieron una puntuación de DockQ mayor a 0.8, lo cual indica que la predicción de estos modelos es de alta calidad con base en esta función de puntuación (ver tabla 7).

Finalmente, se decidió comparar de manera estructural estos modos de unión con respecto a la conformación cristalográfica (ver figura 15B), en donde se observa que las diferencias estructurales son mínimas. Por lo tanto, estos resultados en conjunto indican que los tres servidores lograron reproducir de manera muy similar el modo de unión de la estructura 4A69 y, por consiguiente, estos servidores fueron utilizados para predecir el modo de unión entre AKR1B1 y DAD.



6.3 Acoplamiento molecular AKR1B1 y DAD



**Figura 16.** La figura A muestra el análisis de agrupamiento de los modos de unión de AKR1B1 y DAD predichos por los servidores de acoplamiento molecular PATCHDOCK, ZDOCK y HADDOCK. En la parte derecha se hace énfasis en el subgrupo que contiene algunas de las soluciones mejor calificadas. En la figura B se observan los tres modos de unión mejor calificados y que pertenecen al subgrupo marcado en rojo. La solución predicha por PATCHDOCK se encuentra coloreada de vino, en tanto la de ZDOCK se muestra en color verde y la de HADDOCK en azul. Finalmente, en la tabla se muestran los valores de RMSD entre las conformaciones de DAD y el valor de  $\Delta G$  de unión predicho con el servidor PRODYGY.

Se predijeron los posibles modos de unión de AKR1B1 y DAD en los servidores de PATCHDOCK, ZDOCK y HADDOCK. Los complejos estructurales mejor calificados se sometieron a un análisis más detallado para establecer un modo de unión consensado.

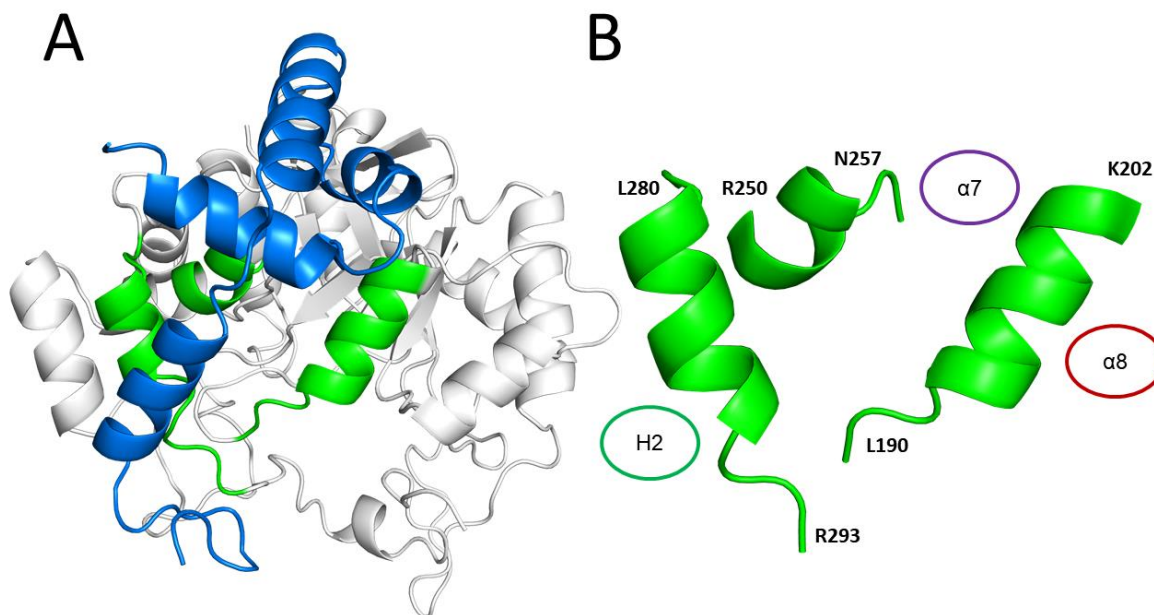
En primer lugar, se realizó una matriz de RMSD para comparar el modo de unión de DAD en todas las soluciones y posteriormente se hizo un agrupamiento jerárquico (ver figura 16A). En donde se observa que existe mucha variación en

los modos de unión generados. En la figura 16A, se observan dos grupos principales bien definidos cuya diferencia de RMSD es mayor a 30.0 Å. En el grupo principal que se encuentra en el centro del mapa de calor se observa que existen dos subgrupos que se superponen. La región de superposición es la que se señala y se hace énfasis en el cuadro rojo. Este grupo contiene 16 de las 90 soluciones y se encuentran algunas de las mejores calificadas por los servidores de acoplamiento: PATCHDOCK-1, PATCHDOCK-2, ZDOCK-1, ZDOCK-2, ZDOCK-3, ZDOCK4, HADDOCK-1, HADDOCK-2, HADDOCK-3 y HADDOCK-4. Estos modos de unión son muy similares debido a la intensidad del color azul en el mapa de calor. Con base al criterio que establece CAPRI para el L-RMSD (RMSD del esqueleto del ligando), el límite de corte para evaluar la similitud entre las predicciones se estableció en 5.0 Å<sup>148</sup>.

Para establecer un modo de unión consenso, se seleccionaron los modos de unión mejor calificados por los tres servidores y con menores diferencias de RMSD (ver figura 16B) y se determinó la energía libre de unión bajo una misma función de energía empleando el servidor PRODYGY. En la figura 16B se representan los tres modos de unión seleccionados y una tabla con los valores de RMSD entre ellos y el  $\Delta G$  de unión calculado con PRODYGY. Se puede apreciar que los tres modos de unión predichos son estructuralmente similares porque sus valores de RMSD entre las estructuras es igual o menor a 5.0 Å. Dado que estos tres algoritmos de búsqueda y funciones de evaluación diferentes arrojan resultados similares, es posible que el modo de unión sea cercano a estas conformaciones. Por lo tanto, con base en la energía de unión calculada, se determinó que la solución HADDOCK-1, sea el modo de unión consenso, debido a que presentó el valor de  $\Delta G$  más favorable (-56.48 kJ/mol).

Finalmente, se analizaron las interacciones de AKR1B1 y DAD de la predicción consenso y se delimitó una región compuesta por tres hélices- $\alpha$  (ver figura 17), que es la responsable de mantener la mayoría de las interacciones con DAD. Esta región está constituida por la hélice  $\alpha 8$  que abarca de los residuos L190-K202, también está presente la hélice  $\alpha 7$  a partir del residuo R250 a N257 y finalmente la

hélice H2 compuesta por los residuos L280-R293, de la cual ya se sabe experimentalmente que uno de sus residuos (L289), es relevante para la interacción con DAD.



**Figura 17.** Sitio de interacción de AKR1B1 y DAD. La imagen A representa el sitio de interacción de AKR1B1 con DAD, el cual se compone de tres hélices- $\alpha$ . En la imagen B se resalta la hélice  $\alpha 8$  que corresponde del residuo 190 al 202 (LTQEKLIQYCQSK); la hélice  $\alpha 7$  que abarca del residuo 250 al 257 (RFPMQRLN). Finalmente, la hélice H2 que está compuesta por los residuos 280 al 293 (LSSQDMTTL~~L~~SYNR).

## 6.4 Simulaciones de dinámica molecular de complejos estructurales

### AKR1B1-DAD

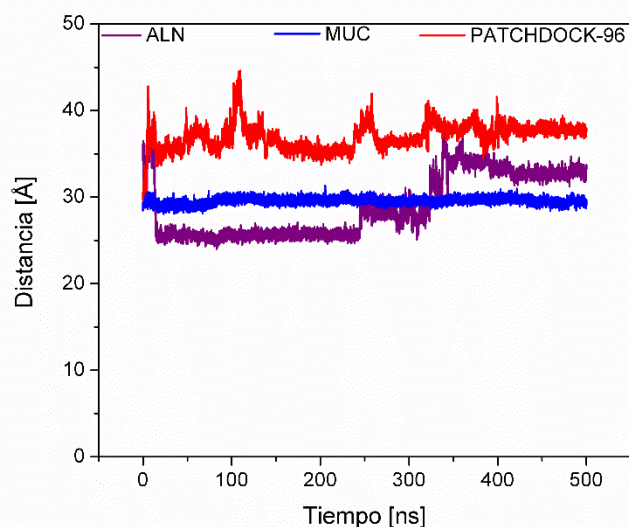
Para caracterizar de manera más detallada el sitio de unión de AKR1B1 con DAD y determinar que el modo de unión seleccionado es el más estable, se realizaron simulaciones de dinámica molecular de diferentes complejos estructurales AKR1B1-DAD.

Dichos complejos estructurales presentan diferencias en el modo de unión de DAD. El primero de ellos llamado ALN, proviene de un alineamiento estructural de la  $\alpha$ -hélice H2 de AKR1B1 con una región de HDAC3 (R28-L41) que interacciona con DAD en el cristal 4A69, en este caso HDAC3 se eliminó del complejo. El

siguiente complejo analizado es el modo de unión consenso (MUC) que fue predicho mediante acoplamientos moleculares, y el último complejo corresponde a una solución con baja calificación por PATCHDOCK (PATCHDOCK-96) en la cual DAD se une a otra región en la proteína. En la siguiente tabla se muestran los valores de RMSD entre estos complejos estructurales y el  $\Delta G$  de unión calculado con PRODYGY.

Complejo	RMSD vs ALN	RMSD vs MUC	RMSD vs PATCHDOCK-96	$\Delta G$ de unión (kJ/mol)
ALN	0.00	18.50	58.67	-46.44
MUC	18.50	0.00	47.35	-56.48
PATCHDOCK-96	58.67	47.35	0.00	-35.14

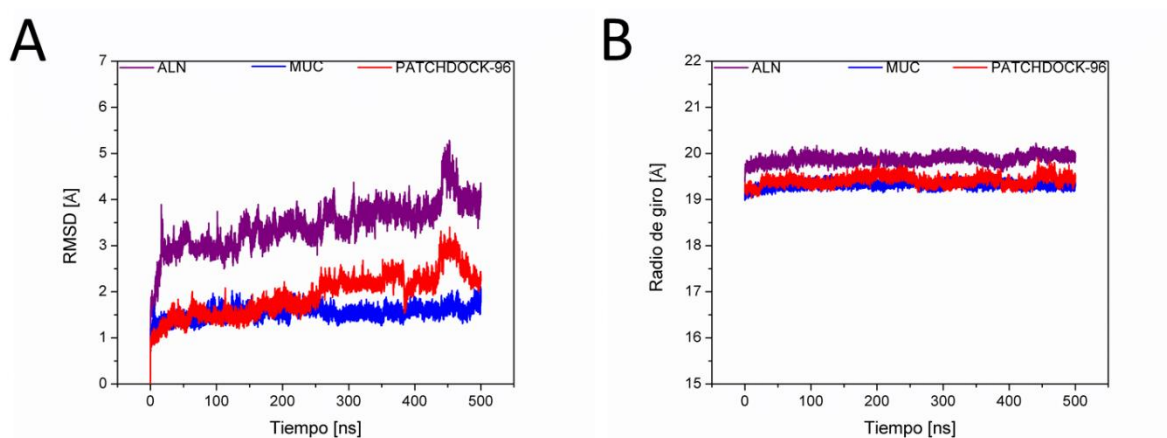
**Tabla 8.** Comparación de RMSD y energía libre de unión entre los complejos estructurales AKR1B1-DAD que se utilizaron en la SDM.



**Figura 18.** Distancia entre los centros de masas de AKR1B1 y DAD durante SDM 500 ns. Las líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

En la figura 18, se presenta un gráfico del análisis de distancia entre centros de masas de AKR1B1 y DAD durante el tiempo de simulación de cada uno de los complejos estructurales. En el complejo estructural ALN la distancia entre los centros de masas se mantiene constante hasta los 250 ns, posteriormente hay un

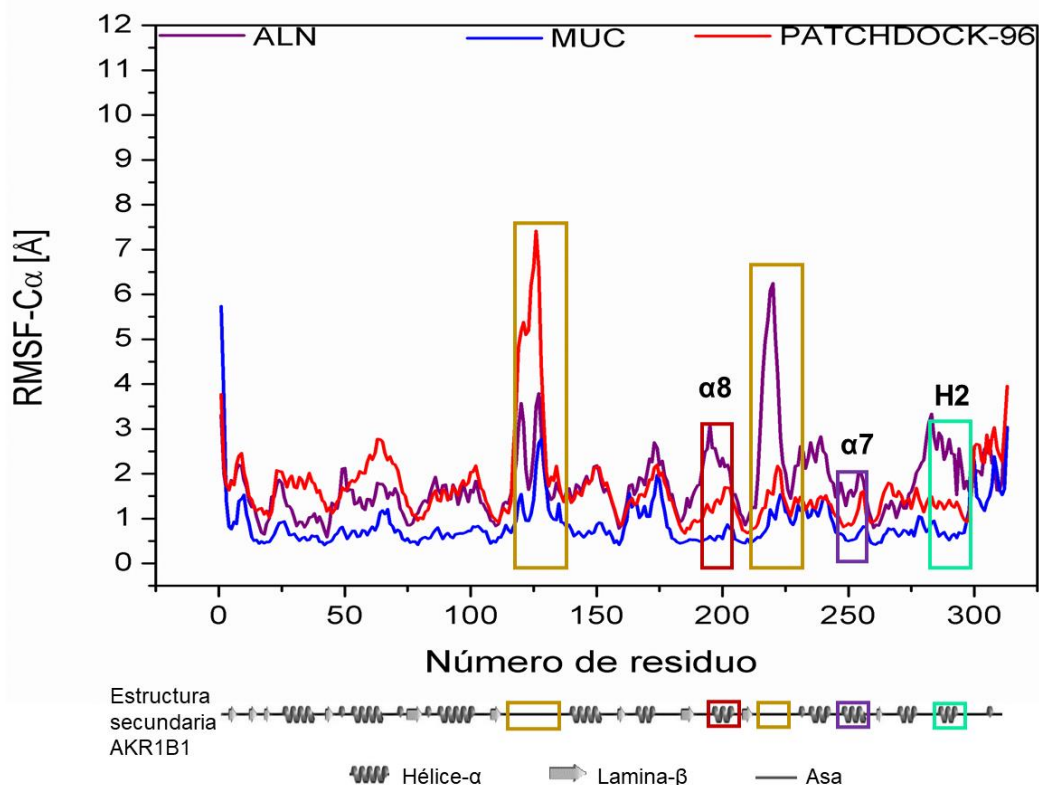
incremento inconstante hasta los 400 ns, esto sugiere que en este periodo de tiempo la unión de AKR1B1 y DAD no es estable. Por otra parte, la distancia de los centros de masas del complejo estructural MUC se mantiene totalmente constante durante los 500 ns, lo cual indica que existe poca variación en la separación entre ligando y receptor, por lo que esta unión podría ser estable. Por el contrario, las distancias en el complejo PATCHDOCK-96 se comportan de manera inconsistente durante todo el tiempo, esto denota que la unión entre las proteínas no es estable.



**Figura 19.** El gráfico A corresponde al RMSD en función del tiempo del esqueleto proteico de AKR1B1 en los diferentes complejos estructurales AKR1B1-DAD. Mientras que el gráfico B indica el radio de giro de AKRB1. La líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

De la misma forma, en la figura 19A se representan los valores de RMSD en función del tiempo para AKR1B1. Se observa que AKR1B1 en el complejo MUC alcanza el equilibrio durante los primeros nanosegundos, y la dinámica estructural de la proteína se mantiene estable durante todo el tiempo de simulación, los valores de RMSD de los microestados no supera los 2.0 Å con respecto a la conformación inicial y el RMSD promedio fue de  $1.55 \pm 0.14$  Å. En contraste, en los dos sistemas restantes se presentan muchos cambios con respecto a la conformación inicial a lo largo de la simulación, para ALN el RMSD promedio fue de  $3.44 \pm 0.50$  Å, mientras que para PATCHDOCK-96 fue de  $1.92 \pm 0.46$  Å. Sin embargo, no se observa alguna transición importante que indique algún cambio

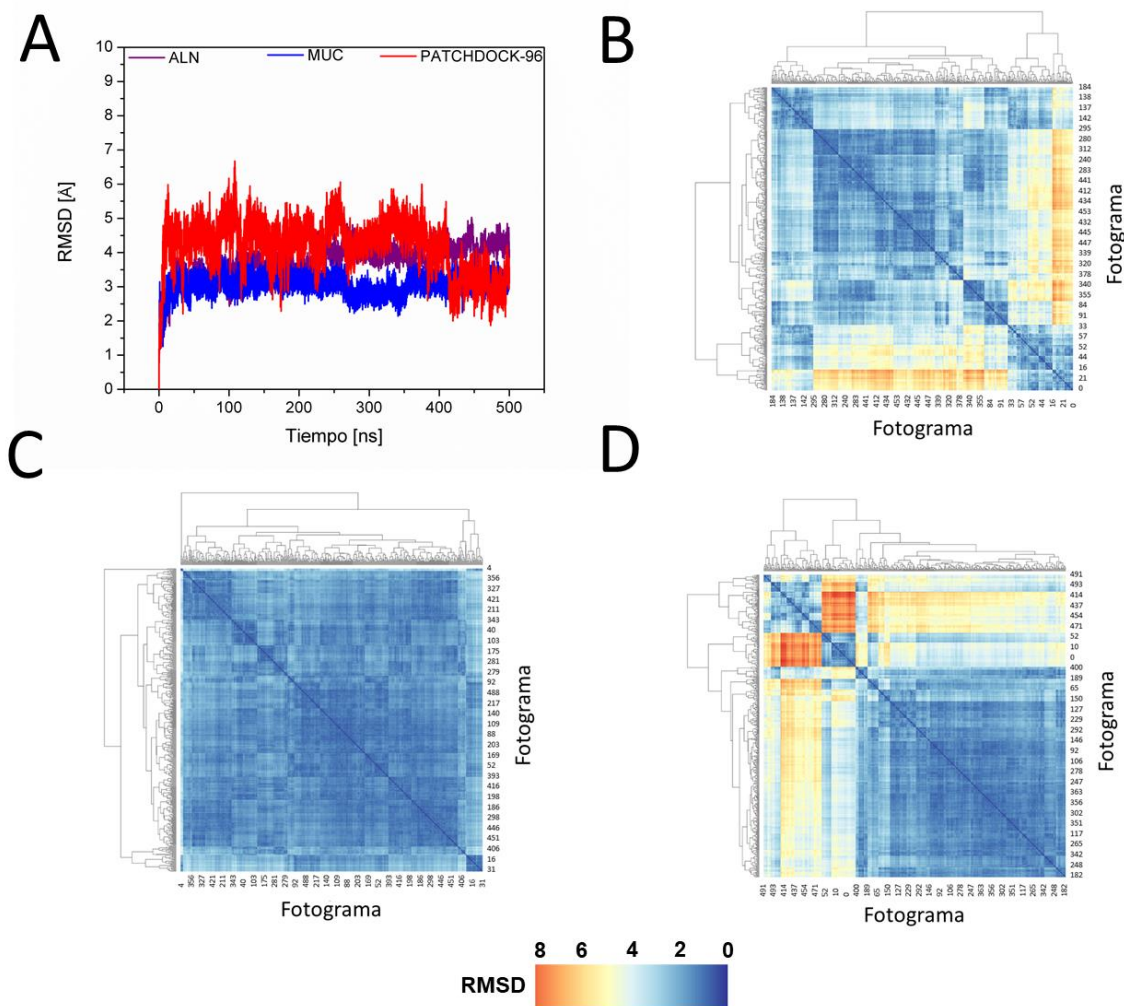
conformacional. En la figura 19B, se muestran los valores del radio de giro en función del tiempo. Estos valores se mantienen constantes en los tres sistemas, lo cual indica que la estructura espacial de la proteína es estable y compacta a lo largo de la SDM. Asimismo, esto sugiere que la interacción con DAD no compromete la estabilidad de la estructura de AKR1B1.



**Figura 20.** El gráfico presenta los valores de RMSF de los Cα de AKR1B1 en los tres complejos estructurales durante la SDM. En los recuadros dorados se seleccionan las regiones de residuos que presentaron mayor fluctuación, mientras que en los recuadros rojo, morado y verde se señalan las regiones correspondientes a las hélices  $\alpha 8$ ,  $\alpha 7$  y H2, respectivamente. Asimismo, se presenta un esquema de la estructura secundaria en la que pertenecen los residuos. Las líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

Para comparar la fluctuación de los residuos de AKR1B1 en los tres complejos estructurales se realizó un gráfico del RMSF (ver figura 20). Donde se nota que el complejo MUC presenta menores fluctuaciones en sus residuos con respecto a los otros sistemas, siendo el complejo ALN en donde se observa las mayores

fluctuaciones en AKR1B1. También se aprecia que una de las regiones más fluctuantes en los tres complejos corresponde a un asa que abarca de los residuos E120-T140, dicho comportamiento ya se había observado en las simulaciones de la proteína sola. Finalmente, en los recuadros marcados en rojo, morado y verde que señalan las hélices  $\alpha 8$ ,  $\alpha 7$  y H2 respectivamente, no hay fluctuaciones grandes, lo cual indica que es una región dinámicamente estable.

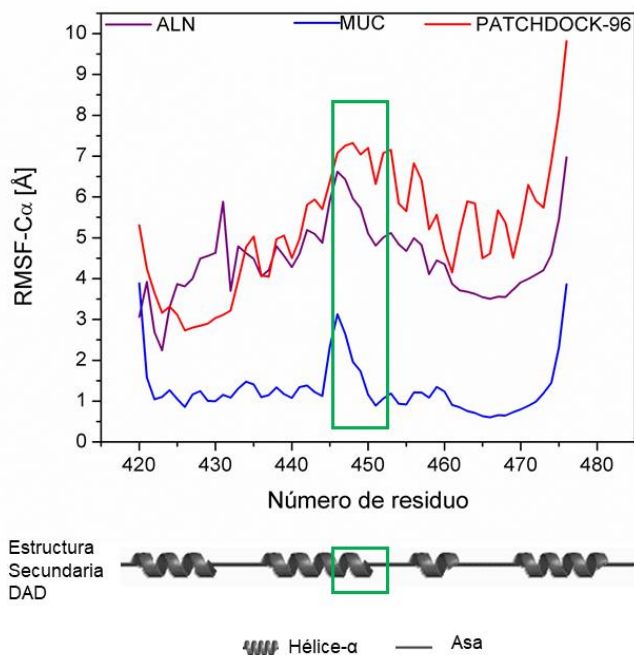


**Figura 21.** En el gráfico A se muestra el RMSD del esqueleto proteico de DAD de la trayectoria con respecto a la conformación inicial. Por otro lado, en el gráfico B se observa una comparación de RMSD y agrupamiento jerárquico entre las conformaciones adoptadas por DAD en SDM de 500 ns del complejo ALN, mientras que C y D corresponden a los complejos MUC y PATCHDOCK-96 respectivamente. En los mapas de calor B, C y D, cada conformación corresponde a un fotograma en intervalos de tiempo de 1 ns. Las líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

En lo que respecta al comportamiento estructural de DAD durante las SDM, en la figura 21A, se observa que el comportamiento del RMSD con respecto a la conformación inicial más estable corresponde al del complejo MUC, ya que las conformaciones adquiridas durante la trayectoria no presentan diferencias mayores a 4.0 Å con respecto a la pose de inicio, en donde el RMSD promedio de DAD fue de  $3.05 \pm 0.28$  Å. Asimismo, el panel C muestra que las variaciones entre conformaciones a diferentes tiempos de la dinámica es baja ( $< 4.0$  Å), lo cual sugiere que el modo de unión de DAD es muy estable durante el periodo de simulación. Por el contrario, el modo de unión del sistema PATCHDOCK-96 se comportó de manera muy inconsistente durante la trayectoria, presentando diferencias ( $>4.0$  Å) en el RMSD entre las conformaciones de inicio con las del final (ver panel D). También se observaron diferencias estructurales entre las conformaciones adquiridas en la trayectoria con respecto a la conformación inicial, con un RMSD promedio de  $4.96 \pm 0.92$  Å, estos datos indican que el modo de unión de DAD en este sistema no es estable.

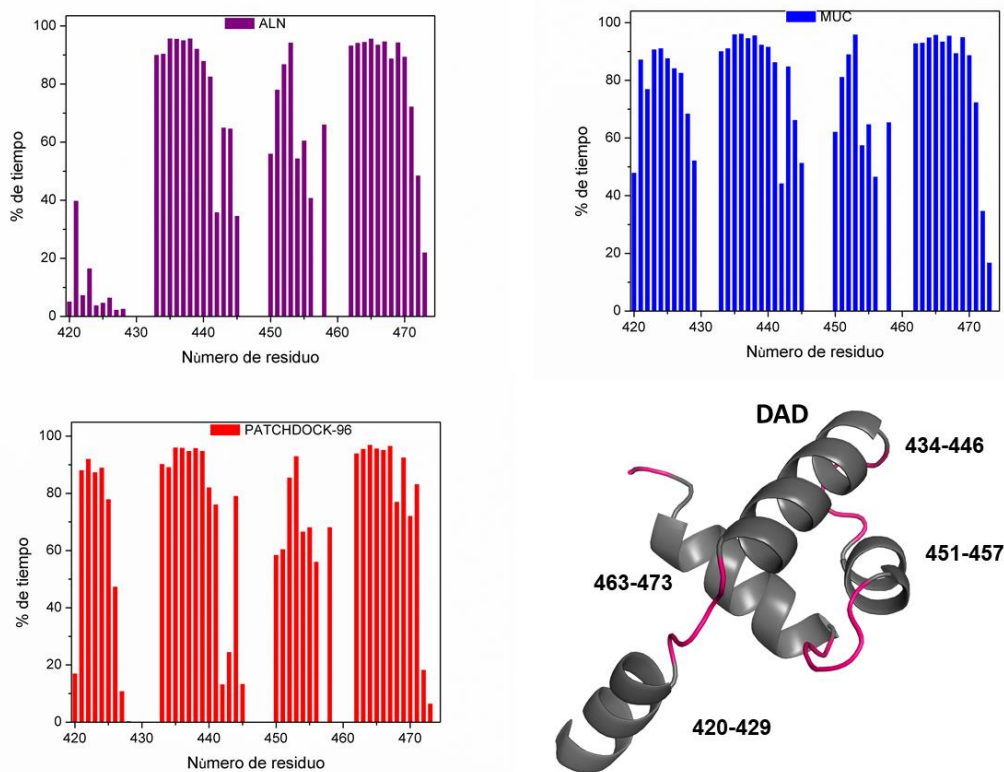
Finalmente, el modo de unión de DAD en el complejo ALN fue ligeramente constante durante dos intervalos de tiempo, el primero durante los primeros 200 ns (0-200 ns) y el segundo durante los últimos 250 ns (250-500). Asimismo, algunas conformaciones a diferentes tiempos presentaron diferencias con respecto a la conformación inicial, obteniéndose un RMSD promedio de  $3.79 \pm 0.48$  Å. Esto sugiere que el modo de unión de DAD no es estable en comparación con el complejo MUC.





**Figura 22.** El gráfico presenta los valores de RMSF de los C $\alpha$  de DAD en los tres complejos estructurales durante la SDM. En el recuadro verde se selecciona la región de residuos que presentaron mayor fluctuación. Asimismo, se presenta un esquema de la estructura secundaria en la que pertenecen los residuos. Las líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

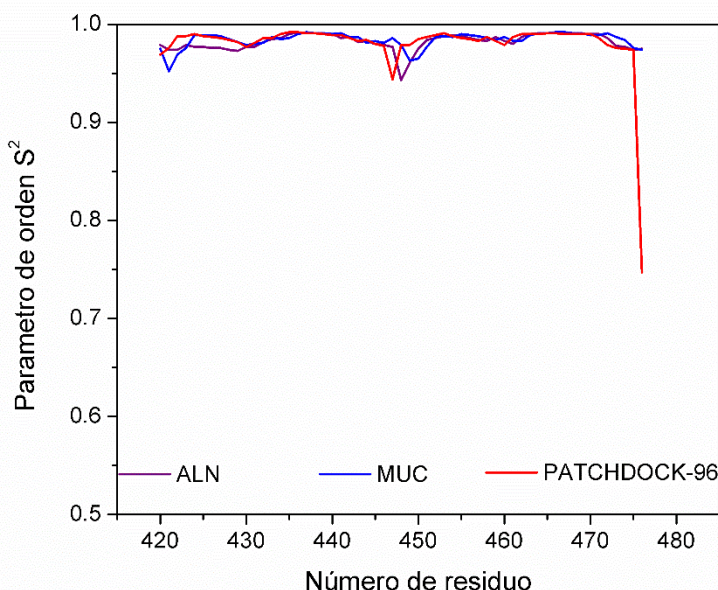
El análisis de RMSF (figura 22) muestra que los residuos de DAD son muy fluctuantes en los complejos ALN y PATCHDOCK-96, mientras que en el complejo MUC la fluctuación es mínima, lo cual indica que DAD presenta más restricciones en su movimiento debido a una unión estable con AKR1B1. En los tres sistemas se observa que la región que presenta mayor flexibilidad es de los residuos 444-450, los cuales están en la región terminal de una hélice- $\alpha$  y al asa a la que está próxima.



**Figura 23.** Porcentaje de tiempo en el que los residuos de DAD presentan helicidad. El gráfico en barras púrpura corresponde al sistema ALN, mientras que los de color azul y rojo son del complejo MUC y PATCHDOCK-96, respectivamente. Finalmente, se presenta una imagen de la estructura de DAD, en la cual se muestran las hélices- $\alpha$  en color gris y las asas en color magenta, asimismo se indican los residuos que componen cada una de estas estructuras secundarias.

El parámetro de helicidad juega un papel importante en la especificidad y actividad biológica de los péptidos. Se ha mostrado que existe una correlación entre la helicidad con una alta especificidad en péptidos<sup>149</sup>. En la figura 23 se presentan gráficos del % de tiempo en el que los residuos de DAD mantienen la helicidad en los tres sistemas. En los gráficos se observan que existen cuatro regiones que presentan helicidad durante un alto porcentaje de tiempo, las cuales corresponden a las cuatro hélices- $\alpha$  presentes en DAD, mientras que los residuos correspondientes a las asas muestran valores de 0. Las hélices- $\alpha$  correspondientes a los residuos 434-446, 451-457 y 463-473 presentan un comportamiento similar en los tres sistemas. Sin embargo, la hélice- $\alpha$  de los residuos 420-429 presentan helicidad en poco porcentaje de tiempo en el sistema

ALN, mientras que en sistema de PATCHDOCK-96 los valores son menores a los de DAD en el sistema MUC. Estos resultados indican que DAD en el complejo MUC presenta mejor helicidad, lo cual sugiere que la unión de DAD es más específica en este sistema.



**Figura 24.** Parámetro de orden generalizado  $S^2$  del enlace N-H en los residuos de DAD de los tres complejos estructurales durante la SDM. Las líneas morada, azul y roja corresponden a los sistemas ALN, MUC y PATCHDOCK-96, respectivamente.

El parámetro de orden generalizado  $S^2$  del esqueleto peptídico se utiliza como una medida del grado de restricción espacial del movimiento, que caracteriza la correlación angular de la dinámica del vector de enlace N-H y sus valores varían entre 0 y 1. Los valores más bajos indican movimientos de gran amplitud de un vector de enlace que no tiene restricciones y los valores más altos corresponden a movimientos restringidos de un enlace que es (en su mayoría) rígido<sup>150,151</sup>.

Con base en esto, se calculó el parámetro de orden del esqueleto peptídico (enlace N-H) de DAD para los tres sistemas (ver figura 24). En los tres casos se observa que el parámetro de orden es cercano a 1, esto indica que los movimientos internos del enlace N-H de DAD están restringidos, lo cual puede deberse a que DAD se conforma de cuatro  $\alpha$ -hélices y estas estructuras generalmente son rígidas. También se observa que la región que presenta mayor

movimiento interno en los tres sistemas corresponde a los residuos 446-450, la cual es la misma región que presentó mayores valores de fluctuación en el RMSF.

Finalmente, se determinó la energía libre de unión de los tres complejos estructurales AKR1B1-DAD, junto con el cálculo de la descomposición de energía por residuo mediante cálculos de MM-PBSA.

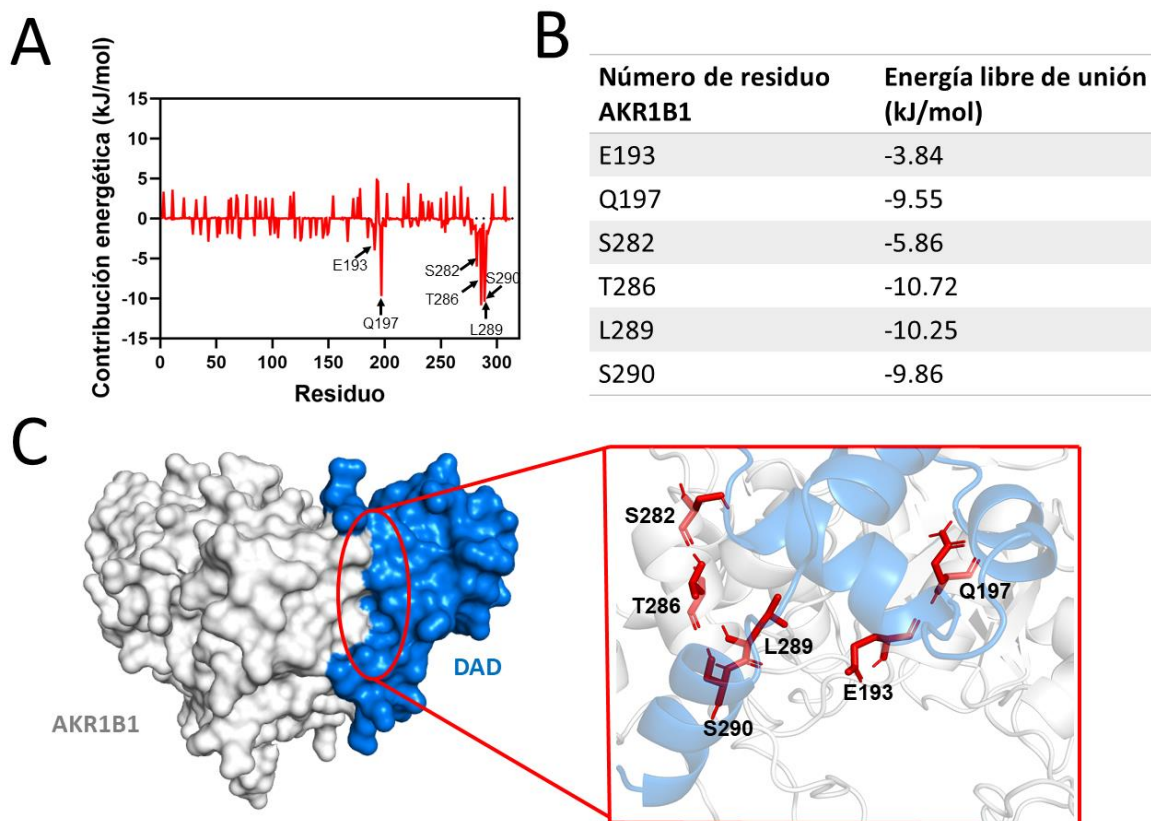
<b>Complejo estructural</b>	<b>Energía libre de unión <math>\Delta G</math> (kJ/mol)</b>	<b>Constante de disociación Kd teórica</b>
ALN	-18.83 $\pm$ 3.23 kJ/mol	1.9*10 <sup>3</sup>
MUC	-172.00 $\pm$ 2.37 kJ/mol	8.8*10 <sup>29</sup>
PATCHDOCK-96	27.93 $\pm$ 3.12 kJ/mol	1.3*10 <sup>-5</sup>

**Tabla 9.** Energía libre de unión de los complejos estructurales AKR1B1-DAD en SDM de 500 ns.

En la tabla 9 se muestran los valores de  $\Delta G$  de unión de los complejos estructurales AKR1B1-DAD mediante cálculos de MM-PBSA. Se observa que el complejo estructural MUC presenta el valor de energía de unión más favorable (-172 kJ/mol), mientras que el complejo PATCHDOCK-96 presentó un valor de energía de unión desfavorable (27.93 kJ/mol). Esto indica que el complejo estructural MUC es el más estable energéticamente, por lo tanto, estos resultados en conjunto con los mencionados anteriormente señalan que el modo de unión predicho por el acoplamiento molecular es el más adecuado. Por consiguiente, este complejo se denominará complejo WT para el análisis de mutagénesis *in silico*. Sin embargo, para conocer de manera exacta cuáles son los residuos que aportan más a la energía de unión, se hizo una descomposición energética por residuo.

En la figura 25 se observa que aquellos residuos que más aportan a la energía libre de unión se encuentran en las hélices  $\alpha 8$  (**E193 y Q197**) y H2 (**S282, T286, L289 y S290**) de AKR1B1, en donde cabe destacar que ya se ha reportado de manera experimental la importancia de la L289 para la IPP. Estos resultados sugieren que la región de residuos compuesta por la hélice H2 y hélice- $\alpha 8$  es la responsable de establecer IPP más estables en AKR1B1, siendo Q197, T286,

L289 y S290 los residuos que aportan más energía de unión. Asimismo, en la tabla 10 se presentan aquellos residuos de DAD que más favorecieron y desfavorecieron la interacción con AKR1B1. Estos resultados en conjunto indican que los residuos mencionados pueden ser relevantes para el diseño de fármacos y péptidos inhibidores de las IPP en AKR1B1.

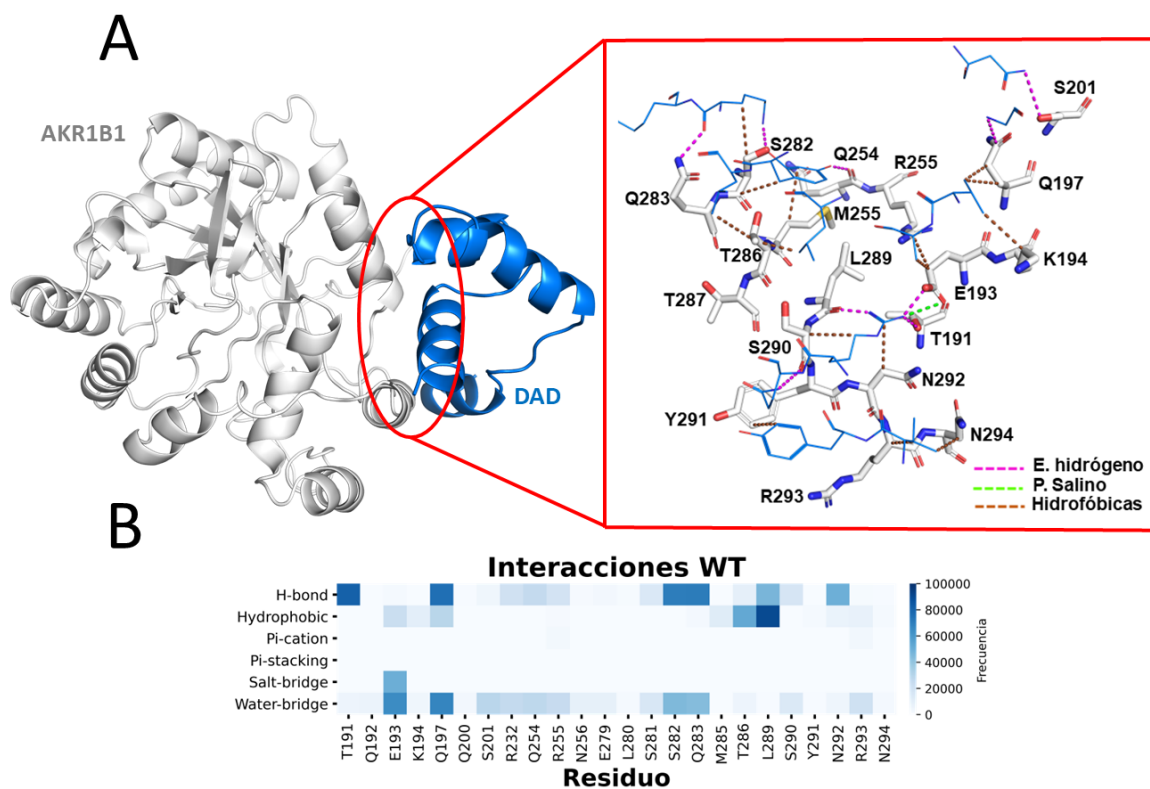


**Figura 25.** Contribución energética por residuo de AKR1B1 en el complejo estructural WT. En el panel A se presenta una gráfica de la contribución energética por residuo, en flechas negras se destacan los residuos que más aportan a la energía libre de unión (E193, Q197, S282, T286, L289 y S290). En el panel B se muestra una tabla con los valores de energía en kJ/mol de los residuos más relevantes. Finalmente, en el panel C se muestra la localización espacial de los residuos en la estructura, presentándolos en forma de bastones de color rojo.

Número de residuo DAD	Energía libre de unión (kJ/mol)
M420	4.73
V422	-11.93
Y423	-14.37
K424	4.26

D425	-9.32
R426	19.02
P451	-7.16

**Tabla 10.** Energía libre de unión de residuos de DAD que más favorecieron y desfavorecieron la unión con AKR1B1.



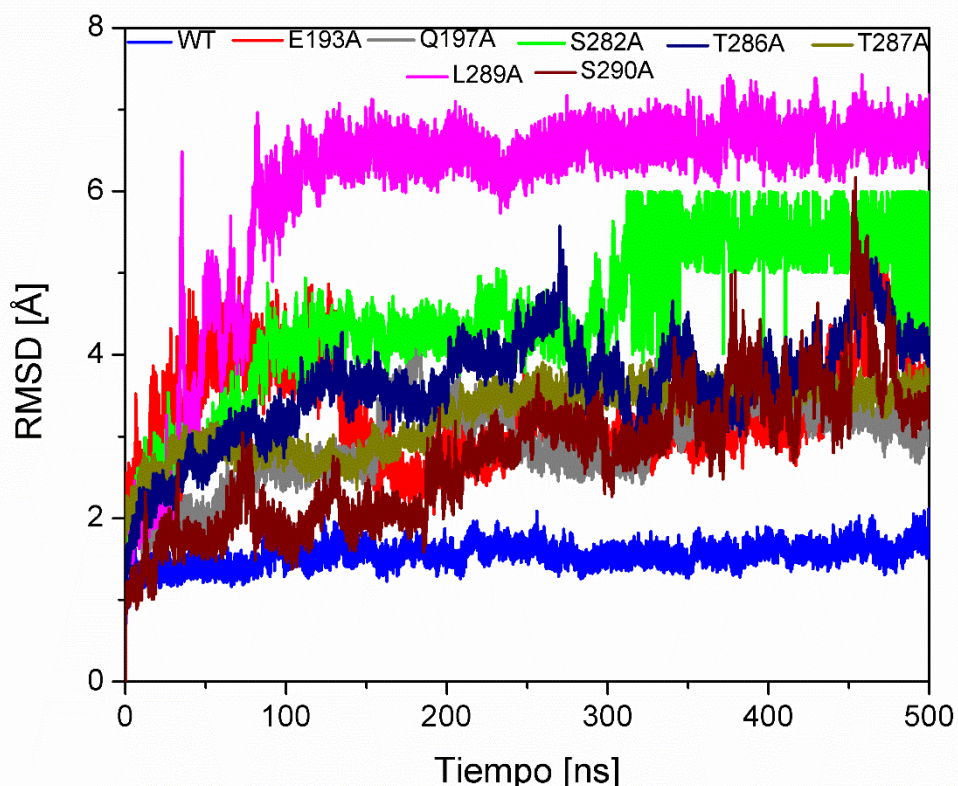
**Figura 26.** Interacciones intermoleculares del complejo WT. En el panel A se muestran las interacciones que establece AKR1B1 de una pose representativa de la SDM de 500. Los residuos de AKR1B1 se muestran en bastones de color blanco, mientras que los residuos de DAD se representan con líneas de color azul. Las interacciones tipo enlace de hidrógeno se denotan con líneas punteadas color magenta, mientras que los puentes salinos se representan con líneas punteadas color verde, finalmente las interacciones hidrofóbicas se muestran con líneas punteadas de color café. En el panel B se presenta un mapa de calor que indica la frecuencia con la que los residuos establecen interacciones de enlace de hidrógeno, hidrofóbicas,  $\pi$ -catión, apilamiento, puente salino y puente de agua durante la SDM de 500 ns.

En el complejo WT, las interacciones con más frecuencia que estableció AKR1B1 durante la SDM fueron enlaces de hidrógeno e interacciones hidrofóbicas. Se observó que los residuos T191, Q197, S282, S283 y L289 fueron los que

establecieron predominantemente interacciones tipo enlace de hidrógeno, mientras que los residuos E193, Q197, T286 y L289 estaban implicados en la formación de interacciones hidrofóbicas con mayor frecuencia. Finalmente, solo el residuo E193 fue capaz de establecer interacciones tipo puente salino.

### 6.5 Mutagénesis de alanina *in silico*

Las mutaciones con ALA es un enfoque para determinar la contribución de un residuo específico a la estabilidad o función de una proteína determinada.



**Figura 27.** Análisis comparativo del RMSD del esqueleto proteico de AKR1B1 en el complejo estructural AK-DAD WT y de los complejos mutantes.

El análisis de RMSD anterior (figura 27) mostró que el complejo estructural WT tiene un RMSD estable a lo largo de los 500 ns de ejecución de producción en comparación con los siete mutantes estructurales. Las proteínas AKR1B1 mutantes S282A, T286A, L289A y S290A fueron los sistemas en donde AKR1B1 presentó mayores cambios estructurales durante la SDM, mostrando RMSD promedios de  $4.46 \pm 0.92$  Å,  $3.61 \pm 0.63$  Å,  $6.02 \pm 1.29$  Å y  $3.30 \pm 0.61$  Å,

respectivamente. Estos resultados indican que estas mutaciones son las que más influyen en la dinámica estructural de la proteína.

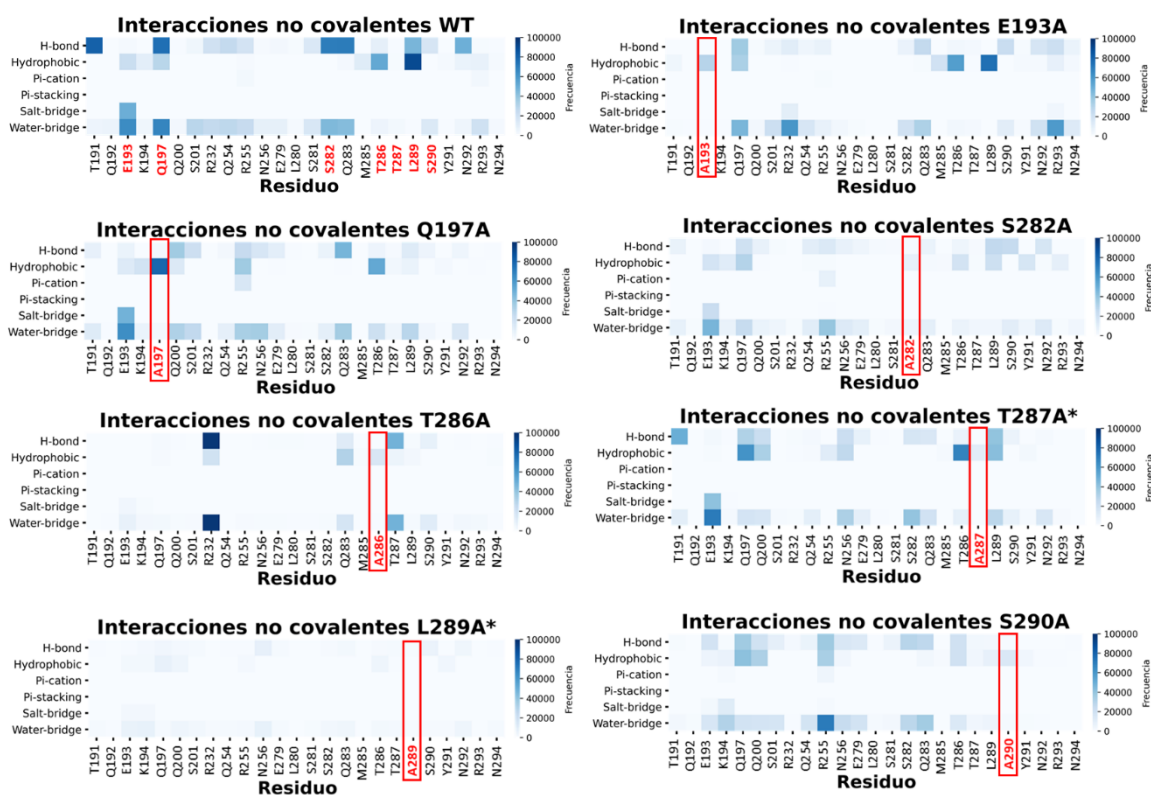
La estabilidad de los complejos estructurales se evaluó mediante la energía libre de unión y el  $\Delta\Delta G$  con respecto al complejo WT (ver tabla 11). Los complejos mutantes presentaron un valor de  $\Delta G$  menos favorable que el complejo WT. Estos datos indican que hay una desestabilización del complejo cuando estos residuos no están presentes, lo cual sugiere que estos residuos de la interfaz determinan una fuerte estabilización y son elementos esenciales para la formación regular del complejo. Se observa que los mutantes que afectaron más la estabilidad del complejo estructural AKR1B1-DAD son S282A, L289A y S290A, lo cual también se puede ver por los valores de  $\Delta\Delta G$ . La mutante L289A es la que desestabiliza más el complejo estructural. Estos datos son consistentes con lo que ya se ha reportado de manera experimental en la literatura por Thiagarajan y colaboradores<sup>10</sup>, en donde se determinó que la mutación de L289A disminuía de manera significativa la IPP con DAD. Asimismo, reportaron que la mutante T287A no afectó de manera importante la interacción con DAD. Esto también se ve reflejado en nuestros resultados, ya que fue el complejo que presentó un valor de  $\Delta G$  más cercano al complejo WT, y su valor de  $\Delta\Delta G$  fue el más bajo entre los complejos mutantes analizados. Por otro lado, los valores de  $K_d$  teóricos indican que los cambios debido a las mutaciones son de grandes ordenes de magnitud, lo cual sugiere que hay una desestabilización de manera significativa de los complejos mutantes.

<b>Complejo de IPP AKR1B1-DAD</b>	<b>Energía libre de unión <math>\Delta G</math> (kJ/mol)</b>	<b><math>\Delta\Delta G</math> (kJ/mol)</b>	<b>Constante de disociación <math>K_d</math> teórica</b>
WT	<b>-172.00 ± 2.37</b>	<b>0.00</b>	$8.8 \cdot 10^{29}$
E193A	-77.31 ± 5.01	94.69	$2.8 \cdot 10^{13}$
Q197A	-88.47 ± 2.67	83.53	$2.5 \cdot 10^{15}$
S282A	-32.29 ± 4.76	150.71	$4.2 \cdot 10^5$
T286A	-63.46 ± 2.57	108.54	$1.1 \cdot 10^{11}$
T287A*	-122.90 ± 2.94	49.10	$1.7 \cdot 10^{21}$
L289A*	17.51 ± 3.68	189.00	$8.9 \cdot 10^{-4}$
S290A	3.49 ± 3.70	175.49	0.25



**Tabla 11.** Determinación de la energía libre de unión mediante cálculos MM/PBSA comparando el complejo tipo salvaje (WT) y los complejos generados con mutantes en residuos de la interfaz de AKR1B1 energéticamente importantes. Nota: \* complejo cuya mutación ya se ha caracterizado de manera experimental.

Las mutaciones con el residuo de alanina se deben a que simula la pérdida de la cadena lateral más allá del carbono β, ya que posee un grupo metilo que es poco voluminoso y no altera la conformación de la cadena principal, además no impone efectos electrostáticos y estéricos extremos<sup>152</sup>. Debido a esto se analizaron las interacciones de los complejos mutantes para analizar la influencia que tienen las cadenas laterales y la posición de los residuos en la IPP.



**Figura 28.** Análisis de interacciones intermoleculares de complejos AKR1B1-DAD WT y mutantes. En la figura se presentan mapas de calor de la frecuencia con la que residuos de AKR1B1 establecieron interacciones tipo puente de hidrógeno, hidrofóbicas, π-catión, apilamiento, puente salino y puente de agua durante la SDM de 500 ns. En los recuadros rojos se resaltan los mutantes en cada sistema.

En la figura 28, se observa un panorama general de las interacciones intermoleculares que se establecen con mayor frecuencia en los diferentes complejos estructurales mutantes. Para la mutación E193A se aprecia que hay menor frecuencia en las interacciones de enlace de hidrógeno e hidrofóbicas con respecto al complejo WT. Asimismo, el residuo 193 perdió las interacciones tipo puente salino que establecía en WT, lo cual indica la importancia del grupo carbonilo de la cadena lateral en el E193. Lo mismo se observó para la mutación S282, en donde se aprecia una pérdida importante de las interacciones de enlace de hidrógeno por parte de la cadena lateral del residuo 282, resaltando la importancia de la presencia del grupo hidroxilo. En el caso de la mutante Q197A, vemos que presenta una pérdida de las interacciones de enlace de hidrógeno. Sin embargo, se observa una mayor frecuencia en las interacciones hidrofóbicas, lo cual indica que en este residuo el grupo amida se encarga de establecer los enlaces de hidrógeno mientras que la cadena principal tiene las interacciones hidrofóbicas.

En la mutante T286A, se observa que el residuo 286 presenta una disminución considerable en la frecuencia de las interacciones hidrofóbicas y una pérdida total de las interacciones de enlace de hidrógeno, por lo que estas interacciones se le atribuyen a la cadena lateral en el complejo WT. El residuo que presentó mayores consecuencias debido a la mutación fue L289A, al observar los resultados, se puede observar que hay una disminución importante en la frecuencia de las interacciones que establecen todos los residuos de AKR1B1. Asimismo, el residuo 289 presentó una pérdida total de las interacciones hidrofóbicas y de enlace de hidrógeno con respecto al complejo WT, destacando que la cadena lateral de este residuo es esencial para el establecimiento de estas interacciones.

Finalmente, en el complejo mutante S290A, vemos que no hay cambios tan considerables en ganancia o pérdida de interacciones, por lo cual la importancia de este residuo se puede deber a su cercanía con la L289, ya que una pequeña desestabilización en las interacciones de este residuo puede afectar la unión de L289 que es un residuo esencial para la interacción con DAD.



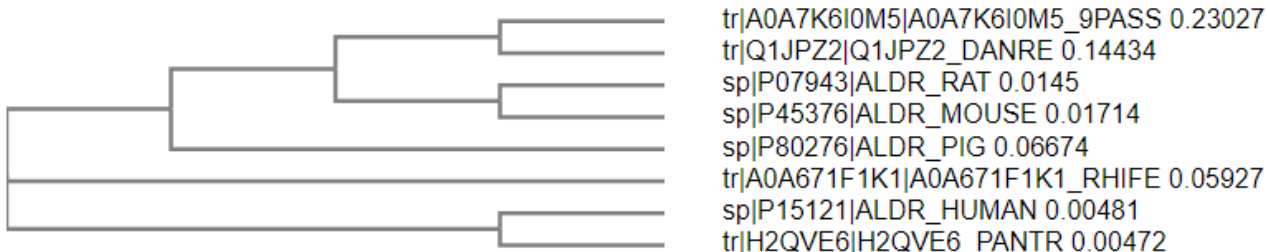
```

tr|A0A7K6I0M5|A0A7K6I0M5_9PASS      HNKTPAQVLR LQIQRN VSVIPK SVTPQRIE ENFKVDF FELTEEMQTL LGFQKRHHICA      300
tr|Q1JPZ2|Q1JPZ2_DANRE              HKKTTAQVLR IRFQIQRN VIVIPK SITPQRIE ENFQVDF FELSEEDMK TILSFNRNFRACP      300
sp|P07943|ALDR_RAT                   YNKTTAQVLR IRFPIQRNL VVIPK SVTPARIA ENFKVDF FELSNEDMAT LLSYNRNWRVCA      300
sp|P45376|ALDR_MOUSE                 YNKTTAQVLR IRFPIQRNL VVIPK SVTPVRIA ENLKVDF EVSSEDMAT LLSYNRNWRVCA      300
sp|P80276|ALDR_PIG                   YNKTTAQVLR IRFPMQRNL VVIPK SVTPERIA ENFQVDF FELSPEDMNT LLSYNRNWRVCA      300
tr|A0A671F1K1|A0A671F1K1_RHIFE      YNKTTAQVLR IRFPMQRNL VVIPK SVTPERIA ENFQVDF FELSTEDMTT LLSYNRNWRICA      300
sp|P15121|ALDR_HUMAN                 HNKTTAQVLR IRFPMQRNL VVIPK SVTPERIA ENFKVDF FELSSQDMTT LLSYNRNWRVCA      300
tr|H2QVE6|H2QVE6_PANTR              HNKTTAQVLR IRFPMQRNL VVIPK SVTPERIA ENFKVDF FELSSQDMTT LLSYNRNWRVCA      300
:.* *****: .***: *****:* * *:*****: :.* *:*.:. : *

tr|A0A7K6I0M5|A0A7K6I0M5_9PASS      LSECKNHKDY PFWED-      315
tr|Q1JPZ2|Q1JPZ2_DANRE              MHMGVKHRDYP FNAEY      316
sp|P07943|ALDR_RAT                   LMSCAKHKDYP FFAEV      316
sp|P45376|ALDR_MOUSE                 LMSCAKHKDYP FFAEV      316
sp|P80276|ALDR_PIG                   LMSCASHKDYP FHEE-      315
tr|A0A671F1K1|A0A671F1K1_RHIFE      LVSCTSHKDYP FKEEF      316
sp|P15121|ALDR_HUMAN                 LLSCTSHKDYP FHEEF      316
tr|H2QVE6|H2QVE6_PANTR              LLSCTSHKDYP FHEE-      315
: .*:***** :

```

B



C

Secuencia	Organismo	% de identidad vs AKR1B1 humana
tr A0A7K6I0M5 A0A7K6I0M5_9PASS	<i>Dasyornis broadbenti</i>	64.13
tr Q1JPZ2 Q1JPZ2_DANRE	<i>Danio rerio</i>	73.10
sp P07943 ALDR_RAT	<i>Rattus norvegicus</i>	85.44
sp P45376 ALDR_MOUSE	<i>Mus musculus</i>	85.44
sp P80276 ALDR_PIG	<i>Sus scrofa</i>	86.98
tr A0A671F1K1 A0A671F1K1_RHIFE	<i>Rhinolophus ferrumequinum</i>	88.92
tr H2QVE6 H2QVE6_PANTR	<i>Pan troglodytes</i>	99.05

**Figura 29.** Alineamiento múltiple de secuencias de AKR1B1. En el panel A se presenta el alineamiento de las secuencias, en donde las identidades de secuencia se indican con asteriscos (\*), las sustituciones conservadas con dos puntos (: ) y las sustituciones semi-conservadas con un punto (.). Asimismo, los aminoácidos están coloreados con base en su clasificación por características fisicoquímicas. En color rojo se muestran los aminoácidos no polares, en verde los polares sin carga, en rosa los aminoácidos polares con carga positiva y en azul los polares con

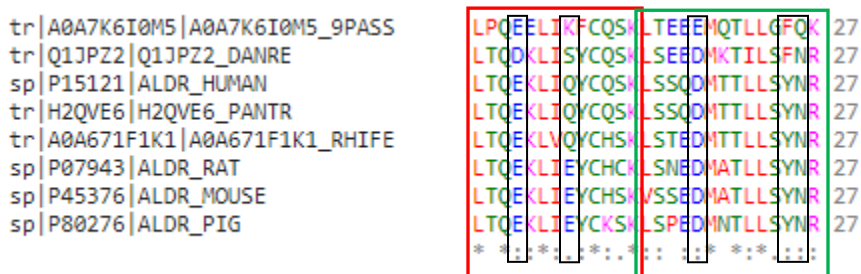
carga negativa. En el panel B se muestra un árbol filogenético con base en la identidad de secuencia. Finalmente, en el panel C se presentan el % de identidad de secuencia con respecto a AKR1B1 humana.

En el análisis se puede observar que 158 de los 316 residuos están totalmente conservados entre las ocho secuencias, 53 de los 316 residuos tienen sustituciones conservadas, 22 de los 316 residuos presentan sustituciones semi-conservadas y 83 residuos no están conservados. Además, se observa que el organismo con el que se presenta mayor identidad de secuencia es con el chimpancé, mientras que el de menor % de identidad de secuencia es del ave *Dasyornis broadbenti*, lo cual indica que el % de identidad de esta proteína disminuye conforme aumenta la diferencia filogenética.

De los 83 residuos que no presentan sustituciones conservadas, se observa que, en la mayoría de los casos, estas sustituciones se encuentran en dos o tres secuencias, lo cual sugiere que estos residuos no podrían ser funcionalmente importantes en ciertos organismos.

Al analizar el alineamiento múltiple de secuencias de manera local, centrándonos en las hélices  $\alpha 8$  y H2 que se determinaron como relevantes mediante el análisis estructural *in silico*, se puede apreciar un patrón similar al alineamiento global. Sin embargo, vemos que hay una ligera disminución en el % de identidad en todas las secuencias, con excepción de la de chimpancé.

A



B

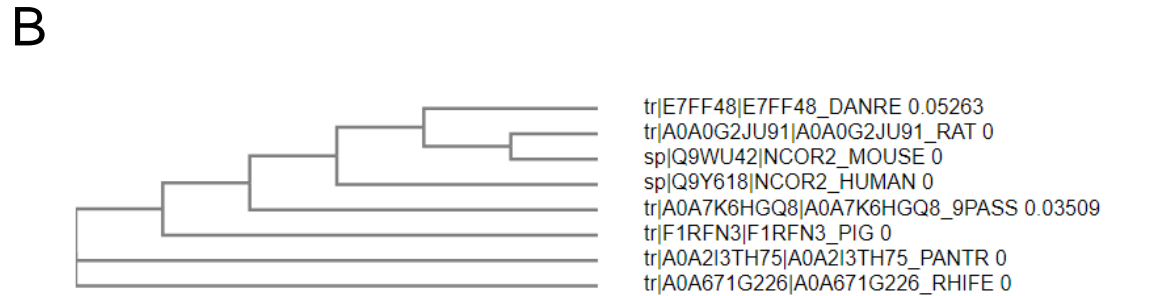
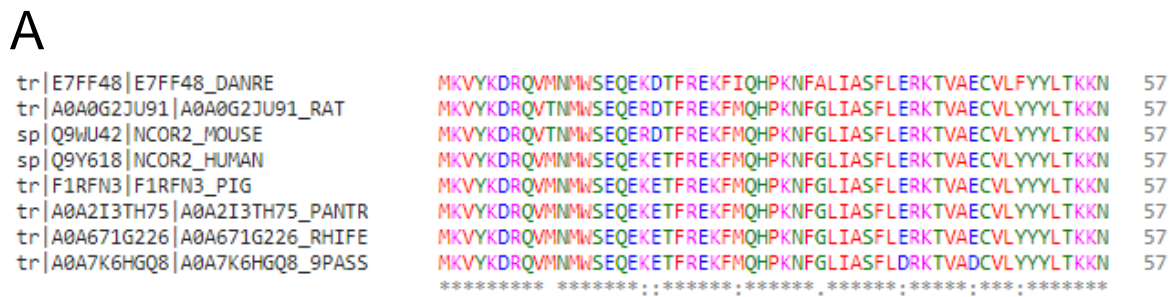
Secuencia	Organismo	% de identidad vs AKR1B1 humana
tr A0A7K6I0M5 A0A7K6I0M5_9PASS	<i>Dasyornis broadbenti</i>	51.85
tr Q1JPZ2 Q1JPZ2_DANRE	<i>Danio rerio</i>	74.07
sp P07943 ALDR_RAT	<i>Rattus norvegicus</i>	77.88
sp P45376 ALDR_MOUSE	<i>Mus musculus</i>	81.48
sp P80276 ALDR_PIG	<i>Sus scrofa</i>	81.48
tr A0A671F1K1 A0A671F1K1_RHIFE	<i>Rhinolophus ferrumequinum</i>	85.19
tr H2QVE6 H2QVE6_PANTR	<i>Pan troglodytes</i>	100.00

**Figura 30.** Alineamiento múltiple de secuencias de los aminoácidos de las hélices  $\alpha 8$  y H2 de AKR1B1. En el panel A se presenta el alineamiento de las secuencias, en donde las identidades de secuencia se indican con asteriscos (\*), las sustituciones conservadas con dos puntos (:) y las sustituciones semi-conservadas con un punto (.). Asimismo, los aminoácidos están coloreados con base en su clasificación por características fisicoquímicas. En color rojo se muestran los aminoácidos no polares, en verde los polares sin carga, en rosa los aminoácidos polares con carga positiva y en azul los polares con carga negativa. Los recuadros de color rojo y verde corresponden a los aminoácidos de las hélices  $\alpha 8$  y H2 respectivamente, mientras que los recuadros negros corresponden a los residuos energéticamente importantes (193, 197, 282, 286, 289 y 290) que fueron determinados con análisis estructural *in silico* en AKR1B1 humana. En el panel B se presentan el % de identidad de secuencia con respecto a AKR1B1 humana.

En los 27 aminoácidos que constituyen las hélices  $\alpha 8$  y H2, solo 8 están totalmente conservados, mientras que 13 tienen sustituciones conservadas, 3 presentan sustituciones semi-conservadas y de la misma forma, 3 aminoácidos no presentan sustituciones conservadas. Con respecto a los aminoácidos remarcados en los recuadros negros, se observa que el aminoácido 193 presenta solamente una sustitución que se da en la secuencia del pez cebra. Dicha sustitución es conservada. Los aminoácidos 197 y 282 tienen tres y cuatro sustituciones respectivamente, con respecto a la secuencia humana. En los organismos donde estos residuos presentan al menos un aminoácido conservado junto con una sustitución conservada o semi-conservada son el chimpancé, murciélago y ratón.

Lo cual sugiere que estos residuos solamente podrían ser relevantes en humano, chimpancé, murciélago y ratón o puede que funcionalmente no sean importantes en ningún organismo. La L289 es un aminoácido totalmente conservado en las ocho secuencias analizadas, lo cual resalta aún más su relevancia funcional. Finalmente, el aminoácido 290 solamente presenta una sustitución semi-conservada en la secuencia del ave *Dasyornis broadbenti*, el cual es un organismo filogenéticamente alejado del humano.

También se realizó un alineamiento múltiple de secuencias de DAD para encontrar si existe alguna relación entre los cambios en la secuencia de AKR1B1 y DAD en los mismos organismos. En la figura 31 se muestran los resultados del alineamiento y se observa que DAD es un dominio altamente conservado en los ocho organismos analizados, en donde el % de identidad de secuencia es mayor al 90% con respecto a DAD humano. El único cambio no conservado se presenta en el aminoácido 429. En humano este aminoácido corresponde a la metionina, y en rata y ratón se sustituye por treonina. Dicho residuo no fue energéticamente favorable o desfavorable para la unión con AKR1B1 mediante análisis estructural *in silico*. Otros cambios que se observan se presentan en los aminoácidos M445 y Y469, en el pez cebra estos aminoácidos están sustituidos por I445 y F469, estas posiciones no parecen ser relevantes en la IPP de la interfaz AKR1B1-DAD predicha.



C

Secuencia	Organismo	% de identidad vs AKR1B1 humana
tr A0A7K6I0M5 A0A7K6I0M5_9PASS	<i>Dasyornis broadbenti</i>	96.49
tr Q1JPZ2 Q1JPZ2_DANRE	<i>Danio rerio</i>	92.98
sp P07943 ALDR_RAT	<i>Rattus norvegicus</i>	94.74
sp P45376 ALDR_MOUSE	<i>Mus musculus</i>	94.74
sp P80276 ALDR_PIG	<i>Sus scrofa</i>	100.00
tr A0A671F1K1 A0A671F1K1_RHIFE	<i>Rhinolophus ferrumequinum</i>	100.00
tr H2QVE6 H2QVE6_PANTR	<i>Pan troglodytes</i>	100.00

**Figura 31.** Alineamiento múltiple de secuencias de DAD. En el panel A se presenta el alineamiento de las secuencias, en donde las identidades de secuencia se indican con asteriscos (\*), las sustituciones conservadas con dos puntos (:) y las sustituciones semi-conservadas con un punto (.). Asimismo, los aminoácidos están coloreados con base en su clasificación por características fisicoquímicas. En color rojo se muestran los aminoácidos no polares, en verde los polares sin carga, en rosa los aminoácidos polares con carga positiva y en azul los polares con carga negativa. En el panel B se muestra un árbol filogenético con base en la identidad de secuencia. Finalmente, en el panel C se presentan el % de identidad de secuencia con respecto a DAD humana.

En resumen, la investigación del presente trabajo condujo a la identificación de residuos energéticamente relevantes para la IPP entre la aldoceto-reductasa AKR1B1 y el dominio DAD, mediante la predicción y caracterización del complejo estructural que forman, empleando métodos computacionales como el acoplamiento molecular, SDM y mutagénesis de alanina *in silico*.

Primeramente, se realizó la predicción del modo de unión entre las dos proteínas, para esto se realizaron acoplamientos moleculares. Al tratarse de métodos de predicción, los programas de acoplamiento tienen sus propias limitaciones, por lo tanto, no existe un algoritmo único que pueda predecir con total certeza todas las conformaciones posibles de complejos proteína-proteína, proteína-ligando, etcétera.



En consecuencia, es necesario utilizar diferentes estrategias que permitan efectuar predicciones confiables. Para este fin, empleamos tres diferentes servidores que trabajan con diferentes algoritmos de búsqueda y funciones de evaluación, y además han mostrado ser eficientes en las evaluaciones CAPRI. La idea de esta estrategia es que, si los algoritmos de búsqueda y funciones de evaluación diferentes dan resultados similares, entonces estas predicciones pueden ser las más probables y de este modo es posible establecer un modo de unión consenso. También, estos servidores fueron sometidos a ensayos de validación, con la finalidad de determinar si son capaces de reproducir la conformación cristalográfica del complejo HDAC3-DAD. La evaluación de estas predicciones se hizo siguiendo los criterios que se establecen en CAPRI (contactos nativos, IRMSD y LRMSD)<sup>65</sup>. Las predicciones mejor evaluadas por los tres servidores lograron una calificación de alta calidad, asimismo, fueron estructuralmente similares a la referencia, lo cual sugiere que estos servidores son confiables para la predicción del complejo estructural AKR1B1-DAD.

Posteriormente, se realizaron los acoplamiento moleculares de AKR1B1 y DAD, en donde las soluciones mejor calificadas por los tres servidores eran muy similares, por lo tanto, las tres soluciones se evaluaron bajo una misma función de energía utilizando PRODYGY y se eligió la solución de menor energía libre de unión como el modo de unión consenso.

Aunque los acoplamiento moleculares permiten una predicción rápida de múltiples complejos proteína-proteína, las limitaciones de la metodología hacen que los resultados solo sean aproximaciones, ya que se requiere un análisis más detalle del espacio conformacional, con la finalidad de demostrar que la unión es estable. Además de observar posibles cambios conformacionales que se pudieran presentar tras la interacción entre ambas proteínas. Por lo tanto, se analizó mediante SDM el complejo AKR1B1-DAD predicho con acoplamiento molecular (MUC) y otros dos complejos que provenían de un alineamiento estructural entre

AKR1B1 y el complejo HDAC3-DAD donde se eliminó HDAC3, y un complejo predicho con acoplamiento molecular que obtuvo una baja calificación.

Los resultados obtenidos mostraron que el complejo MUC es el más estable estructuralmente. Las energías libres de unión para los tres sistemas proteína-proteína se calcularon utilizando el método MMPBSA. La confiabilidad de este método fue confirmada por varias publicaciones en la literatura. En donde se ha reportado la capacidad de predecir afinidades de unión relativas que han sido razonables con los datos experimentales<sup>153</sup>. Sin embargo, también hay ejemplos en los que la aplicación del método MMPBSA condujo a resultados insatisfactorios<sup>153</sup>, ya que su eficiencia depende de varios criterios como complejidad del sistema biológico, los parámetros usados en la SDM, la duración de la simulación, etcétera<sup>154</sup>.

Los cálculos teóricos de la energía libre de unión indicaron que el sistema MUC fue el más estable energéticamente. Además, se determinó que los residuos de AKR1B1 que aportan más a la energía libre de unión son E193, Q197, S282, T286, L289 y S290. Estos residuos fueron los que establecieron con más frecuencia interacciones hidrofóbicas y enlace de hidrógeno.

Por lo general, muy pocos residuos son esenciales para cualquier interfaz entre proteínas, estos residuos cruciales suelen denominarse “*hot-spots*” o puntos calientes funcionales y la mutación de estos residuos afecta significativamente la estabilidad del complejo proteína-proteína. El impacto de las mutaciones puede medirse mediante el cambio en la energía libre de unión entre el complejo tipo silvestre y el mutado<sup>155</sup>. Con relación a lo anterior, se decidió analizar que los residuos energéticamente importantes de AKR1B1 en el complejo MUC fueran puntos calientes funcionales, por lo tanto, se generaron los siguientes complejos proteína-proteína con mutaciones en AKR1B1: E193A, Q197A, S282A, T286A, L289A, S290A y una mutante del residuo T287A que ya se ha determinado experimentalmente que no afecta de manera significativa la formación del

complejo AKR1B1-DAD<sup>10</sup>. Los resultados de las SDM de los complejos mutantes indicaron que las mutaciones afectaban la estabilidad estructural del complejo proteína-proteína, así mismo, los cálculos de energía libre de unión mostraron que todas mutaciones desfavorecieron la energía libre de unión con respecto al WT. Las mutaciones que afectaron de manera drástica la formación del complejo fueron L289A y S290A, mientras que la mutación T287A fue la que menos afectó la formación del complejo. También, el análisis de las interacciones no covalentes mostró que hay una pérdida en la frecuencia de las interacciones hidrofóbicas y de enlace de hidrógeno, lo cual sugiere que la ausencia de las cadenas laterales de estos residuos afecta la estabilidad.

Aunque estos métodos computacionales solo son simulaciones y la interpretación de los resultados debe hacerse con cautela, también debe destacarse que los datos obtenidos son consistentes con observaciones experimentales previas como las que reportan Thiagarajan y colaboradores<sup>10</sup>. Sin embargo, también es necesario complementar los análisis de este proyecto con otras técnicas que permitan la predicción de IPP y de residuos críticos, tales como el enfoque basado en la red de estructura de proteínas (PSN) a nivel de contactos para identificar residuos críticos para la formación de complejo AKR1B1-DAD y los análisis de estabilidad y conservación de la interfaz proteína-proteína para identificar los residuos que son importantes para formar y estabilizar la interacción. Estas técnicas han mostrado ser muy eficientes para predecir residuos funcionales<sup>156,157</sup>.

## **CAPÍTULO VII. CONCLUSIONES**

Partiendo de la información reportada en la literatura y mediante el análisis estructural *in silico*, se predijo y caracterizó el sitio de unión de AKR1B1 con DAD, en el que existen dos  $\alpha$ -hélices encargadas de mediar la interacción. Estas abarcan los residuos 190-202 y 280-293. Asimismo, mediante simulación de dinámica molecular y cálculos de energía MM-PBSA se determinó que los residuos más importantes energéticamente para la interacción son E193, Q197, S282, T286, L289 y S290. Por otra parte, la mutación de estos residuos mediante mutagénesis de alanina *in silico* demostró que provocan una desestabilización en la formación del complejo estructural AKR1B1-DAD, lo cual se ve reflejado en los valores de energía libre de unión y el  $\Delta\Delta G$  con respecto al complejo WT. Del mismo modo, el análisis de las interacciones no covalentes destaca que las interacciones de la L289 y S290 son esenciales para formar el complejo estructural AKR1B1-DAD. Estos datos son consistentes con los resultados experimentales obtenidos por Thiagarajan y colaboradores<sup>10</sup>, donde destacan la relevancia de la L289 para mantener la IPP entre AKR1B1 y DAD.

A pesar de que este trabajo proporciona información valiosa para dilucidar la base estructural de las IPP de AKR1B1 que puede ser utilizada en el diseño eficaz de inhibidores de la IPP. Es necesario complementar la información obtenida en esta tesis con otras estrategias computacionales que son útiles para la predicción de IPP y de residuos críticos, así mismo, se encomienda el desarrollo de trabajos experimentales para validar los resultados obtenidos.

## **CAPÍTULO VIII. PERSPECTIVAS**

- Validar experimentalmente el efecto de las mutaciones de los residuos (determinados como importantes con los análisis *in silico*) de AKR1B1 sobre la formación del complejo AKR1B1-DAD.
- Diseño de inhibidores de la IPP de AKR1B1 mediante estrategias *in silico*.
- Sintetizar u obtener de manera recombinante los inhibidores candidatos y realizar su caracterización fisicoquímica.

- Resolver la estructura cristalográfica del complejo AKR1B1-inhibidor con la finalidad de corroborar de manera experimental el sitio de unión del inhibidor.
- Determinar la eficacia de los inhibidores seleccionados en diferentes líneas celulares de cáncer, mediante el análisis de características distintivas de un fenotipo agresivo como migración, proliferación celular y formación de colonias. Asimismo, determinar de manera experimental la interacción entre AKR1B1 y DAD en presencia de los inhibidores.

**CAPÍTULO IX. REFERENCIAS**

1. Barski, O. A. & Srinivas M. Tipparaju, and A. B. *The Aldo-Keto Reductase Superfamily and its Role in Drug.pdf. Drug Metab Rev* **40**, 553–624 (2009).
2. Tang, W. H., Martin, K. A. & Hwa, J. Aldose reductase, oxidative stress, and diabetic mellitus. *Frontiers in Pharmacology* **3**, 1–8 (2012).
3. Khayami, R., Hashemi, S. R. & Kerachian, M. A. Role of aldo-keto reductase family 1 member B1 (AKR1B1) in the cancer process and its therapeutic potential. *Journal of Cellular and Molecular Medicine* **24**, 8890–8902 (2020).
4. Laffin, B. & Petrash, J. M. Expression of the aldo-ketoreductases AKR1B1 and AKR1B10 in human cancers. *Frontiers in Pharmacology* **3**, 1–7 (2012).
5. Schwab, A. *et al.* Polyol pathway links glucose metabolism to the aggressiveness of cancer cells. *Cancer Research* **78**, 1604–1618 (2018).
6. Wu, X. *et al.* AKR1B1 promotes basal-like breast cancer progression by a positive feedback loop that activates the EMT program. *Journal of Experimental Medicine* **214**, 1065–1079 (2017).
7. Khayami, R., Hashemi, S. R. & Kerachian, M. A. Role of aldo-keto reductase family 1 member B1 (AKR1B1) in the cancer process and its therapeutic potential. *Journal of Cellular and Molecular Medicine* **24**, 8890–8902 (2020).
8. Zhao, J.-X. *et al.* Aldose reductase interacts with AKT1 to augment hepatic AKT/ mTOR signaling and promote hepatocarcinogenesis. *Oncotarget* **8**, (2017).
9. Xiao, M. B. *et al.*  $\beta$ 2-AR regulates the expression of AKR1B1 in human pancreatic cancer cells and promotes their proliferation via the ERK1/2 pathway. *Molecular Biology Reports* **45**, 1863–1871 (2018).
10. Thiagarajan, D. *et al.* Aldose Reductase Acts as a Selective Derepressor of PPAR $\gamma$  and the Retinoic Acid Receptor. *Cell Reports* **15**, 181–196 (2016).
11. Pitre, S., James, A. & Michel, R. G. Computational Methods For Predicting Protein–Protein Interactions. in *Advances in Biochemical Engineering/Biotechnology* 247–267 (2008).
12. Ding, Z. & Kihara, D. Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features. *Curr Protoc Protein Sci* **93**, e62 (2018).
13. Joe, J. & Janin, J. Assessing predictions of protein-protein interaction: The CAPRI experiment. *Protein Science* **14**, 278-283(2005).
14. Janin, J. Welcome to CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins: Structure, Function and Genetics* **47**, 257 (2002).

15. Lensink, M. F., Velankar, S. & Wodak, S. J. Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. *Proteins: Structure, Function and Bioinformatics* **85**, 359–377 (2017).
16. Stephen K. Burley, Helen M. Berman, Gerard J. Kleywegt, John L. Markley, Haruki Nakamura, S. Velankar. Protein Crystallography: Methods and Protocols. *Methods in Molecular Biology*, 627-641 (2017)
17. Gore, S. *et al.* Validation of Structures in the Protein Data Bank. *Structure* **25**, 1916–1927 (2017).
18. Li, X., Wang, X. & Snyder, M. Systematic investigation of protein-small molecule interactions. *IUBMB Life* **65**, 2–8 (2013).
19. H. Jeong, S. P. Mason, A.-L. Barabási & Z. N. Oltvai. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
20. Ahmed, H. *et al.* Network biology discovers pathogen contact points in host protein-protein interactomes. *Nature Communications* **9**, (2018).
21. Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G. & Orengo, C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure* **18**, 1233–1243 (2010).
22. Archakov, A. I. *et al.* Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **3**, 380–391 (2003).
23. Kar, G., Gursoy, A. & Keskin, O. Human cancer protein-protein interaction network: A structural perspective. *PLoS Computational Biology* **5**, (2009).
24. Zhou, M., Li, Q. & Wang, R. Current Experimental Methods for Characterizing Protein-Protein Interactions. *ChemMedChem* **11**, 738–756 (2016).
25. Berggård, T., Linse, S. & James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7**, 2833–2842 (2007).
26. Falconer, R. J., Penkova, A., Jelesarov, I. & Collins, B. M. Survey of the year 2008: Applications of isothermal titration calorimetry. *Journal of Molecular Recognition* **23**, 395–413 (2010).
27. Pattnaik, P. Surface plasmon resonance: Applications in understanding receptor-ligand interaction. *Applied Biochemistry and Biotechnology* **126**, 79–92 (2005).
28. Huber, W. & Mueller, F. Biomolecular Interaction Analysis in Drug Discovery Using Surface Plasmon Resonance Technology. *Current Pharmaceutical Design* **12**, (2006).

29. Duan, X. J., Xenarios, I. & Eisenberg, D. Describing biological protein interactions in terms of protein states and state transitions: the LiveDIP database. *Mol Cell Proteomics* **1**, 104–116 (2002).
30. Alfarano, C. *et al.* The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Research* **33**, 418–424 (2005).
31. von Mering, C. *et al.* STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research* **33**, 433–437 (2005).
32. Zanzoni, A. *et al.* MINT: A Molecular INTeraction database. *FEBS Letters* **513**, 135–140 (2002).
33. Hermjakob, H. *et al.* IntAct: An open-source molecular interaction database. *Nucleic Acids Research* **32**, 452–455 (2004).
34. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, 535–539 (2006).
35. Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Research* **13**, 2363–2371 (2003).
36. Shoemaker, B. A. & Panchenko, A. R. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Computational Biology* **3**, 0337–0344 (2007).
37. Skrabanek, L., Saini, H. K., Bader, G. D. & Enright, A. J. Computational prediction of protein-protein interactions. *Molecular Biotechnology* **38**, 1–17 (2008).
38. Loregian, A., Marsden, H. S. & Palù, G. Protein-protein interactions as targets for antiviral chemotherapy. *Reviews in Medical Virology* **12**, 239–262 (2002).
39. Kovács, I. A. *et al.* Network-based prediction of protein interactions. *Nature Communications* **10**, 1–8 (2019).
40. Chang, J. W., Zhou, Y. Q., Ul Qamar, M. T., Chen, L. L. & Ding, Y. D. Prediction of protein–protein interactions by evidence combining methods. *International Journal of Molecular Sciences* **17**, (2016).
41. Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins: Structure, Function and Bioinformatics* **86**, 7–15 (2018).



42. Kryshchak, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function and Bioinformatics* **87**, 1011–1020 (2019).
43. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
44. Schleinkofer, K., Wang, T. & Wade, R. C. Molecular Docking. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* **443**, 1149–1153 (2006).
45. Dias, Raquel and Azevedo Jr, W. F. Molecular Docking Algorithms. *Current Drug Targets* **9**, 1040–1047 (2008).
46. Lee, J. & Perkins, D. A simulated annealing algorithm with a dual perturbation method for clustering. *Pattern Recognition* **112**, 107713 (2021).
47. Chen, G., Low, C. P. & Yang, Z. Preserving and exploiting genetic diversity in evolutionary programming algorithms. *IEEE Transactions on Evolutionary Computation* **13**, 661–673 (2009).
48. Lensink, M. F., Nadzirin, N., Velankar, S. & Wodak, S. J. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins: Structure, Function and Bioinformatics* **88**, 916–938 (2020).
49. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nature Protocols* **12**, 255–278 (2017).
50. Yan, Y., Tao, H., He, J. & Huang, S. Y. The HDock server for integrated protein-protein docking. *Nature Protocols* **15**, 1829–1852 (2020).
51. Smith, G. R. & Sternberg, M. J. E. Prediction of protein-protein interactions by docking methods. *Current Opinion in Structural Biology* **12**, 28–35 (2002).
52. Katchalski-Katzir, E. *et al.* Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* **89**, 2195–2199 (1992).
53. Padhorny, D. *et al.* Protein-protein docking by fast generalized Fourier transforms on 5D rotational manifolds. *Proc Natl Acad Sci USA* **113**, E4286–E4293 (2016).
54. Woetzel, N., Lindert, S., Stewart, P. L. & Meiler, J. BCL::EM-Fit: Rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. *Journal of Structural Biology* **175**, 264–276 (2011).
55. Maurer, W. D., Lewis, T. G. & Maurer, W. D. Hash table methods. *Computing Surveys* **7**, (1975).

56. Venkatraman, V., Yang, Y. D., Sael, L. & Kihara, D. Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics* **10**, (2009).
57. Su, M. *et al.* Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling* **59**, 895–913 (2019).
58. Bentham Science Publisher, B. S. P. Scoring Functions for Protein-Ligand Docking. *Current Protein & Peptide Science* **7**, 407–420 (2006).
59. Li, J., Fu, A. & Zhang, L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdisciplinary Sciences: Computational Life Sciences* **11**, 320–328 (2019).
60. Yadava, U. Search algorithms and scoring methods in protein-ligand docking. *Endocrinology & Metabolism International Journal* **6**, (2018).
61. Ma, X. H., Wang, C. X., Li, C. H. & Chen, W. Z. A fast empirical approach to binding free energy calculations based on protein interface information. *Protein Engineering* **15**, 677–681 (2002).
62. Yin, S., Biedermannova, L., Vondrasek, J. & Dokholyan, N. v. MedusaScore: An accurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Modeling* **48**, 1656–1662 (2008).
63. Liu, J. & Wang, R. Classification of current scoring functions. *Journal of Chemical Information and Modeling* **55**, 475–482 (2015).
64. Huang, S. Y. & Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins: Structure, Function and Genetics* **72**, 557–579 (2008).
65. Lensink, M. F. *et al.* Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins: Structure, Function and Bioinformatics* **84**, 323–348 (2016).
66. Mashiach, E. *et al.* An integrated suite of fast docking algorithms. *Proteins: Structure, Function and Bioinformatics* **78**, 3197–3204 (2010).
67. Hwang, H., Vreven, T., Pierce, B. G., Hung, J. H. & Weng, Z. Performance of ZDOCK and ZRANK in CAPRI rounds 13-19. *Proteins: Structure, Function and Bioinformatics* **78**, 3104–3110 (2010).
68. Lensink, M. F. & Wodak, S. J. Docking, scoring, and affinity prediction in CAPRI. *Proteins: Structure, Function and Bioinformatics* **81**, 2082–2095 (2013).

69. Duhovny, D., Nussinov, R. & Wolfson, H. J. Efficient unbound docking of rigid molecules. *Lecture Notes in Computer Science* **2452**, 185–200 (2002).
70. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research* **33**, 363–367 (2005).
71. Pierce, B. G., Hourai, Y. & Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* **6**, 0–5 (2011).
72. Mintseris, J. *et al.* Integrating statistical pair potentials into protein complex prediction. *Proteins: Structure, Function and Bioinformatics* **69**, 511–520 (2007).
73. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731–1737 (2003).
74. Van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **428**, 720–725 (2016).
75. Salsbury, F. R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Current Opinion in Pharmacology* **10**, 738–744 (2010).
76. Hansson, T., Oostenbrink, C. & Van Gunsteren, W. F. Molecular dynamics simulations. *Current Opinion in Structural Biology* **12**, 190–196 (2002).
77. Lazim, R., Suh, D. & Choi, S. Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *International Journal of Molecular Sciences* **21**, 1–20 (2020).
78. Rakers, C., Bermudez, M., Keller, B. G., Mortier, J. & Wolber, G. Computational close up on protein-protein interactions: How to unravel the invisible using molecular dynamics simulations? *Wiley Interdisciplinary Reviews: Computational Molecular Science* **5**, 345–359 (2015).
79. Tan, C. W., Sam, I. C., Chong, W. L., Lee, V. S. & Chan, Y. F. Polysulfonate suramin inhibits Zika virus infection. *Antiviral Research* **143**, 186–194 (2017).
80. Somboon, T. *et al.* Computational study on peptidomimetic inhibitors against SARS-CoV-2 main protease. *Journal of Molecular Liquids* **322**, (2021).
81. Lees, W. D., Stejskal, L., Moss, D. S. & Shepherd, A. J. Investigating substitutions in antibody-antigen complexes using molecular dynamics: A case study with broad-spectrum, influenza A antibodies. *Frontiers in Immunology* **8**, (2017).

82. Isa, D. M. *et al.* Dynamics and binding interactions of peptide inhibitors of dengue virus entry. *Journal of Biological Physics* **45**, 63–76 (2019).
83. Ibarra, A. A. *et al.* Predicting and Experimentally Validating Hot-Spot Residues at Protein-Protein Interfaces. *ACS Chemical Biology* **14**, 2252–2263 (2019).
84. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129–1143 (2018).
85. González, M. A. Force fields and molecular dynamics simulations. *Collection SFN* **12**, 169–200 (2011).
86. Monticelli, L. & Tieleman, D. P. Force fields for classical molecular dynamics. *Methods in Molecular Biology* **924**, 197–213 (2013).
87. Clark, M., Cramer, R. D. & Van Opdenbosch, N. Validation of the general purpose tripos 5.2 force field. *Journal of Computational Chemistry* **10**, 982–1012 (1989).
88. Allinger, N. L. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J Am Chem Soc* **99**, 8127–8134 (1977).
89. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a General Amber Force Field. *Journal of Computational Chemistry*, 1157–1174 (2004).
90. Vanommeslaeghe, K. *et al.* CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671–690 (2010).
91. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *American Chemical Society*, 31–41 (1996).
92. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586–3616 (1998).
93. Cornell, W. D. *et al.* A Second-Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **117**, 5179–5197 (1995).
94. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 1656–1676 (2004).

95. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* **78**, 1950–1958 (2010).
96. Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71–73 (2016).
97. Soares, T. A., Daura, X., Oostenbrink, C., Smith, L. J. & van Gunsteren, W. F. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *Journal of Biomolecular NMR* **30**, 407–422 (2004).
98. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545–1614 (2009).
99. Frenkel, D. & Smit, B. Understanding Molecular Simulation. *Computational sciences series* (2002). doi:10.1016/b978-0-12-267351-1.x5000-7.
100. Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98–103 (1967).
101. Gutiérrez, M. G. Simulación molecular del efecto iónico en polímeros de quitosano. (2018).
102. Braun, E. *et al.* Best Practices for Foundations in Molecular Simulations. *Living J Comput Mol Sci*, 1–53 (2019).
103. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
104. Gromacs. <http://www.gromacs.org>.
105. NAMD. <http://www.ks.uiuc.edu/Research/namd>.
106. Case, D. A. *et al.* Amber 2021. University of California, SF. (2021).
107. AMBER. <http://ambermd.org>.
108. CHARMM. <http://www.charmm.org>.
109. William Humphrey, Andrew Dalke, and K. Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
110. R core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Viena, Austria. URL <https://www.R-project.org/>, (2020).

111. Gowers, R. *et al.* MDAAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference*, 98–105 (2016).
112. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319–2327 (2011).
113. Kumari, R., Kumar, R. & Lynn, A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling* **54**, 1951–1962 (2014).
114. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophysical Journal* **98**, 861–871 (2010).
115. Pitera, J. W. Expected distributions of root-mean-square positional deviations in proteins. *Journal of Physical Chemistry B* **118**, 6526–6530 (2014).
116. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 623–628 (2008).
117. Huang, X. & Powers, R. Validity of using the radius of gyration as a restraint in NMR protein structure determination. *J Am Chem Soc* **123**, 3834–3835 (2001).
118. Woo, H. J. & Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* **102**, 6825–6830 (2005).
119. Ruiz, F. X. *et al.* Aldo-keto reductases from the AKR1B subfamily: Retinoid specificity and control of cellular retinoic acid levels. *Chemico-Biological Interactions* **178**, 171–177 (2009).
120. Ruiz, F. X. *et al.* Human and rodent aldo-keto reductases from the AKR1B subfamily and their specificity with retinaldehyde. *Chemico-Biological Interactions* **191**, 199–205 (2011).
121. Rižner, T. L. Enzymes of the AKR1B and AKR1C subfamilies and uterine diseases. *Frontiers in Pharmacology* **3**, 1–13 (2012).
122. Dréanic, M. P., Edge, C. M. & Tuttle, T. New Insights into the Catalytic Mechanism of Aldose Reductase: A QM/MM Study. *ACS Omega* **2**, 5737–5747 (2017).
123. Petrash, J. M. All in the family: Aldose reductase and closely related aldo-keto reductases. *Cellular and Molecular Life Sciences* **61**, 737–749 (2004).

124. Endo, S. *et al.* Kinetic studies of AKR1B10, human aldose reductase-like protein: Endogenous substrates and inhibition by steroids. *Archives of Biochemistry and Biophysics* **487**, 1–9 (2009).
125. Hexokinase. 179–198 doi:10.1007/978-0-387-72891-9\_10.
126. Khayami, R., Hashemi, S. R. & Kerachian, M. A. Role of aldo-keto reductase family 1 member B1 (AKR1B1) in the cancer process and its therapeutic potential. *Journal of Cellular and Molecular Medicine* **24**, 8890–8902 (2020).
127. Sánchez-Gómez, F. J. *et al.* Detoxifying enzymes at the cross-roads of inflammation, oxidative stress, and drug hypersensitivity: Role of glutathione transferase P1-1 and aldose reductase. *Frontiers in Pharmacology* **7**, (2016).
128. Alexiou, P., Pegklidou, K., Chatzopoulou, M., Nicolaou, I. & Demopoulos, V. Aldose Reductase Enzyme and its Implication to Major Health Problems of the 21st Century. *Current Medicinal Chemistry* **16**, 734–752 (2009).
129. Saraswat, M. *et al.* Overexpression of aldose reductase in human cancer tissues. *Medical Science Monitor* **12**, 525–529 (2006).
130. Ramirez, M. A. & Borja, N. L. Epalrestat: An Aldose Reductase Inhibitor for the Treatment of Diabetic Neuropathy. *pharmacotherapy* **28**, 646–655 (2008).
131. Schemmel, K. E., Padiyara, R. S. & D'Souza, J. J. Aldose reductase inhibitors in the treatment of diabetic peripheral neuropathy: A review. *Journal of Diabetes and its Complications* **24**, 354–360 (2010).
132. Sobajima. Erratum: Pharmacological properties of fidarestat, a potent aldose reductase inhibitor, clarified by using sorbitol in human and rat erythrocytes (Pharmacology (2001) 62 (193-199)). *Pharmacology* **63**, 197 (2001).
133. Mylari, B. L. *et al.* Novel, Potent Aldose Reductase Inhibitors: 3,4-Dihydro-4-oxo-3-[[5-(trifluoromethyl)-2-benzothiazolyl]methyl]-1-phthalazineacetic Acid (Zopolrestat) and Congeners. *Journal of Medicinal Chemistry* **34**, 108–122 (1991).
134. Shehzad, M. T. *et al.* Benzoxazinone-thiosemicarbazones as antidiabetic leads via aldose reductase inhibition: Synthesis, biological screening and molecular docking study. *Bioorganic Chemistry* **87**, 857–866 (2019).
135. Wilson, D. K., Tarlet, I., Petrasht, J. M. & Quioco, F. A. Refined 1.8 Å structure of human aldose reductase complexed with the potent inhibitor zopolrestat (active site/diabetic complications/drug design). *Proc. Natl. Acad. Sci. USA* **90** (1993).
136. Srivastava, S. K. *et al.* Aldose reductase inhibition suppresses oxidative stress-induced inflammatory disorders. *Chemico-Biological Interactions* **191**, 330–338 (2011).

137. Tammali, R., K. Srivastava, S. & V. Ramana, K. Targeting Aldose Reductase for the Treatment of Cancer. *Current Cancer Drug Targets* **11**, 560–571 (2012).
138. Equipo de supercómputo “Miztli.” *Laboratorio Nacional de Cómputo de Alto Desempeño* <http://www.lanCAD.mx/?p=59>.
139. Jo, S., Kim, T., Iyer, V. G. & IM, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J Comput Chem* **32**, 174–182 (2012).
140. Basu, S. & Wallner, B. DockQ: A quality measure for protein-protein docking models. *PLoS ONE* **11**, (2016).
141. M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the G. development team. GROMACS User Manual version 5.0.4. (2014).
142. Dong, R., Peng, Z., Zhang, Y. & Yang, J. mTM-align: An algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* **34**, 1719–1725 (2018).
143. Origin(pro). Version 8.0. OriginLab Corporation, Northampton, MA, USA.
144. Xue, L. C., Rodrigues, J. P., Kastiris, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
145. Schrödinger, L. The PyMOL Molecular Graphics System.
146. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004).
147. Dong, G. Q. *et al.* Optimized atomic statistical potentials: Assessment of protein interfaces and loops. *Bioinformatics* **29**, 3158–3166 (2013).
148. Janin, J. Assessing predictions of protein-protein interaction: The CAPRI experiment. *Protein Science* **14**, 278–283 (2005).
149. Bystroff, C., & Garde, S. Helix propensities of short peptides: Molecular dynamics versus bioinformatics. *Proteins: Structure, Function, and Bioinformatics*, **50**, 552–562 (2003)
150. Gu, Y., Li, D. W. & Brüschweiler, R. NMR order parameter determination from long molecular dynamics trajectories for objective comparison with experiment. *Journal of Chemical Theory and Computation* **10**, 2599–2607 (2014).
151. Hoffmann, F., Xue, M., Schäfer, L. V. & Mulder, F. A. A. Narrowing the gap between experimental and computational determination of methyl group



- dynamics in proteins. *Physical Chemistry Chemical Physics* **20**, 24577–24590 (2018).
152. Lefèvre, F., Rémy, M. H. & Masson, J. M. Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function. *Nucleic Acids Research* **25**, 447–448 (1997).
  153. Parenti, M. D. & Rastelli, G. Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnology Advances* **30**, 244–250 (2012).
  154. Ferrari, A. M., Degliesposti, G., Sgobba, M. & Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorganic and Medicinal Chemistry* **15**, 7865–7877 (2007).
  155. Bromberg, Y. & Rost, B. Comprehensive in silico mutagenesis highlights functionally important residues in proteins. in *Bioinformatics* **24**, (2008).
  156. Prabantu, V. M., Naveenkumar, N. & Srinivasan, N. Influence of Disease-Causing Mutations on Protein Structural Networks. *Frontiers in Molecular Biosciences* **7**, (2021).
  157. Høie, M. H., Cagiada, M., Beck Frederiksen, A. H., Stein, A. & Lindorff-Larsen, K. Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. *Cell Reports* **38**, (2022).