



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

**FACULTAD DE INGENIERÍA**

**Simulación In Silico de la  
función del cotransportador de  
Na<sup>+</sup>:Cl<sup>-</sup> sensible a tiazidas  
(NCC)**

**TESIS**

Que para obtener el título de

**Ingeniero en Sistemas Biomédicos**

**P R E S E N T A**

Josue David Díaz Gutiérrez

**DIRECTOR DE TESIS**

Dr. Luis Jiménez Ángeles



**Ciudad Universitaria, Cd. Mx., 2025**



**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 IN SILICO DE LA FUNCION DEL COTRANSPORTADOR DE  $Na^+Cl^-$  SENSIBLE A TIAZIDAS (NCC) 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.

---

**JOSUE DAVID DIAZ GUTIERREZ**  
Número de cuenta: 318177461

---

## ***Agradecimientos***

Agradezco a la convocatoria Alianza 2024 DGTIC-Huawei por brindarme acceso a los servidores y cursos que fueron fundamentales para la realización de este trabajo, así como por el apoyo económico otorgado a lo largo del proceso.

Asimismo, expreso mi más amplio agradecimiento a la Dra. Reyna Erika Moreno Martínez y a la Dra. Eréndira Rojas Ortega, adscritas al Departamento de Nefrología y Metabolismo Mineral, del Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán”; quienes son las investigadoras principales del proyecto “Relación Estructura-Función del Cotransportador Renal Na-Cl sensible a tiazidas.

Reconocemos la capacitación avanzada y los recursos informáticos de alto rendimiento brindados por la DGTIC-UNAM (Alianza UNAM-Huawei 2025, Proyecto ID-05) y el Laboratorio Nacional de Cómputo de Alto Desempeño, UAM-Iztapalapa (LANCAD, Proyecto ID-15-2025).

Mi más sincero agradecimiento al Dr. Luis Jiménez Ángeles, mi director de tesis y guía a lo largo de la carrera. Siempre estaré agradecido por sus enseñanzas y consejos, los cuales contribuyeron enormemente a mi formación como ingeniero. También quiero mencionar a la Dra. Michelin Álvarez Camacho, quien me apoyó y me guió durante la mitad de mi carrera.

Quiero expresar mi gratitud al Dr. Jesus Manuel Alvarez Lopez, quien me aceptó y compartió su conocimiento en el Instituto de Fisiología. Gracias a él, pude vivir experiencias enriquecedoras, como congresos en diversos institutos.

## ***Dedicatoria***

Dedico este trabajo , en primer lugar, a mi familia y a Dios por brindarme la oportunidad de realizar mis estudios sin contratiempos. Mi más profundo reconocimiento va a mis padres, Erika y Miguel, y a mis hermanos, Adriana, Miguel y Emmanuel, quienes siempre han sido un pilar fundamental durante mi formación académica, profesional y personal. Siempre estuvieron ahí con un consejo oportuno en cada etapa de este proceso, el cual culmina con este trabajo. También no quiero dejar fuera de mi familia a mis mascotas que me acompañaron siempre aunque el

---

sueño les pudiera ganar, Malenia, Schon y Hada, son también parte de esto.

También quiero mencionar a todos mis compañeros de la universidad y de la generación de ISB. En particular, agradezco mucho a Eduardo, Luis, Darien, Gilberto y Héctor, quienes me acompañaron durante los desvelos y los retos que atravesamos a lo largo de la carrera. Y a Kevin, Rodrigo, Axel, Godu por acompañarme en mis primeros semestres donde todos nos vimos crecer.

Por último, a los escorpiones de la Facultad de Ingeniería por su apoyo incondicional,IEEE EMBS y al hospital Conde de Valenciana en específico a Marcos, Rubén , Aaron, Adrian, Armando, Adriana, Laura, Erika ,Paola, Fabian y Lorena quienes son también parte de este logro. No puedo excluir de este escrito a todos mis profesores de la universidad, quienes, con sus enseñanzas, no solo me dotaron de conocimiento técnico, sino también de valores que hoy me definen como profesional. Gracias totales Universidad Nacional Autónoma de México y Facultad de ingeniería.

**Sinceramente:**

**Josue David Díaz Gutiérrez**

# Índice general

<b>1. Antecedentes</b>	<b>1</b>
1.1. Estado del Arte: Sistemas In silico . . . . .	1
1.2. Simulación de proteínas . . . . .	5
1.2.1. Estructura biológica . . . . .	6
1.3. Contranportadores . . . . .	10
1.3.1. Definción de contranportador . . . . .	10
1.3.2. Contranportador Na <sup>+</sup> Cl <sup>-</sup> . . . . .	10
1.4. Medicamentos Tiazidas . . . . .	12
<b>2. Definición del problema</b>	<b>14</b>
2.1. Definición del problema . . . . .	14
<b>3. Metodología</b>	<b>16</b>
3.1. Ligando . . . . .	18
3.2. Secuencia . . . . .	18
3.2.1. Relajación y validación estructural . . . . .	19

3.2.2. Identificación del sitio de unión . . . . .	22
3.2.3. Podado de proteína . . . . .	22
3.3. Analisis de la unión proteína-fármaco . . . . .	22
<b>4. Resultados</b>	<b>26</b>
4.1. Analisis Resultados . . . . .	26
<b>5. Conclusiones y trabajo a futuro</b>	<b>34</b>
5.1. Conclusiones . . . . .	34
5.2. Trabajo a futuro . . . . .	35
<b>Bibliografía</b>	<b>37</b>

# Índice de figuras

1.1. Línea del tiempo de modelos computacionales para simulación de proteínas, cada color nos menciona un tipo de modelo diferente. Podemos destacar como los modelos aumentaron en los años 2021-2023 ([13]) . . . . .	2
1.2. Estructura general de los aminoácidos y clasificación según la polaridad de su cadena lateral (R). Panel (A) muestra la estructura genérica de un aminoácido. Los paneles (B), (C) y (D) ilustran ejemplos de aminoácidos con cadenas laterales no polares, polares y aromáticas, respectivamente. (Figura tomada de [1])	7
1.3. Interacciones moleculares, A)Interacciones de Van der Waals B)Enlaces de Hidrogeno, C)Interacciones Iónicas, D)Enlaces covalentes. (Figura obtenida de [1])	8
1.4. El diagrama ilustra el funcionamiento de los transportadores, destacando primero aquellos de tipo unitario. Estos transportadores solo permiten el paso de sustancias en una única dirección a través de la membrana.En la parte derecha del diagrama (A y B), se muestran los componentes de un cotransportador, los cuales incluyen:Simporte: Transporta dos o más sustancias en la misma dirección. Antiporte: Transporta dos o más sustancias en direcciones opuestas.(Diagrama obtenido de [29]) . . . . .	10

1.5. Manejo del sodio en la nefrona distal. El asa gruesa ascendente (TAL) es responsable del 20 %–30 % de la reabsorción de Na <sup>+</sup> a través del cotransportador NKCC2. En el túbulo contorneado distal (DCT), el cotransportador NCC permite la reabsorción del 5 %–10 % del Na <sup>+</sup> . Otros mecanismos incluyen ROMK, Na <sup>+</sup> /K <sup>+</sup> ATPasa y CLC-Kb. (Diagrama obtenido de [30]) . . . . .	11
1.6. Ilustración de la interacción tiazida-NCC. (Obtenido de[30]) . . . . .	13
3.1. Esquema de la metodología empleada para el análisis del cotransportador NCC. (A) Construcción del modelo tridimensional de la politiazida. (B) Relajación estructural del ligando. (C) Generación de los modelos tridimensionales de las proteínas NCC. (D) Relajación y validación estructural de los modelos proteicos. (E) Identificación de cavidades para evaluar las interacciones ligando-proteína. (F) Podado del modelo tridimensional de las proteínas NCC. (G) Análisis de la interacción proteína-fármaco. . . . .	17
3.2. Estructura tridimensional de la politiazida en formato PDB, obtenida de la base de datos PubChem (ID: 4870) y optimizada con Avogadro. [33] . . . . .	18
3.3. HNCC obtenida por medio de CRYO-EM, perteneciente a la familia de transportadores SLC12A3 . . . . .	19
3.4. β-NCC por medio de CRYO-EM, perteneciente a la familia de transportadores SLC12A3 . . . . .	19
4.1. Podemos observar los gráficos de Ramachandra obtenidos, del lado izquierdo tenemos el correspondiente a HNCC y del lado derecho β-NCC. . . . .	27
4.2. Cavidades en HNCC(izquierda) y β-NCC(derecha) en Cavity plus, donde en color verde podemos observar la cavidad presente en la cadena B. . . . .	28
4.3. Podado de proteínas correspondiente, con el objetivo de disminuir el cómputo, se observa del lado izquierdo HNCC y del lado derecho βNCC . . . . .	29
4.4. HNCC . . . . .	31

4.5.  $\beta$ NCC . . . . . 31

# Índice de tablas

4.1. Comparación de resultados Molprobit NCC entre anguila y humano . . . . .	27
4.2. Coordenadas obtenidas de la cavidad en HNCC y $\beta$ -NCC. . . . .	29
4.3. Comparación de resultados entre HNCC y $\beta$ NCC. . . . .	30
4.4. Comparación de resultados entre HNCC y $\beta$ NCC. . . . .	31

## Glosario

- **Protein Data Bank.**- Estilo de formato utilizado en bioinformática que contiene información estructural de moléculas y macromoléculas. También puede ser que se use como base de datos.
- **NCC.**- Transportador  $\text{Na}^+\text{Cl}^-$  de membrana esencial para la función renal.
- **Ligando.**- Molécula, generalmente pequeña, que se une a una biomolécula (como una proteína, un receptor o una enzima) y desencadena o modula su actividad.
- **Simulación *in silico*.**- Tipo de simulación que utiliza modelos y software para estudiar procesos biológicos en un entorno virtual.
- **Contrasportador.**- Subcategoría de proteínas de transporte de membrana que acoplan el movimiento favorable de una molécula con su gradiente de concentración y el movimiento desfavorable de otra molécula contra su gradiente.
- **Tiazidas.**- Clase de medicamentos diuréticos.
- **Docking o unión ligando-proteína.**- Técnica computacional utilizada para predecir la forma en que dos moléculas se acoplan entre sí.
- **Aminoácido.**- Unidad básica de las proteínas.
- **Interacciones de Van der Waals.**- Ocurren cuando moléculas con formas complementarias se acercan entre sí. Estas moléculas pueden no tener carga y solo contener enlaces no polares, pero al entrar en contacto cercano, se puede inducir un dipolo instantáneo que permite interacciones electrostáticas débiles entre grupos con cargas opuestas [1].
- **Enlaces de hidrógeno.**- Ocurren entre moléculas que contienen dipolos, donde la atracción electrostática se produce entre un átomo de N o O parcialmente negativo (con un par de electrones solitario) y un átomo de hidrógeno parcialmente positivo unido a un átomo de N o O diferente.
- **Interacciones iónicas.**- Son interacciones atractivas entre iones con cargas opuestas. Son

la interacción intermolecular más fuerte que involucra carga.

- **Enlaces covalentes entre cisteínas.**- Se forman por la oxidación de dos residuos de cisteína, que pueden encontrarse dentro de una misma cadena proteica o entre dos cadenas separadas.
- **Hélice  $\alpha$ .**- Estructura helicoidal dextrógira estabilizada por puentes de hidrógeno entre nitrógenos y grupos carbonilo de los enlaces peptídicos.
- **Láminas  $\beta$ .**- Estructuras formadas por cadenas polipeptídicas dispuestas en paralelo o antiparalelo, estabilizadas por puentes de hidrógeno entre los grupos carbonilo y amida de cadenas adyacentes.
- **Relajación.**- Proceso en el que se minimiza la energía de la estructura proteica.
- **Energía de unión.**- Representa la cantidad total de energía liberada cuando el ligando se une al receptor. Una mayor energía de unión indica una interacción más fuerte.
- **Eficiencia del ligando.**- Relación entre la energía de unión y el número de átomos no hidrógeno presentes en el ligando. Indica qué tan eficiente es el ligando en función de su tamaño.
- **Constante de inhibición (Ki).**- Valor derivado de la energía de unión que mide la concentración de ligando necesaria para inhibir la actividad del receptor en un 50%.
- **Energía interna del ligando.**- Energía inherente del ligando en su conformación actual, incluyendo la energía torsional y las interacciones intramoleculares.
- **Energía van der Waals y enlaces de hidrógeno.**- Conjunto de interacciones atractivas entre ligando y receptor, esenciales para la estabilidad y especificidad de la unión.
- **Energía electrostática.**- Atracción o repulsión entre cargas opuestas presentes en el ligando y el receptor.
- **Energía torsional.**- Energía requerida para que el ligando cambie su conformación, relacionada con su flexibilidad para adaptarse al receptor.

- **Enlace de hidrógeno.**- Interacción atractiva que contribuye a la especificidad y estabilidad del complejo ligando-receptor.
- **Interacciones de Van der Waals.**- Interacciones atractivas o repulsivas que son fundamentales para la estabilidad del complejo ligando-receptor.
- **Interacciones electrostáticas.**- Interacciones entre átomos cargados de la proteína y del ligando, que afectan la afinidad de la unión.
- **Energía de solvatación.**- Energía asociada a la desolvatación del ligando durante su unión a la proteína.
- **Ruptura de la cadena.**- Penalización aplicada por interrupciones en la cadena de la proteína, que pueden afectar la estabilidad del complejo.
- **Restricciones de coordenadas.**- Penalizaciones por desviaciones en las posiciones atómicas del ligando respecto a las coordenadas deseadas.
- **Restricciones de ángulos torsionales.**- Penalizaciones por desviaciones en los ángulos torsionales del ligando, importantes para mantener su conformación.
- **Ángulo omega.**- Parámetro que evalúa las penalizaciones por desviaciones en los ángulos de enlace peptídico, fundamentales para la estabilidad estructural.
- **Rama.**- Evaluación de la geometría de los ángulos de torsión de la cadena principal de la proteína mediante el gráfico de Ramachandran, asegurando la correcta estabilidad tridimensional.
- **Criomicroscopía electrónica.**-técnica avanzada que permite estudiar la estructura de macromoléculas biológicas, como proteínas y ácidos nucleicos, a resolución cercana al átomo

### Abreviaturas

- HNCC.- Contrtransportador NCC de humano
- $\beta$ NCC.- Contrtransportador NCC de Anguila.

- Cry-em.- Criomicroscopía electrónica
- PDB.- Protein Data Bank

## Simulación de la función del cotransportador de $\text{Na}^+\text{Cl}^-$ sensible a tiazidas (NCC)

por

Josue David Díaz Gutiérrez

### Resumen

La simulación de proteínas ha emergido como una herramienta fundamental en la investigación dado que apoya a reducir tiempo y recursos en difentes líneas de investigación.[2] [3] Con la implementación de modelos de redes neuronales e inteligencia artificial, la simulación *in silico* se ha convertido en un apoyo crucial para diferentes líneas de investigación relacionada a biología celular.

En este trabajo, se propone la simulación *in silico* de la función del cotransportador  $\text{Na}^+\text{Cl}^-$  sensible a tiazidas en dos especies: humano y anguila. Este cotransportador desempeña un papel esencial en el funcionamiento renal y las mutaciones en su estructura pueden estar asociadas con enfermedades relacionadas con la hipertensión debido a alteraciones en el intercambio salino. El objetivo principal es evaluar el efecto del ligando, que en el marco de este trabajo se utilizaran las tiazidas[4], sobre estas proteínas y determinar su capacidad para inhibir su función.

Para ello, se llevó a cabo una metodología que incluyó la obtención de las estructuras proteicas en formato Protein Data Bank y su preparación mediante dos etapas principales: relajación y búsqueda de cavidades. Asimismo, se realizó la relajación de la tiazida para proceder a la simulación de la interacción ligando-proteína.

Los resultados obtenidos indican que HNCC presenta la mejor conformación para la interacción con la tiazida, debido a su mayor estabilidad energética, mejor eficiencia del ligando y un perfil de interacciones más favorable en comparación con  $\beta$ NCC. A pesar de que  $\beta$ NCC exhibe ventajas en términos de flexibilidad y solvatación, los análisis realizados con dos softwares diferentes confirman que HNCC es la conformación óptima para la interacción con la tiazida.

**Simulation of the function of the thiazide-sensitive  $\text{Na}^+\text{Cl}^-$  cotransporter (NCC)**

by

Josue David Díaz Gutiérrez

**Abstract**

Protein simulation has emerged as a fundamental tool in research, as it helps reduce time and resources across different lines of investigation [2] [3]. With the implementation of neural network models and artificial intelligence, *in silico* simulation has become a crucial support for various fields of research related to cell biology.

In this work, we propose the *in silico* simulation of the function of the thiazide-sensitive  $\text{Na}^+\text{Cl}^-$  cotransporter in two species: human and eel. This cotransporter plays an essential role in renal function, and structural mutations can be associated with diseases related to hypertension due to alterations in salt exchange. The main objective is to evaluate the effect of the ligand — thiazides, in the context of this work [4] on these proteins and determine their ability to inhibit the cotransporter's function.

To achieve this, a methodology was employed that included obtaining the protein structures in Protein Data Bank format and preparing them through two main stages: relaxation and cavity detection. Additionally, the thiazide was relaxed in preparation for simulating the ligand–protein interaction.

The results obtained indicate that HNCC exhibits the best conformation for interacting with thiazide, due to its greater energetic stability, better ligand efficiency, and a more favorable interaction profile compared to  $\beta$ NCC. Although  $\beta$ NCC shows advantages in terms of flexibility and solvation, analyses performed with two different software tools confirm that HNCC is the optimal conformation for interaction with thiazide.

## Introducción

El desarrollo de fármacos constituye uno de los pilares fundamentales de la medicina moderna; sin embargo, los métodos tradicionales presentan limitaciones en términos de tiempo, costos y eficiencia, aspectos que han comenzado a optimizarse gracias a los avances tecnológicos recientes [2]. En este contexto, la inteligencia artificial (IA) se posiciona como una herramienta disruptiva, con el potencial de transformar radicalmente las estrategias de descubrimiento y diseño de nuevos tratamientos [3]. Particularmente, el modelado de proteínas asistido por IA está redefiniendo nuestra comprensión de las interacciones moleculares a nivel atómico.

El uso de algoritmos avanzados de aprendizaje automático, incluyendo redes neuronales profundas y técnicas de aprendizaje por refuerzo [5], ha permitido predecir con notable precisión la estructura tridimensional de proteínas y simular su comportamiento en distintos entornos bioquímicos [6]. Estas innovaciones se han materializado en herramientas que han contribuido significativamente a logros científicos de alto impacto, reconocidos incluso con premios de prestigio internacional [7]. De esta manera, la IA habilita a los investigadores a explorar de forma virtual las interacciones entre proteínas y ligandos, acelerando la identificación de candidatos a fármacos prometedores y aumentando la eficiencia del proceso de validación. Además, su capacidad para analizar grandes volúmenes de datos genómicos y proteómicos permite descubrir patrones y relaciones complejas que resultarían inalcanzables mediante enfoques convencionales.[8]

En la presente tesis, se emplearán técnicas de modelado de proteínas impulsadas por IA para simular la función del cotransportador de sodio-cloro (NCC), una proteína esencial en la regulación de la presión arterial y objetivo terapéutico relevante en el tratamiento de ciertas patologías. Para ello, se utilizarán programas computacionales de acceso libre que integran algoritmos de inteligencia artificial para generar modelos estructurales de alta precisión del cotransportador en dos especies diferentes. Posteriormente, se aplicarán simulaciones de dinámica molecular y ensayos de acoplamiento molecular virtual para estudiar las interacciones del NCC con tiazidas, una clase de diuréticos ampliamente empleada en la terapéutica antihipertensiva [4]. Estas inter-

acciones son fundamentales para comprender los mecanismos de inhibición del cotransportador y optimizar el diseño de nuevos agentes terapéuticos.

El objetivo principal de esta investigación es comparar el desempeño de las tiazidas en dos especies diferentes, identificando posibles diferencias en la afinidad y la eficacia de la interacción.

## **Objetivo**

Simular la función del cotransportador de  $\text{Na}^+\text{Cl}^-$  sensible a tiazidas (NCC) utilizando técnicas de IA y validar los resultados con dos algoritmos diferentes.

# Capítulo 1

## Antecedentes

### 1.1. Estado del Arte: Sistemas In silico

Las simulaciones se pueden clasificar según su enfoque, propósito y el tipo de modelo que utilizan. Nos referimos a un modelo como una representación simplificada y abstracta de un sistema o fenómeno. Por ende, estos sistemas se basan en modelos matemáticos, estocásticos, basados en agentes o de redes, los cuales simplifican las características clave de un sistema para predecir su comportamiento bajo distintas condiciones.[9]

Las simulaciones computacionales en la biotecnología, también son conocidas como simulaciones *in silico*, utilizan estos modelos para representar fenómenos biológicos, físicos o químicos y predecir su comportamiento bajo distintas condiciones. [10] [11]

Además de ser herramientas predictivas, permiten optimizar diseños, validar hipótesis científicas y reducir la necesidad de experimentos físicos, lo que ahorra tiempo y recursos.

Con esta definición de sistemas *in silico*, observamos su aplicación en el ámbito de la simulación de proteínas, donde se emplean modelos computacionales para predecir su estructura y función. Esto se puede llevar a cabo de diferentes formas:

- PLM (Protein Language Models): se utilizan para representar secuencias de proteínas en

forma de embeddings. Un embedding es una representación vectorial de una secuencia de proteínas que contiene información relevante sobre la estructura y función de la propia proteína.

- **Secuencia-Estructura:** en este caso, se utiliza un archivo FASTA. Este tipo de archivo contiene la secuencia de aminoácidos de una proteína y, a partir de esta información, se puede predecir o producir su estructura.
- **Estructura-Secuencia:** en este caso, se utiliza un archivo de tipo Protein Data Bank. Un archivo PDB contiene información sobre la estructura tridimensional de una proteína, lo que permite identificar en qué parte de la proteína se encuentra la cadena que se desea localizar.
- **Función:** para estudiar la función de una proteína, es necesario contar previamente con su estructura tridimensional (3D). Esta información permite realizar simulaciones adecuadas mediante técnicas como el Docking[12] [13]

### Protein Models: The Standouts

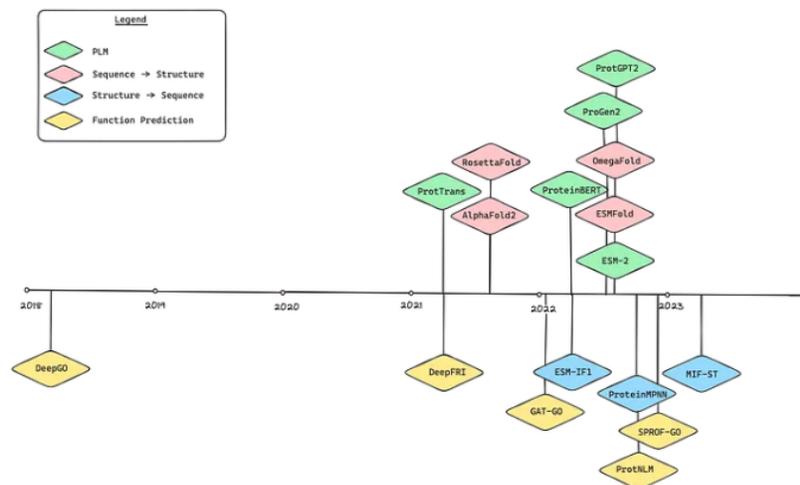


Figura 1.1: Línea del tiempo de modelos computacionales para simulación de proteínas, cada color nos menciona un tipo de modelo diferente. Podemos destacar como los modelos aumentaron en los años 2021-2023 ([13])

Como se puede observar en la figura 1.1, el avance en el modelado computacional de proteínas ha estado fuertemente impulsado por el desarrollo e integración de modelos de aprendizaje automático y redes neuronales profundas. El primer modelo destacado es DeepGO (2018)[14], un software de acceso libre que permite predecir la función de una proteína a partir de su secuencia en formato FASTA, utilizando clasificaciones del Gene Ontology (GO). A partir de ese punto, se ha generado una acelerada evolución en la creación de modelos más complejos, con diferentes enfoques y aplicaciones.

Entre los modelos más influyentes en esta evolución se encuentra AlphaFold2 (2021)[5], desarrollado por DeepMind, que revolucionó la biología estructural al permitir predecir la estructura tridimensional de proteínas con una precisión cercana a la obtenida experimentalmente. Le sigue RoseTTAFold, desarrollado el mismo año por el Instituto de Diseño de Proteínas, el cual, mediante una arquitectura de red neuronal de tipo three-track, ofrece una alternativa más ágil y modular para el modelado estructural. En paralelo, modelos como ESMFold[15] y ESM-2[16], desarrollados por Meta AI, introdujeron el uso de modelos de lenguaje entrenados con grandes corpus de secuencias proteicas para predecir estructuras de forma directa, sin necesidad de alineamientos múltiples.

Por otro lado, herramientas como ProteinMPNN (2022)[17] y ProtGPT2 (2023)[18] han ampliado el enfoque hacia el diseño de proteínas de novo, permitiendo generar secuencias a partir de estructuras deseadas (estructura  $\rightarrow$  secuencia), lo cual resulta valioso para la ingeniería de proteínas y el diseño racional de fármacos.

Durante este periodo de desarrollo en la simulación de proteínas, han destacado tres softwares que han tenido mayor relevancia y uso en comparación con los demás: SWISS MODEL, AlphaFold y RoseTTAFold. Estos programas han sido ampliamente adoptados por la comunidad científica debido a su accesibilidad, precisión y versatilidad en distintos escenarios de modelado estructural y diseño de terapias basadas en proteínas.

- **AlphaFold:** Es un software de simulación de estructuras proteicas basado en técnicas de aprendizaje automático (*machine learning*). Su desarrollo comenzó en 2018 como parte

de la competencia **CASP13** (Critical Assessment of Structure Prediction). No obstante, fue hasta el año 2020, con la participación de **AlphaFold2** en **CASP14**, cuando alcanzó una precisión del 92.4 %, posicionándose como una herramienta altamente confiable para la predicción estructural de proteínas. Esta mejora significativa fue posible gracias a la incorporación de mecanismos de atención, aprendizaje de extremo a extremo, métodos basados en plantillas y sin plantillas, así como un refinamiento iterativo del modelo [6]. Actualmente, se encuentra en desarrollo **AlphaFold3**, que promete ampliar sus capacidades estructurales y funcionales.

- **Rosetta:** Desarrollado bajo la dirección del acreedor al Premio Nobel del 2024 **David Baker**, **Rosetta** es un conjunto integral de herramientas para el modelado de estructuras macromoleculares. Su diseño flexible y modular permite su aplicación en tareas de predicción, diseño y remodelado de proteínas y ácidos nucleicos. Desde su introducción en 1998, los servidores web asociados a Rosetta han ejecutado miles de millones de simulaciones, tanto en predicción estructural como en diseño de proteínas, extendiéndose también al uso masivo en clústeres de supercomputación [19].
- **SWISS-MODEL:** Es una plataforma basada en homología para la modelación estructural de proteínas, desarrollada en 1993. Desde su lanzamiento, ha sido uno de los servidores en línea más utilizados en la comunidad científica. El sistema ha sido objeto de múltiples actualizaciones, destacando la versión de abril de 2023, en la cual se integraron modelos estructurales generados por **AlphaFold**, fortaleciendo así su base de datos y capacidades predictivas [20].

Estos softwares son los más usados en la parte de la simulación de estructuras, pero para la simulación de la función de las proteínas actualmente se cuenta con una técnica que lleva por nombre Docking. El Docking es un método bioinformático que permite predecir y calcular computacionalmente la posición más favorable de interacción entre un ligando y un blanco (usualmente proteico) a partir de sus representaciones tridimensionales. Mientras más estable, específica y favorable sea la unión entre un ligando (fármaco) y su blanco proteico (diana terapéutica), mayor será su actividad biológica (fármaco más efectivo).[21]

Actualmente, el proceso de docking se ha podido simplificar gracias al desarrollo de nuevos softwares como es, Rosie[22], un software en línea desarrollado por el equipo de Rosetta. Este programa facilita la preparación tanto de las proteínas como de los ligandos antes de procesarlos para el análisis. Además, existen otras herramientas similares que también apoyan en la preparación de proteínas y ligandos, optimizando el flujo de trabajo en estudios de interacción molecular. [23]

## 1.2. Simulación de proteínas

La simulación de proteínas ha experimentado avances significativos gracias a la mejora continua de las redes neuronales aplicadas en este campo. Estos avances han permitido obtener modelos estructurales con una precisión cada vez mayor, facilitando la comprensión de sus propiedades funcionales y su interacción con otras biomoléculas.[24]

Para la obtención de estructuras proteicas mediante simulación, se han desarrollado diversos enfoques algorítmicos, entre los cuales destacan dos principales:

- **Homologación:** Este método se basa en el principio de homología, el cual postula que la existencia de una relación evolutiva entre dos secuencias de aminoácidos se traduce en una estructura tridimensional similar. Dicho enfoque ofrece una alta precisión en la generación de modelos cuando se dispone de secuencias con semejanzas conocidas. Sin embargo, su aplicabilidad está limitada, ya que depende de estructuras previamente determinadas mediante experimentación, lo que impide la identificación de nuevos patrones de plegamiento. Para optimizar su desempeño, este método requiere acceso a bases de datos especializadas que proporcionen las plantillas necesarias para la homologación. [25]
- **Ab Initio:** Este enfoque se basa en la información extraída de estructuras tridimensionales depositadas en bases de datos y en principios físicos para generar modelos estructurales. A diferencia de la homologación, los métodos Ab Initio permiten predecir patrones estructurales no recurrentes mediante la comparación de fragmentos. Este método es particularmente útil cuando no existen estructuras homólogas conocidas, lo que lo convierte

en una estrategia clave para el estudio de proteínas con conformaciones novedosas o poco caracterizadas.[13]

Para que los modelos generados sean precisos y confiables, es fundamental contar con una base de datos extensa y bien estructurada. Estas bases de datos pueden clasificarse según el origen de la información que contienen. Por un lado, las bases de datos primarias recopilan datos experimentales directos, como estructuras tridimensionales obtenidas mediante cristalografía de rayos X o resonancia magnética nuclear. Por otro lado, las bases de datos secundarias incluyen información derivada, como anotaciones funcionales y traducciones de secuencias de aminoácidos. En la práctica, ambos tipos de bases de datos suelen combinarse para ampliar el rango de información disponible y mejorar la calidad de los modelos generados. En el contexto de la simulación de proteínas, la base de datos más utilizada es el PDB, que proporciona acceso a una amplia colección de estructuras proteicas determinadas experimentalmente. [26]

Por la parte de la simulación de la función contamos con la técnica de docking donde un primer algoritmo de muestreo predice todas las configuraciones o conformaciones estructurales (poses) que el ligando puede asumir dentro de un dominio o sitio activo en el blanco proteico, y un segundo algoritmo con una función de puntuación que predice las energías de unión entre el ligando y el receptor para cada una de las configuraciones establecidas por el primer algoritmo.[24]

### 1.2.1. Estructura biológica

Debemos de comprender a las proteínas como cadenas de aminoácidos que se pliegan para tener una estructura tridimensional para llevar a cabo diversas funciones.

Un aminoácido es la unidad básica de la proteína. Su denominación responde a la composición química general que presentan, en la que un grupo amino (-NH<sub>2</sub>) y otro carboxilo (-COOH) se unen a un carbono alfa (-C-). Las otras dos valencias de ese carbono quedan saturadas con un átomo de hidrógeno (-H) y con un grupo químico variable al que se denomina cadena lateral (-R) como en la figura 1.2. [27]

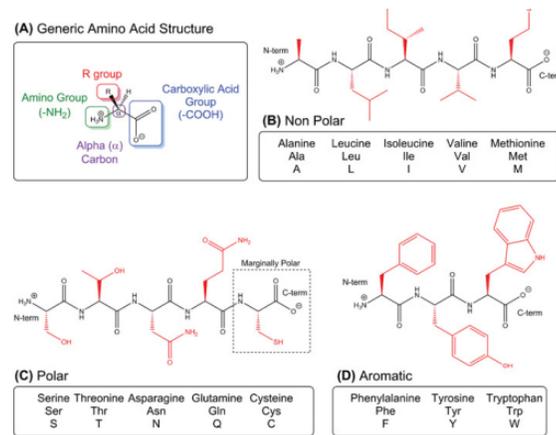


Figura 1.2: Estructura general de los aminoácidos y clasificación según la polaridad de su cadena lateral (R). Panel (A) muestra la estructura genérica de un aminoácido. Los paneles (B), (C) y (D) ilustran ejemplos de aminoácidos con cadenas laterales no polares, polares y aromáticas, respectivamente. (Figura tomada de [1])

Una vez que nosotros sabemos que es un aminoácido esto nos ayuda principalmente a ver las diferentes estructuras que conforman a una proteína.

**Estructura primaria:** La estructura primaria de las proteínas se fundamenta en la cadena de aminoácidos que conforman a la proteína, recordando que las interacciones de estos aminoácidos nos llevan a diferentes estructuras que le van a dar diferentes características a nuestras proteínas, aparte que dichos enlaces nos van a ayudar a inferir algunas cualidades intra e inter moleculares [27][1].

Dichas cualidades principalmente son las diferentes fuerzas que se ven entre las uniones intramoleculares (fig1.3) como pueden ser;

- Interacciones de Van der Waals: ocurren cuando moléculas con formas complementarias se acercan entre sí. Estas moléculas pueden no tener carga y solo contener enlaces no polares, pero al entrar en contacto cercano, se puede inducir un dipolo instantáneo en estos enlaces no polares, permitiendo interacciones electrostáticas débiles entre grupos con cargas opuestas.[1]

- Enlaces de hidrógeno: ocurre cuando dos moléculas que interactúan contienen dipolos (es decir, contienen enlaces covalentes polares), donde la atracción electrostática se produce entre un átomo de N o O parcialmente negativo (con un par de electrones solitario) y un átomo de hidrógeno parcialmente positivo que está unido covalentemente a un átomo de N o O diferente.
- Interacciones iónicas: son interacciones atractivas entre iones con cargas opuestas. Dado que los iones contienen más carga que los otros dipolos discutidos anteriormente, son la interacción intermolecular más fuerte que involucra carga.
- Enlaces covalentes entre cisteína: se forman por la oxidación de dos residuos de cisteína, que pueden formarse dentro de una sola cadena de proteína o entre dos cadenas separadas. Dado que estos enlaces son covalentes, son el enlace intermolecular general más fuerte.

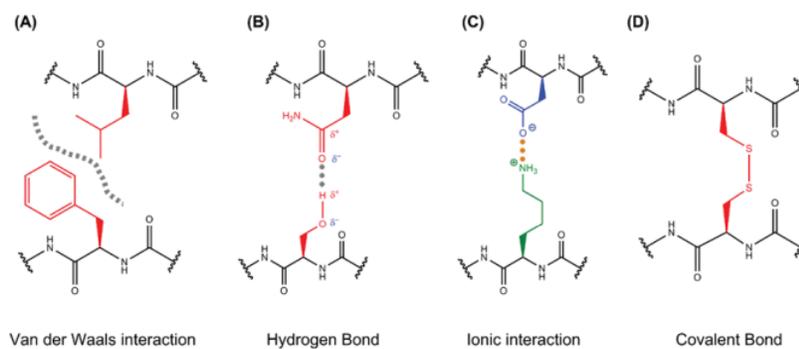


Figura 1.3: Interacciones moleculares, A)Interacciones de Van der Waals B)Enlaces de Hidrogeno, C)Interacciones Iónicas, D)Enlaces covalentes. (Figura obtenida de [1])

**Estructura secundaria:** En esta etapa comienza el plegamiento de la proteína, el cual ocurre cuando su estructura tiende a alcanzar el estado de menor energía. En este proceso la energía de Gibbs está presente, donde un valor negativo indica un plegamiento espontáneo y correcto.

El plegamiento también define los dominios de una proteína. Un dominio proteico se define como "Una región física específica con una función determinada o una secuencia de aminoácidos que se corresponde con un segmento específico de ADN".

Dentro de la estructura secundaria, existen dos motivos estructurales fundamentales: las hélices

$\alpha$  y las láminas  $\beta$ .

La hélice  $\alpha$  es una estructura helicoidal dextrógira estabilizada por puentes de hidrógeno entre los nitrógenos y los grupos carbonilo de los enlaces peptídicos, un enlace peptídico es la unión de dos aminoácidos mediante la pérdida de una molécula de agua entre el grupo amino de un aminoácido y el grupo carboxilo del otro. Estos enlaces son intracatenarios y se establecen entre el oxígeno carbonílico de un residuo ( $n$ ) y el nitrógeno del residuo situado en posición ( $n + 4$ ). Como resultado, la hélice  $\alpha$  adopta una conformación compacta con un giro cada 3.6 residuos de aminoácidos y un avance de aproximadamente 5.4 Å por vuelta.

Por otro lado, las láminas  $\beta$  están formadas por cadenas polipeptídicas dispuestas en paralelo o antiparalelo, estabilizadas por puentes de hidrógeno entre los grupos carbonilo y amida de los enlaces peptídicos de cadenas adyacentes. En esta estructura, las cadenas laterales de los aminoácidos se sitúan de forma alternante a la derecha y a la izquierda del esqueleto de la cadena polipeptídica. Las estructuras  $\beta$  de distintas cadenas polipeptídicas, o bien de diferentes zonas de una misma cadena, pueden interaccionar entre sí mediante puentes de hidrógeno  $\beta$ [1][27][28].

Estructura Terciaria: Se llama estructura terciaria a la disposición tridimensional de todos los átomos que componen la proteína, concepto equiparable al de conformación absoluta en otras moléculas. La estructura terciaria de una proteína es la responsable directa de sus propiedades biológicas, ya que la disposición espacial de los distintos grupos funcionales determina su interacción con los diversos ligandos.[28]

En esta tesis, profundizaremos en la estructura y función de las proteínas de membrana, componentes esenciales de las membranas celulares. Específicamente, analizaremos los cotransportadores.

## 1.3. Contraspportadores

### 1.3.1. Definción de contraspportador

Los contraspportadores son proteínas de membrana que usan gradientes para mover moléculas entre ambos lados de la membrana, en el caso de un contraspportador permite la translación simultánea de dos elementos (dos moléculas, una molécula y un ión, o dos iones) entre ambos lados de la membrana como se observa en la figura(1.4). Si el sentido en el que viajan los dos elementos es opuestos se denomina antiporte y si los dos viajan en el mismo sentido se denomina simporte.

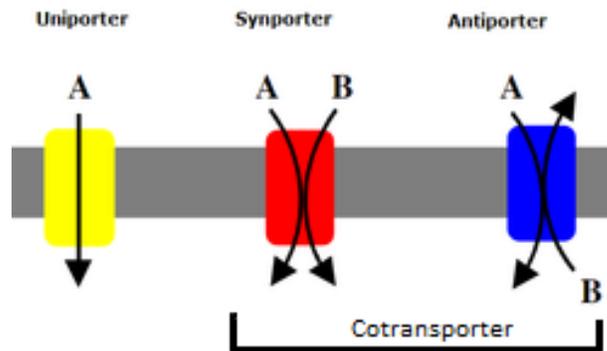


Figura 1.4: El diagrama ilustra el funcionamiento de los transportadores, destacando primero aquellos de tipo unitario. Estos transportadores solo permiten el paso de sustancias en una única dirección a través de la membrana. En la parte derecha del diagrama (A y B), se muestran los componentes de un cotransportador, los cuales incluyen: Simporte: Transporta dos o más sustancias en la misma dirección. Antiporte: Transporta dos o más sustancias en direcciones opuestas. (Diagrama obtenido de [29])

### 1.3.2. Contraspportador $\text{Na}^+\text{Cl}^-$

El cotransportador  $\text{Na}^+\text{Cl}^-$  (NCC) está presente en los riñones, específicamente en la nefrona, donde representa la principal vía de reabsorción de sal en el túbulo contorneado distal (fig. 1.5).

Este proceso es esencial para mantener la homeostasis de sal y agua en el organismo, ya que regula el equilibrio de sales dentro del riñón. Dicho equilibrio, a su vez, es fundamental para otros procesos vitales, como la regulación de la presión arterial y el control de los líquidos corporales. [4]

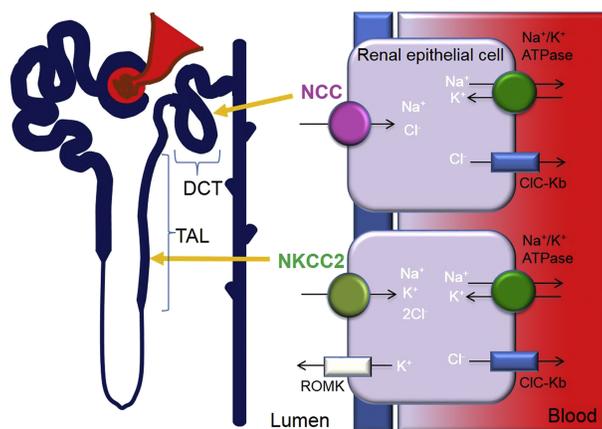


Figura 1.5: Manejo del sodio en la nefrona distal. El asa gruesa ascendente (TAL) es responsable del 20%–30% de la reabsorción de  $\text{Na}^+$  a través del cotransportador NKCC2. En el túbulo contorneado distal (DCT), el cotransportador NCC permite la reabsorción del 5%–10% del  $\text{Na}^+$ . Otros mecanismos incluyen ROMK,  $\text{Na}^+/\text{K}^+$  ATPasa y CLC-Kb.

(Diagrama obtenido de [30])

Las alteraciones en la función del NCC pueden tener consecuencias clínicas significativas. Por ejemplo, una mutación en este transportador puede dar lugar a una enfermedad opuesta al síndrome de Gitelman, manifestándose con hipertensión arterial, acidosis metabólica e hipercalcemia. Esta condición es conocida como síndrome de Gordon, también llamado pseudohipoaldosteronismo tipo II (PHAII) o hipertensión hipercalémica familiar (HHF). [4]

Dado su papel crucial en la regulación de la sal y la presión arterial, el NCC es un objetivo terapéutico clave para ciertos fármacos, como las tiazidas. Estas forman un grupo de diuréticos que inhiben la actividad del NCC en el túbulo contorneado distal, reduciendo la reabsorción de sodio y cloro y, por ende, promoviendo la eliminación de agua. Su alta especificidad por el NCC se debe a la estructura de este transportador, el cual está compuesto por un dominio central hidrofóbico con 12 regiones transmembrana (TM) interconectadas por 6 asas extracelulares y 5

asas intracelulares [31]. Además, presenta un corto extremo amino y un extremo carboxilo más grande, ambos ubicados en el citoplasma. Esta disposición estructural es clave para su función y explica la interacción eficiente con las tiazidas, que bloquean el sitio de unión del  $\text{Cl}^-$ , inhibiendo así su actividad y produciendo su efecto diurético.

## 1.4. Medicamentos Tiazidas

Las tiazidas, diuréticos fundamentales en el tratamiento de la hipertensión arterial, insuficiencia cardíaca, síndrome nefrótico y retención de líquidos, ejercen su acción terapéutica al inhibir la reabsorción de sodio y cloruro en el túbulo contorneado distal del riñón. Esta inhibición, facilitada por su estructura química que incluye un anillo de benzotiadiazina con grupos funcionales como el sulfonamida y, frecuentemente, un cloro en la posición 6, les permite interactuar específicamente con el cotransportador de sodio-cloruro (NCC). [30] Al bloquear el NCC, las tiazidas aumentan la excreción de agua y electrolitos en la orina, reduciendo así el volumen de líquido corporal, lo que a su vez disminuye la presión arterial y alivia la retención de líquidos en diversas condiciones médicas.

El mecanismo de acción primario que se puede observar en la figura 1.6 de las tiazidas, se centra en la inhibición del cotransportador de sodio-cloruro  $\text{Na}^+/\text{Cl}^-$  en el túbulo contorneado distal (DCT) del riñón. Al bloquear este canal, se reduce la reabsorción de sodio, lo que a su vez disminuye la reabsorción de agua, ya que el agua sigue pasivamente al sodio. Este aumento en la excreción de agua y sodio (diuresis) disminuye el volumen plasmático, lo que lleva a una reducción de la presión arterial. Inicialmente, esta disminución de la presión arterial activa el sistema renina-angiotensina-aldosterona (RAAS) y el tono simpático como mecanismos compensatorios. Esta activación del RAAS explica el efecto sinérgico de las tiazidas con los inhibidores de la enzima convertidora de angiotensina (IECA) o los antagonistas del receptor de angiotensina II (ARA-II). Además, las tiazidas inducen una modesta vasodilatación, aunque el mecanismo exacto aún no se comprende completamente.[30]

En cuanto a los efectos de las tiazidas, aumentan la excreción de sodio, cloruro, potasio y agua,

pero también incrementan la reabsorción de calcio. Este aumento en la reabsorción de calcio se debe a un intercambio compensatorio de calcio por sodio a través del intercambiador  $\text{Na}^+\text{Ca}^{2-}$  en la membrana basolateral.

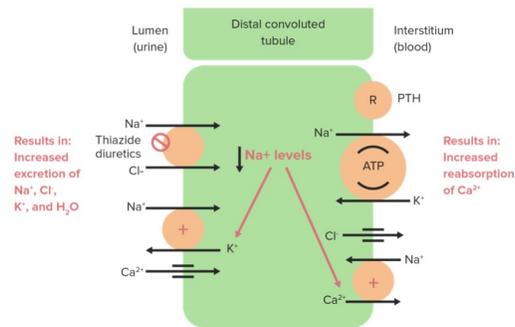


Figura 1.6: Ilustración de la interacción tiazida-NCC. (Obtenido de[30])

Sin embargo, el uso de tiazidas puede llevar al desarrollo de hipocalemia (niveles bajos de potasio en sangre). Esto ocurre porque la menor reabsorción de sodio en el DCT aumenta la cantidad de sodio que llega a los túbulos colectores, lo que estimula la liberación de aldosterona. La aldosterona, a su vez, activa el intercambiador  $\text{Na}^+/\text{K}^+$ , promoviendo la reabsorción de sodio y la excreción de potasio.

Además, las tiazidas pueden causar alcalosis metabólica. Esto se debe a que la aldosterona estimula el intercambiador  $\text{K}^+\text{H}^+$  en los túbulos colectores. Como resultado, parte del potasio se reabsorbe a cambio de hidrógeno, que se excreta, lo que lleva a una pérdida de hidrógeno y, por lo tanto, a alcalosis metabólica.[32]

# Capítulo 2

## Definición del problema

### 2.1. Definición del problema

El cotransportador de sodio y cloro sensible a tiazidas (NCC) desempeña un papel clave en la regulación del equilibrio de electrolitos en los riñones, y su inhibición por tiazidas es fundamental para los efectos diuréticos de estos medicamentos. Sin embargo, los detalles sobre los aminoácidos específicos que intervienen en la unión de las tiazidas al NCC, así como el impacto de esta unión en la conformación de la proteína, aún no están completamente claros. Además, se desconoce si existen interacciones competitivas con los iones que transporta NCC, lo que dificulta la comprensión de los mecanismos precisos involucrados en la inhibición.

A pesar de los avances experimentales, estos procesos moleculares siguen siendo complejos y difíciles de modelar de manera tradicional. Por lo tanto, es crucial contar con herramientas que puedan predecir con mayor precisión las interacciones entre las tiazidas y el NCC, mejorando nuestra capacidad para desarrollar tratamientos más efectivos.

El objetivo principal de esta tesis es simular la función del NCC sensible a tiazidas (NCC) utilizando técnicas de Inteligencia Artificial (IA). Estas simulaciones permitirán modelar la interacción entre las tiazidas y NCC a nivel molecular, identificando los aminoácidos clave involucrados en la unión y analizando cómo las tiazidas inducen cambios conformacionales en

la proteína. Los resultados obtenidos a partir de estos modelos computacionales serán validados con datos experimentales obtenidos mediante técnicas de biología molecular, con el fin de confirmar la precisión y relevancia de las simulaciones en el contexto biológico.

Para alcanzar este objetivo, se emplearán métodos avanzados de modelado molecular y simulación de proteínas, apoyados en técnicas de IA como redes neuronales y algoritmos de aprendizaje automático. Estos modelos computacionales serán complementados con datos experimentales provenientes de estudios de biología molecular que permitan validar las predicciones obtenidas. Así, se busca proporcionar una comprensión más precisa de la función del NCC y su interacción con las tiazidas, contribuyendo al avance en el desarrollo de terapias farmacológicas basadas en este mecanismo.



## Capítulo 3

### Metodología

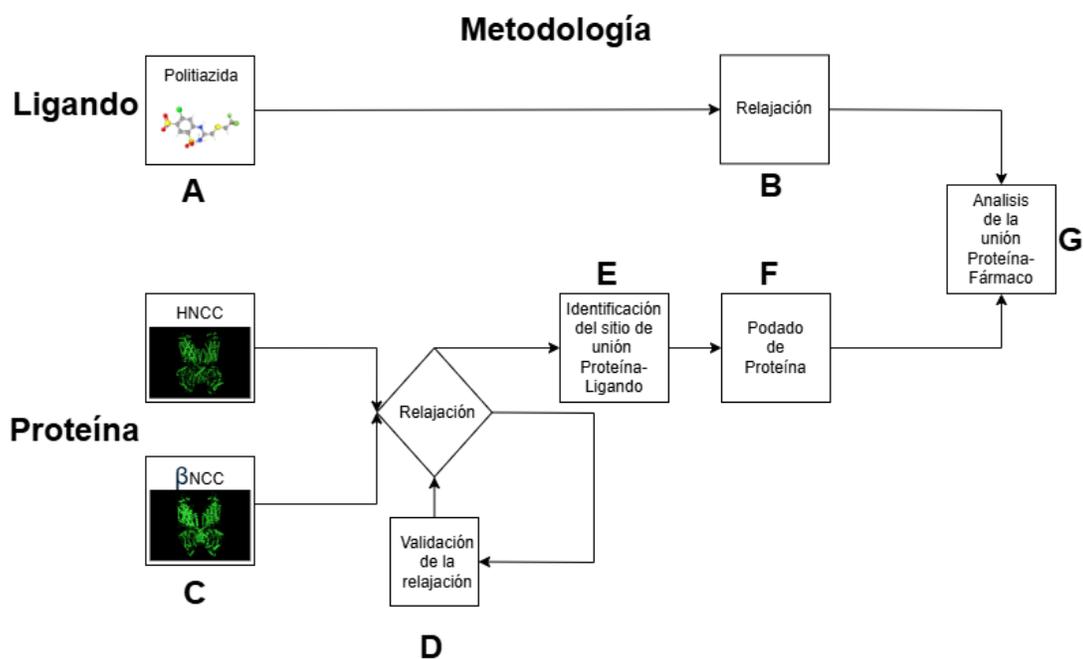


Figura 3.1: Esquema de la metodología empleada para el análisis del cotransportador NCC. (A) Construcción del modelo tridimensional de la politiázida. (B) Relajación estructural del ligando. (C) Generación de los modelos tridimensionales de las proteínas NCC. (D) Relajación y validación estructural de los modelos proteicos. (E) Identificación de cavidades para evaluar las interacciones ligando-proteína. (F) Podado del modelo tridimensional de las proteínas NCC. (G) Análisis de la interacción proteína-fármaco.

El estudio de la unión involucra dos elementos fundamentales como se puede observar en la figura 3.1; el ligando y la proteína. En este caso, la metodología se divide en estos dos componentes para su análisis detallado.

### 3.1. Ligando

El ligando utilizado en este estudio se obtuvo de la base de datos PubChem con el número de identificación 4870[33], donde se descargó el archivo en formato PDB (Fig.3.2 y Fig3.1A). Este formato es ampliamente utilizado en bioinformática y química computacional, ya que almacena información tridimensional sobre la estructura de moléculas y macromoléculas biológicas. Posteriormente, el ligando fue manipulado y procesado en el software Avogadro [34]. En este software, se llevó a cabo una minimización de energía con el objetivo de obtener una conformación más estable(Fig.3.1B), lo que favorece una mejor interacción al momento de realizar la unión.

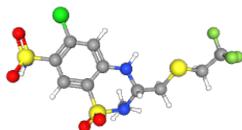


Figura 3.2: Estructura tridimensional de la politiazida en formato PDB, obtenida de la base de datos PubChem (ID: 4870) y optimizada con Avogadro. [33]

### 3.2. Secuencia

En este estudio se emplearon secuencias de la proteína NCC de dos especies: humano (*Homo sapiens*) y anguila (*Anguilla sp.*) (Fig.3.1 C). La estructura de NCC humana se obtuvo de la base de datos Protein Data Bank (PDB) con el identificador 7YGO[35], la cual fue determinada

mediante criomicroscopía electrónica (Cryo-EM). Para efectos de este estudio, esta estructura será referida como H-NCC. (Fig.3.3)

Por otro lado, la estructura de NCC de anguila fue obtenida por el grupo de investigadores del Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán" (INCMNSZ) utilizando también el método de Cryo-EM. En este trabajo, esta estructura será referida como  $\beta$ -NCC (Fig.3.4)

Ambas estructuras fueron visualizadas mediante el software PyMOL[36] y analizadas con Mol-Probity[37], una herramienta que permite la evaluación estructural detallada, proporcionando información sobre la calidad de los modelos y su idoneidad para el estudio.

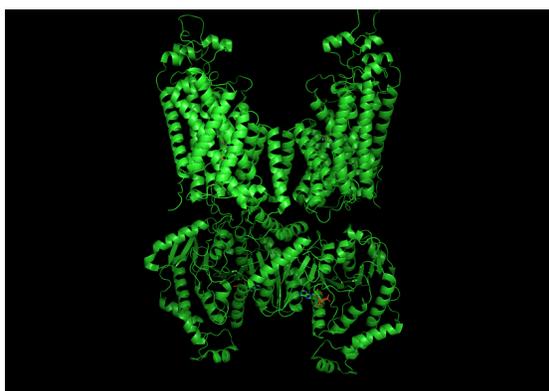


Figura 3.3: HNCC obtenida por medio de CRYO-EM, perteneciente a la familia de transportadores SLC12A3

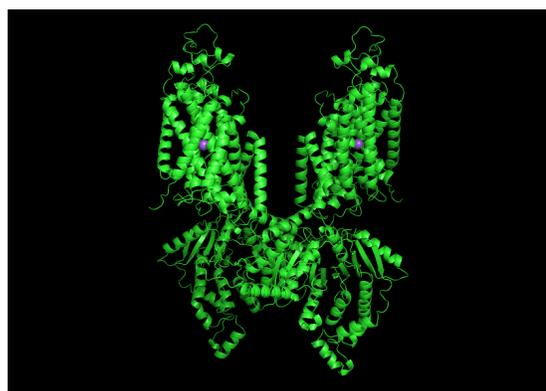


Figura 3.4:  $\beta$ -NCC por medio de CRYO-EM, perteneciente a la familia de transportadores SLC12A3

### 3.2.1. Relajación y validación estructural

La relajación de proteínas (3.1 D) es un proceso en el que se minimiza la energía de la estructura proteica. Este fenómeno está intrínsecamente ligado a las leyes de la termodinámica, específicamente la primera y la segunda ley, y se caracteriza por cambios en la entalpía, la energía libre de Gibbs y la entropía.

La relevancia de la relajación es crucial en el plegamiento de proteínas. Durante este proceso, las

interacciones entre los aminoácidos liberan energía, y el sistema proteico tiende a alcanzar un estado de equilibrio de mínima energía. Este plegamiento está directamente influenciado por la ecuación de Gibbs (3.1), que describe la relación entre la energía libre, la entalpía y la entropía de un sistema:

$$\Delta G = \Delta H - T\Delta S \quad (3.1)$$

donde  $\Delta G$  representa la energía libre de Gibbs,  $\Delta H$  la entalpía,  $T$  la temperatura y  $\Delta S$  la entropía. Dependiendo de los valores de  $\Delta G$  bajo diversas condiciones, se pueden observar diferentes comportamientos en el plegamiento de las proteínas:

- Si  $\Delta G < 0$ , el proceso de plegamiento es espontáneo y la proteína tenderá a adoptar su estructura plegada, lo que indica que la energía liberada es suficiente para superar cualquier barrera energética.
- Si  $\Delta G > 0$ , el proceso no es espontáneo y la proteína permanecerá desplegada a menos que se proporcione energía externa.
- A temperaturas altas, el término  $T\Delta S$  se vuelve más significativo. Dado que  $\Delta S$  generalmente aumenta durante el desplegamiento, un valor alto de  $\Delta S$  puede hacer que  $\Delta G$  sea positivo, favoreciendo el estado desplegado.

Estos principios termodinámicos son esenciales para comprender cómo las proteínas alcanzan su estructura tridimensional funcional.[1]

En el análisis de proteínas, la relajación estructural se realiza de manera cíclica (Fig. 3.1D). Para este propósito, se empleó la herramienta de software abierto **MolProbity**[37], la cual proporciona un análisis detallado de la calidad estructural de las proteínas. El objetivo es obtener parámetros que aseguren la validez y precisión de la estructura modelada. Entre los principales parámetros evaluados se incluyen:

- **Clashscore (Puntuación de colisiones):** Valor deseado  $< 5$  (valores cercanos a 0 son mejores). Mide cuántos átomos están demasiado cerca entre sí, indicando posibles inter-

acciones estéricas desfavorables.

- **Gráfico de Ramachandran (Favorecido):** Valor deseado  $> 95\%$ . Evalúa si los ángulos de torsión de los residuos están dentro de conformaciones permitidas según datos experimentales.
- **C- $\beta$  Deviation (Desviación del carbono beta):** Valor deseado  $< 0,25 \text{ \AA}$  de desviación promedio. Evalúa si los residuos mantienen la geometría esperada en función de su entorno estructural.
- **Rotamer Outliers (Residuos en conformaciones raras):** Valor deseado  $< 1\%$  de residuos en conformaciones desfavorables. Indica si las cadenas laterales adoptan conformaciones improbables.
- **MolProbity Score (Puntuación general):** Valor deseado menor a 1.5. Esta métrica combina Clashscore, Ramachandran y rotámeros en una única puntuación global de calidad estructural.

También se obtuvo el gráfico de Ramachandran, el cual es una herramienta de análisis estructural de proteínas, que permite visualizar la distribución de los ángulos de torsión  $\phi$  (phi) y  $\psi$  (psi) del esqueleto peptídico. Estos ángulos determinan la conformación de la cadena polipeptídica y su estabilidad estructural.

En el gráfico, las regiones sombreadas representan las conformaciones permitidas y energéticamente favorables, basadas en datos experimentales de estructuras proteicas resueltas. Los residuos de la proteína analizada se representan como puntos dentro del gráfico, lo que permite evaluar si su estructura se encuentra dentro de los límites aceptables. La mayoría de los residuos de una proteína correctamente plegada deben ubicarse en las regiones permitidas. La presencia de residuos fuera de estas áreas sugiere posibles errores estructurales o regiones con alta flexibilidad.

### **3.2.2. Identificación del sitio de unión**

Para la identificación de cavidades(3.1E) , se empleó el software de código abierto Cavity-Plus[38], el cual permite localizar las regiones de la proteína donde un ligando puede interactuar de manera óptima, estas regiones fueron seleccionadas apartir de su "druggability", un indicador de la capacidad de interacción de nuestra proteína con fármacos y también su posición. Este software genera coordenadas específicas que posteriormente se utilizan en el proceso de acoplamiento molecular (docking).

### **3.2.3. Podado de proteína**

El podado de la proteína (3.1F) se llevó a cabo utilizando el software PyMOL[36], con el objetivo de reducir la carga computacional del proceso de docking. Este procedimiento consiste en la eliminación de regiones no relevantes de la cadena proteica , optimizando así la eficiencia del análisis. La selección de las regiones a eliminar se realizó con base en la identificación previa de cavidades mediante el software CavityPlus, lo que permitió conservar únicamente las áreas de interés para la interacción con el ligando.

## **3.3. Analisis de la unión proteína-fármaco**

Una vez preparados tanto el ligando como las secuencias, se llevó a cabo el análisis de la interacción proteína-fármaco (3.1G) utilizando dos herramientas computacionales. La primera fue AutoDock[39], un software ampliamente utilizado para predecir la unión de moléculas pequeñas a un receptor. La segunda herramienta empleada fue ROSIE[22], la cual cuenta con su propio algoritmo para realizar acoplamientos moleculares (docking) entre el ligando y la proteína.

En el caso de AutoDock[39], se ejecuta previamente el algoritmo Gridbox, que define la región de búsqueda de la unión del ligando, desempeñando una función similar a la realizada por CavityPlus. Posteriormente, se generan los resultados del acoplamiento. Por otro lado, ROSIE[22] requiere como parámetros de entrada el ligando, la proteína, las coordenadas de la región de interacción y el número de cónformeros, que determina la cantidad de veces que se repetirá el

proceso de docking.

En el desarrollo de este análisis, se manejan una serie de conceptos clave que nos permiten comprender las interacciones entre el ligando y el receptor. Estos conceptos son fundamentales para evaluar la afinidad de unión y la eficiencia del ligando en su interacción con la proteína objetivo. A continuación, se presentan los principales parámetros utilizados en este estudio:

- **Energía de unión:** Este parámetro representa la cantidad total de energía que se libera cuando el ligando se une al receptor. Una mayor energía de unión indica una interacción más fuerte entre el ligando y el receptor, lo cual es crucial para la efectividad de la unión.
- **Eficiencia del ligando:** La eficiencia del ligando se calcula como la relación entre la energía de unión y el número de átomos no hidrógeno presentes en la estructura del ligando. Este índice proporciona una medida de la eficiencia del ligando en términos de su tamaño y su capacidad para interactuar con el receptor de manera efectiva. Una mayor eficiencia sugiere que el ligando se ajusta mejor al receptor con un menor número de átomos.
- **Constante de inhibición (K<sub>i</sub>):** La constante de inhibición es un valor calculado a partir de la energía de unión, y se utiliza para medir la concentración de ligando necesaria para inhibir la actividad del receptor en un 50%. Este valor es esencial para evaluar la potencia del ligando como inhibidor.
- **Energía interna del ligando:** Se refiere a la energía inherente del ligando en su conformación actual. Esta energía incluye tanto la energía torsional como las interacciones intramoleculares del ligando, y es fundamental para comprender su estabilidad y flexibilidad en la interacción con el receptor.
- **Energía van der Waals y enlaces de hidrógeno:** Este concepto engloba las interacciones atractivas entre el ligando y el receptor, como las fuerzas van der Waals y los enlaces de hidrógeno. Estas interacciones juegan un papel clave en la formación de la unión entre ambos, afectando la especificidad y estabilidad de la interacción.

- **Energía electrostática:** La interacción electrostática se refiere a la atracción o repulsión entre cargas opuestas presentes en el ligando y el receptor. Este tipo de interacción es fundamental para la afinidad entre moléculas cargadas, ya que afecta directamente a la fuerza de la unión.
- **Energía torsional:** La energía torsional está asociada con la flexibilidad del ligando. Se refiere a la cantidad de energía requerida para permitir que el ligando cambie su conformación, lo que es relevante para evaluar su capacidad de adaptación al receptor.
- **Enlace de hidrógeno:** Los enlaces de hidrógeno son interacciones atractivas que tienen un papel importante en la especificidad y estabilidad del complejo ligando-receptor. Estos enlaces contribuyen a la formación de estructuras estables entre el ligando y el receptor, facilitando una unión eficaz.

Por otro lado, el **Score Total** es una métrica utilizada para medir la calidad global del acoplamiento entre el ligando y la proteína. Este valor es esencial para evaluar la eficacia y la estabilidad del complejo formado y está determinado por la suma de varias interacciones energéticas entre el ligando y el receptor. A continuación, se describen los principales componentes que contribuyen a la determinación del Score Total:

- **Interacciones de Van der Waals:** Estas interacciones pueden ser tanto atractivas como repulsivas y son cruciales para la formación de la estructura final del complejo. Las interacciones de Van der Waals contribuyen significativamente a la estabilidad de la unión entre el ligando y el receptor.
- **Interacciones electrostáticas:** Las interacciones electrostáticas entre los átomos cargados de la proteína y el ligando afectan directamente la afinidad de la unión. Estas interacciones son esenciales para estabilizar el complejo ligando-receptor.
- **Energía de solvatación:** Este parámetro se refiere a la energía liberada o absorbida cuando un soluto, como el ligando, se disuelve en un solvente. En este contexto, refleja el costo energético asociado con la desolvatación del ligando cuando se une a la proteína, lo cual es un factor importante a considerar para la formación eficiente del complejo.

- **Puentes de hidrógeno:** Los puentes de hidrógeno pueden formarse de dos formas:
  - Entre la cadena principal y la cadena lateral de la proteína.
  - Entre las cadenas laterales de la proteína.

Estas interacciones son esenciales para proporcionar especificidad y estabilidad a la unión entre el ligando y el receptor.

- **Ruptura de la cadena:** Se penaliza cualquier ruptura o interrupción en la cadena de la proteína, ya que esto puede comprometer la estructura del complejo y reducir la efectividad de la unión.
- **Restricciones de coordenadas:** Este parámetro evalúa las penalizaciones por desviaciones en las posiciones atómicas del ligando en relación con las coordenadas deseadas. El cumplimiento de estas restricciones es crucial para asegurar que el ligando se acomode de manera óptima en el receptor.
- **Restricciones de ángulos torsionales:** Las desviaciones en los ángulos torsionales del ligando también son penalizadas. Estos ángulos son importantes para mantener la conformación correcta del ligando en el complejo con el receptor.
- **Ángulo omega:** Este parámetro refleja las penalizaciones por desviaciones en los ángulos de enlace peptídico. Estos ángulos son importantes para mantener la estabilidad estructural del complejo.
- **Rama:** La evaluación de la geometría de los ángulos de torsión de la cadena principal de la proteína se realiza utilizando el gráfico de Ramachandran. Este análisis permite asegurar que los ángulos torsionales se encuentren dentro de los valores permitidos para mantener la estructura tridimensional estable de la proteína.

# Capítulo 4

## Resultados

### 4.1. Analisis Resultados

De acuerdo con la metodología presentada en la Figura 3.1, se obtuvieron los siguientes resultados referentes a distintos pasos.

Los datos presentados en la Tabla 4.1 corresponden al análisis posterior a la etapa de relajación (3.1D), en la que se comparó la calidad estructural de las proteínas HNCC y  $\beta$ -NCC mediante MolProbity[37]. Los resultados indican que ambas proteínas cumplen con los estándares generales de calidad; sin embargo, se identificaron diferencias significativas en parámetros específicos.

La estructura de la proteína de anguila presenta una mejor conformación en términos de rotámeros favorecidos y outliers de Ramachandran. En contraste, la proteína humana exhibe un mayor número de desviaciones estructurales, particularmente en las métricas de CaBLAM y geometría del carbono alfa, lo que sugiere la presencia de regiones menos óptimas en su estructura.

En conclusión, ambas secuencias se encuentran dentro de los rangos esperados, lo que valida su calidad estructural y permite avanzar a la siguiente etapa de la metodología.

A su vez también se muestran sus gráficos de Ramachandra (figs. 4.1) obtenidos en el mismo software mencionado anteriormente, en los graficos se puede observar que la mayoría de las

Tabla 4.1: Comparación de resultados Molprobit NCC entre anguila y humano

Geometría de la Proteína	Anguila			Humano		
	Valor	Porcentaje	Objetivo	Valor	Porcentaje	Objetivo
Rotámetros pobres	0	0.00 %	<0.3 %	32	2.21 %	<0.3 %
Rotámetros favorecidos	1523	99.67 %	>98 %	1395	96.14 %	>98 %
Outliers de Ramachandran	0	0.00 %	<0.05 %	0	0.00 %	<0.05 %
Ramachandran favorecidos	1751	98.48 %	>98 %	1626	95.31 %	>98 %
Puntuación Z de distribución Rama	$0.14 \pm 0.18$		$\text{abs}(Z) < 2$	$-1.79 \pm 0.19$		$\text{abs}(Z) < 2$
Desviaciones $C\beta > 0.25\text{\AA}$	0	0.00 %	0	1	0.06 %	0
Enlaces defectuosos	2 / 14230	0.01 %	0 %	10 / 13896	0.07 %	0 %
Ángulos defectuosos	10 / 19350	0.05 %	<0.1 %	20 / 18890	0.11 %	<0.1 %
<b>Omegas de Péptidos</b>						
Prolinas Cis	0 / 78	0.00 %	$\leq 1$ por cadena, <5 %	0 / 75	0.00 %	$\leq 1$ por cadena, <5 %
<b>Criterios de Baja Resolución</b>						
Outliers CaBLAM	15	0.8 %	<1.0 %	54	3.2 %	<1.0 %
Outliers de Geometría CA	9	0.51 %	<0.5 %	14	0.83 %	<0.5 %
<b>Validaciones Adicionales</b>						
Outliers de Volumen Quiral	0 / 2228			0 / 2150		

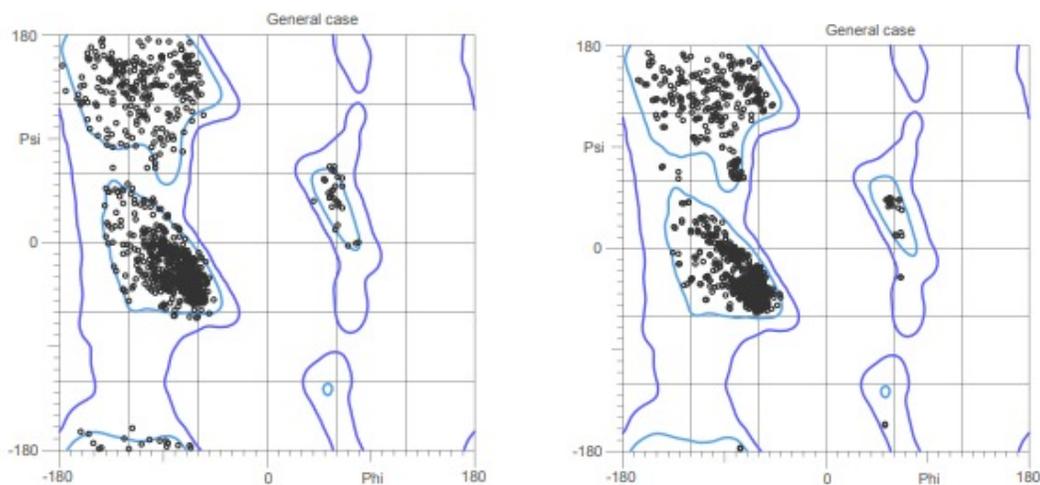


Figura 4.1: Podemos observar los gráficos de Ramachandra obtenidos, del lado izquierdo tenemos el correspondiente a HNCC y del lado derecho  $\beta$ -NCC.

representaciones se encuentran dentro de lo permitido y energéticamente favorable.

Las Figuras 4.2 presentan los resultados obtenidos mediante Cavity Plus[38], donde se visualizan las cavidades identificadas en HNCC y  $\beta$ -NCC. El HNCC muestra una alta "druggability", lo que sugiere una interacción favorable con el ligando. En contraste, el valor de "druggability" de la cavidad en  $\beta$ -NCC es menor, aunque se observa una posición similar a la del HNCC. Esta coincidencia posicional se confirma mediante las coordenadas detalladas en la Tabla 4.2, las cuales son esenciales para un análisis funcional preciso.

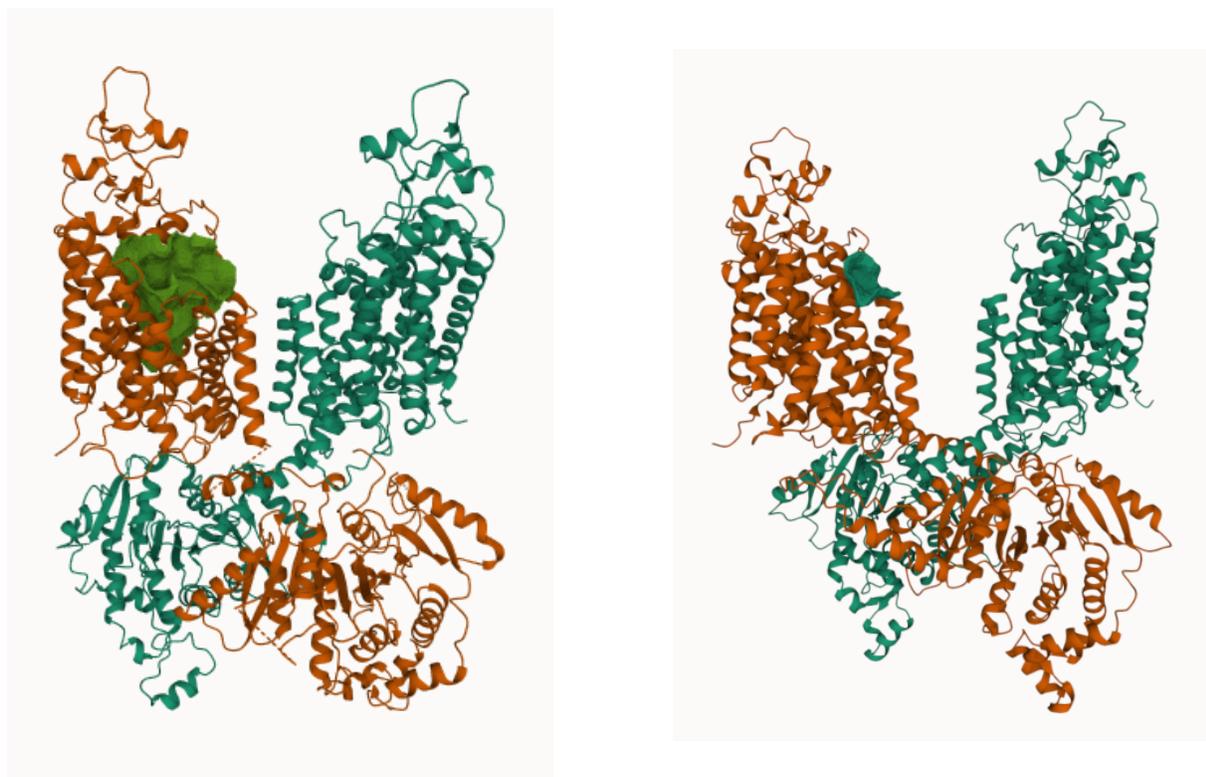


Figura 4.2: Cavidades en HNCC(izquierda) y  $\beta$ -NCC(derecha) en Cavity plus, donde en color verde podemos observar la cavidad presente en la cadena B.

Los resultados del podado 3.1(F) se pueden observar en las figuras 4.3 que eliminamos completamente la cadena A para así solo quedarnos con nuestra cadena B.

Los resultados obtenidos en el analisis de la unión proteína-fármaco mediante el software Au-

Tabla 4.2: Coordenadas obtenidas de la cavidad en HNCC y  $\beta$ -NCC.

Coordenada	HNCC	$\beta$ -NCC
X	131.25	-16.5
Y	139.25	19.5
Z	121	28.5

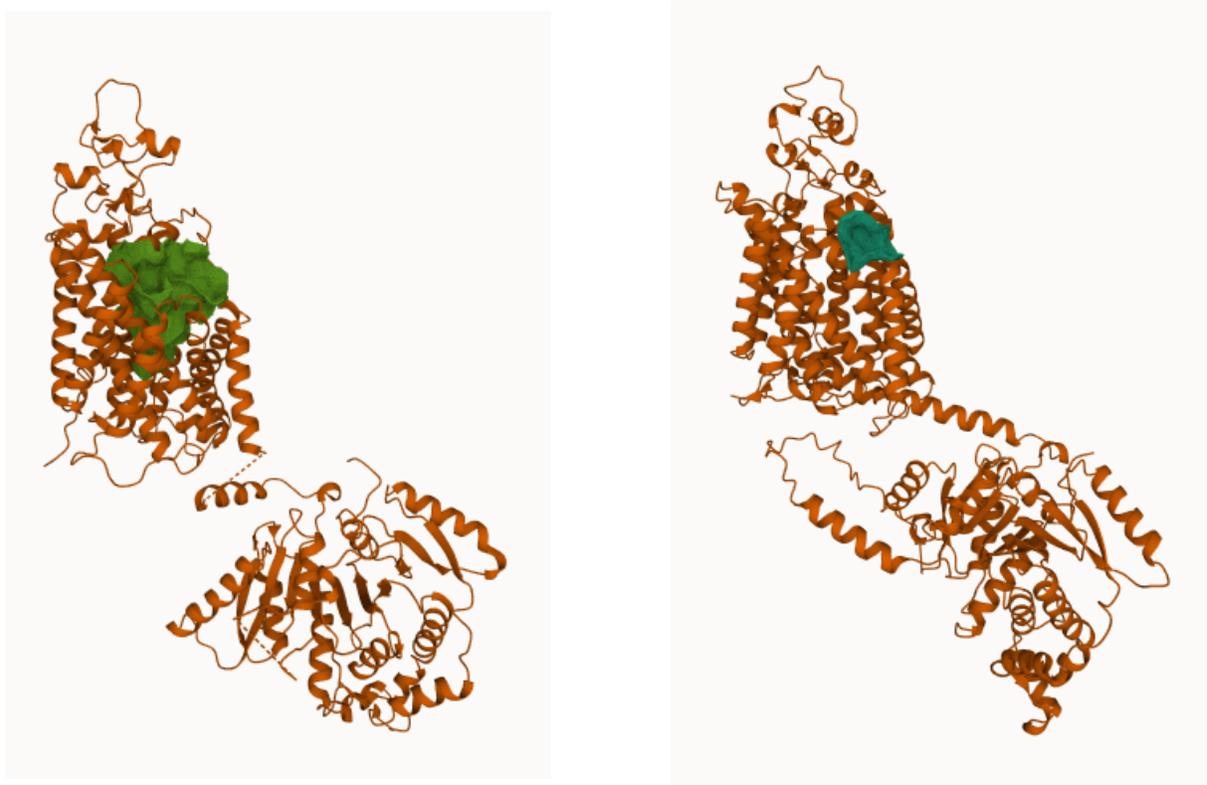


Figura 4.3: Podado de proteínas correspondiente, con el objetivo de disminuir el cómputo, se observa del lado izquierdo HNCC y del lado derecho  $\beta$ NCC

toDock[39] se presentan en la Tabla 4.3. Los datos muestran que el NCC humano exhibe una interacción más fuerte y estable en comparación con el NCC de anguila, con una energía de unión más negativa (-9.09 kcal/mol vs. -5.04 kcal/mol) y una mayor eficiencia del ligando (0.36 vs. 0.22).

Si bien la constante de inhibición es similar en ambas especies, la energía interna del ligando es mayor en humanos (10.88 kcal/mol), lo que sugiere una estructura más estable. Además, la energía de disolución Van der Waals y los enlaces de hidrógeno presentan valores positivos en humanos (10.7 kcal/mol) y negativos en anguilas (-6.84 kcal/mol), lo que indica que las interacciones no covalentes son más favorables en humanos.

Por otro lado, se observan diferencias en la energía electrostática (-0.18 kcal/mol en humanos y 0.58 kcal/mol en anguilas) y en la energía torsional (1.79 kcal/mol en humanos y 1.29 kcal/mol en anguilas), lo que podría influir en la estabilidad estructural del complejo ligando-proteína.

Tabla 4.3: Comparación de resultados entre HNCC y  $\beta$ NCC.

	HNCC	$\beta$ NCC
Energía de unión (Kcal/mol)	-9.09	-5.04
Eficiencia del Ligando	0.36	0.22
Constante de inhibición (nM)	217.8	203.41
Energía interna del ligando (Kcal/mol)	10.88	8.94
E. de Disolución Van der Waals y Enlaces de Hidrógeno (Kcal/mol)	10.7	-6.84
Energía electrostática (Kcal/mol)	-0.18	0.58
Energía Torsional (Kcal/mol)	1.79	1.29

Como se mencionó previamente, se empleó el software de código abierto Rosie[22], el cual utiliza métodos estadísticos para generar un conjunto de resultados. A partir de estos, se seleccionaron los dos modelos más óptimos: HNCC (proteína LG 0671) y  $\beta$ NCC (proteína LG 0003).

Una de las principales ventajas de Rosie es su capacidad para proporcionar el archivo PDB del complejo proteína-ligando, lo que facilita tanto la visualización como el análisis estructural de las interacciones moleculares (Figuras 4.4 y 4.5).

Tabla 4.4: Comparación de resultados entre HNCC y  $\beta$ NCC.

	HNCC	$\beta$ NCC
Score Total	-10.091	-1897.801
Interacciones de Van der Waals	-1733.834	-1963.464
Interacciones Repulsivas de Van der Waals	223.988	181.361
Interacciones Electrostáticas	-37.796	-96.817
Energía de Solvatación	761.999	845.829
Puentes de H entre la C. Principal y la C. Lateral	-11.48	-39.52
Puentes de Hidrógeno entre Cadenas Laterales	-14.803	-36.09
Ruptura de la Cadena	0	0
Restricciones de Coordenadas	6.009	0.897
Restricciones de Ángulos Torsionales	0	0.002
Ángulo Omega	66.512	18.912
Rama	4.658	-18.224

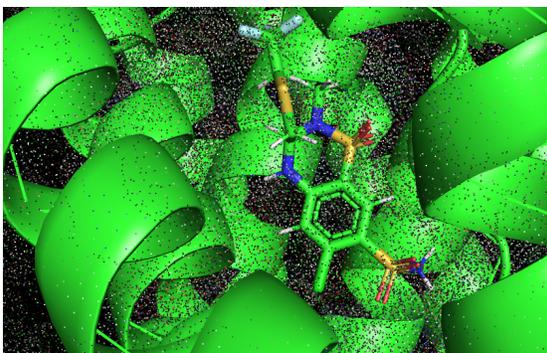
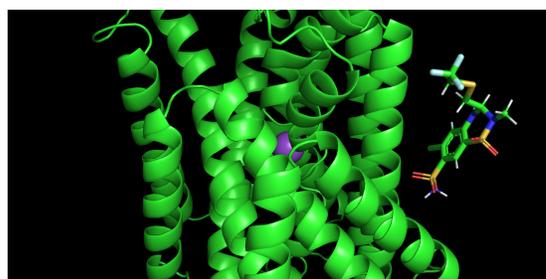


Figura 4.4: HNCC

Figura 4.5:  $\beta$ NCC

Estructuras tridimensionales del complejo proteína-ligando generadas con Rosie, mostrando la interacción del ligando con (fig.4.4) HNCC (proteína LG 0671) y (fig.4.5)  $\beta$ NCC (proteína LG 0003). Los modelos obtenidos en formato PDB permiten un análisis detallado de las interacciones moleculares y la estabilidad conformacional del complejo.

Los resultados obtenidos mediante el docking molecular realizado con Rosie (Tabla 4.4) revelan una notable disparidad en el Score Total. En particular, el  $\beta$ NCC presenta un valor significativamente más negativo (-1897.801) en comparación con el HNCC (-10.091), lo que sugiere que la interacción tiazida- $\beta$ NCC es mucho más termodinámicamente favorable. Esta diferencia de tres órdenes de magnitud resalta la mayor afinidad de la tiazida por el  $\beta$ NCC.

Las interacciones de Van der Waals también favorecen al  $\beta$ NCC, con un valor de -1963.464 frente a -1733.834 para el HNCC. Este hallazgo indica una mayor complementariedad estérica entre la tiazida y el  $\beta$ NCC, lo que refuerza la estabilidad del complejo en términos de la proximidad y las interacciones no covalentes entre las moléculas. Además, las interacciones electroestáticas refuerzan esta tendencia, con un valor de -96.817 para el  $\beta$ NCC frente a -37.796 para el HNCC. Esta diferencia implica una disposición más favorable de las cargas para la interacción entre la tiazida y el  $\beta$ NCC.

Por otro lado, la energía de solvatación, aunque desfavorable en ambos casos, es ligeramente superior para  $\beta$ NCC (845.829 frente a 761.999 para HNCC). Este resultado podría estar relacionado con diferencias en la hidrofobicidad de las cavidades de unión, lo cual influiría en el desplazamiento del agua durante la formación del complejo.

En cuanto a los puentes de hidrógeno, una característica crucial para la estabilidad de los complejos, el  $\beta$ NCC muestra una mayor tendencia a formar estos enlaces tanto entre la cadena principal y lateral (-39.52 frente a -11.48 para HNCC) como entre cadenas laterales (-36.09 frente a -14.803 para HNCC). Este comportamiento sugiere que el complejo tiazida- $\beta$ NCC presenta una mayor especificidad y estabilidad, lo que podría ser un factor determinante en la mayor afinidad del  $\beta$ NCC por la tiazida.

Respecto a las restricciones de coordenadas, estas son significativamente mayores en HNCC (6.009 frente a 0.897 para  $\beta$ NCC), lo que podría indicar una mayor rigidez estructural en la región de unión del HNCC. En contraste, las restricciones de ángulos torsionales son ligeramente mayores en  $\beta$ NCC (0.002 frente a 0 en HNCC), lo que sugiere una mayor flexibilidad conformacional en ciertas regiones del  $\beta$ NCC, lo que podría facilitar su interacción con la tiazida.

Finalmente, los parámetros de ángulo omega y rama muestran diferencias significativas entre los dos transportadores, lo que refleja variaciones en la geometría y la conformación de los residuos de aminoácidos en la región de unión. Estas diferencias estructurales podrían ser determinantes en la selectividad y la afinidad de la tiazida por cada variante del NCC.

# Capítulo 5

## Conclusiones y trabajo a futuro

### 5.1. Conclusiones

El objetivo planteado al inicio de esta tesis fue simular la función del cotransportador de Na<sup>+</sup>:Cl<sup>-</sup> sensible a tiazidas (NCC) utilizando técnicas de inteligencia artificial (IA). Para ello, se llevó a cabo un análisis estructural de las proteínas mediante el software Molprobit, donde ambas proteínas presentaron resultados favorables. Posteriormente, se siguieron los pasos establecidos en la metodología, utilizando herramientas de relajación como Avogadro y Rosie Relax, y otros programas como Cavity Plus y PyMOL para la visualización de cavidades y la representación general de las proteínas. Finalmente, se realizó la simulación de la interacción ligando-proteína utilizando AutoDock y Rosie.

A partir del análisis comparativo de las conformaciones HNCC y  $\beta$ NCC, se concluye que HNCC presenta una mejor interacción con la tiazida, lo que se evidencia en su mayor estabilidad energética y en la naturaleza de sus interacciones con el ligando. Los resultados muestran que HNCC tiene una energía de unión más negativa (-9.09 Kcal/mol) en comparación con  $\beta$ NCC (-5.04 Kcal/mol), lo que indica una mayor estabilidad en la unión con el ligando. Además, la eficiencia del ligando en HNCC (0.36 vs. 0.22) sugiere una interacción más optimizada. Aunque la constante de inhibición es similar en ambas conformaciones (217.8 nM en HNCC y 203.41

nM en  $\beta$ NCC), la diferencia en energía de unión sigue favoreciendo a HNCC.

En cuanto a la energía electrostática, HNCC tiene un valor negativo (-0.18 Kcal/mol), mientras que  $\beta$ NCC tiene un valor positivo (0.58 Kcal/mol), lo que indica que las interacciones de carga son más favorables en HNCC. La energía torsional es menor en  $\beta$ NCC (1.29 Kcal/mol vs. 1.79 Kcal/mol en HNCC), lo que sugiere una mayor flexibilidad en esta conformación, pero no lo suficiente como para superar las ventajas energéticas y estructurales de HNCC.

Las interacciones de Van der Waals también favorecen a HNCC, que presenta valores más favorables (-1733.834 vs. -1963.464 en  $\beta$ NCC) y una mayor repulsión de Van der Waals (223.988 vs. 181.361 en  $\beta$ NCC), lo que sugiere una mejor estabilidad en el sitio de unión. Además, los puentes de hidrógeno entre la cadena principal y lateral son más favorables en HNCC (-11.48 vs. -39.52 en  $\beta$ NCC), lo que indica una mejor fijación del ligando. Aunque los puentes de hidrógeno entre cadenas laterales son más estables en  $\beta$ NCC (-8.225 vs. -14.803 en HNCC), esta ventaja no compensa las diferencias generales en estabilidad de unión. Finalmente, la energía de solvatación es más baja en HNCC (761.999 vs. 845.829 en  $\beta$ NCC), lo que indica que la interacción con el medio acuoso es más favorable en  $\beta$ NCC, aunque esta diferencia no es suficiente para alterar la conclusión general.

En resumen, HNCC es la mejor conformación para la interacción con la tiazida, debido a su mayor estabilidad energética, mejor eficiencia del ligando y un perfil de interacciones más favorable en comparación con  $\beta$ NCC, a pesar de que  $\beta$ NCC presenta algunas ventajas en flexibilidad y solvatación.

## 5.2. Trabajo a futuro

Como trabajo a futuro, se puede implementar el uso de otros softwares que puedan aportar aún más a este tipo de investigación. Uno de los programas que podría ser de vital ayuda es *Rosetta*, el cual, debido a las limitaciones computacionales del trabajo, no se pudo utilizar. Este software es reconocido por su capacidad para predecir y diseñar estructuras de proteínas, lo que podría enriquecer considerablemente los resultados obtenidos.

Además, se podría considerar la incorporación de modelos generados completamente por inteligencia artificial, como los que provienen de *AlphaFold*, un modelo avanzado que ha demostrado ser eficaz para predecir la estructura tridimensional de proteínas a partir de su secuencia aminoácídica. La integración de estos modelos podría proporcionar una mayor precisión y detalle en las simulaciones realizadas.

Finalmente, es importante señalar que este trabajo puede sentar las bases para una futura validación experimental del proyecto. Se plantea como un objetivo a largo plazo realizar experimentos *in vivo* o *in vitro* para confirmar si los resultados obtenidos de manera *in silico* son acertados en la realidad. Este enfoque experimental podría brindar un gran apoyo para futuras generaciones de investigadores, permitiendo una mejor comprensión de los procesos biológicos y abriendo nuevas posibilidades para el desarrollo de terapias y tratamientos innovadores.

# Bibliografía

- [1] E. J. Stollar y D. P. Smith, *Uncovering protein structure*, oct. de 2020. DOI: 10.1042/EBC20190042.
- [2] T. Ongtanasup, A. Mazumder, A. Dwivedi y K. Eawsakul, “Homology Modeling, Molecular Docking, Molecular Dynamic Simulation, and Drug-Likeness of the Modified Alpha-Mangostin against the  $\alpha$ -Tubulin Protein of *Acanthamoeba Keratitis*,” *Molecules*, vol. 27, pág. 6338, 19 sep. de 2022, ISSN: 1420-3049. DOI: 10.3390/molecules27196338.
- [3] J. Selvaraj, “Molecular modeling of cornulin (CRNN) for docking with phytocompounds from *Justicia adhatoda* L.,” *Bioinformation*, vol. 17, págs. 200-205, 1 ene. de 2021, ISSN: 09738894. DOI: 10.6026/97320630017200.
- [4] E. Moreno, D. Pacheco-Alvarez y E. R. Argáiz, “Aspectos estructurales, funcionales y patológicos del cotransportador de NaCl sensible a tiazidas,” *Revista de Investigación Clínica*, vol. 66, págs. 559-567, 6 dic. de 2014.
- [5] H. Xiong, L. Han, Y. Wang y P. Chai, *Evaluating the Reliability of AlphaFold 2 for Unknown Complex Structures with Deep Learning*, jul. de 2022. DOI: 10.1101/2022.07.08.499384.
- [6] Satishlokhande, “What is the story of AlphaFold?” *Medium*, mayo de 2024.
- [7] J. M. Juárez Sánchez e Y. Ramírez Venancio. “El Nobel de Química 2024 reconoce desarrollos en el área de diseño y predicción de estructuras de proteínas.” Facultad de Química, UNAM. Consultado el 28 de abril de 2025. (2024), dirección: <https://quimica.unam.mx/nobel-quimica-2024-proteinas/>.

- [8] T. Sanderson, M. L. Bileschi, D. Belanger y L. J. Colwell, “ProteInfer, deep neural networks for protein functional inference,” *eLife*, vol. 12, feb. de 2023, ISSN: 2050-084X. DOI: 10.7554/eLife.80942.
- [9] B. Lin, X. Luo, Y. Liu y X. Jin, “A comprehensive review and comparison of existing computational methods for protein function prediction,” *Briefings in Bioinformatics*, vol. 25, 4 mayo de 2024, ISSN: 1467-5463. DOI: 10.1093/bib/bbae289.
- [10] R. F. Condori y A. A. V. Aguilar, “Estudios In Silico, simulando vida en un entorno virtual,” *Gaceta Medica Boliviana*, vol. 44, 2 2021, ISSN: 1012-2966. DOI: 10.47993/gmb.v44i2.263.
- [11] N. Pisanti, “Algorithms Foundations,” en *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019, págs. 1-4. DOI: 10.1016/B978-0-12-809633-8.20315-4.
- [12] T. Bepler y B. Berger, “Learning the protein language: Evolution, structure, and function,” *Cell Systems*, vol. 12, 654-669.e3, 6 jun. de 2021, ISSN: 24054712. DOI: 10.1016/j.cels.2021.05.017.
- [13] A. C. da Silva, B. R. P. Nunes y J. Xavier, “Unlocking the Secrets of Life: AI Protein Models Demystified,” *BIOINFO*, vol. 3, pág. 24, 1 sep. de 2023. DOI: 10.51780/bioinfo-03-24.
- [14] M. Kulmanov, M. A. Khan y R. Hoehndorf, “DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier,” *Bioinformatics*, vol. 34, págs. 660-668, 4 feb. de 2018, ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx624.
- [15] Z. Lin, H. Akin, R. Rao et al., “Evolutionary-scale prediction of atomic-level protein structure with a language model,” *Science*, vol. 379, págs. 1123-1130, 6637 mar. de 2023, ISSN: 0036-8075. DOI: 10.1126/science.ade2574.
- [16] Z. Lin, H. Akin, R. Rao et al., *Evolutionary-scale prediction of atomic level protein structure with a language model*, jul. de 2022. DOI: 10.1101/2022.07.20.500902.

- [17] J. Dauparas, I. Anishchenko, N. Bennett et al., “Robust deep learning–based protein sequence design using ProteinMPNN,” *Science*, vol. 378, págs. 49-56, 6615 oct. de 2022, ISSN: 0036-8075. DOI: 10.1126/science.add2187.
- [18] N. Ferruz, S. Schmidt y B. Höcker, “ProtGPT2 is a deep unsupervised language model for protein design,” *Nature Communications*, vol. 13, pág. 4348, 1 jul. de 2022, ISSN: 2041-1723. DOI: 10.1038/s41467-022-32007-7.
- [19] D. Baker, *History of Rosetta*, 1988.
- [20] A. Waterhouse, M. Bertoni, S. Bienert et al., “SWISS-MODEL: homology modelling of protein structures and complexes,” *Nucleic Acids Research*, vol. 46, W296-W303, W1 jul. de 2018, ISSN: 0305-1048. DOI: 10.1093/nar/gky427.
- [21] W. G. B. Paucara y R. E. G. Torrez, “Acomplamiento molecular:: criterios prácticos para la selección de ligandos biológicamente activos e identificación de nuevos blancos terapéuticos,” *Revista CON-CIENCIA*, vol. 7, 2 2019, ISSN: 2310-0265.
- [22] S. Lyskov, F.-C. Chou, S. Conchúir et al., “Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE),” *PLoS ONE*, vol. 8, e63906, 5 mayo de 2013, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0063906.
- [23] G. Tradigo, F. Rondinelli y G. Pollastri, “Algorithms for Structure Comparison and Analysis: Docking,” en *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019, págs. 77-80. DOI: 10.1016/B978-0-12-809633-8.20485-8.
- [24] D. Mariano y B. Silva, “Dinâmica molecular: como mostrar um filme completo em uma folha de papel?” En *BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional*. 2021. DOI: 10.51780/978-6-599-275326-11.
- [25] Contreras-Moreira, “Modelado comparativo de proteínas. Homology modelling,” *Laboratorio de biología computacional*, vol. 8, 40 2011.
- [26] M. Dujana, *Learn Bioinformatics From Scratch (Theory Practical)*, feb. de 2025. dirección: [https://www.udemy.com/share/104AiI3@h\\_YQ9y0Tzv-gW4mlqurRtS89r5xEoAZCOY\\_ECQlr000SzXHOauqE9zi5aYmbGB\\_Pzg==/](https://www.udemy.com/share/104AiI3@h_YQ9y0Tzv-gW4mlqurRtS89r5xEoAZCOY_ECQlr000SzXHOauqE9zi5aYmbGB_Pzg==/).
- [27] M. V. L. Guillén, *ESTRUCTURA Y PROPIEDADES DE LAS PROTEÍNAS*. Universitat Politècnica de Valencia, 2023, págs. 1-17.

- [28] K. K. Yang, Z. Wu, C. N. Bedbrook y F. H. Arnold, “Learned protein embeddings for machine learning,” *Bioinformatics*, vol. 34, 15 2018, ISSN: 14602059. DOI: 10.1093/bioinformatics/bty178.
- [29] D. Lupask, *Trabajo propio, Dominio público*, <https://commons.wikimedia.org/w/index.php?curid=8932281>.
- [30] Lecturio, *Thiazide Diuretics*, 2024. dirección: <https://app.lecturio.com/#/article/3189>.
- [31] K. L. Rios, “Efecto de Diuréticos Tiazida sobre NCC beta de Anguila Europea,” Tesis doct., Universidad Nacional Autonoma de México, 2016, págs. 13-14.
- [32] J. E. Cruz-Aranda, “Fármacos diuréticos: alteraciones metabólicas y cardiovasculares en el adulto mayor,” *Medicina Interna de México*, vol. 34, 4 jul. de 2018, ISSN: 01864866.
- [33] National Center for Biotechnology Information, *PubChem Compound Summary for CID 4870, Polythiazide*, Retrieved March 30, 2025 from <https://pubchem.ncbi.nlm.nih.gov/compound/Renese>, 2025.
- [34] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek y G. R. Hutchison, *Avogadro: An advanced semantic chemical editor, visualization, and analysis platform*, Version 1.XX, <http://avogadro.cc/>, 2012.
- [35] J. Nan, Y. Yuan, X. Yang et al., “Cryo-EM structure of the human sodium-chloride co-transporter NCC,” *Science Advances*, vol. 8, 45 nov. de 2022, ISSN: 2375-2548. DOI: 10.1126/sciadv.add7176.
- [36] Schrödinger, LLC y Warren DeLano, *PyMOL*, ver. 3.1, Updated March 26th, 2025, 2025. dirección: <http://www.pymol.org/pymol>.
- [37] C. J. Williams, J. J. Headd, N. W. Moriarty et al., “MolProbity: More and better reference data for improved all-atom structure validation,” *Protein Science*, vol. 27, págs. 293-315, 1 ene. de 2018, ISSN: 0961-8368. DOI: 10.1002/pro.3330.
- [38] S. Wang, J. Xie, J. Pei y 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, pág. 168 141, 14 jul. de 2023, ISSN: 00222836. DOI: 10.1016/j.jmb.2023.168141.

- [39] G. M. Morris, R. Huey, W. Lindstrom et al., “Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility,” *Journal of Computational Chemistry*, vol. 16, págs. 2785-91, 2009. DOI: 10.1002/jcc.21256.