



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Simulación de la estructura del
cotransportador de Na^+/Cl^- sensible
a tiazidas (NCC) utilizando técnicas
de Inteligencia Artificial**

TESIS

Que para obtener el título de
Ingeniero en Sistemas Biomédicos

P R E S E N T A

Julio Rafael Sandria Sánchez

DIRECTOR DE TESIS

Dr. Luis Jiménez Ángeles



Ciudad Universitaria, Cd. Mx., 2026



**PROTESTA UNIVERSITARIA DE INTEGRIDAD Y
HONESTIDAD ACADÉMICA Y PROFESIONAL
(Titulación con trabajo escrito)**



De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado SIMULACION DE LA ESTRUCTURA DEL COTRANSPORTADOR DE NA⁺/CL⁻ SENSIBLE A TIAZIDAS (NCC) UTILIZANDO TECNICAS DE INTELIGENCIA ARTIFICIAL que presenté para obtener el título de INGENIERO EN SISTEMAS BIOMÉDICOS es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Entidad Académica, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de titulación.

JULIO RAFAEL SANDRIA SANCHEZ
Número de cuenta: 421069811

Simulación de la estructura del cotransportador de Na^+/Cl^- sensible a tiazidas (NCC) utilizando técnicas de Inteligencia Artificial

por

Julio Rafael Sandria Sánchez

Resumen

El cotransportador de Na^+/Cl^- sensible a tiazidas (NCC) es una proteína de membrana clave en la homeostasis del sodio y la regulación de la presión arterial. Su relevancia farmacológica ha motivado la búsqueda de modelos estructurales confiables, a pesar de la limitada disponibilidad de estructuras experimentales. Esta tesis presenta un enfoque computacional comparativo para modelar el NCC en sus formas monomérica y dimérica, en sus variantes humana y de anguila europea, utilizando cuatro plataformas de predicción estructural: AlphaFold, Rosetta, SWISS-MODEL y Modeller.

Se generaron 280 modelos (120 monómeros y 160 dímeros) distribuidos equitativamente entre las cuatro plataformas de predicción, todos estos modelos fueron sometidos a un proceso de optimización estructural (corrección y reacomodación), denominado refinación, por medio de la plataforma RosettaRelax. La evaluación estructural con MolProbity permitió seleccionar 14 modelos de alta calidad estereoquímica. Posteriormente, se caracterizaron cavidades internas mediante CavityPlus, midiendo parámetros de drogabilidad como DrugScore, volumen y área superficial accesible al solvente; estos parámetros nos permiten evaluar los posibles sitios de unión a ligandos (posibles fármacos). Se identificaron tres cavidades principales con propiedades consistentes entre los distintos softwares, observando así que todos los métodos capturaron de manera confiable la cavidad central de unión del NCC, correspondiente al sitio conocido de unión a tiazidas.

Los resultados muestran que AlphaFold y Rosetta generan monómeros con cavidades bien definidas, mientras que Modeller y SWISS-MODEL producen dímeros con cavidades más amplias y recurrentes. El análisis comparativo reveló un solapamiento volumétrico superior al 85 % entre herramientas para las cavidades centrales, validando la coherencia del enfoque. Este traba-

jo propone una estrategia robusta de predicción y análisis estructural que contribuye al estudio de proteínas de membrana de difícil cristalización y apoya al desarrollo de teorías para el diseño racional de nuevos inhibidores del NCC.

Simulation of the Structure of the Thiazide-Sensitive Na^+/Cl^- Cotransporter (NCC) Using Artificial Intelligence Techniques

by

Julio Rafael Sandria Sánchez

Abstract

The thiazide-sensitive Na^+/Cl^- cotransporter (NCC) is a membrane protein critical for sodium homeostasis and blood pressure regulation. Its pharmacological relevance has driven efforts to obtain reliable structural models, despite the limited availability of high-resolution experimental structures. This thesis presents a comparative computational strategy to model the NCC in monomeric and dimeric forms, for both the human and European eel variants, using four structure prediction platforms: AlphaFold, Rosetta, SWISS-MODEL, and Modeller.

A total of 280 models (120 monomers and 160 dimers), evenly distributed among the four prediction platforms, were generated. All models were subjected to a structural optimization process (correction and rearrangement), refinement, using the RosettaRelax platform. Structural evaluation with MolProbity enabled the selection of 14 models with high stereochemical quality. Internal cavities were characterized using CavityPlus by measuring druggability-related parameters such as DrugScore, volume, and solvent-accessible surface area; these parameters allow the assessment of ligand (potential drug) binding sites. Three main cavities with consistent properties across the different software tools were identified, demonstrating that all methods reliably captured the central binding cavity of NCC, corresponding to the known thiazide-binding site.

The results show that AlphaFold and Rosetta generate monomers with well-defined cavities, whereas Modeller and SWISS-MODEL produce dimers with larger and more recurrent cavities. Comparative analysis revealed a volumetric overlap greater than 85 % among tools for the central cavities, validating the consistency of the approach. This work proposes a robust strategy for structural prediction and analysis that contributes to the study of membrane proteins that are difficult to crystallize and supports the proposal of new theories for the rational design of new NCC inhibitors.

Agradecimientos

Agradezco a mis padres, Ana y Julio, así como a mi hermana, Carolina. A mis amigos y a cada miembro de mi familia que me apoyó.

Reconozco y agradezco el entrenamiento y acceso a recursos de cómputo de alto desempeño brindado por la DGTIC-UNAM (Alianza UNAM-HUAWEI, Proyecto ID-05) y al Laboratorio Nacional de Cómputo de Alto Desempeño, UAM-Iztapalapa (LANCAD, Proyecto ID 15-2025). De igual manera, a los profesores que me orientaron y enseñaron durante este proyecto: la Dra. Reyna Erika Moreno Martínez, la Dra. Erendira Rojas Ortega y a mí asesor, el Dr. Luis Jiménez Ángeles.

Dedicatoria

Este trabajo está dedicado a mis padres, Ana y Julio, así como a mi hermana, Carolina. Su apoyo, amor y motivación son lo que ha permitido lograr todas las metas que alguna vez me he propuesto. Las palabras no me alcanzan para expresar mi gratitud y amor, así que espero que una vida de acciones lo logre.

A los miembros de mi familia que siempre han creído en mí y me han apoyado.

A mis amigos, que han estado presentes en todo momento, me han motivado e impulsado.

Sinceramente

Rafael Sandria

Índice general

Introducción	2
Contexto	4
Motivaciones	6
Objetivos	7
Objetivo general	7
Objetivos específicos	7
1. Antecedentes	9
1.1. Proteínas de Membrana	9
1.1.1. Propiedades Estructurales de las Proteínas de Membrana	10
1.1.2. Métodos experimentales para el estudio de estructuras de proteínas de membrana	11
1.2. Cotransportador de Na^+/Cl^- (NCC)	12
1.2.1. Función en el Balance de Sodio y Cloro	13
1.2.2. Papel en la Regulación de la Presión Arterial	13
1.2.3. Implicaciones Clínicas	14
1.3. Diuréticos Tiazídicos y Mecanismo de Acción	15
1.3.1. Mecanismo de Acción	15
1.3.2. Eficacia en el Tratamiento de la Hipertensión	15
1.3.3. Limitaciones en el Diseño y Uso	16
1.4. Técnicas <i>In Silico</i> para el modelado de proteínas	17

1.4.1. Métodos de predicción estructural	17
1.4.2. Herramientas principales	18
1.4.3. Estado del arte en modelado de proteínas mediado por IA	22
2. Definición del Problema	24
2.1. Problemática	24
2.2. Hipótesis	25
3. Metodología	26
3.1. Planteamiento	26
3.2. Enfoque General	27
3.2.1. Descripción del Pipeline	27
3.3. Selección de Datos y Herramientas	29
3.3.1. Bases de datos estructurales	29
3.3.2. Herramientas de Modelado	32
3.4. Proceso de Modelado y Simulación	33
3.4.1. Generación de modelos por AlphaFold	33
3.4.2. Generación de modelos por Rosetta	34
3.4.3. Generación de modelos por SwissModel	35
3.4.4. Generación de modelos por Modeller	37
3.4.5. Cantidad de modelos base generados	38
3.5. Análisis estructural de los modelos y relajación	39
3.5.1. Análisis estructural post relajación con Molprobit	39
3.5.2. Comparación de resultados de cada modelo y proteína	40
3.6. Análisis de cavidades	41
4. Resultados	43
4.1. Impacto cuantitativo del refinamiento iterativo	44
4.2. Modelo estructural del NCC	47
4.2.1. Beta NCC Anguila	47

4.2.2. NCC Humano	50
4.3. Comparación de estructuras	53
4.3.1. β NCC de Anguila — Monómeros	54
4.3.2. NCC Humano — Monómeros	55
4.3.3. β NCC de Anguila — Dímeros	56
4.3.4. NCC Humano — Dímeros	57
4.4. Resultados de Cavidades	59
5. Conclusiones	64
5.1. Limitaciones y Trabajo Futuro	65
Bibliografía	67

Índice de figuras

3.1. Diagrama de flujo para la metodología general propuesta para el modelado In Silico.	30
3.2. Secuencia de aminoácidos de β NCC de Anguilla anguilla, entrada UniProt Q2PDH2; 1043 residuos. Fuente de los datos: UniProt [49]	32
3.3. Secuencia de aminoácidos del NCC humano, entrada UniProt P55017; 1039 residuos. Fuente de los datos: UniProt [49]	32
4.1. Comparación de métricas Ramachandran promedio (Favored y Outliers) para modelos monoméricos conforme a los ciclos de relajación.	46
4.2. Comparación de métricas Ramachandran promedio (Favored y Outliers) para modelos diméricos conforme a los ciclos de relajación.	46
4.3. Monómero de AlphaFold para β NCC de anguila europea posterior a la tercera ronda de relajación.	49
4.4. Dímero de SWISS-MODEL para β NCC de anguila europea, segunda ronda de relajación del modelo.	49
4.5. Monómero de AlphaFold para NCC Humano posterior a la segunda ronda de relajación.	51
4.6. Dímero de AlphaFold para NCC Humano posterior a la segunda ronda de relajación.	52
4.7. MatchMaker de monómeros para β NCC de anguila. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (verde), Rosetta (azul).	54

4.8. MatchMaker de monómeros de NCC humano. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (verde), Rosetta (azul).	55
4.9. MatchMaker de dímeros para β NCC de anguila. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (azul).	56
4.10. MatchMaker de dímeros para NCC humano. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (azul).	57
4.11. Superposición de los modelos diméricos e identificación de cavidades. A) Beta NCC de Anguila y B) NCC de Humano. Se comparan los modelos de AlphaFold (azul) y SWISS-MODEL (rosado). Elaboración propia, publicada previamente en [53, Fig. 4].	61
4.12. Superposición de los modelos monoméricos e identificación de las cavidades con mayor valor de druggability. A) Beta NCC de Anguila y B) NCC de humano. AlphaFold (azul), Modeller (lila), RoseTTaFold (naranja) y SWISS-MODEL (verde). Elaboración propia, publicada previamente en [53, Fig. 3].	61

Índice de tablas

3.1. Principales parámetros de los templates para NCC humano.	37
3.2. Principales parámetros de los templates para β NCC.	37
3.3. Resumen de modelos generados por método, especie y estado oligomérico. . .	38
4.1. Mejora promedio de parámetros estereoquímicos tras el tercer ciclo de relajación (Relax3 vs modelo inicial). Valores positivos en Δ Favored y Δ Outliers indican mejora geométrica. AF: AlphaFold, MD: Modeller, SM: SwissModel, RF: RosettaFold.	45
4.2. Parámetros de interés de los modelos de monómero seleccionados para β NCC de anguila.	48
4.3. Parámetros de interés de los modelos de dímero seleccionados para β NCC de anguila.	48
4.4. Parámetros de interés de los modelos de monómero seleccionados para NCC humano.	50
4.5. Parámetros de interés de los modelos de dímero seleccionados para NCC humano.	51
4.6. Comparación estructural entre modelos monoméricos de β NCC de anguila utilizando MatchMaker de ChimeraX.	54
4.7. Comparación estructural entre modelos monoméricos del NCC humano utilizando MatchMaker de ChimeraX.	55
4.8. Comparación estructural entre modelos diméricos de β NCC de anguila utilizando MatchMaker de ChimeraX.	56

4.9. Comparación estructural entre modelos diméricos del NCC humano utilizando MatchMaker de ChimeraX.	57
4.10. Resumen de comparación estructural por RMSD entre modelos seleccionados y de referencia.	59
4.11. Resumen comparativo de DrugScore por software, tipo de modelo y especie. . .	59

Introducción

La predicción estructural de proteínas ha cobrado una relevancia creciente en la biología computacional moderna, particularmente en el estudio de proteínas de membrana. Estas representan el principal blanco terapéutico de fármacos para el tratamiento de múltiples enfermedades; sin embargo, su análisis experimental mediante métodos como la cristalografía de rayos X o la microscopia crioelectrónica puede resultar técnicamente desafiante y costoso, desembocando en poco conocimiento de estas estructuras [1,2]. En este contexto, las herramientas de modelado In Silico y predicción asistida por inteligencia artificial han permitido extender el conocimiento estructural hacia proteínas carentes de estructuras resueltas experimentalmente, han acelerado la generación y confirmación de hipótesis, ayudado a resolver propuestas de modelado y, sobre todo, han resaltado la importancia y poder de la complementación de métodos experimentales con métodos In Silico en proteínas de alto interés clínico, como es el caso del cotransportador sodio-cloro sensible a tiazidas (NCC) [3,4].

El NCC humano es una proteína de membrana integral perteneciente a la familia de cotransportadores electroneutros del sodio, cuyo papel es fundamental en la reabsorción de sodio a nivel del túbulo distal del nefrón, siendo esta una ubicación clave como sitio de acción de los diuréticos tiazídicos [4,5]. La comprensión de su estructura tridimensional es clave para el desarrollo de inhibidores específicos, con aplicaciones terapéuticas en hipertensión arterial y síndromes de pérdida salina [6,7]. Por otro lado, el estudio del β NCC en anguila europea (*Anguilla anguilla*) ofrece un modelo comparativo de interés evolutivo y funcional, al presentar variantes estructurales conservadas y divergentes respecto a su contraparte humana [8].

A pesar de los avances recientes en la resolución estructural del NCC mediante métodos ex-

perimentales (como la criomicroscopía electrónica) [4], persisten limitaciones en la disponibilidad de modelos completos bajo diferentes estados conformacionales y condiciones comparativas entre especies. Asimismo, aunque herramientas como AlphaFold han demostrado una precisión sobresaliente en múltiples casos [9], se ha señalado que la validación cruzada entre metodologías independientes continúa siendo necesaria, particularmente en proteínas de membrana [3].

Partiendo de esta situación, se emplearon distintas plataformas de modelado estructural, incluyendo herramientas de predicción por homología (SWISS-MODEL, Modeller) [10, 11], métodos híbridos con refinamiento energético (Rosetta) [12] y predicción basada en inteligencia artificial (AlphaFold) [13]. Cada enfoque permite generar modelos tridimensionales con distintos niveles de precisión, los cuales fueron posteriormente evaluados en MolProbity [14] mediante validaciones estereoquímicas (diagramas de Ramachandran, Z-Score), parámetros de calidad geométrica y análisis de cavidades potencialmente drogables. El empleo de múltiples plataformas de predicción estructural permite mitigar los sesgos inherentes a cada algoritmo. El análisis de distintos enfoques permite explorar regiones complementarias del espacio conformacional; además, la convergencia estructural entre modelos generados por metodologías distintas incrementa la confiabilidad y robustez del análisis comparativo [3, 9]

En este contexto, el presente trabajo propone una estrategia comparativa que combina múltiples enfoques de predicción estructural, refinamiento energético y validación estereoquímica para evaluar la consistencia estructural del NCC en sus variantes humana y de anguila europea. Este enfoque permite identificar modelos de alta calidad y analizar la reproducibilidad de cavidades potencialmente drogables entre metodologías independientes, aportando un marco sistemático aplicable al estudio de proteínas de membrana con información experimental limitada.

Este capítulo tiene como objetivo introducir el contexto biológico y estructural de la proteína de interés, justificar la necesidad del modelado computacional en su estudio y establecer las herramientas metodológicas que permitirán la evaluación comparativa de los modelos generados. En los capítulos siguientes se detallan los procedimientos de modelado, los criterios de selección y validación estructural, así como los hallazgos derivados del análisis computacional.

Contexto

El cotransportador de sodio y cloro (NCC, por sus siglas en inglés), codificado por el gen SLC12A3, pertenece a la familia de cotransportadores electroneutros SLC12 y se expresa predominantemente en la membrana apical de las células epiteliales del túbulo contorneado distal del riñón. Su función principal es mediar la reabsorción acoplada de sodio y cloro desde el lumen tubular hacia el interior celular mediante un mecanismo de transporte electroneutro. Este proceso es un paso determinante en el ajuste fino del contenido corporal de sodio; en consecuencia, también regula el volumen extracelular, el equilibrio hidroelectrolítico y el control de la presión arterial sistémica [15].

Desde el punto de vista fisiológico, el NCC es un modulador clave del manejo renal de electrolitos; las variaciones en su actividad pueden producir cambios significativos en la homeostasis del sodio y del potasio. De manera que alteraciones en la función o regulación del NCC se asocian con desequilibrios iónicos que pueden manifestarse como hipertensión arterial, hipokalemia o alcalosis metabólica [15], resaltando su relevancia fisiológica y, por lo tanto, clínica.

En México, la hipertensión arterial es un problema de salud pública significativo, afectando a más del 30% de la población adulta [16]. Este trastorno representa uno de los principales factores de riesgo para enfermedades cardiovasculares, que son la primera causa de mortalidad en el país [17]. Desde el punto de vista fisiopatológico, el incremento sostenido de la presión arterial está estrechamente relacionado con alteraciones en los mecanismos renales de manejo de sodio, entre los cuales, el NCC tiene un papel determinante. En este contexto, aparecen los diuréticos tiazídicos, cuyo efecto terapéutico como inhibidor del NCC en el túbulo contorneado distal de la nefrona lo ubica como uno de los fármacos más empleados para el tratamiento de la hipertensión arterial [5, 6]. Su mecanismo de acción se basa en la reducción de la reabsorción de sodio y cloro, implicando una reducción en la presión arterial sistémica [5]. Este tipo de fármaco se caracteriza por su eficacia clínica, bajo costo y amplio historial de seguridad, de manera que se mantienen con una alta tasa de prescripción. Sin embargo, su uso también ha sido relacionado con efectos adversos como hipokalemia o hiponatremia [6, 15], de manera que se vuelve fundamental comprender a mayor detalle la interacción estructural entre el NCC y sus

inhibidores.

A pesar de su efectividad clínica, el diseño y optimización de diuréticos tiazídicos enfrenta desafíos relacionados con la especificidad hacia el NCC, la variabilidad en la respuesta terapéutica y la aparición de efectos secundarios como hipokalemia o hiponatremia [5, 6]. Estas limitaciones podrían relacionarse con un conocimiento incompleto de los determinantes estructurales que gobiernan la interacción entre el NCC y sus inhibidores.

La caracterización estructural del NCC ha representado históricamente un reto debido a las dificultades experimentales asociadas al estudio de proteínas transmembrana. Si bien avances recientes en técnicas como la criomicroscopía electrónica han permitido obtener información estructural de cotransportadores relacionados [4], aún persisten limitaciones en la disponibilidad de estructuras completas en distintos estados conformacionales. Esta situación resalta la necesidad de enfoques complementarios, como el modelado estructural *in silico*, para explorar con mayor detalle la arquitectura tridimensional del NCC y sus posibles sitios de interacción farmacológica.

Es bajo este contexto que las herramientas de modelado estructural computacional se han consolidado como alternativas robustas para el estudio de proteínas de membrana cuya caracterización experimental es limitada. Métodos de modelado por homología, refinamiento energético y, más recientemente, aproximaciones basadas en inteligencia artificial (AlphaFold) han prometido una notable capacidad para predecir estructuras tridimensionales con alta precisión [9, 13]. Estas estrategias permiten explorar la organización espacial de dominios transmembrana, identificar posibles cavidades funcionales y analizar diferencias estructurales entre ortólogos de distintas especies. En particular, el estudio comparativo entre el NCC humano y su contraparte en anguila ofrece una oportunidad para evaluar la conservación estructural y funcional del cotransportador, así como para identificar variaciones que puedan influir en la interacción con inhibidores farmacológicos.

Motivaciones

La hipertensión arterial y la retención de líquidos constituyen retos clínicos de primer orden, donde los diuréticos tiazídicos han demostrado eficacia terapéutica al inhibir de manera específica el cotransportador de Na^+/Cl^- sensible a tiazidas (NCC) en el túbulo contorneado distal [5, 6]. Sin embargo, el entendimiento a nivel atómico de los mecanismos mediante los cuales estas moléculas modulan la actividad del NCC sigue siendo limitado, en parte debido a la ausencia histórica de estructuras tridimensionales de alta resolución para este cotransportador y a las dificultades inherentes al estudio estructural de proteínas integrales de membrana [18, 19].

En la última década, los avances en criomicroscopía electrónica de partícula única han permitido resolver estructuras de diversos cotransportadores relacionados dentro de la superfamilia SLC12 [4]. No obstante, aún persisten interrogantes respecto a la localización precisa de los residuos responsables de la unión de tiazidas y a los determinantes estructurales que confieren especificidad farmacológica al NCC. Esta laguna estructural limita el diseño racional de inhibidores más selectivos y dificulta la optimización de su perfil terapéutico para minimizar efectos adversos [5].

Paralelamente, el desarrollo de metodologías computacionales para la predicción estructural de proteínas ha experimentado un progreso sustancial. Herramientas basadas en modelado por homología, refinamiento energético y enfoques recientes sustentados en inteligencia artificial han demostrado una alta capacidad predictiva en evaluaciones independientes como CASP [9, 13]. La integración de estas aproximaciones permite generar modelos tridimensionales plausibles, explorar estados conformacionales alternativos e identificar posibles cavidades de unión en ausencia de estructuras experimentales completas. En este contexto, la aplicación de un enfoque multiplataforma para el estudio estructural del NCC se presenta como una estrategia pertinente al desarrollo tecnológico actual para formular hipótesis sobre la interacción con tiazidas y contribuir al entendimiento molecular de su modulación farmacológica.

Objetivos

Objetivo general

Desarrollar y evaluar modelos tridimensionales del cotransportador de Na^+/Cl^- sensible a tiazidas (NCC), en sus formas monomérica y dimérica, para las variantes humana y de anguila europea (*Anguilla anguilla*) mediante herramientas computacionales basadas en inteligencia artificial y modelado comparativo, con el propósito de identificar cavidades estructurales relevantes y comparar el desempeño de distintas plataformas de predicción y refinamiento estructural en términos de calidad geométrica, estabilidad conformacional y propiedades de drogabilidad.

Objetivos específicos

1. Generar modelos estructurales del NCC humano y de anguila europea, tanto en su forma monomérica como dimérica, utilizando métodos de predicción basados en inteligencia artificial (AlphaFold, RosettaFold), modelado por homología (SwissModel) y modelado comparativo (Modeller).
2. Refinar las estructuras obtenidas mediante relajación energética con el protocolo Rosetta-Relax, optimizando la geometría local de cada modelo. Posteriormente, evaluar su calidad estereoquímica mediante análisis de Ramachandran y otros indicadores estructurales, seleccionando las estructuras con mejores parámetros geométricos.
3. Comparar la consistencia conformacional entre modelos a través de superposición estructural y cálculo del RMSD, evaluando la similitud tanto entre métodos de predicción como entre especies (humano vs. anguila), en sus estados monomérico y dimérico.
4. Identificar y caracterizar cavidades internas relevantes en los modelos tridimensionales utilizando CavityPlus, cuantificando parámetros asociados a la drogabilidad tales como DrugScore, volumen, área superficial accesible al solvente y número de cavidades clasificadas como fuertes.

5. Analizar la variabilidad intermetodológica en la predicción de cavidades, así como la conservación estructural de cavidades entre especies y la posible aparición de cavidades específicas en la interfase dímera o en regiones periféricas del cotransportador.

Capítulo 1

Antecedentes

1.1. Proteínas de Membrana

Las proteínas de membrana se definen como proteínas asociadas a las membranas celulares, siendo esta una ubicación estratégica que les permite desempeñar funciones esenciales en el transporte de sustancias, la transducción de señales y el mantenimiento del metabolismo, así como la homeostasis celular. Se localizan tanto en la membrana plasmática como en complejas estructuras subcelulares, aunque también pueden llegar hasta el ambiente extracelular en asociación con vesículos extracelulares. Las proteínas de membrana determinan la función precisa de cada membrana y, por extensión, la función de cada tipo de célula [20].

Desde el punto de vista fisiológico y farmacológico, las proteínas de membrana son de particular relevancia debido a que median el intercambio selectivo de sustancias, juegan un papel crucial en los mecanismos fisiológicos para la supervivencia de las células y por lo tanto, actúan como blancos terapéuticos debido a su implicación en un amplio rango de condiciones patológicas [21]. Las proteínas de membrana representan alrededor del 20-30 % del proteoma humano y por lo tanto más del 60 % de objetivos terapéuticos [20]. Su adecuada comprensión estructural y funcional resulta fundamental para el diseño de fármacos y el desarrollo de estrategias terapéuticas más eficientes.

No obstante, el estudio estructural y bioquímico de estas proteínas representa un desafío

significativo debido a su sofisticada naturaleza. Algunos de estos obstáculos se encuentran en características como su poca abundancia en las membranas biológicas, la estructura anclada a la membrana en la bicapa lipídica y la insolubilidad en fracciones acuosas debido a las diferencias en hidrofobicidad de los componentes protéicos; su estudio requiere de complejos protocolos de aislamiento que permitan un acercamiento por medios bioquímicos y biofísicos. Esta complejidad ha limitado históricamente la obtención de estructuras tridimensionales completas, lo que ha impulsado el desarrollo de enfoques experimentales y computacionales complementarios para su caracterización [22, 23].

1.1.1. Propiedades Estructurales de las Proteínas de Membrana

Las proteínas de membrana presentan características estructurales particulares que les permiten interactuar simultáneamente con la bicapa lipídica y con el medio acuoso circundante. En términos generales, poseen regiones hidrofóbicas que se insertan en el núcleo lipídico de la membrana y dominios hidrofílicos que se proyectan hacia los espacios intra y extracelulares, manteniendo así su estabilidad estructural y funcional [19].

En las proteínas integrales de membrana, los segmentos transmembrana adoptan predominantemente conformaciones de hélice α , caracterizadas por regiones hidrofóbicas que favorecen su inserción en la bicapa lipídica. Estudios de análisis genómico han mostrado que estas hélices transmembrana constituyen el motivo estructural más común en proteínas de membrana de organismos eucariotas, bacterianos y arqueanos. El número y la disposición de estos segmentos determinan la topología de la proteína, incluyendo la orientación de los extremos amino y carboxilo respecto a la membrana. En contraste, algunas proteínas de la membrana externa de bacterias presentan estructuras tipo barril β , que forman poros o canales [24].

Existen diferentes tipos de proteínas de membrana clasificadas acorde a su modo de asociación con la bicapa lipídica:

- **Proteínas integrales:** Se encuentran embebidas en la membrana y pueden atravesar completamente la bicapa mediante uno o múltiples dominios transmembrana. Muchas de ellas forman complejos oligoméricos que modulan su actividad funcional.

- **Proteínas periféricas:** Se asocian a la superficie de la membrana mediante interacciones electrostáticas o con otras proteínas, sin penetrar el núcleo hidrofóbico de la bicapa [25].

Además de su organización estructural primaria, muchas proteínas de membrana experimentan cambios conformacionales dinámicos como parte de su mecanismo funcional, especialmente aquellas involucradas en transporte activo o pasivo de iones y solutos. Estas transiciones estructurales dependen tanto de interacciones intramoleculares como del microentorno lipídico, el cual puede influir directamente en su estabilidad y actividad biológica. La interacción específica de estas proteínas con ciertos lípidos ayuda a estabilizar su estructura tridimensional, afectando directamente su funcionalidad y permitiendo que puedan responder adecuadamente a señales externas [23]

1.1.2. Métodos experimentales para el estudio de estructuras de proteínas de membrana

El estudio estructural de las proteínas de membrana ha sido un desafío debido a su complejidad y a su naturaleza hidrofóbica, que dificulta su extracción y cristalización en entornos acuosos; el reto, en términos experimentales, consiste en expresar y purificar a estas proteínas en condiciones de estabilidad y funcionalidad fuera de la bicapa lipídica [22]. Sin embargo, los avances en métodos experimentales como la cristalografía de rayos X, la resonancia magnética nuclear (RMN) y la microscopía electrónica de alta resolución (EM) han permitido obtener imágenes de alta resolución de estas proteínas, proporcionando información clave sobre su estructura y función.

- **Cristalografía de rayos X:** Este método ha sido el más exitoso en la obtención de estructuras detalladas de proteínas de membrana. Sin embargo, requiere que las proteínas estén en forma cristalina, lo cual es complicado para las proteínas de membrana debido a su baja solubilidad en agua y su dependencia de un entorno lipídico. Los desarrollos recientes, como el uso de difracción de rayos X con microcristales y la ingeniería de proteínas para mejorar su estabilidad, han facilitado el avance en este campo [18].

- **Resonancia Magnética Nuclear (RMN):** La RMN permite estudiar proteínas de membrana en un ambiente que se asemeja más al fisiológico, como en detergentes o en bicelas (membranas sintéticas), lo cual es útil para observar interacciones dinámicas. Sin embargo, la RMN es más adecuada para proteínas pequeñas o de tamaño medio, debido a las limitaciones técnicas actuales para proteínas de gran tamaño [23].
- **Microscopía electrónica de alta resolución (cryo-EM):** La microscopía electrónica ha emergido como una de las técnicas más poderosas para la determinación estructural de proteínas de membrana. Este método permite analizar proteínas en estado casi nativo, vitrificadas rápidamente en hielo amorfo, evitando la necesidad de cristalización. Los avances en detectores electrónicos directos y en algoritmos de procesamiento de imágenes han permitido alcanzar resoluciones cercanas al nivel atómico. La cryo-EM es particularmente ventajosa para el estudio de complejos macromoleculares grandes y proteínas de membrana que son difíciles de cristalizar, ampliando significativamente el número de estructuras disponibles en bases de datos estructurales. [26]

En conjunto, estas técnicas experimentales han permitido avances significativos en la caracterización estructural de proteínas de membrana, aunque cada una presenta ventajas y un gran desarrollo en años recientes, también se caracterizan por sus limitaciones dependiendo del tamaño, estabilidad y complejidad del sistema bajo estudio, además de aún ser consideradas como técnicas costosas, con un alto consumo de tiempo y en ocasiones poco factibles [21]

1.2. Cotransportador de Na^+/Cl^- (NCC)

El cotransportador de sodio y cloro sensible a tiazidas (**NCC**, por sus siglas en inglés) es una proteína de membrana ubicada en la porción apical de las células epiteliales del túbulo conorneado distal del riñón, perteneciente a la familia de transportadores SLC12 cloro-catiónicos. Este transportador es responsable de la reabsorción de sodio y cloro desde la orina en formación hacia el interior de la célula, un proceso crucial para el mantenimiento del equilibrio hidroeletrolítico y el control de la presión arterial. Dependiendo de la especie de donde fue clonado

puede contener entre 1002 a 1028 aminoácidos y se compone por un dominio hidrofóbico central que contiene 12 regiones transmembranales interconectadas por seis asas extracelulares y cinco intracelulares [15].

1.2.1. Función en el Balance de Sodio y Cloro

El NCC trabaja mediante un mecanismo de cotransporte electroneutro, acoplando la entrada de un ion de sodio (Na^+) y un ion de cloro (Cl^-) al interior celular. Este proceso constituye un transporte activo secundario, dependiente del gradiente electroquímico de sodio generado por la actividad de la Na^+/K^+ ATPasa en la membrana basolateral, la cual expulsa sodio hacia el espacio intersticial y mantiene baja su concentración intracelular [27].

El túbulo contorneado distal, por acción del NCC principalmente, contribuye aproximadamente al 5-10 % de la reabsorción total de sodio en el riñón, un porcentaje pequeño pero fundamental, ya que las alteraciones en su actividad pueden tener efectos significativos en la homeostasis del volumen extracelular y, en consecuencia, en la presión arterial [27, 28].

1.2.2. Papel en la Regulación de la Presión Arterial

El NCC constituye el blanco farmacológico de los diuréticos tiazídicos, fármacos ampliamente utilizados en el tratamiento de la hipertensión arterial. Estos compuestos inhiben directamente la actividad del cotransportador, disminuyendo la reabsorción de sodio y cloro en el túbulo contorneado distal. Como consecuencia, se incrementa la excreción urinaria de sodio (natriuresis) y agua, lo que conduce a una reducción del volumen extracelular y, en última instancia, de la presión arterial [27].

La actividad del NCC está finamente regulada por mecanismos moleculares que modulan su estado de fosforilación y su abundancia en la membrana apical. Un eje central de esta regulación es la vía de señalización WNK–SPAK/OSR1, la cual fosforila y activa al NCC en respuesta a cambios en el estado del volumen extracelular y en la concentración de electrolitos [27].

Entre los moduladores hormonales, la aldosterona puede aumentar la actividad del NCC de manera indirecta mediante la activación de la quinasas SGK1 y la modulación de la vía

WNK–SPAK/OSR1, favoreciendo la reabsorción de sodio en condiciones de hipovolemia. Por su parte, la hormona antidiurética (ADH) puede influir en la función del túbulo distal a través de mecanismos dependientes de AMPc, con posibles efectos moduladores sobre el NCC [27].

1.2.3. Implicaciones Clínicas

Mutaciones inactivantes en el gen **SLC12A3**, que codifica al NCC, causan el síndrome de Gitelman, un trastorno autosómico recesivo caracterizado por hipokalemia, hipomagnesemia, hipocalciuria y alcalosis metabólica. Estas alteraciones derivan de una disminución en la reabsorción de sodio y cloro en el túbulo contorneado distal, lo que conduce a una activación compensatoria del sistema renina–angiotensina–aldosterona y a un aumento en la secreción distal de potasio [15, 27].

Desde el punto de vista molecular, las mutaciones en *SLC12A3* pueden afectar la expresión, el tráfico hacia la membrana apical, la estabilidad estructural o la capacidad de fosforilación del cotransportador, comprometiendo su función. La variabilidad fenotípica observada en pacientes con síndrome de Gitelman sugiere que distintos tipos de mutaciones pueden alterar de manera diferencial la estructura y regulación del NCC [15].

En contraste, el aumento en la actividad del NCC se ha asociado con formas de hipertensión sensibles a sal, así como con trastornos relacionados con una activación excesiva de la vía WNK–SPAK/OSR1. De hecho, mutaciones en quinasas reguladoras como WNK1 y WNK4 pueden incrementar la fosforilación y activación del NCC, favoreciendo la retención de sodio y la expansión del volumen extracelular [27].

En este contexto, el estudio detallado de la estructura tridimensional y los mecanismos regulatorios del NCC no solo posee relevancia fisiológica, sino también implicaciones clínicas directas, al permitir comprender cómo alteraciones estructurales específicas pueden traducirse en disfunción del transporte iónico y en el desarrollo de hipertensión o síndromes de pérdida renal de sal.

1.3. Diuréticos Tiazídicos y Mecanismo de Acción

Los diuréticos tiazídicos son una clase de medicamentos ampliamente utilizados como primera línea para el tratamiento de la hipertensión arterial y otros trastornos relacionados con la retención de líquidos [29]. Estos compuestos actúan específicamente sobre el cotransportador de sodio y cloro (NCC), inhibiendo su función y modulando el balance hidroelectrolítico [15,27].

1.3.1. Mecanismo de Acción

El mecanismo de acción de los diuréticos tiazídicos se basa en la inhibición directa del cotransportador de sodio y cloro (NCC) en la membrana apical del túbulo contorneado distal. Al bloquear el transporte electroneutro de Na^+ y Cl^- , se reduce su reabsorción tubular, aumentando la cantidad de sodio que permanece en la luz tubular y favoreciendo la natriuresis [30].

El incremento en la carga distal de sodio también potencia su reabsorción en segmentos más distales del nefrón, particularmente en el túbulo colector, a través de canales epiteliales de sodio (ENaC). Este fenómeno genera un gradiente eléctrico luminal negativo que favorece la secreción de potasio (K^+), lo que explica la aparición de hipokalemia como uno de los efectos adversos más frecuentes asociados al uso de tiazidas [6].

En términos hemodinámicos, la pérdida inicial de sodio y agua conduce a una reducción del volumen extracelular y del volumen plasmático, lo que contribuye a la disminución de la presión arterial. A largo plazo, se han descrito además efectos vasculares adicionales que participan en la respuesta antihipertensiva sostenida [15].

1.3.2. Eficacia en el Tratamiento de la Hipertensión

Los diuréticos tiazídicos son considerados fármacos de primera línea en el tratamiento de la hipertensión arterial de acuerdo con guías clínicas internacionales [29]. Su eficacia terapéutica se fundamenta en múltiples mecanismos:

- Reducción sostenida de la presión arterial secundaria a la disminución inicial del volumen extracelular y del gasto cardíaco.

- Efectos vasculares a largo plazo, incluyendo disminución de la resistencia vascular periférica y mejora de la distensibilidad arterial, mecanismos parcialmente independientes del volumen intravascular.
- Bajo costo, amplia disponibilidad y extensa experiencia clínica acumulada.

Ensayos clínicos de gran escala, como el estudio ALLHAT, han demostrado que los diuréticos tiazídicos son efectivos en la prevención de eventos cardiovasculares mayores, incluyendo accidente cerebrovascular e insuficiencia cardíaca, particularmente en poblaciones de alto riesgo y pacientes de edad avanzada. Se considera que los diuréticos tiazídicos deben ser priorizados como la primera opción para el control de la hipertensión [31].

1.3.3. Limitaciones en el Diseño y Uso

A pesar de su eficacia clínica, los diuréticos tiazídicos presentan limitaciones y efectos adversos que deben considerarse en su uso terapéutico:

- **Hipokalemia:** Complicación frecuente secundaria al aumento del flujo distal de sodio y la consecuente secreción de potasio en el túbulo colector, lo que puede predisponer a arritmias cardíacas [6].
- **Hiponatremia:** Particularmente prevalente en pacientes de edad avanzada, asociada a una alteración en la capacidad de dilución urinaria y a la pérdida sostenida de sodio [28].
- **Hiperglucemia:** Se considera que la hipokalemia es causa directa de hiperglucemia inducida por diuréticos [6].
- **Eficacia reducida en enfermedad renal crónica avanzada:** La actividad de los tiazídicos disminuye cuando la tasa de filtración glomerular es significativamente baja, lo que limita su utilidad en estadios avanzados de insuficiencia renal [32].

Algunos otros efectos secundarios que se han encontrado son reacciones de hipersensibilidad, hiperuricemia y gota, posible daño en los riñones, hipopotasemia y alcalosis metabólica [28]. Desde el punto de vista del diseño farmacológico, los diuréticos tiazídicos fueron

desarrollados antes de que se conociera con detalle la estructura tridimensional del NCC. Recientemente se han resuelto estructuras del cotransportador mediante criomicroscopía electrónica, incluyendo complejos con diuréticos tiazídicos [33]. Sin embargo, no se dispone aún de una caracterización completa de todos los estados conformacionales y dinámicas estructurales relevantes para su función fisiológica y regulación. Esta información incompleta limita la identificación exhaustiva de los determinantes estructurales responsables de la afinidad, especificidad y modulación farmacológica del transportador.

El análisis estructural detallado del NCC mediante herramientas *in silico*, como las empleadas en el presente trabajo, representa una estrategia prometedora para caracterizar los sitios de unión, explorar dinámicas conformacionales y contribuir al diseño racional de nuevas moléculas con mayor especificidad y mejor perfil terapéutico.

1.4. Técnicas *In Silico* para el modelado de proteínas

El modelado de proteínas mediante métodos *in silico* constituye una herramienta central en la biología estructural contemporánea, permitiendo predecir, analizar y refinar estructuras tridimensionales cuando los datos experimentales son limitados o inexistentes. Estos enfoques abarcan desde el modelado por homología y métodos *de novo* hasta estrategias basadas en inteligencia artificial (IA) y aprendizaje profundo [21]. Su relevancia es particularmente significativa en el estudio de proteínas de membrana, cuya caracterización experimental suele ser compleja debido a su naturaleza hidrofóbica y dependencia del entorno lipídico.

1.4.1. Métodos de predicción estructural

Las técnicas de predicción estructural buscan generar estructuras, recibiendo como entrada principal la secuencia de aminoácidos correspondiente a la proteína de interés. Estos métodos generalmente se engloban por su tipo de enfoque:

- **Modelado por homología:** Utiliza proteínas homólogas, nombradas plantillas, para construir modelos de proteínas con estructura desconocida, partiendo del conocimiento de que

las secuencias evolutivamente relacionadas entre sí (familias) presentan estructuras similares.

- **Reconocimiento de plegamiento (threading):** A falta de homólogos cercanos, se utilizan métodos de reconocimiento de plegamiento (threading) que alinean la secuencia de interés contra una base de datos de estructuras conocidas, minimizando una función de energía o puntuación. Estos métodos asumen que el número de plegamientos estructurales es limitado, por lo que nuevas proteínas probablemente compartan arquitectura con estructuras ya determinadas.
- **De novo:** Construye estructuras tridimensionales directamente a partir de la secuencia de aminoácidos, sin necesidad de plantillas experimentales.
- **Híbridos:** Métodos integrativos que buscan la combinación de procesos experimentales con los distintos tipos de enfoques computacionales.

Estas técnicas permiten modelar proteínas de membrana, complejos macromoleculares y proteínas sin estructura cristalográfica conocida, contribuyendo al diseño racional de fármacos y a la comprensión de funciones biológicas. La predicción estructural por métodos computacionales abre la posibilidad al campo de la bioinformática estructural a gran escala, dejando atrás el cuello de botella que ha representado en la investigación la cobertura estructural puramente experimental [13,21].

1.4.2. Herramientas principales

1.4.2.1. AlphaFold

AlphaFold, desarrollado por DeepMind, constituye uno de los avances más significativos en la predicción estructural de proteínas. Su arquitectura se basa en modelos de aprendizaje profundo tipo *transformer*, capaces de integrar información proveniente de alineamientos múltiples de secuencia (MSA) y bases de datos estructurales para predecir relaciones espaciales entre residuos y generar modelos tridimensionales de alta precisión [13].

Su desempeño en la competencia CASP14 (Critical Assessment of Structure Prediction) demostró niveles de exactitud comparables a los obtenidos mediante métodos experimentales en numerosos casos, marcando un cambio paradigmático en la biología estructural computacional.

Ventajas:

- Alta precisión estructural incluso en ausencia de homólogos cercanos con identidad de secuencia elevada.
- Generación de métricas internas de confianza, como el *predicted Local Distance Difference Test* (pLDDT) y el *Predicted Aligned Error* (PAE), que permiten evaluar la calidad global y local del modelo.
- Disponibilidad de una base de datos pública (AlphaFold Protein Structure Database) que amplía significativamente la cobertura estructural del proteoma.

Limitaciones: Diversos estudios han señalado las limitaciones inherentes a este enfoque [3]:

- Aunque AlphaFold y AlphaFold-Multimer permiten modelar complejos proteicos, presentan limitaciones en la predicción de estados conformacionales alternativos, transiciones dinámicas y estructuras dependientes de ligandos o cofactores.
- Puede mostrar menor confiabilidad en regiones intrínsecamente desordenadas o altamente flexibles.
- No sustituye estudios dinámicos, como simulaciones de dinámica molecular, necesarios para analizar estabilidad conformacional y mecanismos funcionales.

1.4.2.2. SWISS-MODEL

SWISS-MODEL es una plataforma automatizada de modelado por homología ampliamente utilizada para la generación de estructuras tridimensionales a partir de la alineación de secuencias con plantillas experimentales disponibles en bases de datos como el Protein Data Bank (PDB) [10]. El sistema integra herramientas de búsqueda de plantillas, alineamiento y evaluación estructural en un entorno accesible vía web.

Ventajas:

- Plataforma intuitiva y accesible para usuarios sin experiencia avanzada en modelado estructural.
- Integración con bases de datos estructurales y funcionales como PDB y UniProt.
- Generación automatizada y rápida de modelos con métricas internas de calidad.

Limitaciones:

- La precisión del modelo depende directamente del grado de identidad de secuencia y cobertura estructural entre la plantilla y la proteína objetivo, en concordancia con la relación establecida entre divergencia de secuencia y variación estructural [34].
- Presenta menor confiabilidad cuando no existen plantillas homólogas cercanas o cuando las regiones de interés corresponden a bucles flexibles o dominios poco conservados.

1.4.2.3. Rosetta

Rosetta es una suite integral de herramientas para modelado estructural y diseño macromolecular que combina estrategias de ensamblaje de fragmentos, muestreo conformacional basado en Monte Carlo y funciones de energía empíricas para predecir y refinar estructuras proteicas [12,35].

El marco metodológico de Rosetta permite abordar tanto modelado por homología como predicción *de novo*, refinamiento estructural, diseño de proteínas y modelado de interacciones macromoleculares. Su arquitectura modular facilita la integración de diferentes protocolos según la naturaleza del problema estructural.

Ventajas:

- Alta versatilidad en predicción estructural, refinamiento y diseño racional de proteínas.
- Capacidad para modelar interacciones proteína-proteína, proteína-ligando y complejos macromoleculares.

- Uso de funciones de energía físicas y estadísticas que permiten optimización estructural mediante muestreo conformacional avanzado.

Limitaciones:

- Elevada demanda computacional, particularmente en protocolos *de novo* y de diseño.
- Curva de aprendizaje pronunciada debido a la complejidad de sus módulos y parametrización.
- Sensibilidad a la calidad de las restricciones iniciales y parámetros energéticos empleados.

1.4.2.4. Modeller

Modeller es un programa ampliamente utilizado para el modelado comparativo de proteínas basado en homología estructural. Su metodología se fundamenta en la satisfacción de restricciones espaciales derivadas de alineaciones entre la secuencia objetivo y estructuras proteicas conocidas [11, 36].

El algoritmo genera modelos tridimensionales mediante la optimización de una función objetivo que integra restricciones geométricas (distancias interatómicas, ángulos diedros, restricciones estereoquímicas) y términos energéticos estadísticos derivados de estructuras experimentales. Este enfoque permite construir modelos consistentes incluso cuando existen regiones variables o inserciones respecto a la plantilla.

Ventajas:

- Capacidad de generar modelos utilizando múltiples plantillas estructurales.
- Alta flexibilidad y personalización mediante scripts en Python.
- Especialmente útil para modelar mutaciones puntuales, regiones faltantes y bucles.
- Permite generar múltiples modelos y seleccionar el óptimo mediante funciones de evaluación como DOPE score.

Limitaciones:

- Dependencia crítica de la calidad del alineamiento secuencia-estructura.
- Precisión limitada en regiones con baja identidad de secuencia o alta flexibilidad estructural.
- No incorpora de manera nativa protocolos extensivos de refinamiento dinámico comparables a simulaciones completas de dinámica molecular.

1.4.3. Estado del arte en modelado de proteínas mediado por IA

El surgimiento de herramientas basadas en inteligencia artificial, como AlphaFold2 y RoseTTAFold, ha redefinido el panorama del modelado estructural al alcanzar niveles de precisión comparables a métodos experimentales en numerosos casos [13, 37]. Posteriormente, desarrollos como AlphaFold-Multimer y modelos basados en aprendizaje profundo a escala evolutiva (por ejemplo, ESMFold) ampliaron la capacidad de predicción hacia complejos macromoleculares y estructuras derivadas de modelos de lenguaje entrenados sobre bases de datos masivas de secuencias [13, 38]. Más recientemente, AlphaFold3 ha extendido estas capacidades hacia la predicción integrada de interacciones proteína–proteína, proteína–ligando y proteína–ácido nucleico [39].

No obstante, persisten desafíos relevantes, particularmente en la caracterización de estados conformacionales múltiples, transiciones dinámicas, ensamblajes transitorios y proteínas de membrana cuya estabilidad depende de lípidos específicos o modificaciones postraduccionales [40, 41]. Asimismo, la predicción estructural estática no sustituye la caracterización termodinámica ni cinética necesaria para comprender mecanismos moleculares complejos [42].

En el contexto biomédico, estos avances han impactado significativamente el diseño racional de fármacos, la ingeniería de proteínas y el análisis estructural de variantes patogénicas asociadas a enfermedades humanas. Sin embargo, la integración de predicción estructural basada en IA, refinamiento energético y simulaciones dinámicas continúa siendo un área activa de investigación, particularmente en sistemas de transporte de membrana [43].

En el caso particular del cotransportador Na^+/Cl^- (NCC), la aplicación combinada de herramientas de modelado estructural, evaluación de calidad y análisis de interacciones moleculares

permite explorar determinantes estructurales relevantes para su función, regulación y modulación farmacológica. Estas aproximaciones computacionales constituyen el fundamento metodológico del presente trabajo, enmarcado dentro del paradigma contemporáneo de biología estructural asistida por inteligencia artificial.

Capítulo 2

Definición del Problema

2.1. Problemática

El conocimiento estructural detallado de proteínas de membrana es fundamental para comprender sus mecanismos de transporte, regulación y reconocimiento molecular, así como para su explotación como blancos terapéuticos. No obstante, la determinación experimental de estructuras tridimensionales de este tipo de proteínas continúa enfrentando limitaciones significativas asociadas a su hidrofobicidad, inestabilidad fuera del entorno lipídico y flexibilidad conformational.

En el caso específico del cotransportador sodio-cloro sensible a tiazidas (NCC), la disponibilidad de información estructural de alta resolución es limitada, lo que restringe el análisis preciso de sus determinantes estructurales, sus posibles estados oligoméricos y sus sitios de interacción con fármacos. Esta carencia dificulta la comprensión mecánica de su función y la racionalización estructural de su modulación farmacológica.

Si bien las herramientas de predicción estructural basadas en inteligencia artificial han demostrado alta precisión en múltiples sistemas proteicos, la generación de modelos confiables para proteínas de membrana y complejos oligoméricos requiere evaluación comparativa, validación estructural rigurosa y análisis funcional complementario.

Por lo tanto, el problema central en el cual se enfoca este trabajo consiste en determinar

si es posible obtener modelos tridimensionales estructuralmente consistentes y funcionalmente relevantes del NCC por medio de métodos In Silico, que presenten características geométricas y fisicoquímicas compatibles con sitios de unión potencialmente drogables.

2.2. Hipótesis

Se plantea que la integración de herramientas de predicción estructural basadas en inteligencia artificial y modelado comparativo permitirá generar modelos tridimensionales confiables del cotransportador Na^+/Cl^- sensible a tiazidas (NCC), tanto en su forma monomérica como dimérica, los cuales presentarán estabilidad estructural, consistencia estereoquímica y cavidades conservadas con propiedades de drogabilidad favorables entre métodos de modelado independientes.

Capítulo 3

Metodología

3.1. Planteamiento

La metodología empleada en este trabajo se estructura en seis fases principales que permiten abordar de manera integral el modelado estructural del cotransportador NCC, así como el análisis comparativo entre distintas plataformas computacionales:

1. **Obtención y análisis de secuencias:** Se recuperaron las secuencias de aminoácidos del NCC humano (SLC12A3_HUMAN, UniProt: P55017) y de la variante β NCC de anguila europea (UniProt: Q2PDH2), verificando su integridad y características generales.
2. **Predicción estructural:** Se generaron modelos tridimensionales del NCC en sus formas monomérica y dimérica utilizando cuatro plataformas de predicción: AlphaFold (IA), RoseTTAFold (IA), SWISS-MODEL (modelado por homología) y Modeller (modelado comparativo). Cabe mencionar que en la generación de dímeros se excluyó al software RoseTTAFold.
3. **Refinamiento estructural:** Todos los modelos fueron sometidos al protocolo RosettaRelax mediante la plataforma ROSIE, con el objetivo de optimizar la geometría local, reducir tensiones energéticas y mejorar la calidad estereoquímica de las estructuras [44].
4. **Validación estereoquímica:** Se evaluaron los modelos refinados mediante MolProbity,

utilizando como criterios principales el análisis del diagrama de Ramachandran, la detección de rotámeros desfavorables y la identificación de colisiones atómicas. Solo los modelos con más del 96 % de residuos en regiones favorecidas fueron considerados de alta calidad [45].

5. **Caracterización de cavidades internas:** Se analizaron las estructuras seleccionadas utilizando CavityPlus, cuantificando parámetros relevantes de drogabilidad como DrugScore, volumen de cavidades y número de cavidades clasificadas como "fuertes" [46].
6. **Comparación estructural:** Se realizó una superposición estructural entre modelos utilizando ChimeraX y el algoritmo de alineamiento Needleman-Wunsch, con cálculo de RMSD global. Se analizaron tanto las similitudes intraespecie (humano vs anguila) como intermetodológicas (entre softwares) [47].

Cada una de estas etapas fue diseñada para responder a dos ejes fundamentales del estudio: (1) generar modelos de alta fidelidad del NCC en sus formas monomérica y dimérica para ambas especies, y (2) evaluar la consistencia, precisión y posible utilidad farmacológica de los modelos generados por diferentes herramientas computacionales.

Este enfoque metodológico no solo permite explorar la aplicabilidad de la inteligencia artificial y del modelado comparativo en proteínas de membrana de difícil cristalización, sino que también ofrece criterios cuantificables (como RMSD, DrugScore y volumen de cavidad) para validar los modelos en términos estructurales y funcionales.

3.2. Enfoque General

3.2.1. Descripción del Pipeline

Este trabajo adopta un enfoque metodológico computacional que integra predicción estructural, refinamiento estereoquímico, validación geométrica y análisis funcional de proteínas de membrana. La estrategia está diseñada para generar y caracterizar modelos tridimensionales

confiables del cotransportador de sodio y cloro sensible a tiazidas (NCC), tanto en su forma monomérica como dimérica, en dos especies: *Homo sapiens* y *Anguilla anguilla*.

El enfoque se sustenta en el uso combinado de herramientas clásicas de modelado por homología (SWISS-MODEL, Modeller) y plataformas basadas en inteligencia artificial (AlphaFold, RoseTTAFold), con el objetivo de comparar la calidad y consistencia estructural de los modelos generados por cada método. Posteriormente, se emplea un protocolo de relajación atómica (RosettaRelax) para refinar la geometría de los modelos y asegurar una conformación estereotópicamente óptima.

Una parte central del análisis es la identificación y caracterización de cavidades internas mediante la herramienta **CavityPlus**. Este software permite detectar regiones estructurales con potencial farmacológico por medio de distintas métricas que, al ser evaluadas en conjunto, permiten estimar un valor denominado como Druggability (drogabilidad) [46].

El análisis de cavidades permite no solo identificar sitios potenciales de unión a diuréticos tiazídicos, sino también comparar la recurrencia y localización de dichas cavidades entre modelos generados por diferentes métodos, así como entre especies. Este componente funcional del estudio aporta evidencia complementaria al análisis estructural puro, facilitando la evaluación de modelos en términos de *drogabilidad* y posible relevancia terapéutica.

En conjunto, este enfoque busca generar modelos estructurales de alta fidelidad para el NCC al tiempo que se establece un marco computacional robusto que pueda ser aplicado al estudio de otras proteínas de membrana con interés farmacológico.

La estrategia metodológica se estructuró en dos ramas principales desarrolladas de manera paralela, las cuales son descritas a continuación.

Por un lado, se realizó la preparación del ligando de interés que sería evaluado frente a los modelos estructurales generados. El compuesto fue obtenido de la base de datos PubChem [48] y posteriormente sometido a un proceso de optimización geométrica y minimización de energía utilizando el software Avogadro, con el objetivo de obtener una conformación estable adecuada para los análisis de interacción molecular.

De forma simultánea, se llevó a cabo el modelado estructural del cotransportador. En primer lugar, se obtuvieron las secuencias de aminoácidos necesarias a partir de la base de datos de Uni-

Prot [49]. Posteriormente, dicha secuencia fue empleada como entrada en distintos programas de modelado estructural —AlphaFold, Rosetta, SWISS-MODEL y Modeller— cada uno basado en enfoques metodológicos diferentes, incluyendo predicción basada en inteligencia artificial y modelado comparativo por homología.

Una vez generados los modelos tridimensionales, estos fueron sometidos a un proceso iterativo de evaluación y refinamiento. Inicialmente, cada modelo fue analizado en términos de calidad estereoquímica y geométrica mediante MolProbity. Posteriormente, se realizó un proceso de relajación estructural utilizando el protocolo ROSIE Relax, orientado a optimizar la conformación energética del sistema. Este ciclo de evaluación–refinamiento se repitió en tres ocasiones para cada modelo generado.

La visualización, inspección estructural y comparación entre modelos se realizaron mediante ChimeraX. Tras seleccionar los modelos con mejores parámetros de calidad, estos fueron analizados en CavityPlus para la identificación y caracterización de posibles sitios de unión con potencial farmacológico. Finalmente, las regiones relevantes de la proteína fueron delimitadas para llevar a cabo el análisis de interacción proteína–fármaco utilizando el ligando previamente preparado.

En la Figura 3.1 se presenta el diagrama de flujo que resume la metodología general descrita en esta sección.

3.3. Selección de Datos y Herramientas

3.3.1. Bases de datos estructurales

Para el modelado tridimensional del cotransportador de sodio y cloro sensible a tiazidas (NCC), se utilizaron bases de datos biomoleculares especializadas que permiten el acceso a secuencias primarias y estructuras cristalográficas de proteínas homólogas.

- **UniProtKB (Universal Protein Resource Knowledgebase):** Se utilizaron las secuencias completas de aminoácidos del NCC humano (*SLC12A3*, entrada P55017) y del NCC β

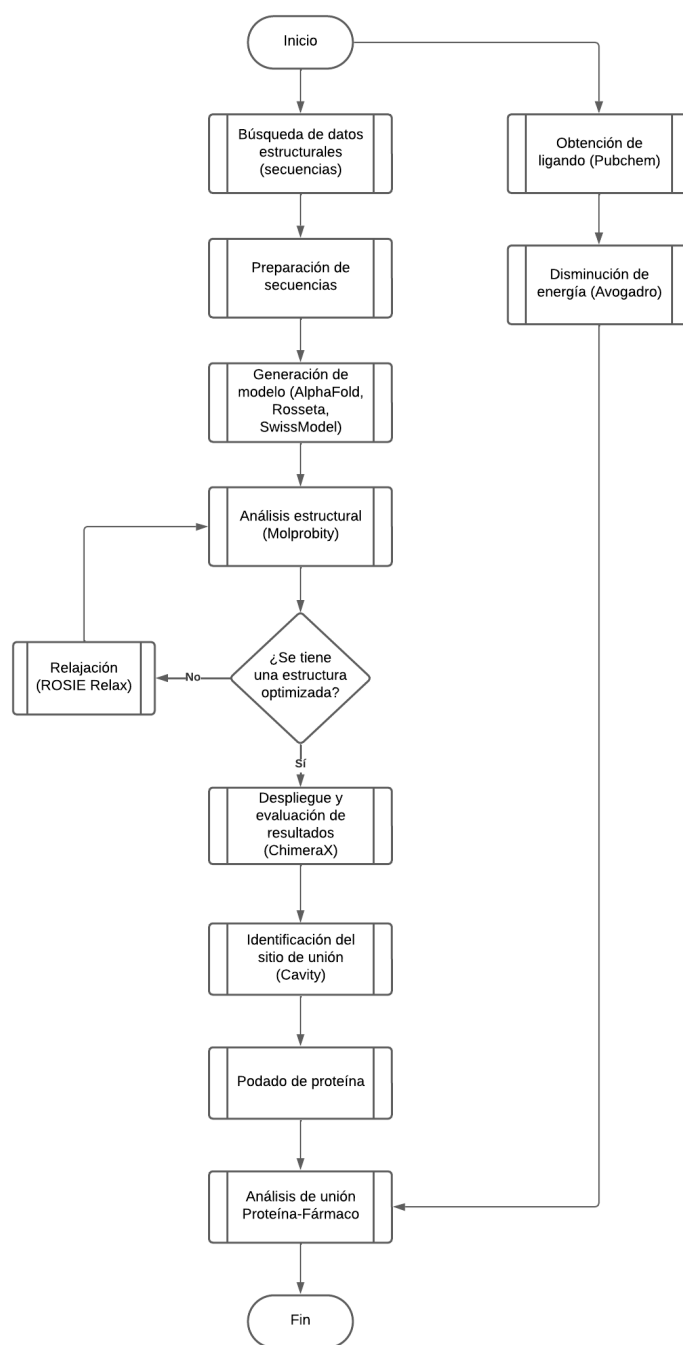


Figura 3.1: Diagrama de flujo para la metodología general propuesta para el modelado In Silico.

de *Anguilla anguilla* (entrada Q2PDH2). Estas secuencias sirvieron como punto de partida para todas las predicciones estructurales.

■ **Protein Data Bank (PDB):** Para los métodos de modelado por homología (SWISS-MODEL y Modeller), se seleccionaron estructuras experimentales de proteínas relacionadas con el NCC como plantillas de referencia. Entre los identificadores PDB empleados se encuentran:

- 8FHO, 8FHN — estructuras de transportadores electroneutros de la familia SLC12 en conformaciones alternantes determinadas por criomicroscopía electrónica [50].
- 8VPP, 8VPN, 9BWT — estructuras obtenidas por criomicroscopía electrónica del NCC humano en complejos con inhibidores [33].

Estas bases de datos proporcionan información crítica tanto para la predicción de estructuras *de novo* (cuando no existen estructuras homólogas cercanas) como para la selección racional de plantillas en metodologías de modelado comparativo. La calidad de la plantilla y la cobertura secuencial son factores determinantes en la precisión de los modelos obtenidos por homología.

En el caso particular de AlphaFold y RoseTTAFold, no fue necesaria la utilización explícita de plantillas, ya que estos algoritmos aplican redes neuronales profundas que aprenden representaciones estructurales a partir de múltiples alineamientos de secuencia y distancias evolutivas, generando predicciones tridimensionales directamente desde la secuencia primaria.

A continuación, se presentan las cadenas de aminoácidos empleadas para las dos secuencias principales (Beta NCC y NCC de humano). En la Figura 3.2 se visualiza la cadena de aminoácidos correspondiente a β NCC de anguila europea desplegada en un archivo tipo FASTA.

```

>EEL-EUROPEA-BETA-NCC
MGQRISGPEWPAGSQHGRFSVSEEAPPRYSCSHGDGRTLNMNLSNNTDGLVLPMTV
TMETEASDGVSERQDSRRPTLYSTIDAVPTVDFYNTMATGRNRRSRPSLETLRKVMD
DGETGSSMEDLGVDGYAGNQEQHSHEDRSQQNSQPVRFGWVTGVMIRCMNIW
GVILFLRLPWITSQAGIILTWIISMSVIVTSITALSVSAIASNGKVSSGGAYFMISRTLGP
MGGPIGMVFSFANALACAMNTVGFSETVRDLLIEYGAVIVDPINDVRIVGVATVTLILLIS
LAGMEWEAKTQILFFLVLLSFTNYFVGTVPASTEKQAMGFFSYRREIFLTNLLPDWRG
PEGNFFQMFAIFFPSAIGILSGANICGDLKDPETAIPKGTLMIAFWTTVSYLGISATIGGSV
LRDASGSLNDSLELNSSAVCEGVACSLGWNFSHCEVTHTCSFGLANNFQVLMSVSGF
GPLITAGIFAASLSSALAFVLVSAPKVFQCLCKDNIPYIGFFAKGYGKNDPLRAYFLTYLI
AVAFILIAQLNAIAPLISNFFLCSYALINFSCFHASITNSPGWRPSFRYYSHWTALFGAII
VLMFLLTWWAALISFSIILFLLGYVSYKKPEVNWGSSVQAGTYNMALSYSVSLAGVEDH
VKNFRPQCLVLTGPPNLRPALVDFVGSFTKNISLMICGDILMEDGTAVLPQRNVARLVK
WLNHRKVRAFYTPTSNDLREGASHLLQATGLGKLPNTLIMGFKTNWQECSPHSMED
YITITSDTFDSNYGVCLLRMMDGLDISEEIEGEVNHAFETESTTAPEQAAPGRNSEDRN
GDVGPTVQIRTVFQNKQGGKTIDVYISDDGGLTLLIPYLLRRRRWRLSKVRVFI
QHSMEDQHKEMVLLQKFRLDVHDVIVMTDSERPPNSKNLKRFDIAPFRLREGLHG
EAVLQDLKRDCPWKVSQKDLALRLKSERKVRLEIRKNSQHAALVLSPLVPQTDCP
SSLYMAWLDLTLYGLHCPSSLIRGNQQNVLTFCYQ

```

Figura 3.2: Secuencia de aminoácidos de β NCC de Anguilla anguilla, entrada UniProt Q2PDH2; 1043 residuos. Fuente de los datos: UniProt [49]

Por otro lado, para el NCC de humano se utilizó la siguiente cadena de aminoácidos, compuesta por 1039 residuos. En la Figura 3.3 se observa su contenido en un archivo tipo FASTA.

```

>sp|P55017|S12A3_HUMAN Solute carrier family 12 member 3
OS=Homo sapiens OX=9606 GN=SLC12A3 PE=1 SV=3
MAELPTTETPGDATLCSGRFTISTLLSSDEPSPAAYDSSHPSHLTHSTFCMRTFGYNT
IDVVPITYEHYANSTQPGEPKVRPTLADLHSLFKQEGRHLHALAFDSRPSHEMTDGLVEG
EAGTSSEKNPEEPVRFGWKGVMIKRLNIWGVILYLRLPWITAQAGIVLTWIIILLSVT
VTSITGLSISAISTNGKVSQGGTYFLISRSLGPELGGSIGLIFAFANAVGVAMHTVGF
TVRDLLQEYGAPIVDPINDIRIIAVSVTVLLAISLAGMEWESKAQVLFLLVIMVSFANY
LVGTLIPPSSEDKASKGFFSYRADIFVQNLVPDWRGPDGTFGFMFSIFFPSATGILAGANI
SGDLKDPALIAIPKGTLMIAFWTTISYLAISATIGSCVVRDASGVLDNTPGWDGACEGLA
CSYGNWTFECTQQHSCHYGLINYYQTMSMVSGFAPLITAGIFGATLSSALACLVSAKVF
QCLCEDQLYPLIGFFGKGYGKNKEPVRGYLLAYIAVAFIIIAELNTIAPIIISNFFLCSY
ALINFSCFHASITNSPGWRPSFYNNKWAALFGAIIISVIMFLLTWAAALIAIGVVLFL
LYVIYKKPEVNWGSSVQAGSYNLALSYSVGLNEVEDHIKNYRPQCLVLTGPPNFRPALV
FVGTFTRNLSLMICGHVLIQPHKQRMPELQLIANGHTKWLNRKIKAFYSVDVIAEDLRRG
VQILMQAAGLGRMKPNILVVGFKKNWQSAHPATVEDYIGILHDAFDNYGVCVMRMREGL
NVSKMMQAHINPVFDPADGKEASARVDPKALVKEEQATTIFQSEQGGKTIIDYIWLFD
GLTLLIPYLLGRKRRWSKCKIRVVGQINRMDQERKAIIISLSKFRLGFHEVHILPDIN
QNPRAEHTKRFEDMIAPFRLNDGFKDEATVNMRRDCPWKISDEEITKNRVKSLRQVRLN
EIVLDYSRDAALIVITLPIGRKGKCPSSLYMAWLETLSQDLRPPVILIRGNQENVLTFYC
Q

```

Figura 3.3: Secuencia de aminoácidos del NCC humano, entrada UniProt P55017; 1039 residuos. Fuente de los datos: UniProt [49]

3.3.2. Herramientas de Modelado

Se seleccionaron herramientas avanzadas para la simulación y análisis estructural del NCC, tomando en cuenta su precisión, facilidad de uso y respaldo en la literatura científica.

- AlphaFold: Una herramienta de predicción estructural basada en inteligencia artificial desarrollada por DeepMind. AlphaFold se eligió debido a su capacidad para generar modelos de alta precisión, incluso en ausencia de estructuras experimentales previas. Su desempeño ha sido validado en experimentos como el CASP (Critical Assessment of Structure Prediction) [13].
- SWISS-MODEL: Plataforma dedicada al modelado por homología, útil para comparar la estructura del NCC basada en proteínas homólogas resueltas experimentalmente. SWISS-MODEL fue seleccionado por su enfoque simplificado y su integración con bases de datos como UniProt [10].
- Rosetta: Software utilizado para el modelado de estructuras y refinamiento molecular. Su versatilidad lo convierte en una herramienta poderosa para optimizar modelos generados por homología y abordar problemas específicos de flexibilidad estructural [12].
- Modeller: Herramienta especializada en modelado por homología que emplea técnicas de satisfacción de restricciones espaciales para generar modelos estructurales a partir de alineaciones con proteínas conocidas. Se eligió por su precisión en la construcción de modelos tridimensionales cuando se dispone de plantillas confiables, así como por su capacidad para automatizar el modelado de múltiples variantes estructurales [36].

3.4. Proceso de Modelado y Simulación

3.4.1. Generación de modelos por AlphaFold

Para la predicción estructural basada en inteligencia artificial, se utilizó la plataforma AlphaFold, desarrollada por DeepMind. En particular, se empleó la versión AlphaFold Multimer (v3), diseñada para modelar tanto proteínas individuales como complejos multimoleculares. A partir de las secuencias de aminoácidos obtenidas desde UniProt, se generaron modelos tridimensionales del cotransportador NCC en dos variantes taxonómicas (humano y anguila europea) y en

dos estados oligoméricos (monómero y dímero). Para cada combinación de especie y forma estructural se generaron cinco réplicas independientes, totalizando 20 modelos:

- 10 modelos monoméricos (5 humano + 5 anguila).
- 10 modelos diméricos (5 humano + 5 anguila).

El proceso de predicción incluyó la construcción de alineamientos múltiples de secuencia (MSA), la detección de homólogos remotos y la inferencia estructural mediante redes neuronales profundas. Las predicciones fueron acompañadas de dos métricas internas [51]:

- **pLDDT (Predicted Local Distance Difference Test):** estima la fiabilidad local de la predicción para cada residuo. Se consideraron aceptables los modelos con valores promedio superiores al 80 %.
- **PAE (Predicted Aligned Error):** representa la incertidumbre esperada en la distancia entre pares de residuos, útil para interpretar la orientación relativa de dominios o subunidades.

Los modelos generados fueron posteriormente sometidos a refinamiento estereoquímico mediante el protocolo RosettaRelax, descrito en la sección correspondiente. Las estructuras refinadas se utilizaron en los análisis de validación geométrica y caracterización funcional detallados en capítulos posteriores.

3.4.2. Generación de modelos por Rosetta

Como segundo enfoque basado en inteligencia artificial, se empleó la plataforma RosettaFold, desarrollada por el Baker Lab. Esta herramienta utiliza una arquitectura basada en redes neuronales tridimensionales para predecir la estructura de proteínas a partir de su secuencia primaria, integrando información evolutiva derivada de múltiples alineamientos de secuencia (MSA) y predicciones de contacto.

Para este trabajo se utilizaron modelos disponibles a través del servidor *Robetta*, el cual permite predicciones automatizadas mediante la versión pública de RosettaFold para estructuras

monoméricas. Debido a las limitaciones computacionales actuales del sistema, esta herramienta no fue empleada para generar modelos diméricos.

Se generaron un total de **10 modelos monoméricos**, distribuidos de la siguiente forma:

- 5 modelos del NCC humano (*Homo sapiens*).
- 5 modelos del β NCC de anguila europea (*Anguilla anguilla*).

Los modelos generados se acompañan de una evaluación interna basada en puntajes de confianza derivados de la probabilidad de contacto y de la consistencia de la topología secundaria. Sin embargo, a diferencia de AlphaFold, RosettaFold no genera puntuaciones estandarizadas como pLDDT o PAE. Por esta razón, los modelos fueron directamente evaluados estructuralmente tras el refinamiento con RosettaRelax.

Una vez generados, los modelos fueron preparados en formato PDB para su posterior procesamiento mediante protocolos de relajación y validación estereoquímica, descritos en secciones siguientes.

A diferencia de AlphaFold, Rosetta no se basa en aprendizaje profundo, sino en principios físico-químicos y estadísticos. Sin embargo, su capacidad de personalización, su uso extensivo en el refinamiento estructural y su aplicabilidad en problemas complejos como mutaciones, diseño de proteínas o interacciones moleculares lo convierten en una herramienta robusta y ampliamente validada en la comunidad científica [12].

3.4.3. Generación de modelos por SwissModel

SWISS-MODEL es una plataforma en línea que permite la predicción estructural de proteínas mediante modelado por homología. Esta técnica se basa en la suposición de que dos proteínas con alta identidad de secuencia comparten una arquitectura tridimensional similar. Para ello, se requieren estructuras cristalográficas conocidas (plantillas), que permiten construir modelos de proteínas homólogas cuya estructura experimental no ha sido resuelta.

En este trabajo, SWISS-MODEL se utilizó para generar modelos tanto monoméricos como diméricos del cotransportador NCC para ambas especies. Se generó la siguiente cantidad de

modelos:

- 10 modelos monoméricos (5 humano + 5 anguila).
- 10 modelos diméricos (5 humano + 5 anguila).

Durante el proceso de modelado, SWISS-MODEL identificó automáticamente las mejores plantillas disponibles en el Protein Data Bank (PDB), priorizando estructuras con alta cobertura, identidad y resolución experimental. Entre las principales plantillas seleccionadas se encuentran las entradas PDB: 8VPP, 8VPN, 9BWT, 8FHO y 8FHN.

Cada modelo generado fue evaluado preliminarmente por la plataforma mediante métricas internas como el GMQE (Global Model Quality Estimation) y el QMEAN (Qualitative Model Energy Analysis), que permiten estimar la confiabilidad del modelo construido. No obstante, para efectos de comparación estandarizada con los otros métodos, todos los modelos fueron posteriormente refinados mediante RosettaRelax y evaluados estructuralmente con MolProbity.

En el modelado por homología la elección de plantillas es un paso fundamental para la correcta predicción estructural. En esta metodología, los principales parámetros de interés formados por la relación entre la plantilla y la secuencia base fueron los siguientes [10]:

- **GMQE (Global Model Quality Estimate):** Valores entre 0 y 1, indican la calidad estimada del modelo a generar.
- **QSQE (Quaternary Structure Quality Estimate):** indica la calidad de las interacciones entre cadenas dado el alineamiento. Se buscan valores $QSQE > 0.7$.
- **Identity:** Porcentaje de aminoácidos exactamente alineados entre sí al comparar las dos secuencias.
- **Method:** Tipo de método por el cual fue obtenida la secuencia propuesta como plantilla. Se da preferencia a aquellas obtenidas por métodos experimentales.

En la Tabla 3.1 se presentan los parámetros de GMQE, QMEAN, Identity (identidad), Method (método de obtención de la estructura) y Oligo-State (estado oligomérico) de los templates (plantillas) seleccionados para el modelado por plantilla del NCC humano.

Template	GMQE	QSQE	Identity	Method	Oligo State
8vpp.1. A	0.76	0.76	100	EM	Homo-dimer
9bwt.1. A	0.77	0.77	99	EM	Homo-dimer
8vpn.1. A	0.73	0.72	100	EM	Homo-dimer
7y6i.1. A	0.57	0.78	99.90	EM	Homo-dimer
8vpn.1. A	0.70	0.72	100	EM	Homo-dimer

Tabla 3.1: Principales parámetros de los templates para NCC humano.

En la Tabla 3.2 se presentan los parámetros de GMQE, QMEAN, Identity (identidad), Method (método de obtención de la estructura) y Oligo-State (estado oligomérico) de los templates (plantillas) seleccionados para el modelado por plantilla del β NCC de anguila europea.

Template	GMQE	QSQE	Identity	Method	Oligo State
8fho.1. B	0.75	0.49	61.78	EM	Homo-dimer
9bwt.1. A	0.78	0.78	61.00	EM	Homo-dimer
8fhn.1. A	0.75	0.51	61.78	EM	Homo-dimer
8vpp.1. A	0.76	0.79	62.96	EM	Homo-dimer
8vpn.1. A	0.75	0.79	62.96	EM	Homo-dimer

Tabla 3.2: Principales parámetros de los templates para β NCC.

Para la elección de ambos tipos de plantillas se dio prioridad a aquellas donde GMQE era el valor más alto entre todas las posibles opciones.

Los archivos estructurales generados por SWISS-MODEL fueron exportados en formato PDB y almacenados para su análisis posterior en los flujos de validación y caracterización descritos en capítulos siguientes.

3.4.4. Generación de modelos por Modeller

Modeller es una herramienta de modelado por homología ampliamente utilizada para la predicción estructural de proteínas a partir de una o varias plantillas experimentales. Su principio de funcionamiento consiste en alinear la secuencia blanco con las plantillas disponibles, identificar regiones conservadas y generar una conformación tridimensional satisfaciendo restricciones espaciales derivadas del alineamiento.

Para este trabajo se utilizó la versión 10.1 de Modeller, en combinación con un entorno Python. Se generaron modelos estructurales tanto para el NCC humano como para el β NCC de anguila europea, en sus formas monomérica y dimérica. A diferencia de las herramientas basadas en inteligencia artificial, Modeller requiere la selección explícita de estructuras plantilla. En este caso, se utilizaron las siguientes entradas del Protein Data Bank (PDB): 8FHN, 8FHO, 8VPP, 8VPN y 9BWT. Estas fueron elegidas por su relevancia funcional y nivel de identidad con las secuencias blanco.

Se generaron cinco modelos monoméricos y cinco modelos diméricos para cada especie. Los modelos se generaron mediante scripts personalizados en Python que especifican los alineamientos, las plantillas y los parámetros de construcción. Posteriormente, todos los modelos fueron almacenados en formato PDB y preparados para su refinamiento estructural mediante RosettaRelax.

Cabe destacar que, a diferencia de otras plataformas empleadas en este estudio, Modeller no incluye un módulo nativo de refinamiento estereoquímico posterior al modelado, además, esta plataforma es la que más permite personalizar y seleccionar criterios y opciones para el desarrollo de las estructuras, de manera que requiere un mayor entendimiento de los fundamentos químicos de las proteínas así como de la computación para el desarrollo de los scripts. Por ello, el protocolo RosettaRelax resultó esencial para optimizar la geometría local de las estructuras y permitir su validación estructural en condiciones equivalentes a las de los otros métodos.

3.4.5. Cantidad de modelos base generados

En la Tabla 3.3 se observa a manera de resumen la relación de modelos generados por cada software empleado, descomponiendo las cantidades en monómeros y dímeros por especie.

Tabla 3.3: Resumen de modelos generados por método, especie y estado oligomérico.

Método	Monómeros NCC Humano	Monómeros β -NCC (Anguila)	Dímeros NCC Humano	Dímeros β -NCC (Anguila)	Total
AlphaFold	5	5	5	5	20
RosettaFold	5	5	0	0	10
SWISS-MODEL	5	5	5	5	20
Modeller	5	5	5	5	20

3.5. Análisis estructural de los modelos y relajación

3.5.1. Análisis estructural post relajación con Molprobability

Una vez que se tienen los modelos bases, corresponden las etapas iterativas de refinamiento y evaluación de los mismos. Para ello se emplean dos softwares principales:

1. Molprobability: evaluar las características geométricas y energéticas del modelo.
2. ROSIE: relajación de la estructura.

Primero, es necesario obtener las evaluaciones correspondientes al primer modelo adquirido, de manera que este se ingresa a Molprobability, a partir de estos resultados se podrá definir si las relajaciones posteriores verdaderamente mejoran la calidad de la estructura. MolProbability permite evaluar la estructura por medio de los ángulos generados y la geometría. Para el análisis de los ángulos, se emplea el gráfico de Ramachandran, que examina los ángulos phi y psi de los enlaces en la cadena principal de aminoácidos de la proteína. Este gráfico ayuda a identificar las conformaciones posibles y estables, señalando los ángulos que minimizan la energía y son comunes en proteínas nativas. Otros parámetros en los que se hace énfasis son [45]:

- Ramachandran Outliers: Este valor representa los residuos de aminoácidos cuyos ángulos de torsión, se encuentran fuera de las regiones permitidas en el diagrama de Ramachandran. Un valor bajo de outliers (menor al 0.05) indica que la mayoría de los residuos están en conformaciones energéticamente favorables, lo cual sugiere que la estructura es de buena calidad. Tener un bajo porcentaje de outliers significa que hay pocas distorsiones en la estructura de la proteína.
- Ramachandran Favored: Un valor alto en esta métrica (mayor al 98%) sugiere que la mayoría de los ángulos de torsión están en las regiones más estables del diagrama. Estas regiones representan conformaciones de baja energía, lo cual es ideal para la estabilidad estructural de la proteína. Cuanto mayor sea este valor, mejor es la calidad conformacional y estructural de la proteína. En este estudio, se propuso considerar como modelos de alta calidad aquellos con valores superiores al 96%.

- Rama Distribution Z-Score: Este valor mide la conformación global de la proteína en comparación con una estructura típica. Un valor cercano a cero sugiere que la conformación de la proteína es común y estable, mientras que un valor bajo (menor a 2) indica que la estructura tiene desviaciones mínimas, reforzando la estabilidad estructural. Un Z-score alto, por el contrario, podría implicar desviaciones significativas y potencialmente problemas estructurales.

La iteración entre evaluación - relajación consistió en que cada modelo base fue sometido a 3 rondas de relajación, cada una de ellas con su respectiva evaluación por Molprobity. Esto dio lugar a la identificación y selección de las estructuras cuyos parámetros indican la calidad más alta. Cada ronda de relajación en la plataforma de ROSIE proporciona como salida un archivo tipo pdb con la estructura optimizada, de manera que por cada modelo base se llegaron a obtener 4 versiones (1 base + 3 relajadas).

Los modelos seleccionados tras este proceso cumplieron con los criterios mínimos de calidad estereoquímica establecidos, con porcentajes elevados de residuos en regiones favorecidas del diagrama de Ramachandran, ausencia significativa de colisiones atómicas y conformaciones rotaméricas permitidas. Estos resultados validan la eficacia del protocolo RosettaRelax en la optimización geométrica de los modelos, estableciendo una base estructural confiable para los análisis posteriores. En particular, la calidad alcanzada permite avanzar con confianza hacia la caracterización funcional de las estructuras, incluyendo el análisis de cavidades internas y propiedades de drogabilidad.

3.5.2. Comparación de resultados de cada modelo y proteína

Los resultados arrojados por Molprobity para la evaluación estructural de cada modelo obtenido fueron registrados en una tabla en excel, recordando que los parámetros de interés son: Ramachandran Outliers, Ramachandran Favored y Ramachandran Z-Score. La generación de esta tabla permite identificar los mejores modelos generados por cada software, así como para cada cadena de aminoácidos (humano y anguila), de manera que por cada uno de ellos se encontraron los modelos con mayor porcentaje en los parámetros de interés definidos. Por cada

software y para cada una se seleccionó un modelo, bajo el criterio de que, en primer lugar, el valor de Ramachandran favored fuese mayor al 96 %. En el caso de que varios cumplieren con el criterio anterior, se optó por aquel con un menor valor en el parámetro de Ramachandran Outliers. A partir de ello, se encontraron las que aparentemente son las mejores 14 predicciones estructurales entre las 280 generadas.

Estos modelos seleccionados fueron ingresados al software ChimeraX para la visualización de su estructura, así como para realizar un análisis de matchmaking. Posteriormente, se procedió al análisis de cavidades.

3.6. Análisis de cavidades

Una vez seleccionados los mejores modelos arrojados por cada software, se realizó un análisis de las cavidades por medio de CavityPlus para determinar información en relación al DrugScore de cada modelo.

Para evaluar la capacidad de unión a ligandos de los modelos tridimensionales generados, se utilizó la plataforma en línea **CavityPlus**, una herramienta computacional desarrollada para la detección, evaluación y análisis funcional de cavidades en estructuras proteicas. CavityPlus combina métodos geométricos y energéticos para identificar potenciales sitios activos, caracterizarlos estructuralmente, y predecir su afinidad con compuestos tipo fármaco (*drug-like molecules*) [52].

La herramienta permite un análisis integral que incluye la predicción de cavidades, el cálculo de su superficie accesible, volumen, y diversos descriptores relacionados con su posibilidad de ser blanco farmacológico. Entre los principales parámetros reportados por CavityPlus se incluyen:

- **DrugScore:** Es una puntuación cuantitativa que estima la capacidad de una cavidad para unirse a fármacos. Se calcula combinando características geométricas (volumen, forma, accesibilidad) y propiedades fisicoquímicas (polaridad, hidrofobicidad, carga) de la cavidad. Un valor alto sugiere una mayor probabilidad de que la cavidad pueda alojar compuestos bioactivos.

- **Druggability promedio:** Se refiere al promedio de la métrica denominada *Druggability Score* para todas las cavidades identificadas en un modelo. Este valor, que oscila típicamente entre 0 y 3, clasifica las cavidades según su potencial de unión a fármacos:
 - <1.0: poco drogable.
 - 1.0–2.0: moderadamente drogable.
 - >2.0: altamente drogable.
- **Cavidades “strong”:** Número total de cavidades en un modelo cuya puntuación de *druggability* supera el umbral definido (por ejemplo, *Druggability Score* > 2,5), y que, por lo tanto, son candidatas sólidas a sitios funcionales o blancos terapéuticos.
- **Área superficial:** Representa la superficie accesible al solvente (SASA, por sus siglas en inglés) de la cavidad. Se mide en Å² y permite inferir el tamaño y accesibilidad del sitio. Cavidades con mayor área superficial son, en general, más favorables para interacciones con moléculas grandes o flexibles.
- **Volumen:** Corresponde al volumen interno de la cavidad medido en Å³. Aunque no siempre se reporta en los archivos de salida resumidos, es un descriptor importante para determinar la capacidad espacial de una cavidad.

Estos parámetros fueron calculados para todos los modelos seleccionados (monómeros y dímeros de NCC humano y de anguila), y posteriormente comparados para determinar qué software de modelado genera estructuras con cavidades con mayor potencial para posibles aplicaciones farmacológicas con el ligando de interés, en este caso, tiazida.

Capítulo 4

Resultados

Parte de los resultados presentados en este capítulo, particularmente los relacionados con la comparación estructural y la identificación de cavidades, fue publicada previamente en forma resumida en [53]. La presente tesis desarrolla con mayor amplitud la metodología, los resultados individuales y su análisis, mientras que [53] presenta una síntesis de los principales hallazgos mostrados aquí y se enfoca mayormente en el análisis de las cavidades.

4.1. Impacto cuantitativo del refinamiento iterativo

Con el objetivo de evaluar el efecto del protocolo de relajación estructural, se compararon cuantitativamente las métricas estereoquímicas de los modelos iniciales (crudos) con aquellas obtenidas tras el tercer (último) ciclo de refinamiento.

En términos de porcentaje de residuos en regiones favorecidas del diagrama de Ramachandran, se observó una mejora promedio que varió según el software y la forma oligomérica. En los modelos diméricos, Modeller presentó el mayor incremento, con mejoras promedio cercanas al 9–10 %, mientras que AlphaFold mostró incrementos más moderados (4–6 %), indicando que sus modelos iniciales ya presentaban mayor coherencia geométrica.

En monómeros, las mejoras fueron más homogéneas entre AlphaFold, SWISS-MODEL y Modeller ($\cong 3\%$), mientras que RosettaFold mostró variaciones mínimas, lo que sugiere una menor sensibilidad al refinamiento bajo las condiciones aplicadas.

La reducción en el porcentaje de residuos outliers fue especialmente notable en modelos diméricos generados por Modeller, con disminuciones promedio superiores al 8 %, lo que evidencia la efectividad del refinamiento para aliviar tensiones estéricas en estructuras complejas.

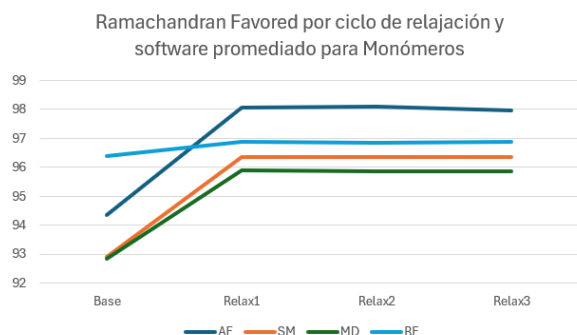
En conjunto, estos resultados demuestran que el impacto del refinamiento depende del método de modelado inicial y de la organización oligomérica de la proteína. Los modelos derivados de enfoques comparativos tradicionales se benefician en mayor medida del proceso iterativo, mientras que las predicciones basadas en inteligencia artificial muestran mayor estabilidad geométrica desde etapas tempranas.

Tabla 4.1: Mejora promedio de parámetros estereoquímicos tras el tercer ciclo de relajación (Relax3 vs modelo inicial). Valores positivos en Δ Favored y Δ Outliers indican mejora geométrica. AF: AlphaFold, MD: Modeller, SM: SwissModel, RF: RosettaFold.

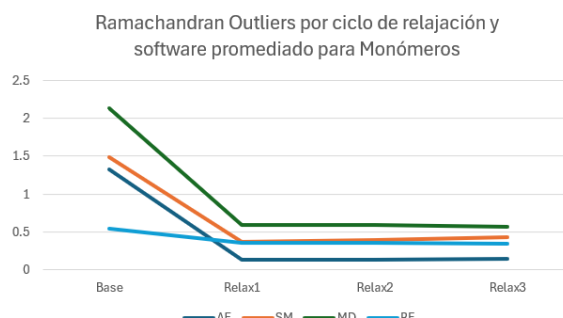
Forma	Software	Especie	Δ Favored (%)	Δ Outliers (%)	$\Delta Z $
DIM	AF	Anguila	6.58	2.30	0.07
DIM	AF	Humano	4.71	1.50	-0.02
DIM	MD	Anguila	9.73	8.67	0.49
DIM	MD	Humano	9.55	10.11	0.90
DIM	SM	Anguila	3.76	1.06	-0.13
DIM	SM	Humano	2.86	0.68	0.44
MON	AF	Anguila	3.92	1.36	0.28
MON	AF	Humano	3.34	1.02	-0.14
MON	MD	Anguila	3.07	1.55	-0.08
MON	MD	Humano	3.01	1.57	0.06
MON	RF	Anguila	0.17	0.17	-0.19
MON	RF	Humano	0.82	0.22	0.18
MON	SM	Anguila	3.39	1.06	0.25
MON	SM	Humano	3.52	1.04	0.25

Como se observa en la Tabla 4.1, los modelos diméricos generados por Modeller presentaron el mayor incremento promedio en residuos favorecidos ($\approx 9\text{--}10\%$), así como la mayor reducción de outliers (hasta 10.11%), lo que indica una mayor dependencia del refinamiento iterativo. En contraste, los modelos generados por AlphaFold mostraron mejoras más moderadas, reflejando una mayor estabilidad geométrica inicial. Los modelos monoméricos exhibieron incrementos más homogéneos entre plataformas ($\approx 3\text{--}4\%$), mientras que RosettaFold mostró el

menor impacto tras la relajación.

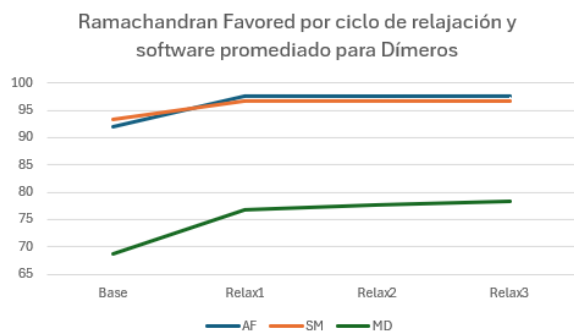


(a) Ramachandran favored promedio para monómeros

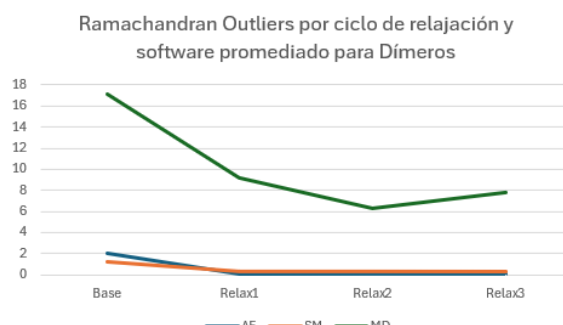


(b) Ramachandran outliers promedio para monómeros

Figura 4.1: Comparación de métricas Ramachandran promedio (Favored y Outliers) para modelos monoméricos conforme a los ciclos de relajación.



(a) Ramachandran favored promedio para monómeros



(b) Ramachandran outliers promedio para monómeros

Figura 4.2: Comparación de métricas Ramachandran promedio (Favored y Outliers) para modelos diméricos conforme a los ciclos de relajación.

En las Figuras 4.1 y 4.2 se puede visualizar el análisis de la evolución de las métricas a lo largo de los ciclos de relajación. Esto mostró que la mayor mejora geométrica ocurrió durante el primer ciclo (Relax1) sin importar el software, observándose posteriormente incrementos progresivamente menores en los ciclos subsecuentes. Esta tendencia sugiere un proceso de con-

vergencia estructural, en el cual las tensiones estéricas iniciales son corregidas tempranamente y la estructura alcanza un estado geoméricamente estable tras las iteraciones finales. Es relevante destacar que, a pesar de las mejoras en parámetros locales como el porcentaje de residuos favorecidos y la reducción de outliers, no se observaron cambios drásticos en la arquitectura global de los modelos seleccionados, lo que indica que el refinamiento optimizó la estereoquímica sin alterar significativamente la organización tridimensional general. En conjunto, estos resultados validan el protocolo iterativo empleado como un paso necesario para garantizar la calidad estructural antes de los análisis comparativos y de cavidades realizados en las secciones posteriores.

4.2. Modelo estructural del NCC

Al concluir los ciclos de relajación + evaluación se obtuvieron como resultado 280 modelos:

- **Monómeros:** 20 modelos (5 de base más sus 3 versiones optimizadas) por software y por especie (20x4x2), resultando en 160 modelos de monómeros.
- **Dímeros:** 20 modelos (5 de base más sus 3 versiones optimizadas) por software y por especie (20x3x2), resultando en 120 modelos de dímeros.

A partir de los métricas evaluadas con Molprobity se seleccionaron aquellos para las siguientes etapas del estudio. En las siguientes subsecciones se presentaran los modelos seleccionados para las dos especies analizadas.

4.2.1. Beta NCC Anguila

En la Tabla 4.2 se presentan los parámetros de evaluación obtenidos con MolProbity para los cuatro modelos seleccionados (uno por cada software) del β NCC de anguila en su forma monomérica. Estas estructuras fueron elegidas por cumplir con los criterios propuestos de calidad estereoquímica establecidos para el análisis estructural.

Tabla 4.2: Parámetros de interés de los modelos de monómero seleccionados para β NCC de anguila.

Beta NCC Monómeros	Ramachandran Outliers ($<0.05\%$)	Ramachandran Favored ($>96\%$)	Ramachandran Z- Score (abs(ZScore) <2)	“-+”
AF-model01-relaxed3	0.10	98.37	-0.19	0.24
SM-model05-relaxed1	0.00	97.76	-0.37	0.26
MD-model04-relaxed1	0.19	97.79	-0.49	0.23
RF-model04-relaxed2	0.19	98.17	0.00	0.25

De igual manera con respecto a los dímeros se presentan los modelos seleccionados por software junto con sus respectivos parámetros de evaluación en la Tabla 4.3:

Tabla 4.3: Parámetros de interés de los modelos de dímero seleccionados para β NCC de anguila.

Beta NCC Dímeros	Ramachandran Outliers ($<0.05\%$)	Ramachandran Favored ($>96\%$)	Ramachandran Z- Score (abs(ZScore) <2)	“-+”
AF-model5-relaxed2	0.05	97.60	0.04	0.18
SM-model3-relaxed2	0.28	98.02	-0.49	0.18
MD-model2-relaxed2	0.11	97.92	-0.40	0.19

A continuación, se presenta el modelo de mayor calidad para el monómero de β NCC de anguila europea. Este modelo fue generado usando AlphaFold y corresponde a la tercera ronda de relajación. Como parámetro notable se tiene un valor de 98.37 % de residuos en regiones favorecidas del diagrama de Ramachandran. La visualización del modelo se realizó en ChimeraX y se presenta en la Figura 4.3:

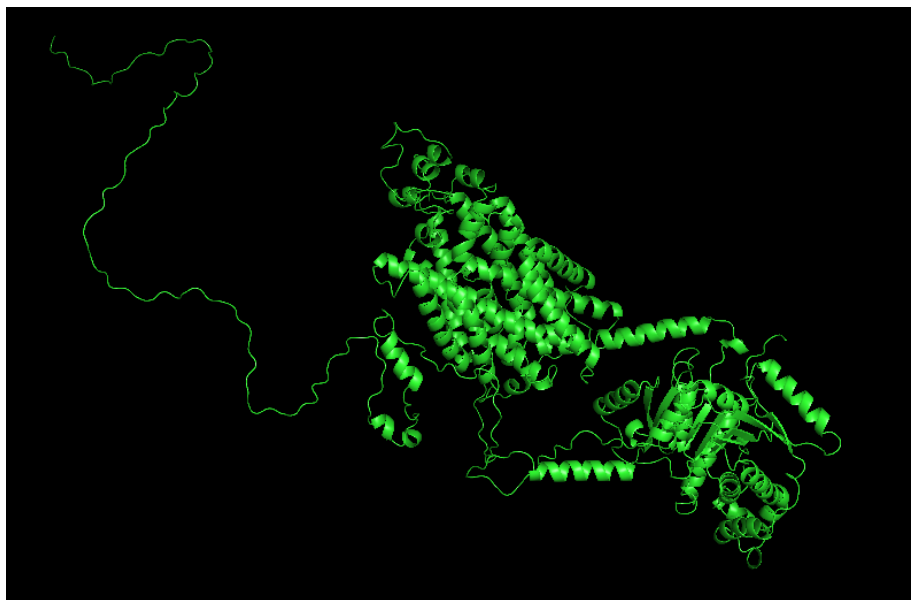


Figura 4.3: Monómero de AlphaFold para β NCC de anguila europea posterior a la tercera ronda de relajación.

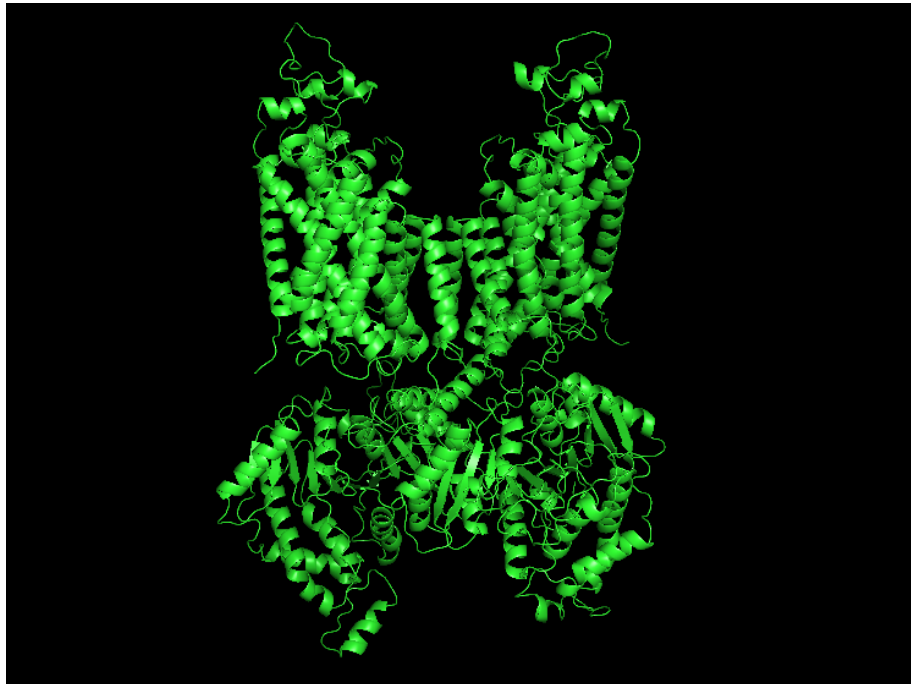


Figura 4.4: Dímero de SWISS-MODEL para β NCC de anguila europea, segunda ronda de relajación del modelo.

Para el dímero del β NCC, el modelo determinado como aquel de mayor calidad fue generado por SWISS-MODEL, específicamente el modelo 3 tras su segunda ronda de relajación, alcanzando un 98.02 % de residuos en regiones favorecidas del diagrama de Ramachandran. En la Figura 4.4 se presenta su visualización en ChimeraX:

4.2.2. NCC Humano

En la Tabla 4.4 se muestran los parámetros de evaluación obtenidos con MolProbity para los modelos seleccionados del NCC humano en su forma monomérica. Por cada software se seleccionó el modelo con los mejores puntajes. Las estructuras fueron seleccionadas por cumplir con los criterios de validación estructural establecidos.

Tabla 4.4: Parámetros de interés de los modelos de monómero seleccionados para NCC humano.

NCC Humano Monómeros	Ramachandran Outliers ($<0.05\%$)	Ramachandran Favored ($>96\%$)	Ramachandran Z- Score ($\text{abs}(\text{ZScore}) < 2$)	“-+”
AF-model15-relaxed2	0.00	98.63	0.10	0.25
SM-model2-relaxed2	0.31	96.59	-0.05	0.26
MD-model4-relaxed3	0.29	97.45	-0.09	0.24
RF-model04-relaxed3	0.20	97.64	-0.10	0.26

Se utilizaron los mismos criterios para determinar los modelos diméricos de mayor calidad, tomando uno por cada software. Los parámetros de ellos se despliegan en la Tabla 4.5:

Tabla 4.5: Parámetros de interés de los modelos de dímero seleccionados para NCC humano.

NCC Humano Dímeros	Ramachandran Outliers ($<0.05\%$)	Ramachandran Favored ($>96\%$)	Ramachandran Z- Score ($\text{abs}(\text{ZScore}) < 2$)	“-+”
AF-model3-relaxed2	0.00	98.14	-0.09	0.18
SM-model4-relaxed3	0.11	98.07	-0.41	0.26
MD-model2-relaxed3	2.11	90.23	-2.59	0.17

El modelo monomérico del NCC humano con mayor proporción de residuos en regiones favorecidas fue generado por AlphaFold tras la segunda ronda de relajación, alcanzando un 98.63 %. Su visualización se realizó en ChimeraX y se despliega en la Figura 4.5:

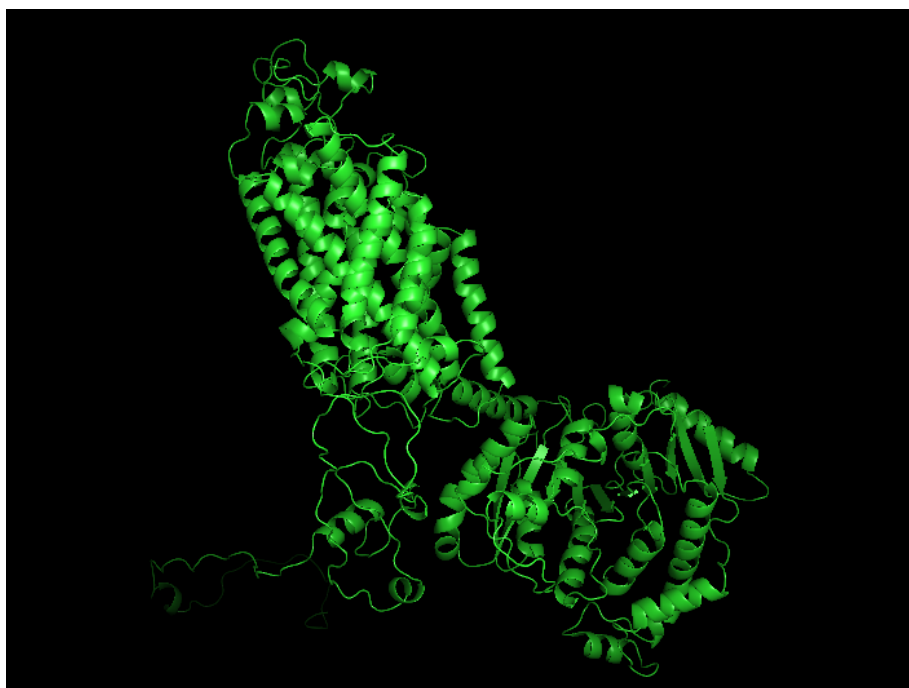


Figura 4.5: Monómero de AlphaFold para NCC Humano posterior a la segunda ronda de relajación.

El dímero seleccionado para el NCC humano fue generado por AlphaFold posterior a la segunda ronda de relajación, con un porcentaje de residuos favorecidos de 98.14 %. En la Figura

4.6 se visualiza la captura de pantalla de este modelo en ChimeraX:

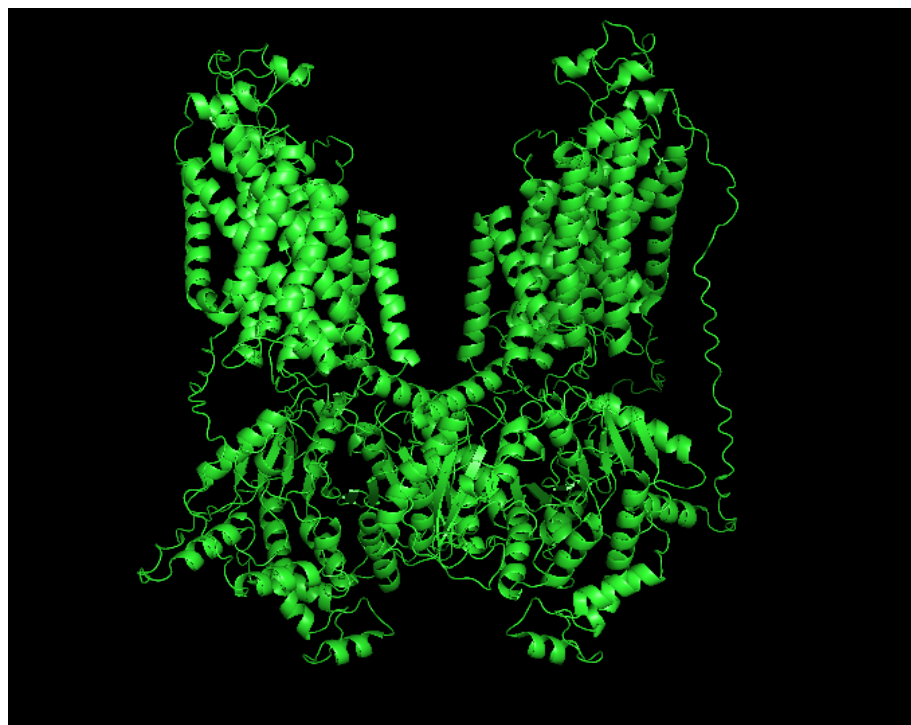


Figura 4.6: Dímero de AlphaFold para NCC Humano posterior a la segunda ronda de relajación.

De los cuatro modelos seleccionados con las características estereoquímicas más favorables —considerando ambas especies y formas oligoméricas— tres fueron generados mediante AlphaFold, lo cual refleja una alta consistencia de este método bajo los criterios aplicados en este estudio.

Resumen de calidad estructural

El análisis de calidad estructural mediante MolProbity permitió evaluar la geometría estereoquímica de los modelos generados por cada herramienta.

En términos generales, los modelos generados mediante AlphaFold presentaron la mayor proporción de residuos en regiones favorecidas y el menor número de outliers, tanto en el NCC humano como en el de anguila, particularmente en las formas monoméricas. Estos resultados

indican una mejor coherencia geométrica en los ángulos Phi y psi del esqueleto peptídico, lo cual se traduce en una menor probabilidad de tensiones conformacionales artificiales dentro del modelo.

El desempeño consistente de AlphaFold puede atribuirse a su arquitectura basada en aprendizaje profundo, la cual incorpora información evolutiva y restricciones espaciales implícitas derivadas de patrones estructurales aprendidos a partir de grandes bases de datos experimentales. En contraste, los modelos generados por métodos de modelado comparativo tradicional dependen en mayor medida de la calidad y similitud de las plantillas disponibles, lo que puede introducir desviaciones locales cuando la identidad de la secuencia de la plantilla es limitada.

Es relevante señalar que, tras los ciclos iterativos de refinamiento y relajación estructural, se observó una mejora en las métricas geométricas en la mayoría de los modelos, lo que confirma la utilidad del protocolo de validación empleado. Este proceso permitió minimizar conflictos estéricos y optimizar la orientación de cadenas laterales, incrementando la confiabilidad de las estructuras finales seleccionadas para análisis posteriores.

En conjunto, la validación estereoquímica respalda la selección de los modelos con mayor calidad geométrica como candidatos para el análisis de cavidades y estudios de interacción proteína–ligando, estableciendo una base estructural sólida para las siguientes etapas del estudio.

4.3. Comparación de estructuras

En el apartado previo se determinaron los modelos con mejores características estereoquímicas para cada proteína y conformación. A continuación, se realiza una comparación estructural entre dichos modelos utilizando la herramienta MatchMaker del software UCSF ChimeraX.

MatchMaker permite alinear cadenas proteicas con base en coincidencia de estructura secundaria y alineamiento de secuencia (Needleman-Wunsch con matriz BLOSUM-62), evaluando la similitud estructural mediante la desviación cuadrática media (RMSD), tanto local (pares bien alineados) como global (todos los residuos).

4.3.1. β NCC de Anguila — Monómeros

El modelo monomérico de referencia para β NCC de anguila fue el generado por AlphaFold. Los modelos obtenidos por los demás métodos fueron comparados con este utilizando MatchMaker. En la Figura 4.7 se presenta el resultado del MatchMaker para los monómeros de Beta NCC:

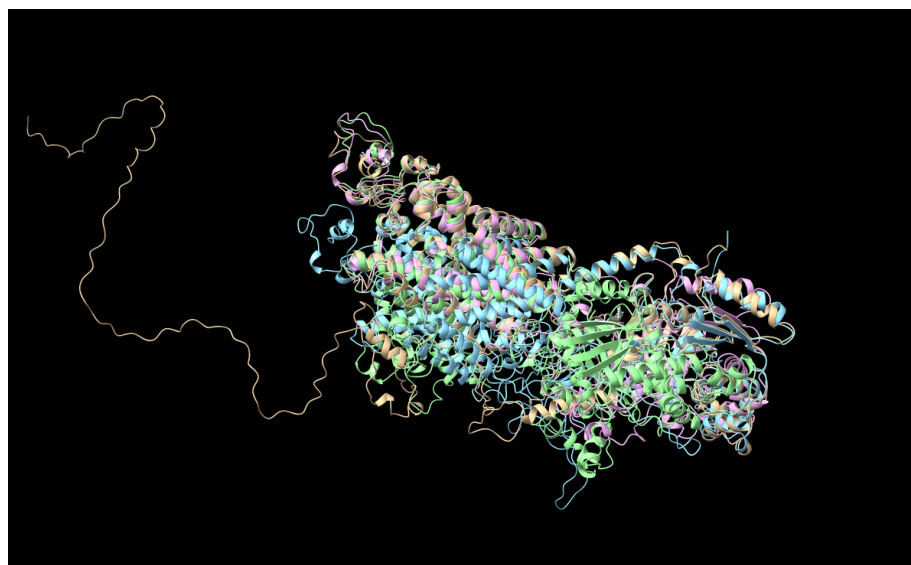


Figura 4.7: MatchMaker de monómeros para β NCC de anguila. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (verde), Rosetta (azul).

Tabla 4.6: Comparación estructural entre modelos monoméricos de β NCC de anguila utilizando MatchMaker de ChimeraX.

Comparación	RMSD (pruned) [Å]	RMSD (total) [Å]	Coincidencia local	Coincidencia global
AF vs Modeller	0.862	39.67	Excelente	Baja
AF vs SWISS-MODEL	1.029	10.70	Muy buena	Buena
AF vs Rosetta	1.182	32.46	Aceptable	Baja

En la Tabla 4.6 se observaron las relaciones presentadas entre las estructuras con el modelo base (generado por AlphaFold). SWISS-MODEL mostró la mayor similitud estructural global con AlphaFold (RMSD total de 10.70 Å, RMSD local de 1.03 Å). Modeller presentó una coin-

cidencia local excelente (0.86 Å) pero con una desviación global alta (39.67 Å), mientras que Rosetta tuvo una coincidencia moderada.

4.3.2. NCC Humano — Monómeros

El modelo monomérico de referencia para el NCC humano fue también el de AlphaFold. En la Figura 4.8 se tiene la captura de pantalla del MatchMaker realizado en ChimeraX:

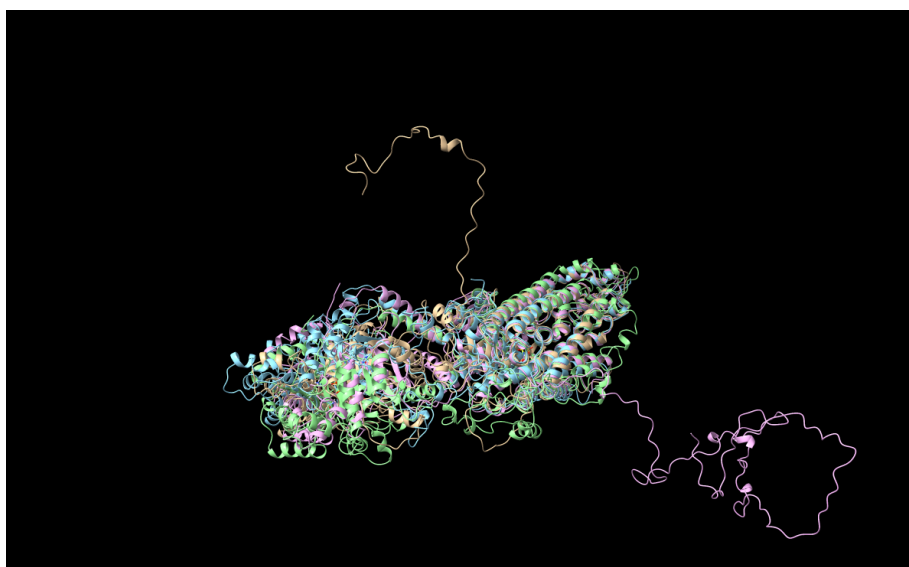


Figura 4.8: MatchMaker de monómeros de NCC humano. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (verde), Rosetta (azul).

Tabla 4.7: Comparación estructural entre modelos monoméricos del NCC humano utilizando MatchMaker de ChimeraX.

Comparación	RMSD (pruned) [Å]	RMSD (total) [Å]	Coincidencia local	Coincidencia global
AF vs SWISS-MODEL	0.908	8.879	Excelente	Buena
AF vs Modeller	1.022	38.407	Muy buena	Baja
AF vs Rosetta	1.200	25.730	Aceptable	Baja

En la Tabla 4.7 se presentaron los parámetros de comparación estructural de los modelos generados por cada software en relación al modelo que se estableció como base (generado

por AlphaFold). SWISS-MODEL mostró la mayor similitud estructural con AlphaFold, con un RMSD global de 8.88 Å y local de 0.91 Å. Modeller conservó el núcleo estructural (1.02 Å) pero divergió globalmente. Rosetta presentó la mayor variabilidad.

4.3.3. β NCC de Anguila — Dímeros

El modelo de referencia fue el generado por SWISS-MODEL. En la Figura 4.9 se presenta el MatchMaker de los demás modelos con aquel creado por SWISS-MODEL:



Figura 4.9: MatchMaker de dímeros para β NCC de anguila. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (azul).

Tabla 4.8: Comparación estructural entre modelos diméricos de β NCC de anguila utilizando MatchMaker de ChimeraX.

Comparación	RMSD (pruned) [Å]	RMSD (total) [Å]	Coincidencia local	Coincidencia global
SM vs AlphaFold	1.093	10.63	Muy buena	Buena
SM vs Modeller	1.235	10.03	Aceptable	Buena

En la Tabla 4.8 se presentaron los parámetros de comparación estructural entre modelos diméricos con respecto al generado por SWISS-MODEL. Ambos modelos alternativos mos-

traron buena conservación estructural con respecto al modelo de SWISS-MODEL. AlphaFold conservó mejor el núcleo estructural (RMSD local menor), mientras que Modeller se aproximó más en el total de la estructura (RMSD global menor).

4.3.4. NCC Humano — Dímeros

El modelo de referencia fue el generado por AlphaFold. En la Figura 4.10 se presenta el MatchMaker con las demás estructuras:

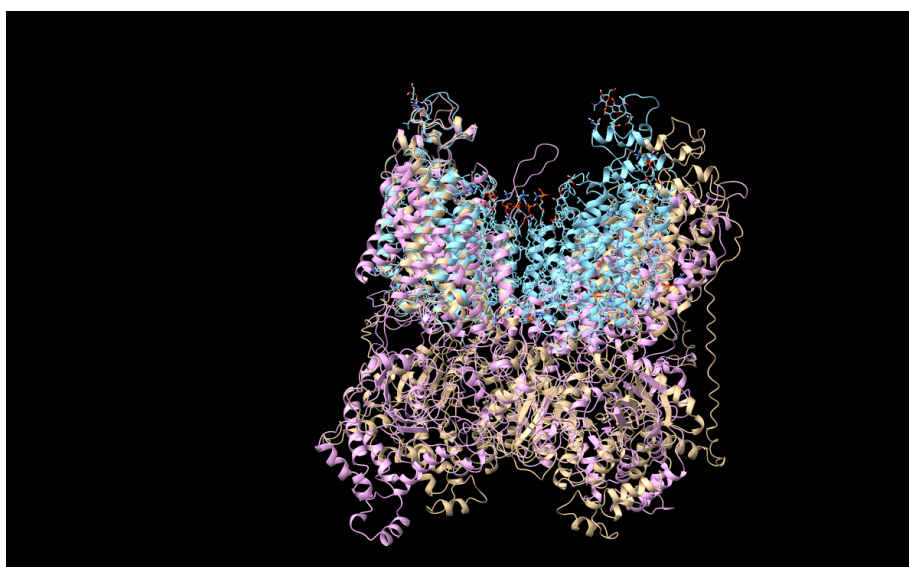


Figura 4.10: MatchMaker de dímeros para NCC humano. AlphaFold (amarillo), SWISS-MODEL (rosado), MODELLER (azul).

Tabla 4.9: Comparación estructural entre modelos diméricos del NCC humano utilizando MatchMaker de ChimeraX.

Comparación	RMSD (pruned) [Å]	RMSD (total) [Å]	Coincidencia local	Coincidencia global
AF vs SWISS-MODEL	0.789	1.150	Excelente	Excelente
AF vs Modeller	1.278	23.309	Aceptable	Baja

A partir de la Tabla 4.9, donde se presenta la comparación estructural, se observa que SWISS-MODEL fue el modelo más similar a AlphaFold, con un RMSD global y local muy

bajo, indicando una coincidencia casi completa entre ambas estructuras. Modeller, por su parte, mostró desviaciones significativas a nivel global.

Resumen de comparación estructural

La superposición estructural entre los modelos generados mediante AlphaFold, Rosetta, SWISS-MODEL y Modeller reveló diferencias conformacionales apreciables en la comparación global de las estructuras. Los valores de RMSD obtenidos oscilaron entre 20 y 40 Å en algunos casos, particularmente en comparaciones entre modelos derivados de enfoques metodológicos distintos.

Si bien estos valores podrían interpretarse como discrepancias estructurales significativas, es importante considerar que el NCC es una proteína de membrana con dominios citosólicos extensos y regiones potencialmente flexibles. En este contexto, pequeñas variaciones en la orientación relativa de dominios periféricos pueden amplificarse en el cálculo global del RMSD, incrementando el valor total sin necesariamente comprometer la arquitectura del núcleo transmembrana.

Asimismo, las diferencias metodológicas entre herramientas —como el uso de inteligencia artificial frente a modelado comparativo basado en plantillas— pueden favorecer conformaciones alternativas en regiones con menor restricción estructural. En consecuencia, los valores elevados de RMSD global deben interpretarse con cautela y no necesariamente como evidencia de inconsistencia en el núcleo funcional del cotransportador.

En la Tabla 4.10 se presentan todos los parámetros obtenidos como resultado de la aplicación de la herramienta MatchMaker para ambas especies y conformaciones.

Tabla 4.10: Resumen de comparación estructural por RMSD entre modelos seleccionados y de referencia.

Proteína y Estado	Comparación	RMSD (pruned) [Å]	RMSD (total) [Å]	Similitud Global
β NCC Anguila Monómero	AF vs Modeller	0.862	39.67	Baja
	AF vs SWISS-MODEL	1.029	10.70	Buena
	AF vs Rosetta	1.182	32.46	Baja
NCC Humano Monómero	AF vs SWISS-MODEL	0.908	8.88	Buena
	AF vs Modeller	1.022	38.41	Baja
	AF vs Rosetta	1.200	25.73	Baja
β NCC Anguila Dímero	SM vs AlphaFold	1.093	10.63	Buena
	SM vs Modeller	1.235	10.03	Buena
NCC Humano Dímero	AF vs SWISS-MODEL	0.789	1.15	Excelente
	AF vs Modeller	1.278	23.31	Baja

4.4. Resultados de Cavidades

Tabla 4.11: Resumen comparativo de DrugScore por software, tipo de modelo y especie.

Modelo	Software	DrugScore Máx.	DrugScore Prom.	Druggability Prom.	Strong Cavities
Monómero Anguila	Rosetta (RF)	6012.0	1867.33	2.78	7
	SWISS-MODEL (SM)	3212.0	849.71	2.43	3
	Modeller (MD)	1840.0	651.36	2.43	6
	AlphaFold (AF)	3627.0	915.50	2.40	4
Monómero Humano	AlphaFold (AF)	2876.0	1343.78	2.78	7
	SWISS-MODEL (SM)	3385.0	1126.38	2.63	5
	Modeller (MD)	2623.0	1070.85	2.54	7
	Rosetta (RB)	3796.0	896.43	2.50	7
Dímero Anguila	Modeller (MD)	13260.0	1712.00	2.69	11
	SWISS-MODEL (SM)	3956.0	1383.00	2.67	4
	AlphaFold (AF)	6445.0	1359.32	2.37	7
Dímero Humano	SWISS-MODEL (SM)	12300.0	2044.22	2.56	5
	AlphaFold (AF)	8383.0	1424.67	2.57	12
	Modeller (MD)	7252.0	1217.43	2.49	17

El análisis de cavidades mediante CavityPlus, desplegado en la Tabla 4.11, reveló diferencias cuantitativas entre los modelos generados por cada herramienta de predicción estructural, tanto en monómeros como en dímeros y en ambas especies estudiadas.

En modelos monoméricos, Rosetta y AlphaFold mostraron valores elevados de DrugScore promedio y un mayor número de cavidades clasificadas como “strong”, lo que sugiere la presencia de bolsillos bien definidos geométricamente. Sin embargo, esta tendencia no fue uniforme en todas las configuraciones, ya que SWISS-MODEL y Modeller presentaron valores comparables en ciertos casos, especialmente en el NCC humano.

En modelos diméricos, Modeller exhibió los valores máximos de DrugScore, particularmente en el β NCC de anguila. Este comportamiento podría asociarse a una mayor irregularidad superficial o a la generación de cavidades intersubunitarias más amplias tras el refinamiento estructural. No obstante, el hecho de que distintos softwares identifiquen cavidades en regiones espacialmente coincidentes —con superposiciones volumétricas superiores al 85 % en los bolsillos centrales— sugiere que los sitios principales no son artefactos metodológicos, sino características estructurales consistentes.

Es relevante destacar que la calidad estereoquímica no se correlacionó linealmente con el número de cavidades detectadas. Algunos modelos con parámetros geométricos moderados generaron mayor cantidad de cavidades, lo que indica que la detección de bolsillos depende también de la topología superficial y dominios transmembrana.

En conjunto, los resultados muestran que, aunque existen variaciones cuantitativas entre métodos, los principales sitios de unión identificados presentan estabilidad estructural y consistencia intermetodológica, lo que respalda su potencial relevancia funcional. Estos resultados respaldan la robustez del flujo de trabajo implementado, ya que los principales determinantes estructurales asociados a posibles sitios de unión fueron identificados de manera consistente a través de diferentes enfoques de modelado. Así, aunque las métricas cuantitativas específicas pueden variar entre softwares, la convergencia en la localización de cavidades clave refuerza la confiabilidad del análisis estructural realizado.

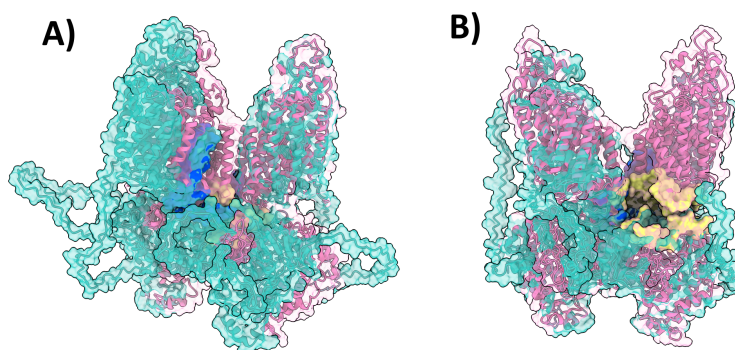


Figura 4.11: Superposición de los modelos diméricos e identificación de cavidades. A) Beta NCC de Anguila y B) NCC de Humano. Se comparan los modelos de AlphaFold (azul) y SWISS-MODEL (rosado). Elaboración propia, publicada previamente en [53, Fig. 4].

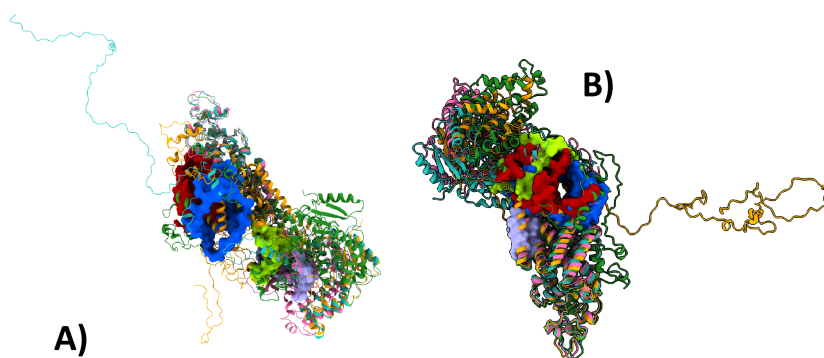


Figura 4.12: Superposición de los modelos monoméricos e identificación de las cavidades con mayor valor de druggability. A) Beta NCC de Anguila y B) NCC de humano. AlphaFold (azul), Modeller (lila), RoseTTaFold (naranja) y SWISS-MODEL (verde). Elaboración propia, publicada previamente en [53, Fig. 3].

Las figuras 4.11 y 4.12 fueron elaboradas como parte de esta investigación y publicadas previamente en [53].

En las Figuras 4.11 y 4.12 se puede apreciar la convergencia de las cavidades principales entre método y especie, permitiendo concluir que estas se conservan.

Adicionalmente, se realizaron las siguientes observaciones:

- El conteo total de cavidades con DrugScore ≥ 0.5 varió entre 3 y 7 según el software, sugiriendo múltiples sitios de posible interacción con ligandos.
- Se observó una superposición volumétrica superior al 85 % en los bolsillos centrales al comparar modelos generados por distintas plataformas, lo que respalda la estabilidad estructural de dichas regiones.
- Las variaciones observadas en volumen y SASA fueron moderadas ($\pm 10\%$), atribuibles a diferencias menores en la orientación de bucles o segmentos helicoidales.

Resumen de los resultados en función con las preguntas base planteadas en la sección de objetivos:

1. ¿Cuál software proporciona el mejor modelo base?

En términos estereoquímicos, Rosetta generó los modelos monoméricos más robustos, alcanzando más del 96 % de residuos en regiones favorecidas en la métrica Ramachandran. Para modelos diméricos, SWISS-MODEL mostró mayor estabilidad estructural y consistencia, mientras que Modeller presentó alta variabilidad. AlphaFold mantuvo un rendimiento estable, sin liderar, pero siempre dentro de parámetros aceptables.

2. ¿Cuál software se beneficia más de las relajaciones?

AlphaFold fue el que más mejoró con las rondas de relajación, alcanzando valores superiores al 97.5 % en Ramachandran Favored en todos los casos. SWISS-MODEL también mostró mejoras importantes, especialmente en modelos diméricos. Aunque Rosetta produjo los mejores modelos base, su mejora tras relajación fue limitada en comparación con AlphaFold.

3. **¿Qué características presentan las cavidades?**

En monómeros, RosettaFold y AlphaFold generaron modelos con cavidades bien definidas y valores destacados en DrugScore. En dímeros, Modeller y SWISS-MODEL sobresalieron.

4. **¿Cuál es la forma y volumen de las cavidades?**

Los bolsillos principales fueron caracterizados con consistencia entre métodos, mostrando un solapamiento superior al 85 % y variaciones menores entre plataformas. Esto confirma la fiabilidad del análisis estructural de cavidades en este contexto.

Capítulo 5

Conclusiones

El presente trabajo demostró que la integración de herramientas de predicción estructural basadas en inteligencia artificial, modelado comparativo, refinamiento iterativo y análisis de cavidades constituye un pipeline robusto para el estudio estructural del cotransportador Na^+/Cl^- (NCC) en sus formas monomérica y dimérica.

Los principales hallazgos se pueden resumir en los siguientes puntos:

- Se generaron modelos tridimensionales con alta calidad estereoquímica, alcanzando porcentajes de residuos en regiones favorecidas del diagrama de Ramachandran superiores al 96 % tras el refinamiento estructural iterativo.
- El protocolo de relajación mostró un efecto cuantificable en la mejora geométrica de los modelos, particularmente en estructuras diméricas generadas por métodos comparativos tradicionales, evidenciando convergencia estructural sin alterar significativamente la arquitectura global.
- El análisis de cavidades permitió identificar de manera consistente entre métodos sitios estructurales con altos valores de drugabilidad.
- Si bien los valores cuantitativos de DrugScore y número de cavidades variaron entre softwares, la localización espacial de los principales bolsillos mostró consistencia intermetodológica, lo que respalda la robustez del enfoque empleado.

- La caracterización geométrica de estos sitios —en términos de volumen, área superficial y puntuación de druggability— sugiere un potencial estructural para interacción con ligandos, proporcionando bases para estudios posteriores de acoplamiento molecular y dinámica.

En conjunto, la metodología desarrollada establece un marco computacional reproducible e independiente del software de predicción específico, fortaleciendo la viabilidad del modelado estructural del NCC como herramienta para estudios funcionales y futuras estrategias de diseño racional de moduladores farmacológicos dirigidos a proteínas de membrana.

5.1. Limitaciones y Trabajo Futuro

A pesar de los resultados obtenidos, es importante reconocer ciertas limitaciones inherentes al enfoque computacional empleado. En primer lugar, las estructuras generadas corresponden a predicciones estáticas y no capturan completamente la dinámica conformacional del NCC, particularmente relevante en proteínas transportadoras que alternan entre distintos estados funcionales.

Asimismo, aunque el análisis de cavidades proporciona estimaciones geométricas y energéticas preliminares sobre el potencial de interacción con ligandos, estos resultados no sustituyen a estudios de acoplamiento molecular, simulaciones de dinámica molecular o validación experimental directa. La clasificación de cavidades como “druggable” se basa en parámetros computacionales que requieren corroboración adicional en contextos bioquímicos y farmacológicos.

Como trabajo futuro, se propone:

- Realizar estudios de acoplamiento molecular y dinámica molecular para evaluar la estabilidad de los complejos proteína–ligando en los sitios identificados.
- Explorar la predicción de estados conformacionales alternativos del NCC que puedan revelar sitios transitorios o alostéricos adicionales.
- Comparar las predicciones estructurales con futuros datos experimentales disponibles para validar y refinar el pipeline computacional.

En este contexto, el pipeline desarrollado constituye una base metodológica sólida que puede ampliarse e integrarse con aproximaciones experimentales para profundizar en la caracterización estructural y funcional del NCC.

Bibliografía

- [1] C. Zhou, Y. Zheng, and Y. Zhou, “Structure prediction of membrane proteins,” *Genomics, Proteomics & Bioinformatics*, vol. 2, no. 1, pp. 1–5, 2004. 2
- [2] R. Bajaj, “Hurdles and advancements in experimental membrane protein structural biology,” *Journal of Structural Biology*, vol. 217, no. 4, p. 108251, 2025. [Online]. Available: <https://doi.org/10.1016/j.jsb.2025.108251> 2
- [3] M. Q. C. Li, S. Wang, S.-R. Lin, L. E. N. Ting, Z.-H. Wan, G. Xie, and J. Zhang, “Advantages and Limitations of AlphaFold in Structural Biology: Insights from Recent Studies,” *The Protein Journal*, vol. 45, no. 1, pp. 22–38, 2026. [Online]. Available: <https://doi.org/10.1007/s10930-025-10310-8> 2, 3, 19
- [4] J. Nan, Y. Yuan, X. Yang, Z. Shan, H. Liu, F. Wei, W. Zhang, and Y. Zhang, “Cryo-EM structure of the human sodium-chloride cotransporter NCC,” *Science Advances*, vol. 8, p. eadd7176, 2022. 2, 3, 5, 6
- [5] G. Gamba, “The thiazide-sensitive Na⁺-Cl⁻ cotransporter: molecular biology, functional properties, and regulation,” *American Journal of Physiology-Renal Physiology*, vol. 297, no. 4, pp. F838–F848, 2009. [Online]. Available: <https://doi.org/10.1152/ajprenal.00159.2009> 2, 4, 5, 6
- [6] D. H. Ellison and J. Loffing, “Thiazide effects and adverse effects: Insights from molecular genetics,” *Hypertension*, vol. 54, no. 2, pp. 196–202, 2009. [Online]. Available: <https://doi.org/10.1161/HYPERTENSIONAHA.109.129171> 2, 4, 5, 6, 15, 16

- [7] D. B. Simon, C. Nelson-Williams, M. J. Bia, D. Ellison, F. E. Karet, A. M. Molina *et al.*, “Gitelman’s variant of Bartter’s syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter,” *Nature Genetics*, vol. 12, no. 1, pp. 24–30, 1996. 2
- [8] C. P. Cutler and G. Cramb, “Molecular physiology of osmoregulation in eels,” *Journal of Experimental Biology*, vol. 210, pp. 2369–2378, 2007. 2
- [9] A. Kryshchak, T. Schwede, M. Topf, K. Fidelis, and J. Moult, “Critical assessment of methods of protein structure prediction (CASP14): summary and results,” *Proteins: Structure, Function, and Bioinformatics*, vol. 89, no. 12, pp. 1607–1617, 2021. 3, 5, 6
- [10] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede, “Swiss-model: homology modelling of protein structures and complexes,” *Nucleic Acids Research*, vol. 46, no. W1, pp. W296–W303, 2018. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6030848/> 3, 19, 33, 36
- [11] A. Šali and T. L. Blundell, “Comparative protein modelling by satisfaction of spatial restraints,” *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993. 3, 21
- [12] A. Leaver-Fay *et al.*, “Rosetta3: an object-oriented software suite for the simulation and design of macromolecules,” *Methods in Enzymology*, vol. 487, pp. 545–574, 2011. 3, 20, 33, 35
- [13] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, 2021. 3, 5, 6, 18, 22, 33
- [14] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral *et al.*, “MolProbity: all-atom structure validation for macromolecular crystallography,” *Acta Crystallographica Section D*, vol. 66, no. 1, pp. 12–21, 2010. 3

- [15] E. Moreno, D. Pacheco Alvarez, and E. Ríos Argaiz, “Aspectos estructurales, funcionales y patológicos del cotransportador de nacl sensible a tiazidas,” *Revista de Investigación Clínica*, vol. 66, no. 6, pp. 559–567, Nov. 2014, publicado en noviembre-diciembre de 2014. 4, 13, 14, 15
- [16] Instituto Nacional de Salud Pública, “Encuesta nacional de salud y nutrición 2021 sobre COVID-19: Resultados nacionales,” Instituto Nacional de Salud Pública, Cuernavaca, México, Tech. Rep., 2022. [Online]. Available: https://ensanut.insp.mx/encuestas/ensanutcontinua2021/doctos/informes/220804_Ensa21_digital_4ago.pdf 4
- [17] Instituto Nacional de Estadística y Geografía, “Estadísticas de defunciones registradas 2021,” Instituto Nacional de Estadística y Geografía, Aguascalientes, México, Tech. Rep., 2022. [Online]. Available: https://www.inegi.org.mx/contenidos/programas/mortalidad/doc/defunciones_registradas_2021_nota_tecnica.pdf 4
- [18] R. M. Bill, P. J. F. Henderson, S. Iwata, E. R. S. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, and H. Vogel, “Overcoming barriers to membrane protein structure determination,” *Nature Biotechnology*, vol. 29, no. 4, pp. 335–340, Apr. 2011, published April 2011. 6, 11
- [19] Z. Cournia, T. W. Allen, I. Andricioaei, B. Antonny, D. Baum, G. Brannigan, N. Buchete, J. T. Deckman, L. Delemotte, C. Del Val, R. Friedman, P. Gkeka, H. Hege, J. Hénin, M. A. Kasimova, A. Kolocouris, M. L. Klein, S. Khalid, M. J. Lemieux, N. Lindow, M. Roy, J. Selent, M. Tarek, F. Tofoleanu, S. Vanni, S. Urban, D. J. Wales, J. C. Smith, and A. Bondar, “Membrane protein structure, function, and dynamics: a perspective from experiments and theory,” *Journal of Membrane Biology*, vol. 248, no. 4, pp. 611–640, Aug. 2015, epub 11 June 2015. 6, 10
- [20] N. Palomar-Alonso, M. Lee, and M. Kim, “Exosomes: Membrane-associated proteins, challenges and perspectives,” *Biochemistry and Biophysics Reports*, vol. 37, p. 101599, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405580823001802> 9

- [21] J. G. L. Mullins, “Computational approaches for membrane protein structure prediction,” in *Membrane Proteins—Production and Functional Characterization*, ser. Methods in Enzymology, A. K. Shukla, Ed. Elsevier, 2015, vol. 556, pp. 85–104. 9, 12, 17, 18
- [22] T. Aristotelous, A. L. Hopkins, and I. Navratilova, “Surface plasmon resonance analysis of seven-transmembrane receptors,” in *Membrane Proteins—Production and Functional Characterization*, ser. Methods in Enzymology, A. K. Shukla, Ed. Elsevier, 2015, vol. 556, pp. 499–525. 10, 11
- [23] Coskun and K. Simons, “Cell membranes: the lipid perspective,” *Structure*, vol. 19, no. 11, pp. 1543–1548, Nov. 2011, published online 9 November 2011. 10, 11, 12
- [24] E. Wallin and G. von Heijne, “Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms,” *Protein Science*, vol. 7, no. 4, pp. 1029–1038, 1998. 10
- [25] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. Garland Science, 2015. 11
- [26] Y. Cheng, “Single-particle cryo-em—how did it get here and where will it go,” *Science*, vol. 361, no. 6405, pp. 876–880, 2018. 12
- [27] A. R. Subramanya and D. H. Ellison, “Distal convoluted tubule,” *Clinical Journal of the American Society of Nephrology*, vol. 9, no. 12, pp. 2147–2163, 2014. 13, 14, 15
- [28] S. B. Beckmann, C. F. Rudolphi, P. H. Imenez Silva, R. A. Fenton, and E. J. Hoorn, “Thiazide-induced hyponatremia,” *Best Practice & Research Clinical Endocrinology & Metabolism*, vol. 40, no. 1, p. 102063, 2026. [Online]. Available: <https://doi.org/10.1016/j.beem.2025.102063> 13, 16
- [29] M. Zuin *et al.*, “Advances in hypertension management: Insights from the latest european guidelines,” *Journal of Cardiovascular Development and Disease*, vol. 12, no. 4, p. 155, 2025. 15

- [30] G. Gamba, “Molecular physiology and pathophysiology of electroneutral cation-chloride cotransporters,” *Physiological Reviews*, vol. 85, no. 2, pp. 423–493, 2005. [Online]. Available: <https://doi.org/10.1152/physrev.00011.2004> 15
- [31] ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group, “Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic,” *JAMA*, vol. 288, no. 23, pp. 2981–2997, 2002. 16
- [32] G. C. Roush and D. A. Sica, “Diuretics for hypertension: A review and update,” *American Journal of Hypertension*, vol. 29, no. 10, pp. 1130–1137, 2016. [Online]. Available: <https://doi.org/10.1093/ajh/hpw030> 16
- [33] Y. Zhao, H. Schubert, A. Blakely, B. Forbush, M. D. Smith, J. Rinehart, and E. Cao, “Structural bases for Na^+/Cl^- cotransporter inhibition by thiazide diuretic drugs and activation by kinases,” *Nature Communications*, vol. 15, p. 7006, 2024. 17, 31
- [34] C. Chothia and A. M. Lesk, “The relation between the divergence of sequence and structure in proteins,” *The EMBO Journal*, vol. 5, no. 4, pp. 823–826, 1986. 20
- [35] J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth *et al.*, “Macromolecular modeling and design in rosetta: recent methods and frameworks,” *Nature Methods*, vol. 17, pp. 665–680, 2020. 20
- [36] B. Webb and A. Sali, “Comparative protein structure modeling using modeller,” *Current Protocols in Bioinformatics*, vol. 54, pp. 5.6.1–5.6.37, 2016, epub 2016; PMID: 27322406; PMCID: PMC5031415. 21, 33
- [37] M. Baek, F. DiMaio *et al.*, “Accurate prediction of protein structures and interactions using a three-track neural network,” *Science*, vol. 373, pp. 871–876, 2021. 22

- [38] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido *et al.*, “Evolutionary-scale prediction of atomic-level protein structure with a language model,” *Science*, vol. 379, pp. 1123–1130, 2023. 22
- [39] J. Abramson, J. Adler, J. Dunger, T. Green, A. Pritzel *et al.*, “Accurate structure prediction of biomolecular interactions with alphafold 3,” *Nature*, 2024. 22
- [40] D. del Alamo, D. Sala, H. S. Mchaourab, and J. Meiler, “Sampling alternative conformational states of transporters and receptors with alphafold2,” *eLife*, vol. 11, p. e75751, 2022. 22
- [41] T. Hegedűs, M. Geisler, G. L. Lukács, and B. Farkas, “Ins and outs of alphafold2 transmembrane protein structure predictions,” *Cellular and Molecular Life Sciences*, vol. 79, no. 1, p. 73, 2022. [Online]. Available: <https://doi.org/10.1007/s00018-021-04112-1> 22
- [42] M. A. Pak, K. A. Markhieva, M. S. Novikova, D. S. Petrov, I. S. Vorobyev, E. S. Maksimova, F. A. Kondrashov, and D. N. Ivankov, “Using alphafold to predict the impact of single mutations on protein stability and function,” *PLOS ONE*, vol. 18, no. 3, p. e0282689, 2023. [Online]. Available: <https://doi.org/10.1371/journal.pone.0282689> 22
- [43] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, and S. Zhao, “Applications of machine learning in drug discovery and development,” *Nature Reviews Drug Discovery*, vol. 18, no. 6, pp. 463–477, 2019. [Online]. Available: <https://doi.org/10.1038/s41573-019-0024-5> 22
- [44] D. F. Thieker, J. B. Maguire, S. T. Kudlacek, A. Leaver-Fay, S. Lyskov, and B. Kuhlman, “Stabilizing proteins, simplified: A rosetta-based webtool for predicting favorable mutations,” *Protein Science*, vol. 31, no. 10, p. e4428, 2022. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9490798/> 26
- [45] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson, “Molprobity: all-atom contacts and structure validation for proteins and nucleic acids,” *Nucleic Acids*

- Research*, vol. 35, no. Web Server issue, pp. W375–W383, 2007. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1933162/> 27, 39
- [46] S. Wang, J. Xie, J. Pei, and L. Lai, “Cavityplus 2022 update: An integrated platform for comprehensive protein cavity detection and property analyses with user-friendly tools and cavity databases,” *Journal of Molecular Biology*, vol. 435, no. 14, p. 168141, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S002228362300219X> 27, 28
- [47] E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, and T. E. Ferrin, “Ucsf chimeraX: Tools for structure building and analysis,” *Protein Science*, vol. 32, no. 11, p. e4792, 2023. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10588335/> 27
- [48] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant, “Pubchem substance and compound databases,” *Nucleic Acids Research*, vol. 44, no. D1, pp. D1202–D1213, 2016. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4702940/> 28
- [49] The UniProt Consortium, “Uniprot: the universal protein knowledgebase in 2025,” *Nucleic Acids Research*, vol. 53, no. D1, pp. D609–D617, 2025. [Online]. Available: <https://academic.oup.com/nar/article/53/D1/D609/7902999> VIII, 29, 32
- [50] M. Fan, J. Zhang, C.-L. Lee, J. Zhang, and L. Feng, “Structure and thiazide inhibition mechanism of the human Na–Cl cotransporter,” *Nature*, vol. 614, no. 7949, pp. 788–793, 2023. [Online]. Available: <https://doi.org/10.1038/s41586-023-05718-0> 31
- [51] H.-B. Guo, A. Perminov, S. Bekele, G. Kedziora, S. Farajollahi, V. Varaljay, K. Hinkle, V. Molinero, K. Meister, C. Hung, P. Dennis, N. Kelley-Loughnane, and B. Rajiv, “AlphaFold2 models indicate that protein sequence determines both structure and dynamics,” *Scientific Reports*, vol. 12, no. 1, p. 10696, 2022. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9226352/> 34

- [52] Y. Xu, S. Wang, Q. Hu, S. Gao, X. Ma, W. Zhang, Y. Shen, F. Chen, L. Lai, and J. Pei, “Cavityplus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction,” *Nucleic Acids Research*, vol. 46, no. W1, pp. W374–W379, Jul. 2018, epub July 2 2018; PMID: 29750256; PMCID: PMC6031032. 41
- [53] J. R. Sandria Sánchez, G. Rodríguez Cárdenas, C. B. Gotchev Chtereveva, E. Moreno, and L. Jimenez-Angeles, “A comparative pipeline for computational protein modeling: A guide for biomedical engineering students,” in *XLVIII Mexican Conference on Biomedical Engineering*, ser. IFMBE Proceedings, vol. 137. Cham, Switzerland: Springer, 2026, pp. 184–192. [Online]. Available: https://doi.org/10.1007/978-3-032-13729-6_20 IX, 43, 61, 62