



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN
CIENCIAS QUÍMICAS**

Simulaciones de dinámica molecular de la modulación alostérica en el
receptor para péptido similar a glucagón de tipo 1 (GLP-1R)

TESIS

PARA OPTAR POR EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

Q.F.B. José Guadalupe Rosas Jiménez

Dra. Karina Martínez Mayorga
Instituto de Química

Ciudad de México, Abril de 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN
CIENCIAS QUÍMICAS

Simulaciones de dinámica molecular de la modulación alostérica en el
receptor para péptido similar a glucagón de tipo 1 (GLP-1R)

TESIS

PARA OPTAR POR EL GRADO DE
MAESTRO EN CIENCIAS

PRESENTA

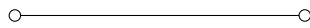
Q.F.B. José Guadalupe Rosas Jiménez



Ciudad de México, 2021

Dedicatoria

Para mis padres y mis hermanas, el fuerte pilar que ha mantenido a flote este barco durante un largo camino.



Para mis amigos, Lola, Blanquita, Mema y todos los demás, que luego de todo el tiempo y la distancia, siguen estando presentes en mis pensamientos.



Para Wendolyn, por los tiempos que siempre nos van a pertenecer.
"Pero en el fondo de tus ojos no hay alegría, sino pura tristeza; como si tus ojos supieran que no existe la dicha y que todo lo bello y amado es efímero. Tienes los más hermosos ojos que puede haber, y también los más tristes."

3 6ñ mm mē nō jñ

Índice general

1	Resumen	1
2	Introducción	2
3	Marco Teórico	4
3.1	GLP-1R como objetivo biológico en la <i>Diabetes mellitus</i>	4
3.1.1	Vías de señalización y efectos fisiológicos	5
3.1.2	Estructura y agonismo	7
3.1.3	Modulación alostérica	11
3.2	Dinámica molecular y Modelos de Estado de Markov	15
3.2.1	Cadenas de Markov Ergódicas y su Descomposición Espectral	16
3.2.2	Análisis de clúster de Perron y Teoría de Rutas de Transición	18
4	Hipótesis	20
5	Objetivos	20
5.1	Objetivo general	20
5.2	Objetivos específicos	20
6	Métodos	21
6.1	Preparación de la estructura del GLP-1R	21
6.2	Acoplamiento molecular covalente	22
6.3	Dinámica Molecular	24
6.3.1	Parametrización de los ligandos	24
6.3.2	Hidratación del canal del receptor mediante el modelo de sitios de interacción de referencia (RISM)	25
6.3.3	Empaquetado del receptor en la membrana	26
6.3.4	Equilibrio y producción	26
6.4	Modelado de Estados de Markov	28
6.4.1	Extracción de variables	28
6.4.2	Reducción de dimensiones	29
6.4.3	Generación de los microestados (discretización)	29
6.4.4	Estimación de las matrices de transición y validación de los modelos	30
6.4.5	Análisis de los modelos de estado de Markov	30

6.5	Cálculos de energía libre con <i>Umbrella Sampling</i>	31
7	Resultados	32
7.1	Preparación de la estructura y validación	32
7.2	Acoplamiento molecular covalente	34
7.3	Dinámica molecular	35
7.3.1	Parametrización de los ligandos	35
7.3.2	Hidratación del canal del receptor	38
7.4	Modelos de Estado de Markov	40
7.4.1	Extracción de variables	40
7.4.2	Reducción de dimensiones	41
7.4.3	Generación de microestados, estimación y validación de los modelos de Markov	44
7.4.4	Análisis de los modelos	48
8	Discusión	59
8.1	Comparación entre los subespacios de componentes independientes . .	59
8.2	Sistema sin modulador	61
8.3	Modulación positiva	61
8.4	Efecto de la presencia del agonista	67
8.5	Participación del modulador negativo	69
8.6	Concordancia con las relaciones estructura - actividad y datos expe- rimentales	71
9	Conclusiones	76
10	Perspectivas	77
A	Métodos en MOE	80
A.1	Asignación de protómeros y tautómeros con Protonate 3D	80
A.2	Acoplamiento molecular covalente	81
B	Método VAMP	82
B.1	Método VAMP	82

Resumen

La diabetes mellitus de tipo 2 es uno de los mayores problemas de salud pública mundial en décadas recientes. El desarrollo de fármacos capaces de ejercer un control glucémico eficaz a largo plazo, así como un efecto preventivo sobre las complicaciones crónicas, representa un área de interés activo en la investigación. Actualmente, uno de los blancos moleculares explorados con este objetivo es el receptor para el péptido similar a glucagón de tipo 1 (GLP-1R) debido a que es capaz de ejercer diversos efectos biológicos útiles en el tratamiento de la diabetes. Los fármacos en uso clínico, basados en agonistas sobre este receptor, son péptido miméticos diseñados para ser resistentes a degradación enzimática o para alargar su vida media en el plasma, aunque el uso de moléculas pequeñas es atractivo debido a que presenta diversas ventajas en cuanto a producción, costos, estabilidad y vía de administración.

En este trabajo se utilizaron simulaciones de dinámica molecular del GLP-1R, en presencia y ausencia de moduladores alostéricos conocidos, con el objetivo de determinar los cambios estructurales importantes que ejercen los moduladores para lograr regular la activación del receptor. Los resultados obtenidos se utilizaron para generar modelos de estado de Markov con el fin de caracterizar las transiciones entre los estados conformacionales explorados en la simulación. Mediante la interpretación de los modelos de estado de Markov y de cálculos de energía libre, se observó que los cambios más importantes ocurren al nivel de la red de residuos polares ubicada en el extremo intracelular del receptor. La presencia del modulador alostérico positivo favorece la ruptura de la red, facilitando el movimiento de la hélice transmembranal 6 requerido para adquirir la conformación activa. La disrupción de la red polar ocurre por la interacción favorable del modulador positivo con una serie de residuos hidrofóbicos ubicados en el extremo intracelular de las hélices 3, 5 y 6. Las conclusiones derivadas de estos resultados son útiles para identificar características estructurales necesarias para que un ligando pueda tener actividad moduladora sobre el receptor GLP-1R y, de esta forma guiar el desarrollo o selección de moléculas con el potencial de ser utilizadas como fármacos antidiabéticos.

Introducción

La diabetes mellitus de tipo 2 es una enfermedad crónica de prevalencia mundial cuya morbilidad se ha incrementado en décadas recientes. La Organización Mundial de la Salud informa que la incidencia de diabetes se incrementó de un 4.7 % en 1980 hasta un 8.5 % en 2014, con una mayor tendencia al incremento en países de bajo y mediano desarrollo. En México, la diabetes fue reconocida como la segunda causa de muerte durante 2016.

Las complicaciones relacionadas con la diabetes involucran daños en diversos órganos y tejidos, llegando a comprometer la vida del paciente. La primera estrategia en el tratamiento de la diabetes mellitus tipo 2 es la modificación de hábitos en el estilo de vida, mediante dieta y ejercicio, aunado a la farmacoterapia basada en hipoglucemiantes orales. Frecuentemente, la terapia es modificada e individualizada con el avance de la enfermedad, llegando a utilizarse combinaciones de hipoglucemiantes u otro tipo de fármacos para tratar las complicaciones o patologías que suelen desarrollarse a la par. Existe interés en el desarrollo de medicamentos con nuevos blancos biológicos que, además de ofrecer un control glucémico eficaz, sean capaces de disminuir los riesgos de los efectos crónicos de la enfermedad. En ensayos clínicos recientes, los inhibidores del co-transportador de sodio-glucosa (SGLT) y los agonistas del receptor para péptido similar al glucagón de tipo 1 (GLP-1R), han mostrado eficacia en la reducción de los principales factores de riesgo cardiovasculares asociados con esta enfermedad crónica.

El GLP-1R es un receptor acoplado a proteína G de clase B que reconoce a la hormona peptídica del mismo nombre, el péptido similar a glucagón de tipo 1 (GLP-1). Esta hormona, en conjunto con el péptido insulínico dependiente de glucosa (GIP), es responsable del efecto incretina. Se conoce como efecto incretina a un incremento en la secreción de insulina cuando se administra glucosa por vía oral, en comparación con la secreción estimulada bajo una misma dosis de glucosa por vía intravenosa. La activación del GLP-1R media efectos pleiotrópicos que no solo están involucrados en la regulación de la glucemia, sino también en el control de otros factores relacionados con la diabetes, como el control de peso, la minimización del riesgo de hipoglucemia y la supervivencia de las células β pancreáticas.

A pesar de su importancia, la estructura del GLP-1R no había podido determinarse hasta hace pocos años, al igual que otros receptores de clase B. Debido a los avances en microscopía crio-electrónica, se han logrado determinar las estructuras de múltiples GPCR de clase B, incluido el GLP-1R, en complejo con las proteínas G. La disponibilidad de estas estructuras ha revelado los cambios estructurales in-

volucrados en la activación del receptor en presencia de agonistas. También se han identificado sitios alostéricos a los que se unen moléculas pequeñas capaces de modular la activación del receptor. Los moduladores alostéricos negativos (NAM) impiden la activación del receptor o disminuyen su capacidad de respuesta en presencia de agonista. Por su parte, los moduladores alostéricos positivos (PAM) incrementan la respuesta producida por la activación del receptor como resultado de la unión con un agonista. En el contexto de la diabetes mellitus de tipo 2, es la modulación positiva del GLP-1R el efecto biológico de interés terapéutico.

La elucidación de los mecanismos moleculares involucrados en el agonismo y la modulación alostérica es de importancia para la búsqueda o el diseño de moléculas pequeñas capaces de activar o modular la activación del receptor. En este trabajo, utilizamos simulaciones de dinámica molecular del GLP-1R en presencia de moduladores alostéricos, en conjunto con modelos de estado de Markov, para identificar estados metaestables y de transición con el fin de proponer un mecanismo por el cual ocurre el fenómeno de modulación alostérica positiva sobre este receptor.

Marco Teórico

3.1. GLP-1R como objetivo biológico en la *Diabetes mellitus*

La diabetes mellitus de tipo 2 es una enfermedad crónica reconocida por la Organización Mundial de la Salud como uno de los grandes problemas de este siglo (1). La incidencia de diabetes se incrementó de un 4.7% en 1980 hasta un 8.5% en 2014, y fue identificada como una de las principales causas de muerte durante 2016 (1). La Encuesta Nacional de Salud y Nutrición de 2018 mostró que el 10.3% de la población adulta ha sido diagnosticada con diabetes, lo que implica una proporción más alta que la global y un incremento respecto a los resultados en 2016 (2). Los riesgos a la salud a largo plazo asociados con la diabetes involucran daño cardíaco, vascular, renal y neuropatías. Estas complicaciones pueden comprometer diversos órganos o tejidos, por lo que el uso de fármacos capaces de actuar tanto sobre el control glucémico como en la prevención o tratamiento de los daños producidos de forma crónica, es de interés como opción terapéutica en esta enfermedad.

Entre las estrategias utilizadas en el tratamiento farmacológico de la diabetes mellitus, la sensibilización al efecto incretina es de interés debido a los efectos fisiológicos a mediano y largo plazo que se producen. Se conoce como efecto incretina a un incremento en la secreción de insulina cuando se administra glucosa por vía oral, en comparación con la secreción estimulada bajo una misma dosis de glucosa por vía intravenosa (3, 4). Dos hormonas, producidas por células del intestino, son responsables de este incremento en la liberación de insulina: el péptido similar a glucagón de tipo 1 (*Glucagon-Like Peptide 1*, GLP-1) y el péptido insulínico dependiente de glucosa (*Glucose-dependent Insulinotropic Peptide*, GIP).

Tras la ingesta de alimentos, las células K y L del intestino delgado son estimuladas para liberar las hormonas GIP y GLP-1, respectivamente. Estas hormonas viajan hasta las células β del páncreas, donde se unen a sus receptores homónimos y provocan un incremento en la liberación de insulina dependiente de glucosa. Existe evidencia que muestra que los niveles de GIP son mayores que los de GLP-1 y, en individuos sanos, el efecto incretina se debe primordialmente a GIP (4). En pacientes con diabetes mellitus tipo 2, el efecto incretina se ve reducido, a pesar de que los niveles de ambas hormonas no son significativamente distintos a los de individuos sanos. Sin embargo, en pacientes con diabetes mellitus tipo 2 se ha observado una desensibilización del receptor de GIP, pero la respuesta a GLP-1 se mantiene (4, 5).

En pacientes con diabetes mellitus tipo 2 (con efecto incretina disminuido, debido a la pérdida de acción de GIP), la estimulación farmacológica del receptor para GLP-1 mediante agonistas, es suficiente para lograr un control glucémico eficaz, además de los efectos fisiológicos favorables que produce a mediano y largo plazo, debido a las vías de señalización que desencadena (4, 5). Por lo tanto, la activación del receptor para GLP-1, nombrado de forma homónima como GLP-1R, ha resultado atractivo en el tratamiento farmacológico de la diabetes mellitus. Actualmente existen 6 fármacos en uso clínico que actúan directamente sobre la activación del receptor, aunque todos ellos son peptidomiméticos de administración generalmente parenteral, derivados del GLP-1 resistentes a degradación enzimática. Por otra parte, la experiencia clínica con estos fármacos ha comprobado la eficacia en el control de los efectos esperados como la normoglucemia, la cardio y neuroprotección, entre otros (5). Asimismo, la incidencia de efectos secundarios se ha caracterizado a mediano y largo plazo (5), colocando al uso de incretina-miméticos como una opción válida en el tratamiento de la diabetes.

3.1.1. Vías de señalización y efectos fisiológicos

El mecanismo de transducción de señales tras la activación de GLP-1R en células pancreáticas ha sido ampliamente estudiado y explica los diferentes efectos asociados con su agonismo. La figura 3.1 resume de forma esquemática las vías de señalización involucradas en la liberación y biosíntesis de insulina. De manera independiente a la activación del GLP-1R, un incremento en los niveles de glucosa plasmática es suficiente para originar la liberación de insulina. La glucosa es introducida en las células β pancreáticas a través del receptor GLUT2, con una cinética lo suficientemente rápida para inducir un incremento en la producción de ATP mediante estimulación de la glucólisis y la fosforilación oxidativa (3, 6). El incremento en los niveles de ATP es detectado por el canal de potasio sensible a ATP (K-ATP), inhibiendo el transporte de K^+ al medio extracelular, lo que origina una despolarización en la membrana. Este cambio en la diferencia de potencial local ocasiona la apertura de los canales de calcio activados por voltaje de tipo L, con un subsecuente incremento en la concentración de Ca^{2+} intracelular. El aumento en los niveles de Ca^{2+} promueve la exocitosis de vesículas con insulina (3, 6, 7).

El GLP-1R recluta de manera preferente a proteínas de tipo $G_{\alpha s}$ que se encargan de activar la enzima adenilil ciclasa para producir un incremento en los niveles de cAMP intracelulares (3, 7). El cAMP es capaz de promover la activación de diferentes efectores intracelulares que se encargan de regular diversas funciones. En particular, el cAMP libera la subunidad catalítica de la proteína cinasa A (PKA) y activa la proteína de intercambio activada por cAMP (Epac). La PKA se encarga de fosforilar la subunidad receptora de sulfonil urea (SUR) del canal K-ATP, cerrándolo y evitando el exporte de potasio, favoreciendo la despolarización de la membrana. Epac, por otra parte, promueve el transporte intracelular de Ca^{2+} desde el retículo endoplásmico hasta el citoplasma. El incremento de Ca^{2+} citoplásmico no solo favorece la exocitosis de insulina, sino que activa diversas proteínas dependientes de calcio, como la calcineurina. Esta proteína defosforila al factor nuclear de células T activadas (NFAT) que actúa como factor de transcripción promoviendo la expresión

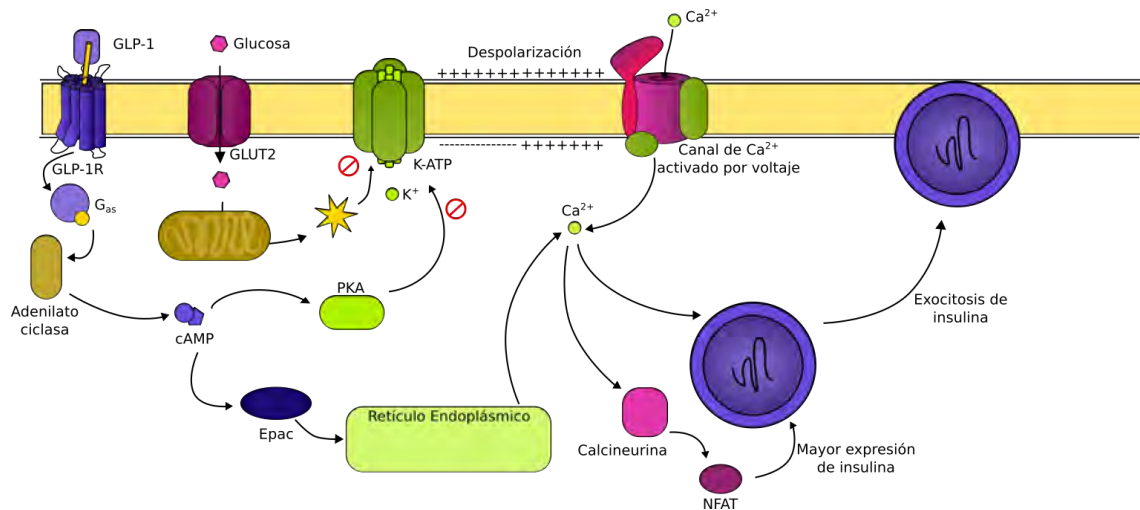


Figura 3.1: Diagrama de las vías de señalización activadas por GLP-1R involucradas en la estimulación de la exocitosis y producción de insulina en las células β del páncreas.

del gen de la insulina (3).

La activación de PKA y Epac es un punto de ramificación importante para la activación de otras vías que originan diversos efectos. La ruta de la PKA activa principalmente 2 ramificaciones: la ruta de las proteína-quinasas asociadas a mitógenos (MAPK/ERK) y la ruta de la fosfoinosítidos-3-quina (PI3K/AKT) (8). La activación de la ruta MAPK/ERK ocasiona un incremento en la ciclina D1, un factor fundamental en la progresión del ciclo celular que conduce a la proliferación de las células β (3). Por su parte, los diversos efectores activados por AKT se encargan de regular procesos relacionados con el envejecimiento, inhibiendo factores pro-apoptóticos o la liberación de citocinas proinflamatorias (8). Esto produce una mayor supervivencia de las células β con una disminución en el estrés oxidativo y de procesos inflamatorios. En pacientes con diabetes mellitus de tipo 2, el tejido pancreático sufre una atrofia de forma crónica que origina la necesidad de administrar insulina exógena. La promoción de la supervivencia de las células β es, por tanto, un efecto deseable en una estrategia farmacológica a largo plazo.

Además del acoplamiento a las proteínas G_{as} , el GLP-1R también activa vías dependientes de β -arrestina-1. En esta ruta alterna, una de las proteínas efectoras que desencadena las respuestas celulares es la proteína quina pERK1/2. Como resultado de la activación de pERK1/2, la proteína Bad es fosforilada, inactivándola y originando un efecto antiapoptótico (3). La proteína Bad pertenece a la familia BCL-2 (*B Cell Linfofoma 2*), encargadas del control de la apoptosis (9, 10). Bad inhibe a otras proteínas de la familia BCL-2 de efecto anti-apoptótico, promoviendo la permeabilización de la membrana externa mitocondrial como un punto de inicio en la apoptosis. Su fosforilación por parte de pERK1/2 y de PKA inhibe su función, promoviendo la supervivencia de las células β (9, 11). Además de su participación en la apoptosis, experimentos con ratones *knock-out* al gen que codifica la proteína Bad muestran que las células β pancreáticas tienen una disminución importante en la capacidad de secretar insulina, en comparación con células obtenidas de ratones

sin delección (10, 12). También, la efectividad en la liberación de insulina vuelve a incrementarse al introducir de forma exógena el gen codificante de Bad en células derivadas de los ratones *knock-out* (12). De esta manera, la fosforilación de Bad es otro factor sobre el que el GLP-1R promueve la secreción de insulina. La activación del GLP-1R también se ha asociado a un incremento en la proteólisis del precursor transmembranal de la β -celulina. El péptido resultante actúa como agonista del receptor para el factor de crecimiento epidérmico, ocasionando su transactivación y promoviendo la proliferación celular.

Se conoce que la expresión del GLP-1R ocurre también en tejido extrapancreático, pero los mecanismos involucrados en su activación a nivel de otros órganos están menos caracterizados. En el hígado, el GLP-1R puede estar involucrado en la inhibición de la gluconeogénesis, la lipogénesis y el incremento en la producción de glucógeno (3). Asimismo, se ha observado que el tratamiento con el agonista exendina-4 incrementa los receptores ativados por proliferadores peroxisómicos (PPAR) (3). En estudios clínicos, el uso de terapia con incretina-miméticos produce una mejora en el perfil lipídico de los pacientes, con una disminución significativa en los niveles de colesterol LDL y colesterol total (5). En el sistema nervioso central, este receptor participa en el proceso de saciedad, originando una reducción sobre la ingesta de alimentos y contribuye en el control del peso corporal (3). Los estudios clínicos con el uso de diversos agonistas de GLP-1R como la exenatida y la liraglutida, en pacientes con diabetes y obesidad, muestran una reducción neta en el peso corporal que puede incrementarse en terapias combinadas con otros agentes antihiperglucémicos (5). Otro efecto relevante del receptor de GLP-1 y que ha llamado la atención en años recientes, es que su activación en el sistema nervioso central se ha relacionado con neuroprotección y efectos antiinflamatorios, que pueden ser útiles en la enfermedad de Parkinson (8).

3.1.2. Estructura y agonismo

El GLP-1R es un receptor acoplado a proteína G (*G-Protein Coupled Receptor*, GPCR) de clase B1 o similar a secretina, de la familia del receptor de glucagón. Hasta la fecha, sólo se conocen 15 receptores de clase B en seres humanos. Todos los receptores de esta clase reconocen hormonas peptídicas y se encargan de mediar respuestas fisiológicas relacionadas con la homeostasis de Ca^{2+} , la regulación del apetito, la homeostasis de glucosa, el metabolismo de los aminoácidos, el tono cardiovascular, el desarrollo cardiovascular y gastrointestinal, el metabolismo en el tejido óseo y la respuesta inmune (13, 14). Por sus funciones biológicas, estos receptores son importantes blancos moleculares para el desarrollo de fármacos contra diversas enfermedades.

Estructuralmente, los receptores de la clase B poseen, al igual que todos los GPCR, siete hélices transmembranales ordenadas en forma de cáliz, con una cavidad ortostérica prominente en la cara extracelular. Una característica particular de los miembros de la clase B es la presencia de un dominio extracelular (*Extracellular Domain*, ECD) de entre 120-160 residuos de aminoácido, directamente unido al dominio transmembranal (*Transmembranal Domain*, TMD). A pesar de poseer un

bajo porcentaje de identidad en sus secuencias, la estructura del ECD es muy similar entre los distintos receptores pertenecientes a esta clase, que consiste en tres capas con un plegamiento $\alpha-\beta-\beta-\alpha$, estabilizadas mediante tres puentes disulfuro (15). La participación del ECD en la activación del receptor no se conoce con certeza, pero se ha propuesto que es responsable de reconocer de manera específica al péptido agonista en un paso previo a la activación del receptor.

Se ha propuesto el *modelo de unión en dos dominios* para explicar la activación del receptor por parte del agonista. En un primer paso, el extremo C-terminal de la hormona peptídica interacciona con el ECD de su receptor. Posteriormente, el extremo N-terminal del péptido se une a la cavidad ortostérica, donde ejerce diversas interacciones para finalmente activar al receptor (3, 14, 15). La figura 3.2 resume de manera esquemática este proceso.

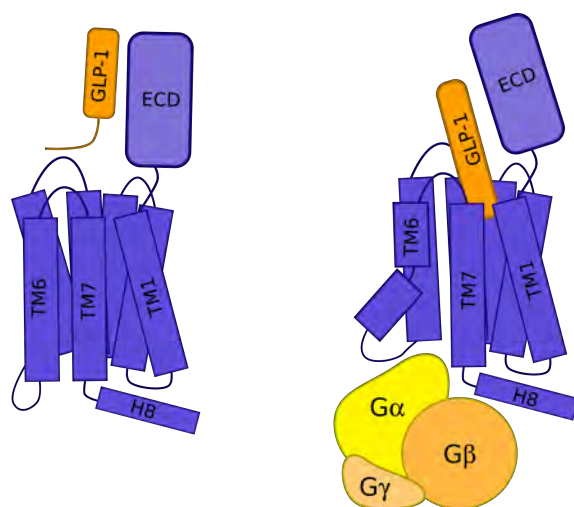


Figura 3.2: Modelo de unión *en dos dominios*: el extremo C-terminal del péptido primero es reconocido por el ECD, luego el extremo N-terminal del péptido interactúa con la cavidad ortostérica para activar al receptor.

La determinación estructural de los GPCRs de la familia B ha sido elusiva. Aunque se han obtenido estructuras cristalográficas de los dominios extracelulares en complejo con péptidos agonistas, los receptores completos se han reportado tan solo desde hace poco tiempo, debido a los avances en la microscopía crio-electrónica (13, 16, 17). La estructura del GLP-1R ha sido reportada en conformación inactiva (sólo el dominio transmembranal) (18), en conformación parcialmente activa, unido a un péptido truncado que no interacciona con el ECD (19), y en forma totalmente activa, en presencia de los agonistas peptídicos GLP-1 y exentidina-P5, en complejo con la proteína G (16, 17). Gracias a la comparación entre estos estados funcionales, así como con otros miembros de la misma clase, se han logrado determinar características estructurales clave en el proceso de activación.

Los dominios transmembranales de distintos miembros de GPCRs de la clase B en el estado inactivo son estructuralmente similares, teniendo valores de RMSD bajos, de alrededor de 1.1 Å. En contraparte, la comparación de las estructuras en los estados activos resulta en valores de RMSD de alrededor de 3.0 Å. Esto indica que se generan cambios conformacionales significativos y diferenciables (15). Se han

reconocido tres zonas formadoras de redes de interacción que se modifican de gran manera luego de la activación del receptor (15):

- **Red polar central.** El residuo S^{1.50b} forma un enlace de hidrógeno con S^{7.47b}, mientras los residuos R^{2.60b}, N^{3.43b} y Q^{7.49b} forman una red que debe romperse para originar la conformación activa.
- **Interruptor de transmisión.** Está formado por el conjunto de residuos hidrofóbicos altamente conservados Y/F^{3.44b}, M/L^{3.47b}, L^{6.45b} y L/F^{6.49b}. Esta red tiende a mantener al receptor en conformación inactiva.
- **Redes intracelulares.** Incluye el cierre iónico entre H^{2.50b} y E^{3.50b}, así como la red de enlaces de hidrógeno entre T^{6.42b} y Y^{7.57b}. En GLP-1R, la red formada entre R^{2.46b}, R^{6.37b}, N^{8.47b} y E^{8.49b} forman una red de interacción que se rompe en el estado activo.

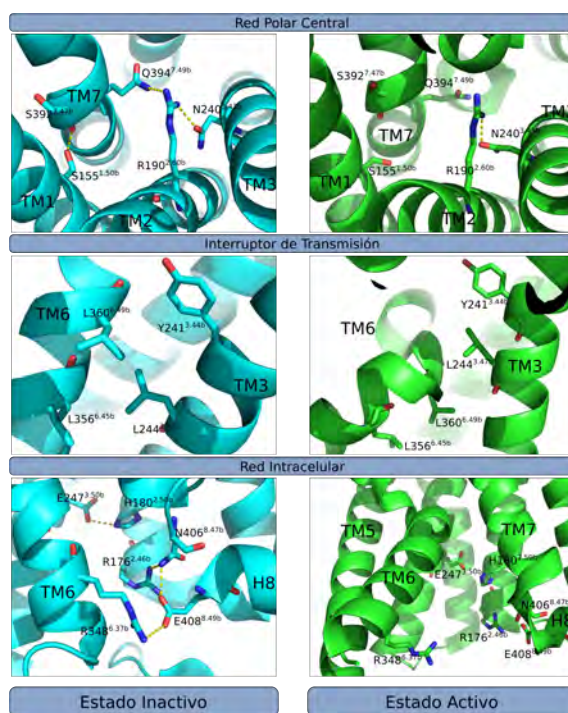


Figura 3.3: Cambios significativos en el GLP-1R, desde una conformación inactiva hacia una conformación activa. El interruptor de transmisión y la red intracelular son las regiones con cambios más pronunciados. Las interacciones entre las cadenas laterales de R176^{2.46b}, E408^{8.49b} y N406^{8.47b} son de suma importancia, ya que se conservan en ambos estados. Por su parte, el residuo R348^{6.37b} debe abandonar la red polar en un primer paso en el mecanismo de activación, por lo que la ruptura de su participación en el puente salino es una de las barreras energéticas que se oponen a la activación.

En la descripción anterior se ha utilizado la numeración de aminoácidos de acuerdo con la extensión de la nomenclatura de Ballesteros-Weinstein a las otras clases de GPCRs, en la que se agrega la letra minúscula de la familia a la que se hace referencia (3). En la figura 3.3 se muestran los cambios estructurales más significativos en las

zonas descritas. Es notable la ruptura de interacciones específicas en estos puntos. En el interruptor de transmisión, las interacciones entre los residuos de aminoácido son principalmente hidrofóbicas. Las cadenas se desplazan debido al movimiento de la hélice TM6, con una formación de un pivote a una altura alrededor del centro de la hélice. Los cambios más pronunciados ocurren en la red intracelular, donde el amplio movimiento de la hélice TM6 termina por romper el puente salino entre los residuos R348^{6.37b} y E408^{8.49b}, que se alejan hasta una distancia de alrededor de 10 Å.

El cambio de mayor amplitud es el movimiento de la hélice TM6. La hélice se dobla a partir de un pivote que se forma aproximadamente a la mitad del TMD, en el sitio conservado P^{6.47b}xxG^{6.50b}. La prolina y la glicina son aminoácidos con poca capacidad para formar hélices estables, por lo que su presencia en este punto induce el estrés necesario para permitir la deformación de la hélice TM6 (14, 15). Este movimiento permite la apertura de una cavidad en la cara intracelular del receptor donde interacciona la proteína G.

En la cara extracelular del receptor, los cambios más notables ocurren a nivel de las hélices TM1 y TM7, así como en el asa ECL1. La figura 3.4 muestra una superposición de las estructuras del dominio transmembranal del GLP-1R en estado activo e inactivo. Las hélices TM1, TM2 y TM7 se mueven hacia el centro de la cavidad, de manera que entran en contacto con el péptido (14-16). De manera interesante, en el receptor inactivo, la ECL1 se muestra desestructurada, mientras que la unión a GLP-1 favorece la extensión de las hélices TM2 y TM3 mediante interacción directa con el péptido a través de residuos hidrofóbicos con la Leu20 del GLP-1 (15, 19). Estudios de mutagénesis de residuos hidrofóbicos y aromáticos en esta región confirman una pérdida significativa en la actividad del receptor (3).

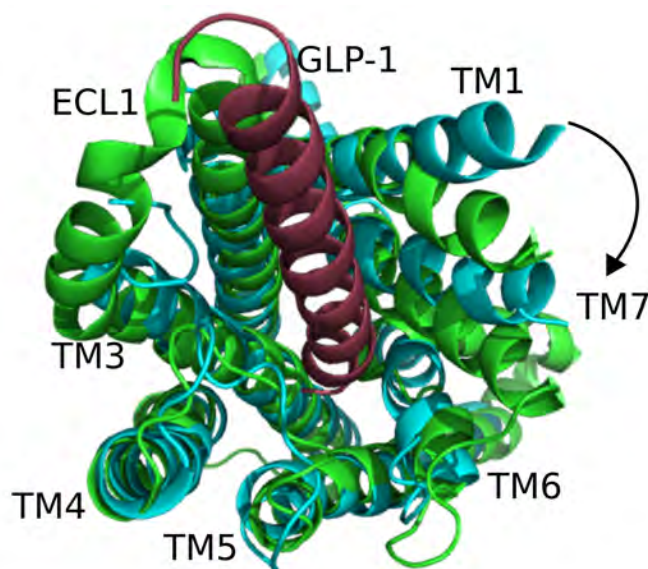


Figura 3.4: Comparación entre las estructuras del GLP-1R en estado inactivo (cian) y activo (verde). El GLP-1 se muestra en color rojo oscuro.

La interfaz entre la unión del GLP-1R y la proteína G se ha logrado determinar mediante microscopía crio-electrónica, en conjunto con otros miembros de la clase

B. La figura 3.5 muestra los principales contactos entre el GLP-1R y la proteína G_s . La unión entre estas proteínas es mediada principalmente por interacciones polares. Al igual que en otros miembros de la clase B, la hélice H5 de la subunidad α de la proteína G se empaqueta en estrecho contacto con la cavidad formada luego del movimiento de la hélice TM6 (13, 15). El residuo S352^{6.41b} del GLP-1R establece contactos polares con la hélice H5 y está involucrado en un sitio de reconocimiento importante para la unión de moduladores alostéricos negativos (NAM) (15, 18). Una diferencia importante con los GPCR de clase A es la orientación de la hélice 8, que se encuentra totalmente en el sitio intracelular sin penetrar en la membrana. En los receptores de clase A, la hélice 8 es casi paralela a la membrana, mientras que en las estructuras conocidas de los GPCR clase B, esta hélice tiene un ligero ángulo de inclinación (13-15). Las estructuras de los receptores en complejo con las proteínas G, incluyendo el GLP-1R, muestran que la inclinación de la hélice 8 permite establecer contactos con la subunidad β y por lo tanto, podría estar involucrada en el reclutamiento de la proteína G (15, 16).

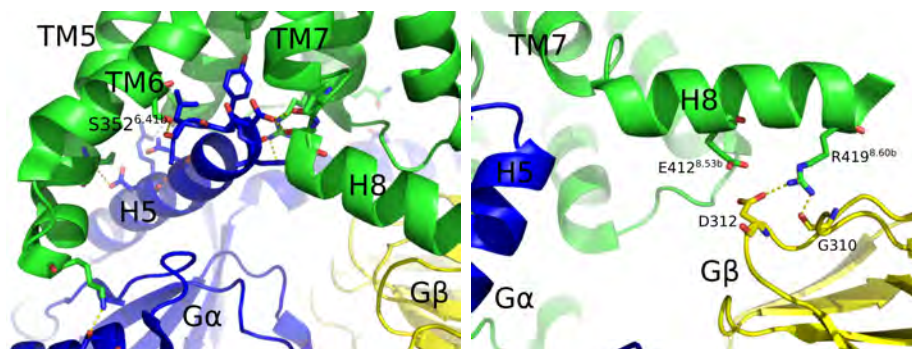


Figura 3.5: Principales puntos de interacción en la interfaz entre el GLP-1R y la proteína G. En azul se muestra la subunidad α de la proteína G, en amarillo la subunidad β , y en verde el GLP-1R

3.1.3. Modulación alostérica

La modulación de GPCRs de clase B es atractiva debido a los múltiples efectos biológicos que son capaces de mediar. No obstante, la identificación de moléculas pequeñas con efecto agonista a través de su unión al sitio ortostérico no ha sido fructífera. La principal razón radica en el gran número de interacciones que se requieren en el sitio ortostérico para lograr un efecto comparable al ejercido por moléculas de naturaleza peptídica (15, 20). Una alternativa es la búsqueda de moléculas pequeñas capaces de modular la actividad del receptor mediante su unión a cavidades alostéricas. El desarrollo de moduladores alostéricos pequeños es un área de gran interés farmacéutico debido a que ofrece diversas ventajas. En receptores cuyo sitio ortostérico es muy similar, los sitios alostéricos ofrecen puntos de mayor variación o divergencia, permitiendo una mayor selectividad (21). En este sentido, una de las principales diferencias entre receptores de la misma clase radica precisamente en los mecanismos de modulación alostérica (14, 15).

En los GPCR de la familia del receptor de glucagón, Song *et al* (18) lograron identificar y describir, mediante difracción de rayos X, los sitios de unión para mo-

duladores alostéricos negativos (NAM). Se encontró que los NAM se unen cerca de la cara intracelular del receptor, anclándose entre dos puntos de las interfaces en las hélices TM5 y TM6 y las TM6 y TM7, como se muestra en la figura 3.6. Aunque no se ha descrito el mecanismo de modulación, se acepta que estos moduladores favorecen la permanencia del receptor en el estado inactivo por el bloqueo del movimiento de la hélice TM6 (14, 18). No obstante, es interesante notar que los NAM establecen interacciones polares con la red intracelular (figura 3.3), que requiere romperse para lograr la activación. Más aun, estos ligandos forman un enlace de hidrógeno con el residuo S352^{6.41b}, que está involucrado en la interfaz con la proteína G del receptor activo. De esta manera, además de restringir el movimiento de la TM6, el NAM podría estabilizar la red de interacciones intracelulares, haciendo que la transición al estado activo sea menos favorable.

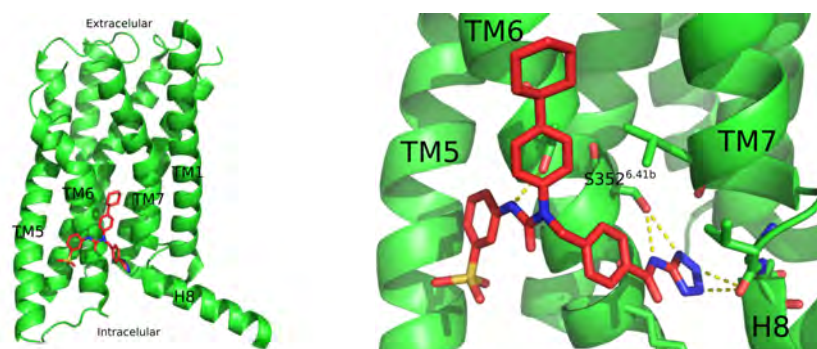


Figura 3.6: Modulador alostérico negativo NNC0640 en complejo con el GLP-1R. Se resaltan las interacciones del modulador con el receptor, que coinciden con la red intracelular y con residuos involucrados en la interacción con la proteína G.

Aunque en el contexto de la diabetes mellitus tipo 2 la activación del receptor es de mayor importancia terapéutica, el descubrimiento del sitio de modulación negativa es importante porque es cercano a un punto de modulación positiva. En 2007, luego de un cribado de alto rendimiento conducido por Teng *et al* (22), una familia de 6,7-dicloro-2-sulfonilquinoxalinas fue descubierta capaz de potenciar la capacidad del GLP-1 para activar al receptor. El compuesto líder resultante de estos experimentos, la N-tert-butil-6,7-dicloro-3-metanosulfonilquinoxalin-2-amina, posteriormente conocida como compuesto 2 en la literatura, es considerado como referencia en estudios de modulación alostérica sobre el GLP-1R. En ausencia de GLP-1, la serie de quinoxalinas es capaz de ejercer una respuesta débil sobre el receptor. Por esta actividad agonista débil, los autores clasifican a los ligandos encontrados como moduladores *ago*-alostéricos. Los estudios de estructura-actividad mostraron que la presencia de un grupo sulfónico es necesaria para mantener la capacidad moduladora. De forma independiente, el compuesto 4-(3-(benziloxi)fenil)-2-etilsulfinil-6-(trifluorometil)pirimidina (BETP) fue descubierto con capacidades similares para modular el GLP-1R, mediante una amplificación de la respuesta en la producción de cAMP en presencia de agonista. Las estructuras de estos dos compuestos modelo se muestran en la figura 3.7.

Una característica en común entre los moduladores es la presencia de grupos electrofílicos, como la metilsulfona. Los estudios farmacocinéticos del compuesto 2 y el BETP mostraron muy baja estabilidad en fluidos biológicos debido a que el grupo

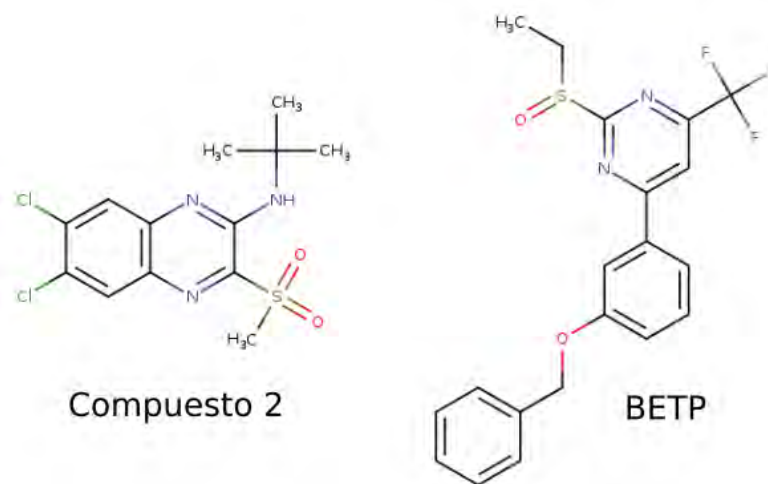


Figura 3.7: Estructura de dos moduladores alostéricos positivos modelo del GLP-1R, el compuesto 2 y el BETP

sulfónico o sulfinilo reaccionan fácilmente con residuos de cisteína libres en proteínas, o sufren una rápida conjugación con glutatión (20, 23). Este comportamiento llevó a la formulación de la hipótesis de que estos moduladores podrían modificar de manera covalente al receptor. Posteriormente, utilizando experimentos de mutagénesis en combinación con análisis por *western blot*, espectrometría de masas y por marcaje con ³H, se demostró que los ligandos electrofílicos se unen de manera covalente a los residuos de C347^{6.36b}, en el extremo intracelular de la TM6, y de C438 en el extremo C-terminal del receptor (20, 23). Sin embargo, los estudios de mutagénesis revelan que sólo el residuo C347^{6.36b} es esencial para mantener la capacidad moduladora de los PAM, puesto que la sustitución C438A no afecta su actividad biológica (23). De manera notable, el residuo C347^{6.36b} es una característica única en el GLP-1R, pues entre otros miembros de la clase B, este residuo no está conservado. La figura 3.8 muestra un alineamiento de la hélice TM6 de diversos miembros de la clase B, demostrando que la posición 6.36b es un punto de divergencia importante.

El residuo de C347^{6.36b} forma parte de una cavidad ubicada en la interfaz de las hélices TM5 y TM6. Este sitio de unión se traslapa con el punto de interacción del anillo m-metanosulfonil-fenilo del NAM NNC0640, mostrado en la figura 3.6. A pesar del avance en la determinación de las estructuras del GLP-1R, no se han reportado en presencia de moduladores positivos que actúen sobre esta cavidad. Se ha propuesto que, de manera análoga a la acción de los NAM, los PAM son capaces de favorecer la transición hacia la conformación activa al facilitar el movimiento de la hélice TM6 (15, 18). La figura 3.9 resume de manera esquemática los modelos de modulación negativa y positiva propuestos (18).

En un estudio reciente (25), se utilizó cribado virtual basado en acoplamiento molecular en la búsqueda de PAMs nuevos. Utilizando la herramienta SiteFinder en el programa Molecular Operating Environment (MOE), identifican una cavidad que coincide con el sitio de interacción del anillo de tetrazol del NAM NNC0640 (figura 3.6). Utilizando esta cavidad como objetivo, los autores lograron descubrir una molécula pequeña que se comporta de manera similar a los ago-PAMs, es decir,

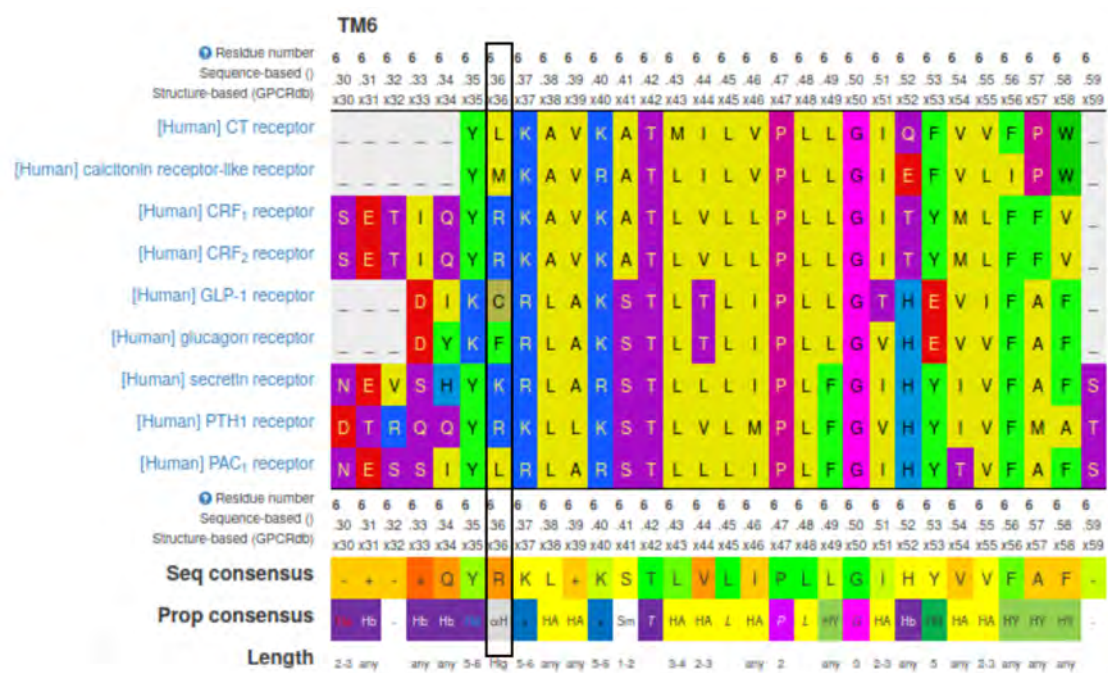


Figura 3.8: Alineamiento de secuencias de la hélice TM6 entre distintos GPCR de clase B. El residuo C347^{6.36b} es único en el GLP-1R. El alineamiento fue generado en la herramienta GPCRdb (24)

puede activar débilmente al receptor y potenciar su respuesta en presencia de GLP-1 (25). Es notable que la conformación calculada con el acoplamiento molecular predice la formación de una interacción con el residuo S352^{6.41b}. Este aminoácido está involucrado tanto en la unión de los NAM en las estructuras cristalizadas como en la interacción con la proteína G (figuras 3.5 y 3.6). Aunque el sitio de unión de esta molécula no es exactamente el mismo que el punto de anclaje del compuesto 2 y el BETP, el traslape entre los sitios es notable. Estos resultados confirman que la región intracelular de la hélice TM6 está involucrada en el mecanismo de modulación alostérica ejercida por estos compuestos.

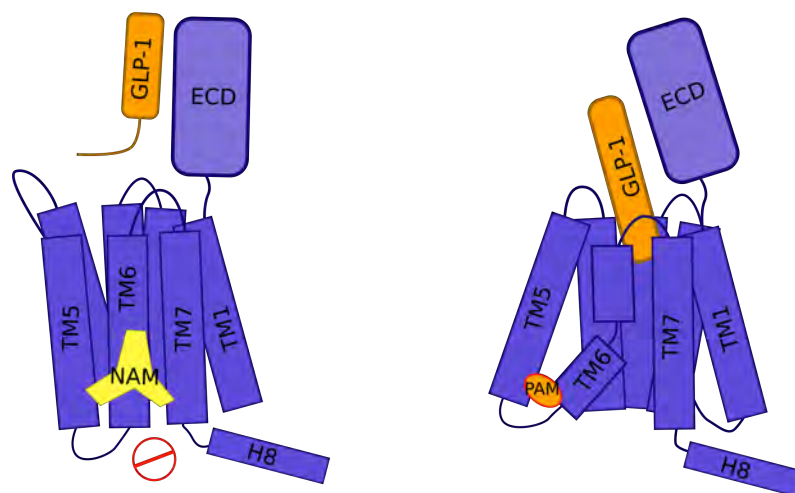


Figura 3.9: Modelos de modulación alostérica negativa y positiva propuestos sobre el GLP-1R. Diagrama adaptado de (18)

A pesar de su relevancia, no se han reportado estructuras cristalográficas o por microscopía crio-electrónica del GLP-1R en presencia de moduladores alostéricos positivos que actúen sobre la región intracelular de la TM6. La elucidación del mecanismo por el que las moléculas conocidas ejercen su efecto modulador es determinante para identificar el tipo de cambios estructurales que inducen sobre el receptor. Las interacciones entre el ligando y la proteína pueden revelar la presencia de elementos farmacofóricos importantes para guiar en el diseño o selección de moléculas pequeñas con capacidad para potenciar la activación del GLP-1R. De esta manera, el efecto incretina podrá ser utilizado como una estrategia en la terapia contra la diabetes mellitus de tipo 2, utilizando moléculas *no peptídicas con biodisponibilidad oral*.

3.2. Dinámica molecular y Modelos de Estado de Markov

En el área de la bioquímica estructural, las simulaciones de dinámica molecular se han convertido en una valiosa herramienta para caracterizar diversos aspectos biofísicos en proteínas, tales como la estabilidad estructural, los cambios conformacionales o los fenómenos de interacción con ligandos (26, 27). Frecuentemente, los métodos experimentales relacionados con la estructura de proteínas son complementados con análisis basados en dinámica molecular, pues proveen de una base teórica para la interpretación de sus resultados. En el caso de los GPCRs, el uso de metodologías basadas en dinámica molecular han permitido encontrar relaciones entre las conformaciones de los receptores y su capacidad para propiciar respuestas biológicas como el reclutamiento de diversos efectores intracelulares. También ha sido posible identificar residuos de aminoácido que participan de forma diferencial en interacciones esenciales para el agonismo, antagonismo, agonismo sesgado y modulación alostérica. En este sentido, ante la ausencia de una estructura experimental del GLP-1R en complejo con moduladores alostéricos positivos, el uso de simulaciones por dinámica molecular es una estrategia atractiva para identificar los cambios inducidos por el ligando sobre la estructura conocida del receptor. Además, se tienen datos experimentales, como las relaciones de estructura y actividad en ligandos conocidos, sobre los cuales formar una interpretación del fenómeno de reconocimiento molecular que pueda coincidir con las observaciones derivadas de las simulaciones.

No obstante, un cálculo de dinámica molecular suele originar una gran cantidad de datos que requieren ser integrados e interpretados para lograr obtener conclusiones útiles. En años recientes, los Modelos de Estado de Markov (*Markov State Model*, MSM) han ganado popularidad por su capacidad de identificar transiciones importantes en diversos fenómenos relacionados con la estructura y función de proteínas y otras macromoléculas. En un MSM, la trayectoria obtenida a partir de dinámica molecular se divide en estados conformacionales discretos, transformándose en una serie de "saltos" sobre el espacio conformacional discreto, asumiendo que el sistema cumple con la propiedad de Markov. Como resultado de esta simplificación, es posible identificar las estructuras donde el sistema pasa una mayor cantidad de tiempo, así como las preferencias de transición entre los estados conformacionales

y el tiempo esperado para que las transiciones ocurran. De esta manera, el uso de esta metodología permite caracterizar los aspectos termodinámicos y cinéticos que dominan el comportamiento del sistema.

3.2.1. Cadenas de Markov Ergódicas y su Descomposición Espectral

El principal objetivo del método de análisis conocido como Modelos de Estado de Markov, en el contexto de las simulaciones de dinámica molecular, es transformar el espacio de configuración, \mathbf{x} , donde se desarrolla la dinámica, en un conjunto de n estados discretos, \mathbf{S} , tal que la evolución sobre este conjunto de estados posee la propiedad de Markov, ilustrada en la ecuación 3.1 (28). Es decir, la probabilidad de que el sistema evolucione a un nuevo estado depende únicamente del paso inmediato anterior (29). Esto implica que al conocer las probabilidades de transición entre cada uno de los estados se puede caracterizar el proceso completo.

$$\mathbb{P}(X_t = S_t | X_{t-1} = S_{t-1}, X_{t-2} = S_{t-2}, \dots, X_0 = S_0) = \mathbb{P}(X_t = S_t | X_{t-1} = S_{t-1}) \quad (3.1)$$

Si p_{ij} es la probabilidad de que el sistema evolucione al estado S_j , partiendo del estado S_i , se induce que el conjunto de valores p_{ij} forman una matriz \mathbf{P} , conocida como matriz de transición (29). La matriz de transición es el objeto matemático fundamental en el análisis de Markov, pues contiene toda la información dinámica del sistema. Si \mathbf{p}_t es un vector en el que cada uno de sus componentes representa la probabilidad de que el sistema se encuentre en cada uno de los estados en \mathbf{S} , para un tiempo t , el producto matricial de la ecuación 3.2 representa la misma probabilidad m pasos adelante (29, 30). Es decir, la potencia m de la matriz de transición representa la propagación en el tiempo de la probabilidad de encontrar el sistema en cada uno de sus estados posibles.

$$\mathbf{p}_t^T \mathbf{P}^m = \mathbf{p}_{t+m}^T \quad (3.2)$$

El espacio de fases sobre el que evoluciona un sistema termodinámico es continuo y, dada una cantidad suficiente de tiempo, todos los estados son accesibles. Las cadenas de Markov que mejor representan este comportamiento son de la clase *irreducible* y *ergódica* (30). Una cadena de Markov es irreducible si cada uno de sus estados está conectado a por lo menos otro y es ergódica si cada uno de los estados es accesible desde cualquier otro en un intervalo finito de tiempo (29). Es decir, la matriz \mathbf{P}^k es estrictamente positiva para algún valor de k (29). Para una matriz de transición que representa una cadena de Markov ergódica, el límite dado por la ecuación 3.3 existe (29, 31). Es decir, la potencia m de la matriz de transición tiende a un límite cuando el número de pasos es muy largo. En particular, \mathbf{W} es una matriz cuyas filas son todas iguales. Esto implica que, para una cadena ergódica,

la probabilidad de visitar cada estado, luego un tiempo largo, se vuelve independiente del estado inicial. En otras palabras, si el proceso evoluciona en un intervalo de tiempo infinito, cada uno de los estados será visitado de manera recurrente e independientemente del estado en que comenzó.

$$\lim_{m \rightarrow \infty} \mathbf{P}^m = \mathbf{W} \quad (3.3)$$

Para una matriz de transición que representa una cadena ergódica, existe un único vector $\boldsymbol{\pi}$ que cumple la ecuación 3.4 (29, 30). De manera equivalente, $\boldsymbol{\pi}$ es un vector propio a la izquierda de \mathbf{P} asociado con un valor propio de 1. Puede demostrarse que $\boldsymbol{\pi}$ es un vector de probabilidad y es igual a las filas de \mathbf{W} , la matriz a la que \mathbf{P} converge (29, 31). La ecuación 3.4 implica una distribución que permanece constante en el tiempo. Dicha distribución se conoce como distribución estacionaria o de equilibrio. Los componentes de $\boldsymbol{\pi}$ son las probabilidades de encontrar al sistema en cada uno de los estados en el equilibrio. De manera equivalente, el componente π_i representa la fracción de veces que el sistema visitará al estado S_i en un tiempo de evolución infinitamente largo. Es decir, partiendo de cualquier estado, el sistema alcanzará eventualmente y, de manera asintótica, la distribución de equilibrio.

$$\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T \quad (3.4)$$

Como se mencionó anteriormente, para una matriz ergódica, \mathbf{P}^k es estrictamente positiva para algún valor de k . Este tipo de matrices también se conocen como *primitivas* (31). El teorema de Perron-Frobenius (31, 32) establece que toda matriz primitiva tiene un valor propio máximo que es real y de multiplicidad 1. Asimismo, su vector propio asociado puede seleccionarse de manera que todos sus componentes sean positivos (31, 32). Esto implica que el vector de probabilidad estacionaria es único y que el resto de los valores propios de \mathbf{P} son, en valor absoluto, todos menores a 1. Asumiendo¹ que el resto de los valores propios de \mathbf{P} también tienen multiplicidad 1, la matriz \mathbf{P}^m admite la descomposición espectral dada en la ecuación 3.5, donde \mathbf{U} es la matriz cuyas columnas son los vectores propios a la derecha de \mathbf{P} , $\mathbf{\Pi}$ es la matriz cuyas filas son los respectivos vectores propios a la izquierda y $\mathbf{\Lambda}$ es la matriz diagonal de valores propios (30).

$$\mathbf{P}^m = \mathbf{U} \mathbf{\Lambda}^m \mathbf{\Pi} = \lambda_1^m \mathbf{u}_1 \boldsymbol{\pi}_1^T + \lambda_2^m \mathbf{u}_2 \boldsymbol{\pi}_2^T + \dots \quad (3.5)$$

La suma de la ecuación 3.5 puede re-escribirse de manera más clara utilizando notación de Dirac como se muestra en la ecuación 3.6. Con esto, la propagación de la probabilidad durante la evolución del sistema se descompone en contribuciones dadas por cada uno de los términos en la suma. En toda matriz ergódica, $\lambda_1 = 1$, por lo que el primer término es persistente durante todo el tiempo de evolución. De hecho, dado que $\boldsymbol{\pi}_1$ es el vector de probabilidad estacionaria, la proyección $|\mathbf{u}_1\rangle \langle \boldsymbol{\pi}_1|$ es la matriz límite \mathbf{W} y es el único término de la suma que sobrevive cuando m es grande (31). Esto ocurre porque $|\lambda_i| < 1$ para el resto de los valores propios, de

¹Esta suposición no se requiere si se utiliza la aproximación de rango r de \mathbf{P} (30, 33)

manera que la contribución de cada término se desvanece exponencialmente con m , pues todas las proyecciones van ponderadas por λ_i^m (31).

$$\mathbf{P}^m = \lambda_1^m |\mathbf{u}_1\rangle \langle \boldsymbol{\pi}_1| + \lambda_2^m |\mathbf{u}_2\rangle \langle \boldsymbol{\pi}_2| + \dots \quad (3.6)$$

La descomposición espectral de \mathbf{P}^m indica que los vectores propios de la matriz de transición representan los procesos dinámicos que ocurren a medida que el sistema se acerca al equilibrio (30). Esta expresión también implica que la magnitud de los valores propios son una medida de la persistencia o decaimiento de los procesos que sus vectores asociados describen. En particular, si partimos de una distribución arbitraria $\langle \mathbf{p}_0|$, la propagación m pasos adelante se expresa según la ecuación 3.7. El resultado es una combinación lineal de los vectores $\langle \boldsymbol{\pi}_i|$, donde los coeficientes están dados por el producto entre λ_i^m y $\langle \mathbf{p}_0|\mathbf{u}_i\rangle$. Ésta última expresión es la proyección ortogonal de la probabilidad inicial sobre el vector propio a la derecha \mathbf{u}_i . De manera intuitiva, los signos de los componentes del vector propio a la derecha indican el "flujo" de probabilidad entre los estados a medida que se acercan a la distribución de equilibrio. Por tanto, la estructura de los signos entre los componentes de los vectores propios a la derecha de la matriz de transición explican de manera cualitativa la transiciones esperadas que persisten por un mayor tiempo, de acuerdo con la magnitud de su valor propio, durante la evolución del sistema.

$$\langle \mathbf{p}_0|\mathbf{P}^m = \lambda_1^m \langle \mathbf{p}_0|\mathbf{u}_1\rangle \langle \boldsymbol{\pi}_1| + \lambda_2^m \langle \mathbf{p}_0|\mathbf{u}_2\rangle \langle \boldsymbol{\pi}_2| + \dots \quad (3.7)$$

3.2.2. Análisis de clúster de Perron y Teoría de Rutas de Transición

En el análisis de trayectorias de dinámica molecular utilizando modelos de estado de Markov, el espacio de variables que representa al sistema de interés se divide en n estados discretos mediante algoritmos de clustering. Normalmente, se requieren de varios cientos de clusters para capturar correctamente el comportamiento cinético del sistema (28, 30, 34). Para describir de forma cualitativa el modelo calculado, es útil agrupar los clusters de acuerdo con su metaestabilidad. Un subconjunto de estados en \mathbf{S} forman un grupo metaestable si el sistema pasa un largo tiempo intercambiándose entre los estados que forman parte del grupo antes de evolucionar hacia otro subconjunto (34). De esta forma, los clusters que pertenecen a un mismo grupo metaestable son cinéticamente equivalentes y las barreras de energía entre ellos son bajas. Asimismo, se puede relacionar a los grupos metaestables con valles en la superficie de energía libre (34).

El análisis de clusters de Perron (PCCA++) es un algoritmo de clustering espectral capaz de agrupar una serie de estados en conjuntos metaestables, basado en una matriz de transición (34, 35). El método se basa en que las entradas de esta matriz son una medida de similitud entre los estados, puesto que si la probabilidad de transición para dos de ellos es alta, entonces pueden intercambiarse fácilmente. El algoritmo utiliza la distribución de los estados sobre los vectores propios de la

matriz de transición para calcular vectores de pertenencia a cada uno de los grupos metaestables (34). Los componentes de los vectores de pertenencia indican la factibilidad de que un estado pertenezca a cada uno de los grupos metaestables. En la sección anterior se mostró que los vectores propios de la matriz de transición representan los procesos más lentos que ocurren en el transcurso de la evolución del sistema rumbo a la distribución estacionaria. La distancia entre los estados proyectados sobre cada vector propio es, entonces, proporcional a la frecuencia esperada en que éstos se intercambien. Por lo tanto, realizar el agrupamiento sobre el espacio propio de la matriz de transición equivale a dividir el conjunto de estados en grupos con máxima metaestabilidad (34, 35).

Otro método comúnmente utilizado en el análisis de los modelos de estado de Markov es la teoría de rutas de transición (*Transition Path Theory*, TPT) (36, 37). Retomando que \mathbf{S} representa el conjunto de todos los estados discretos, si se tienen dos subconjuntos de interés \mathbf{A} y \mathbf{B} ($\mathbf{A}, \mathbf{B} \subset \mathbf{S}$), entonces puede haber un cierto número de estados intermedios que los conecten. Una trayectoria reactiva es una realización posible de la cadena de Markov que describe a la dinámica, tal que parte de \mathbf{A} , evoluciona entre los estados intermedios y finalmente llega a \mathbf{B} sin haber regresado a \mathbf{A} (34, 36, 37). Uno de los objetivos de la TPT es describir si existen trayectorias reactivas preferentes en la evolución del sistema desde \mathbf{A} hasta \mathbf{B} . Los subconjuntos \mathbf{A} y \mathbf{B} pueden ser estados conformacionales de interés en una proteína o bien, ser grupos metaestables predichos por PCCA++.

Uno de los conceptos fundamentales en la TPT es la *función de compromiso*, q_i^+ , definida como la probabilidad de que, estando en un estado intermedio $S_i \notin \mathbf{A} \cup \mathbf{B}$, se alcance primero \mathbf{B} antes de \mathbf{A} , (34, 36, 37). Cada estado intermedio tiene un valor de probabilidad de compromiso y su distribución en las rutas que conectan \mathbf{A} y \mathbf{B} describe preferencias entre las posibles trayectorias. De esta manera, es posible calcular el flujo neto por el que el sistema evoluciona entre los estados de interés, identificando los estados de transición más probables a partir de las funciones de compromiso (36). La identificación de los estados de transición mediante este método permite estimar las principales barreras energéticas que deben superarse para observar una evolución desde el estado inicial hasta el final. Es decir, se trata de una forma de caracterizar el mecanismo por el que una determinada transformación ocurre.

En este trabajo se realizaron simulaciones de dinámica molecular utilizando el GLP-1R en presencia y ausencia de moduladores alostéricos. Los resultados fueron utilizados para construir modelos de estado de Markov e identificar, mediante PCCA++, los estados metaestables inducidos por los moduladores. Se utilizó TPT para analizar las relaciones cinéticas entre los estados metaestables de interconversión más lenta, para lograr proponer un mecanismo molecular por el que ocurre la modulación alostérica sobre este receptor, de importancia farmacológica en la diabetes mellitus de tipo 2.

Hipótesis

La presencia de modulador alostérico positivo N-tert-butil-6,7-dicloro-3-metano sulfonilquinoxalin-2-amina (compuesto 2) podría favorecer las transiciones del receptor GLP-1R hacia una conformación activa, mediante cambios inducidos en las interacciones de los residuos de aminoácido en la región intracelular. La barrera de energía libre que mantiene el receptor en conformación inactiva puede disminuir debido a la unión del compuesto 2.

Objetivos

5.1. Objetivo general

- Identificar las interacciones ligando-proteína que inducen los cambios estructurales relacionados con la modulación alostérica en el receptor GLP-1R.

5.2. Objetivos específicos

- Modelar la unión de la N-tert-butil-6,7-dicloro-3-metanosulfonilquinoxalin-2-amina (compuesto 2) al receptor GLP-1R para obtener una estructura del complejo receptor-modulador.
- Efectuar las simulaciones de dinámica molecular del receptor unido al agonista, en presencia y ausencia del compuesto 2 como modulador positivo y en complejo con el modulador negativo NNC0640.
- Analizar de manera comparativa las trayectorias de cada sistema, generando modelos de estado de Markov, para identificar las diferencias que puedan estar asociadas con el efecto modulador de los ligandos.

Métodos

6.1. Preparación de la estructura del GLP-1R

1. La estructura del dominio transmembranal del GLP-1R fue obtenida del Protein Data Bank con el PDBID: 5VEX (18). Para favorecer el aislamiento y la determinación de la estructura por difracción de rayos X, los autores introdujeron 10 mutaciones termoestabilizantes, mostradas en la tabla 6.1. El puente disulfuro entre los residuos I317C – G361C mantiene al receptor en una conformación inactiva. En esta estructura no se resolvió la posición de diversos fragmentos: el extremo N-terminal, que corresponde con el conector entre el ECD y el TMD en el receptor completo, un fragmento del ECL1, la mayor parte del ECL3 y el extremo C-terminal.
2. La secuencia nativa del GLP-1R humano fue obtenida de UniProt, con el código identificador P43220. La secuencia del receptor mutado, contenida en la estructura 5VEX, fue alineada contra la secuencia nativa en el servidor ClustalW.
3. Utilizando el programa Modeller 9.23 y con el alineamiento obtenido de ClustalW, se calcularon 100 modelos por homología utilizando la estructura 5VEX como molde. Esto permitió revertir las mutaciones introducidas por los autores para obtener la estructura con la secuencia nativa del receptor. El programa también permite modelar los fragmentos no resueltos en la difracción de rayos X.
4. Para los sistemas en complejo con el GLP-1, la unión del péptido al receptor se obtuvo utilizando Modeller, tratando la cadena peptídica del ligando como otra subunidad. La estructura molde para el emplazamiento del GLP-1 se obtuvo del PDB con el PDBID:5VAI (16).
5. Los modelos calculados fueron evaluados utilizando PROCHECK (38, 39). Las estructuras con los mejores puntajes, utilizando como criterio los factores G de PROCHECK, se seleccionaron para ser utilizadas en los análisis posteriores.
6. En total se prepararon 4 sistemas, como se describen en la tabla 6.2.

Mutación	Posición
S193C	2.63b
I196F	2.66b
S225A	3.28b
M233C	3.36b
S271A	4.47b
I317C	5.47b (Puente disulfuro)
G318I	5.48b
K346A	6.35b
C347F	6.36b
G361C	6.50b (Puente disulfuro)

Tabla 6.1: Mutaciones termoestabilizantes encontradas en la estructura del receptor, introducidas por los autores del estudio para facilitar la cristalización. La posición se define con la nomenclatura de Ballesteros-Weinstein, donde el primer número indica la hélice transmembranal a la que pertenece el residuo y el segundo número es la posición relativa al residuo más conservado, que ocupa la posición número 50.

Sistema	Componentes
GLP-1	Receptor y péptido agonista (GLP-1)
GLP-1 + CM2	Receptor, péptido agonista y compuesto 2
CM2	Receptor y compuesto 2
NAM	Receptor y modulador negativo NNC0640

Tabla 6.2: Sistemas utilizados para las simulaciones por dinámica molecular.

6.2. Acoplamiento molecular covalente

1. Las estructuras seleccionadas fueron preparadas para los cálculos de acoplamiento molecular utilizando el programa Molecular Operating Environment 2019.01. Los estados de protonación a 300 K y pH 7.0 fueron asignados utilizando la herramienta Protonate 3D. Los detalles generales de este algoritmo pueden consultarse en el apéndice A.
2. Las estructuras con los átomos de hidrógeno agregados fueron sometidas a una minimización de energía, utilizando el campo de fuerzas Amber10:EHT. Este campo combina los parámetros de Amber10 para proteínas y ácidos nucleicos y utiliza una parametrización de moléculas pequeñas basada en la teoría extendida de Hueckel. El algoritmo de minimización de MOE utiliza tres estrategias consecutivas: primero realiza pasos de descenso en dirección contraria

al gradiente (*Steepest Descent*), luego, cuando el valor del gradiente disminuye, cambia a pasos por gradiente conjugado y, finalmente, cuando el gradiente tiene un valor muy pequeño, utiliza el método truncado de Newton hasta alcanzar la convergencia. Se impusieron restricciones armónicas respecto a las posiciones iniciales de todos los átomos para prevenir desviaciones muy grandes de la estructura original. Por defecto, las restricciones agregan el término $(3/2)kT/\sigma^2(x_0 - x)^2$ a la función de energía, con un valor de $\sigma = 0.5$, donde x_0 representa las coordenadas iniciales de cada átomo.

3. La estructura del compuesto 2 fue construida con la herramienta Builder de MOE. Se efectuó una optimización de geometría utilizando el campo de fuerzas MMFF94x, mediante el mismo algoritmo que la minimización del receptor.
4. El protocolo de acoplamiento molecular covalente requiere que la reacción sea definida *a priori* con un mapeo de átomos para poder ser identificados en la estructura que corresponderá a los productos. Dado que la reacción de sustitución nucleofílica aromática involucrada en la unión del compuesto 2 al residuo de cisteína no se incluye en la biblioteca de MOE, la reacción fue definida en Marvin Sketch de ChemAxon. La reacción fue guardada como archivo .rdf y cargada en la utilidad de acoplamiento molecular covalente de MOE. En la figura 6.1 se muestra la reacción de sustitución nucleofílica aromática involucrada.

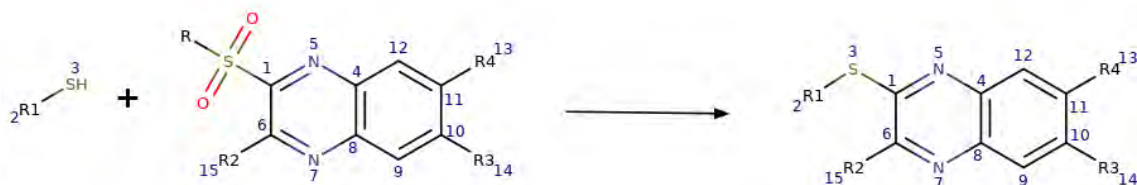


Figura 6.1: Reacción de sustitución nucleofílica aromática involucrada en la unión covalente del compuesto 2 al residuo C347^{6.36b}. Se muestra el mapeo de átomos utilizado por MOE para identificar la formación del aducto.

5. El sitio de unión fue especificado mediante la selección del residuo C347^{6.36b} y una extensión de 8 Å alrededor de él.
6. El algoritmo de acoplamiento molecular implementado en MOE se describe en el apéndice A. Se utilizó el método de ajuste inducido (que permite el movimiento de las cadenas laterales de los aminoácidos) para el refinamiento de las conformaciones. La función de puntaje utilizada fue el valor de dG de London.
7. Se calculó un máximo de 20 conformaciones por ejecución del acoplamiento molecular. El cálculo se repitió 3 veces de manera independiente. La conformación con el menor valor de la función de puntaje fue utilizada para proceder a la dinámica molecular. Las estructuras con los complejos fueron guardadas en formato pdb.

6.3. Dinámica Molecular

6.3.1. Parametrización de los ligandos

1. En las simulaciones de dinámica molecular, el ligando no covalente fue modelado utilizando el campo de fuerzas general de AMBER (*General AMBER Force Field*, GAFF). Este campo de fuerzas cubre una amplia variedad de moléculas orgánicas y es compatible con otros campos de fuerzas de AMBER para biomacromoléculas. El residuo resultante del aducto cisteína-compuesto 2 se modeló con los parámetros disponibles en el campo de fuerzas ff14SB.
2. Los parámetros para el ligando NNC0640, que se une de manera no covalente al receptor, fueron asignados utilizando el programa *antechamber*. Posteriormente, se utilizó el programa *parmchk2* para verificar la integridad de los parámetros en la molécula. De acuerdo con el manual de usuario, el programa *parmchk2* se encarga de buscar en la base de datos del campo de fuerza especificado los parámetros requeridos para modelar un ligando particular. Si hay algún parámetro (como una constante de elongación de enlace o angular) que no se encuentre disponible en el campo de fuerzas, el programa trata de asignar un valor equivalente, más general, acompañado de un valor de penalización que mide el grado de similitud entre el parámetro requerido y el asignado. Si no es posible asignar tampoco un valor por similitud, el programa genera una advertencia con la leyenda "ATTN, need revision", indicando que el campo de fuerzas no es capaz de modelar a la molécula, por lo que se requeriría la parametrización *ab initio* del ligando (40).
3. Para realizar la parametrización del residuo de cisteína modificado de manera covalente, primero se construyó la estructura del aducto cisteína-compuesto 2 en el programa Avogadro. Los grupos amino y carboxilo de la cadena principal del aminoácido fueron extendidos con un grupo acetilo y N-metilo, respectivamente.
4. La asignación de los parámetros del campo de fuerza se realizó utilizando el programa *antechamber*, con los valores disponibles en la biblioteca del campo de fuerzas ff14SB. El molde con la estructura y los parámetros se guardó como archivo .ac.
5. En un archivo .mc fueron definidos los átomos del residuo molde que pertenecerán a la cadena principal, de acuerdo con el manual de AMBER. El molde del nuevo residuo y el archivo con la definición de la cadena principal se utilizaron como entradas del programa *prepgen*, que se encarga de identificar la cadena principal en el nuevo residuo y origina un archivo de entrada para generar los parámetros del residuo modificado.
6. El archivo generado por *prepgen* se utilizó como entrada en el programa *parmchk2*. El programa busca y asigna los parámetros para cada tipo de átomo del molde en las bases de datos de los campos de fuerza indicados y escribe un archivo .frcmod para que el nuevo aminoácido pueda ser identificado y

modelado por AMBER, de la misma manera en que se describió previamente. Se verificó que todos los parámetros requeridos para modelar el residuo modificado estuvieran disponibles en el campo de fuerzas.

6.3.2. Hidratación del canal del receptor mediante el modelo de sitios de interacción de referencia (RISM)

1. Antes de los cálculos de dinámica molecular, el canal del receptor fue hidratado. La posición óptima de las moléculas de agua en las cavidades intra y extracelulares se calculó mediante el modelo de sitios de interacción de referencia (3D-RISM). En este modelo, se resuelven de forma numérica las ecuaciones integrales de Ornstein-Zernik para obtener la función de distribución radial del disolvente alrededor del soluto. La función de distribución radial es proporcional a la probabilidad de que una molécula de agua se encuentre en determinada posición alrededor de la proteína. De esta forma, el patrón de hidratación se puede predecir mediante los máximos de la función de distribución radial calculada.
2. Primero se obtuvo la susceptibilidad del disolvente mediante un cálculo de RISM 1D utilizando el programa *rism1d* en AMBER. Básicamente, la susceptibilidad se calcula a partir de las funciones de correlación de las moléculas de agua consigo mismas y es necesario para el cálculo en presencia de cualquier otro soluto. El cálculo se efectuó con una temperatura de 300 K, en presencia de NaCl a una concentración de 0.150 M, utilizando el modelo TIP3P para el agua y la relación de cierre de Kovalenko-Hirata. El modelo TIP3P consiste en una molécula rígida con 3 cargas puntuales y se trata de uno de los modelos de agua explícitos más utilizados en simulación molecular porque tiene un balance aceptable entre precisión y exigencia computacional. Aunque modelos con mayores grados de libertad ofrecen algunas mejoras en la reproducción de propiedades experimentales, suele incrementarse la cantidad de esfuerzo computacional requerido en el cálculo, por lo que el modelo TIP3P es adecuado en particular en sistemas grandes como el estudiado en este trabajo. La susceptibilidad es guardada en un archivo .xvv.
3. Las estructuras de los sistemas en formato pdb fueron utilizadas para generar topologías y coordenadas de inicio, mediante el programa *tLeap* de AMBER. Las proteínas se modelaron utilizando el campo de fuerzas ff14SB de AMBER y los ligandos mediante GAFF, como se describe en la sección de parametrización.
4. Las topologías, coordenadas iniciales y la susceptibilidad del disolvente fueron utilizadas como entrada para el cálculo de 3D-RISM, utilizando el programa *rism3d.snglpt* de AMBER.
5. Las funciones de distribución radial para cada tipo de átomo del disolvente se guardaron en formato .dx. Estos archivos contienen el valor de la función evaluada en una red de puntos extendida alrededor de la proteína.

6. Para identificar los puntos de máxima probabilidad, se generó un mapa con el laplaciano de la función de distribución radial del átomo de oxígeno del agua. El valor del laplaciano fue calculado con el programa *metatwist* de AMBER.
7. Los átomos de oxígeno de las moléculas de agua se posicionaron en los centroides de las regiones con valores negativos del laplaciano. Estas regiones involucran zonas donde la función de distribución radial se concentra, por lo que la probabilidad es máxima. El posicionamiento de las moléculas de agua se realizó con *metatwist*, utilizando un valor límite de 0.5. El valor límite controla el umbral de valores del laplaciano que habrán de considerarse para posicionar las moléculas de agua. Todas las regiones con un valor del laplaciano entre el valor mínimo encontrado y el producto de este valor mínimo por el límite serán seleccionadas para agregar moléculas de agua. Por tanto, el valor límite controla, de manera indirecta, la cantidad de moléculas de agua que se agregarán, pero esta cantidad depende de la forma de la función de distribución radial y de su laplaciano.
8. Las estructuras del receptor hidratado se guardaron en formato pdb.

6.3.3. Empaquetado del receptor en la membrana

1. Los sistemas del receptor en la membrana fueron preparados utilizando el programa *packmol-memgen* de AMBER (41). El programa calcula la orientación de la proteína de acuerdo con el algoritmo del programa *membed* (42) y después utiliza *Packmol* para empaquetar todos los componentes.
2. Se utilizó una composición de la membrana de POPC:POPE:Coolesterol en proporción 2:2:1. Fue incluido NaCl en una concentración de 0.150 M. La distancia mínima entre la proteína y el límite de la membrana y las celdas de agua fue de 10 Å. Los sistemas empaquetados en la membrana fueron guardados en formato pdb.
3. Las estructuras producidas por *packmol-memgen* se utilizaron para generar las topologías y coordenadas iniciales para las simulaciones, mediante el programa *tleap*. El campo de fuerzas para modelar la proteína fue el ff14SB (43), para los lípidos se utilizó Lipid14 (44) y el modelo de agua fue el TIP3P. Los ligandos se modelaron con el campo GAFF (45) como se indica en la sección de parametrización.

6.3.4. Equilibrio y producción

1. Los sistemas en membrana fueron utilizados como punto de partida para simulaciones por dinámica molecular, utilizando el programa *pmemd.cuda* de AMBER (40). Las simulaciones se calcularon en los nodos con soporte para GPU de la supercomputadora Miztli, utilizando tarjetas gráficas NVIDIA Tesla M2090.

2. Los pasos de minimización y equilibrio se detallan en la tabla 6.3. Los sistemas fueron sometidos a un proceso de minimización de geometría, utilizando 10000 pasos iniciales de descenso de gradiente seguidos por 5000 pasos de gradiente conjugado. La temperatura del sistema se elevó hasta 310 K en un paso de calentamiento durante 25 ps. Posteriormente se efectuaron una serie de 6 etapas de equilibrio, como se detallan en la tabla 6.3, con un total de 10.35 ns de simulación. En todos los casos se impusieron restricciones armónicas, relajando paulatinamente los valores de las constantes de fuerza hasta liberar el sistema por completo. Los valores de las constantes de fuerza para cada etapa se muestran en la tabla 6.3.

Etapa	Duración	Constantes de fuerza ($\text{kcal mol}^{-1} \text{Å}^{-2}$)				
		SC	MC	Lig	P	CHL
Minimización	10000 pasos SD 5000 pasos CG	4.78	9.56	9.56	0.0	0.0
Calentamiento	25 ps	4.78	9.56	9.56	2.39	2.39
Equilibrio 1	25 ps NVT	2.39	4.78	4.78	2.39	2.39
Equilibrio 2	25 ps NVT	1.20	2.39	2.39	0.96	0.96
Equilibrio 3	100 ps NPT	0.48	1.20	1.20	0.48	0.48
Equilibrio 4	100 ps NPT	0.12	0.48	0.48	0.10	0.10
Equilibrio 5	100 ps NPT	0.0	0.12	0.12	0.0	0.0
Equilibrio 6	10 ns NPT	0.0	0.0	0.0	0.0	0.0

Tabla 6.3: Duración y condiciones de las etapas de minimización, calentamiento y equilibrio, previo a la producción de la dinámica molecular. SD, descenso de gradiente; GC, gradientes conjugados; SC, átomos de las cadenas laterales de los residuos de aminoácido; MC, átomos de la cadena principal en la proteína; Lig, átomos del ligando; P, átomos de fósforo de los fosfolípidos en la membrana; CHL, moléculas de colesterol en la membrana.

3. En todos los casos, se utilizó un tiempo de integración de 2 fs, aplicando el algoritmo SHAKE para restringir el movimiento de los enlaces que involucran hidrógeno. Las simulaciones se llevaron a cabo bajo condiciones periódicas, introduciendo un valor de corte para interacciones de no enlace de 10 Å. Las correcciones de largo alcance se calcularon mediante el método de Ewald (PME).
4. Para el control de la temperatura a 310 K se utilizó el termostato de Langevin, con una frecuencia de colisión de 1.0 ps^{-1} . La presión fue regulada a 1 bar mediante el barostato de Berendsen con un tiempo de relajación de 4.0 ps.
5. Después de las etapas de minimización, calentamiento y equilibrio, se ejecutaron rondas de producción de 500 ns para cada sistema. Las estructuras se guardaron cada 0.1 ns en el archivo de trayectoria. Fueron efectuadas 3 repeticiones independientes para cada uno de los cuatro sistemas, obteniendo un total de 6 μs de simulación.

6.4. Modelado de Estados de Markov

Se generó un modelo de estado de Markov por cada uno de los sistemas estudiados en las simulaciones, como se enlistan en la tabla 6.2. Los valores de los hiperparámetros se ajustaron de acuerdo con cada sistema y éstos se presentan en la sección de resultados. La generación y análisis de los modelos se calcularon utilizando la biblioteca PyEMMA en Python (46). El error en las variables calculadas se estimó utilizando *bootstrapping*, con excepción de las cantidades derivadas de la matriz de transición, cuya incertidumbre se calcula por muestreo de Gibbs como se detalla más adelante.

6.4.1. Extracción de variables

1. Las variables seleccionadas para generar los modelos de Markov fueron las contribuciones por aminoácido al área superficial accesible al disolvente de los residuos en contacto cercano ($<5 \text{ \AA}$) con la proteína G, observados en la conformación activa. La lista de los residuos involucrados se muestra en la tabla 6.4. El área superficial accesible al disolvente mide la exposición de los residuos de aminoácido hacia la superficie de la proteína, por lo que es un buen indicador de la formación de cavidades en la estructura. Además, esta variable correlaciona con los contactos de las cadenas laterales de aminoácido respecto a otros residuos o ante la presencia de ligandos.

Residuos de aminoácido involucrados				
R176	L254	S261	K334	L356
H180	L255	E262	L335	L359
E247	A256	V327	R348	Y402
Y250	S258	I330	S352	N406
L251	V259	V331	T355	N407

Tabla 6.4: Lista de residuos de aminoácido del GLP-1R que entran en contacto con la proteína G. Su contribución al área superficial accesible al disolvente fue utilizada como espacio de variables en el modelado de Markov.

2. El cálculo de los valores de SASA se realizó en el programa VMD. Para cada una de las estructuras en las trayectorias de los sistemas utilizados en las simulaciones, se calculó el valor de área superficial total para la cavidad, $\mathbf{s}_T(\mathbf{t})$, y la contribución por aminoácido, $\mathbf{s}_i(\mathbf{t})$. De esta manera, las observaciones del espacio de parámetros se representan mediante la matriz \mathbf{S} de $n \times m$, con n pasos de tiempo y m valores de área superficial.

6.4.2. Reducción de dimensiones

1. El espacio de las variables seleccionado para generar el modelo se proyectó a un subespacio de menor dimensionalidad utilizando análisis de componentes independientes con desfase en el tiempo (*Time-lagged Independent Component Analysis*, TICA) (47). TICA resuelve el problema generalizado de valores propios mostrado en la ecuación 6.1, donde \mathbf{C}_0 es la matriz de covarianza para las variables del modelo y \mathbf{C}_τ es la matriz de covarianza entre las variables y sus valores con un desfase de τ pasos.

$$\mathbf{C}_\tau \mathbf{r}_i = \mathbf{C}_0 \lambda_i \mathbf{r}_i \quad (6.1)$$

El conjunto de vectores $\{\mathbf{r}_i\}$, llamados *componentes independientes*, forman una nueva base sobre la que se proyectan las variables originales.

2. La selección del desfase se realizó mediante el análisis de la convergencia de las escalas de tiempo implicadas; que se relacionan con los valores propios de los componentes independientes por la ecuación 6.2.

$$t_i = -\frac{\tau}{\ln |\lambda_i|} \quad (6.2)$$

3. El número de dimensiones por mantener se eligió de forma que diera origen a una prueba de Chapman-Kolmogorov válida, es decir, preservara la propiedad de Markov.

6.4.3. Generación de los microestados (discretización)

1. Cada una de las observaciones proyectadas en el espacio de componentes independientes fue asignada a un microestado utilizando el algoritmo de *k-medias* (*k-means*). Este algoritmo separa el espacio en grupos mutuamente exclusivos, ($A_i \cap A_j = \emptyset, \forall i, j \in k$), basado en la distancia entre los puntos (un parámetro puramente geométrico).
2. El número de microestados (*clusters*) se seleccionó de manera que maximizara el valor *VAMP-2* (33) para el modelo de Markov estimado. De manera general, el valor VAMP-2 es una extensión del principio variacional para la aproximación de los valores propios más altos de una matriz, dado que estos representan una cota superior cuando los vectores propios no son conocidos (33). Es decir, un valor mayor de VAMP-2 implica una mejor aproximación de los valores y vectores propios para la matriz de transición que representa el proceso de Markov inherente al fenómeno de estudio. En el apéndice B se presentan mayores detalles acerca del método VAMP.

6.4.4. Estimación de las matrices de transición y validación de los modelos

1. Los elementos de la matriz de transición se estimaron mediante el algoritmo de inferencia Bayesiana implementado en PyEMMA (48). Para un modelo de Markov con espacio de estados discreto, los parámetros a estimar son las entradas de la matriz de transición. La distribución *a posteriori* de los parámetros del modelo, condicionada a las transiciones observadas en la simulación, está dada por el teorema de Bayes de la ecuación 6.3, donde P representa las entradas de la matriz de transición a estimar (los parámetros del modelo) y C son los conteos de las transiciones observadas. Los detalles de la función de verosimilitud ($\mathbb{P}[C|P]$) y de la distribución *a priori* ($\mathbb{P}[P]$) se pueden consultar en (48).

$$\mathbb{P}[P|C] \propto \mathbb{P}[C|P] \mathbb{P}[P] \quad (6.3)$$

2. El algoritmo usa un muestreo de Gibbs de la distribución *a posteriori* para generar un conjunto de matrices de transición con la que se pueden calcular valores esperados e intervalos de confianza.
3. La selección del desfase se realizó mediante el análisis de la convergencia de las escalas de tiempo implicadas, que cumplen la misma relación de la ecuación 6.2, para los valores propios de la matriz de transición.
4. Los modelos se validaron utilizando la prueba de Chapman – Kolmogorov, en la que se verifica el cumplimiento de la igualdad 6.4. Es decir, si el sistema cumple la propiedad de Markov, la potencia k de la matriz de transición debe ser equivalente a la matriz estimada con $k\tau$ pasos de tiempo (28, 30, 34).

$$\mathbf{P}(k\tau) = \mathbf{P}^k(\tau) \quad (6.4)$$

6.4.5. Análisis de los modelos de estado de Markov

1. Los microestados se agruparon por su similitud cinética, predicha por la matriz de transición, utilizando el análisis de clúster de Perron (*Perron Cluster Cluster Analysis*, PCCA) mediante el algoritmo PCCA++ (35). Este algoritmo genera un vector de pertenencia para cada microestado. Cada componente i del vector es la probabilidad de que el microestado pertenezca al grupo i . Los grupos son cúmulos de microestados cinéticamente equivalentes, es decir, hay un intercambio rápido entre ellos. Por su parte, los microestados en diferentes grupos se interconvierten muy lentamente. Los grupos calculados por PCCA++ son identificados como los macroestados del sistema o grupos metaestables.
2. A partir de la agrupación en estados metaestables, se calcularon las trayectorias reactivas más frecuentes y los valores de la función de compromiso para los estados intermedios, entre los estados de interconversión más lenta, de acuerdo con lo predicho con el segundo vector propio de la matriz de transición,

utilizando los algoritmos de la teoría de rutas de transición en PyEMMA (36, 37).

3. Para identificar las diferencias entre las conformaciones en términos de las estructuras del receptor y el ligando, se muestrearon las estructuras de las trayectorias de acuerdo con los valores de pertenencia obtenidos en el análisis de clústers de Perron. Se calcularon los contactos e interacciones de la red polar intracelular y entre los ligandos y la proteína, utilizando VMD.

6.5. Cálculos de energía libre con *Umbrella Sampling*

Con el fin de caracterizar el cambio de energía libre asociado con la ruptura del candado iónico en la red polar intracelular, se realizó un muestreo con el método *Umbrella Sampling* de acuerdo con el siguiente protocolo:

1. La coordenada de reacción fue definida como la distancia entre el átomo de carbono ζ del residuo R348^{6.35b} y el átomo de carbono δ del residuo E408^{8.49b}, y fue dividida en 25 ventanas, partiendo de una distancia de 4.0 Å hasta 10.0 Å, en intervalos de 0.25 Å.
2. Partiendo de las estructuras equilibradas de las simulaciones, se calcularon dinámicas moleculares sobre cada una de las ventanas, imponiendo una restricción armónica a la energía de acuerdo con la ecuación 6.5. En esta ecuación, k es el valor de la constante de fuerza sobre la restricción, d es el valor de la coordenada de reacción observada en la simulación y d_0 es el centro de la ventana. En todas las ventanas se utilizó un valor de $k = 100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Fueron calculados 20 ns de simulación en cada ventana, descartando los primeros 10 ns como equilibrio.

$$V_{bias} = \frac{1}{2}k(d - d_0)^2 \quad (6.5)$$

3. El sesgo fue removido para reconstruir el cambio de energía libre a lo largo de la coordenada de reacción mediante el método de análisis de histogramas ponderados ([Weighted Histogram Analysis Method], WHAM). El cálculo se realizó con el programa WHAM, desarrollado por el Dr. Grossfield (49).
4. Se realizó un cálculo de energía libre para el sistema en ausencia de modulador alostérico y otro para el sistema en presencia de modulador.

Resultados

7.1. Preparación de la estructura y validación

La estructura reportada en el Protein Data Bank con el PDBID: 5VEX corresponde con el dominio transmembranal del GLP-1R humano. Los autores obtuvieron los cristales en fase cúbica lipídica y determinaron la estructura mediante difracción de rayos X, con una resolución de 3 Å. Las mutaciones presentes en esta estructura incluyen un puente disulfuro artificial entre los residuos I317^{5.47b}C - G361^{6.50b}C y la sustitución C347^{6.36b}F. Estas modificaciones fueron hechas experimentalmente por los autores para lograr cristalizar al receptor y se encuentran presentes en la estructura reportada en el PDB. El puente disulfuro artificial conecta las hélices TM5 y TM6, restringiendo al receptor en una conformación inactiva. Por su parte, el residuo C347^{6.36b} es el punto de unión covalente del compuesto 2, por lo que fue necesario revertir las mutaciones de acuerdo con la secuencia nativa del GLP-1R.

Las mutaciones fueron revertidas con el programa Modeller, utilizando como molde la estructura cristalográfica y como objetivo la secuencia nativa del receptor. Este proceso permite optimizar los conformeros de los residuos no compartidos entre las secuencias y modelar las regiones no resueltas en la difracción de rayos X. En la figura 7.1 se muestra una gráfica con el valor de puntaje de Modeller para cada uno de los 100 modelos calculados. La selección del número de modelos se basa en que la estructura del GLP-1R es conocida y el protocolo de modelado por homología implementado en Modeller se utilizó para hacer los cambios de los residuos de aminoácido mutados sin introducir choques o geometrías desfavorables, tratando de mantener la geometría del modelo cercana a la reportada experimentalmente. Además, cualquier desviación o conformación de alta energía podrá relajarse finalmente en las etapas de minimización y dinámica molecular subsecuentes. La función objetivo utilizada por Modeller, conocida como DOPE (*Discrete Optimized Protein Energy*), mide el grado de desviación de las estructuras generadas respecto del molde y de restricciones geométricas impuestas basados en parámetros empíricos (50). La función DOPE es un potencial estadístico desarrollado para medir la desviación de la geometría de un modelo respecto a lo comúnmente observado en estructuras cristalográficas, con base en teoría de probabilidades. Tanto la escala como las dimensiones de esta función son arbitrarias y no suele representarse mediante unidades físicas (50, 51). De esta forma, los valores más bajos de la función de puntaje indican modelos con un mejor ajuste a la estructura molde y geometrías moleculares normalmente observadas.

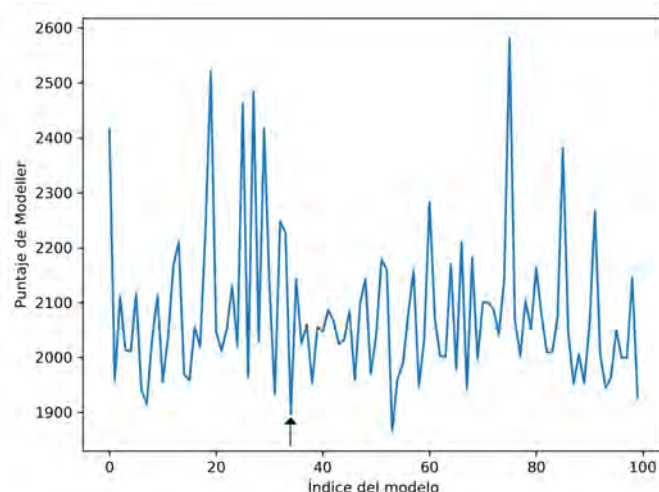


Figura 7.1: Puntaje de modeller para cada uno de los modelos generados. Los modelos son identificados mediante un índice del 1 al 100

La calidad de las estructuras modeladas se evaluaron en grupo, de forma comparativa, mediante los factores G de PROCHECK. Los factores G son proporcionales a la probabilidad¹ de observar cada parámetro conformacional, como los ángulos de torsión o las distancias de enlace, en relación con lo regularmente observado para estructuras de alta resolución (39). PROCHECK calcula los valores G para cada uno de los ángulos de torsión de la cadena principal y la cadena lateral, y para los ángulos y distancias de enlace de la cadena principal, reportando los valores promedio para cada estructura. Los resultados de los factores G para cada uno de los modelos se grafican en la figura 7.2. En esta gráfica, la línea marcada como Diedros hace referencia a los factores G para los ángulos diedros de la cadena principal y las cadenas laterales, mientras que la línea marcada como Covalente se refiere a los factores G para las distancias de enlace. La línea Promedio indica el valor promedio de estos valores (39).

Valores de factores G superiores a -0.5 indican la presencia de geometrías usualmente observadas en las estructuras de proteínas. Valores menores se relacionan con parámetros geométricos inusuales. La figura 7.2 muestra que las 100 estructuras generadas poseen factores G mayores a -0.5 y son muy similares entre sí. El modelo con el índice 35 posee uno de los menores valores del puntaje de Modeller en conjunto con un valor alto de factores G. Por esta razón, se seleccionó esta estructura para utilizar en los acoplamientos moleculares y en la dinámica molecular. En la figura 7.3 se muestra el gráfico de Ramachandran para el modelo 35. Este modelo tiene una proporción de 94.1 % de los residuos en las regiones más favorecidas, como se espera de estructuras con buena calidad. Solamente uno de los residuos, la N304^{5,34b}, se encuentra en una región prohibida y se ubica en el extremo extracelular de la hélice TM5. De acuerdo con estos parámetros, la geometría calculada por Modeller es satisfactoria para su uso en simulación, indicando que el número de modelos generados es suficiente para este propósito. Las pocas desviaciones encontradas podrán relajarse en las posteriores etapas de minimización y dinámica molecular.

¹Específicamente se trata de un logaritmo de razón de probabilidades *odds ratio* y es, por tanto, adimensional

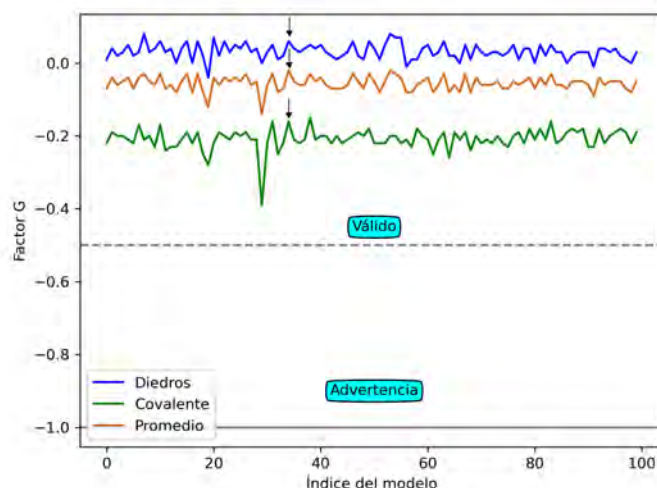


Figura 7.2: Resultado de los factores G de ángulos dihedros, de parámetros covalentes y globales para cada uno de los modelos generados por Modeller

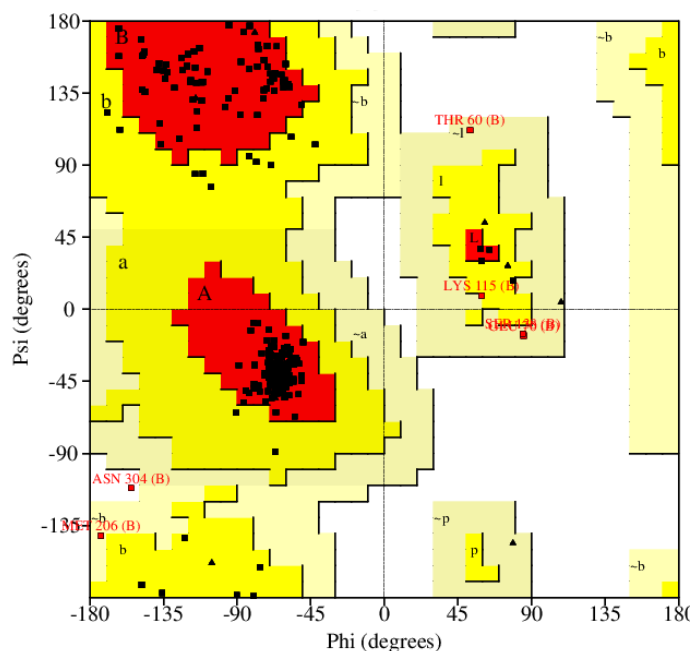


Figura 7.3: Gráfico de Ramachandran para el modelo 35. Los residuos de glicina se muestran como triángulos. En rojo se muestran las regiones más favorecidas y en amarillo las regiones permitidas.

7.2. Acoplamiento molecular covalente

A partir del cálculo de acoplamiento molecular, se obtuvieron 40 conformaciones del complejo ligando-receptor. En el histograma de la figura 7.4 se muestra la distribución de los valores de la energía de unión de London de las conformaciones calculadas. El algoritmo de acoplamiento molecular converge en dos mínimos de la superficie de energía de unión, con valores alrededor de $-6.5 \text{ kcal mol}^{-1}$ y $-3.0 \text{ kcal mol}^{-1}$. Para comparar las conformaciones encontradas, se realizó un análisis de componentes principales a partir de las coordenadas cartesianas de los átomos

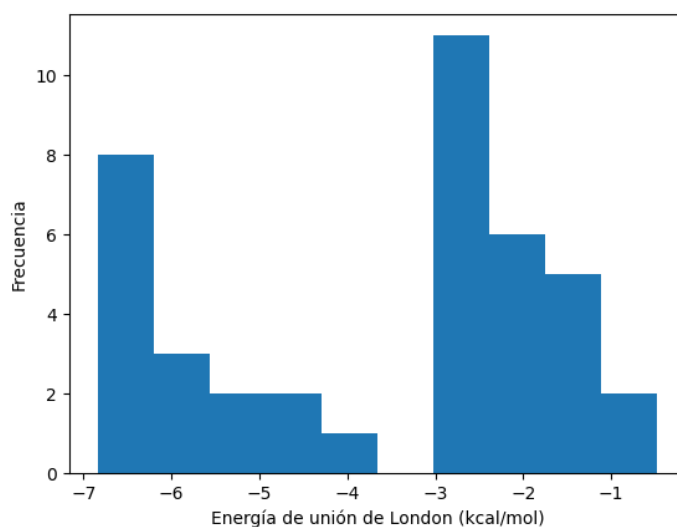


Figura 7.4: Resultados de la energía de unión de London, calculada para las 40 conformaciones obtenidas.

del ligando. La proyección de los dos primeros componentes principales se muestra en la figura 7.5, junto con los valores de la energía de unión para cada conformero.

La gráfica de la figura 7.5 muestra la distribución en el espacio conformacional explorado mediante el acoplamiento. Las conformaciones que se agrupan en una misma región también comparten valores similares de energía de unión y corresponden a mínimos locales en la función de puntaje. Los resultados de los cálculos independientes siempre convergen hacia la región de menor energía, alrededor de los $-6.0 \text{ kcal mol}^{-1}$, ubicada en la parte inferior del gráfico. Esto muestra que los resultados del acoplamiento molecular son reproducibles. La estructura con el menor valor de energía de London, con $-6.8 \text{ kcal mol}^{-1}$, fue seleccionada para utilizarla en los cálculos de dinámica molecular.

7.3. Dinámica molecular

7.3.1. Parametrización de los ligandos

El ligando covalente fue modelado utilizando los parámetros disponibles en el campo de fuerzas para proteínas ff14SB. Este campo fue desarrollado para mejorar y abordar las limitaciones de su antecesor, el ff99SB (43). En ff14SB, los parámetros de los ángulos de torsión para la cadena principal fueron recalculados para reproducir constantes de acoplamiento observadas en resonancia magnética nuclear y para representar de manera más precisa la formación de elementos de estructura secundaria. Resultados obtenidos por los desarrolladores y por otros muestran que a partir del campo de fuerzas ff14SB se obtienen resultados de entalpías de hidratación (52), coeficientes de difusión (52), coeficientes de partición (53), reproducción de estruc-

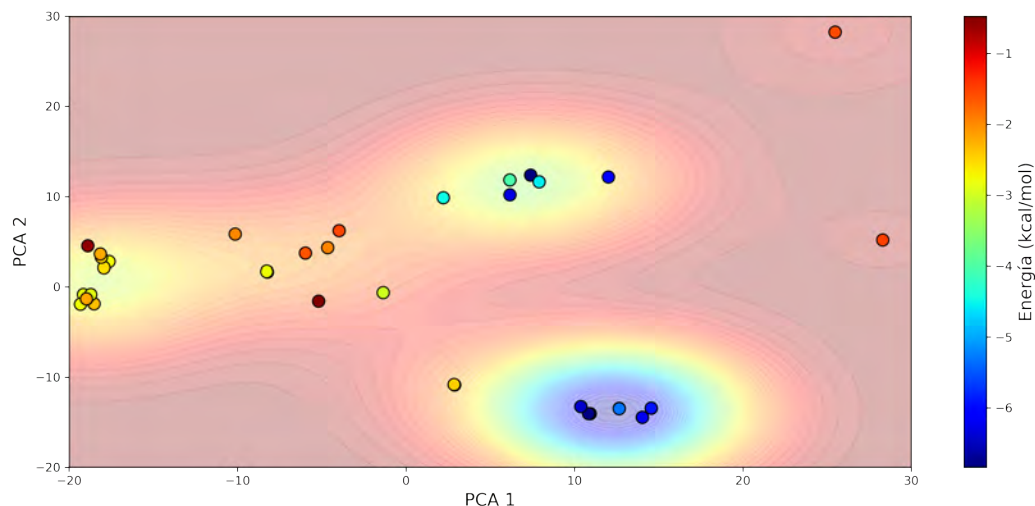
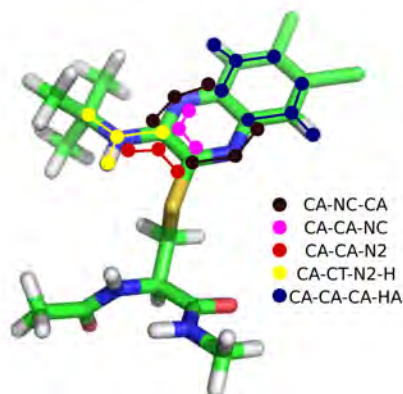


Figura 7.5: Mapeo en dos dimensiones de los conformeros obtenidos luego del cálculo de acoplamiento molecular. El color del marcador y del fondo es proporcional al valor de la energía de London para esa conformación.

turas nativas en simulaciones de alto rendimiento (54) y observables de resonancia magnética nuclear (43) comparables con resultados experimentales, representando una mejora respecto a sus antecesores. Por su parte, se utilizó el campo de fuerzas GAFF para modelar el ligando de unión no covalente. GAFF tiene prácticamente la misma forma funcional que los campos de fuerzas de AMBER dedicados a biomacromoléculas, pero sus parámetros han sido desarrollados para una gran cantidad de moléculas orgánicas. El programa *parmchk2* verifica que en la base de GAFF y de ff14SB se encuentren los parámetros para la molécula a modelar y asigna valores generales por defecto si no se encuentran los específicos. La figura 7.6 muestra la estructura del aducto del compuesto 2 con la cisteína. La mayor parte de los parámetros fueron encontrados en el campo ff14SB, y sólo algunos fueron asignados al valor general por defecto, principalmente de ángulos de enlace y ángulos de torsión impropios. Ninguno de los parámetros requeridos estuvo ausente en la base del campo de fuerzas. En la figura 7.6 se muestran los valores generales asignados a los parámetros de enlace en el compuesto 2, en conjunto con su valor de penalización.

Con el objetivo de determinar si los parámetros asignados son satisfactorios, a pesar de los valores de penalización, se calculó el perfil de energía potencial en función de los ángulos de enlace señalados en la figura 7.6, utilizando la parametrización mediante el campo de fuerzas y comparando con cálculos de estructura electrónica con el nivel de teoría B3LYP y la base cc-pVDZ. Los perfiles de energía se muestran en conjunto en la figura 7.7. Estas gráficas muestran que los valores de la energía respecto al cambio en el ángulo de enlace son muy cercanos a lo predicho por el método basado en química cuántica.

Además de obtener el perfil de energía, la geometría de la molécula se optimizó, comparando los resultados del campo de fuerzas con los cálculos de estructura electrónica. La superposición de ambas geometrías se muestra en la figura 7.8. El valor de RMSD entre ambas es de tan solo 0.088 Å. La parte de la molécula que corresponde con el compuesto 2 sigue una geometría de equilibrio muy similar a la



Ángulo				
Nombre	k_θ (kcal mol rad ⁻²)	θ_{eq} (°)	Penalización	
CA-NC-CA	70.0	118.6	0.3	
CA-CA-NC	70.0	121.5	1.1	
CA-CA-N2	70.0	120.1	1.1	
Torsión impropia				
Nombre	V_n (kcal mol ⁻¹)	Fase (°)	Periodicidad	Penalización
CA-CT-N2-H	1.0	180.0	2.0	6.0
CA-CA-CA-HA	1.1	180.0	2.0	6.0

Figura 7.6: Parámetros para interacciones de enlace que no fueron encontradas en el campo de fuerzas para el aducto cisteína - compuesto 2 y se asignaron los valores por defecto de grupos generales, acompañados de su valor de penalización.

predicha con B3LYP. La principal diferencia entre las estructuras radica en la sección que forma parte de la cadena principal. No obstante, en este fragmento se agregó un grupo N-metilo y acetilo para simular el enlace peptídico, que en la proteína dependerá de la parametrización de la cadena principal de acuerdo con el campo de fuerzas ff14SB. La parametrización, por tanto, es satisfactoria para los propósitos de este trabajo.

De manera similar, el modulador negativo NNC0640 fue modelado con el campo GAFF. En este caso, todos los parámetros fueron encontrados y asignados correctamente, con excepción de algunos ángulos impropios, como se muestra en la figura 7.9. Los términos referentes a ángulos dihedros impropios suelen ser accesorios para forzar la planaridad de algunos sistemas como carbonilos o anillos aromáticos. Los valores genéricos suelen ser suficientes y son generalmente idénticos para una gran cantidad de tétradas que pertenecen a sistemas planos. De hecho, es el único conjunto de parámetros cuya ausencia es permitida en AMBER (40).

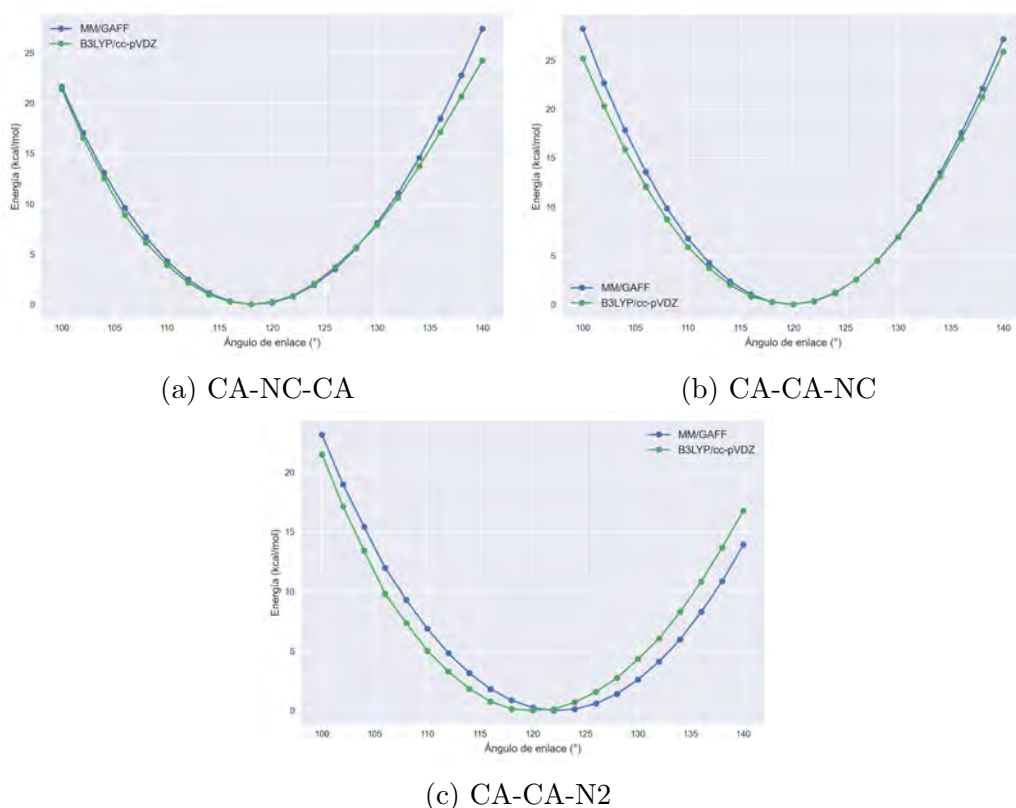


Figura 7.7: Perfiles de energía potencial en función de los ángulos de enlace, para los parámetros que fueron asignados con penalización. Se comparan los resultados de la parametrización con GAFF (azul) contra los resultados correspondientes a cálculos de estructura electrónica con el nivel de teoría B3LYP/cc-pVDZ

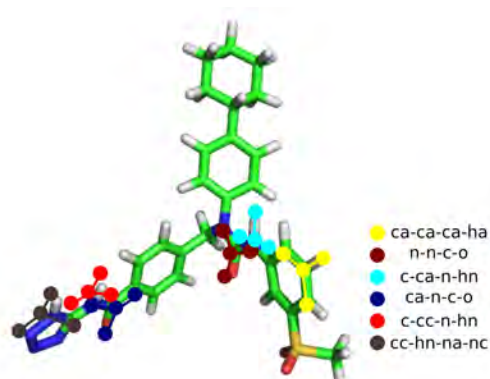
7.3.2. Hidratación del canal del receptor

El modelo 3D-RISM permite calcular la función de distribución radial sitio a sitio entre la proteína y el átomo de oxígeno de las moléculas de agua, utilizadas como disolvente. De manera análoga a una función de distribución radial, la función de distribución sitio a sitio es proporcional a la probabilidad de que los sitios de referencia del soluto y el disolvente se encuentren a una distancia r . En sistemas que no poseen simetría esférica, esta función es anisotrópica y puede calcularse al resolver de forma numérica la adaptación de las ecuaciones de Ornstein-Zernike, de acuerdo con el modelo 3D-RISM. En la figura 7.10a, se muestra una isosuperficie de la función de distribución sitio a sitio para el átomo de oxígeno de la molécula de agua alrededor del receptor GLP-1R.

El laplaciano de la función de distribución sitio a sitio fue calculado con el objetivo de identificar las zonas de máxima probabilidad para ubicar las moléculas de agua. La isosuperficie con estos resultados se muestra en la figura 7.10b. Dado que el laplaciano de una función es equivalente a la divergencia de su gradiente, los puntos cuyo valor es negativo corresponden con puntos hacia donde convergen los vectores del gradiente. Es decir, en estas zonas hay una concentración de la función de densidad. Como puede observarse en la figura 7.10b, las zonas de máxima probabilidad se distribuyen alrededor de la proteína y en sitios y cavidades a través del



Figura 7.8: Resultado de la optimización de la geometría molecular, utilizando la parametrización por GAFF (en azul) y mediante estructura electrónica con el nivel de teoría B3LYP y la base cc-pVDZ



Torsión impropia				
Nombre	V_n (kcal mol ⁻¹)	Fase (°)	Periodicidad	Penalización
ca-ca-ca-ha	1.1	180	2.0	6.0
n-n-c-o	10.5	180	2.0	6.0
ca-n-c-o	10.5	180	2.0	6.0
c-ca-n-hn	1.1	180	2.0	6.0
c-cc-n-hn	1.1	180	2.0	6.0
cc-hn-na-nc	1.1	180	2.0	31.9

Figura 7.9: Parámetros del campo de fuerzas que fueron asignados con valores generales por defecto, en ausencia de los parámetros particulares. La figura muestra la estructura del modulador y los ángulos dihedros impropios que corresponden con estos parámetros

canal del receptor. Normalmente, la llegada de moléculas de agua a estas zonas de difícil acceso puede verse entorpecida por la presencia de barreras energéticas que no puedan superarse fácilmente en simulaciones estándar. Por esta razón, la función de distribución es útil para posicionar las moléculas de agua como punto de partida,

en sitios donde la interacción sea favorable pero el acceso pueda ser complicado. Las moléculas de agua se colocaron en los centroides de las zonas mostradas en la figura 7.10b, manteniendo solamente aquellas ubicadas en el canal interno, puesto que en las zonas externas se encontrará el modelo de membrana.

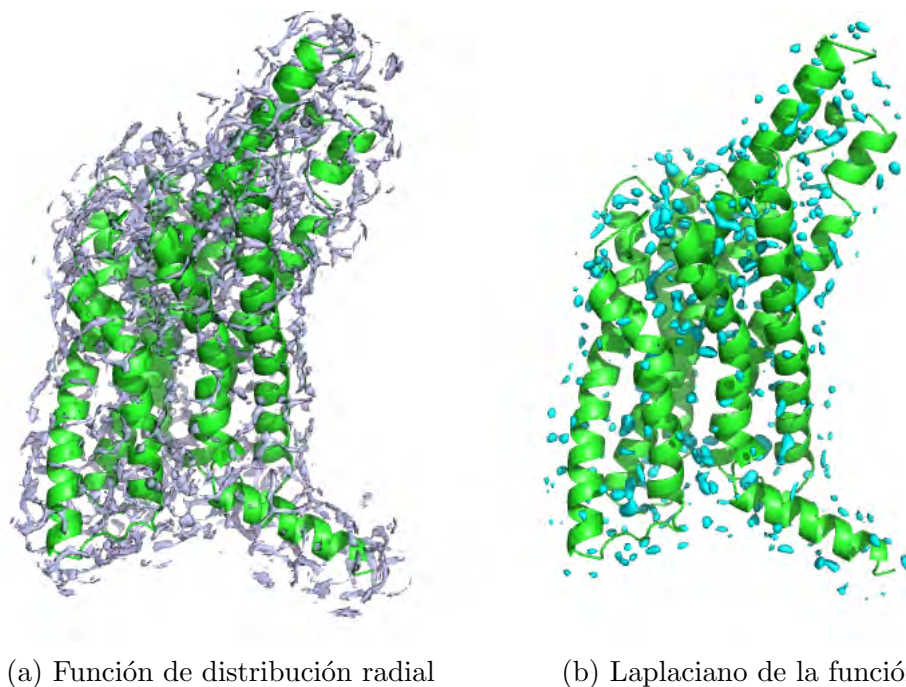


Figura 7.10: Resultados de la función de distribución radial sitio a sitio para el átomo de oxígeno de la molécula de agua alrededor del receptor, calculada mediante la teoría 3D-RISM.

7.4. Modelos de Estado de Markov

7.4.1. Extracción de variables

Los resultados del área accesible a disolvente en función del tiempo para la cavidad total y para cada uno de los sistemas, se muestran en la figura 7.11, considerando el promedio de las tres repeticiones. Estos datos muestran que la presencia de los moduladores puede modificar el grado de exposición de la cavidad de reconocimiento para la proteína G. Los sistemas en complejo con el compuesto 2, tanto en presencia como en ausencia del agonista, inducen un mayor grado de exposición, con valores promedio de aproximadamente 1114 \AA^2 . En ausencia de modulador, el GLP-1 puede generar una exposición de alrededor de 1059 \AA^2 en promedio, mientras que con el modulador negativo, el valor promedio del área disminuye hasta 985 \AA^2 . Como se esperaría, el compuesto 2 es capaz de promover la apertura de la cavidad de unión para la proteína G en presencia de agonista, mientras que el modulador negativo tiende a restringir la exposición de los aminoácidos involucrados en la interacción. Los moduladores que actúan por modificación covalente, con el mismo mecanismo que el compuesto 2, también tienen actividad agonista intrínseca débil. Esto puede

explicar la similitud de los valores del área observados para los dos sistemas con modulador positivo.

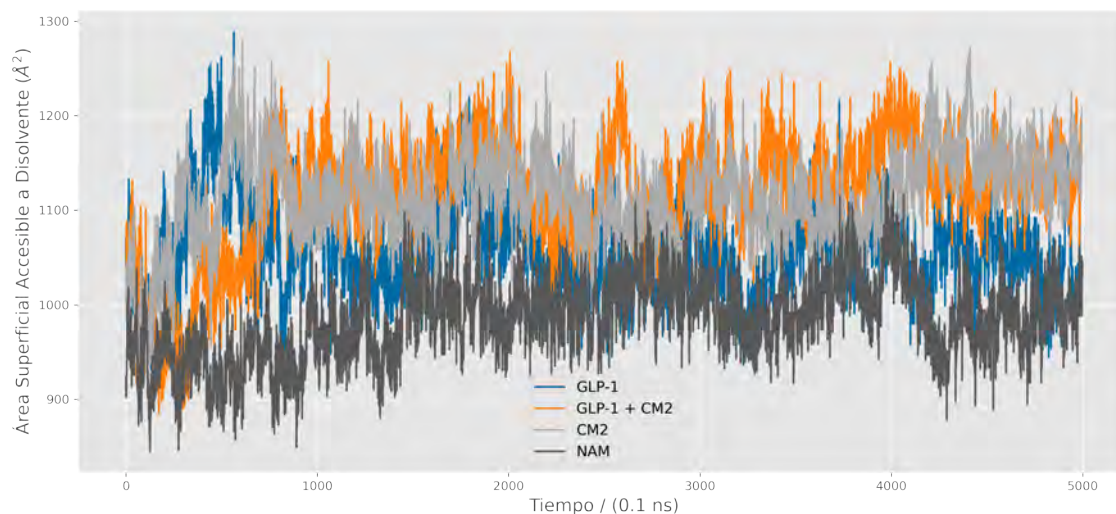


Figura 7.11: Área Superficial Accesible al Disolvente (SASA) promedio como función del tiempo de simulación para cada uno de los sistemas

Para determinar si los cambios en el área superficial exhiben un comportamiento estacionario o presentan alguna tendencia, se calcularon los correlogramas para cada sistema. Los valores de la función de autocorrelación estimada a partir de los datos del área como series de tiempo, se muestran en la figura 7.12. En todos los casos, hay una correlación persistente, aun con un desfase de 60 pasos, con un decaimiento paulatino. Este comportamiento es típico de procesos que siguen una tendencia en el tiempo (55). Es decir, la exposición de los aminoácidos participantes en la interfaz receptor – proteína G no permanece constante, sino que se incrementan con el tiempo. La tendencia al incremento se deduce por el valor positivo de todos los valores de la función de autocorrelación.

Puesto que la presencia de modulador influye sobre el grado de exposición de la cavidad de interés en las trayectorias observadas, se utilizaron los valores de área superficial por residuo de aminoácido para el modelado de Markov. La caracterización del comportamiento dinámico sobre esta cavidad también puede ser útil como objetivo de moduladores alostéricos con mecanismo similar a los exhibidos por la familia del compuesto 2. Los histogramas de los valores de área superficial por cada residuo de aminoácido involucrado se muestran en la figura 7.13. Nótese que estas gráficas no muestran los cambios en los valores de las variables con el tiempo.

7.4.2. Reducción de dimensiones

Las variables originales fueron proyectadas en un espacio de baja dimensionalidad mediante análisis de componentes independientes con desfase en el tiempo. De manera análoga al análisis de componentes principales, los componentes independientes son una combinación lineal de las variables iniciales, pero en lugar de describir los cambios de mayor varianza, representan las variables con cambios más

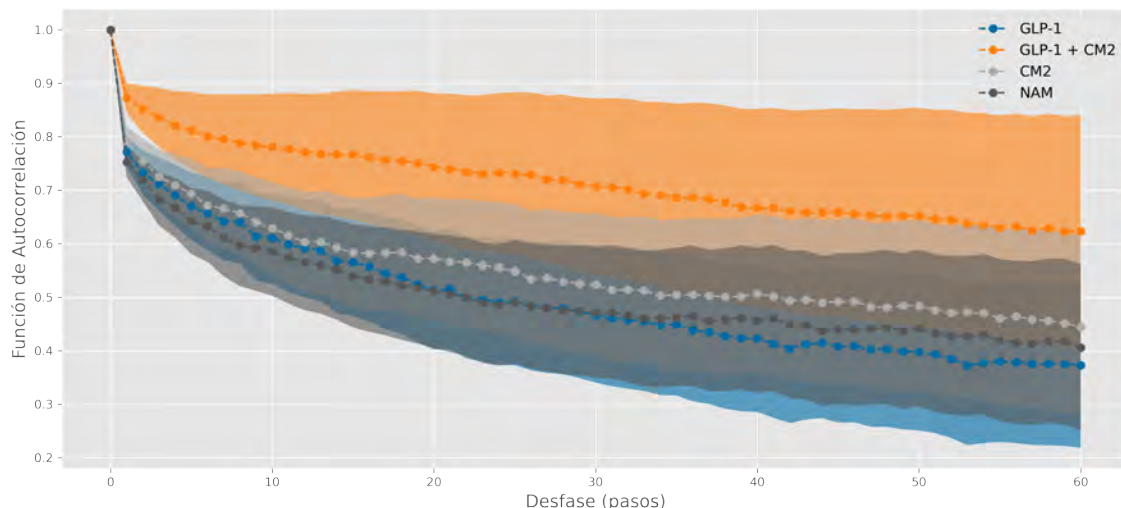


Figura 7.12: Correlogramas para las series de los valores de SASA en función del tiempo. Los marcadores unidos con líneas discontinuas indican el valor estimado de la función de autocorrelación y las áreas sombreadas representan los intervalos de confianza al 95 %.

lentos. Así mismo, los valores propios se relacionan con las escalas de tiempo, mediante la ecuación 6.2, implicadas en la relajación del proceso descrito por cada componente. Los hiperparámetros de este algoritmo son el tiempo de desfase y la cantidad de dimensiones a mantener. En la figura 7.14 se muestran los valores de las primeras 4 escalas de tiempo estimadas con distintos desfases. El valor apto para el desfase es el mínimo a partir del cual las escalas de tiempo tienden a permanecer aproximadamente constantes. En estas gráficas, los cambios son aproximadamente lineales a partir de entre los 20 ns y los 30 ns. Para los sistemas en presencia de GLP-1 y de GLP-1 + CM2, el desfase seleccionado fue 20 ns, mientras que para aquellos en presencia de CM2 y NAM fue de 30 ns. El resumen con todos los hiperparámetros de cada modelo se presentan en la tabla 7.1 más adelante.

Cada componente independiente describe un proceso dinámico asociado a la escala de tiempo relacionada con su valor propio. De esta manera, los componentes independientes con las escalas de tiempo menores al valor de desfase contienen información dinámica de procesos rápidos que no pueden resolverse de forma precisa. Utilizando este razonamiento, se seleccionaron los componentes cuyo valor asociado de escala de tiempo fuera mayor al desfase. Las gráficas de la figura 7.15 muestran los valores de las primeras 10 escalas de tiempo para las proyecciones TICA, con un valor de desfase fijo. Las líneas horizontales muestran el valor límite para los procesos que pueden resolverse con ese número de componentes. De esta forma, se mantuvieron 4 dimensiones para el sistema con GLP-1, 4 para el GLP-1 + CM2, 3 para el CM2 y 3 para el sistema con el NAM.

En la figura 7.16 se muestran las proyecciones sobre el espacio de los primeros componentes independientes. Los histogramas de las frecuencias marginales para cada componente (izquierda) muestran distintos máximos que desaparecen al disminuir la escala de tiempo involucrada. Éstos máximos representan los procesos dinámicos descritos por el componente y corresponden muy de cerca con los estados

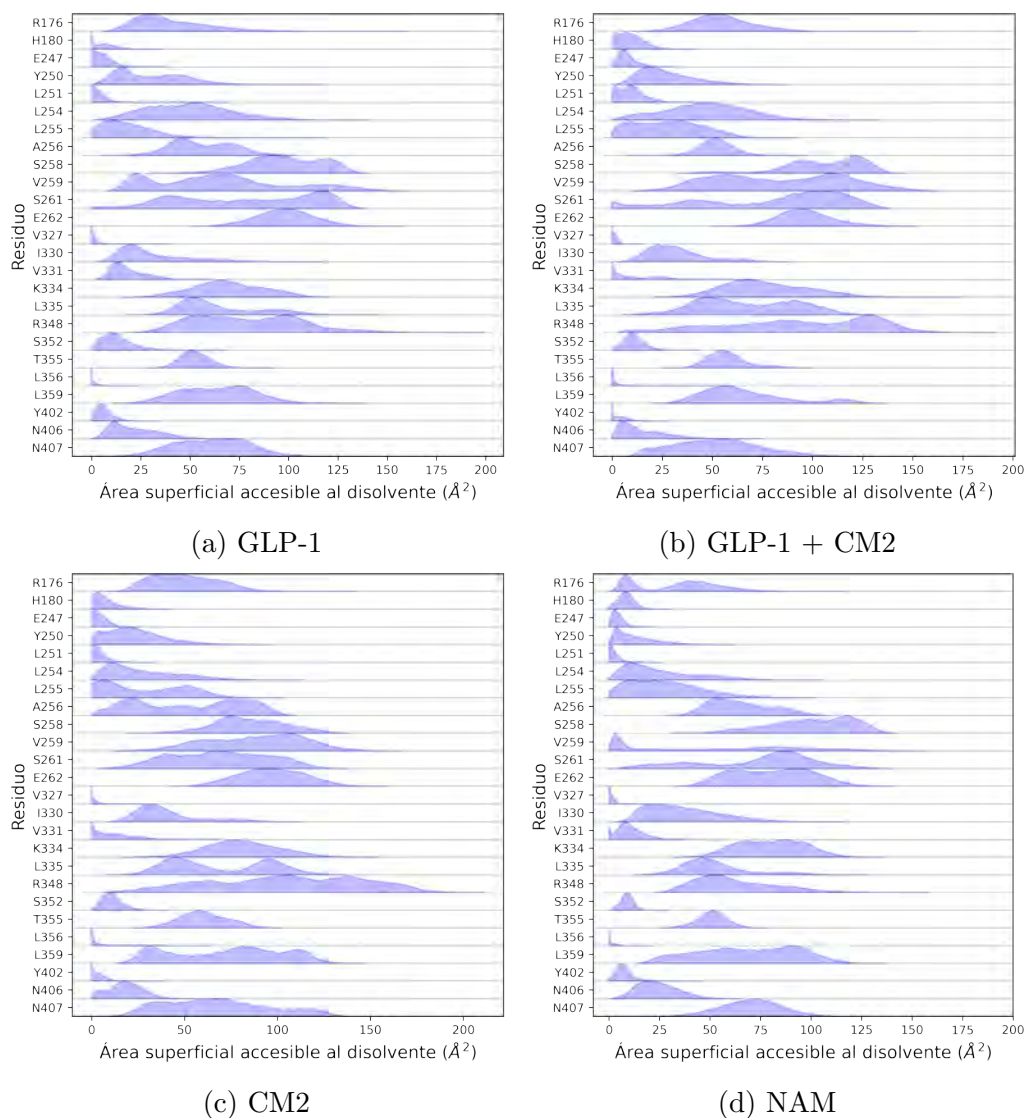


Figura 7.13: Histogramas del área superficial accesible al disolvente por aminoácido, para cada uno de los sistemas. Estas variables fueron utilizadas como descriptores para la generación de los modelos de Markov

metaestables del fenómeno, aunque su densidad no es necesariamente la del equilibrio. Las gráficas al centro y a la izquierda son las frecuencias conjuntas observadas para la proyección de los dos primeros componentes. Su interpretación es similar a los histogramas pero incluyen el comportamiento conjunto de las dos variables.

Los histogramas de la proyección para el sistema en presencia de modulador negativo muestran que hay un conjunto de estados sin conectar con el resto del muestreo. Esto indica que una de las trayectorias no tiene un traslape suficiente con el resto, es decir, ocupa estados que no fueron muestreados por las otras simulaciones. Si no hay traslape entre las trayectorias, la cadena de Markov que representa la evolución del sistema no es irreducible y, por lo tanto, no es posible definir una distribución estacionaria única. Ante la imposibilidad de estimar un modelo de Markov con la totalidad de los datos, para este sistema sólo se utilizaron las trayectorias intercomunicadas.

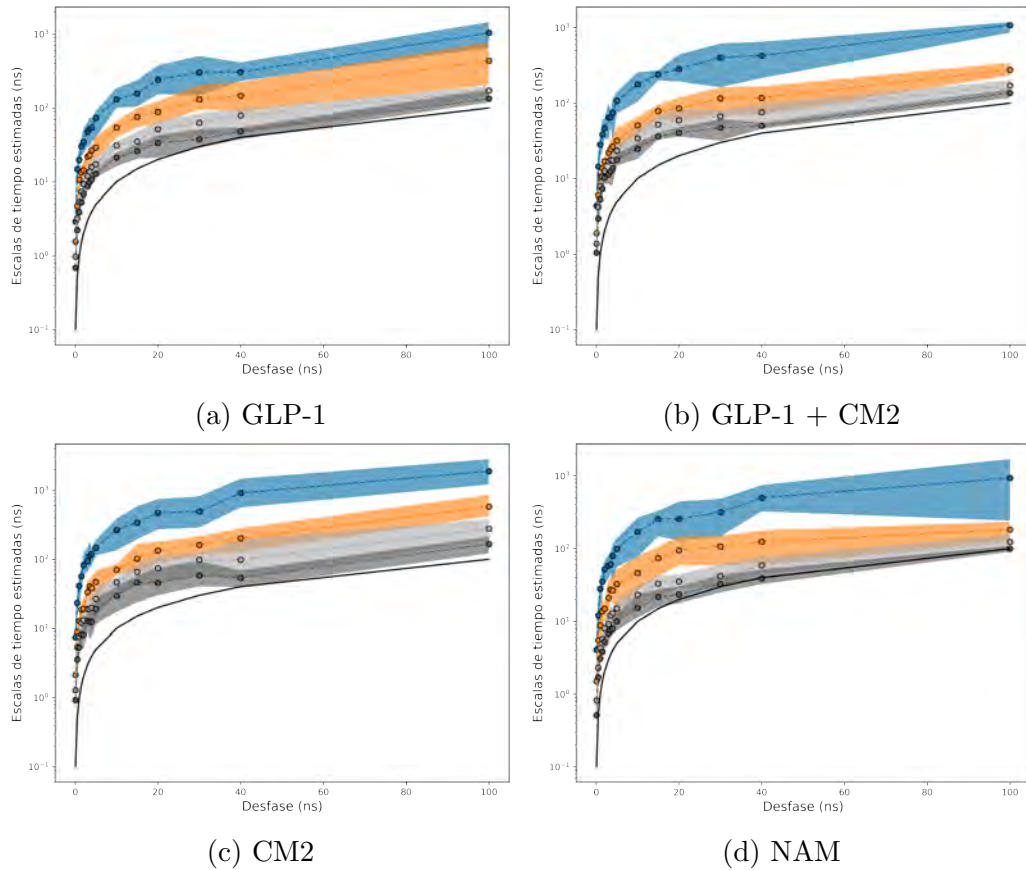


Figura 7.14: Escalas de tiempo de TICA en función del desfase. Las zonas sombreadas representan los intervalos de confianza al 95 %. La línea negra continua es el tiempo de desfase.

7.4.3. Generación de microestados, estimación y validación de los modelos de Markov

Los datos proyectados en el espacio de componentes independientes fueron agrupados en microestados mediante el algoritmo de clustering k -medias. El número de microestados (clústers) es determinante en la estimación del modelo porque a partir de las transiciones entre los clústers se calculan los parámetros de la matriz de transición. La figura 7.17 muestra el valor del puntaje VAMP-2 para diferente número de clústers en cada uno de los sistemas. El VAMP-2 se obtiene a partir del cociente de Rayleigh para la aproximación a los vectores propios de la matriz de transición que representa la evolución del sistema. Este cociente es máximo para los valores exactos de los vectores propios y, en ausencia de una solución analítica, la maximización de este cociente es una guía para optimizar el modelo. Por ende, se seleccionó el número de clústers para el cual se obtienen los valores más altos de este puntaje. En la figura 7.17, las líneas discontinuas delimitan el intervalo del número de microestados para los que el puntaje VAMP-2 es máximo. En general, se deduce que el número de microestados con un valor máximo de puntaje VAMP-2 es entre 100 y 200; más allá de este intervalo, el puntaje cae, lo que indica sobreajuste (33).

Las matrices de transición se calculan partiendo de los conteos observados de los

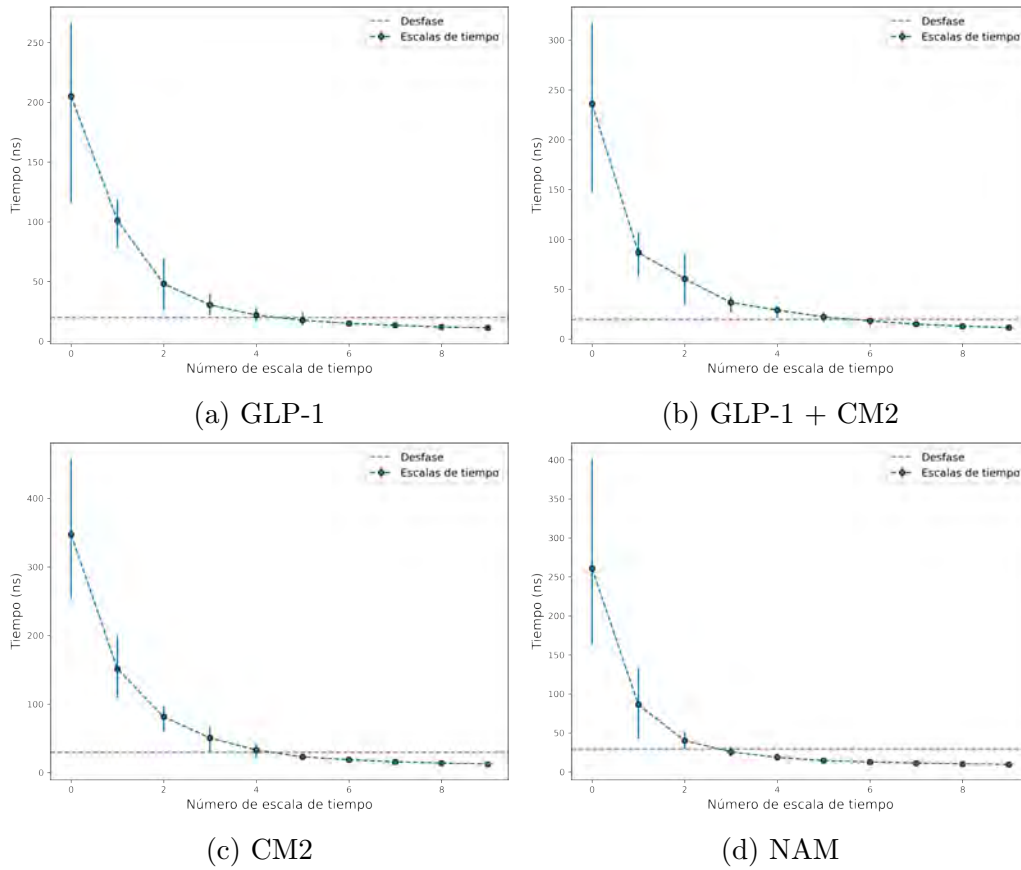


Figura 7.15: Valores para las primeras escalas de tiempo de cada proyección TICA con desfase fijo. La línea horizontal continua representa el valor del desfase. Las barras que acompañan a los marcadores representan los intervalos de confianza.

pasos entre cada microestado, luego de un desfase τ . En la figura 7.18 se muestran los valores de las escalas de tiempo estimadas para distintos valores del desfase. Las escalas de tiempo se relacionan con los valores propios de la matriz de transición de la misma forma que en el caso de la proyección TICA, de acuerdo con la ecuación 6.2, presentada en la sección de métodos. El valor óptimo del tiempo de desfase debe ser el mínimo a partir del cual las escalas de tiempo son aproximadamente constantes, pero lo suficientemente pequeños para conservar la propiedad de Markov. Para todos los sistemas, el cambio en las escalas de tiempo es menor a partir de aproximadamente los 15 ns. Para los sistemas en presencia de GLP-1 y CM2 únicamente, los valores de las escalas de tiempo más lentas siguen incrementándose al aumentar el desfase, lo que indica que el muestreo de algunos microestados puede estar subestimado. La calidad numérica de estos modelos puede mejorarse enriqueciendo las zonas poco visitadas por las trayectorias de estos sistemas. En la tabla 7.1 se presenta un resumen con los hiperparámetros seleccionados para generar los modelos de Markov.

Para verificar que las series de tiempo sigan la propiedad de Markov, se validaron los modelos generados mediante la prueba de Chapman-Kolmogorov y los resultados se muestran en la figura 7.19. En este método se compara la potencia n de la matriz de transición con el desfase original τ , contra un modelo estimado con un desfase $n\tau$. Si la propiedad de Markov se cumple, estas probabilidades deberían ser aproxi-

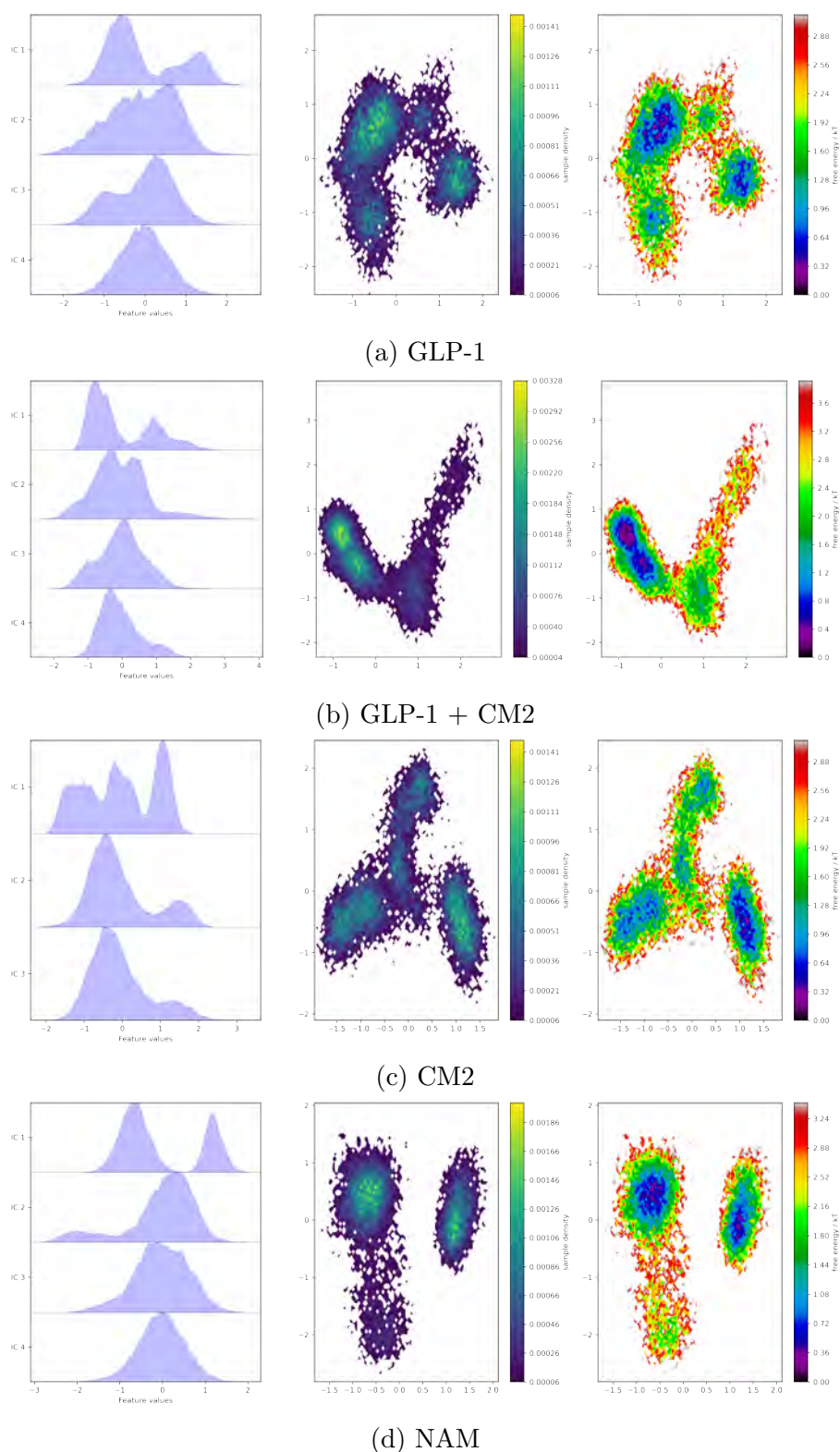


Figura 7.16: Proyección de los sistemas en los primeros componentes independientes: *Izquierda:* histogramas marginales de cada uno de los componentes mantenidos para el análisis. *Centro:* densidad de estados muestral, conjunta para los dos primeros componentes independientes. *Derecha:* diagramas de pseudo-energía libre para los dos primeros componentes independientes.

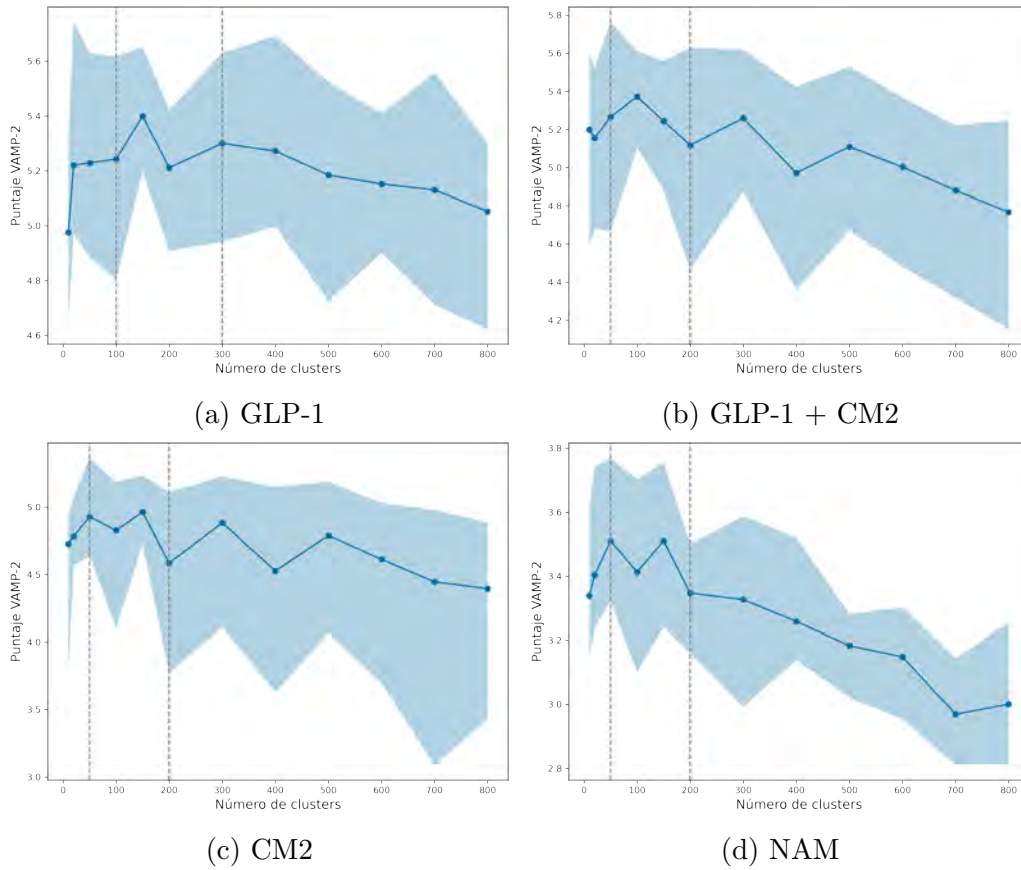


Figura 7.17: Puntaje VAMP-2 con respecto al número de microestados (clusters) en el espacio de componentes independientes. Las líneas discontinuas verticales indican las regiones donde el VAMP-2 es máximo.

Sistema	TICA		Clustering	MSM
	Desfase / ns	Dimensiones	k	Desfase / ns
GLP-1	20	4	150	15
GLP-1 + CM2	30	4	100	15
CM2	30	3	150	15
NAM	20	3	150	10

Tabla 7.1: Valores de los hiperparámetros seleccionados para generar los modelos de estado de Markov. TICA, análisis de componentes independientes con desfase en el tiempo; k, número de clústers; MSM, modelos de estado de Markov

madamente iguales, es decir, la potencia n de la matriz de transición debe describir la evolución del sistema n pasos adelante. Las gráficas de la figura 7.19 muestran las matrices correspondientes a las transiciones entre los estados metaestables. En estas gráficas, las líneas azules continuas representan la probabilidad predicha por la potencia n de la matriz de transición, mientras que las líneas negras discontinuas muestran las transiciones observadas. Puede observarse que los valores predichos y estimados tienen correspondencia estrecha y sólo hay algunas desviaciones menores

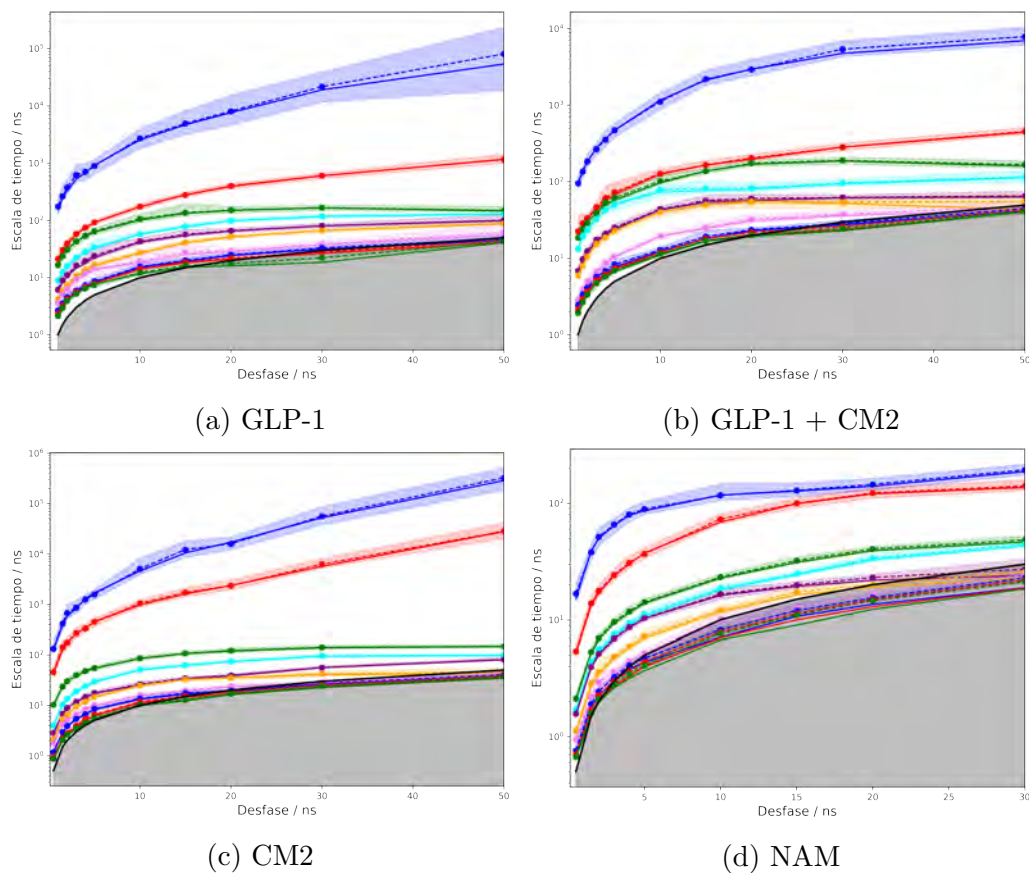


Figura 7.18: Valores de las escalas de tiempo estimadas a partir de las matrices de transición para distintos valores de desfase.

a valores largos de desfase. Esto puede ocurrir porque al incrementarse el valor de desfase, también se reduce la cantidad de datos disponibles para estimar la matriz de transición, dado que siempre se trabaja con trayectorias de longitud finita. Estos resultados indican que los sistemas cumplen con la propiedad de Markov con los valores de desfase seleccionados.

7.4.4. Análisis de los modelos

En el marco teórico se mostró que para una matriz de transición ergódica, sus vectores propios contienen la información cinética acerca del fenómeno descrito. El vector propio asociado con el valor propio de 1 contiene la distribución de probabilidad estacionaria o en el equilibrio, es decir, la proporción de veces en que el sistema visitará cada microestado en una simulación infinitamente larga corresponde con el valor de la probabilidad estacionaria en este vector. Las distribuciones estacionarias calculadas para la matriz de transición de cada sistema se muestran en la figura 7.20, proyectadas en los dos primeros componentes independientes. Los gráficos de la derecha corresponden con la distribución de probabilidad y los de la izquierda con los valores de energía libre. Estos diagramas muestran zonas de probabilidad máximas separadas por regiones de baja probabilidad y representa una aproximación al intercambio cinético entre estados metaestables esperado para una simulación

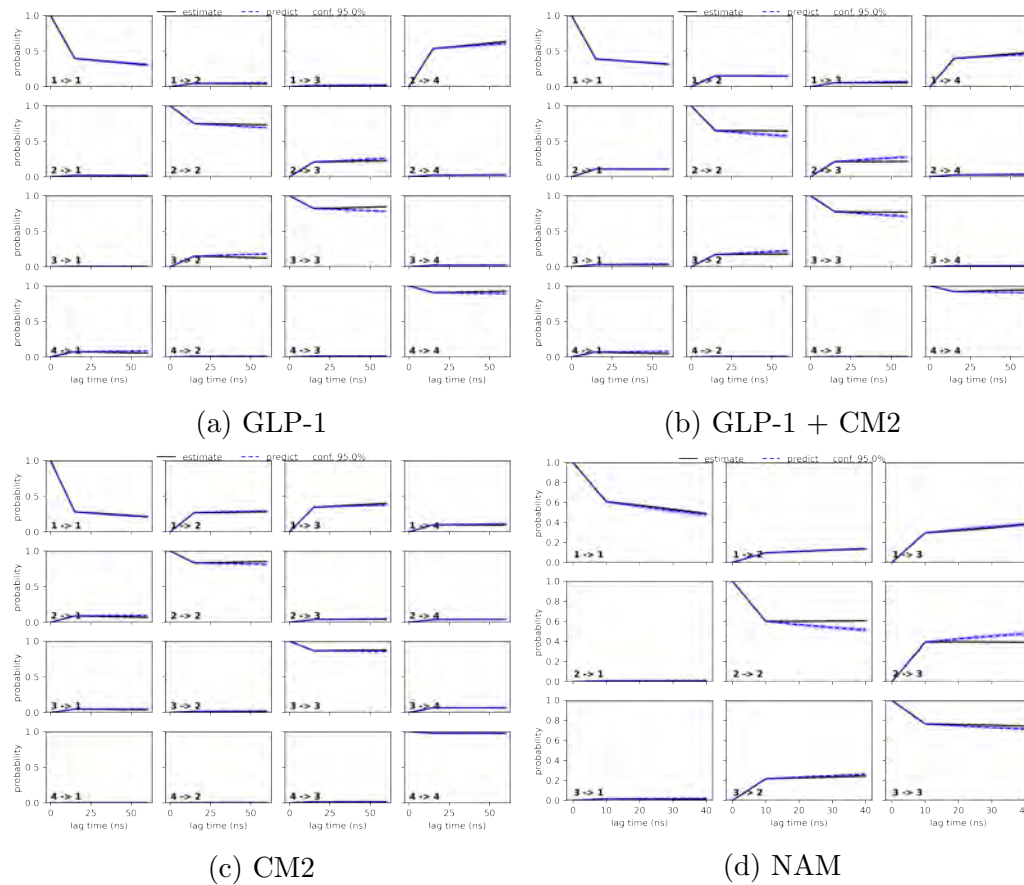
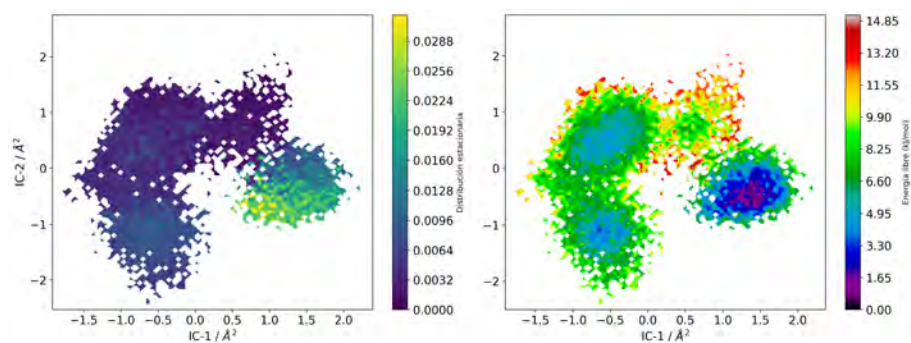


Figura 7.19: Resultados de las pruebas de Chapman-Kolmogorov para los modelos generados. La línea negra continua corresponde al valor estimado con el modelo a un desfase de $n\tau$ pasos, mientras la línea azul discontinua representa el valor predicho por la potencia n de la matriz de transición del modelo a prueba.

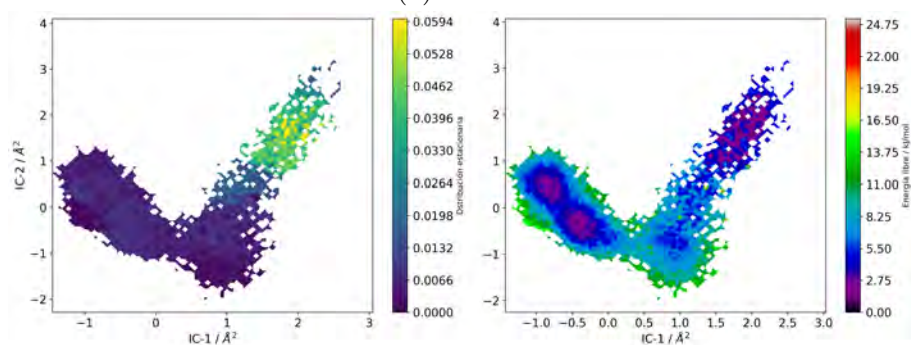
infinitamente larga.

El resto de los vectores propios contiene una descomposición de la evolución del sistema de acuerdo con los procesos más lentos. La figura 7.21 muestra los valores de los primeros 3 vectores propios después de la distribución estacionaria. En esta figura, las gráficas están ordenadas de manera que el proceso más lento se encuentra a la izquierda. El fenómeno descrito por cada vector es un transporte en el espacio de variables entre los puntos donde ocurre un cambio de signo. En todos los casos, el primer vector después de la distribución estacionaria divide el sistema en dos grupos principales cuyo intercambio es el fenómeno más lento observado en las simulaciones. Los siguientes vectores describen el transporte dentro de los grupos principales, que ocurren con mayor frecuencia.

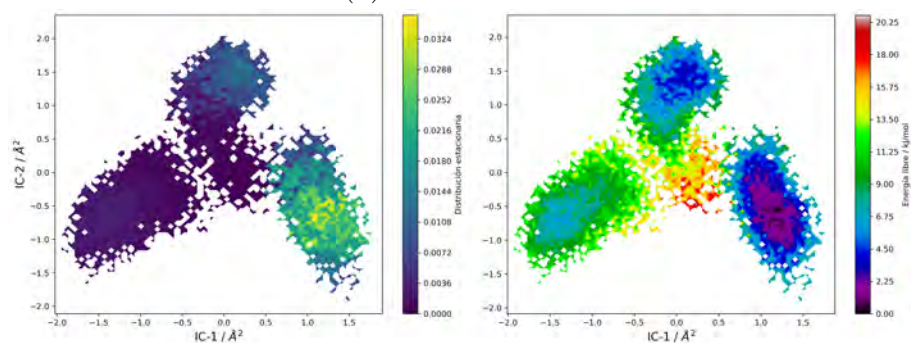
El agrupamiento en microestados utilizando k -medias genera conjuntos cuyos integrantes son similares geoméricamente, en este caso, con valores similares de área superficial expuesta. No obstante, la similitud geométrica entre estados no necesariamente está relacionada con la similitud cinética. Es decir, microestados separados con distancias geométricas similares no necesariamente se encuentran separados por barreras energéticas similares y, por tanto, su intercambio cinético tampoco es equi-



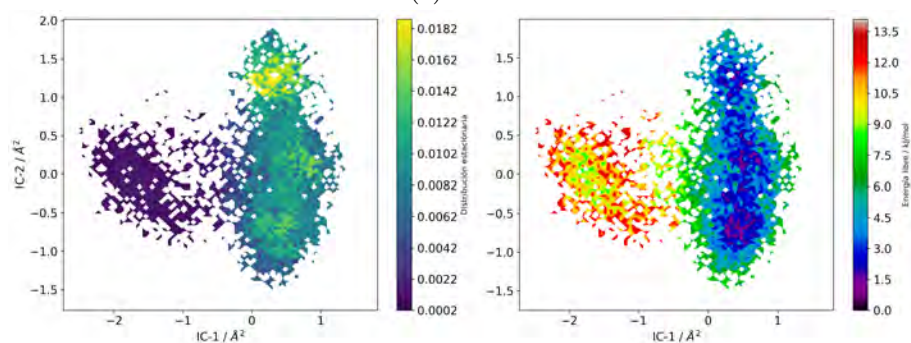
(a) GLP-1



(b) GLP-1 + CM2



(c) CM2



(d) NAM

Figura 7.20: Vector de distribución estacionaria para las matrices de transición estimadas. A la derecha se presentan las gráficas con los valores de probabilidad y en la izquierda se encuentran los mapas de energía libre.

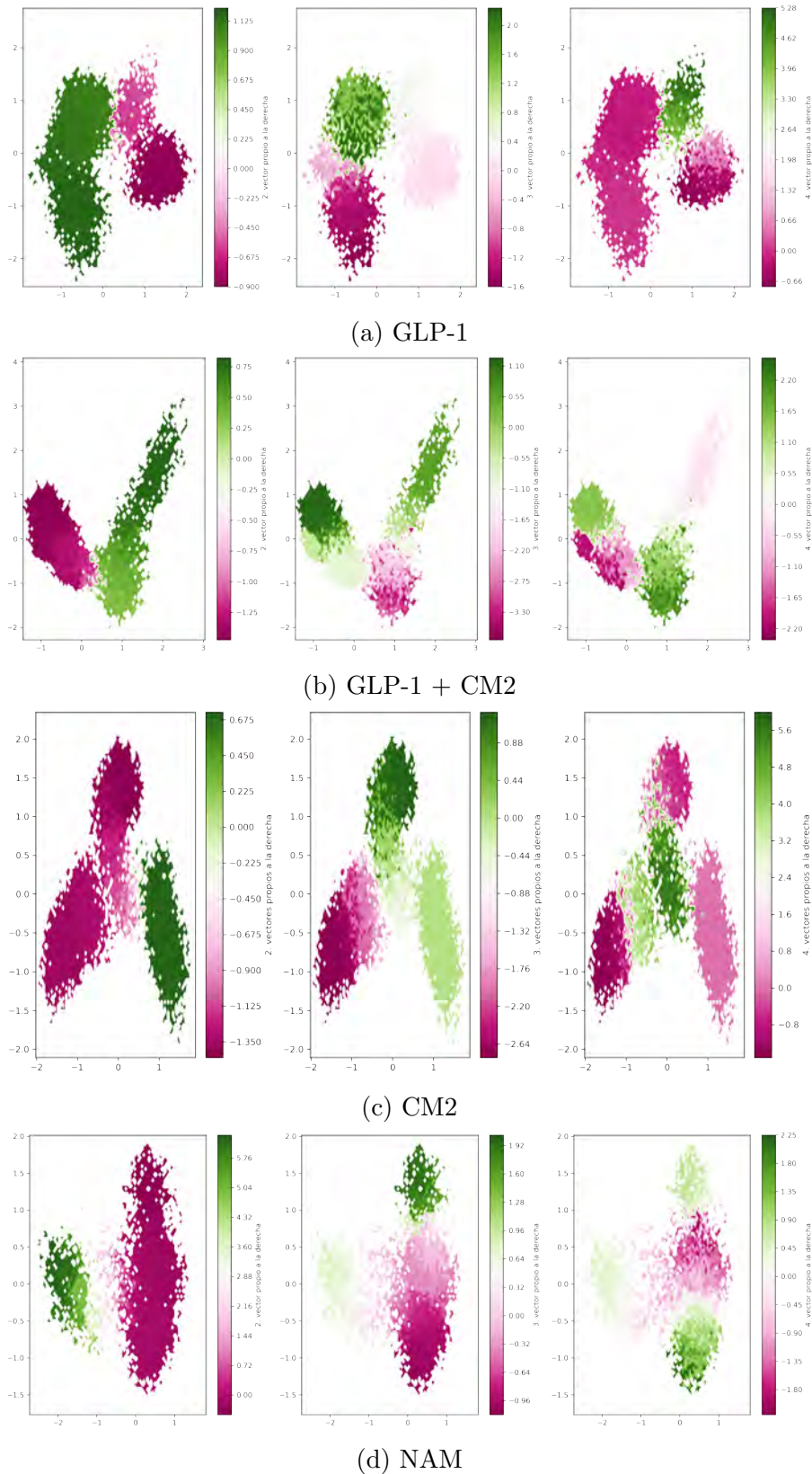


Figura 7.21: Primeros vectores propios de la matriz de transición, después de la distribución estacionaria. Cada vector representa los procesos más lentos en la simulación. El proceso descrito por cada vector es la transición entre los microestados donde ocurre un cambio de signo.

valente. En el contexto del modelado de estados de Markov, un *macroestado* es una agrupación de microestados con un rápido intercambio cinético y corresponde con los grupos metaestables de la cadena de Markov. De esta manera, los microestados que pertenecen a un mismo macroestado se intercambian frecuentemente en la simulación, mientras que una transición entre microestados en diferente macroestado es un evento raro o de poca frecuencia. Además, los macroestados pueden relacionarse con estados funcionales de la proteína. Los macroestados no corresponden con una estructura en particular, sino con un ensamble de estructuras geométrica y cinéticamente similares. Para identificar los macroestados en las simulaciones, se utilizó el análisis de clúster de Perron. Este algoritmo realiza un agrupamiento de los microestados sobre el espacio propio de la matriz de transición. Como resultado, asigna a cada microestado un vector cuyas componentes son las probabilidades de pertenecer a cada macroestado. No obstante, no existe un criterio para elegir de forma definitiva el número de macroestados en que pueden dividirse los datos. En este caso, se decidió utilizar un número de macroestados de acuerdo con el número de grupos en que los primeros tres vectores propios dividen el sistema, como se ilustra en la figura 7.21. De esta manera, los primeros tres sistemas se dividieron en 4 macroestados y el último en 3. Nótese que el primer vector propio divide el conjunto de datos en dos y los siguientes vuelven a dividir cada grupo en al menos otros dos subgrupos.

Las figuras 7.22 a 7.25 muestran los resultados del agrupamiento por PCCA++, donde el color de cada una de las gráficas es proporcional a la pertenencia de los microestados a cada macroestado. En la esquina superior izquierda de cada gráfica se muestra un identificador para cada macroestado. Estos resultados muestran una segregación completa del espacio de variables con un mínimo de traslape entre los macroestados, lo que refuerza la justificación del número seleccionado. En algunas fuentes, este tipo de agrupación se conoce como *representación en grano grueso* de la matriz de transición. Esta representación facilita la interpretación del modelo pero limita su descripción cuantitativa, de manera que el cálculo de cualquier propiedad debe realizarse a partir de la matriz de transición completa, para los microestados. La representación en grano grueso de la matriz de transición agrupa los cientos de microestados en pocos estados metaestables, con lo que la comparación entre ellos es más fácil de determinar. Así mismo, los microestados tienen diferencias estructurales más sutiles que son más difíciles de identificar. La agrupación en grano grueso permite reconocer fácilmente los cambios estructurales mayores e identificar sus relaciones cinéticas, como las transformaciones más lentas entre una conformación y otra. Esta caracterización permite encontrar los valles en la superficie de energía libre del sistema que finalmente permite describir su comportamiento cinético y termodinámico.

En las gráficas de pertenencia, todos los microestados sombreados dentro de un mismo grupo se intercambian con facilidad durante la simulación. Por su parte, aquellos estados separados tardan un mayor tiempo en interconvertirse y, por tanto, son las etapas determinantes en la evolución del sistema. A partir de la representación en grano grueso del modelo de Markov, es posible aproximar las distribuciones en el equilibrio de los macroestados y sus respectivos valores relativos de energía libre. Las tablas 7.2 a 7.5 muestran estos resultados para cada sistema.

Si bien los macroestados de los diferentes sistemas no son necesariamente equi-

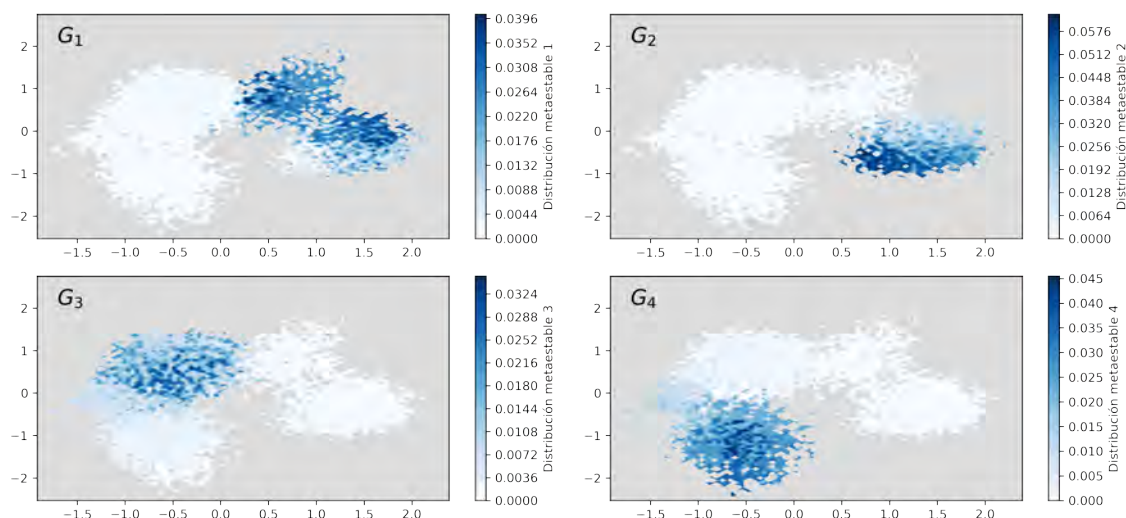


Figura 7.22: Valores de pertenencia de los microestados obtenida mediante PC-CA++ para el sistema en presencia de GLP-1. Cada gráfica representa uno de los macroestados y el color es proporcional a la probabilidad de que cada microestado se encuentre en éste. Los identificadores de los macroestados se muestran en la esquina superior izquierda de cada gráfica.

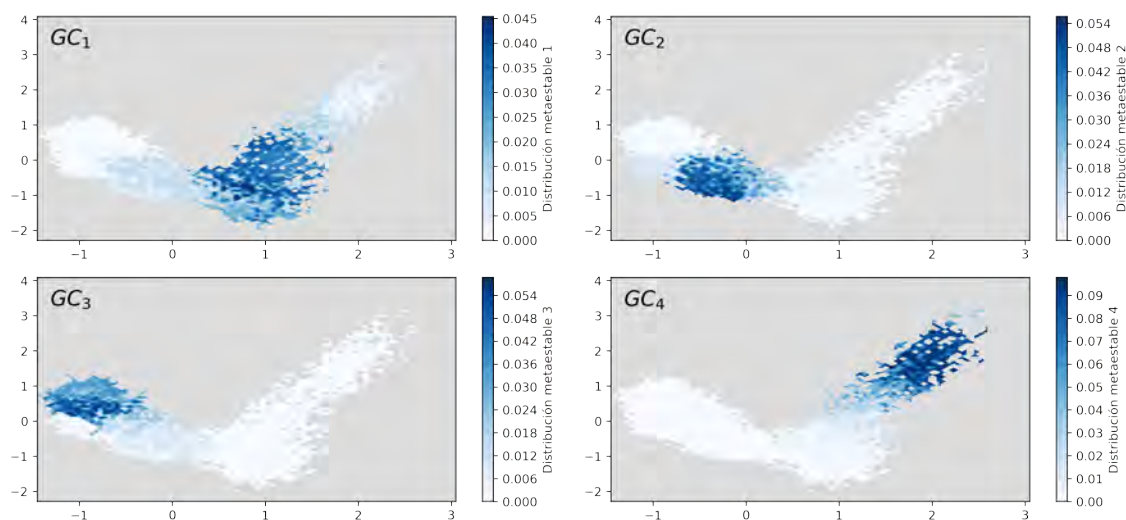


Figura 7.23: Valores de pertenencia de los microestados obtenida mediante PC-CA++ para el sistema en presencia de GLP-1 + CM2. Cada gráfica representa uno de los macroestados y el color es proporcional a la probabilidad de que cada microestado se encuentre en éste. Los identificadores de los macroestados se muestran en la esquina superior izquierda de cada gráfica.

valentes, su evolución se enfrenta con barreras energéticas de distinta altura. En ausencia de moduladores, la diferencia de energía más alta es de alrededor de 7 kJ/mol, mientras que en presencia de modulador positivo, la barrera de energía más alta que logró observarse fue de aproximadamente 4.5 kJ/mol. El sistema con modulador positivo y en ausencia de agonista se enfrenta a barreras energéticas de orden similar al sistema con agonista y sin modulador. Por su parte, en presencia de modulador negativo, la diferencia de energía más alta es alrededor de 8 kJ/mol, aunque no parece

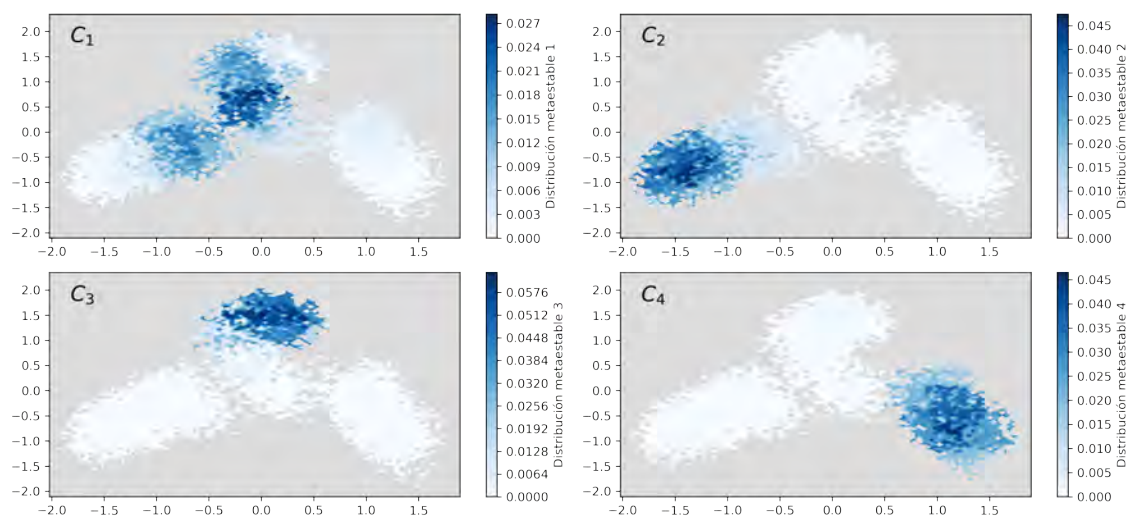


Figura 7.24: Valores de pertenencia de los microestados obtenida mediante PC-CA++ para el sistema en presencia de CM2. Cada gráfica representa uno de los macroestados y el color es proporcional a la probabilidad de que cada microestado se encuentre en éste. Los identificadores de los macroestados se muestran en la esquina superior izquierda de cada gráfica.

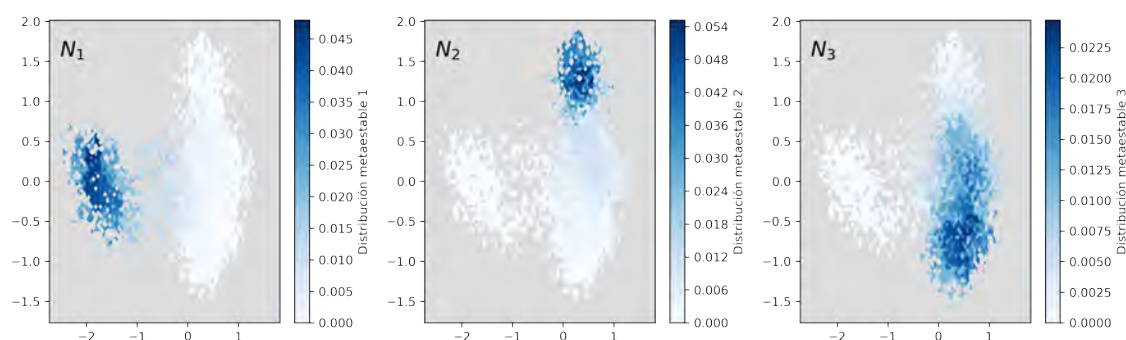


Figura 7.25: Valores de pertenencia de los microestados obtenida mediante PC-CA++ para el sistema en presencia de NAM. Cada gráfica representa uno de los macroestados y el color es proporcional a la probabilidad de que cada microestado se encuentre en éste. Los identificadores de los macroestados se muestran en la esquina superior izquierda de cada gráfica.

extenderse sobre el espacio de conformaciones con la misma amplitud de los sistemas anteriores. Estos resultados sugieren que la presencia de modulador positivo y de agonista favorecen que el receptor evolucione más fácilmente hacia conformaciones distintas a la inicial, por la disminución en las barreras de alta energía que se oponen a estas transiciones.

GLP-1				
Estado	Probabilidad	IC 95 %	G_i kJmol ⁻¹	IC 95 %
G_1	0.03	± 0.02	9.16	± 1.03
G_2	0.46	± 0.33	2.20	± 0.98
G_3	0.23	± 0.14	3.95	± 0.80
G_4	0.34	± 0.19	2.89	± 0.73

Tabla 7.2: Valores calculados de probabilidad estacionaria y energía libre relativa entre los macroestados de acuerdo con la representación en grano grueso del sistema con GLP-1. G_i , energía libre para el estado i ; IC 95 %, intervalos de confianza al 95 %

GLP-1 + CM2				
Estado	Probabilidad	IC 95 %	G_i kJmol ⁻¹	IC 95 %
GC_1	0.10	± 0.07	6.13	± 0.93
GC_2	0.20	± 0.14	4.34	± 0.94
GC_3	0.19	± 0.11	4.46	± 0.76
GC_4	0.57	± 0.36	1.61	± 0.84

Tabla 7.3: Valores calculados de probabilidad estacionaria y energía libre relativa entre los macroestados de acuerdo con la representación en grano grueso del sistema con GLP-1 y CM2. G_i , energía libre para el estado i ; IC 95 %, intervalos de confianza al 95 %

Otro aspecto interesante de la distribución de macroestados es su relación cinética, es decir, cómo y con qué velocidad ocurren las transiciones entre ellos. En el sistema con GLP-1, el estado de mayor energía es el G_1 y su vector de pertenencia se presenta en la figura 7.22. Este estado es un punto de comunicación entre los macroestados con mayor probabilidad estacionaria. De manera análoga, los estados GC_1 y C_1 son los de mayor energía en los sistemas en presencia de GLP-1 + CM2 y de CM2, respectivamente. Los vectores de pertenencia para estos estados muestran que se concentran entre los de mayor probabilidad. Estas observaciones sugieren que G_1 , GC_1 y C_1 son estados de transición para cada sistema, pues representan las principales barreras energéticas en el intercambio entre las conformaciones metaestables observadas.

Con la identificación de los estados metaestables, es de utilidad conocer las relaciones dinámicas entre ellos. Los flujos sobre el espacio de variables entre los estados

CM2				
Estado	Probabilidad	IC 95 %	G_i kJmol ⁻¹	IC 95 %
C ₁	0.02	± 0.01	10.54	± 0.97
C ₂	0.12	± 0.03	5.66	± 0.75
C ₃	0.25	± 0.10	3.76	± 1.08
C ₄	0.65	± 0.15	1.18	± 0.62

Tabla 7.4: Valores calculados de probabilidad estacionaria y energía libre relativa entre los macroestados de acuerdo con la representación en grano grueso del sistema con CM2. G_i , energía libre para el estado i ; IC 95 %, intervalos de confianza al 95 %

NAM				
Estado	Probabilidad	IC 95 %	G_i kJmol ⁻¹	IC 95 %
N ₁	0.04	± 0.02	8.68	± 1.10
N ₂	0.26	± 0.07	3.62	± 0.75
N ₃	0.74	± 0.18	0.86	± 0.65

Tabla 7.5: Valores calculados de probabilidad estacionaria y energía libre relativa entre los macroestados de acuerdo con la representación en grano grueso del sistema con NAM. G_i , energía libre para el estado i ; IC 95 %, intervalos de confianza al 95 %

de intercambio más lento fueron calculados mediante los algoritmos derivados de la teoría de rutas de transición. En este análisis, se calcularon las trayectorias reactivas entre los macroestados cuya interconversión es la más lenta, de acuerdo con lo predicho por el primer vector propio a la derecha de la matriz de transición de la figura 7.21. Las trayectorias reactivas más probables de cada sistema se muestran en la figura 7.26, en conjunto con los valores de su probabilidad de compromiso, definida por la posición en x de los marcadores en estas gráficas.

En la figura 7.26, el tamaño de los nodos es proporcional a la probabilidad del estado respectivo en el equilibrio, mientras que los números que acompañan a las flechas representan la frecuencia esperada de las transiciones entre los estados conectados, por unidad de tiempo. En presencia de GLP-1, los estados de interconversión más lenta son G_4 y G_2 , siendo G_2 el estado de menor energía. El mayor flujo entre ellos ocurre utilizando a los estados G_3 y G_1 como intermedios, aunque la trayectoria directa también tiene un peso importante. La posición en x de los nodos indica el valor de la probabilidad de compromiso de los estados intermedios. De esta forma, G_3 tiene un mayor grado de compromiso para evolucionar a G_4 y G_1 para llegar a G_2 . En este caso, G_1 es el estado de menor probabilidad en el equilibrio, correspondiendo con una barrera de energía importante. La observación del sistema en el estado G_1 es, por tanto, un evento raro y toda vez que se alcanza, es muy probable que evolucione hacia G_2 .

Por su parte, en el sistema del receptor en presencia de modulador positivo y

GLP-1, el flujo más frecuente entre los estados $GC_3 \rightarrow GC_4$ ocurre a través de GC_2 y GC_1 . Nuevamente, GC_1 respresenta un estado de alta energía pero forma parte importante de la trayectoria reactiva entre los estados de interconversión más lenta. Es decir, dado su bajo valor de probabilidad en el equilibrio, el sistema tarda en visitar este estado, pero al llegar a él puede desembocar rápidamente hacia el estado de mayor probabilidad, GC_4 . En ausencia de GLP-1, pero estando presente el modulador positivo, se observaron dos rutas alternas entre los estados de interconversión lenta $C_2 \rightarrow C_3$, una a través de C_4 y otra a través de C_1 como intermedios. No obstante, la trayectoria más frecuente es la que pasa por C_1 . La probabilidad en el equilibrio de C_4 es intermedia entre los estados C_2 y C_3 , por lo que representa también un estado de baja energía. Cuando el sistema alcanza C_4 , tiende a permanecer en él, más que evolucionar hacia uno u otro estado. Por el contrario, C_1 es de baja probabilidad y, por tanto, de alta energía; al llegar a este estado, el sistema evoluciona rápidamente hacia un estado metaestable de menor energía y por ello la mayor parte del flujo ocurre a través de él.

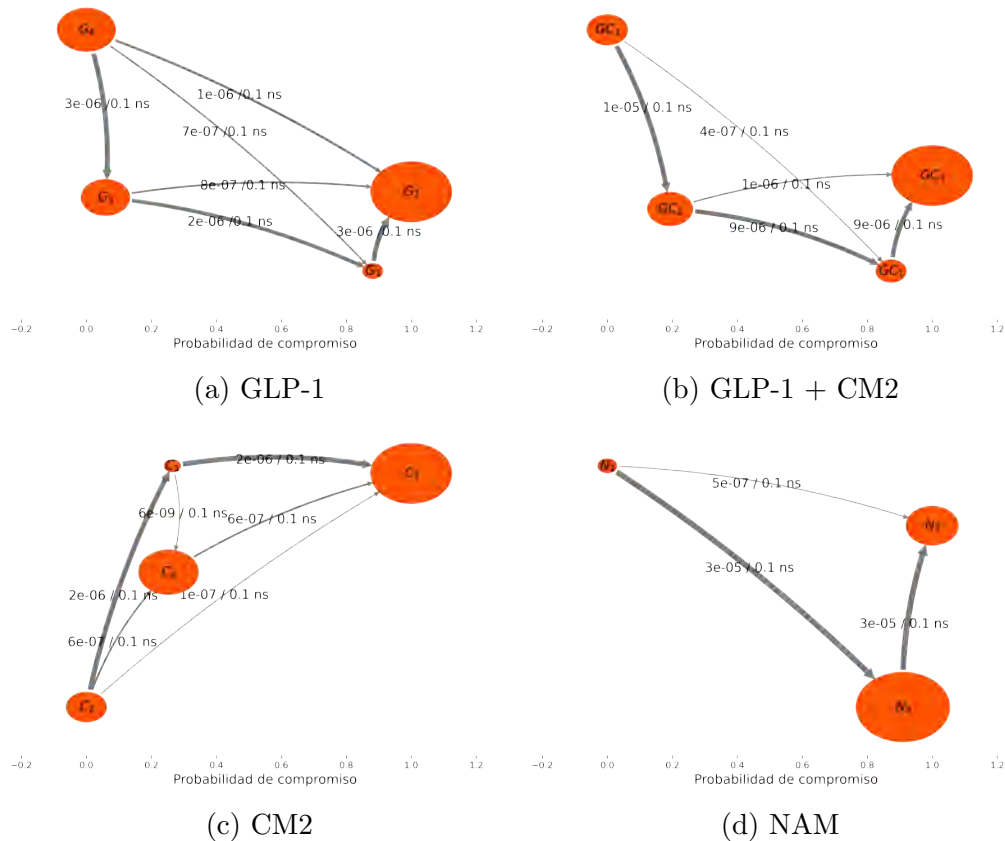


Figura 7.26: Trayectorias reactivas preferentes entre los macroestados de interconversión más lenta. El tamaño de marcador es proporcional a la distribución estacionaria del macroestado, mientras el tamaño de las flechas y sus etiquetas muestran la frecuencia esperada de la transición. Las posiciones en x representan los valores de la función de compromiso de cada estado intermedio, los valores en y no tienen un significado particular.

Finalmente, en presencia de modulador negativo, el sistema evoluciona muy rápidamente desde N_1 hacia los estados de menor energía N_2 y N_3 . En este caso, aunque

N_1 también es un estado de muy baja probabilidad como en los sistemas anteriores, no comunica a los estados de más baja energía, sino que representa una barrera energética que el sistema no puede superar y decae hacia un estado de menor energía donde tiende a permanecer por el resto de la simulación. Esto sugiere que el modulador negativo tiende a capturar al receptor en un estado donde la exposición de los residuos de aminoácido permanecen casi constantes, formando un impedimento para la apertura de la cavidad reclutadora de la proteína G.

Los resultados del análisis de Markov permitieron descomponer las trayectorias observadas en las simulaciones en un proceso estocástico de saltos entre estados discretos, sobre un espacio de baja dimensionalidad. La agrupación en estados metaestables y el análisis de trayectorias reactivas indican que la mayor parte del flujo en el sistema ocurre a través de los estados de probabilidad más baja. De esta manera, los estados G_1 , GC_1 y C_1 son de alta energía, alcanzarlos representa la etapa determinante en los procesos y, por tanto, pueden ser agrupaciones de estados de transición que comunican los valles de energía libre observados en las simulaciones. No obstante, dado que la proyección de componentes independientes se realizó para cada sistema por separado, las agrupaciones no corresponden necesariamente con estados similares entre los sistemas. En la siguiente sección analizaremos de manera cualitativa los resultados generados por los estados metaestables y sus relaciones cinéticas, desde la perspectiva de los cambios estructurales y en la interacción con el ligando relacionados con las transiciones observadas.

Discusión

8.1. Comparación entre los subespacios de componentes independientes

Aunque los sistemas analizados cumplen con la propiedad de Markov y se lograron calcular matrices de transición que describen su evolución en el tiempo, los subespacios donde se proyectaron las trayectorias fueron distintos para cada caso. De esta forma, un punto en el espacio de variables para el sistema GLP-1 no es equivalente al mismo punto en el espacio proyectado para el GLP-1 + CM2. Para cuantificar la similitud o grado de traslape entre las proyecciones, se calculó la raíz cuadrada de la suma de productos internos al cuadrado (Root Mean Squared Inner Product, RMSIP) entre los componentes independientes, es decir, entre los vectores propios de acuerdo con la ecuación 6.1, que define el método TICA. La expresión para el RMSIP se muestra en la ecuación 8.1, donde $\mathbf{u}_i \cdot \mathbf{v}_j$ indica la operación producto interno entre dos vectores. El traslape máximo se obtiene cuando $\mathbf{u}_i = \mathbf{v}_j$, es decir, cuando los dos subespacios son iguales. Por el contrario, si el producto interno es de 0, los vectores son ortogonales, indicando que no hay traslape entre ellos. En la figura 8.1 se muestra la matriz con los valores de RMSIP entre los sistemas analizados, normalizados a un traslape máximo de 1. De acuerdo con esta matriz, los sistemas que tienen un mayor traslape mutuo son CM2 y NAM, por lo que las coordenadas sobre estos espacios identifican estados que son más similares entre sí, en comparación del resto. GLP-1 + CM2 tiene una similitud muy baja con el resto, por lo que la participación de los residuos de aminoácido sobre los fenómenos descritos por cada componente independiente son distintas a los demás sistemas.

$$RMSIP_{u,v} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \mathbf{u}_i \cdot \mathbf{v}_j} \quad (8.1)$$

La tabla 8.1 muestra los residuos de aminoácido que tienen mayor contribución a los primeros 4 componentes independientes para cada uno de los análisis. Los residuos de aminoácido que más influencia tienen sobre los procesos descritos en los modelos de Markov son los residuos polares GLU247^{3.50b}, HIS180^{2.50b}, SER352^{6.41b} y los hidrofóbicos LEU356^{6.45b}, VAL327^{5.57b} y VAL331^{5.61b}. Los primeros forman parte importante de la red polar intracelular, cuyas interacciones representan una barrera energética contra el paso del receptor hacia una conformación activa. Los

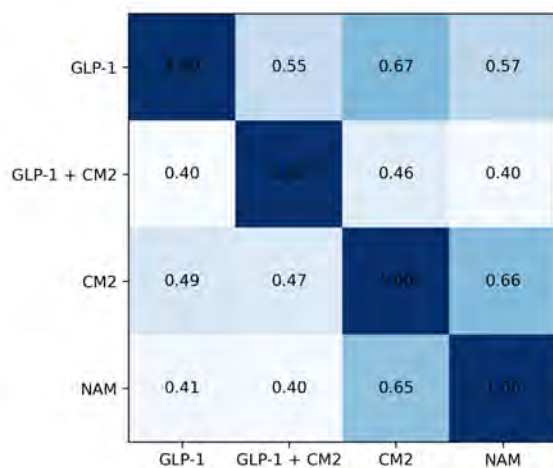


Figura 8.1: Matriz de RMSIP entre las proyecciones de componentes independientes para cada uno de los sistemas analizados. Los valores de RMSIP están normalizados a que el valor de traslape máximo sea de 1

residuos hidrofóbicos, por su parte, son cercanos al punto de unión del compuesto 2 e interaccionan de manera estrecha con este ligando. El hecho de que estos residuos tengan alto peso sobre los primeros componentes independientes implica que son protagónicos en el proceso dinámico observado en las simulaciones. El modulador positivo induce cambios importantes sobre la red polar intracelular y podría desestabilizarla para lograr un primer paso en la evolución hacia el estado activo. Para verificar esto, se tomó una muestra de las estructuras representativas de los estados metaestables y se identificaron los contactos de los residuos involucrados en la tabla 8.1. Las estructuras representativas se muestrean a partir de los valores de pertenencia calculados por PCCA++, es decir, se seleccionan las estructuras con mayor valor de pertenencia para cada estado metaestable. El valor del área superficial accesible al disolvente, además de indicar el grado de exposición de un residuo de aminoácido, se correlaciona con el número de contactos que ejerce. Es decir, un residuo de aminoácido tiene bajo nivel de exposición (un área superficial más pequeña) cuantas más interacciones forme con sus vecinos. Por ende, los modelos de Markov describen, de manera indirecta, la evolución de los contactos e interacciones en las que participan los residuos seleccionados para su definición.

Sistema	IC 1	IC 2	IC 3	IC 4
GLP-1	GLU247 ^{3.50b}	GLU247 ^{3.50b}	VAL327 ^{5.57b}	VAL327 ^{5.57b}
GLP-1 + CM2	VAL331 ^{5.61b}	LEU356 ^{6.45b}	VAL331 ^{5.61b}	LEU356 ^{6.45b}
CM2	HIS180 ^{2.50b}	GLU247 ^{3.50b}	VAL327 ^{5.57}	GLU247 ^{3.50b}
NAM	SER352 ^{6.41b}	GLU247 ^{3.50b}	GLU247 ^{3.50b}	LEU356 ^{6.45b}

Tabla 8.1: Residuos de aminoácido con la mayor contribución sobre los procesos descritos por cada uno de los primeros componentes independientes

8.2. Comportamiento de la red polar en ausencia de modulador

La figura 8.2 muestra una serie de estructuras representativas de los estados metaestables del receptor en ausencia de modulador, con énfasis en los residuos de aminoácido que participan en la red polar intracelular. Los dos estados de más baja energía y de intercambio más lento fueron G_2 y G_4 . En G_4 , los residuos E247^{3.50b}, Y402^{7.57b} y R176^{2.46b} forman una triada estabilizada por un enlace de hidrógeno E-COO⁻-HO-Y y un puente salino R⁺/E⁻, manteniendo en contacto a las hélices TM2, TM3 y TM7. Hacia la superficie del medio intracelular, el candado iónico entre los residuos R348^{6.37b} y E408^{8.49b} mantiene unidas a las hélices TM6 y TM7. Se ha propuesto que este candado iónico es una de las interacciones importantes para mantener a la hélice TM6 en posición inactiva. Este estado también es el de mayor similitud con los contactos nativos observados en la estructura cristalográfica.

En el estado de mayor probabilidad, G_2 , el residuo R176^{2.46b} pierde su interacción con E247^{3.50b} y adquiere una conformación expuesta hacia la superficie intracelular. Este cambio ocasiona una mayor exposición de los residuos E247^{3.50b}, S352^{6.41b} e H180^{2.50b}. Además, en este estado, R176^{2.46b} forma una triada iónica con el par E408^{8.49b} / R348^{6.37b}, que también está presente en el estado G_3 de probabilidad intermedia. El estado G_1 , de baja probabilidad y por ende, de alta energía, implica una separación entre los residuos cargados R176^{2.46b}, R348^{6.37b} y E408^{8.49b}. La baja probabilidad estacionaria de G_1 y el valor de su probabilidad de compromiso indican que este estado se visita raramente y tiende a decaer hacia G_2 . Es decir, la separación de estos residuos iónicos es un proceso que requiere alta energía y, en ausencia de modulador, no se observó una disrupción en la red polar.

8.3. Participación del modulador positivo e interacciones involucradas

De manera análoga al caso sin modulador, se investigaron las conformaciones de la red polar intracelular en relación con los estados metaestables en los sistemas con modulador positivo. La figura 8.3 muestra las estructuras con detalle sobre las interacciones polares investigadas en el sistema en presencia de agonista y modulador positivo. En este caso, la interconversión más lenta ocurre entre los estados $GC_3 \rightarrow GC_4$. En los paneles de la figura 8.3 puede apreciarse que la principal diferencia entre estas estructuras, a nivel de la red polar, es la separación del residuo R348^{6.37b} del par R176^{2.46b} y E408^{8.49b}, es decir, el paso más lento involucra la ruptura de esta interacción iónica.

Los estados GC_2 y GC_3 , de probabilidad intermedia y muy similar, involucran un cambio en la conformación de la cadena lateral del residuo R348^{6.37b}, con un incremento de su exposición hacia el espacio intracelular. De acuerdo con los flujos netos calculados, el sistema evoluciona desde GC_2 hasta GC_1 , donde rápidamente se

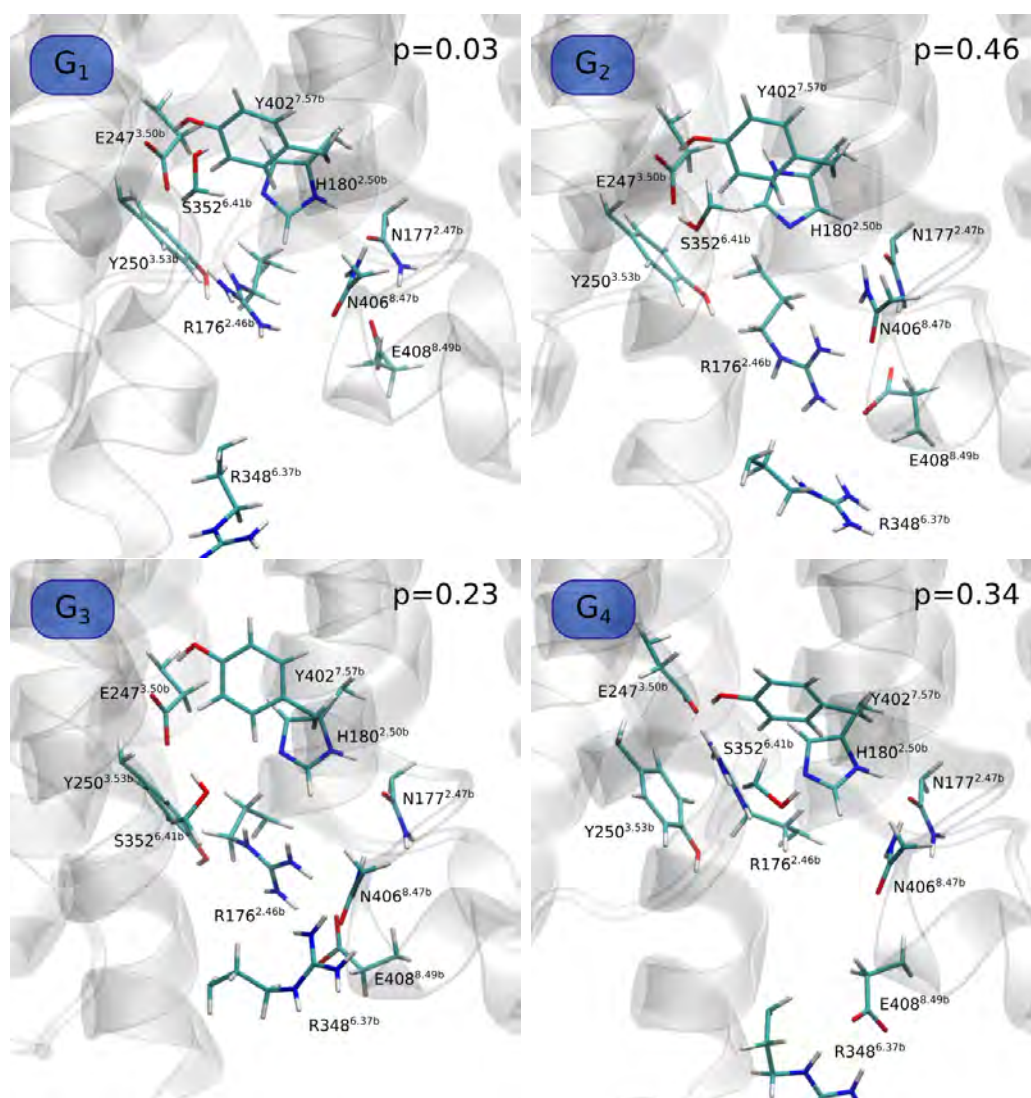


Figura 8.2: Conformaciones de la red polar intracelular para estructuras representativas de los estados metaestables del receptor en ausencia de modulador. Dentro de las imágenes se incluyen los valores de la probabilidad estacionaria para cada estado.

convierte hacia GC_4 , cuyo valor de probabilidad es el más grande de los estados metaestables. En este estado, la cadena lateral del residuo de $R348^{6.37b}$ ya no participa en la interacción iónica en el centro de la superficie intracelular, aunque el contacto $R176^{2.46b} / E408^{8.49b}$ es persistente. Así mismo, el residuo de $S352^{6.41b}$ en la hélice TM6, en todos los estados se orienta con el grupo hidroxilo expuesto, dejando de participar en la red intracelular central. De esta manera, en presencia del compuesto 2, las interacciones polares que anclan a la hélice TM6 con el resto, se debilitan, lo que puede favorecer el movimiento requerido para su activación.

El residuo de $C347^{6.34b}$ es el punto de unión covalente del compuesto 2 y es inmediatamente aledaño al residuo $R348^{6.35b}$. La cadena lateral modificada y más voluminosa del residuo $C347^{6.34b}$ podría permitir contactos que induzcan estrés suficiente sobre la región intracelular de la hélice TM6 para superar las barreras energéticas del candado iónico y la red polar. Para observar la participación del compuesto 2

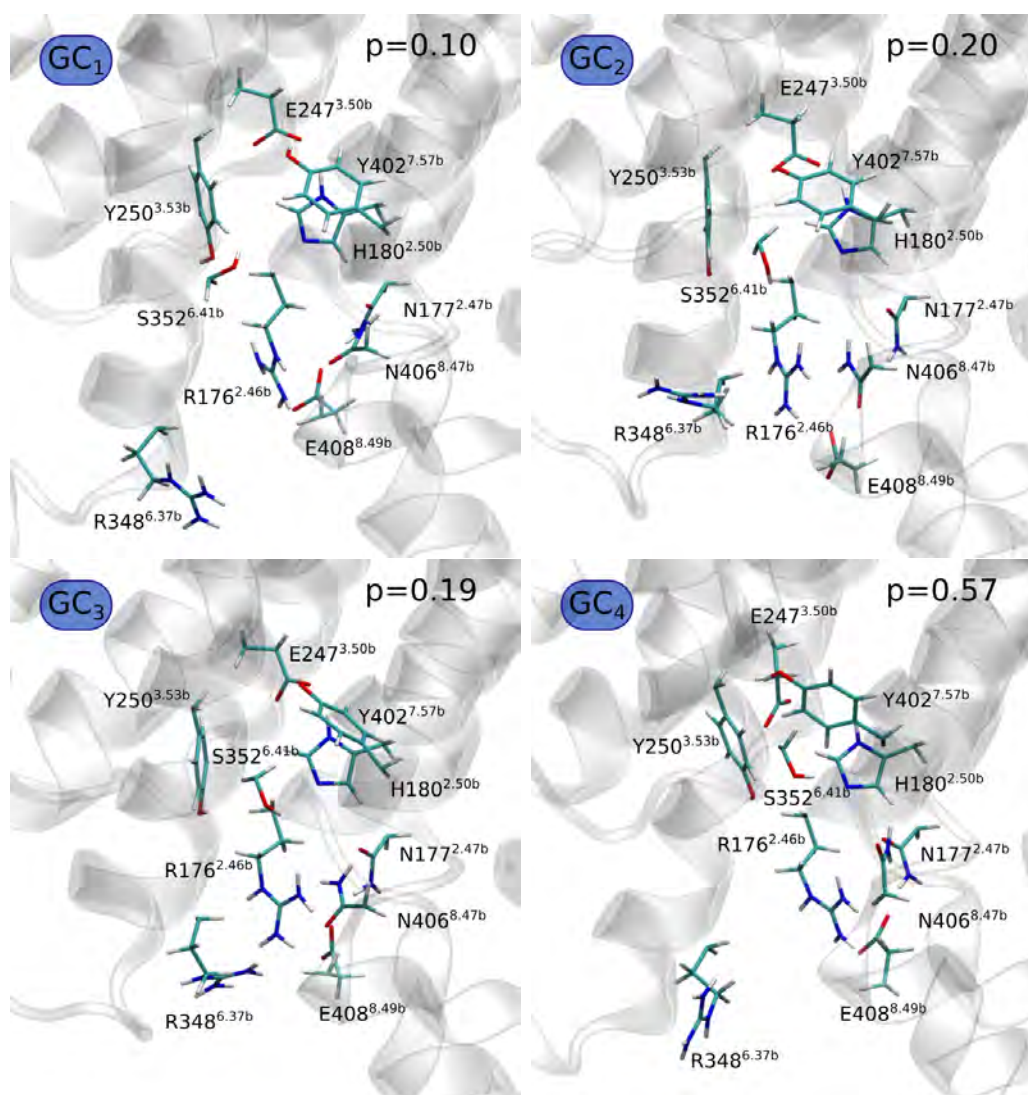


Figura 8.3: Conformaciones de la red polar intracelular para estructuras representativas de los estados metaestables del receptor con modulador positivo y agonista. Dentro de las imágenes se incluyen los valores de la probabilidad estacionaria para cada estado.

sobre las conformaciones de esta región, la figura 8.4 muestra las estructuras representativas en cada estado metaestable. En estas conformaciones, sólo en GC_3 el compuesto 2 se ubica con su grupo ter-butilo expuesto hacia el espacio intracelular. En esta posición, la cadena lateral del residuo $R348^{6.35b}$ se encuentra orientada hacia el par iónico $R176^{2.46b} / E408^{8.49b}$, formando una triada. En el resto de los estados, el compuesto 2 se orienta de forma que entra en contacto con los residuos hidrofóbicos del extremo intracelular de las hélices TM5 y TM6. De particular importancia es el estado GC_4 que, en este sistema, es el de menor energía, de acuerdo con su probabilidad estacionaria. En GC_4 , el grupo ter-butilo entra en contacto estrecho con los residuos hidrofóbicos $L255^{3.58b}$, $I330^{5.60b}$, $V331^{5.61b}$ y $A350^{6.39b}$. Por su parte, el anillo aromático tiene interacciones cercanas con las cadenas alifáticas de los residuos $K334^{5.64b}$ y $K342^{6.31b}$. Esta conformación está favorecida por el colapso de los grupos no polares en el compuesto 2 hacia el interior del ambiente hidrofóbico formado por los residuos no polares ubicados en esta región de la proteína. Además,

dado que el modulador se une de manera covalente a la C347^{6.34b}, para que adquiriera esta conformación favorable es necesario reubicar el resto del residuo de cisteína, con lo que se ocasiona un giro en la hélice que influye sobre la orientación de su residuo vecino, la R348^{6.35b}, generando la ruptura de la red. El efecto hidrofóbico y la unión covalente pueden ser suficientes para contrarrestar la barrera energética ejercida por las interacciones iónicas y polares de la red central. Esta hipótesis concuerda con el hecho de que los estados donde los residuos iónicos están a cortas distancias tienen una mayor probabilidad estacionaria en ausencia de modulador, pero cuando está presente, ésta probabilidad incrementa para los estados en que dichos residuos están alejados.

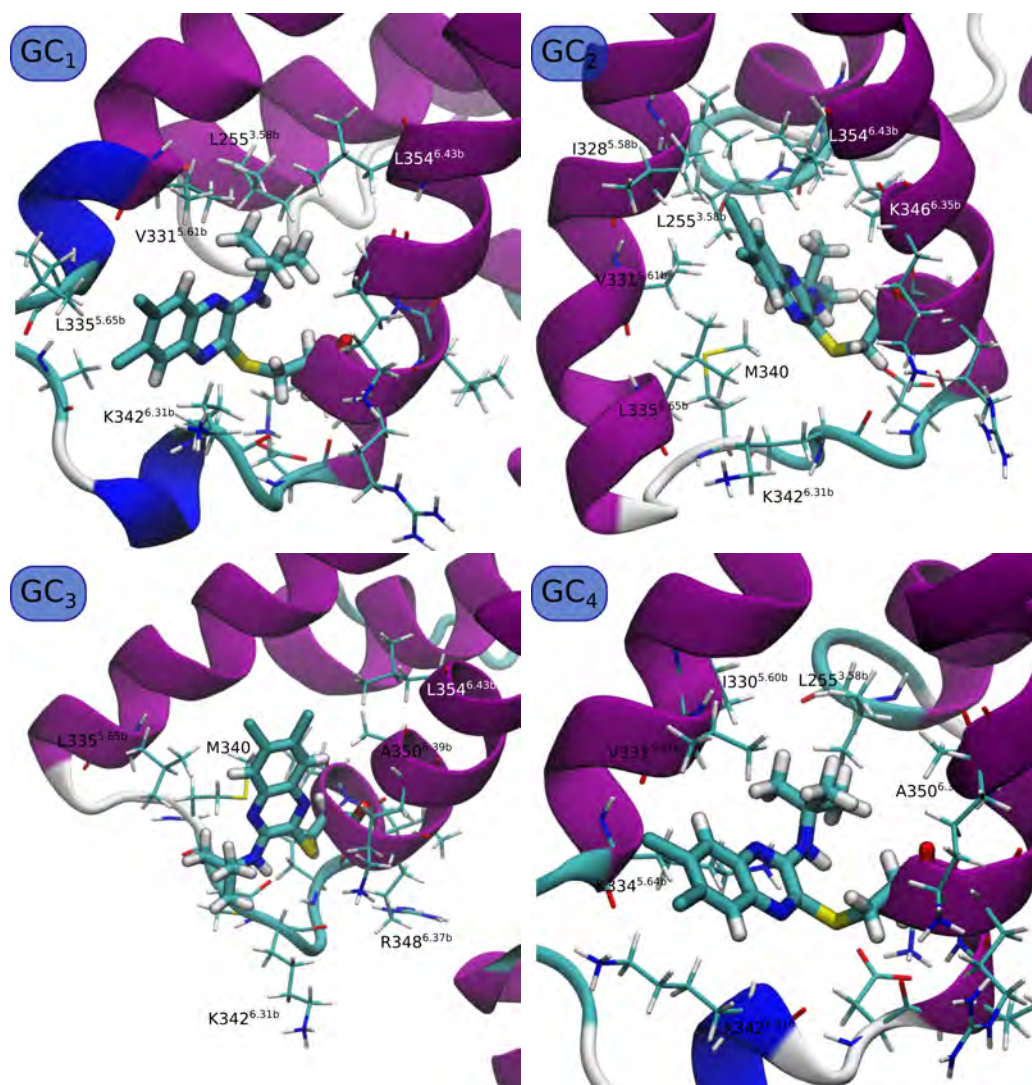


Figura 8.4: Conformaciones del compuesto 2 en las estructuras representativas de los estados metaestables predichos por los modelos de Markov. Dentro de las imágenes se incluyen los valores de la probabilidad estacionaria para cada estado.

Con el propósito de esclarecer si la energía involucrada en la separación de este par iónico se ve afectada por la presencia de modulador, se realizó un cálculo de energía libre con *Umbrella Sampling* (US), utilizando como coordenada de reacción la distancia entre el átomo de carbono ζ del residuo R348^{6.35b} y el átomo de carbono δ del residuo E408^{8.49b}, como se describe en la sección de métodos. Éstos átomos

representan los centros de los grupos guanidinio y carboxilo, que les confieren la carga eléctrica a los residuos de aminoácido involucrados, por lo que esta distancia se puede utilizar para medir la interacción electrostática que ocurre entre ellos. Las gráficas en la figura 8.5 muestran el valor de energía libre como función de la coordenada de reacción, estimando la variabilidad en los datos mediante *bootstrapping*. En concordancia con los modelos de Markov y nuestra hipótesis, la separación del par iónico en ausencia de modulador es un proceso que implica un incremento inicial en la energía libre. De acuerdo con este cálculo, se requieren de aproximadamente 25 kJ/mol para separar estos grupos de 4.0 Å hasta 10.0 Å.

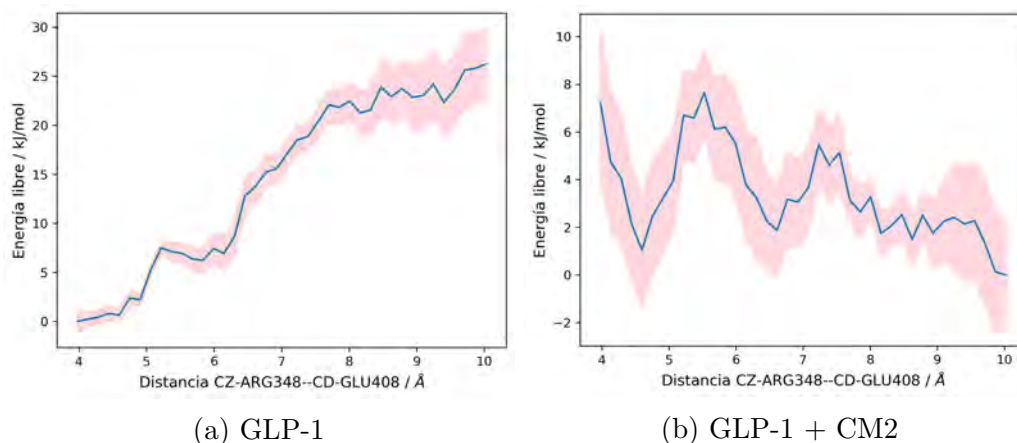


Figura 8.5: Cambios de energía libre como función de la distancia R348^{6.35b}-Cζ-E408^{8.49b}-Cδ, calculados mediante *Umbrella Sampling*. Las regiones sombreadas estiman el error utilizando *bootstrapping*

En presencia del compuesto 2, el perfil de la energía libre cambia completamente. La interacción a 4.0 Å se convierte en un estado de alta energía, desfavorable. A partir de este valor, la energía ya no asciende de manera escarpada, sino que pasa por dos mínimos alrededor de 4.5 Å y 6.5 Å, topándose con dos barreras en, aproximadamente 5.5 Å y 7.5 Å, aunque ninguna de estas barreras supera los 8.0 kJ/mol. Incluso, hay un descenso de energía al aproximarse a los 10.0 Å, por lo que este estado ya no es desfavorable.

Para verificar que lo observado en las estructuras representativas se cumple en el resto de la simulación, se calculó la misma coordenada de reacción utilizada en el *Umbrella Sampling*, para cada uno de los pasos en la dinámica. Con los datos obtenidos, se calculó la función de densidad de probabilidad en torno a la coordenada de reacción, utilizando un estimador de densidad con kernel gaussiano y re-ponderando la distribución por las probabilidades estacionarias predichas por los modelos de Markov. A partir de la función de densidad se obtuvo el perfil de energía libre correspondiente. Los resultados, para el sistema con y sin modulador se muestran en la figura 8.6. Los perfiles de energía libre se comportan de manera similar a lo observado en el *Umbrella Sampling*. En ausencia de modulador, el histograma muestra que el estado en que ambos residuos están cerca, alrededor de los 4.0 Å, se encuentra favorecido, mientras que a mayores separaciones, la probabilidad es más baja. El cambio en energía libre indica un incremento inicial para lograr la separación de los residuos iónicos. Además, este diagrama también coloca al estado inicial en un valle de energía libre, mientras aquellos con distancias más grandes de 6.0 Å, tienen

energías más altas. Cuando el compuesto 2 está presente, el histograma se invierte y es más fácil observar estados donde el puente salino se rompe, como se observa en el histograma. De acuerdo con el diagrama de energía libre y en concordancia con los resultados del Umbrella Sampling, el contacto entre estos dos grupos se convierte en un estado de alta energía, mientras que una separación de hasta 12 Å está ahora energéticamente favorecida. Los cálculos de energía indican que el compuesto 2 permite disminuir la barrera energética que ancla a la hélice TM6 en el candado iónico, principalmente debido al colapso hidrofóbico de este ligando al interior de la interfaz TM5-TM6 y su unión covalente al residuo de C347^{6.34b}.

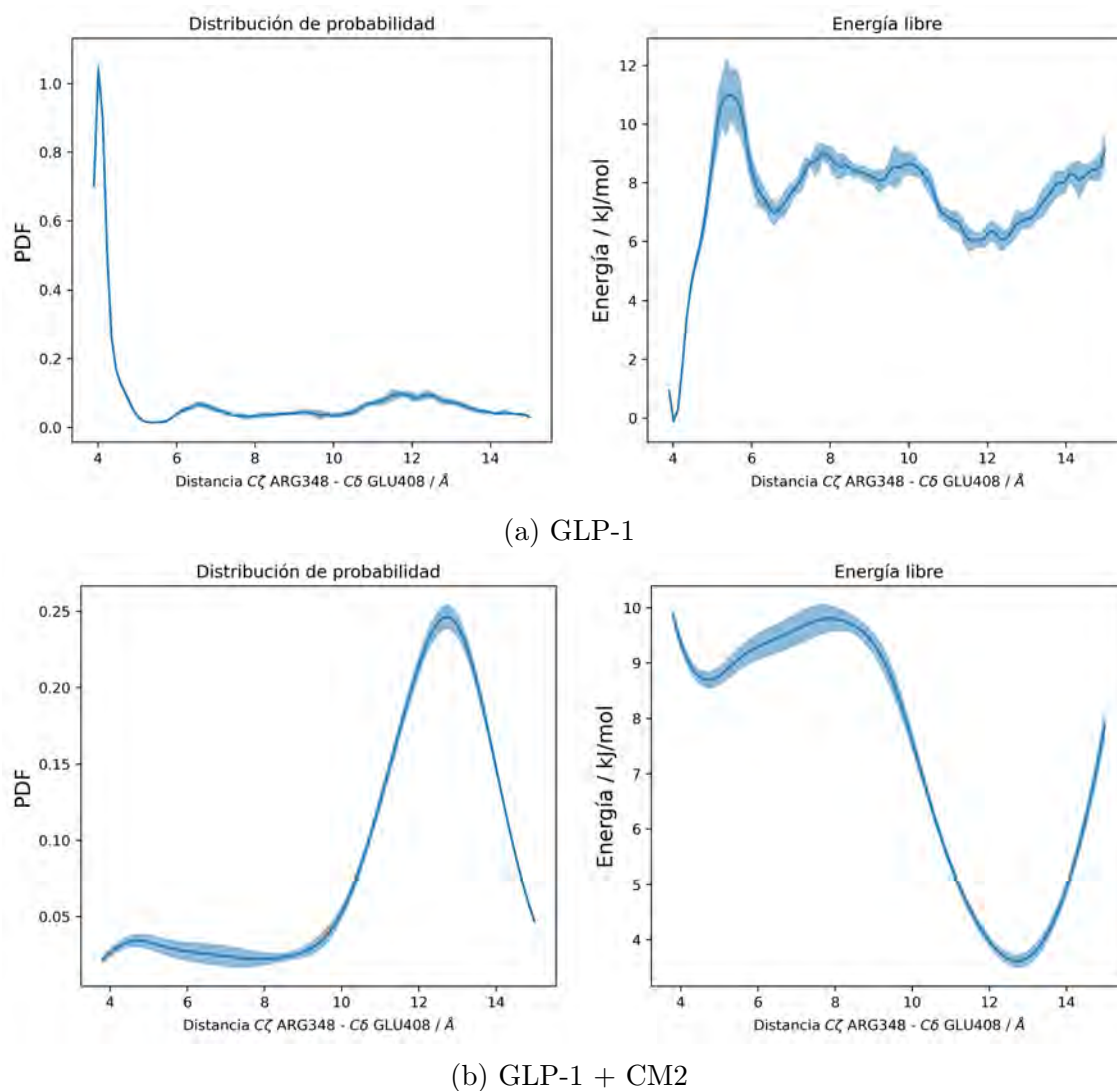


Figura 8.6: Funciones de densidad de probabilidad estimadas a partir de los histogramas de las distancias entre los átomos R348^{6.35b}-C ζ -E408^{8.49b}-C δ , observadas en las trayectorias de dinámica molecular. Las distribuciones fueron ponderadas por los valores de probabilidad en el equilibrio estimadas por los modelos de Markov. A la izquierda se muestra la gráfica correspondiente de energía libre.

8.4. Efecto de la presencia del agonista

Dado que los fenómenos observados ocurren en el extremo intracelular del receptor, alejados del sitio ortostérico, se investigó si este comportamiento también se cumple en ausencia de agonista. Los moduladores alostéricos de la familia del compuesto 2, además de amplificar la respuesta biológica en presencia de GLP-1, poseen actividad agonista débil. Por esta razón, en la literatura suelen describirse como moduladores *ago*-alostéricos (22). La figura 8.7 muestra las conformaciones de la red polar intracelular en ausencia de agonista para las estructuras representativas de los estados metaestables. Nuevamente, en los estados de mayor probabilidad, C_4 y C_3 , los residuos que participan en el candado iónico tienden a separarse, indicando que esta conformación sigue siendo favorable aun en ausencia de agonista. Sin embargo, las diferencias más interesantes con el sistema con GLP-1 no ocurren en el candado iónico, sino en la parte más profunda de la red polar intracelular.

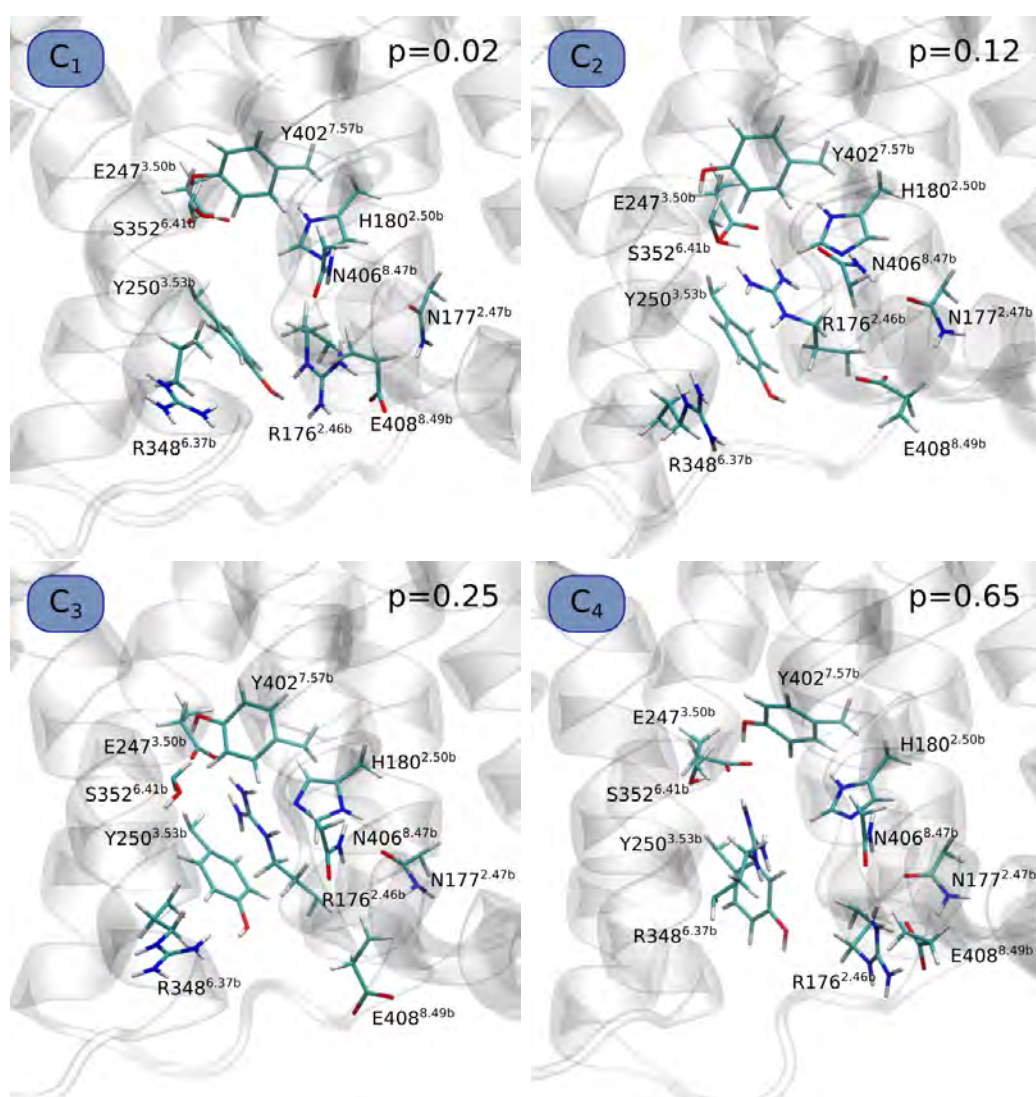


Figura 8.7: Conformaciones de la red polar intracelular para estructuras representativas de los estados metaestables del receptor con modulador positivo y en ausencia de agonista

Cuando el agonista GLP-1 está presente, la cadena lateral del residuo R176^{2.46b} se orienta de manera preferente expuesto hacia el espacio intracelular, con la capacidad de interactuar con el residuo de E408^{8.49b}. De hecho, en GC_4 , el estado de menor energía en presencia conjunta de modulador y agonista, E408^{8.49b} y R176^{2.46b} forman un puente salino en el que no participa R348^{6.35b}. El estado de mayor probabilidad y, por ende, de más baja energía, en ausencia de GLP-1 es C_3 . La figura 8.7 muestra que, en este estado, la cadena lateral de R176^{2.46b} permanece hacia el interior del receptor, apuntando en dirección al residuo E247^{3.50b}. Esta conformación también se observa en el estado C_2 que, aunque es menos frecuente que C_3 , también tiene una contribución importante. Un estado que no se observa en los demás sistemas es C_4 , donde el residuo R348^{6.35b}, en lugar de ubicarse expuesto hacia el espacio intracelular, se oculta hacia el interior de la red polar en dirección al residuo E247^{3.50b}, tal como lo hace R176^{2.46b} en C_2 y C_3 . La matriz de traslape entre los subespacio mostraba que el sistema en presencia sólo de modulador positivo divergía de aquellos en presencia de agonista, por lo que es posible considerar que este estado es uno de los puntos de divergencia. Además de estas interacciones, el residuo E247^{3.50b} presenta una interacción cercana con Y402^{7.57b} que es persistente en todos los estados.

Para averiguar si hay cambios en las interacciones ejercidas por el compuesto 2 cuando no está presente el agonista, se buscaron los contactos cercanos en las estructuras representativas de los estados metaestables y los resultados se ilustran en la figura 8.8. En los estados de menor probabilidad estacionaria, C_1 y C_2 , el compuesto 2 tiene un alto grado de exposición hacia el espacio intracelular, mientras que lo contrario se observa en los estados de menor energía C_3 y C_4 . Nuevamente, en las estructuras de mayor probabilidad estacionaria, el grupo ter-butilo del modulador se oculta mediante interacciones con las cadenas alifáticas de V331^{5.61b}, L254^{3.57b}, L251^{3.54b}, L255^{3.58b} y L354^{6.43b}. Estos contactos son los mismos o muy similares a los que se observan en presencia de agonista, también en los estados de menor energía. Los resultados indican que la conformación más favorecida del compuesto 2 en su punto de unión es equivalente tanto en presencia como en ausencia del GLP-1.

Las conformaciones observadas en el sistema sin GLP-1, así como sus valores de probabilidad estacionaria, muestran que las diferencias principales ocurren en las conformaciones de la parte interna de la red polar intracelular. En ausencia de agonista, la interacción polar E247^{3.50b} / R176^{2.46b} es de mayor estabilidad en comparación con el sistema en presencia de GLP-1. Esta conformación tiende a mantener a los residuos que participan en interacciones importantes con la proteína G en un estado de difícil acceso a la superficie intracelular. No obstante, la distancia entre los residuos que participan en el candado iónico, R348^{6.37b} y E408^{8.49b}, se ve afectada de manera muy similar en ambos casos. Inclusive, las interacciones y modos de unión de mayor probabilidad para el compuesto 2 parecen ser independientes de la presencia de agonista, por lo que su influencia sobre esta interacción es una característica propia de este modulador.

Al igual que en los sistemas anteriores, se calcularon las distribuciones de probabilidad y los valores de energía libre para la distancia entre los aminoácidos que forman el candado iónico. Los resultados de estos cálculos se muestran en las gráficas de la figura 8.9. Es interesante notar que el contacto estrecho entre los residuos de aminoácido vuelve a ser un estado de baja energía. Sin embargo, las barreras

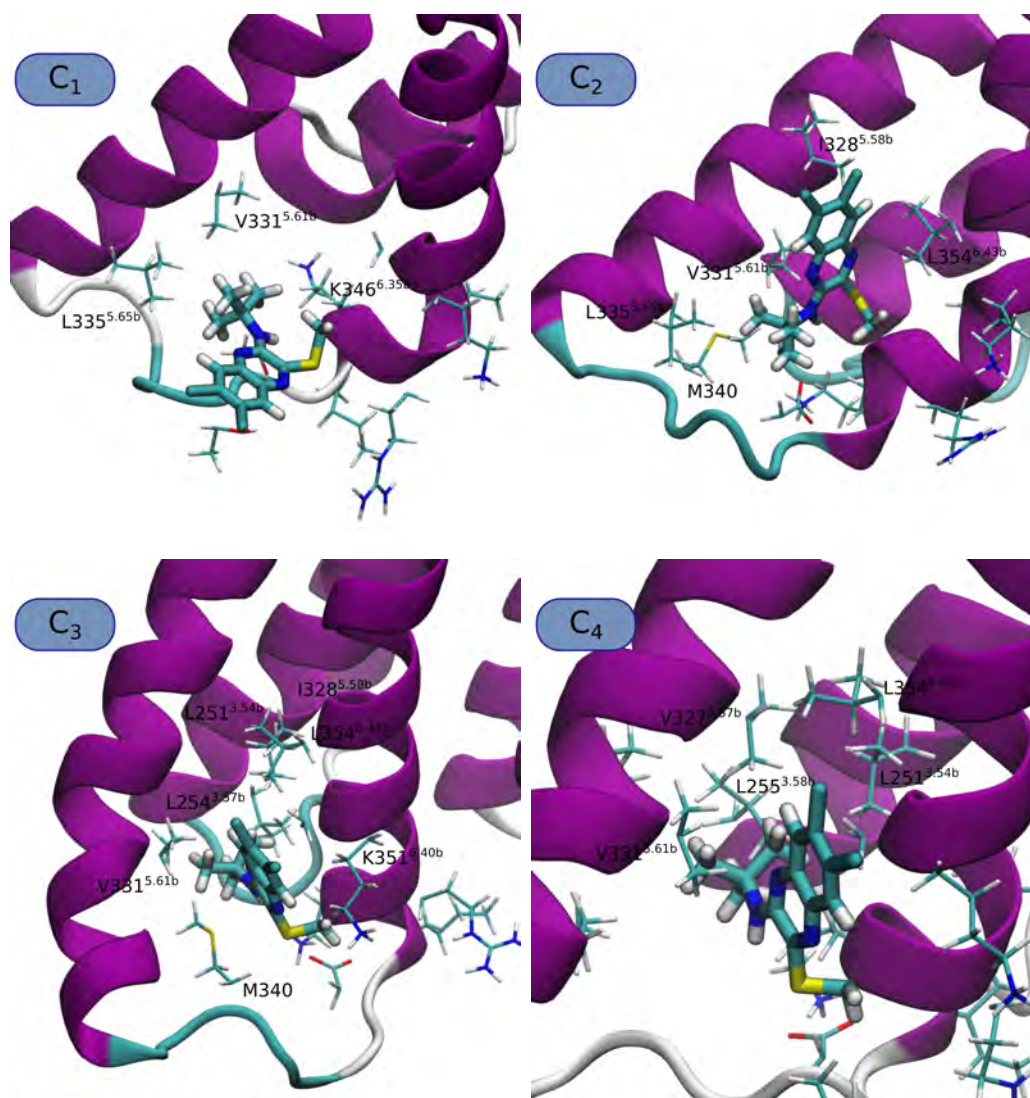


Figura 8.8: Conformaciones del compuesto 2 en las estructuras representativas de los estados metaestables predichos por los modelos de Markov, en ausencia de agonista

presentes durante la separación del puente salino no son tan pronunciadas como en ausencia de modulador. Esto indica que la ruptura de la red polar intracelular inducida por la presencia del agonista en el sitio ortostérico también tiene influencia sobre la ruptura del candado iónico. Es decir, aunque el compuesto 2 es capaz de disminuir, por sí mismo, la barrera de energía impuesta por la interacción entre los residuos de la red polar, es la acción sinérgica entre el agonista y el modulador el factor más importante para que la ruptura de las interacciones que mantienen al receptor en conformación inactiva sean un proceso favorable.

8.5. Participación del modulador negativo

Los moduladores alostéricos negativos conocidos para el GLP-1R tienen dos puntos de anclaje no covalentes, uno en la interfaz de la hélice TM5-TM6, cercano al

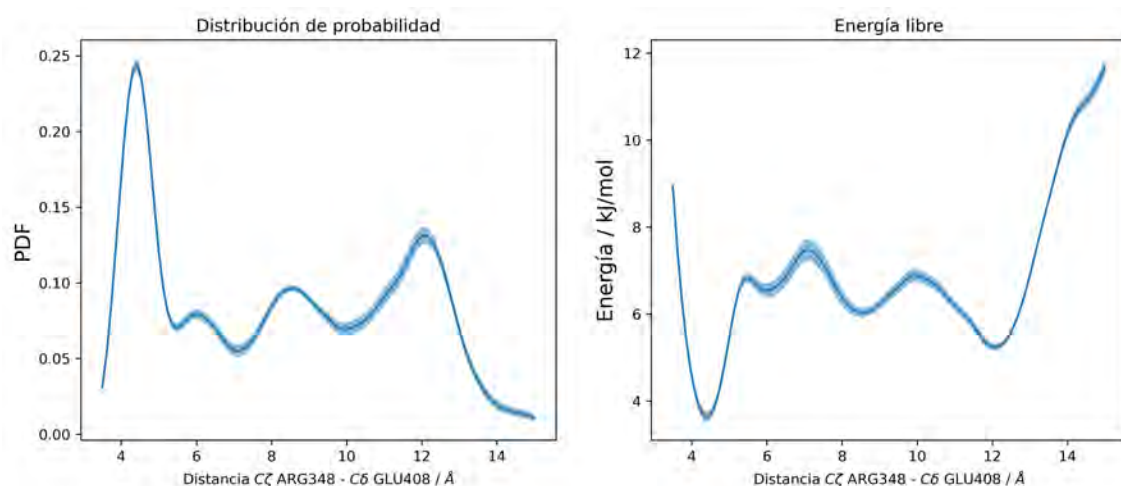


Figura 8.9: Funciones de densidad de probabilidad estimadas a partir de los histogramas de las distancias entre los átomos R348^{6.35b}-C ζ -E408^{8.49b}-C δ , observadas en las trayectorias de los sistemas en presencia de modulador positivo, pero ausencia de agonista. Las distribuciones fueron ponderadas por los valores de probabilidad en el equilibrio estimadas por los modelos de Markov.

punto de unión de los moduladores covalentes, y otro en la interfaz TM6-TM7, en contacto directo con la red polar intracelular. Se conocen moduladores negativos que actúan mediante un mecanismo similar en el receptor para glucagón, el receptor representativo de esta familia. En la figura 8.10 se muestran las estructuras representativas para los estados metaestables observados en las simulaciones en presencia de modulador negativo. En este caso, N_1 es un estado de muy baja probabilidad que decae muy rápidamente hacia N_3 y N_2 . En N_1 , el anillo tetrazólico y el grupo amida del modulador negativo se unen en la profundidad de la red polar, ejerciendo interacciones con el grupo hidroxilo del residuo S352^{6.41b} y la cadena lateral de R348^{6.37b}. Al mismo tiempo, este residuo forma la triada iónica con E408^{8.49b} y R176^{2.46b}. En N_2 y N_3 , el anillo tetrazólico se aleja de los residuos de la red profunda, H180^{2.50b} y Y402^{7.57b}, pero el residuo S352^{6.41b} se reorienta y ejerce una interacción persistente con el grupo amida, adyacente al tetrazol. S352^{6.41b} es uno de los residuos involucrados en interacciones polares con la proteína G, por lo que su interacción favorable con el modulador puede restringir su accesibilidad a la superficie intracelular, impidiendo que pueda entrar en contacto con la proteína.

Así mismo, la conformación de la triada iónica difiere entre los estados de mayor probabilidad. En N_2 , los residuos E408^{8.49b} y R176^{2.46b} forman un puente salino, pero en N_3 son E408^{8.49b} y R348^{6.37b} quienes forman este tipo de interacción. De esta manera, el sistema tiende a oscilar entre este intercambio, pero con una presencia más frecuente del estado N_3 . Por tanto, en estas condiciones, la formación del candado iónico en la superficie intracelular se ve favorecida y tiene mayor persistencia que en los sistemas anteriores. El modulador negativo favorece la captura del receptor en la conformación inactiva porque es capaz de estabilizar la red polar intracelular. Tanto S352^{6.41b}, como R348^{6.37b} pertenecen a la hélice TM6, por lo que estas interacciones tienden a anclar la hélice hacia el interior del dominio.

Nuevamente, los valores de las distancias entre los residuos que forman el candado

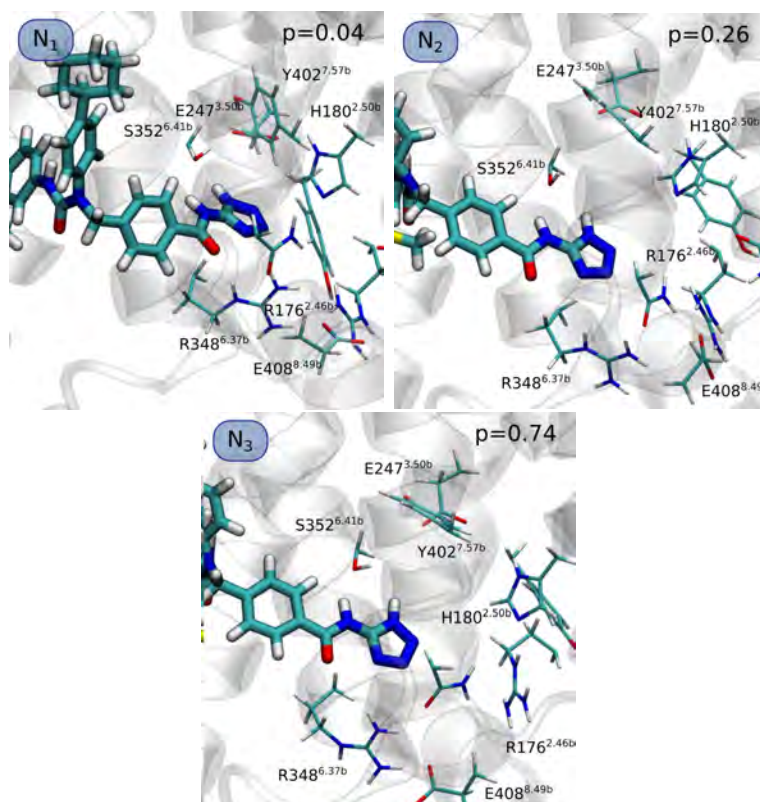


Figura 8.10: Conformaciones de la red polar intracelular y sus interacciones con el modulador negativo para estructuras representativas de los estados metaestables del receptor. Dentro de las imágenes se incluyen los valores de la probabilidad estacionaria para cada estado.

iónico se calcularon para cada uno de los pasos de la simulación, con el fin de generar los histogramas ponderados por la distribución estacionaria. Los resultados se presentan en las gráficas de la figura 8.11. En este caso, los estados donde los residuos R348^{6.37b} y E408^{8.49b} están en contacto estrecho tienen el menor valor de energía libre, similar a lo observado en el sistema en ausencia de modulador positivo. Al incrementar la distancia de separación, la energía se incrementa de manera casi monótona, lo que indica que, al contrario de lo observado para el compuesto 2, el modulador negativo tiende a incrementar la barrera energética que captura a la red polar en la conformación inactiva. El estado N_3 , de mayor probabilidad estacionaria, muestra que el modulador negativo se une mediante interacción con la cadena lateral del residuo de R348^{6.37b}, a través del grupo amida unido al anillo tetrazólico. Este contacto puede favorecer la orientación del residuo de manera que se encuentre más fácilmente disponible para formar el puente salino.

8.6. Concordancia con las relaciones estructura - actividad y datos experimentales

De esta manera, las distribuciones estacionarias calculadas a partir de los modelos de Markov para cada uno de los sistemas, predicen que las conformaciones de la

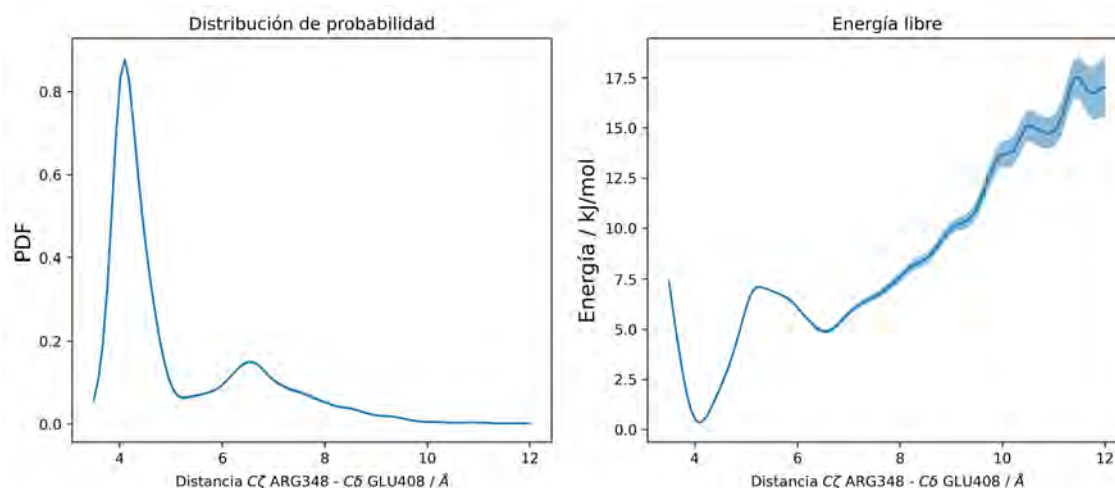


Figura 8.11: Funciones de densidad de probabilidad estimadas a partir de los histogramas de las distancias entre los átomos R348^{6.35b}-Cζ-E408^{8.49b}-Cδ, observadas en las trayectorias de los sistemas en presencia de modulador negativo. Las distribuciones fueron ponderadas por los valores de probabilidad en el equilibrio estimadas por los modelos de Markov.

red polar central tienen diferentes pesos probabilísticos. Los cambios de estabilidad, relacionados con las probabilidades estacionarias, pueden atribuirse a la presencia de los moduladores alostéricos y esto se refleja en los cambios de energía libre que cada ligando induce. Los ligandos covalentes de la familia del compuesto 2 podrían actuar mediante mecanismos muy similares a los descritos en este trabajo. De hecho, las relaciones estructura-actividad sobre este tipo de compuestos muestra que los puntos críticos para su capacidad moduladora son (22): (1) la presencia de sulfóxidos o sulfonas unidas al anillo, (2) sensibilidad a la presencia de grupos electro-donadores en el anillo de benceno y (3) sensibilidad a la presencia de grupos polares en la posición del ter-butilo.

Una de las principales fuerzas motoras en la disrupción del candado iónico fue la retracción del compuesto 2 de manera que su grupo ter-butilo se encontrara en el núcleo hidrofóbico del extremo intracelular de las hélices TM3, TM5 y TM6. Este comportamiento debe mantenerse para cualquier otro grupo no polar, siempre que su tamaño no supere el volumen disponible para que haya una repulsión estérica. La tabla 8.2 muestra algunos análogos estructurales del compuesto 2, acompañados por su valor de pEC₅₀ sobre la producción de cAMP tras la activación del GLP-1R (22). Esta tabla muestra que los sustituyentes con grupos polares en el sitio correspondiente al ter-butilo se asocian con valores menores de actividad biológica. Por el contrario, la presencia de grupos hidrofóbicos y aromáticos incrementan los valores de actividad hasta dos unidades logarítmicas. De esta forma, el comportamiento observado por modificación estructural de los compuestos concuerdan con el mecanismo propuesto.

Además, el hecho de que el ligando se une de manera covalente al átomo de azufre en la C347^{6.36b} puede contribuir a que se ejerza la torsión sobre la hélice para romper el candado iónico. Es decir, la separación de un par iónico en agua es más favorable que la ruptura del enlace covalente C-S. Si el ligando se uniera de forma

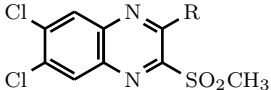
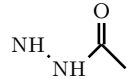
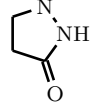
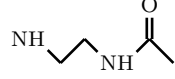

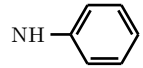
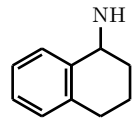
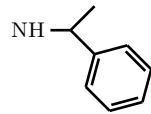

Estructura base	R	pEC ₅₀
		4.85
		4.85
		5.15
		5.20
		6.14
		6.36
		6.76
		6.81

Tabla 8.2: Modificaciones estructurales sobre algunos análogos estructurales del compuesto 2 y valores de actividad biológica sobre la activación del GLP-1R. Valores obtenidos de (22)

no covalente, podría asumir una conformación estable sin necesidad de modificar la orientación de la cisteína. Esto implica que la formación del aducto puede ser un paso esencial para la modulación alostérica, explicando la relación (1). Finalmente, no se encontró un rol persistente en las interacciones de los átomos de cloro con el receptor, pues se encuentran expuestos al disolvente en varios de los estados meta-estables o ejercen pocos contactos, principalmente con los residuos más superficiales del extremo intracelular de la hélice TM5. El papel de este sustituyente sobre el mecanismo de modulación puede ser de naturaleza electrónica, pues la presencia de grupos electroattractores unidos al anillo aromático favorece las reacciones de sustitución nucleofílica aromática.

Por otra parte, los resultados obtenidos también sugieren una participación importante de la red polar intracelular sobre la modulación alostérica positiva. En particular, los modelos y perfiles de energía libre predicen la ruptura de la interacción iónica entre el residuo R348^{6.37b} y E408^{8.49b}, pero la interacción de E408^{8.49b} con otros residuos polares en la red es persistente. Por ejemplo, se observó que, en el sistema en presencia de agonista y modulador alostérico positivo, el estado meta-estable con mayor valor de probabilidad estacionaria involucra la interacción entre la cadena lateral de E408^{8.49b} con R176^{2.46b}. La figura 8.12 muestra las conformaciones de la red intracelular para una serie de estructuras experimentales del receptor

en conformación activa e intermedia en presencia de distintos agonistas. En la mayor parte de estas estructuras, tras perderse la participación de R348^{6.37b} debido al movimiento de la hélice TM6, el residuo E408^{8.49b} que formaba parte del candado iónico en la conformación inactiva, presenta interacciones persistentes con los residuos R176^{2.46b} y N406^{8.47b}. De esta forma, los estados predichos por la simulación en la región formada por la red polar intracelular es similar a las conformaciones observadas mediante métodos experimentales en los estados activos.

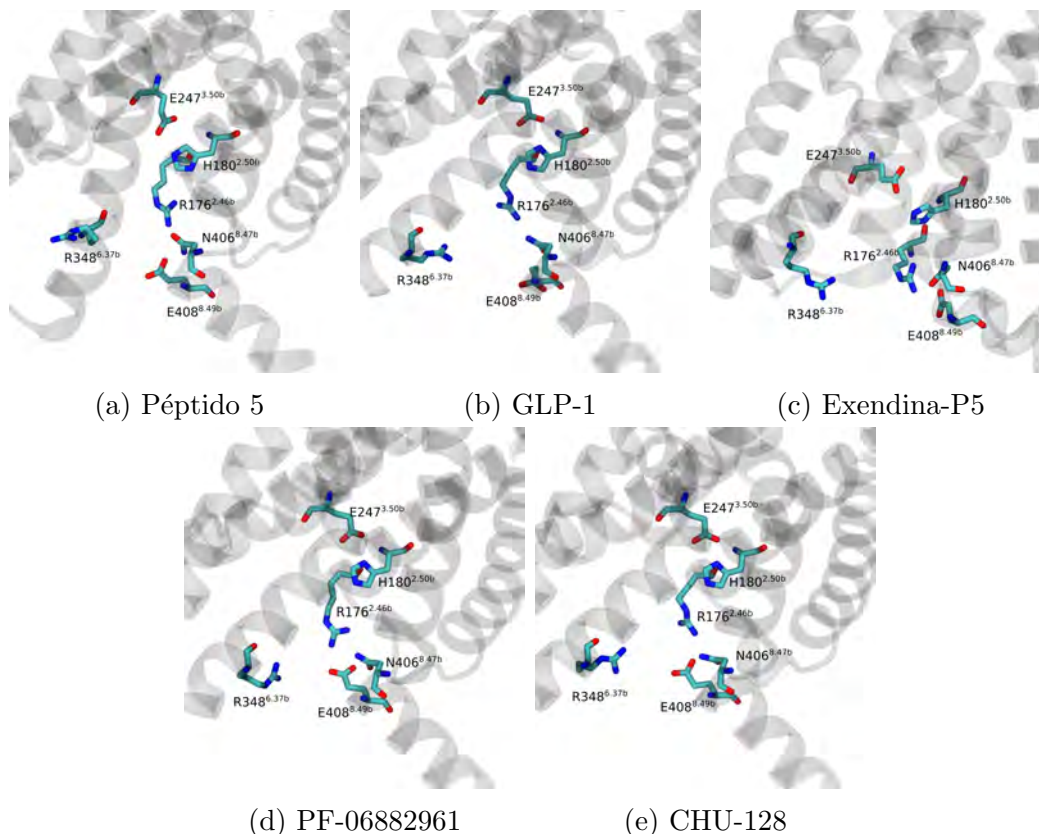


Figura 8.12: Conformaciones de la red polar intracelular en el GLP-1R en estado activo, en presencia de diversos agonistas: Péptido 5 (19), GLP-1 (56), Exendina-P5(17), PF-06882961 (56), CHU-128 (56)

Aunado a esto, en un estudio de cribado virtual (25) publicado durante la elaboración de este trabajo, se encontró una nueva molécula pequeña, llamada C-1, con actividad moduladora que no posee grupos reactivos y muy probablemente se une al receptor de manera no covalente. Los resultados de acoplamiento molecular de estos autores sugieren que el ligando forma interacciones con los residuos de S352^{6.41b} y N406^{8.47b}, ambos participantes en la red polar intracelular. Esto coincide con el mecanismo propuesto respecto a que la ruptura de las interacciones polares en esta red puede favorecer la activación del receptor. El ligando también es capaz de producir una respuesta biológica en ausencia de agonista, al igual que las moléculas similares al compuesto 2. El descubrimiento de C-1 y la similitud de su efecto con los moduladores alostéricos conocidos es una prueba de la participación de la red polar intracelular sobre la activación del receptor, como se observó en los resultados de este trabajo.

Las observaciones derivadas de los estados metaestables y de los perfiles de

energía libre permiten explicar el mecanismo por el que ocurre la modulación alostérica sobre el GLP-1R. El mecanismo deducido a partir de los resultados coincide con las relaciones estructura-actividad que se han observado sobre las moléculas moduladoras de unión covalente al residuo de C347^{6.36b} y puede extenderse sobre esta familia de moduladores positivos. De esta forma, se pueden tomar en consideración las interacciones hidrofóbicas necesarias en las hélices TM5 y TM6, así como la presencia de grupos reactivos para buscar ligandos que ofrezcan una mayor estabilidad en medios biológicos, pero puedan mantener su capacidad moduladora sobre el GLP-1, en campañas posteriores de búsqueda o diseño, como se propone en la sección de Perspectivas.

Conclusiones

Los resultados de las simulaciones de dinámica molecular, en conjunto con los modelos de estado de Markov y los perfiles de energía libre calculados, indican que la presencia del modulador alostérico positivo analizado se relaciona con la disrupción de la red polar intracelular que se opone a la activación del receptor. Los principales cambios se deben a la unión covalente con el residuo de cisteína reactivo y a las interacciones hidrofóbicas del compuesto 2 con cadenas laterales de residuos de aminoácido no polares en el sitio de unión. Las observaciones coinciden con las relaciones de estructura y actividad observadas en análogos estructurales del compuesto 2, permitiendo explicarlas en términos de la interacción ligando-proteína. Ante la naturaleza covalente del mecanismo de unión de estos ligandos, la selectividad frente a otros residuos de cisteína expuestos o su inactivación prematura es un reto para lograr su uso como fármaco. No obstante, la búsqueda de mecanismos moderadamente exergónicos y la optimización de los fragmentos de la molécula que interactúan de forma no covalente con el receptor en el sitio de unión encontrado, pueden contribuir a conseguir la selectividad deseada, empleando diversos métodos de modelado molecular. En este escenario, y en ausencia de estructuras experimentales del receptor GLP-1R en complejo con este tipo de moduladores, se propone un modelo de unión plausible del que se puede partir para la búsqueda de nuevas moléculas con capacidad moduladora, con el objetivo de mejorar el perfil farmacológico de este tipo de compuestos.

Perspectivas

En este trabajo se logró describir, de manera teórica y basado en los valores de probabilidad predichas por los modelos de Markov, un mecanismo plausible para la modulación alostérica positiva sobre el GLP-1R. Se identificaron interacciones en las que participan residuos de aminoácido particulares que pueden ser claves en este mecanismo. De manera interesante, no hay muchos estudios de mutagénesis sobre los residuos de aminoácido identificados, en el contexto de modulación alostérica (24, 57). Con los datos derivados de este trabajo se pueden plantear experimentos para introducir perturbaciones sobre la red polar intracelular, utilizando, por ejemplo, mutagénesis dirigida. Dada la cercanía de los residuos que participan en el candado iónico, se pueden introducir modificaciones covalentes artificiales para mantenerlos unidos. De acuerdo con nuestros resultados, el receptor debería permanecer capturado en el estado inactivo, por lo que su actividad debería verse disminuida. De manera análoga, se puede anclar el residuo de cisteína reactivo a la hélice TM6, de forma que adquiera una conformación similar a la adoptada en presencia del compuesto 2. Esta nueva conformación debería tener un efecto similar a la presencia de modulador, por lo que el receptor podría ser más sensible a la acción del agonista. Los estudios propuestos servirían para aportar evidencia experimental sobre los mecanismos propuestos, operantes en la modulación alostérica del receptor GLP-1R.

En relación con el diseño de moduladores, el enlace covalente del ligando al residuo de cisteína reactivo parece ser necesario para este tipo de mecanismo. Si bien, las moléculas que actúan por modificación covalente se han evitado en el desarrollo e investigación de fármacos debido a varias desventajas relacionadas con su toxicidad y promiscuidad, en años recientes ha incrementado el interés en el diseño dirigido de modificadores covalentes (58, 59). La fracción de fármacos aprobados y en pruebas clínicas que actúan por mecanismos de unión covalente es significativa, lo que demuestra que es posible lograr especificidad y selectividad sobre blancos biológicos, así como seguridad en su uso (58, 59). Uno de los principales problemas de las moléculas análogas al compuesto 2 y al BETP es su baja estabilidad en líquidos biológicos. Una estrategia para optimizar su perfil farmacológico sin comprometer su actividad, puede ser la búsqueda de otros grupos reactivos capaces de unirse al grupo tiol de la cisteína por otros mecanismos de reacción moderadamente exergónicos que permitan su unión covalente reversible (60, 61). Por ejemplo, entre los grupos reactivos más utilizados en fármacos de unión a cisteína se encuentran las acrilamidas, alquilamidas, fumaratos y vinilsulfonas (59). Es posible realizar una selección de compuestos con este tipo de sustituyentes y someterlos a cribado virtual, utilizando un protocolo de acoplamiento molecular covalente como el empleado en este

trabajo, usando las estructuras representativas de los estados más estables predichos por los modelos de Markov, con el fin de seleccionar ligandos prometedores para su evaluación experimental. Posteriormente, se puede optimizar la "cadena lateral" del ligando para ejercer contactos no covalentes favorables en la cavidad de unión y, de esta manera, maximizar la selectividad sobre el receptor (60, 61).

Por otra parte, dado el protagonismo de la red polar intracelular en el mecanismo descrito, puede darse énfasis a la cavidad formada por estos residuos. En esta región del receptor no hay residuos que puedan ser potencialmente reactivos y las moléculas dirigidas sobre la cavidad se pueden unir de manera no covalente. Si una molécula puede interferir con la formación de la red polar o del candado iónico, posiblemente muestre actividad moduladora. Una prueba que contribuye a esta hipótesis es el descubrimiento del ligando C-1, que potencialmente se une en la red polar. Por lo tanto, este sitio puede explorarse con mayor profundidad para buscar moduladores alostéricos que actúen mediante mecanismos de unión no covalentes.

Agradecimientos

El autor agradece a CONACyT por el financiamiento mediante la beca otorgada al número de CVU 957490, para hacer posible el cumplimiento de estos estudios de posgrado.

Al Instituto de Química de la UNAM, por otorgar los recursos y facilidades de software y hardware requeridos para llevar a cabo este trabajo de investigación.

Los cálculos fueron realizados gracias al apoyo de la Dirección General de Cómputo y de Tecnologías de la Información y Comunicación (DGTIC)-UNAM, por proveer acceso a los recursos de cómputo de Miztli con el proyecto LANCAD-UNAM-DGTIC-367.

Al equipo de atención al usuario de Miztli, por sus respuestas atentas, rápidas y precisas y por sus valiosos consejos para aprovechar de la mejor manera los recursos y tiempo de cómputo.

A la Dra. Karina, por su asesoría, consejos y confianza durante el desarrollo del trabajo y los estudios de maestría. Al Dr. Abraham, por sus valiosos consejos en la asesoría técnica de los métodos computacionales y de modelado utilizados en el proyecto.

Resumen de métodos utilizados por MOE

A.1. Asignación de protómeros y tautómeros con Protonate 3D

Como paso previo a todo cálculo de acoplamiento molecular, es necesario calcular las posiciones de los átomos de hidrógeno que frecuentemente no se resuelven en la difracción de rayos X por la que se obtiene la estructura de una proteína. Protonate 3D es una herramienta implementada en MOE para optimizar la posición de los átomos de hidrógeno teniendo en cuenta el estado de protonación más probable para las cadenas laterales de los residuos de aminoácido ionizables. El algoritmo que utiliza Protonate 3D se resume en las siguientes etapas:

1. Se genera una partición del sistema de acuerdo con grupos funcionales definidos en un diccionario. Para cada grupo ionizable se generan todos los posibles estados de protonación y de tautómeros, así como una colección de rotámeros.
2. Se calcula una matriz de energía de interacción entre todos los grupos ionizables calculados. La función de energía incluye términos para interacciones de Van der Waals, electrostáticas y de solvatación mediante el modelo generalizado de Born.
3. Se calcula un vector con las auto energías para cada uno de los grupos calculados. La función de auto energía incluye términos para interacciones electrostáticas, de solvatación y de pK_a . El pK_a se calcula mediante un ciclo termodinámico.
4. Se realiza una pre-selección de los protómeros, eliminando aquellos de muy alta energía.
5. Mediante un algoritmo de optimización, se seleccionan los protómeros, tautómeros y rotámeros que minimizan la energía del sistema.

A.2. Algoritmo de acoplamiento molecular covalente

El acoplamiento molecular covalente implementado en MOE requiere que la reacción sea definida de antemano. El programa cuenta con una biblioteca limitada de reacciones, pero se puede definir cualquier reacción de interés mediante un archivo .rdf que se puede generar con un programa de edición de estructuras químicas 2D como Marvin Sketch o Chem Draw. El algoritmo de acoplamiento molecular covalente es:

1. El aducto entre el ligando y el residuo reactivo es generado, basado en el archivo de reacción.
2. El complejo residuo-ligando es utilizado para generar confórmeros y éstos se usan como punto de partida para calcular una serie de rotámeros.
3. Los rotámeros del residuo modificado se re-colocan en la cavidad. En la cavidad se realizan pequeñas rotaciones y traslaciones en torno al centro de masas para obtener conformaciones de baja energía, utilizando la función Alpha HB, que solo toma en cuenta parámetros geométricos y de enlace de hidrógeno.
4. Si se solicita un refinamiento, el programa busca un mínimo en la función de puntaje, modificando los valores de los grados de libertad mediante tres algoritmos sucesivos: los primeros pasos en dirección contraria al gradiente (*Steepest Descent*), luego cambia al método de gradientes conjugados y finalmente utiliza un método truncado de Newton hasta encontrar un mínimo o superar un determinado número de iteraciones. Si es utilizado un refinamiento por ajuste inducido, las cadenas laterales de los residuos de aminoácido en la cavidad también se incluyen en el proceso de optimización.

Enfoque variacional para procesos de Markov

B.1. Enfoque variacional para procesos de Markov (VAMP)

La parte medular del análisis por modelos de estado de Markov consiste en estimar una matriz \mathbf{P} que mejor reproduce las trayectorias observadas en las simulaciones de dinámica molecular, sobre una partición discreta del espacio de estados. Esto equivale a representar el sistema, cuya dinámica es inherentemente no lineal, con un modelo lineal sobre un nuevo espacio de variables discretas. Es necesario, entonces, medir el ajuste de \mathbf{P} a los datos observados, de manera análoga a cómo el valor de R^2 mide el ajuste en un problema de regresión. Noé y Wu (33) desarrollaron un método para evaluar la calidad de un modelo de estado de Markov, dada la matriz de transición estimada a partir de los datos. El método fue llamado enfoque variacional para procesos de Markov (*Variational Approach for Markov Processes*, VAMP) debida a que se fundamenta en un principio variacional para estimar los valores propios más altos de una matriz.

El teorema de Eckart-Young establece que la mejor aproximación para un operador lineal \mathfrak{T} (que puede representarse mediante una matriz \mathbf{T}) está dada por la descomposición en valores singulares de \mathbf{T} truncada hasta los k primeros términos (62). La ecuación B.1 muestra dicha descomposición utilizando notación de Dirac. Una matriz de transición es la representación del operador que propaga un vector de probabilidad sobre un intervalo de tiempo. Si el proceso de Markov es reversible, la descomposición en valores singulares coincide con su descomposición en vectores propios como se mostró en la ecuación 3.6.

$$\mathbf{T}_k = \sum_{i=1}^k \sigma_i |\mathbf{u}_i\rangle \langle \mathbf{v}_i| \quad (\text{B.1})$$

Para cualquier matriz \mathbf{T} y dos vectores arbitrarios $|\mathbf{x}\rangle$ y $|\mathbf{y}\rangle$, el cociente generalizado de Rayleigh se define de acuerdo con la ecuación B.2.

$$R = \frac{\langle \mathbf{x} | \mathbf{T} | \mathbf{y} \rangle}{\langle \mathbf{x} | \mathbf{x} \rangle^{1/2} \langle \mathbf{y} | \mathbf{y} \rangle^{1/2}} \quad (\text{B.2})$$

Se puede mostrar (62) que el cociente R es máximo si los vectores $|\mathbf{x}\rangle$ y $|\mathbf{y}\rangle$ coinciden con los vectores singulares de \mathbf{T} relacionados con el valor singular máximo. Además, el valor de R para este caso coincide con el valor singular máximo. Más aun, se puede extender este comportamiento hasta los primeros k valores singulares de la matriz. En este caso, si restringimos los vectores de prueba a que estén normalizados¹, se cumple la desigualdad de la ecuación B.3. La igualdad es estricta sólo si los vectores que participan en el cociente de Rayleigh son los primeros k vectores singulares. De esta manera, cualquier par de vectores de prueba generan una cota inferior para los valores singulares verdaderos de la matriz \mathbf{T} . Esto implica que, mientras mayor es el término de lado derecho en la desigualdad B.3, se obtiene una mejor aproximación a los vectores y valores propios de la matriz de interés y, por consecuencia, a la propia matriz \mathbf{T} . Éste término, reescrito en la ecuación B.4, se conoce como puntaje VAMP- r . Dado que la ecuación es válida para cualquier entero r , en teoría se puede utilizar cualquier valor, pero en PyEMMA se tiene implementado únicamente para $r = 2$ (VAMP-2).

$$\sum_{i=1}^k \sigma_i^r \geq \sum_{i=1}^k \langle \mathbf{x}_i | \mathbf{T} | \mathbf{y}_i \rangle^r \quad (\text{B.3})$$

$$R_r[\mathbf{x}, \mathbf{y}] = \sum_{i=1}^k \langle \mathbf{x}_i | \mathbf{T} | \mathbf{y}_i \rangle^r \quad (\text{B.4})$$

Aunque el VAMP-2 indica una medida de la aproximación a la matriz de transición que mejor reproduce las trayectorias observadas, su valor no está acotado. La mayor utilidad de esta función radica en que es un criterio para comparar entre modelos para el mismo sistema o para decidir el mejor valor algunos hiperparámetros cuando es requerido seleccionar un modelo.

¹Esto se puede lograr para cualquier vector al dividir entre su norma

Referencias

- (1) Organización Mundial de la Salud. Informe Mundial sobre la Diabetes, inf. téc., Organización Mundial de la Salud, 2016.
- (2) Shama-Levy, T., Vielma-Orozo, E., Hereida-Hernandez, O., Romero-Martinez, M., Mojica-Cuevas, J., Cuevas-Nasu, L., Santaella-Castel, J. y Rivera-Dommarco, J. Encuesta Nacional de Salud y Nutrición 2018-19: Resultados Nacionales, inf. téc., Cuernavaca, México: Instituto Nacional de Salud Pública, 2020.
- (3) De Graaf, C., Donnelly, D., Wootten, D., Lau, J., Sexton, P. M., Miller, L. J., Ahn, J.-M., Liao, J., Fletcher, M. M., Yang, D., Brown, A. J. H., Zhou, C., Deng, J. y Wang, M.-W. (2016). Glucagon-Like Peptide-1 and Its Class B G Protein-Coupled Receptors: A Long March to Therapeutic Successes. *Pharmacological Reviews* 68, 954-1013.
- (4) Nauck, M. A. y Meier, J. J. (2018). Incretin hormones: Their role in health and disease. *20*, 5-21.
- (5) Hinnen, D. (2017). Glucagon-like peptide 1 receptor agonists for type 2 diabetes. *Diabetes Spectrum*, DOI: 10.2337/ds16-0026.
- (6) Pabreja, K., Mohd, M. A., Koole, C., Wootten, D. y Furness, S. G. (2014). Molecular mechanisms underlying physiological and receptor pleiotropic effects mediated by GLP-1R activation. *British Journal of Pharmacology*, DOI: 10.1111/bph.12313.
- (7) Koole, C., Savage, E. E., Christopoulos, A., Miller, L. J., Sexton, P. M. y Wootten, D. (2013). Minireview: Signal Bias, Allosterism, and Polymorphic Variation at the GLP-1R: Implications for Drug Discovery. *Molecular Endocrinology*, DOI: 10.1210/me.2013-1116.
- (8) Athauda, D. y Foltynie, T. (2016). The glucagon-like peptide 1 (GLP) receptor as a therapeutic target in Parkinson's disease: mechanisms of action. *Drug discovery today* 21, 802-18.
- (9) Matsuura, K., Canfield, K., Feng, W. y Kurokawa, M. en *International Review of Cell and Molecular Biology*, Jeon, K. y Galluzzi, L., eds.; Academic Press: San Diego, US, 2016, págs. 43-87.
- (10) Gross, A. y Katz, S. (2017). Non-apoptotic functions of BCL-2 family proteins. *Cell Death and Differentiation* 24, 1348-1358.

- (11) Ljubcic, S., Polak, K., Fu, A., Wiwczar, J., Szlyk, B., Chang, Y., Alvarez-Perez, J. C., Bird, G. H., Walensky, L. D., Garcia-Ocaña, A. y Danial, N. N. (2015). Phospho-BAD BH3 mimicry protects β cells and restores functional β cell mass in diabetes. *Cell Reports* 10, 497-504.
- (12) Danial, N. N. y col. (2008). Dual role of proapoptotic BAD in insulin secretion and beta cell survival. *Nature Medicine* 14, 144-153.
- (13) Liang, Y. L., Belousoff, M. J., Zhao, P., Koole, C., Fletcher, M. M., Truong, T. T., Julita, V., Christopoulos, G., Xu, H. E., Zhang, Y., Khoshouei, M., Christopoulos, A., Danev, R., Sexton, P. M. y Wootten, D. (2020). Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Molecular Cell* 77, 656-668.e5.
- (14) Cao, C., Zhang, H., Yang, Z. y Wu, B. (2018). Peptide recognition, signaling and modulation of class B G protein-coupled receptors. *Current Opinion in Structural Biology* 51, 53-60.
- (15) de Graaf, C., Song, G., Cao, C., Zhao, Q., Wang, M. W., Wu, B. y Stevens, R. C. (2017). Extending the Structural View of Class B GPCRs. *Trends in Biochemical Sciences* 42, 946-960.
- (16) Zhang, Y., Sun, B., Feng, D., Hu, H., Chu, M., Qu, Q., Tarrasch, J. T., Li, S., Sun Kobilka, T., Kobilka, B. K. y Skiniotis, G. (2017). Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature*, DOI: 10.1038/nature22394.
- (17) Liang, Y. L. y col. (2018). Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor-Gs complex. *Nature*, DOI: 10.1038/nature25773.
- (18) Song, G. y col. (2017). Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. *Nature* 546, 312-315.
- (19) Jazayeri, A. y col. (2017). Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature*, DOI: 10.1038/nature22800.
- (20) Bueno, A. B., Showalter, A. D., Wainscott, D. B., Stutsman, C., Marín, A., Ficorilli, J., Cabrera, O., Willard, F. S. y Sloop, K. W. (2016). Positive Allosteric Modulation of the Glucagon-like Peptide-1 Receptor by Diverse Electrophiles. *The Journal of biological chemistry* 291, 10700-15.
- (21) Latorraca, N. R., Venkatakrishnan, A. J. y Dror, R. O. (2017). GPCR dynamics: Structures in motion. *Chemical Reviews* 117, 139-155.
- (22) Teng, M. y col. (2007). Small molecule ago-allosteric modulators of the human glucagon-like peptide-1 (hGLP-1) receptor. *Bioorganic and Medicinal Chemistry Letters*, DOI: 10.1016/j.bmcl.2007.06.086.
- (23) Nolte, W. M., Fortin, J. P., Stevens, B. D., Aspnes, G. E., Griffith, D. A., Hoth, L. R., Ruggeri, R. B., Mathiowetz, A. M., Limberakis, C., Hepworth, D. y Carpino, P. A. (2014). A potentiator of orthosteric ligand activity at GLP-1R acts via covalent modification. *Nature Chemical Biology* 10, 629-631.
- (24) Kooistra, A., Mordalski, S., Pandey-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keseru, G. y Gloriam, D. (2021). GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Research* 49, D335-D343.

- (25) Redij, T., Ma, J., Li, Z., Hua, X. y Li, Z. (2019). Discovery of a potential positive allosteric modulator of glucagon-like peptide 1 receptor through virtual screening and experimental study. *Journal of Computer-Aided Molecular Design*, DOI: 10.1007/s10822-019-00254-4.
- (26) Deganutti, G., Moro, S. y Reynolds, C. A. (2019). Peeking at G-protein-coupled receptors through the molecular dynamics keyhole. *Future Medicinal Chemistry* 11, 599-615.
- (27) Liao, C., May, V. y Li, J. en *G Protein-Coupled Receptor Signaling*, Tiberi, M., ed.; Humana Press: New York, USA, 2019, págs. 3-19.
- (28) Bowman, G. R. en *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G., Pande, V. y Noe, F., eds.; Springer: New York, USA, 2014, págs. 7-22.
- (29) Beichelt, F., *Applied Probability and Stochastic Processes*, 2.^a ed.; CRC Press: Florida, US, 2016.
- (30) Sarich, M., Prinz, J. y Schütte, C. en *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G., Pande, V. y Noé, F., eds.; Springer: New York, USA, 2014, págs. 23-44.
- (31) Horn, R. y Johnson, C., *Matrix Analysis*, 2.^a ed.; Cambridge University Press: New York, USA, 2013.
- (32) Serre, D., *Matrices. Theory and Applications*, 1.^a ed.; Springer: New York, USA, 2002.
- (33) Wu, H. y Noe, F. (2020). Variational Approach for Learning Markov Processes from Times Series Data. *Journal of Nonlinear Science* 30, 23-66.
- (34) Noé, F. y Prinz, J. en *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G., Pande, V. y Noé, F., eds.; Springer: New York, USA, 2014, págs. 75-90.
- (35) Roebnitz, S. y Weber, M. (2013). Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classif.* 7, 147-179.
- (36) Metzner, P. y Vanden-Eijnden, E. (2009). Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* 7, 1192-1219.
- (37) Vanden-Eijnden, E. en *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G., Pande, V. y Noé, F., eds.; Springer: New York, USA, 2014, págs. 91-100.
- (38) Laskowski, R., MacArthur, M. W., Moss, D. y Thornton, J. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.
- (39) Laskowski, R., MacArthur, M., Smith, D., Jones, D., Hutchinson, E., Morris, A., Moss, D. y Thornton, J. PROCHECK v.3.5. Programs to check the Stereochemical Quality of Protein Structures. Operating Manual; European Bioinformatics Institute, EMBL-EBI, Cambridgeshire, UK.
- (40) Case, D., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham, T., Cruzeiro, V., Darden, T., Duke, R., Ghoreishi, D., Giambasu, G., Giese, T., Gilson, M. y Gholke, H. AMBER 2019, ver. 2019, San Fransisco, USA, 2019.

- (41) Schott-Verdugo, S. y Gohlke, H. (2019). PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *Journal of Chemical Information and Modeling* 59, 2522-2528.
- (42) Nugent, T. y Jones, D. (2013). Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics* 14, 276.
- (43) Maier, J., Martinez, C., K., K., Wickstrom, L., Hauser, K. E. y Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* 11, 3696-3713.
- (44) Dickson, C., Madej, B., Skjevik, A., Betz, R., Teigen, K., Gould, I. y Walker, R. (2014). Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* 10, 865-879.
- (45) Wang, J., Wolf, R., Caldwell, J., Kollman, P. y Case, D. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157-74.
- (46) Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., Wehmeyer, C., Prinz, J.-H. y Noé, F. (2015). PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* 11, 5525-5542.
- (47) Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritis, G. y Noe, F. (2013). Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139, 015102.
- (48) Trendelkamp-Shroer, B., Wu, H., Paul, F. y Noe, F. (2015). Estimation and uncertainty of reversible Markov models. *Journal of Chemical Physics* 143, 143.
- (49) Grossfield, A. WHAM.
- (50) Shen, M.-Y. y Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science* 15, 2507-2524.
- (51) Sali, A. MODELLER A Program for Protein Structure Modeling Release 8v1; University of California San Francisco, San Francisco, USA, 2005.
- (52) Zhang, H., Yin, C., Jiang, Y. y van der Spoel, D. (2018). Force Field Benchmark of Amino Acids I: Hydration and Diffusion in Different Water Models. *Journal of Chemical Information and Modeling* 58, 1037-1052.
- (53) Zhang, H., Jiang, Y., Cui, Z. y Yin, C. (2018). Force Field Benchmark of Amino Acids. 2. Partition Coefficients between Water and Organic Solvents. *Journal of Chemical Information and Modeling* 58, 1669-1681.
- (54) Kamenik, A. S., Handle, P., Hofer, F., Kahler, J., U. Krami y Liedl, K. (2020). Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding. *The Journal of Chemical Physics* 153, 185102.
- (55) Montgomery, D., Jennings, C. L. y Kulahci, M., *Introduction to Time Series Analysis and Forecasting*; Wiley-Interscience: New Jersey, USA, 2008.
- (56) Zhang, X. y col. (2020). Differential GLP-1R Binding and Activation by Peptide and Non-peptide Agonist. *Molecular Cell* 80, 485-500.

-
- (57) Munk, C., Mutt, E., Isberg, V., Nikolajsen, L., Bibbe, J., Flock, T., Hanson, M., Stevens, R., Deupi, X. y Gloriam, D. (2019). An online resource for GPCR structure determination and analysis. *Nature Methods* 16, 151-162.
- (58) Halford, B. (2020). Covalent drugs go from fringe field to fashionable endeavor. *Chemical & Engineering News* 98.
- (59) De Vita, E. (2020). 10 years into the resurgence of covalent drugs. *Future Medicinal Chemistry* 13.
- (60) Awoonor-Williams, E., Walsh, A. G. y Rowley, C. N. (2017). Modeling covalent-modifier drugs. *BBA - Proteins and Proteomics* 1865, 1664-1675.
- (61) Blanco, G., Goodsell, D. S. y Forli, S. (2020). Selective and Effective: Current Progress in Computational Structure-Based Drug Discovery of Targeted Covalent Inhibitors. *Trends in Pharmacological Sciences* 41, 1038-1049.
- (62) Hsing, T. y Eubank, R., *Theoretical Foundations of Functional Data Analysis, with an INtroduction to Linear Operators*; Wiley: Chichester, UK, 2015.

Me siento junto al fuego y pienso
en todo lo que he visto,
en flores silvestres y mariposas
de veranos que han sido.

En hojas amarillas y telarañas,
en otoños que fueron,
la niebla en la mañana, el sol de plata,
y el viento en mis cabellos.

Me siento junto al fuego y pienso
cómo el mundo será,
cuando llegue el invierno sin una primavera
que yo pueda mirar.

Pues hay todavía tantas cosas
que yo jamás he visto:
en todos los bosques y primaveras
hay un verde distinto.

Me siento junto al fuego y pienso
en las gentes de ayer,
y en gentes que verán un mundo
que no conoceré.

Y mientras estoy aquí sentado
pensando en otras épocas
espero oír unos pasos que vuelven
y voces en la puerta.

El Señor de los Anillos, La Comunidad del Anillo
J. R. R. Tolkien