



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS QUÍMICAS Y CENTRO DE QUÍMICA-
INSTITUTO DE CIENCIAS

POSGRADO DE CIENCIAS QUÍMICAS

INVESTIGACIÓN COMPUTACIONAL DE AGENTES CAPACES DE MODULAR
AL DOMINIO DE ENLACE A NUCLEÓTIDOS CÍCLICOS (CNBD) DE LOS
CANALES HCN (CANALES REGULADOS POR NUCLEÓTIDOS CÍCLICOS
ACTIVADOS POR HIPERPOLARIZACIÓN)

ENERO DE 2025

TESIS

PARA OBTENER EL GRADO DE: MAESTRÍA EN QUÍMICA

ESPECIALIDAD EN FÍSICOQUÍMICA

PRESENTA:

OMAR MÉNDEZ FERNÁNDEZ

DIRECTOR DE TESIS:

JOSÉ MANUEL PÉREZ AGUILAR

ASESOR DE TESIS:

FRANCISCO JAVIER MELÉNDEZ BUSTAMANTE



BUAP

Oficio PCQ/438/2024

Mtro. Alfredo Avendaño Arenaza
Director General de Bibliotecas
Presente

A través de este medio me dirijo a usted para informarle que la tesis con título **Investigación computacional de agentes capaces de modular al dominio de enlace a nucleótidos cíclicos (CNBD) de los canales HCN**, ha sido avalada bajo los preceptos de integridad académica; por tanto, se autoriza la impresión de tesis del alumno:

Nombre del alumno: **Omar Méndez Fernández**

Matrícula: **222470375**

Facultad: **Facultad de Ciencias Químicas – Instituto de Ciencias**

Programa Educativo: **Maestría en Ciencias Químicas**

Licenciatura o Posgrado: **Posgrado**

Sin más por el momento, agradezco la atención al presente.

ATENTAMENTE
“Pensar bien, para vivir mejor”
H. Puebla de Z., a 06 de diciembre de 2024


Dr. Juan Manuel Solano Altamirano
Secretario Académico




Dr. José Manuel Pérez Aguilar
Director de Tesis


Dr. Francisco Javier Méndez Bustamante
Codirector

Agradecimientos

A Dios (Al-lāh, الله), por permitirme terminar este proyecto tan importante en mi formación.

A mis familiares, especialmente a mi mamá (María Teresa de Jesús Fernández Sánchez), a mi papá (Felipe Méndez Huerta) y a mi hermano (Abraham Méndez Fernández) por el apoyo que me brindaron a lo largo de estos dos años y medio.

A mi director de tesis, el Dr. José Manuel Pérez Aguilar; a mi codirector de tesis, el Dr. Francisco Javier Meléndez Bustamante; a las doctoras Brenda L. Sánchez Gaytán, María Eugenia Castro Sánchez, Norma A. Caballero Concha, integrantes de mi comité tutorial.

A mis compañeros, el doctor Eliud Morales Dávila, el M.C. Juan C. García Hernández y el M.C. Mario F. Salamanca Vera, por su apoyo constante.

Al Laboratorio Nacional de Supercómputo (LNS), y al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) por la beca otorgada.

“El amor es el camino y la vía de nuestro profeta”.

-- Rumi

Resumen

Simulaciones de dinámica molecular (SDM) fueron realizadas para los sistemas CNBD (apo), CNBD/cAMP y CNBD/cGMP con una escala de tiempo mayor a 500 nanosegundos en cada caso. Para ello, tres estructuras representativas (correspondientes a 0, 250 y 500 nanosegundos, respectivamente) fueron sometidas a un algoritmo de propagación de perturbaciones. Los resultados fueron comparados y, tras excluir los valores de tiempo y cadenas para los cuales había una diferencia estadísticamente significativa (según pruebas de ANOVA y de t de Student), se obtuvo un promedio. Hecho lo anterior, fueron escogidos algunos de los valores intensidad de acoplamiento alostérico más importantes (*ACI hotspots*), y luego se definió, para cada sistema, el sitio alostérico más relevante.

Posteriormente se determinó, mediante el software Ohm, cuáles eran las vías alostéricas más probables y los residuos críticos en dichas vías. Se procedió haciendo un análisis de los residuos críticos principales.

Más tarde, se realizó una medición de RMSF para los tres sistemas a lo largo de la trayectoria de 400 a 500 nanosegundos, haciendo énfasis en los resultados que arrojaron diferencias estadísticamente significativas entre el sistema CNBD (*apo*) y los sistemas CNBD/cAMP y CNBD/cGMP.

Se concluyó que es conveniente hacer Simulaciones de Dinámica Molecular antes de emplear el algoritmo de propagación de perturbaciones del software Ohm. Además, se concluyó que la unión a nucleótidos cíclicos conduce a una pérdida de la comunicación alostérica entre el sitio activo y los residuos más próximos al segmento transmembranal, a la vez que ocasiona mayores fluctuaciones en su posición relativa (brinda mayores valores de RMSF), lo cual podría indicar una pérdida en la conectividad entre motivos estructurales.

Índice

Lista de figuras	7
Lista de tablas	8
1. Introducción	9
1.1 Las proteínas	10
1.2 Proteínas membranales y canales iónicos	11
1.3 Nucleótidos cíclicos	13
1.4 Técnicas computacionales	14
2. Marco teórico	17
2.1 Los canales activados por hiperpolarización modulados por nucleótidos cíclicos (HCN).	17
2.2 Comunicación alostérica	20
2.3 El estudio de la comunicación alostérica a través del uso de un algoritmo de propagación de perturbaciones.	21
2.4 Simulaciones de Dinámica Molecular (SDM)	26
3. Hipótesis	30
4. Objetivos	31
4.1 General	31
4.2 Particulares	31
5. Metodología	32
6. Discusión de resultados	34
6.1 Simulaciones de Dinámica Molecular (SDM)	34
6.2 Determinación de los residuos que forman el sitio de unión a ligante	36
6.3 Análisis estadístico de los resultados de Ohm	37
6.4 Comunicación alostérica en el sistema CNBD (apo)	40
6.5 Comunicación alostérica en el sistema CNBD/cAMP	43
6.6 Comunicación alostérica en el sistema CNBD/cGMP	44
6.7 Mediciones de RMSF	49
7. Conclusiones	51
8. Perspectivas	52

9. Bibliografía	53
10. Apéndices	56
10.1 Apéndice 1. Funciones que describen el comportamiento cinético de la corriente de hiperpolarización	56
10.2 Apéndice 2. Función de partición del ensamble isotérmico-isobárico	66
10.3 Apéndice 3. Inferencias bayesianas a partir de las SDM	69
10.4 Apéndice 4. Análisis estadístico	71

Lista de figuras

Figura 1. Estructura de un canal iónico tetramérico (el HCN)	12
Figura 2. Estructura tetramérica de los HCN	13
Figura 3. Algoritmo de Ohm para el cálculo de los valores de Intensidad de Acoplamiento Alostérico (ACI)	25
Figura 4. Algoritmo de Ohm para la determinación de los residuos críticos y de sus importancias.....	26
Figura 5. Estructura homotetramérica del CNBD del HCN/cAMP	34
Figura 6. Valores de RMSD para el sistema CNBD (apo)	35
Figura 7. Valores de RMSD para el sistema CNBD/cAMP	35
Figura 8. Valores de RMSD para el sistema CNBD/cGMP	36
Figura 9. Estructura de los HCN	37
Figura 10. Residuos (a) comunes a los sitios activos, (b) del sitio activo de CNBD/cAMP. (c) del sitio activo de CNBD/cGMP.....	39
Figura 11. CNBD de HCN1 (a) sistema <i>apo</i>, (b) CNBD/cAMP	40
Figura 12. CNBD de HCN1 (a) sistema <i>apo</i>, (b) CNBD/cGMP y (c) CNBD/cGMP (vista alternativa)	45
Figura 13. Valores de RMSF para los diferentes residuos de la hélice 1 en los diferentes sistemas, CNBD (apo), CNBD/cAMP y CNBD/cGMP	50

Lista de tablas

Tabla 1. Diferencias estadísticamente significativas entre medias de ACI	38
Tabla 2. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD (apo).....	41
Tabla 3. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD/cAMP.....	43
Tabla 4. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD/cGMP	46
Tabla 5. Análisis ulterior acerca de los sitios alostéricos más relevantes y otros potenciales sitios alostéricos	47
Tabla 6. Análisis ulterior acerca de los residuos críticos principales	48

1. Introducción

1.1 Las proteínas.

Las proteínas son macromoléculas presentes en los sistemas biológicos y se ha estimado que alrededor del 50% de la masa seca de las células está compuesta por este tipo de biomoléculas. Las proteínas ayudan a las células a llevar a cabo funciones tan diversas como la replicación de ADN, la biosíntesis, el transporte de vesículas para secretar desechos, el transporte de gases como el oxígeno, así como la realización de reacciones químicas mediante las cuales se obtiene energía [1].

Las proteínas se componen por subunidades a las que designamos *residuos aminoácidos*, por tratarse de una secuencia de aminoácidos enlazados mediante un tipo de enlace denominado *enlace peptídico*. Específicamente, se trata de al menos 22 aminoácidos los que pueden estar presentes en las proteínas [2]. Los aminoácidos presentan diferentes características químicas dependiendo de la naturaleza de los átomos que forman su cadena lateral. Por ejemplo, existen aminoácidos hidrofóbicos porque su cadena lateral es alifática, como la glicina (GLY, G), la alanina (ALA, A), prolina (PRO, P), valina (VAL, V), leucina (LEU, L), isoleucina (ILE, I) y metionina (MET, M). Existen otros que son categorizados como aromáticos debido a los grupos que presentan en su cadena lateral, como la fenilalanina (PHE, F), el triptófano (TRP, W) y la tirosina (TYR, Y). Estos dos últimos tiene un carácter más polar que la fenilalanina, a pesar de que su cadena lateral es aromática, e incluso presentan un comportamiento anfifílico (es decir: que prefieren la interfaz entre la membrana y el agua, en vez de sólo el interior de la membrana). Por otra parte, hay también aminoácidos polares cuya cadena lateral no tiene carga, como la serina (SER, S) y la treonina (THR, T), cuyas cadenas laterales incluyen el grupo alcohol, la cisteína (CYS, S), cuya cadena lateral incluye el grupo tiol, la asparagina (ASN, N) y la glutamina (GLN, Q), cuya cadena lateral incluye el grupo amida. Por otra parte, hay otros aminoácidos polares cuya cadena lateral está cargada positivamente a valores de pH neutros, como la lisina (LYS, K), la arginina (ARG, R) y en algunas ocasiones, la histidina (HIS, H). Finalmente, existen aminoácidos que a valores de pH neutro presentan una carga negativa, como el aspartato (ASP, D) y el glutamato (GLU, E) [3].

Existen diferentes niveles estructurales de las proteínas, siendo el primero el que indica la identidad de los diferentes residuos aminoacídicos presentes y denominado *estructura primaria*; el segundo nivel, llamado *estructura secundaria*, se refiere a la disposición espacial relativa de los diferentes residuos y es principalmente determinado por un patrón de puentes de hidrógeno formados por átomos de las cadenas principales. En este contexto, en los años 50, por ejemplo, se postuló la existencia de las *hélices alfa*, que son secuencias de aminoácidos que tienden a enrollarse en espiral de manera periódica, dejando su cadena lateral expuesta hacia el exterior de la espiral. En algunas ocasiones, dichos residuos son capaces de establecer puentes de hidrógeno a través de su cadena lateral expuesta y, con ello, dar estabilidad al segmento; esto último indica que los aminoácidos serina (SER, S), treonina (THR, T), cisteína (CYS, S), asparagina (ASN, N) y glutamina (GLN, Q) pueden jugar un papel fundamental en la estabilización de la estructura de las hélices alfa. Hoy se sabe que, aunque las hélices alfa pueden estar constituidas por menos de 10 aminoácidos, también pueden tener hasta más de 25 aminoácidos; de manera adicional, se sabe que éstas pueden ser lineales, curvas o torcidas dependiendo esto último de varios aspectos, desde la secuencia de aminoácidos hasta distorsiones de los puentes de hidrógeno o del enlace peptídico inducidas por agentes externos [4]. Además de las hélices alfa, existen otros motivos estructurales designados *hojas beta plegadas*. Éstas también tienden a estabilizarse por puentes de hidrógeno, pero se caracterizan por el hecho de que se disponen a modo de hebras de la cadena de aminoácidos. Existen hojas beta plegadas paralelas y antiparalelas, siendo las segundas más estables que las primeras. En las hojas beta plegadas paralelas, los extremos C-terminal y N-terminal de cada hebra tiene la misma orientación; en las hojas beta antiparalelas, ocurre lo contrario: la dirección se va alternando, lo cual confiere mayor estabilidad estructural. Se observa en las hojas beta-plegadas que son sumamente usuales los aminoácidos aromáticos y los aminoácidos ramificados; por ejemplo, por fenilalanina (PHE, F), tirosina (TYR, Y) y triptófano (TRP, W), por una parte, e isoleucina (ILE, I), treonina (THR, T) y valina (VAL, V) por otra parte. Además, se sabe que hay ciertos pares aminoacídicos que tienden a interactuar en mayor medida entre sí en hojas beta antiparalelas; por ejemplo, PHE-PHE, PHE-TYR, GLU-ARG y GLU-LYS [5]. Aparte de hélices alfa y hojas beta-plegadas, otros motivos estructurales presentes en las proteínas son los siguientes: giros y vueltas, caracterizados por mayor flexibilidad estructural y menor estabilidad

también; otras hélices menos estables que las hélices alfa, como las hélices 3(10) y las hélices π .

La estructura terciaria de una proteína, por otra parte, se refiere a la estructura tridimensional de dicha biomacromolécula, resultante del conjunto de interacciones entre residuos, entre las cuales están incluidas no sólo los puentes de hidrógeno, sino también interacciones hidrofóbicas, puentes disulfuro, interacciones electrostáticas, fuerzas de van der Waals débiles, etc. [6].

Cuando una proteína no es homomérica, sino dimérica u oligomérica, sus monómeros establecen sus enlaces con los mismos tipos de interacciones entre residuos antes mencionados (esto es: puentes de hidrógeno, puentes disulfuro, interacciones hidrofóbicas, interacciones electrostáticas y fuerzas de van der Waals). A esto último se le conoce como la estructura cuaternaria de una proteína [6].

1.2 Proteínas membranales y canales iónicos.

Dentro del conjunto de las proteínas, destaca un grupo denominado como *proteínas membranales, transmembrana o integrales de la membrana*. Además de lo que su nombre sugiere, es decir: que son proteínas que atraviesan la membrana plasmática, tienen las siguientes características: representan entre un 20 y un 30% de todas las proteínas, y desempeñan generalmente una función en “la comunicación y el transporte entre las células y el mundo exterior” [7].

Desde el siglo XIX, se especulaba que debían existir en las células *canales* que ayudaran el transporte de determinadas sustancias hacia el interior o exterior de la célula. En 1952, el trabajo de Hodgkin y Huxley propició que avanzara la aceptación de la idea de canales a través de los cuales fluyen iones. Entre las décadas de los sesenta y setenta, sin embargo, B. Hille y C. Armstrong comenzaron su estudio en los canales iónicos, defendiendo la idea de que eran poros moleculares. Fue hasta 1998 cuando R. McKinnon pudo cristalizar y caracterizar un canal de ion potasio, K^+ , probando la validez de las suposiciones hechas previamente [8].

Los canales iónicos dependientes de voltaje se caracterizan, entre otras cosas, por tener una hélice transmembranal (S4) cargada positivamente, la cual es un *sensor de voltaje* que está

enlazado con las hélices transmembranales que forman el poro, ver figura 1. Además, muchos de éstos son oligoméricos o tienen una estructura semejante a la de un oligómero [8].

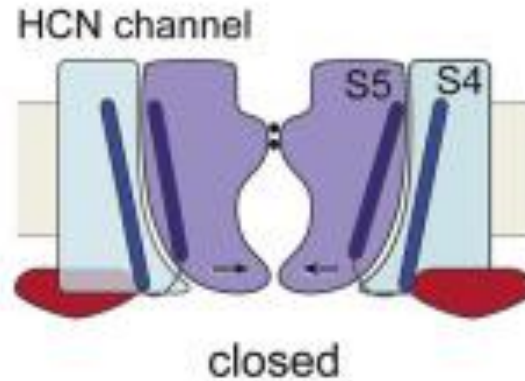


Figura 1. Estructura de un canal iónico tetramérico (el HCN). Imagen extraída de Lee, C.-H.; MacKinnon, R. Structures of the Human HCN1 Hyperpolarization-Activated Channel. Cell 2017, 168(1–2), 111–120.e11

Una característica importante que es compartida por varios canales iónicos dependientes de voltaje, es que éstos pueden cambiar su permeabilidad a determinado ion rápidamente (por ejemplo, los canales de sodio y potasio pueden cambiar su permeabilidad a sodio o potasio, respectivamente, en menos de un milisegundo). Además, si bien su conductancia eléctrica puede variar significativamente, generalmente ésta es de pocos a cientos de picosiemens (pS). Respecto a qué tan estrechos pueden ser los poros, se encuentra que un poro abierto generalmente tiene como región más estrecha un “diámetro” de pocos Angstroms (por ejemplo, en su sección más estrecha, un canal de sodio presenta una sección transversal de $3 \times 5 \text{ \AA}$) [9].

Existe una familia de canales iónicos designados HCN (canales regulados por nucleótidos cíclicos activados por hiperpolarización). Éstos se caracterizan por ser tetraméricos, tener su sensor de voltaje en una hélice transmembranal conocida como S4, tener una escasa selectividad a iones como potasio y sodio, y por controlar la función de marcapasos en células cardiacas y neuronales [10].

Los niveles de cAMP y de cGMP citosólicos magnifican la respuesta al voltaje de los HCN mediante la interacción con un sitio de unión designado CNBD (dominio de enlace a nucleótidos cíclicos) [10].

1.3 Nucleótidos cíclicos

Los nucleótidos cíclicos son compuestos que pueden obtenerse como parte de la degradación enzimática de los trifosfatos de nucleósido (ATP, por ejemplo). Se sabe que el monofosfato de adenosina 3',5'-cíclico (cAMP), el monofosfato de guanosina 3',5'-cíclico (cGMP) y el monofosfato de citidina 3',5'-cíclico (cCMP) se encuentran de manera natural en los organismos vivos, y que tanto el cAMP como el cGMP, desempeñan la función de mensajeros secundarios [11].

Además de lo anterior, se sabe que ciertos nucleótidos cíclicos pueden regular la función de algunos canales iónicos, como es el caso del cAMP y su papel como modulador de un canal de potasio, el cual, como se observa en la figura 2, presenta una estructura tetramérica cuyo arreglo tridimensional le permite formar un poro en la región central” [11]. El objeto de estudio de este trabajo forma parte de un canal iónico modulado por nucleótidos cíclicos (HCN1).

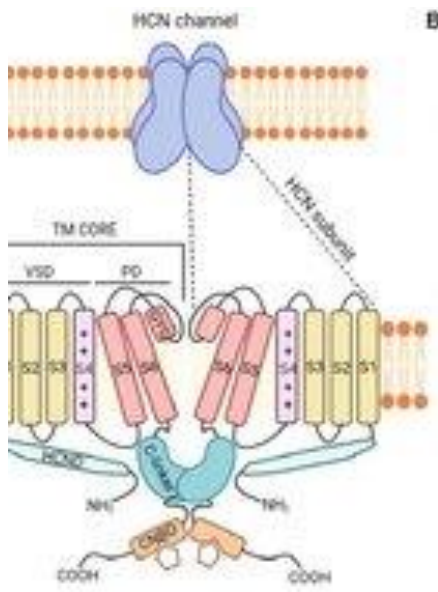


Figura 2. Estructura tetramérica de los HCN.
Imagen extraída de Kim, D.; Roh, H.; Lee, H. M.; Kim, S. J.; Im, M. Localization of Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels in the Vertebrate Retinas Across Species and Their Physiological Roles. Front. Neuroanat. 2024, 18, 1385932.

1.4 Técnicas computacionales

Como parte de las técnicas computacionales, se encuentran los experimentos *in silico* y los experimentos *in virtuo*. Éstos y los experimentos *in vitro* se caracterizan por estar sujetos a menos restricciones éticas y técnicas que los experimentos *in vivo*, ya que emplean una representación artificial de un sistema vivo entero, y se pone a prueba el comportamiento de dicha representación ante diferentes procesos. Los métodos computacionales, sin embargo,

pueden manejar condiciones más parecidas a las fisiológicas y tienen por ende una limitación menor al extrapolar los resultados a un organismo entero que los experimentos *in vitro*, los cuales son llevados a cabo de manera aislada [12].

Los métodos *in silico* han sido ampliamente empleados en la experimentación farmacocinética y farmacodinámica; sin embargo, también han sido empleados en otras áreas como la fisiología cardio-pulmonar [12]. Éstos se basan en modelos matemáticos de sistemas fisiológicos, ya sea en ecuaciones diferenciales ordinarias o parciales, así como en herramientas estocásticas [13]. “La validez de estos modelos se basa tanto en la validez del marco del modelo como de las suposiciones correspondientes” [12]. Los métodos de análisis para los modelos matemáticos incluyen “el análisis de bifurcación” y “la estimación de parámetros”, así como “el cálculo k”, “el cálculo p” y “el cálculo brane” [13].

Respecto a los experimentos *in virtuo*, puede decirse lo siguiente: éstos se caracterizan por el hecho de que “el usuario puede probar la reactividad y la adaptabilidad del modelo en progreso. Y entonces puede tomar la mayor cantidad de propiedades del comportamiento de los modelos numéricos que pueblan los universos virtuales” [13]. Además, estos experimentos buscan tomar en cuenta (1) el ambiente y las perturbaciones en el sistema ocasionadas por aquél, (2) la aleatoriedad, (3) diferentes modelos para diferentes niveles estructurales, entre otros aspectos [13].

Entre las técnicas computacionales utilizadas para describir distintos aspectos relacionados con biomoléculas, destacan las Simulaciones de Dinámica Molecular (SDM), que “capturan el comportamiento de proteínas y otras biomoléculas a escala atómica-molecular con una resolución temporal muy fina” [14]; el acoplamiento molecular, una herramienta útil en la predicción de los modos de enlace de un ligante con una proteína cuya estructura tridimensional es conocida [15]; así como el modelado por homología, cuyo objetivo es predecir una estructura con base en su secuencia “con una precisión que es comparable a los mejores resultados obtenidos experimentalmente”. [16]

Acerca de las Simulaciones de Dinámica Molecular, puede decirse lo siguiente: son capaces de predecir el comportamiento de cada partícula (átomo) que compone al sistema a lo largo del tiempo con base en un modelo que utiliza conceptos de Física Clásica para evaluar las distintas

interacciones interatómicas. Dada la naturaleza atómica de las partículas que componen al sistema, la resolución temporal es de apenas pocos femtosegundos (tamaño de paso). Así, las Simulaciones de Dinámica Molecular permiten predecir el efecto de cambios en la estructura, como pueden ser mutaciones, cambios postraduccionales y unión a ligante, por mencionar algunos. Además, nos permiten investigar una amplia variedad de procesos que afectan la función de diferentes biomoléculas, incluyendo cambios conformacionales, interacción con compuestos de bajo peso molecular, así como interacción con otras biomoléculas [14].

Respecto al acoplamiento molecular, también conocido como *docking* molecular, puede decirse que su objetivo es predecir el reconocimiento molecular, encontrar modos de interacción probables y estimar la afinidad de enlace de dos moléculas (de las cuales al menos una suele ser una macromolécula). De manera particular, los modos de enlace no sólo se refieren a los sitios de unión, sino también a la orientación y a la conformación con la cual interactúan con alta probabilidad [15].

Es necesario escoger los métodos de acoplamiento adecuados para el objeto de estudio en cuestión, ya que se basan en varias aproximaciones que hacen que unos métodos funcionen mejor para ciertos sistemas. Existen diferentes funciones para calificar los modos de enlace; por ejemplo, existen funciones empíricas y funciones basadas en un campo de fuerza. También existen diferentes métodos de búsqueda empleados en el acoplamiento molecular; éstos pueden ser sistemáticos (deterministas) o estocásticos (que involucran cambios aleatorios iterativos); además, los métodos de búsqueda pueden ser locales (que tienden a encontrar un mínimo de energía local o más cercano) o globales (que buscan el mínimo de energía global) [15].

La técnica computacional de modelado por homología se basa en la premisa de que la estructura de una proteína está determinada por la secuencia de aminoácidos (estructura primaria) y que secuencias similares adoptan estructuras similares [16]. El modelado por homología sigue un proceso que puede constar de los siguientes pasos: identificación de la o las plantillas, alineamiento entre las plantillas y la secuencia blanco, generación de la estructura tridimensional de la cadena principal, refinamiento de los bucles, modelado de la cadena lateral, optimización del modelo y validación del modelo [16]; en muchas ocasiones, desde la generación de estructuras hasta la optimización del modelo, se realizan de manera simultánea.

2. Marco teórico

2.1 Los canales activados por hiperpolarización modulados por nucleótidos cíclicos (HCN).

Los HCN son canales iónicos escasamente selectivos dependientes del voltaje, esto es: su apertura requiere de una diferencia de potencial negativa (de al menos -50 ± 10 milivoltios). La conductividad de estos canales es generalmente baja (del orden de 10^{-11} siemens), por lo que, tras la activación de los HCN, el flujo de iones ocasiona una corriente interna negativa. Durante la despolarización diastólica, por ejemplo, se observan, en mayor medida, el flujo de iones sodio hacia adentro, y, en menor medida, el de iones potasio hacia afuera. Estos canales permiten incluso el flujo de iones más grandes, como ion bario y el tetraetilamonio, aunque no de ion cesio [17].

Es posible identificar 4 isoformas de los HCN (designadas HCN1, HCN2, HCN3 y HCN4). Cada una de éstas tiene una diferente dependencia de la conductividad con el voltaje; por ejemplo, el umbral de HCN1 y HCN4 es de -40 milivoltios, mientras que el de HCN2 es aproximadamente 20 milivoltios más negativo [17], y también difiere el valor de $V_{1/2}$ (la diferencia de potencial para la cual la conductividad alcanza la mitad de su valor máximo). Además de lo anterior, cada isoforma tiene un comportamiento cinético diferente; por ejemplo, HCN1 presenta una activación más rápida que HCN2 y HCN4. Por tanto, se puede notar que la conductancia de estos canales depende del voltaje como lo describe la siguiente ecuación:

$$G(V) = G_{\min} + \frac{G_{\max} - G_{\min}}{1 + \exp[(V - V_{1/2})/s]} \quad (1),$$

donde G_{\min} , G_{\max} , $V_{1/2}$ y s son parámetros por determinar experimentalmente para cada isoforma (G_{\max} es la máxima conductancia posible, G_{\min} es la mínima, $V_{1/2}$ es una constante para la cual $G = \frac{G_{\max} + G_{\min}}{2}$). Análogamente, la siguiente ecuación representa bastante bien el comportamiento cinético de varios canales:

$$\frac{I_h}{I_{\max}} \approx 1 - 2^{-(t-r)/(t_{1/2}-r)} \quad (2),$$

donde r es una medida del retardo en reaccionar y $t_{1/2}$ es una constante que depende de cuál isoforma se trata y de la presencia o ausencia del ligante. No obstante, al suponer que $\frac{I_h}{I_{\max}}$ es una combinación lineal de términos del tipo $1 - 2^{-(t-r)/(t_{1/2}-r)}$, es decir:

$$\frac{I_h}{I_{max}} \approx \sum_j c_j \cdot [1 - 2^{-(t-r_j)/(\tau_j-r_j)}] \quad (3),$$

se obtiene una mejoría, al igual que al suponer que

$$\frac{I_h}{I_{max}} = \frac{1}{\beta-1} \cdot \left\{ \beta - \beta \left[\log_{\beta} \left(\frac{\beta+1}{2} \right) \right]^{(t-r)/(\tau-r)} \right\} \quad (4),$$

donde β es un parámetro por determinar experimentalmente, o

$$\frac{I_h}{I_{max}} = 1 - \frac{1}{\alpha} \cdot \alpha \left[\log_{\alpha} \left(\frac{\alpha}{2} \right) \right]^{(t-r)/(\tau-r)} \quad (5),$$

donde α es un parámetro análogo. Para más detalles respecto a esto, véase el Apéndice 1.

Es importante mencionar que, de las 4 isoformas, la más estudiada es HCN1 [17].

Respecto a la estructura de los HCN, se encuentran casi siempre como tetrámeros, ya sean heterotetrámeros u homotetrámeros en su forma funcional, y poseen seis hélices transmembranales. Éstas se abrevian S1, S2... S6, siendo S1 la más cercana al extremo N-terminal, mientras que S6 es la más cercana al extremo C-terminal [17].

Las hélices S5 y S6 son las que forman el poro del canal iónico, mientras que la hélice S4 (rica en aminoácidos cargados eléctricamente) contiene el sensor de voltaje. Respecto al poro, conviene resaltar que éste tiene 2 sitios de filtro (donde abundan tirosinas) y que su secuencia de aminoácidos es muy parecida a la del poro del canal de calcio K_v . Respecto al sensor de voltaje, conviene resaltar que se ha propuesto que éste experimenta un movimiento ante el umbral de voltaje, lo cual conduce a que se muevan las hélices S4 y S5, ocasionando así la apertura del poro [17].

Cerca del extremo C-terminal de los HCN, se encuentra el dominio de unión a nucleótidos cíclicos (CNBD), el cual es un dominio citosólico con alta afinidad a cAMP y a cGMP. Actualmente, se sabe que la unión del CNBD a cAMP o a cGMP acelera la activación y desacelera la desactivación de los HCN, además de que desplaza el umbral de voltaje, y, con ello, incrementa la probabilidad de apertura del poro. A pesar de lo anterior, la activación de estos canales no depende de nucleótidos cíclicos [17].

Los HCN son de suma relevancia para el correcto funcionamiento del sistema nervioso (central y periférico) y del corazón de muchas especies de animales. Destaca que, en el sistema nervioso, impulsan la generación y la transmisión de señales nerviosas, mientras que, en el corazón,

contribuyen a la función de marcapasos de células nodales [17], y diversos eventos que incrementan o reducen significativamente su abundancia o su actividad están correlacionados con problemas de salud; por ejemplo: (1) algunas mutaciones de HCN están correlacionadas con la epilepsia hereditaria (se cree que se debe a que los HCN modulan la excitabilidad neuronal, y la epilepsia se ocasiona por descargas excesivas y repentinas); (2) la enfermedad de Parkinson está correlacionada con la pérdida de neuronas dopaminérgicas, donde abundan las isoformas HCN2 y HCN4; (3) dicha enfermedad también está correlacionada con la disminución en la actividad de los HCN [18]; (4) la disfunción del nodo sinoauricular está correlacionada con una disminución de las isoformas HCN2 y HCN4 en dichos tejidos; (5) en personas que padecen fibrilación auricular, han sido encontradas mutaciones en HCN4; (6) el incremento en la síntesis de los HCN en las aurículas cardíacas está correlacionado con la susceptibilidad a padecer fibrilación auricular; (7) es posible inducir un bloqueo auriculo-ventricular en animales si se inhibe la síntesis de HCN4; (8) la hipertrofia cardíaca, la cardiopatía isquémica y la insuficiencia cardíaca pueden ser acompañadas por un incremento en la corriente de HCN en los ventrículos (lo cual indica un posible desplazamiento del umbral de voltaje hacia valores menos negativos). Esto es de suma relevancia, ya que diversas enfermedades del corazón se encuentran entre las principales causas de muerte a nivel mundial [19].

Lo anterior hace que sea de interés científico hallar ligantes específicos para los HCN que puedan controlar a conveniencia su funcionamiento y que puedan ser empleados como fármacos. Por ejemplo, se ha planteado que un agente capaz de inhibir selectivamente la isoforma HCN2 podría funcionar como analgésico debido al papel que dicha isoforma desempeña en la percepción del dolor [18]; además, existe ya en el mercado un fármaco (la Ivabradina) que inhibe la I_h de los canales HCN, es decir: inhibe activación de los HCN y, con ello, el flujo de iones a través de éste; debido a ello, funciona como un agente bradicárdico específico y se emplea para la cardiopatía coronaria y ante ataques cardíacos [17].

2.2 Comunicación alostérica.

Se designa *sitio alostérico* a una región de una proteína que, al experimentar un cambio conformacional, facilita o propicia otro cambio conformacional en otro sitio de la proteína. Análogamente, se designa *proteína alostérica* a aquella que tiene al menos un sitio alostérico.

Existe un mecanismo en que la unión de una molécula en un sitio de una proteína puede afectar otro sitio estereoquímicamente distinto, sin que los dos sitios puedan superponerse (en contraste con el mecanismo de inhibición por exclusión mutua debida a un impedimento estérico); dicho mecanismo se conoce como *alostería* o *comunicación alostérica*, y hoy se sabe que las interacciones entre sitios son “mediadas por alteraciones discretas, reversibles de la estructura molecular de la proteína” [20].

A lo largo del tiempo, han sido propuestos diversos modelos para explicar la comunicación alostérica. Por ejemplo, en 1959, surgió un modelo (el de ajuste inducido) que explicaba la especificidad de la actividad enzimática basándose en la suposición de que es necesario un ajuste estérico para que una reacción enzimática pueda ocurrir; según dicho modelo, tal ajuste estérico ocurriría exclusivamente después de un cambio en la forma de la enzima inducido por el sustrato; dicho de otro modo, se hacía la suposición de que ligantes particulares pueden inducir una conformación adecuada en una proteína. En 1963, dicho modelo fue retomado en el intento de explicar la actividad de ciertas enzimas bacterianas; el fundamento de lo anterior era que, en efecto, se observaba un cambio en la conformación de dichas enzimas asociado a la transducción de señal; sin embargo, desde 1965, el modelo fue puesto en duda, y diversos avances tecnológicos de las últimas décadas han permitido obtener resultados experimentales que indican que, en realidad, las diferentes conformaciones de las proteínas alostéricas existen independientemente de si el ligante está unido o no [21].

En 1966, Koshland, Némethy y Filmer [21] propusieron un modelo (KNF) para explicar la alostería basándose en la suposición de que, en efecto, sucede un ajuste inducido; dicho modelo propone múltiples estados conformacionales para las proteínas, incluyendo intermediarios con cambios en la estructura terciaria que dependen de la estructura del ligante y de la ocupación; sin embargo, en varios casos, como el de la lactato deshidrogenasa, no son observados los intermediarios que predice este modelo, lo cual parece indicar un mecanismo concertado [21].

También en esa década, surgió uno de los modelos de alostería más importantes (Monod-Wyman-Changeux, MWC), el cual planteó que las proteínas alostéricas deben ser oligoméricas y deben cumplir con ciertas propiedades de simetría (p. ej. mínimo un eje de simetría); para dichos oligómeros, existen ciertos estados que difieren en la posición o energía de los enlaces

entre protómeros, y, por ende, también en ciertas restricciones conformacionales (es decir: difieren en la estructura cuaternaria); cuando se experimenta una transición entre dichos estados, la afinidad de los diferentes sitios, a sus respectivos ligantes, cambia [20]. Dichos estados conformacionales existen independientemente de la estructura del ligante y de si están o no ocupados [21]. Según este modelo, la transición entre estados debe ocurrir de un modo tal que la simetría molecular se conserve; además, en contraste con el modelo KNF, éste parte de la suposición de que las alteraciones estructurales no son ocasionadas por la unión con el ligante, sino que cada ligante estabiliza selectivamente el estado al que se une preferentemente, razón por la cual, según Changeux, este modelo es análogo a la teoría de Darwin respecto a la evolución, mientras que el modelo KNF es análogo a la teoría de Lamarck. El modelo MWC inicialmente permitió explicar la retroalimentación inhibitoria ejercida por ciertos ligantes reguladores sobre ciertas enzimas y la cooperación positiva entre las subunidades de la hemoglobina, y hoy se sabe que describe la mayor parte de casos de alostería documentados. No obstante, también hay excepciones; por ejemplo, ciertos receptores acoplados a proteína G (GPCRs) funcionan sin experimentar oligomerización y ciertas proteasas presentan alostería siendo monoméricas. También se han puesto en duda algunos de los supuestos requisitos para la experimentación de alostería; por ejemplo, se sabe que hay conjuntos moleculares sumamente complejos que son escasamente simétricos y, sin embargo, muchos aspectos de su comportamiento concuerdan con los predichos por este modelo; análogamente, algunas simulaciones de dinámica molecular (por ejemplo, de canales iónicos de potasio tetraméricos) han predicho transiciones en las que la simetría no se conserva. [20].

2.3 El estudio de la comunicación alostérica a través del uso de un algoritmo de propagación de perturbaciones.

En la actualidad, se cree que la estructura secundaria y terciaria de una proteína está determinada por su estructura primaria, y que, como consecuencia, existe información acerca de la dinámica de plegamiento en la estructura primaria [16]. La alostería ha sido considerada análoga al anterior problema en cuanto una perturbación en un sitio (el sitio alostérico) introduce perturbaciones en otros sitios, entendiéndose por perturbación, un cambio producido por la unión de un ligante, una modificación postraducciona, la absorción de luz, etc. [22].

Para predecir el comportamiento a nivel de residuos, herramientas computacionales que emplean los resultados de las simulaciones de dinámica molecular ofrecen la posibilidad de estudiar las diferentes vías de transmisión de la información de una región de la macromolécula a otra distante [22].

En 2020, Wang *et al* [22], lograron construir mapas de acoplamientos dinámicos en proteínas con base exclusivamente en la estructura de las mismas y empleando el software Ohm, desarrollado por el laboratorio del Dr. Dokholyan. Ohm es capaz de predecir vías y sitios alostéricos, estructuras críticas y correlaciones entre pares de residuos. Dicho software se basa en un proceso estocástico iterativo de propagación de perturbaciones [22], es decir que se describe con base en la probabilidad una secuencia de eventos (la propagación de las perturbaciones) que dependen de múltiples variables aleatorias.

El algoritmo de dicho software para calcular los valores de ACI (intensidad de acoplamiento alostérico) se describe a continuación.

Inicialmente, se obtiene la matriz de contactos. Para ello, se estima la cantidad de contactos entre pares de residuos no adyacentes de la proteína; por definición, los átomos de la cadena principal de dos residuos adyacentes (i y j tal que $|i - j| = 1$) no cuentan como contactos; tampoco cuenta como contacto un par de átomos a y b si la distancia entre éstos ($|\mathbf{r}_a^i - \mathbf{r}_b^j|$) es mayor a determinado valor crítico (denotado mediante r_0 , su valor es arbitrario; de manera predeterminada, se le asigna el valor de $r_0 = 3.4 \text{ \AA}$, pero puede modificarse si el usuario lo considera necesario). Si denotamos mediante $H(r_0 - |\mathbf{r}_a^i - \mathbf{r}_b^j|)$ la función escalón de Heaviside evaluada en $x = r_0 - |\mathbf{r}_a^i - \mathbf{r}_b^j|$ es posible obtener la cantidad de contactos entre dos residuos por medio de la siguiente suma: $C_{ij} = \sum H(r_0 - |\mathbf{r}_a^i - \mathbf{r}_b^j|)$, donde la suma toma en cuenta todos los posibles valores de a y b . Posteriormente, se calcula la cantidad promedio de contactos que cada átomo del i –ésimo residuo forma con el j –ésimo residuo mediante la ecuación $N_{ij} = C_{ij}/C_i$, donde C_i es la cantidad de átomos del i –ésimo residuo [22].

Más tarde, se obtiene la matriz de probabilidades. Lo anterior sirve para estimar la probabilidad de que una perturbación en el i –ésimo residuo no se propague al j –ésimo residuo (p_{ij}), pues

se hace la suposición de que $p_{ij} = \exp(-\alpha N_{ij})$, donde α es otra constante arbitraria que inicialmente adopta el valor predeterminado de $\alpha = 3.0$, pero también puede ser modificada por el usuario en caso de considerarse preciso. Más tarde, se calcula la probabilidad de que una perturbación en el i –ésimo residuo sí se propague al j –ésimo residuo (P_{ij}) mediante la ecuación $P_{ij} = 1 - p_{ij}$ [22].

Hecho lo anterior, el software “induce una perturbación” en cada residuo del sitio activo, con lo que comienza el algoritmo de propagación de perturbaciones. Para ello, define un par de vectores de dimensión N , donde N es la cantidad total de residuos de la proteína en cuestión; dichos vectores son \mathbf{V} y \mathbf{W} , y se asigna el valor de 1 para V_n y W_n para los residuos del sitio activo (identificados dentro de la proteína mediante el número n). Más tarde, es generado un número aleatorio r entre 0 y 1, el cual es comparado con la probabilidad P_{nm} si existe contacto entre n y m (m es un número para identificar dentro de la proteína a cada residuo “vecino” del sitio activo): sólo si $P_{nm} > r$ se considera que la perturbación en el n –ésimo residuo se ha propagado al m –ésimo residuo, y entonces se le asigna a una variable V_m el valor de 1; de lo contrario, se le asigna el valor de 0; por otra parte, a W_m se le asigna el valor de 1 independientemente de cuál sea el valor de V_m para todos los residuos vecinos del sitio activo. El proceso se repite hasta que se ha asignado 1 a cada entrada del vector \mathbf{W} . Además, se define T_j a una variable a la que se le suma 1 cada que V_j ha adoptado el valor de 1. El anterior proceso se repite 10,000 veces. Tras normalizarse el vector cuya j –ésima entrada es T_j , cada una de estas magnitudes es considerada una medida de la intensidad de acoplamiento alostérico [22].

Después de obtener las intensidades de acoplamiento alostérico, es posible determinar cuáles son los potenciales sitios alostéricos, y, una vez hecho lo anterior, es posible identificar las diferentes rutas alostéricas y sus probabilidades. Tras determinar lo anterior, es posible también identificar los residuos críticos y sus importancias relativas, donde siempre el inicio y el fin de la ruta tienen un valor de 1.

A continuación, en las Figuras 3 y 4, se expone este algoritmo a modo de diagrama de flujo.

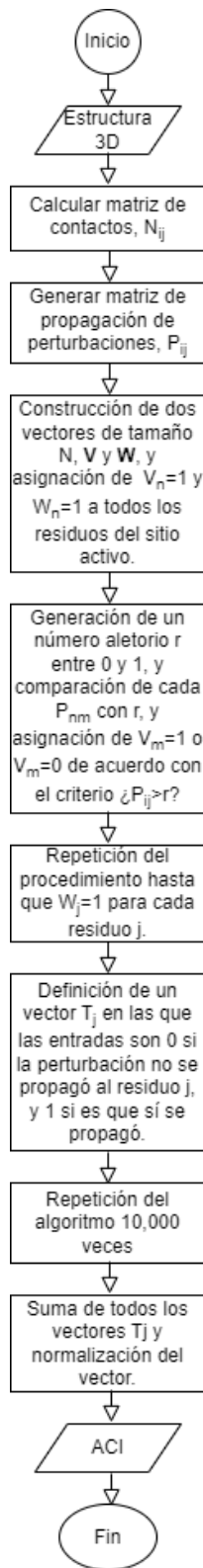


Figura 3. Algoritmo de Ohm para el cálculo de los valores de Intensidad de Acoplamiento Alostérico (ACI).

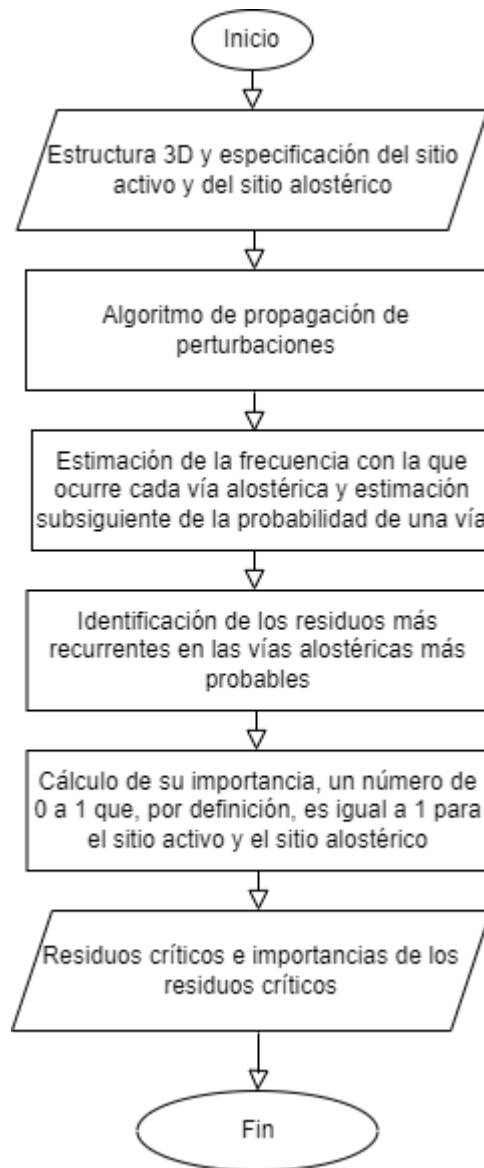


Figura 4. Algoritmo de Ohm para la determinación de los residuos críticos y de sus importancias.

2.4 Simulaciones de dinámica molecular (SDM).

Cualquier simulación computacional es un modelo matemático que recrea computacionalmente un proceso [23]; se trata de una representación de la realidad compleja [23].

Quien usa las simulaciones computacionales adopta una postura empirista, es decir: asume “la posición filosófica de que la experiencia es una fuente de información acerca de nuestro mundo,

y nuestra única fuente” [24]. Esto no debe confundirse, sin embargo, con el hecho de que sólo los experimentos materiales (*in vivo* e *in vitro*) sean válidos; la experiencia puede también alcanzarse a través de los experimentos no-materiales, que incluyen los experimentos mentales (*in mente*) y las técnicas computacionales (*in virtuo* e *in silico*). Estos últimos difieren en el aspecto de que los experimentos *in silico* permiten que siga su curso un proceso fijo, mientras que los experimentos *in virtuo* tienen “la posibilidad de perturbar un modelo que está corriendo” [23].

Las simulaciones computacionales se basan en una epistemología empirista constructivista, es decir en una epistemología que parte del planteamiento de que la ciencia nos proporciona teorías empíricamente adecuadas, en vez de teorías literalmente reales, y que aceptar una teoría implica únicamente creer que es empíricamente adecuada [24], de acuerdo con la perspectiva del filósofo neerlandés Bastiaan Cornelis van Fraassen. Como consecuencia, quien las usa admite la necesidad de que exista una concordancia entre los fenómenos observables y los predichos por una teoría para que esta última sea aceptada; por eso mismo, utiliza modelos teóricos previamente validados.

Si bien las simulaciones no son completamente aceptadas todavía entre gran parte de la comunidad científica, es un error pensar que son menos reales que los experimentos materiales [23].

Actualmente, las simulaciones de dinámica molecular (SDM) son empleadas para obtener, desde el nivel atómico, propiedades de biomoléculas [25]. Los resultados obtenidos son considerados cualitativos, no cuantitativos, pues, a menudo éstas prueban conformaciones únicamente en pocos mínimos locales energéticos en una escala de tiempo generalmente corta, lo cual dificulta la estimación certera de ciertas magnitudes, si bien existen algunos artificios matemáticos y algunas técnicas computacionales útiles para estimar diferencias de energía libre. Por ejemplo, para calcular diferencias de energía de Gibbs, se emplea el ensamble isotérmico-isobárico. Para más detalles al respecto, véase el Apéndice 2.

Específicamente, los softwares para llevar a cabo las SDM se basan en la resolución numérica de ecuaciones diferenciales basadas en la Mecánica Clásica y a veces en aproximaciones Mecánico-Cuánticas para describir algunas interacciones, para lo cual se emplean usualmente

pasos de tiempo de pocos picosegundos o incluso tan pequeños como 1 femtosegundo (matemáticamente, esto es: $1 \times 10^{-12} \text{s} \geq O(\Delta t) \geq 1 \times 10^{-15} \text{s}$) [25]. La escala de tiempo solía ser hasta hace poco del orden de los picosegundos o de los nanosegundos (lo cual implicaba el uso de una cantidad de saltos de un orden menor a 1×10^9) [25]; sin embargo, innovaciones en el área permiten extender dicha escala al orden de los microsegundos o incluso arriba de los milisegundos (lo cual puede implicar el uso de una cantidad de saltos de un orden mayor a 1×10^9 [si se emplea $\Delta t = 1 \times 10^{-15} \text{s}$] pero no mayor a 1×10^{15}) [25]. Si bien el límite superior de la escala del tiempo es arriba de los milisegundos, incluso escalas de cientos de nanosegundos sirven para obtener suficientes estructuras representativas y una idea general de diversos fenómenos que ocurren a gran velocidad en un sistema fisiológico; con lo anterior, se evita caer en extrapolaciones demasiado aventuradas.

En concreto, los pasos de tiempo son empleados para calcular las nuevas posiciones y las nuevas velocidades mediante la resolución numérica de la segunda ley de Newton, que, para una partícula individual (j), sería

$$\mathbf{F}_j = m_j \cdot \frac{d\mathbf{v}_j}{dt} = m_j \cdot \frac{d^2\mathbf{r}_j}{dt^2},$$

a modo que

$$(\mathbf{v}_j)_{t+\Delta t} = (\mathbf{v}_j)_t + \frac{1}{m_j} \cdot \mathbf{F}_j \cdot \Delta t,$$

así como

$$(\mathbf{r}_j)_{t+\Delta t} = (\mathbf{r}_j)_t + (\mathbf{v}_j)_t \Delta t,$$

donde los subíndices j indican que las propiedades físicas correspondientes le corresponden a la j -ésima partícula.

A su vez, para estimar \mathbf{F}_j se hace la suposición de que las fuerzas son conservativas, a modo que es posible asociar una función U (correspondiente a la energía potencial) tal que $\mathbf{F}_j = -\nabla_j U$, donde ∇_j toma en cuenta las coordenadas especiales de la j -ésima partícula.

CHARMM (siglas en inglés para *Mecánica Macromolecular de Química en Harvard*) propone la siguiente ecuación para U , misma que se retoma en este trabajo:

$$\begin{aligned}
 U = & \sum_{\text{enlaces}} K_b \cdot (b - b_0)^2 + \sum_{\text{ángulos}} K_\theta \cdot (\theta - \theta_0)^2 + \\
 & + \sum_{\text{dihedros}} K_\chi \cdot [1 + \cos(n\chi - \delta)] + \sum_{\text{dihedros}} K_{UB} \cdot (S - S_0)^2 + \sum_{\text{impropios}} K_\phi \cdot (\phi - \phi_0)^2 + \\
 & + \sum_{\text{pares de átomos no enlazados}} \left(\varepsilon_{ij} \cdot \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{R_{\min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon_D r_{ij}} \right)
 \end{aligned}$$

Es posible señalar que las SDM pueden ser útiles para realizar inferencias, desde el punto de vista bayesiano. El análisis bayesiano requiere de la elección *a priori* de una distribución estadística que tiene el objetivo de representar lo que se conoce acerca de ciertos parámetros. Escogiendo cuidadosamente el modelo el análisis bayesiano puede evitar que lo previo a la experiencia domine y distorsione “lo que la información está tratando de decir” [26]. Una vez que se pone a prueba la adecuación de la distribución de probabilidad escogida *a priori* para los datos, es posible conocer la distribución de probabilidad que siguen los parámetros definidos cuando los datos varían, y, de ese modo, es posible hacer inferencias acerca de los parámetros. Para más detalles al respecto, revisar el Apéndice 3.

3. Hipótesis

La unión a los ligantes nativos (cAMP o cGMP) afecta de manera significativa la comunicación alostérica dentro del dominio CNBD (por sus siglas en inglés *cyclic nucleotide-binding domain*) del canal HCN1, misma que debe ocurrir en condiciones normales del sitio activo a la región citosólica (en lo subsiguiente llamada “hélice 1”) más cercana al dominio transmembranal S6. Los métodos computacionales nos ayudan a entender cómo se ve afectada la comunicación alostérica y cómo se ve afectado el movimiento de la hélice 1 a lo largo de una trayectoria, lo cual puede ayudar a buscar un fármaco con efectos similares para incrementar la actividad de dicho canal iónico.

4. Objetivos

4.1 General

Utilizar métodos computacionales para incrementar la comprensión respecto a la comunicación alostérica dentro del sistema proteico y los factores que pueden afectarla, al igual que la comprensión respecto a las fluctuaciones de los residuos de ciertos dominios. De ese modo, determinar qué efectos, en la comunicación alostérica y en las fluctuaciones de ciertos dominios, puede tener un fármaco que incremente la función de los canales iónicos bajo estudio.

4.2 Particulares

Describir la comunicación alostérica dentro del CNBD del HCN1 en el sistema *apo* (en ausencia de ligantes), así como en los sistemas en los que dicha proteína está ligada a cAMP o a cGMP.

Encontrar, si es posible, una relación entre la polaridad de los residuos más involucrados en la comunicación alostérica.

Realizar mediciones de RMSF para medir las fluctuaciones que ocurren en los residuos que forman parte de la hélice 1 a lo largo de una trayectoria de Simulaciones de Dinámica Molecular tanto en el sistema *apo* como en el sistema ligado a cAMP y cGMP.

Realizar una comparación entre el comportamiento del sistema en presencia y en ausencia de ligantes, y obtener conclusiones respecto a la posible forma en que un ligante o un fármaco podría afectar la comunicación alostérica dentro del sistema y el movimiento de la hélice 1.

5. Metodología

Fueron obtenidas de la base de datos *Protein Data Bank* (PDB) las estructuras determinadas experimentalmente mediante cristalografía de rayos X correspondientes al homotetrámero de HCN1 (como complejo con cAMP con el código 5U6O); de éste, se obtuvo también la estructura apo y el complejo con cGMP.

Más tarde, fueron seleccionados 207 residuos pertenecientes al dominio de enlace a nucleótidos cíclicos (denominado CNBD por las siglas en inglés de *cyclic nucleotide-binding domain*) de cada monómero (segmento SER 402-ASP 608), y se desarrolló un código en el lenguaje TCL para preparar cada sistema tetramérico (apo, CNBD/cAMP y CNBD/cGMP) en un ambiente fisiológico. Posteriormente, se introdujo el sistema proteico en una caja de agua prismática a una distancia mínima de 18 ± 2 Å del sistema homotetramérico y se incluyó una solución 0.15 M de NaCl. Dichos sistemas corresponden al estado inicial que será investigado mediante la técnica computacional de simulaciones de dinámica molecular. Los cálculos se realizaron con NAMD a una temperatura de 37 °C y a una presión de 1 atm. Para estimar las interacciones de los átomos del sistema, se empleó el campo de fuerzas generalizado (en el caso en que fueron incluidos los ligantes cAMP y cGMP) y el campo de fuerzas para proteínas de CHARMM36; los pasos de tiempo empleados fueron de 2.0 femtosegundos y la escala de tiempo de la trayectoria fue mayor a 500 nanosegundos para cada sistema.

Se calculó el RMSD para los diferentes sistemas para comprobar que se alcanzaba estabilidad estructural a lo largo de 500 nanosegundos de la simulación.

Utilizando la estructura inicial del complejo ligante/proteína, se determinó cuáles eran los residuos que formaban parte del sitio activo, donde se seleccionaron todos aquellos residuos que tenían alguno de sus átomos a un máximo de 4 Å de distancia de cualquier átomo de los ligantes.

Una vez determinados los residuos que forman el sitio de unión a ligante, se procedió a calcular las intensidades de acoplamiento alostérico (ACI) para cada sistema, CNBD (apo), CNBD/cAMP y CNBD/cGMP para las estructuras tridimensionales a 0 nanosegundos, 250 nanosegundos y 500 nanosegundos de simulación.

Los resultados fueron comparados y se obtuvo un promedio de las cuatro cadenas del homotetrámero (A, B, C y D) donde se excluyeron los valores de las cadenas y de tiempo de simulación para los cuales había una diferencia estadísticamente significativa (según pruebas de ANOVA y de t de Student). Para más detalles al respecto, véase el Apéndice 4.

Basado en el análisis anterior, se escogió el tiempo de simulación más apropiado para estudiar los valores de ACI de cada sistema, y se determinó, para cada sistema, cuáles eran los 3 residuos más cercanos al segmento transmembrana S6 que, con mayor probabilidad, podían ser considerados sitios alostéricos.

Hecho lo anterior, se procedió a determinar cuáles eran las rutas alostéricas más probables hacia dichos residuos, y cuáles eran los residuos críticos (con sus respectivas importancias).

Adicionalmente, se midió el RMSF de cada sistema para la trayectoria durante el tiempo de 400 a 500 nanosegundos, y luego se buscó una correlación entre dicha magnitud y la importancia de los residuos críticos.

6 Discusión de resultados

6.1 Simulaciones de Dinámica Molecular.

Para cada sistema, se simuló una trayectoria de más de 500 nanosegundos del homotetrámero correspondiente. A continuación, en la figura 5, se muestra la estructura homotetramérica del CNBD del HCN1/cAMP.

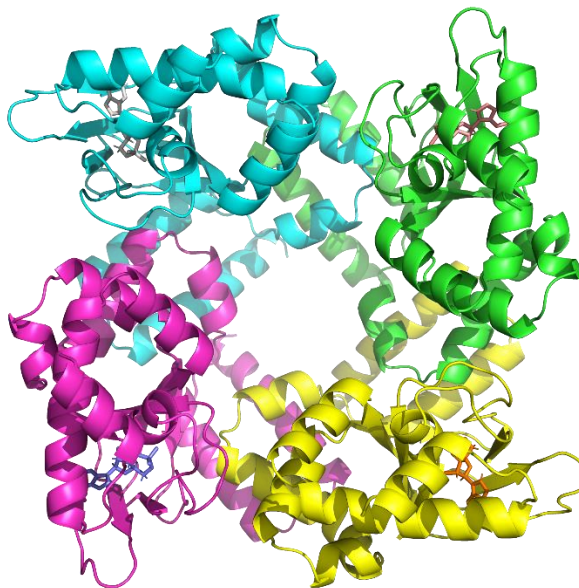


Figura 5. Estructura homotetramérica del CNBD del HCN1/cAMP. Las cadenas son indicadas con diferentes colores. La cadena A en verde; la cadena B en cian; la, cadena C en magenta, y la cadena D en amarillo. La estructura del ligante cAMP está indicada en morado.

Uno de los análisis que podemos realizar a partir de los resultados de simulaciones de dinámica molecular es el del cálculo de la *desviación de la raíz cuadrática media* (RMSD por las siglas en inglés de *Root Mean Square Deviation*), la cual es una medida de la distancia promedio entre un grupo de átomos. En este caso, se empleó para comparar la similitud de las diferentes estructuras obtenidas de la simulación de dinámica molecular, relativas a la estructura inicial, utilizando las posiciones de sus átomos de carbono alfa.

Se observó para el sistema CNBD (apo) que alcanza una meseta aproximadamente a 150 nanosegundos de la simulación, lo cual indica que el sistema homotetramérico ha alcanzado

estabilidad estructural dadas las condiciones de las simulaciones de dinámica molecular. Los valores de RMSD relativos a la estructura inicial (estructura cristalográfica), se encuentran por debajo de los 4 Å. Véase al respecto la figura 6.

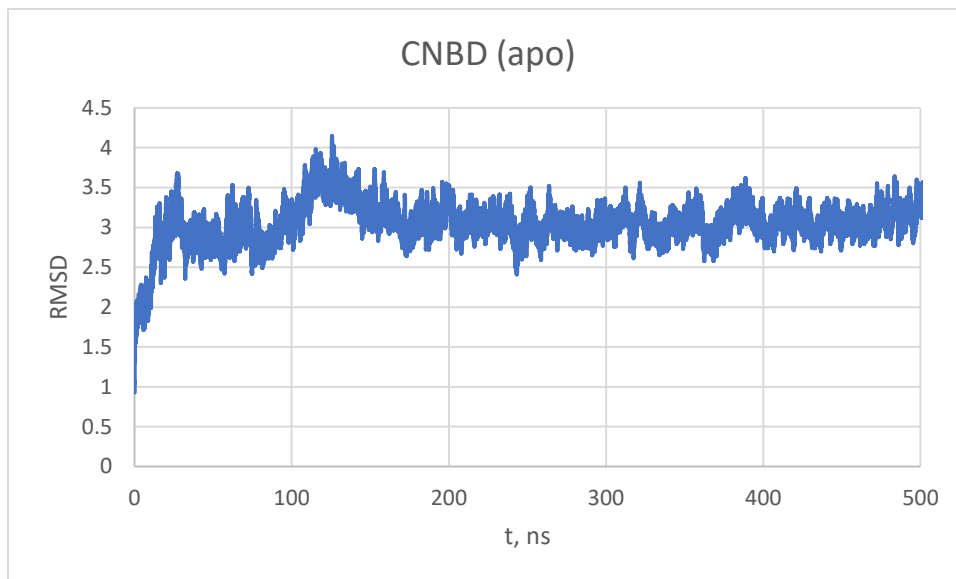


Figura 6. Valores de RMSD para el sistema CNBD (apo).

Para el sistema CNBD/cAMP, análogamente, el análisis de RMSD nos indica que el sistema alcanza una meseta aproximadamente desde los 250 nanosegundos, lo cual indica cierta estabilidad estructural del sistema homotetramérico. Véase al respecto la figura 7.

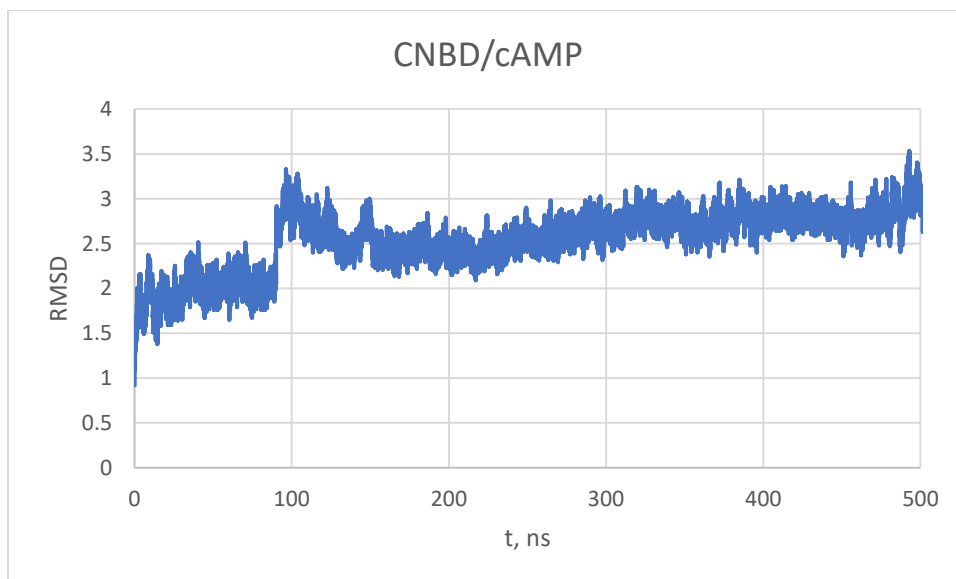


Figura 7. Valores de RMSD para el sistema CNBD/cAMP.

Por otra parte, para el sistema CNBD/cGMP, se observa una meseta aproximadamente desde los 150 nanosegundos, indicando que el homotetrámero alcanza desde entonces cierta estabilidad estructural. Véase al respecto la figura 8.

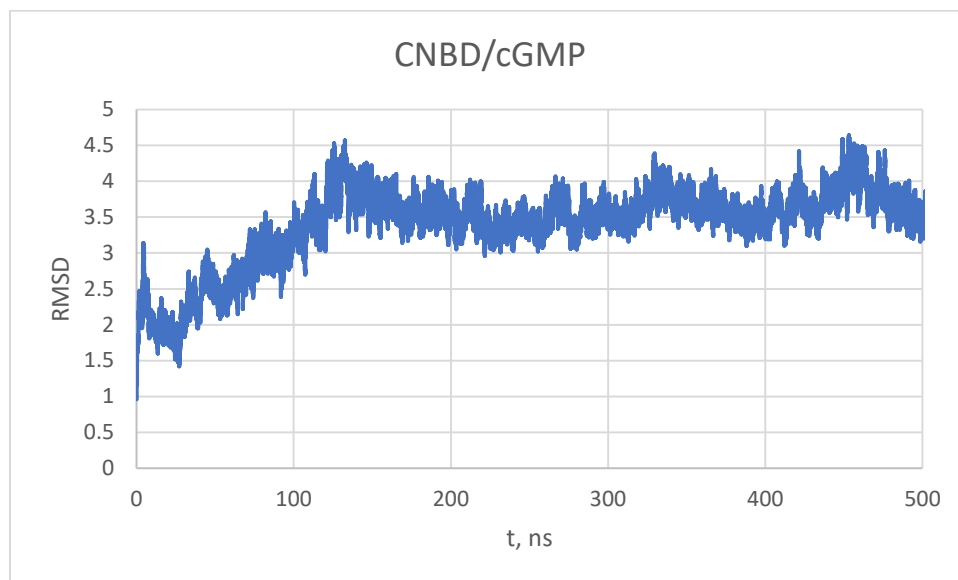


Figura 8. Valores de RMSD para el sistema CNBD/cGMP.

6.2 Determinación de los residuos que forman el sitio de unión a ligante.

Utilizando el criterio de 4 Å descrito en la sección de metodología, se determinó que los sitios activos de los sistemas CNBD/cAMP y CNBD/cGMP compartían los siguientes residuos: ILE 503, MET 530, PHE 538, GLY 539, GLU 540, ILE 541, CYS 542, ARG 549, THR 550, ALA 551, VAL 553, ARG 590, ILE 594, los cuales están resaltados en azul en la figura 9a. Sin embargo, el sitio activo del sistema CNBD/cAMP incluía, además de los antes mencionados, los residuos LEU 532 y ARG 593 (resaltados en verde en la Figura 9b), mientras que el sitio activo del sistema CNBD/cGMP incluía el residuo LEU 591 (resaltado en azul en la figura 9c).

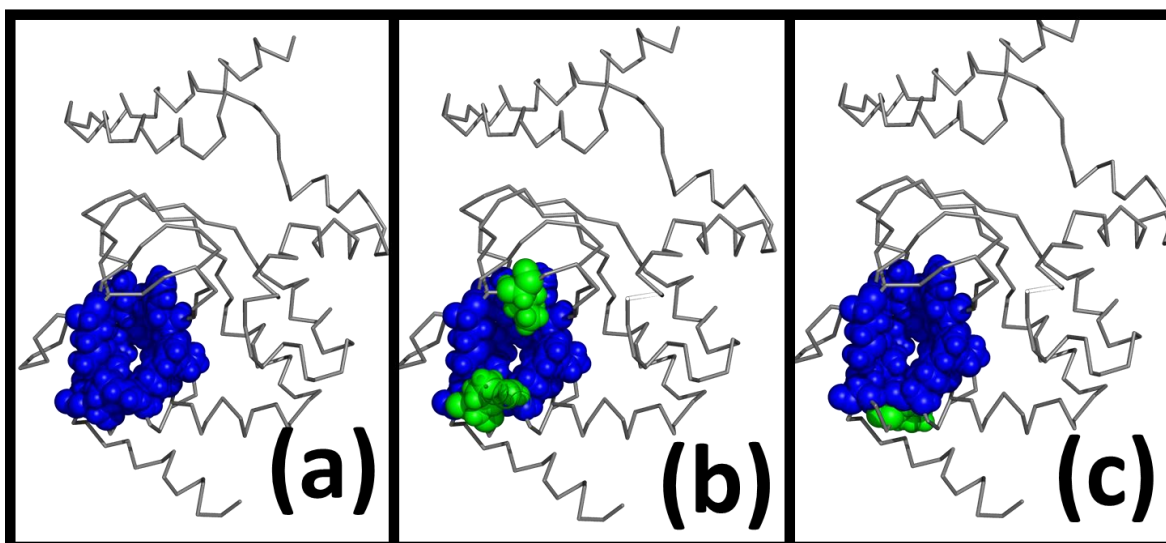


Figura 9. Residuos (a) comunes a los sitios activos, (b) del sitio activo de CNBD/cAMP. (c) del sitio activo de CNBD/cGMP.

Debido a los protocolos de Ohm, el software empleado para estimar los valores de intensidad de acoplamiento alostérico (ACI), es necesario especificar los residuos que formarán el denominado *sitio activo*, se decidió incluir a los residuos comunes a los sitios activos de CNBD/cAMP y CNBD/cGMP como aquellos que son considerados “el sitio activo” para el sistema CNBD *apo*.

A partir del cálculo de simulaciones de dinámica molecular del sistema homotetramérico CNBD, se decidió realizar los análisis de comunicación alostérica en 3 sistemas, el inicial (denominado 0 ns), uno intermedio (denominado 250 ns) y uno final (denominado 500 ns), donde las estructuras representativas de los sistemas intermedio y final fueron tomadas a los 250 ns y 500 ns, respectivamente, mientras que la estructural inicial es la estructura obtenida experimentalmente. Dicha selección nos generó un total de 9 estructuras homotetraméricas a analizar utilizando el protocolo de Ohm, CNBDapo_0ns, CNBDapo_250ns, CNBDapo_500ns, CNBDcAMP_0ns, CNBDcAMP_250ns, CNBDcAMP_500ns, CNBDcGMP_0ns, CNBDcGMP_250ns y CNBDcGMP_500ns.

6.3 Análisis estadístico de los resultados de Ohm.

Tras realizar las pruebas estadísticas ANOVA y de t de Student para las diferentes estructuras, se encontró que, para el sistema *apo* entre los tiempos de simulación CNBDapo_0ns y CNBDapo_250ns, había diferencias estadísticamente significativas para el 13.9% de residuos a la distancia antes mencionada de la hélice 1; al comparar los resultados para las estructuras CNBDapo_0ns y CNBDapo_500ns, el porcentaje incrementaba a 16.5%, mientras que, al comparar los sistemas CNBDapo_250ns y CNBDapo_500ns, el porcentaje disminuía a 2.06%. Para el sistema CNBD/cAMP, se encontraba algo muy semejante: al comparar las estructuras CNBDcAMP_0ns y CNBDcAMP_250ns, el porcentaje era de 13.0%; al comparar los sistemas CNBDcAMP_250ns y CNBDcAMP_500ns, el porcentaje incrementaba 22.4%, mientras que, al comparar los resultados para las estructuras CNBDcAMP_0ns y CNBDcAMP_500ns, el porcentaje disminuía a 8.85%. Finalmente, para el sistema CNBD/cGMP, se repetía el mismo patrón: al comparar los resultados para las estructuras CNBDcGMP_0ns y CNBDcGMP_250ns, el porcentaje era de 13.5%; para los sistemas CNBDcGMP_250ns y CNBDcGMP_500ns, el porcentaje incrementaba 20.7%, mientras que, al comparar los resultados para las estructuras CNBDcGMP_0ns y CNBDcGMP_500ns, el porcentaje disminuía a 6.74%. Lo anterior se resume con la siguiente tabla:

Tabla 1. Diferencias estadísticamente significativas entre medias de ACI

Sistema	(0ns, 250ns)	(0ns, 500ns)	(250ns, 500ns)
CNBD (apo)	13.9%	16.5%	2.06%
CNBD/cAMP	13.0%	22.4%	8.85%
CNBD/cGMP	13.5%	20.7%	6.74%

De lo anterior, se concluyó que es preferible, efectivamente, realizar simulaciones de dinámica molecular antes de someter las estructuras al algoritmo de propagación de perturbaciones. Esto debido a que las estructuras obtenidas a partir de las simulaciones para los tres sistemas (CNBD(apo), CNBD/cAMP y CNBD/cGMP) representa una mejor descripción de las conformaciones que el sistema proteico explora bajo un ambiente fisiológico de presión, temperatura y concentración salina. Dichas conformaciones difieren de la estructura inicial (véase al respecto la tabla 1), debido a las condiciones en las que fue obtenida la estructura inicial (estructura experimental). Adicionalmente, con base en estas observaciones, se

determinó que podría ser prácticamente indistinto si se escogía el marco correspondiente a 250 nanosegundos o si se escogía el correspondiente a 500 nanosegundos para el siguiente paso, que consiste en la determinación de las rutas alostéricas y de los residuos críticos. Sólo para aprovechar que la escala de tiempo es mayor a 500 nanosegundos para cada sistema, se escogieron las estructuras correspondientes a 500 nanosegundos, es decir, CNBDapo_500ns, CNBDcAMP_500ns y CNBDcGMP_500ns.

Las cuatro cadenas (monómeros) de cada sistema (homotetrámero) fueron tomadas en cuenta en los cálculos. Sin embargo, notamos que una vez definido el sitio activo dentro alguna cadena

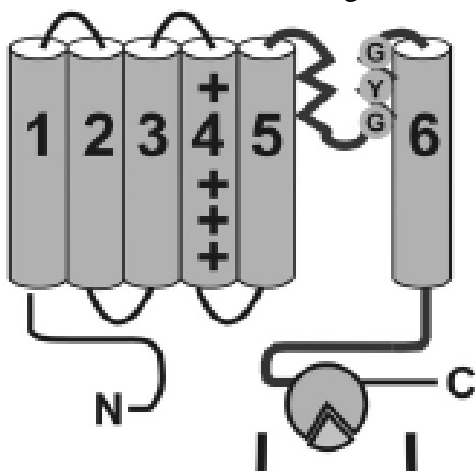


Figura 10. Estructura de los HCN. Imagen extraída de Sartiani, L.; Mannaioni, G.; Masi, A.; Romanelli, M. N.; Cerbai, E. *The Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels: From Biophysics to Pharmacology of a Unique Family of Ion Channels*. *Pharmacol. Rev.* 2017, 69 (4), 354–395.

(cadena A, por ejemplo), los residuos que presentaban mayores valores de ACI eran los que se encontraban en la misma cadena (cadena A en este caso) y con una contribución baja de residuos de otras cadenas (cadena B, C y D en este caso). Por lo tanto, nos restringimos a buscar potenciales sitios alostéricos dentro de la misma cadena en la que estaba el sitio activo. Las cuatro cadenas (monómeros) de cada sistema (homotetrámero) fueron tomadas en cuenta en los cálculos. Sin embargo, notamos que una vez definido el sitio activo dentro alguna cadena (cadena A, por ejemplo), los residuos que presentaban mayores valores de ACI eran los que se encontraban en la misma cadena (cadena A en este caso) y con una contribución baja de residuos de otras cadenas (cadena B, C y D en este caso). Por lo tanto, nos restringimos a buscar potenciales sitios alostéricos dentro de la misma cadena en la que estaba el sitio

activo.

Además de lo anterior, se buscó en cada caso los potenciales sitios alostéricos más cercanos al segmento transmembrana S6, el cual es el segmento más próximo al dominio CNBD y que constituye parte del dominio de formación del poro del canal (ver figura 10). Para ello, se midió

la distancia de los potenciales sitios alostéricos con el centro de masa de la hélice 1, la cual está constituida por el segmento ARG 404 – PHE 420 y marcada en amarillo en las figuras 11 y 12.

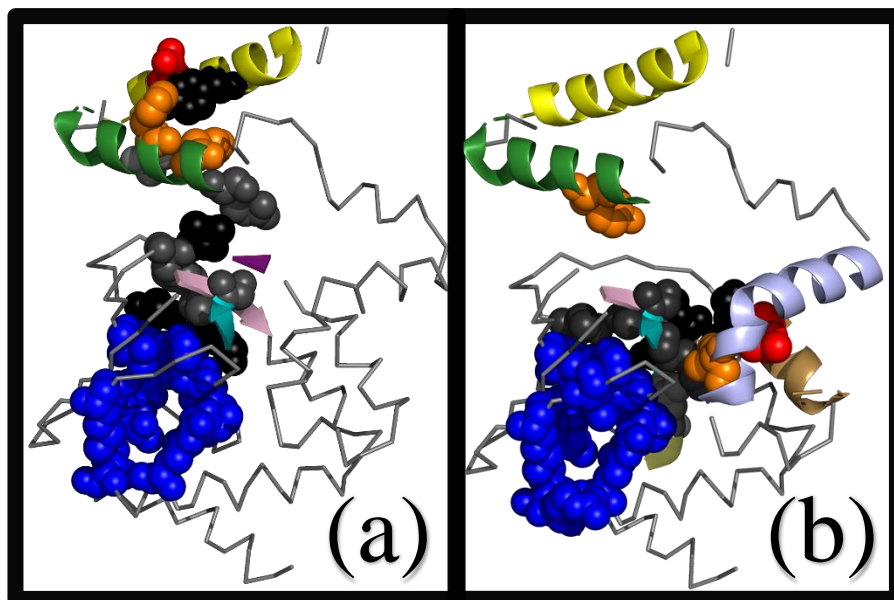


Figura 11. CNBD de HCN1 (a) sistema apo, (b) CNBD/cAMP. En amarillo, se señala la hélice 1. En verde, la hélice 2. En lavanda, la hélice 4. En café arena, la hélice 5. En amarillo pálido, la hélice 6. En púrpura profundo, la hoja beta plegada 1. En cian, la hoja beta plegada 2. En rosa, la hoja beta plegada 7. Como un conjunto de esferas azules, se encuentran los residuos que forman parte del sitio activo; como un conjunto de esferas rojas y naranjas, los sitios alostéricos más relevantes para cada sistema, y, como un conjunto de esferas negras y grises oscuras, los principales residuos críticos en la comunicación alostérica.

6.4 Comunicación alostérica en el sistema CNBD (apo)

Una vez definido el sitio activo para el caso del sistema CNBD (apo), mostrado como esferas azules en las figuras 11a y 12a, se procedió a determinar los sitios alostéricos más relevantes y con cierta proximidad a la hélice 1, marcada en amarillo en la misma figura. Así, se determinó que el sitio alostérico más relevante (GLU 415, marcado como un conjunto de esferas rojas en la Figura 11a) estaba localizado en la hélice 1 (marcada en amarillo en la Figura 11, apenas a 5.3 Å del centro de masa de dicha hélice). Por otra parte, otros dos potenciales sitios alostéricos,

MET 418 e HIS 432 (marcados como un conjunto de esferas anaranjadas en la Figura 11a), se encontraban a menos de 10.4 Å del centro de masa de dicha hélice (MET 418 se encuentra dentro de la hélice 1, a 9.4 Å del centro de masa, mientras que HIS 432 se encuentra dentro de la hélice 2, marcada en verde en la Figura 11, a 10.3 Å del centro de masa de la hélice 1). Al intentar encontrar una relación entre la naturaleza de los aminoácidos en cuestión y su carácter de sitio alostérico más relevante o de potenciales sitios alostéricos, no hay una relación clara: dos son polares mientras que el otro tiene un carácter alifático no polar.

Nota: para este primer sistema, se dio prioridad a los residuos con valores altos de ACI que se encontraban a menos de 30 Å del centro de masa de la hélice 1. Esto, en parte, porque, de 11 valores de ACI que cumplían con los criterios para ser tomados en cuenta, 9 (82%) se encontraban a menos de esa distancia. A continuación, en la tabla 2, son mencionados los valores medios de ACI para los potenciales sitios alostéricos. Los valores promedio son considerados tomando en cuenta los valores para las 4 cadenas del homotetrámero de CNBD investigado.

Tabla 2. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD (apo).

Residuo	r, Å	(ACI)_promedio
TYR 411	3.7	0.72
VAL 414	4.7	0.71
GLU 415	5.3	0.74
MET 418	9.4	0.72
ILE 431	11.6	0.76
HIS 432	10.3	0.74
TYR 434	13.7	0.73
TYR 435	9.0	0.74
LEU 469	32.7	0.86
ALA 486	35.2	0.82
PRO 498	23.4	0.74

Lo anterior indica que los residuos LEU 469, ALA 486 y PRO 498, mostrados al final de la tabla anterior, también experimentan comunicación alostérica con el sitio activo (sitio de unión a nucleótidos cíclicos); no obstante, se trata de residuos muy lejanos al centro de masa de la

hélice 1, la cual se ha sugerido relevante debido a su proximidad con la hélice transmembranal S6.

A continuación, se definirá “contribución” (C_j) a la razón aritmética de la importancia (I_j) de un residuo crítico (o conjunto de residuos críticos) y la suma de las importancias de los residuos críticos ($\sum_{j=1}^n I_j$). Dicho de otro modo, se tiene

$$C_j = \frac{I_j}{\sum_{j=1}^n I_j} \quad (1).$$

Cabe resaltar que, por definición, $I_j = 1.0$ para el sitio activo y para el sitio alostérico, pero adopta valores entre 0 y 1 para el resto de residuos críticos.

También es preciso mencionar que, para escoger el conjunto de “residuos críticos principales”, se ha tomado el conjunto de residuos críticos distintos al sitio activo y al sitio alostérico con valores de I_j más altos (por ejemplo, $I_j = 1.0$, $I_j = 0.8$ o $I_j = 0.6$) de tal modo que la suma de sus contribuciones es apenas mayor a 50%:

$$\sum_{j=1}^m C_j = 50\% + E \quad (2),$$

donde E debe ser un número positivo lo más pequeño posible (por ejemplo, para el conjunto de sistemas estudiados, se cumple $4.0\% \leq E \leq 22.0\%$).

Por otra parte, se determinó que, de los 17 residuos críticos distintos al sitio activo y al sitio alostérico más relevante, sólo 8 aportaban una contribución del 59.2%. Estos residuos (los residuos críticos principales, están marcados como un conjunto de esferas negras o gris oscuro en la Figura 11a).

De éstos, el 87.5% (es decir: 7 de 8) se encontraba en estructuras secundarias bien definidas: uno de estos (TYR 411) se encontraba dentro de la hélice 1, a 3.73 Å del centro de masa; dos (TYR 435 y ARG 438) se encontraban dentro de la hélice 2 (marcada en verde en la Figura 11); dos (VAL 495 y PHE 496) se encontraban dentro de la hoja beta plegada 1 (marcada en púrpura profundo en la Figura 11a); uno (TYR 513) se encontraba dentro de la hoja beta plegada 2 (marcada en cian en la Figura 11), y uno (LEU 561) se encontraba dentro de la hoja beta plegada 7 (marcada en rosa en la Figura 11). El residuo crítico principal restante, ILE 502, es hidrofóbico y alifático y se encuentra en un bucle. Al intentar encontrar una relación entre la naturaleza de los aminoácidos y su carácter de residuos críticos principales, se encuentra que

la mitad son polares (uno cargado eléctricamente) y la otra mitad es hidrofóbica (3 de 4 alifáticos); dicho de otro modo, no hay una relación evidente.

6.5 Comunicación alostérica en el sistema CNBD/cAMP

Para el sistema CNBD/cAMP, por otra parte, se determinó que el sitio alostérico más relevante (LEU 488, marcado como un conjunto de esferas rojas en la Figura 11b) se localizaba en la hélice 5 (marcada en café arena en la Figura 11b) a una distancia de entre 31.5 Å del centro de masa de la hélice 1; otro potencial sitio alostérico (TYR 434, marcado como un conjunto de esferas anaranjadas en la Figura 11b) se localizaba en la hélice 4 (marcada en color lavanda en la Figura 11b) a tan sólo 13.7 Å del centro de masa de la hélice 1, pero otro (LEU 469, marcado como un conjunto de esferas anaranjadas en la Figura 11b) se encontraba a 32.7 Å del centro de masa de la hélice 1. Nuevamente, al intentar encontrar una relación entre la naturaleza de los aminoácidos en cuestión y su carácter de sitio alostérico más relevante o de potenciales sitios alostéricos, no hay una relación clara: dos son hidrofóbicos y alifáticos, pero uno es polar sin carga.

Nota: para el segundo sistema, se dio prioridad a los residuos con valores altos de ACI que se encontraban a menos de 35 Å del centro de masa de la hélice 1. Esto, en parte, porque, de 7 valores de ACI que cumplían con los criterios para ser tomados en cuenta, 4 (57%) se encontraban a menos de esa distancia. Dos residuos son excluidos del análisis (ARG 548 y ASP 589) por tratarse de residuos adyacentes al sitio activo (es decir: directamente conectados, mediante enlace covalente, a éste). A continuación, en la tabla 3, son mencionados los valores medios de ACI para los potenciales sitios alostéricos. De manera similar, los valores promedio son considerados tomando en cuenta los valores para las 4 cadenas del homotetrámero de CNBD/cAMP investigado.

Tabla 3. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD/cAMP.

Residuo	r, Å	(ACI)_promedio
TYR 434	13.7	0.72
LEU 469	32.7	0.80
ASN 482	41.3	0.75

PHE 483	39.5	0.81
LEU 488	31.5	0.86
ILE 523	35.4	0.84
ARG 548	37.7	0.93
TYR 558	21.6	0.80
ASP 589	46.8	0.93

Lo anterior indica que los residuos ASN 482, PHE 483 e ILE 523 también experimentan comunicación alostérica con el sitio activo; no obstante, se trata de residuos muy lejanos a la hélice 1. De hecho, se encuentran a más de 15 Å del centro de masa de cualquier residuo de la hélice 1, a diferencia de los residuos que sí fueron tomados en cuenta. Por otra parte, como se ha comentado ARG 548 y ASP 589 están unidos covalentemente al sitio activo y fueron excluidos del análisis.

Por otra parte, se determinó que, de los 11 residuos críticos distintos al sitio activo y al sitio alostérico más relevante, 7 aportaban una contribución del 54%. Los “residuos críticos principales” están marcados como un conjunto de esferas negras o gris oscuro en la Figura 11b. De éstos, el 71.4% (es decir: 5 de 7) se encontraba en estructuras secundarias bien definidas: dos (LYS 511 y PHE 514) se encontraban dentro de la hoja beta plegada 2 (marcada en cian en la Figura 11), uno (LEU 543) se encontraba dentro de la hélice 6 (marcada en amarillo pálido en la Figura 11b) y dos (LEU 561 y SER 563) se encontraban dentro de la hoja beta plegada 7 (marcada en rosa en la Figura 11). Los aminoácidos críticos principales restantes, LEU 491 e ILE 502, son hidrofóbicos y alifáticos. Nuevamente, al intentar encontrar una relación entre la naturaleza de los aminoácidos y su carácter de residuos críticos principales, se encuentra que 5 de 7 son hidrofóbicos, mientras que sólo 2 son polares (uno de ellos cargado eléctricamente).

6.6 Comunicación alostérica en el sistema CNBD/cGMP

Análogamente, para el sistema CNBD/cGMP, el sitio alostérico más relevante (PHE 514, marcado como un conjunto de esferas rojas en las figuras 12b y 12c) se encontraba a una distancia de 30.8 Å del centro de masa de la hélice 1, dentro de la hoja beta plegada 2 (marcada en cian en la Figura 12). Por otra parte, otro potencial sitio alostérico, ASN 463, marcado como un conjunto de esferas anaranjadas en las figuras 12b y 12c, se encontraba a 32.7 Å del centro de masa de la hélice 1, dentro de la hélice 4 (marcada en beige claro en la Figura 12), mientras

que otro potencial sitio alostérico, CYS 559, marcado como un conjunto de esferas anaranjadas en las figuras 12b y 12c, se encontraba apenas a 24.6 Å del centro de masa de la hélice 1, dentro de la hoja beta plegada 7 (marcada en rosa en la Figura 12). Al intentar encontrar una relación entre la naturaleza de los aminoácidos en cuestión y su carácter de sitio alostérico más relevante o de potenciales sitios alostéricos, no hay una relación clara: dos son polares sin carga (uno de los cuales es azufrado), pero uno es hidrofóbico y aromático.

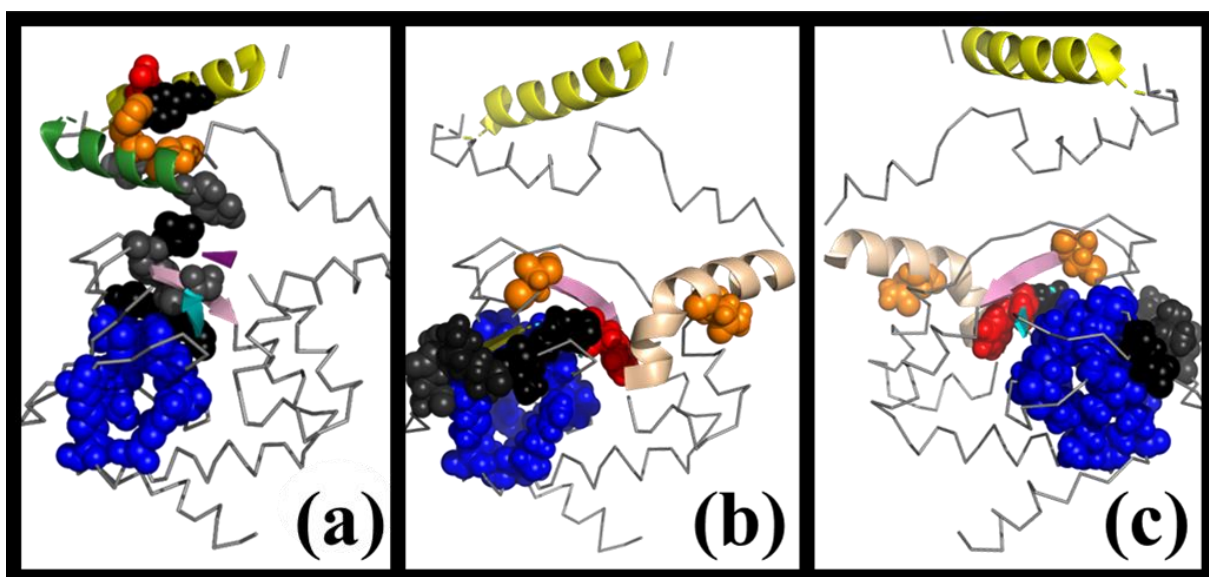


Figura 12. CNBD de HCN1 (a) sistema apo, (b) CNBD/cGMP y (c) CNBD/cGMP (vista alternativa). En amarillo, se señala la hélice 1. En verde, la hélice 2. En beige claro, la hélice 4. En púrpura profundo, la hoja beta plegada 1. En cian, la hoja beta plegada 2. En rosa, la hoja beta plegada 7. Como un conjunto de esferas azules, se encuentran los residuos que forman parte del sitio activo; como un conjunto de esferas rojas y naranjas, los sitios alostéricos más relevantes para cada sistema, y, como un conjunto de esferas negras y grises oscuras, los principales residuos críticos en la comunicación alostérica.

Nota: para el tercer sistema, nuevamente se dio prioridad a los residuos con valores altos de ACI que se encontraban a menos de 35 Å del centro de masa de la hélice 1. Un residuo es excluido del análisis (SER 552) por tratarse de un residuo adyacente al sitio activo (es decir: directamente conectado, mediante enlaces covalentes, a dos residuos del sitio activo). A continuación, en la tabla 4, son mencionados los valores medios de ACI para los potenciales sitios alostéricos.

Tabla 4. Potenciales sitios alostéricos, su distancia del centro de masa de la hélice 1 y su valor medio de ACI para el sistema CNBD/cGMP.

Residuo	r, Å	(ACI)_promedio
ASN 463	32.7	0.80
PHE 476	38.0	0.83
PRO 481	41.6	0.82
PHE 483	39.5	0.87
LEU 488	31.5	0.89
PHE 514	30.8	0.87
SER 552	33.7	0.93
CYS 559	24.6	0.84
VAL 572	37.4	0.89
LEU 573	40.4	0.83
TYR 576	42.7	0.81
PHE 583	40.6	0.85

Lo anterior indica que los residuos PHE 476, PRO 481, PHE 483, VAL 572, LEU 573, TYR 576 y PHE 583 también experimentan comunicación alostérica con el sitio activo; no obstante, se trata de residuos muy alejados del centro de masa de la hélice 1. De hecho, se encuentran a más de 15 Å del centro de masa de cualquier residuo de la hélice 1, a diferencia de los residuos que sí fueron incluidos. Además, como se ha comentado, SER 552 está unido, mediante enlaces covalentes, a dos residuos del sitio activo, por lo que se decidió no tomarlo en cuenta.

Por otra parte, se determinó que, de los 8 residuos críticos distintos al sitio activo y al sitio alostérico, 7 aportaban una contribución del 72%. Los residuos críticos principales están marcados como un conjunto de esferas negras o gris oscuro en las figuras 12b y 12c.

De éstos, el 71.4% (es decir: 5 de 7) se encontraba en estructuras secundarias bien definidas: uno (ILE 515) se encontraba dentro de la hoja beta plegada 2 (marcada en cian en la Figura 12), dos (VAL 522 e ILE 523) se encontraban dentro de la hélice 3 (marcada en verde oscuro en las Figuras 12a y 12b) y dos (LYS 531 y LEU 532) se encontraban dentro de la hoja beta plegada 4. Los dos residuos críticos principales restantes, GLU 505 y GLU 529, ambos tienen una carga eléctrica negativa. Nuevamente, al intentar encontrar una relación entre la naturaleza de los aminoácidos y su carácter de residuos críticos principales, se encuentra que 4 de 7 son hidrofóbicos y alifáticos, mientras que sólo 3 son polares y cargados eléctricamente.

Los anteriores hallazgos son resumidos en la tabla 5.

Tabla 5. Análisis ulterior acerca de los sitios alostéricos más relevantes y otros potenciales sitios alostéricos.

Sistema	CNBD (apo)	CNBD/cAMP	CNBD/cGMP
Distancia entre el sitio alostérico más relevante y la hélice 1	0 Å	> 10 Å, <15 Å	> 10 Å, <15 Å
Distancia entre el sitio alostérico más relevante y el centro de masa de la hélice 1	5.3 Å	31.5 Å	30.8 Å
Naturaleza del sitio alostérico más relevante	GLU, cargado eléctricamente	LEU, hidrofóbico y alifático	PHE, hidrofóbico y aromático
Tipo de motivo estructural en el que se encuentra	Hélice alfa	Hélice alfa	Hoja beta plegada
Distancia entre los potenciales sitios alostéricos y la hélice 1	<5 Å	> 5 Å, <15 Å	> 10 Å, <15 Å
Distancia entre los potenciales sitios alostéricos y el centro de masa de la hélice 1	9.4 Å, 10.3 Å	13.7 Å, 32.7 Å	32.7 Å, 24.6 Å
Naturaleza de los potenciales sitios alostéricos	MET, hidrofóbico y azufrado HIS, cargado eléctricamente	TYR, polar sin carga LEU, hidrofóbico y alifático	ASN, polar sin carga CYS, polar sin carga y azufrado
Tipo de motivo estructural en el que se encuentran	Hélice alfa	Hélice alfa	Hélice alfa y hoja beta plegada

Lo anterior muestra que la presencia de ligante desplaza el sitio alostérico hacia regiones más lejanas a la hélice 1, y hacia regiones más cercanas al sitio activo. Se muestra, de igual modo, que todos los sitios alostéricos más relevantes y los potenciales sitios alostéricos se encuentran dentro de motivos estructurales bien definidos (como hélices alfa u hojas beta plegadas). Análogamente, no se encuentra alguna relación entre la naturaleza química de los aminoácidos y su carácter de sitios alostéricos más relevantes o potenciales sitios alostéricos. Lo anterior se resume en la tabla 6.

Tabla 6. Análisis ulterior acerca de los residuos críticos principales.

Sistema	CNBD (apo)	CNBD/cAMP	CNBD/cGMP
Porcentaje de residuos críticos principales	$\left(\frac{8}{19}\right) \times (100\%) = 42.1\%$	$\left(\frac{7}{13}\right) \times (100\%) = 53.8\%$	$\left(\frac{7}{10}\right) \times (100\%) = 70.0\%$
Contribución	$\left(\frac{5.8}{9.8}\right) \times (100\%) = 59.2\%$	$\left(\frac{3.37}{6.25}\right) \times (100\%) = 54.0\%$	$\left(\frac{5.6}{7.8}\right) \times (100\%) = 72.0\%$
Contribución / Porcentaje	$\left(\frac{8}{19}\right) \times \left(\frac{9.8}{5.8}\right) = 0.71$	$\left(\frac{7}{13}\right) \times \left(\frac{6.25}{3.37}\right) = 1.00$	$\left(\frac{7}{10}\right) \times \left(\frac{7.8}{5.6}\right) = 0.98$
Naturaleza de los residuos críticos principales	ARG, cargado eléctricamente TYR (3 de 8), polar sin carga VAL, PHE, LEU e ILE, hidrofóbicos	LYS, cargado eléctricamente SER, polar sin carga LEU (3 de 7), PHE e ILE, hidrofóbicos	GLU (2 de 7) y LYS, cargados eléctricamente ILE (2 de 7), LEU y VAL, hidrofóbicos y alifáticos
Porcentaje de residuos críticos principales en motivos estructurales bien definidos	$\left(\frac{7}{8}\right) \times (100\%) = 87.5\%$	$\left(\frac{5}{7}\right) \times (100\%) = 71.4\%$	$\left(\frac{5}{7}\right) \times (100\%) = 71.4\%$

Por otra parte, se encuentra que la relación entre la contribución de los residuos críticos principales y el porcentaje que éstos representan tiende a ser 1 en presencia de ligantes. Nuevamente, no existe una relación entre el carácter polar o hidrofóbico de los aminoácidos y su condición de residuos críticos principales. Además, la gran mayoría de los residuos críticos principales se encuentran en motivos estructurales bien definidos.

Las observaciones respecto a la presencia de los sitios alostéricos y de los residuos críticos principales sobre hélices alfa y hojas beta, que confieren rigidez a la proteína, parecen indicar que una rigidez estructural facilita la propagación de perturbaciones.

Por otra parte, si bien ni todos los sitios alostéricos más relevantes ni todos los potenciales sitios alostéricos ni todos los residuos críticos son polares, se puede especular que, cuando sí lo son, se debe a que éstos son susceptibles a experimentar cambios conformacionales ante movimientos de otros aminoácidos polares que se encuentren cerca, ya sea a través de una atracción o una repulsión electrostática. Por otra parte, cuando son no-polares, la comunicación

alostérica podría verse favorecida ante movimientos de otros residuos en el interior de la proteína que propicien el contacto del agua con los residuos hidrófobos.

6.7 Mediciones de RMSF

Las mediciones de la raíz de la fluctuación cuadrática media (RMSF por las siglas en inglés de *Root Mean Square Fluctuation*) sirven para medir la fluctuación promedio, a lo largo de una trayectoria, de un conjunto de átomos respecto a una posición de referencia. Valores bajos de RMSF para un conjunto de átomos indican mayor estabilidad estructural y mayor rigidez de determinados motivos estructurales.

Al comparar los valores de RMSF obtenidos para los diferentes sistemas, fue posible demostrar que los sistemas CNBD/cAMP y CNBD/cGMP presentan valores de RMSF significativamente mayores que el sistema CNBD (apo) para todos los residuos de la hélice 1, mientras que, entre los valores de RMSF de los sistemas CNBD/cAMP y CNBD/cGMP no hay una diferencia estadísticamente significativa.

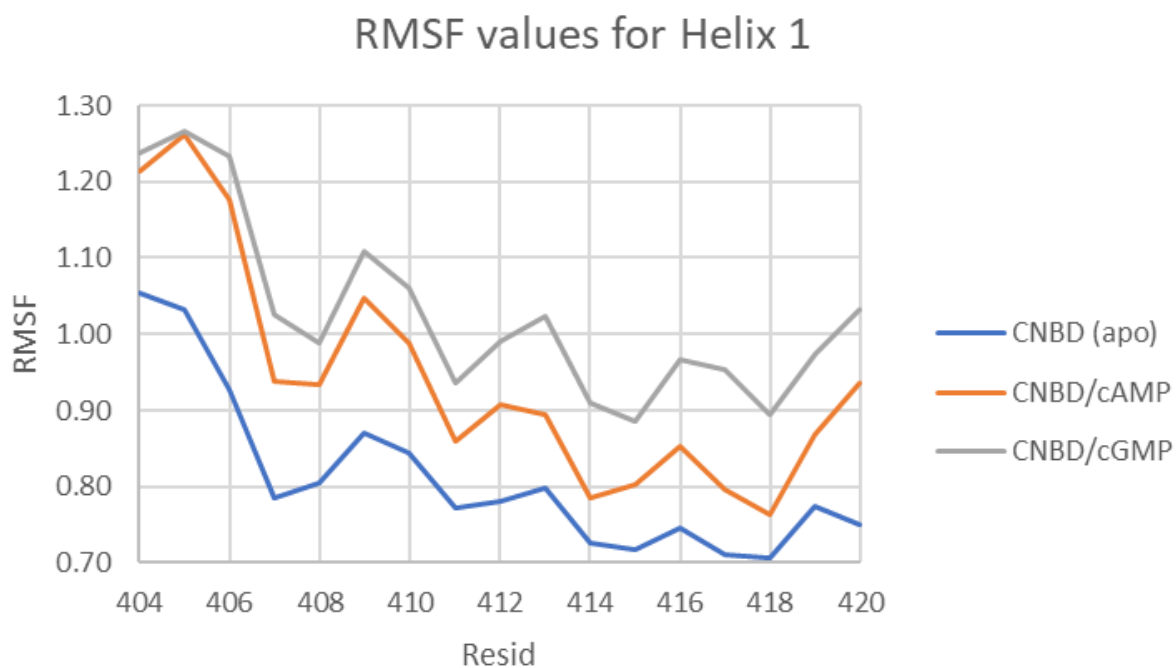


Figura 13. Valores de RMSF para los diferentes residuos de la hélice 1 en los diferentes sistemas, CNBD (apo), CNBD/cAMP y CNBD/cGMP.

Los últimos resultados expuestos son interpretados del siguiente modo: los valores de ACI para la hélice 1 de los sistemas CNBD/cAMP y CNBD/cGMP son considerablemente menores que para la hélice 1 del sistema CNBD (apo), mientras que el patrón contrario se observa en los valores de RMSF. Como consecuencia, la unión a ligante ocasiona una disminución en la conectividad de la hélice 1 con otros motivos estructurales de la proteína.

7. Conclusiones

1. Las simulaciones de dinámica molecular generan conformaciones en concordancia con las condiciones fisiológicas de presión, temperatura y concentración salina. Dichas conformaciones del sistema proteico son más adecuadas para utilizar el algoritmo de propagación de perturbaciones como el incluido en el programa Ohm.
2. La comunicación alostérica ocurre generalmente siguiendo un camino que se aprovecha de la rigidez estructural que confieren determinados motivos estructurales, como lo son las hélices alfa y las hojas beta, a diferencia de segmentos sin una estructura secundaria definida. Por ello, la mayoría de sitios alostéricos más relevantes y potenciales sitios alostéricos, al igual que la mayoría de los residuos críticos principales se encuentran ubicados en hélices alfa u hojas beta.
3. Tanto residuos polares como residuos no-polares contribuyen a la comunicación alostérica en una medida prácticamente igual. Los residuos polares con mayor probabilidad experimentan desplazamientos ante el desplazamiento de otros residuos polares o ante su exposición ante éstos, mientras que los residuos no-polares con mayor probabilidad experimentan desplazamientos ante su exposición al agua.
4. La presencia de ligantes, cAMP y cGMP, cambia considerablemente las vías alostéricas, haciendo que se pierda la comunicación alostérica entre la hélice 1 (contigua a la hélice transmembrana S6) y el sitio activo (sitio de unión a nucleótido cíclicos).
5. A su vez, la presencia de ligantes se ve reflejada en mayores fluctuaciones de los residuos de la hélice 1, según puede observarse en el análisis de valores de RMSF.
6. Lo anterior parece indicar que un fármaco capaz de modular al CNBD de manera similar a como lo hacen el cAMP y el cGMP (que incrementan la probabilidad de abertura del poro acelerando la activación y desacelerando la desactivación) deben, a su vez, ocasionar una pérdida significativa de la comunicación alostérica entre el sitio activo y la hélice 1, a la vez que ocasionan un movimiento significativamente mayor en los residuos de dicha hélice.

8. Perspectivas

1. Es posible evaluar en un futuro próximo la forma en la que la unión de otros varios ligantes que afectan la actividad de los HCN modifican también la comunicación alostérica dentro del CNBD.
2. La forma de análisis propuesta para la comunicación alostérica se puede automatizar por medio del desarrollo de un software adecuado que tome como entrada la información de salida brindada por el software Ohm.
3. También se desea emplear este método de análisis de la comunicación alostérica para medir el efecto de ciertas mutaciones en diferentes proteínas, incluidas algunos canales iónicos y la proteína globular epóxido-hidrolasa soluble (sEH).

9. Referencias.

- [1] **Whitford, D.** *Proteins: Structure and Function*; John Wiley & Sons: Chichester, **2013**.
- [2] **Rose, W. C.** The Nutritive Significance of the Amino Acids. *Physiol. Rev.* **1938**, *18*(1), 109–136.
- [3] **Mitaku, S.; Hirokawa, T.; Tsuji, T.** Amphiphilicity Index of Polar Amino Acids as an Aid in the Characterization of Amino Acid Preference at Membrane-Water Interfaces. *Bioinformatics* **2002**, *18*(4), 608–616.
- [4] **Kumar, S.; Bansal, M.** Structural and Sequence Characteristics of Long Alpha Helices in Globular Proteins. *Biophys. J.* **1996**, *71*(3), 1574–1586.
- [5] **Smith, C. K.** Studies on the Factors Influencing Beta-Sheet Stability. Ph.D. Thesis, Yale University, **1997**.
- [6] **Rehman, I.; Kerndt, C. C.; Botelho, S.** Biochemistry, Tertiary Protein Structure. *StatPearls [Internet]*, StatPearls Publishing: Treasure Island, FL, **2022**.
- [7] **Tusnády, G. E.; Dosztányi, Z.; Simon, I.** Transmembrane Proteins in the Protein Data Bank: Identification and Classification. *Bioinformatics* **2004**, *20*(17), 2964–2972.
- [8] **Hille, B.; Armstrong, C.; MacKinnon, R.** Ion Channels: From Idea to Reality. *Nat. Med.* **1999**, *5*, 1105–1109.
- [9] **Armstrong, C. M.** Voltage-Dependent Ion Channels and Their Gating. *Physiol. Rev.* **1992**, *72*(suppl_4), S5–S13.
- [10] **Lee, C.-H.; MacKinnon, R.** Structures of the Human HCN1 Hyperpolarization-Activated Channel. *Cell* **2017**, *168*(1–2), 111–120.e11
- [11] **Newton, R. P.; Smith, C. J.** Cyclic Nucleotides. *Phytochemistry* **2004**, *65*(17), 2423–2437.
- [12] **Colquitt, R. B.; Colquhoun, D. A.; Thiele, R. H.** In Silico Modelling of Physiologic Systems. *Best Pract. Res. Clin. Anaesthesiol.* **2011**, *25*(4), 499–510.

- [13] **Desmeulles, G.; Querrec, G.; Redou, P.; Kerdelo, S.; Misery, L.; Rodin, V.; Tisseau, J.** The Virtual Reality Applied to Biology Understanding: The *In Virtuo* Experimentation. *Expert Syst. Appl.* **2006**, *30*(1), 82–92.
- [14] **Hollingsworth, S. A.; Dror, R. O.** Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*(6), 1129–1143.
- [15] **Morris, G. M.; Lim-Wilby, M.** Molecular Docking. In *Molecular Modeling of Proteins*; Humana Press: Totowa, NJ, **2008**; pp 365–382.
- [16] **Krieger, E.; Nabuurs, S. B.; Vriend, G.** Homology Modeling. In *Structural Bioinformatics*; Humana Press: Totowa, NJ, **2003**; Vol. 44, pp 509–523.
- [17] **Sartiani, L.; Mannaioni, G.; Masi, A.; Romanelli, M. N.; Cerbai, E.** The Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels: From Biophysics to Pharmacology of a Unique Family of Ion Channels. *Pharmacol. Rev.* **2017**, *69* (4), 354–395.
- [18] **DiFrancesco, J. C.; DiFrancesco, D.** Dysfunctional HCN Ion Channels in Neurological Diseases. *Front. Cell. Neurosci.* **2015**, *6*, 174.
- [19] **Rivolta, I.; Binda, A.; Masi, A.; DiFrancesco, J. C.** Cardiac and Neuronal HCN Channelopathies. *Pflügers Arch. Eur. J. Physiol.* **2020**, *472*, 931–951.
- [20] **Changeux, J. P.** Allosterity and the Monod-Wyman-Changeux Model after 50 Years. *Annu. Rev. Biophys.* **2012**, *41*, 103–133.
- [21] **Changeux, J. P.; Edelstein, S.** Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *FI000 Biol. Rep.* **2011**, *3*.
- [22] **Wang, J.; Jain, A.; McDonald, L. R.; Gambogi, C.; Lee, A. L.; Dokholyan, N. V.** Mapping Allosteric Communications within Individual Proteins. *Nat. Commun.* **2020**, *11* (1), 3862.
- [23] **Vallverdú, J.** What Are Simulations? An Epistemological Approach. *Procedia Technol.* **2014**, *13*, 6–15.

[24] **Rosen, G.** What Is Constructive Empiricism? *Philos. Stud.* **1994**, *74* (2), 143–178.

[25] **Yamauchi, M.; Mori, Y.; Okumura, H.** Molecular Simulations by Generalized-Ensemble Algorithms in Isothermal–Isobaric Ensemble. *Biophys. Rev.* **2019**, *11*, 457–469.

[26] **Box, G. E.; Tiao, G. C.** *Bayesian Inference in Statistical Analysis*; John Wiley & Sons: Hoboken, NJ, 2011.

10. Apéndices

10.1 Apéndice 1. Funciones que describen el comportamiento cinético de la corriente de hiperpolarización

Las funciones del tipo $I_h = f(t)$ cumplen con las siguientes condiciones:

$$\frac{dI_h}{dt} > 0 \forall t \geq 0 \quad (1),$$

$$\lim_{t \rightarrow \infty} I_h = I_{max} \quad (2),$$

$$I_h(t = 0) = 0 \quad (3),$$

$$I_h(t = \tau) = \frac{I_{max}}{2} \quad (4).$$

Ecuación propuesta

$$I_h = I_{max} \cdot (1 - 2^{-t/\tau}) \quad (5).$$

Prueba de la condición 1:

$$\frac{dI_h}{dt} = -I_{max} \cdot \frac{d}{dt} 2^{-t/\tau} = I_{max} \cdot (\ln 2 / \tau) \cdot 2^{-t/\tau} > 0 \forall t \geq 0 \quad (6).$$

Prueba de la condición 2:

$$\lim_{t \rightarrow \infty} I_h = I_{max} \cdot \lim_{t \rightarrow \infty} (1 - 2^{-t/\tau}) = I_{max} \cdot \left(1 - \lim_{t \rightarrow \infty} 2^{-t/\tau}\right) = I_{max} \quad (7).$$

Prueba de la condición 3:

$$I_h(t = 0) = I_{max} \cdot (1 - 2^{-0/\tau}) = I_{max} \cdot (1 - 1) = 0 \quad (8).$$

Prueba de la condición 4:

$$I_h(t = \tau) = I_{max} \cdot (1 - 2^{-\tau/\tau}) = I_{max} \cdot (1 - 2^{-1}) = \frac{I_{max}}{2} \quad (9).$$

Ahora bien, se propondrá a continuación una nueva función $I_h = f(t)$.

$$I_h = c_1 + c_2 \cdot \exp[c_3 \cdot \exp(c_4 t)] \quad (10),$$

donde ni c_2 , ni c_3 , ni c_4 son iguales a 0, además de que c_1 , c_2 , c_3 y c_4 adoptan valores forzosamente reales y finitos.

Se tiene

$$\frac{d}{dt} I_h = c_2 \cdot \frac{d \exp[c_3 \cdot \exp(c_4 t)]}{d[\exp(c_4 t)]} \cdot \frac{d}{dt} \exp(c_4 t) = c_2 c_3 c_4 \cdot \exp[c_3 \cdot \exp(c_4 t)] \cdot \exp(c_4 t) \quad (11),$$

a modo que $\frac{d}{dt} I_h > 0$ si y sólo si $c_2 c_3 c_4 > 0$, lo cual implica que c_2 , c_3 y c_4 tienen todos valores positivos, o que sólo una constante entre c_2 , c_3 y c_4 es positiva, mientras que las demás son negativas.

Por otra parte, se tiene

$$\lim_{t \rightarrow \infty} I_h = c_1 + c_2 \cdot \lim_{t \rightarrow \infty} \exp[c_3 \cdot \exp(c_4 t)] \quad (12).$$

Evidentemente, si $c_3 > 0$ y $c_4 > 0$, entonces

$$\lim_{t \rightarrow \infty} \exp[c_3 \cdot \exp(c_4 t)] = \infty \quad (13),$$

lo cual no tendría sentido (pues I_h es finito, al igual que c_1 , c_2 y c_3).

Por otra parte, si $c_3 < 0$ y $c_4 > 0$, entonces

$$\lim_{t \rightarrow \infty} \exp[c_3 \cdot \exp(c_4 t)] = \lim_{x \rightarrow \infty} \exp(-x) = 0 \quad (14),$$

a modo que, para que la condición 2 se cumpla, debería también tenerse

$$\lim_{t \rightarrow \infty} I_h = c_1 = I_{max} \quad (15).$$

Si, por otra parte, $c_4 < 0$, entonces

$$\lim_{t \rightarrow \infty} \exp[c_3 \cdot \exp(c_4 t)] = 1 \quad (16),$$

por lo que debería cumplirse

$$\lim_{t \rightarrow \infty} I_h = c_1 + c_2 = I_{max} \quad (17).$$

Esto limita del siguiente modo la naturaleza de las constantes: c_3 y c_4 no pueden ser ambas positivas. Y, por otra parte, c_2 debe ser negativa si $c_3 c_4 < 0$, mientras que debe ser positiva si $c_3 c_4 > 0$.

De acuerdo con la ecuación propuesta, se tiene

$$I_h(t = 0) = c_1 + c_2 \cdot \exp(c_3) \quad (18).$$

Por ende, para que se cumpliera la condición 3, debería tenerse

$$c_3 = \ln\left(-\frac{c_1}{c_2}\right) \quad (19).$$

Recordando que c_3 es distinto de 0 y que, al igual que c_1 , c_2 y c_4 , adopta un valor finito y real, deducimos que debe cumplirse la inecuación $\frac{c_1}{c_2} < 0$ (de otro modo, c_3 tendría que ser imaginario o complejo); esto implica que c_1 y c_2 no pueden ser ambos positivos ni ambos negativos. Además, se infiere que $-\frac{c_1}{c_2} \neq 1$ (de lo contrario, se tendría $c_3 = 0$). Adicionalmente, c_1 , al igual que c_2 , c_3 y c_4 , debe ser distinto de 0 (de otro modo, se tendría $c_3 = -\infty$).

Por otra parte, para que c_3 adopte un valor negativo, debe cumplirse la inecuación $|c_1| < |c_2|$, mientras que, para que adopte un valor positivo debe cumplirse la inecuación $|c_1| > |c_2|$.

Recapitulando, para que se cumplan las condiciones 1 a 3, debe cumplirse lo siguiente:

$$\begin{aligned} & \left[(c_1, c_2, c_3, c_4) \in \mathbb{R}^4 \wedge c_1 c_2 c_3 c_4 \neq 0 \wedge c_3 = \ln\left(-\frac{c_1}{c_2}\right) \right] \wedge \\ & \wedge [(c_1 = I_{max} \wedge c_2 < 0 \wedge c_3 < 0 \wedge c_4 > 0) \vee \\ & \vee (c_1 + c_2 = I_{max} \wedge c_2 c_3 < 0 \wedge c_4 < 0)] \end{aligned}$$

Lo anterior puede reescribirse del siguiente modo exclusivamente en términos de c_1 , c_2 y c_4 :

$$\begin{aligned} & [(c_1, c_2, c_4) \in \mathbb{R}^3 \wedge c_1 c_2 c_4 \neq 0 \wedge c_1 c_2 < 0] \wedge \{(c_1 = I_{max} \wedge c_2 < 0 \wedge |c_1| < |c_2| \wedge c_4 > 0) \vee \\ & \vee \{(c_1 + c_2 = I_{max} \wedge c_4 < 0) \wedge [(c_2 < 0 \wedge |c_1| > |c_2|) \vee (c_2 > 0 \wedge |c_1| < |c_2|)]\}\} \end{aligned}$$

Además, la ecuación 10 puede ser reescrita del siguiente modo:

$$I_h = c_1 + c_2 \cdot \exp \left[\ln \left(-\frac{c_1}{c_2} \right) \cdot \exp(c_4 t) \right] = c_1 + c_2 \cdot \left(-\frac{c_1}{c_2} \right)^{\exp(c_4 t)} \quad (20).$$

Sólo por conveniencia, se definirá una constante c_5 mediante la ecuación $c_5 = -\frac{c_1}{c_2}$, a modo que la ecuación 20 pueda ser reescrita del siguiente modo:

$$I_h = c_1 \cdot \left[1 - c_5^{\exp(c_4 t) - 1} \right] \quad (21),$$

donde se sabe que c_5 es una magnitud positiva diferente de 1.

Las condiciones antes expuestas para c_1 , c_2 y c_4 pueden expresarse del siguiente modo en términos de c_1 , c_4 y c_5 :

$$\begin{aligned} & [(c_1, c_4, c_5) \in \mathbb{R}^3 \wedge c_1 c_4 c_5 \neq 0 \wedge c_5 > 0] \wedge \{(c_1 = I_{max} \wedge c_4 > 0 \wedge c_5 < 1) \vee \\ & \vee \left\{ \left[c_1 \cdot \left(1 - \frac{1}{c_5} \right) = I_{max} \wedge c_4 < 0 \right] \wedge [(c_1 > 0 \wedge c_5 > 1) \vee (c_1 < 0 \wedge c_5 < 1)] \right\} \} \end{aligned}$$

Finalmente, se evaluará la compatibilidad de las ecuaciones 4 y 21. En caso de ser completamente compatibles, desde luego, debería cumplirse lo siguiente:

$$c_1 \cdot \left[1 - c_5^{\exp(c_4 \tau) - 1} \right] = \frac{I_{max}}{2} \quad (22).$$

Evidentemente, si $c_4 > 0$ y $c_5 < 1$, luego la ecuación 22 implica la siguiente:

$$c_5^{\exp(c_4 \tau) - 1} = \frac{1}{2} \quad (23),$$

que, a su vez, implica la siguiente:

$$(\ln c_5)[\exp(c_4 \tau) - 1] = -\ln 2 \quad (24).$$

Es posible, entonces, inferir que

$$\exp(c_4 \tau) - 1 = -\frac{\ln 2}{\ln c_5} \quad (25).$$

Como consecuencia,

$$\exp(c_4\tau) = \frac{\ln(c_5/2)}{\ln c_5} = \log_{c_5}(c_5/2) \quad (26).$$

De lo anterior se infiere con facilidad que

$$c_4 = \frac{1}{\tau} \cdot \ln[\log_{c_5}(c_5/2)] \quad (27).$$

Lo anterior implica, desde luego lo siguiente:

$$\exp(c_4t) = [\log_{c_5}(c_5/2)]^{t/\tau} \quad (28).$$

De este modo,

$$c_5^{\exp(c_4t)-1} = \frac{1}{c_5} \cdot c_5^{[\log_{c_5}(c_5/2)]^{t/\tau}} \quad (29),$$

y entonces

$$I_h = I_{max} \cdot \left\{ 1 - \frac{1}{c_5} \cdot c_5^{[\log_{c_5}(c_5/2)]^{t/\tau}} \right\} \quad (30),$$

donde $0 < c_5 < 1$.

En el caso en que $c_1 \cdot \left(1 - \frac{1}{c_5}\right) = I_{max}$, por otra parte, tenemos que la ecuación 22 implica la siguiente:

$$1 - c_5^{\exp(c_4\tau)-1} = \frac{1}{2} \cdot \left(1 - \frac{1}{c_5}\right) \quad (31).$$

Como consecuencia,

$$c_5^{\exp(c_4\tau)} = \frac{c_5+1}{2} \quad (32)$$

y luego

$$c_4 = \frac{1}{\tau} \cdot \ln \left[\log_{c_5} \left(\frac{c_5+1}{2} \right) \right] \quad (33).$$

Por ende,

$$\exp(c_4 t) = \left[\log_{c_5} \left(\frac{c_5 + 1}{2} \right) \right]^{t/\tau} \quad (34)$$

y luego

$$I_h = \frac{I_{max}}{c_5 - 1} \cdot \left\{ c_5 - c_5^{\left[\log_{c_5} \left(\frac{c_5 + 1}{2} \right) \right]^{t/\tau}} \right\} \quad (35).$$

Si definimos a mediante la ecuación $a = c_5$, es fácil inferir que, en el caso especial en que $a \gg 1$, se cumple de manera aproximada

$$1 - \frac{I_h}{I_{max}} = \frac{1}{a} \cdot a^{\left[\log_a(a/2) \right]^{t/\tau}} \quad (36).$$

Es evidente que esta misma ecuación podría ser obtenida desde la ecuación 30, así que se considerará que, en dicha ecuación, a también puede ser mayor a 1.

Introduciendo una corrección igual a $-r$ en t y τ , resulta

$$1 - \frac{I_h}{I_{max}} = \frac{1}{a} \cdot a^{\left[\log_a(a/2) \right]^{(t-r)/(\tau-r)}} \quad (37).$$

De este modo, se obtiene con facilidad

$$\ln \left(1 - \frac{I_h}{I_{max}} \right) = (\ln a) \cdot \left\{ \left[\log_a(a/2) \right]^{\frac{t-r}{\tau-r}} - 1 \right\} \quad (38).$$

Derivando ambas funciones respecto a t , resulta

$$\frac{-1}{I_{max} - I_h} \cdot \frac{dI_h}{dt} = \left\{ \frac{(\ln a) \ln \left[\log_a(a/2) \right]}{\tau - r} \right\} \cdot \left[\log_a(a/2) \right]^{\frac{t-r}{\tau-r}} \quad (39).$$

Como consecuencia,

$$\begin{aligned} \ln \left(\frac{dI_h}{dt} \right) - \ln(I_{max} - I_h) &= \frac{t}{\tau - r} \cdot \ln \left[\log_a \left(\frac{a}{2} \right) \right] + \ln \left\{ \frac{-(\ln a) \cdot \ln \left[\log_a \left(\frac{a}{2} \right) \right]}{\tau - r} \right\} + \\ &+ (-1) \frac{r}{\tau - r} \cdot \ln \left[\log_a(a/2) \right] \end{aligned} \quad (40).$$

Evidentemente, si se define m mediante la siguiente ecuación:

$$m = \frac{1}{\tau-r} \cdot \ln[\log_a(a/2)] \quad (41),$$

a la vez que b se define del siguiente modo:

$$b = \ln \left\{ \frac{-(\ln a) \cdot \ln[\log_a(a/2)]}{\tau-r} \right\} - \frac{r}{\tau-r} \cdot \ln[\log_a(a/2)] \quad (42),$$

entonces la ecuación 40 puede ser reescrita del siguiente modo:

$$\ln(dI_h/dt) - \ln(I_{max} - I_h) = mt + b \quad (43).$$

Es fácil demostrar, desde la ecuación 41, las siguientes ecuaciones:

$$\ln[\log_a(a/2)] = (\tau - r)m \quad (44)$$

y

$$\ln a = \frac{\ln 2}{1 - \exp[(\tau-r)m]} \quad (45).$$

Sustituyendo en la ecuación 42 $\ln[\log_a(a/2)]$ y $\ln a$ por los miembros derechos de las ecuaciones 44 y 45, resulta

$$b = \ln \left\{ \frac{(-\ln 2)m}{1 - \exp[(\tau-r)m]} \right\} - mr \quad (46).$$

De lo anterior se deduce con facilidad

$$\exp(b) = \frac{(-\ln 2)m}{\exp(mr) - \exp(m\tau)} \quad (47)$$

y luego

$$r = \frac{1}{m} \cdot \ln[\exp(m\tau) - (\ln 2)m \exp(-b)] \quad (48).$$

Desde las ecuaciones 45 y 48 es posible, entonces, inferir

$$\ln a = \frac{(\ln 2) \exp(mr)}{\exp(mr) - \exp(m\tau)} = \ln 2 - \frac{1}{m} \cdot \exp(m\tau + b) \quad (49).$$

Dicho de otro modo,

$$a = 2 \cdot \exp \left[-\frac{1}{m} \cdot \exp(m\tau + b) \right] \quad (50).$$

Las ecuaciones 48 y 50 ofrecen una forma de calcular los parámetros a y r partiendo del conocimiento de los parámetros m , b y τ .

Desde luego, aunque el valor experimental de τ puede distar mucho del real debido al error aleatorio, se sugiere tomar dicho valor como verdadero como una primera aproximación.

Por otra parte, en el caso en que la dependencia $I_h = f(t)$ puede ser descrita por medio de la ecuación 35, se cumple la siguiente ecuación si $\frac{1}{\alpha} = c_5 \ll 1$:

$$\frac{I_h}{I_{max}} = \left(\frac{1}{\alpha} \right)^{[\log_{\alpha}(2)]^{(t-r)/(\tau-r)}} \quad (51).$$

Como consecuencia,

$$\frac{1}{I_{max}} \cdot \frac{dI_h}{dt} = \left(\frac{1}{\alpha} \right)^{[\log_{\alpha}(2)]^{(t-r)/(\tau-r)}} \cdot \frac{(\ln \alpha) \{-\ln[\log_{\alpha}(2)]\}}{\tau-r} \cdot [\log_{\alpha}(2)]^{(t-r)/(\tau-r)} \quad (52).$$

De lo anterior se deduce con facilidad que

$$\frac{dI_h/dt}{I_h} = \frac{(\ln \alpha) \{-\ln[\log_{\alpha}(2)]\}}{\tau-r} \cdot [\log_{\alpha}(2)]^{(t-r)/(\tau-r)} \quad (53),$$

a modo que

$$\ln \left(\frac{dI_h/dt}{I_h} \right) = \frac{t}{\tau-r} \cdot \ln[\log_{\alpha}(2)] + \ln \left\{ \frac{(\ln \alpha) \{-\ln[\log_{\alpha}(2)]\}}{\tau-r} \right\} - \frac{r}{\tau-r} \cdot \ln[\log_{\alpha}(2)] \quad (54).$$

Evidentemente, si definimos m mediante la siguiente ecuación:

$$m = \frac{1}{\tau-r} \cdot \ln[\log_{\alpha}(2)] \quad (55),$$

a la vez que b se define del siguiente modo:

$$b = \ln \left\{ \frac{(\ln \alpha) \{-\ln[\log_{\alpha}(2)]\}}{\tau-r} \right\} - \frac{r}{\tau-r} \cdot \ln[\log_{\alpha}(2)] \quad (56),$$

luego se tiene

$$\ln[\log_{\alpha}(2)] = (\tau - r)m \quad (57),$$

$$\ln \alpha = \ln 2 \cdot \exp[-(\tau - r)m] \quad (58),$$

Sustituyendo $\ln[\log_{\alpha}(2)]$ por $(\tau - r)m$ y $\ln \alpha$ por $(\ln 2) \cdot \exp[-(\tau - r)m]$, en la ecuación 56, resulta

$$b = \ln[(\ln 2)(-m)] - m\tau \quad (59).$$

Evidentemente, la anterior ecuación permite calcular el valor de τ . Específicamente, se tiene

$$\tau = \frac{1}{m} \ln[(\ln 2)(-m)] - \frac{b}{m} \quad (60).$$

Por otra parte, para determinar el valor de α , se sugiere el uso de la siguiente ecuación:

$$\alpha = 2^{\exp(mr)/\exp(m\tau)} = 2^{\exp(mr+b)/[(\ln 2)(-m)]} \quad (61).$$

Evidentemente, este método no es útil para calcular el valor de r . Por ende, se propondrá para su determinación un ajuste a la siguiente ecuación:

$$\ln\left(1 - \frac{I_h}{I_{max}}\right) = -\frac{t}{\tau-r} \cdot \ln(2) + \frac{r}{\tau-r} \cdot \ln(2) \quad (62).$$

Desde luego, si se define m mediante

$$m = -\frac{\ln(2)}{\tau-r} \quad (63),$$

así como b mediante

$$b = \frac{\ln(2)}{\tau-r} \cdot r \quad (64),$$

luego es posible determinar el valor de r mediante la siguiente ecuación:

$$r = -\frac{b}{m} \quad (65).$$

Adicionalmente, debido a que

$$\tau = -\frac{1}{m} \ln(2) + r \quad (66),$$

luego

$$\tau = -\frac{1}{m} \cdot [\ln(2) + b] \quad (67).$$

Por otra parte, para los valores de c_5 que no sean tan diferentes de 1, se propondrá realizar un ajuste no lineal respecto a los parámetros. La ecuación de ajuste será la siguiente:

$$\frac{I_h}{I_{max}} = \frac{1}{\beta-1} \cdot \left\{ \beta - \beta \left[\log_{\beta} \left(\frac{\beta+1}{2} \right) \right]^{\frac{t-r}{t-r}} \right\} \quad (68).$$

10.2 Apéndice 2. Función de partición del ensamble isotérmico-isobárico

Para las simulaciones de Dinámica Molecular, han sido desarrollados “algoritmos de ensamble generalizado”. Estos fueron desarrollados originalmente para el ensamble canónico, en el cual “el volumen de la caja de simulación es constante, y la presión no es controlada” [25]. Sin embargo, las condiciones experimentales generalmente involucran un control de la presión y de la temperatura. Lo anterior indica la importancia del ensamble isotérmico-isobárico para la descripción de los resultados experimentales.

Además de lo anterior, cabe mencionar que la energía libre de Gibbs se relaciona con la función de partición del ensamble isotérmico isobárico a través de la siguiente ecuación:

$$G = -k_B T \cdot \ln \Delta \quad (1).$$

A continuación, se ofrece una descripción matemática de cómo la función de partición del ensamble isotérmico-isobárico puede ser aplicada a sistemas tan complejos como una macromolécula rodeada de disolvente, de iones y posiblemente embebida por lípidos.

Es preciso recordar que la función de partición correspondiente a dicho ensamble es la siguiente:

$$\Delta(N, T, P) = \sum_V Q(N, V, T) \exp[-PV/(kT)] \quad (2),$$

donde $Q(N, V, T)$ es la función de partición del ensamble canónico.

Es posible suponer que la energía de cada estado disponible para los sistemas del ensamble canónico está determinada por la siguiente ecuación:

$$E = E_k + U = \frac{1}{2} \cdot \sum_{i=1}^N m_i \cdot (v_{ix}^2 + v_{iy}^2 + v_{iz}^2) + U(\{\mathbf{r}_i\}, V) \quad (3);$$

como consecuencia, es posible expresar

$$Q(N, V, T) = \frac{1}{N!} \cdot \frac{1}{h^{3N}} (\prod_{i=1}^N m_i^3) \times [\prod_{i=1}^N (\int_{-\infty}^{\infty} dv_{ix}) (\int_{-\infty}^{\infty} dv_{iy}) (\int_{-\infty}^{\infty} dv_{iz}) (\prod_{i=1}^N \int d\mathbf{r}_i)] \times \\ \times \left\{ \exp \left[-\frac{\sum_{i=1}^N m_i (v_{ix}^2 + v_{iy}^2 + v_{iz}^2)}{2kT} \right] \right\} \left[(\prod_{i=1}^N \int d\mathbf{r}_i) \exp \left(-\frac{U}{kT} \right) \right] \quad (4),$$

donde h es la constante de acción de Planck y cada $\int_{-\infty}^{\infty} dv_{ij}$, al igual que cada $\prod_{i=1}^N \int d\mathbf{r}_i$ es un operador.

Ahora bien, la integral múltiple presente en el miembro derecho de la ecuación 4 no puede simplificarse de manera significativa debido a que no es posible sostener que cada \mathbf{v}_i sea independiente entre sí cuando el sistema descrito es una proteína en una solución salina y posiblemente embebido en una membrana lipídica, del mismo modo en que no es posible sostener que cada \mathbf{r}_i sea independiente entre sí para el mismo sistema.

Además, no es posible afirmar que cada \mathbf{v}_i es completamente independiente de \mathbf{r}_i debido a las restricciones que la posición específica de una partícula impone sobre la velocidad de éste (por ejemplo, si está embebido por lípidos, su velocidad será diferente).

El valor de la integral anterior depende de diversos aspectos, como la geometría del sistema. Debido a lo anterior y a la forma de la ecuación de $U = F(\{\mathbf{r}_i\})$, su solución no es trivial.

A pesar de lo anterior, sí es posible encontrar otras expresiones posiblemente útiles para la anterior integral múltiple. Por ejemplo, si tenemos un sistema confinado a una caja prismática de dimensiones $K \times L \times M$ con centro en el origen, entonces es posible obtener la siguiente igualdad:

$$\prod_{i=1}^N \int d\mathbf{r}_i = \prod_{i=1}^N \left(\int_{-J/2}^{J/2} dx_i \right) \left(\int_{-L/2}^{L/2} dy_i \right) \left(\int_{-M/2}^{M/2} dz_i \right) \quad (5),$$

donde cada $\int_{-J/2}^{J/2} dx_i$, $\int_{-L/2}^{L/2} dy_i$ y $\int_{-M/2}^{M/2} dz_i$ es un operador, no una integral definida en sí.

Si se define por conveniencia $s_{3i-2} = \frac{2x_i}{K}$, $s_{3i-1} = \frac{2y_i}{L}$, $s_{3i} = \frac{2z_i}{M}$, entonces es posible expresar

$$\prod_{i=1}^N \int d\mathbf{r}_i = \frac{8^N}{V^N} \cdot \prod_{i=1}^{3N} \left(\int_{-1}^1 ds_i \right) \quad (6),$$

donde nuevamente cada $\int_{-1}^1 ds_i$ es un operador, no una integral definida en sí.

Sólo por conveniencia, se definirá la función $F(N, V, T)$ mediante la siguiente ecuación:

$$F(N, V, T) = \left[\prod_{i=1}^N \left(\int_{-\infty}^{\infty} dv_{ix} \right) \left(\int_{-\infty}^{\infty} dv_{iy} \right) \left(\int_{-\infty}^{\infty} dv_{iz} \right) \right] \left[\prod_{i=1}^{3N} \left(\int_{-1}^1 ds_i \right) \right] \times$$

$$\times \exp \left[-\frac{\sum_{i=1}^N m_i (v_{ix}^2 + v_{iy}^2 + v_{iz}^2)}{2kT} \right] \exp(-U/kT) \quad (7),$$

de tal modo que se tenga $Q(N, V, T) = \left(\frac{8}{V}\right)^N \cdot F(N, V, T)$ y luego

$$\Delta(N, T, P) = 8^N \cdot \int_{\min(V)}^{\max(V)} \frac{1}{V^N} \cdot G(N, V, T) \exp[-PV/(kT)] dV \quad (8),$$

donde $\min(V)$ y $\max(V)$ son los valores mínimo y máximo, respectivamente, posibles para el sistema durante la simulación de dinámica molecular.

Para encontrar una mejor expresión para la función de partición del sistema en cuestión, es necesario considerar las condiciones periódicas de contorno, así como las restricciones impuestas a las posiciones de las diferentes partículas (por ejemplo, por estar enlazadas unas a otras o por estar embebidas en un lípido).

Es posible también introducir correcciones que contemplen cómo las moléculas de disolvente pueden limitar los posibles valores de posiciones y velocidades (para ello, es necesario introducir correlaciones entre los grados de libertad de las partículas de la proteína y el disolvente).

10.3 Apéndice 3. Inferencias bayesianas a partir de las SDM

Para realizar inferencias bayesianas desde las Simulaciones de Dinámica Molecular (SDM), es posible obtener como variables observables el conjunto de posiciones y de velocidades, $\mathbf{r} \cup \mathbf{v} = \{\mathbf{r}_1, \dots, \mathbf{r}_n, \mathbf{v}_1, \dots, \mathbf{v}_n\}$; por otra parte, es posible denotar mediante $\boldsymbol{\theta}$ el conjunto de parámetros de los cuales depende la distribución de probabilidad de cada r_1 y v_1 ; por ejemplo, si, *a priori*, asignamos para la posición en x de la j –ésima partícula una distribución del tipo

$$p(x_j) \propto \exp \left[-\frac{(x_j - \langle x \rangle_j)^2}{2s_{x_j}^2} \right] \quad (1)$$

y para las velocidades en x para la j –ésima partícula una distribución del tipo

$$p[(v_x)_j] \propto (v_x)_j^2 \cdot \exp \left\{ -\frac{[(v_x)_j - \langle v_x \rangle_j]^2}{2s_{(v_x)_j}^2} \right\} \quad (2),$$

entonces los parámetros $(\boldsymbol{\theta}_x)_j$ para el eje x y la j –ésima partícula serían $\langle x \rangle_j$, $s_{x_j}^2$, $\langle v_x \rangle_j$ y $s_{(v_x)_j}^2$. Dicho de otro modo, $(\boldsymbol{\theta}_x)_j = \{\langle x \rangle_j, s_{x_j}^2, \langle v_x \rangle_j \text{ y } s_{(v_x)_j}^2\}$.

Además,

$$p(\mathbf{y}) = p(\mathbf{r} \cup \mathbf{v}) \propto \prod_{j=1}^N \exp \left\{ -\frac{(x_j - \langle x \rangle_j)^2}{2s_{x_j}^2} \right\} \exp \left\{ -\frac{(y_j - \langle y \rangle_j)^2}{2s_{y_j}^2} \right\} \exp \left\{ -\frac{(z_j - \langle z \rangle_j)^2}{2s_{z_j}^2} \right\} \times \\ \times \prod_{j=1}^N (v_x)_j^2 (v_y)_j^2 (v_z)_j^2 \exp \left\{ -\frac{[(v_x)_j - \langle v_x \rangle_j]^2}{2s_{(v_x)_j}^2} \right\} \exp \left\{ -\frac{[(v_y)_j - \langle v_y \rangle_j]^2}{2s_{(v_y)_j}^2} \right\} \exp \left\{ -\frac{[(v_z)_j - \langle v_z \rangle_j]^2}{2s_{(v_z)_j}^2} \right\} \quad (3).$$

Y es posible también suponer *a priori* que $\boldsymbol{\theta}$ seguirá la siguiente distribución

$$p(\boldsymbol{\theta}) \propto \prod_{j=1}^N \exp \left[-\frac{(\langle x \rangle_j - \mu_{x_j})^2}{2\sigma_{x_j}^2} \right] \exp \left[-\frac{(\langle y \rangle_j - \mu_{y_j})^2}{2\sigma_{y_j}^2} \right] \exp \left[-\frac{(\langle z \rangle_j - \mu_{z_j})^2}{2\sigma_{z_j}^2} \right] \times \\ \prod_{j=1}^N \exp \left[-\frac{(s_{x_j} - \sigma_{x_j})^2}{\sigma_{x_j}^2/k_{x_j}} \right] \exp \left[-\frac{(s_{y_j} - \sigma_{y_j})^2}{\sigma_{y_j}^2/k_{y_j}} \right] \exp \left[-\frac{(s_{z_j} - \sigma_{z_j})^2}{\sigma_{z_j}^2/k_{z_j}} \right] \times$$

$$\begin{aligned}
& \prod_{j=1}^N \exp \left\{ -\frac{[(v_x)_j - \mu(v_x)_j]^2}{2\sigma_{(v_x)_j}^2} \right\} \exp \left\{ -\frac{[(v_y)_j - \mu(v_y)_j]^2}{2\sigma_{(v_y)_j}^2} \right\} \exp \left\{ -\frac{[(v_z)_j - \mu(v_z)_j]^2}{2\sigma_{(v_z)_j}^2} \right\} \times \\
& \prod_{j=1}^N \exp \left\{ -\frac{[s(v_x)_j - \sigma(v_x)_j]^2}{\sigma_{(v_x)_j}^2 / k(v_x)_j} \right\} \exp \left\{ -\frac{[s(v_y)_j - \sigma(v_y)_j]^2}{\sigma_{(v_y)_j}^2 / k(v_y)_j} \right\} \exp \left\{ -\frac{[s(v_z)_j - \sigma(v_z)_j]^2}{\sigma_{(v_z)_j}^2 / k(v_z)_j} \right\} \quad (4)
\end{aligned}$$

Haciendo lo anterior, el teorema de Bayes, puede expresarse mediante la siguiente ecuación:

$$\begin{aligned}
p(\boldsymbol{\theta}|\mathbf{y}) & \propto p(\mathbf{y}|\boldsymbol{\theta}) \times \left\{ \prod_{j=1}^N \exp \left[-\frac{(\langle x \rangle_j - \mu_{x_j})^2}{2\sigma_{x_j}^2} \right] \exp \left[-\frac{(\langle y \rangle_j - \mu_{y_j})^2}{2\sigma_{y_j}^2} \right] \exp \left[-\frac{(\langle z \rangle_j - \mu_{z_j})^2}{2\sigma_{z_j}^2} \right] \right\} \times \\
& \times \prod_{j=1}^N \exp \left[-\frac{(s_{x_j} - \sigma_{x_j})^2}{\sigma_{x_j}^2 / k_{x_j}} \right] \exp \left[-\frac{(s_{y_j} - \sigma_{y_j})^2}{\sigma_{y_j}^2 / k_{y_j}} \right] \exp \left[-\frac{(s_{z_j} - \sigma_{z_j})^2}{\sigma_{z_j}^2 / k_{z_j}} \right] \times \\
& \times \prod_{j=1}^N \exp \left\{ -\frac{[(v_x)_j - \mu(v_x)_j]^2}{2\sigma_{(v_x)_j}^2} \right\} \exp \left\{ -\frac{[(v_y)_j - \mu(v_y)_j]^2}{2\sigma_{(v_y)_j}^2} \right\} \exp \left\{ -\frac{[(v_z)_j - \mu(v_z)_j]^2}{2\sigma_{(v_z)_j}^2} \right\} \times \\
& \times \prod_{j=1}^N \exp \left\{ -\frac{[s(v_x)_j - \sigma(v_x)_j]^2}{\sigma_{(v_x)_j}^2 / k(v_x)_j} \right\} \exp \left\{ -\frac{[s(v_y)_j - \sigma(v_y)_j]^2}{\sigma_{(v_y)_j}^2 / k(v_y)_j} \right\} \exp \left\{ -\frac{[s(v_z)_j - \sigma(v_z)_j]^2}{\sigma_{(v_z)_j}^2 / k(v_z)_j} \right\} \quad (5)
\end{aligned}$$

donde $p(\boldsymbol{\theta}|\mathbf{y})$ es la distribución de probabilidad *a posteriori* de $\boldsymbol{\theta}$, a la vez que $p(\mathbf{y}|\boldsymbol{\theta})$ es la distribución de probabilidad *a posteriori* de \mathbf{y} .

10.4 Apéndice 4. Análisis estadístico

Para un sistema en específico, el valor de ACI correspondiente a la i –ésima cadena, al j –ésimo residuo y al tiempo de simulación t se denotará mediante $x_{i,j,t}$

Para cada sistema, hay 4 cadenas y entre 193 ± 1 residuos por cadena; además, se consideran 3 valores de t . La cantidad de residuos varía según la cantidad de residuos en el sitio activo. Por ejemplo, para el sistema CNBD (apo), hay 13 residuos (ILE 503, MET 530, PHE 538, GLY 539, GLU 540, ILE 541, CYS 542, ARG 549, THR 550, ALA 551, VAL 553, ARG 590, ILE 594); para el sistema CNDB/cAMP, 15 (los anteriores, así como LEU 532 y ARG 593), y, para el sistema CNBD/cGMP, 14 (los 13 del sistema CNBD(apo) y LEU 591).

Por ende, para cada sistema fueron obtenidos $(193 \pm 1) \times 4 \times 3 = 2,316 \pm 12$ valores de $x_{i,j,t}$: 2,328 para el sistema CNBD (apo), 2,304 para el sistema CNBD/cAMP y 2,316 para el sistema CNBD/cGMP.

En principio, se determinaría cuál es el sitio alostérico encontrando el número de residuo al que le corresponde el valor más alto de $x_{i,j,t}$, pero probablemente diferencias estadísticamente significativas entre los valores de $x_{i,j,t}$ según el valor de i (la cadena en cuestión) y el valor de t (el tiempo de simulación).

Para la presente discusión, se definirán las siguientes magnitudes: μ_t (la media del valor de ACI para determinado sistema y determinado tiempo de simulación, promediando sobre todas las cadenas y todos los residuos), $\mu_{i,t}$ (la media del valor de ACI para determinado sistema, determinada cadena y determinado tiempo de simulación, promediando sobre todos los residuos) y $(\bar{x})_{j,t}$ (la media del valor de ACI para determinado sistema, determinado residuo y determinado tiempo de simulación, promediando sobre todas las cadenas). Sus ecuaciones correspondientes son las siguientes:

$$\mu_t = \frac{\sum_{i,j}(\text{ACI})_{i,j,t}}{N_{\text{estructuras}}} \quad (1),$$

$$\mu_{i,t} = \frac{\sum_j(\text{ACI})_{i,j,t}}{N_{\text{residuos}}} \quad (2)$$

y

$$(\bar{x})_{j,t} = \frac{\sum_i(\text{ACI})_{i,j,t}}{N_{\text{cadenas}}} \quad (3).$$

Para empezar, para cada sistema y para cada valor de t , se puso a prueba la hipótesis nula $H_0 : \mu_{A,t} = \mu_{B,t} = \mu_{C,t} = \mu_{D,t}$ por medio de una prueba ANOVA de un factor, donde el factor era la cadena.

Si la anterior hipótesis nula resultaba ser verdadera, entonces podría incluirse cada cadena al momento de obtener los promedios $\mu_t = \frac{\sum_{i,j}(\text{ACI})_{i,j,t}}{N_{\text{estructuras}}}$ y $(\bar{x})_{j,t} = \frac{\sum_i(\text{ACI})_{i,j,t}}{N_{\text{cadenas}}}$. De lo contrario, debían ser excluidas determinadas cadenas hasta obtener el mayor valor posible de P para la hipótesis nula.

Tras excluir ciertas cadenas con base en el anterior criterio, para cada sistema, se puso a prueba la siguiente hipótesis nula $H_0 : \mu_{t=0\text{ns}} = \mu_{t=250\text{ns}} = \mu_{t=500\text{ns}}$. Esto se realizó por medio de una prueba ANOVA de un factor, donde el factor era el tiempo de simulación. Dicha hipótesis resultó ser improbable para todos los sistemas, por lo cual se rechazó.

Más tarde, se comparó el valor de \bar{x}_j obtenido para diferentes tiempos de simulación (el correspondiente a 0, 250 y 500 nanosegundos). Por ejemplo, para determinar si existía una diferencia estadísticamente significativa entre \bar{x}_j para 0 y 250 nanosegundos, se comparó el

valor de $t_{v,\alpha}$ con el valor de la magnitud $\sqrt{4} \cdot \frac{|(\bar{x}_j)_{0\text{ns}} - (\bar{x}_j)_{250\text{ns}}|}{\sqrt{(s_j^2)_{0\text{ns}} + (s_j^2)_{250\text{ns}}}}$, donde los subíndices indican

el tiempo de simulación para cada marco, y donde

$$s_j^2 = \frac{(\bar{x}_j - x_{A,j})^2 + (\bar{x}_j - x_{B,j})^2 + (\bar{x}_j - x_{C,j})^2 + (\bar{x}_j - x_{D,j})^2}{3} \quad (4).$$

Desde luego, α , en $t_{v,\alpha}$, es el valor máximo admisible del error del tipo I, y es un parámetro escogido a conveniencia (para este paso y para cada sistema, se escogió $\frac{\alpha}{2} = 0.025$), mientras que v es la cantidad de grados de libertad, que, en este caso, se calcula por medio de la siguiente ecuación:

$$v = \text{redon} \left\{ \frac{(3)[(s_j^2)_{0\text{ns}} + (s_j^2)_{250\text{ns}}]^2}{[(s_j^2)_{0\text{ns}}]^2 + [(s_j^2)_{250\text{ns}}]^2} \right\} \quad (5).$$

Si, para el j –ésimo residuo, se cumplía

$$\sqrt{4} \cdot \frac{|(\bar{x}_j)_{0\text{ns}} - (\bar{x}_j)_{250\text{ns}}|}{\sqrt{(s_j^2)_{0\text{ns}} + (s_j^2)_{250\text{ns}}}} - t_{v,\alpha} > 0 \quad (6),$$

entonces se aceptaba la hipótesis positiva, es decir: $H_1 : (\bar{x}_j)_{0\text{ns}} \neq (\bar{x}_j)_{250\text{ns}}$. Haciendo un procedimiento de esta naturaleza, pudo determinarse, por ejemplo, que había múltiples diferencias estadísticamente significativas entre los marcos correspondientes a 0 y 250 nanosegundos de simulación para cada sistema individual, así como entre los tiempos 0 y 500 nanosegundos de simulación. En contraste, sólo se encontraron diferencias estadísticamente significativas entre los marcos correspondientes a 250 y 500 nanosegundos para pocos segmentos de los diferentes sistemas cerca de la hélice.

En concreto, se encontró que, para el sistema *apo* entre los tiempos de simulación 0 y 250 nanosegundos, había diferencias estadísticamente significativas para el 13.9% de residuos a la distancia antes mencionada de la hélice 1; al comparar los tiempos 0 y 500 nanosegundos, el porcentaje incrementaba a 16.5%, mientras que, al comparar los tiempos 250 y 500 nanosegundos, el porcentaje disminuía a 2.06%. Para el sistema CNBD/cAMP, se encontraba algo muy semejante: al comparar los tiempos 0 y 250 nanosegundos, el porcentaje era de 13.0%; al comparar los tiempos 250 y 500 nanosegundos, el porcentaje incrementaba 22.4%, mientras que, al comparar los tiempos 0 y 500 nanosegundos, el porcentaje disminuía a 8.85%. Y, para el sistema CNBD/cGMP, se repetía el mismo patrón: al comparar los tiempos 0 y 250 nanosegundos, el porcentaje era de 13.5%; al comparar los tiempos 250 y 500 nanosegundos, el porcentaje incrementaba 20.7%, mientras que, al comparar los tiempos 0 y 500 nanosegundos, el porcentaje disminuía a 6.74%. Lo anterior se resume con la tabla 6.

De lo anterior, se concluyó que es preferible, efectivamente, realizar simulaciones de dinámica molecular antes de someter las estructuras al algoritmo de propagación de perturbaciones. Adicionalmente, con base en estas observaciones, se determinó que podría ser prácticamente indistinto si se escogía el marco correspondiente a 250 nanosegundos o si se escogía el correspondiente a 500 nanosegundos para el siguiente paso, que consiste en la determinación de las rutas alostéricas y de los residuos críticos. Sólo para aprovechar que la escala de tiempo

es mayor a 500 nanosegundos para cada sistema, se escogió el marco correspondiente a 500 nanosegundos.

Más tarde, se procedió obteniendo un promedio para los valores de \bar{x}_j en el que se tomaba la mayor cantidad de marcos posible (se tenía cuidado en que sus valores no tuvieran una diferencia estadísticamente significativa).

Si, para algún residuo, eran aceptadas las hipótesis nulas

$$H_0 : (\bar{x}_j)_{0ns} = (\bar{x}_j)_{250ns} \wedge H_0 : (\bar{x}_j)_{0ns} = (\bar{x}_j)_{500ns} \wedge H_0 : (\bar{x}_j)_{250ns} = (\bar{x}_j)_{500ns} \quad (7),$$

entonces, para dicho residuo, era obtenido un nuevo promedio, \bar{X}_j calculado mediante la siguiente ecuación:

$$\bar{X}_j = \frac{1}{3} \cdot [(\bar{x}_j)_{0ns} + (\bar{x}_j)_{250ns} + (\bar{x}_j)_{500ns}] \quad (8).$$

Por otra parte, si, para algún residuo de algún sistema, eran aceptadas las siguientes hipótesis

$$[H_1 : (\bar{x}_j)_{0ns} \neq (\bar{x}_j)_{250ns} \vee H_1 : (\bar{x}_j)_{0ns} \neq (\bar{x}_j)_{500ns}] \wedge H_0 : (\bar{x}_j)_{250ns} = (\bar{x}_j)_{500ns} \quad (9),$$

entonces el nuevo promedio \bar{X}_j tenía la siguiente ecuación:

$$\bar{X}_j = \frac{1}{2} \cdot [(\bar{x}_j)_{250ns} + (\bar{x}_j)_{500ns}] \quad (10),$$

es decir: no se tomaba en cuenta el valor correspondiente a los 0 nanosegundos; se daba prioridad a las estimaciones realizadas sobre estructuras obtenidas tras realizar simulaciones de dinámica molecular.

Por otra parte, si se cumplía la hipótesis

$$H_1 : (\bar{x}_j)_{250ns} \neq (\bar{x}_j)_{500ns} \quad (11),$$

entonces, independientemente que se aceptaran o no las hipótesis nulas

$$H_0 : (\bar{x}_j)_{0ns} = (\bar{x}_j)_{250ns} \wedge H_0 : (\bar{x}_j)_{0ns} = (\bar{x}_j)_{500ns} \quad (12),$$

se tomaba en cuenta exclusivamente el valor correspondiente a $(\bar{x}_j)_{500ns}$.

Posteriormente, para determinar cuáles residuos eran, con mayor probabilidad los sitios alostéricos, se evaluó, a través de otra prueba basada en la distribución de t de Student, la hipótesis positiva

$$H_1 : \bar{X}_j > \mu \quad (13),$$

donde

$$\mu = \sum_j \bar{X}_j / N_{residuos} \quad (14).$$

El valor de α (el máximo valor del error del tipo I permitido) escogido para este paso era de 5×10^{-4} .

Finalmente, el sitio alostérico sería determinada buscando cuál residuo, entre aquellos que cumplen la anterior hipótesis nula, arroja un mayor valor para la siguiente magnitud:

$$\text{Intensidad} \times \text{Agudeza} = (\text{ACI})_j \cdot \left\{ (\text{ACI})_j - \frac{[(\text{ACI})_{j-1} + (\text{ACI})_{j+1}]}{2} \right\} \quad (15).$$

Nótese que

$$\text{Agudeza} = (\text{ACI})_j - \frac{[(\text{ACI})_{j-1} + (\text{ACI})_{j+1}]}{2} \quad (16)$$

es una medida de qué tan alto es el valor de $(\text{ACI})_j$ en comparación con los valores de su entorno inmediato. Dicho de otro modo, no sólo se toma en cuenta qué tan alto es el pico, sino también qué tanto sobresale en su entorno.

Los potenciales residuos alostéricos restantes fueron 2 en cada caso, y se escogieron de acuerdo con qué tan alto fuera su valor de Intensidad \times Agudeza.