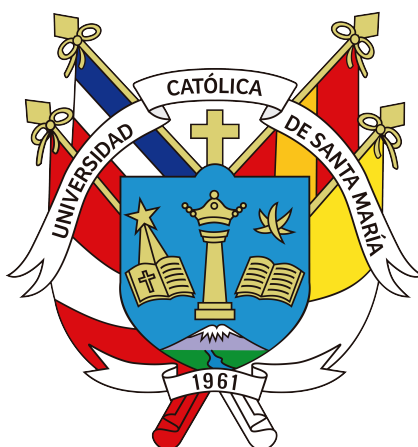


Universidad Católica de Santa María
Facultad de Ciencias Farmacéuticas, Bioquímicas y
Biotecnológicas
Escuela Profesional de Ingeniería Biotecnológica



Diseño y Construcción de la Estructura Completa del
Receptor de Insulina Humana Mediante Métodos
Computacionales

Tesis presentada por la Bachiller:

Valderrama Zamudio, Lesly Estefanía

ORCID: 0000-0001-8184-4034

para optar el Título Profesional de Ingeniera Biotecnóloga

Asesor:

Dr. Gómez Valdez, Badhin

ORCID: 0000-0001-6539-1207

**Arequipa – Perú
2025**

UCSM-ERP

UNIVERSIDAD CATÓLICA DE SANTA MARÍA

INGENIERIA BIOTECNOLOGICA

TITULACIÓN CON TESIS

DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 03 de Noviembre del 2023

Dictamen: 002990-C-EPIB-2023

Visto el borrador del expediente 002990, presentado por:

2015102002 - VALDERRAMA ZAMUDIO LESLY ESTEFANÍA

Titulado:

**DISEÑO Y CONSTRUCCIÓN DE LA ESTRUCTURA COMPLETA DEL RECEPTOR DE INSULINA
HUMANA MEDIANTE MÉTODOS COMPUTACIONALES**

Nuestro dictamen es:

APROBADO

Titulo Profesional/Titulo de Segunda Especialidad/Grado Académico a optar:

INGENIERO BIOTECNOLOGO

**29389056 - PAZ ALIAGA CARLOS EITEL IVAN
DICTAMINADOR**



**41404327 - BARREDA DEL CARPIO JAIME ERNESTO
DICTAMINADOR**



**44361522 - BARAZORDA CCAHUANA HARUNA LUZ
DICTAMINADOR**



Diseño y Construcción de la Estructura Completa del Receptor de Insulina Humana Mediante Métodos Computacionales

INFORME DE ORIGINALIDAD

14%

INDICE DE SIMILITUD

10%

FUENTES DE INTERNET

12%

PUBLICACIONES

11%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Vikas Dubey, Behruz Bozorg, Daniel Wüstner, Himanshu Khandelia. "Cholesterol binding to the sterol-sensing region of Niemann Pick C1 protein confines dynamics of its N-terminal domain", PLOS Computational Biology, 2020 Publicación	<1 %
2	repositorio.uo.edu.cu Fuente de Internet	<1 %
3	Submitted to Imperial College of Science, Technology and Medicine Trabajo del estudiante	<1 %
4	Submitted to Trinity College Dublin Trabajo del estudiante	<1 %
5	idoc.pub Fuente de Internet	<1 %
6	Submitted to University of Newcastle upon Tyne Trabajo del estudiante	<1 %

Dedicatoria

Quiero comenzar dedicando este trabajo a Dios, quien siempre ha sido mi soporte emocional y espiritual en los momentos difíciles. Pero sobretodo a mis padres, Martín y Luz, quienes me apoyado en todas mis decisiones y, a pesar de las dificultades, me han permitido alcanzar mis sueños y cultivar mi confianza. Los amo de aquí hasta las estrellas.



Agradecimientos

Quiero agradecer a Dios, a mis padres y mi familia por nunca dejarme sola y apoyarme día a día en mis decisiones, sin ustedes esto seguiría siendo un sueño. A la Universidad Católica de Santa María, institución donde llevé a cabo mis estudios durante 5 años. Destaco aquí mis agradecimientos al Centro de Investigación en Ingeniería Molecular, lugar en el que desarrollé mis habilidades como investigadora y me permitió trabajar en esta tesis. En especial quiero agradecer al PhD. Badhin Gomez por acogerme en el centro, ser un excelente mentor y darme la oportunidad de tener una gran familia académica, llena de aprendizajes y oportunidades. Mis agradecimientos al PhD. Jorge Aguilar por sus incalculables aportes y apoyo durante el desarrollo de mi tesis, y sobretodo a Pamela Gamero, quien me invitó a conocer el mundo de la biología computacional y ha sido un apoyo confiable en todo momento. Por último, quiero agradecer a mis amigas más cercanas Fabi, Nicol, Joanne y Vale sin quienes estos años hubieran sido muy difíciles. Gracias por ser ese apoyo constante, por compartir sus conocimientos y por los grandes recuerdos que hemos creado. A todos mis compañeros, ¡gracias!, porque es imposible mencionar a todos los que de alguna forma me han permitido crecer en la persona que soy hoy.

Resumen

El receptor de insulina (IR) humana pertenece a la familia de los receptores tirosina quinasa, siendo la Isoforma B de este receptor la más abundante en los tejidos periféricos y la que presenta mayor afinidad por la insulina. Esto último hace que dicha isoforma se encuentre más involucrada en los efectos metabólicos. Por lo que el objetivo del presente trabajo de investigación fue diseñar y construir la estructura completa del receptor de insulina Isoforma B mediante el uso de métodos computacionales, considerando que estos últimos contribuyen al desarrollo de estructuras viables cuando los métodos experimentales no son suficientes. Esto ha sido observado con el IR, principalmente, pero no limitado a, su segmento transmembrana y juxtamembrana, los cuales presentan fragmentos de su secuencia sin estructura predicha. Es así que, este estudio se ha realizado de forma teórica utilizando servidores basados en métodos de predicción por comparación y *ab initio* para obtener los fragmentos faltantes de su estructura. El dímero completo del receptor se ensambló mediante el uso de los visualizadores Chimera y Pymol, y la adición de N-glicanos al mismo se realizó con doGlycans. La dinámica para la optimización de la estructura final se realizó con Gromacs 2019.2 durante 100 ns usando una temperatura de 36.5°C y 1 bar de presión. Finalmente, el uso de diferentes métodos para la evaluación de la estructura permitió concluir que se requiere de un mayor tiempo de dinámica molecular para confirmar que la estructura se encuentra en su estado nativo. No obstante, la estructura obtenida en formato .pdb puede ser utilizada para futuras dinámicas.

Palabras clave: Dinámica molecular, estructura, insulina

Abstract

The human insulin receptor (IR) belongs to the family of tyrosine kinases receptors, with Isoform B of this receptor being the most abundant in peripheral tissues and the one with the highest affinity for insulin. The latter means that said isoform is more involved in metabolic effects. Therefore, the objective of this research work is to design and build the complete structure of the insulin receptor isoform B through the use of computational methods, considering that the latter contribute to the development of viable structures when experimental methods are not sufficient. This has been observed with the IR, mainly, but not limited to, its transmembrane and juxtamembrane segments, which present fragments of its sequence with no predicted structure. Thus, this study has been carried out theoretically using servers based on prediction methods by comparison and *ab initio* to obtain the missing fragments of its structure. The complete receptor dimer was assembled using Chimera and Pymol viewers, and the addition of N-glycans to it was done with doGlycans. Dynamics for final structure optimization was performed with Gromacs 2019.2 for 100 ns using a temperature of 36.5°C and 1 bar pressure. Finally, the use of different methods for the evaluation of the structure allowed us to conclude that a longer molecular dynamics time is required to confirm that the structure is in its native state. However, the structure obtained in .pdb format can be used for future dynamics.

Keywords: Molecular dynamics, structure, insulin

Índice

Dedicatoria	I
Agradecimientos	II
Resumen	III
Abstract	IV
Índice	V
Índice de figuras	VIII
Índice de tablas	X
Glosario	XI
Introducción	1
Hipótesis	3
Objetivos	4
1. Marco Teórico	5
1.1. Receptor de insulina	5
1.1.1. Estructura	6
Ectodominio	7
Segmento transmembrana	8
Componente intracelular	9
Puentes disulfuro	9
Glicanos asociados	11
1.2. Bioinformática y Dinámica Molecular	13
1.2.1. Selección del campo de fuerza	16
1.2.2. Diseño del sistema de trabajo	17

Ensamblados	17
Potencial de Interacción	19
Algoritmos de Integración	21
Parámetros de los archivos	22
1.2.3. Predicción de estructuras de proteínas	22
1.2.4. Predicción por comparación	23
1.2.5. Predicción <i>de novo</i>	24
Análisis de estructuras proteicas	24
2. Metodología	28
2.1. Lugar de ejecución	28
2.2. Detalles Computacionales	28
2.2.1. Hardware	28
2.2.2. Softwares	28
Gromacs 2019.2	28
UCSF Chimera Alpha Linux	29
VMD	29
MEGA X	29
PyMOL	30
DoGlycans	30
2.2.3. Servidores Web	30
NCBI	30
UniProt	31
Protein Data Bank	31
SWISS-MODEL	31
PEP-FOLD3	32
Ab Initio Domain Assembly Server (AIDA)	32
3DRefine	32
CHARMM-GUI	33
Quark Online	33
GlyProt	33
MolProbity	34
2.3. Métodos	34
2.3.1. Obtención de las secuencias de interés	34
2.3.2. Construcción de segmentos	34
2.3.3. Optimización con una bicapa lipídica	38
2.3.4. Descripción de enlaces	42
2.3.5. Construcción del dímero	43

2.3.6. Adición de glicanos	44
2.3.7. Optimización de la estructura completa del IR	44
2.3.8. Evaluación de la estructura terciaria optimizada	45
2.4. Diagrama de flujo de actividades	45
3. Resultados y Discusión	47
3.0.1. Búsqueda y selección de dominios existentes	47
3.0.2. Construcción del segmento transmembrana y juxtamembrana .	51
3.0.3. Optimización del segmento transmembrana y juxtamembrana .	54
3.0.4. Construcción del ectodominio	56
3.0.5. Construcción del segmento intracelular	58
3.0.6. Ensamblado de la estructura terciaria	59
Construcción del monómero	59
Construcción del dímero	61
Adición de glicanos	62
3.0.7. Optimización del Receptor de Insulina	66
3.0.8. Análisis de la estructura optimizada del Receptor de Insulina . .	71
Raíz de la Desviación Media Cuadrática (RMSD)	71
Raíz de la Fluctuación Media Cuadrática (RMSF)	72
Radio de giro	74
Enlaces de Hidrógeno	74
Diagrama de Ramachandran	75
Movimiento e influencia de los glicósidos	84
Conclusiones	88
Recomendaciones	90
Referencias Bibliográficas	91

Índice de figuras

1.1. Cambio conformacional del receptor de insulina en presencia de un ligando	8
1.2. Dominios en el receptor de insulina	9
1.3. Segmento transmembrana y componente intracelular	10
1.4. Puente disulfuro entre cisteínas	11
2.1. Interfaz de AIDA	36
2.2. Cuadro de diálogo de la opción Build Structure en UCSF Chimera.	38
2.3. Configuración de la orientación de la proteína en la membrana	40
2.4. Configuración de la distribución de lípidos en la bicapa lipídica	41
2.5. Diagrama de flujo de actividades	46
3.1. Alineamiento de las secuencias del ectodominio con IR completa.	50
3.2. Alineamiento de las secuencias del segmento intracelular.	51
3.3. Alineamiento del segmento transmembrana con el IR y ectodominio seleccionado.	52
3.4. Modelo obtenido con Swiss-Model	52
3.5. Construcción del dominio transmembrana y juxtamembrana	53
3.6. Calificación de los modelos en 3DRefine	54
3.7. Alineamiento del segmento transmembrana en la bicapa lipídica	54
3.8. Tamaño del sistema que incluye el segmento transmembrana y juxtamembrana embebido en una bicapa lipídica	55
3.10. Parámetros para la optimización y equilibrado de la proteína en la membrana	56
3.9. Construcción de la bicapa lipídica	56
3.11. Construcción del ectodominio	57
3.12. Construcción del segmento intracelular	59
3.13. Modelado de enlaces disulfuro	60
3.14. Diferencia en la protonación de los residuos de histidina	62
3.15. Alineamiento de monómeros al ectodominio 6SOF	63
3.16. Construcción del dímero del Receptor de Insulina	64

3.17. Búsqueda de glicanos en Glyprot	65
3.18. Sistema construido para la optimización	68
3.19. Parámetros registrados en la Dinámica Molecular	70
3.20. Estructura del Receptor de Insulina optimizada	70
3.21. RMSD del Receptor de Insulina tras 100 ns	71
3.22. RMSF del Receptor de Insulina tras 100 ns	73
3.23. Análisis del factor b usando superficies	73
3.24. Radio de giro del Receptor de Insulina tras 100 ns	74
3.25. Enlaces de hidrógenos presentes durante la simulación del Receptor de Insulina	75
3.26. Gráfica de Ramachandran Genérica del Receptor de Insulina después de la dinámica de 100 ns	76
3.27. Estructura del IR resaltando en rojo los residuos ubicados en zonas no permitidas de la Gráfica de Ramachandran genérica	77
3.28. Comparación del ectodominio del IR usado como molde y el obtenido después de la simulación	79
3.29. Gráfica de Ramachandran de Glicina del Receptor de Insulina después de la dinámica de 100 ns	80
3.30. Gráficas de Ramachandran de Prolina del Receptor de Insulina después de la dinámica de 100 ns	81
3.31. Gráfica de Ramachandran de pre-Prolina del Receptor de Insulina después de la dinámica de 100 ns	82
3.32. Gráfica de Ramachandran de Isoleucina y Valina del Receptor de Insulina después de la dinámica de 100 ns	83
3.33. Estructura del IR glicosilada después de la optimización	85
3.34. Análisis de la influencia de los glicósidos en la estructura del IR	86

Índice de tablas

3.1. Alineamiento de las secuencias FASTA de las 2 isoformas del IR	48
3.2. Glicósidos más abundantes en cada residuo glicosilado del IR	65



Glosario

1. aa: aminoácido(s)
2. AMBER: campo de fuerza ampliamente usado que contiene la información para todos los aminoácidos y nucleótidos comunes.
3. ARG: arginina
4. ASN: asparagina
5. CHARMM: Chemistry at Harvard Macromolecular Mechanics, conjunto de campos de fuerza y paquete de software de análisis computacional asociado con los campos de fuerza del mismo nombre.
6. CIIM: Centro de Investigación en Ingeniería Molecular
7. CYS: cisteína
8. DPPC: dipalmitoilfosfatidilcolina
9. GDT-HA: Global Distance Test - High Accuracy (Prueba de Distancia Global - Alta precisión)
10. GDT-TS: Global Distance Test - Total Score (Prueba de Distancia Global - Puntaje total)
11. GLU: ácido glutámico
12. HIS: histidina
13. IGF: Factor de crecimiento insulínico (IGF, por sus siglas en inglés “Insulin-like Growth Factor”)
14. ILE: isoleucina
15. IR: Receptor de Insulina (IR, por sus siglas en inglés “Insulin Receptor”)

16. IRS: Sustrato del Receptor de Insulina (IRS, por sus siglas en inglés “Insulin Receptor Substrate”)
17. kDa: kilodalton
18. LEU: leucina
19. LYS: lisina
20. Na: Sodio
21. NMR: Resonancia Magnética Nuclear (NMR, por sus siglas en inglés “Nuclear Magnetic Resonance”)
22. ns: nanosegundo
23. OPLS: campo de fuerza con “Potenciales optimizados para simulaciones de líquidos” (OPLS, por sus siglas en inglés “Optimized Potentials for Liquid Simulations”)
24. ps: picosegundos
25. SER: serina
26. sOPEP: energía *coarse grained* de PEP-FOLD
27. TB: terabyte
28. THR: treonina
29. TYR: tirosina
30. VAL: valina

Introducción

El receptor de insulina humana posee la estructura característica de los receptores de la superfamilia Tirosina Quinasa a la cual pertenece. Estos presentan un ectodominio extracelular N-terminal, un segmento transmembrana y un dominio intracelular en el que se encuentra la enzima tirosina quinasa, finalizando en una cola no estructurada¹.

La familia de los receptores de insulina resaltan principalmente por encontrarse covalentemente unidos en su estado dimérico en la superficie celular². Esto es posible gracias a los puentes disulfuro que se forman entre sus dos cadenas homodiméricas desde su síntesis. Además, su ectodominio es muy complejo y sufre cambios conformacionales en presencia de sus ligandos para poder activar a la enzima tirosina quinasa. En este proceso es importante resaltar el rol del segmento juxtamembrana, el cual influye en la activación de la quinasa pues presenta roles inhibitorios y activadores³.

Como lo indica Diwanji et al.³ *Una estructura de alta resolución de la estructura completa de un receptor de la familia de receptores de insulina sería invaluable para revelar los mecanismos adicionales que controlan la activación dependiente del ligando de esta clase de receptores Tirosina Quinasa.* Por ello, su estudio a nivel computacional ayudará a obtener una posible estructura tridimensional que pueda ser utilizada en ensayos teóricos para determinar el mecanismo por el cuál el receptor de insulina puede ser activado.

Actualmente, la estructura completa del IR no ha podido ser dilucidada por los métodos experimentales convencionales. Por lo que se ha recurrido al uso de tecnologías computacionales, las cuales han permitido obtener fragmentos más completos. No obstante, al realizar una investigación de la literatura es posible apreciar que el uso de este tipo de tecnologías para determinar la estructura completa del receptor insulínico es casi nulo. Por lo que el objetivo de este estudio es obtener la estructura tridimensional

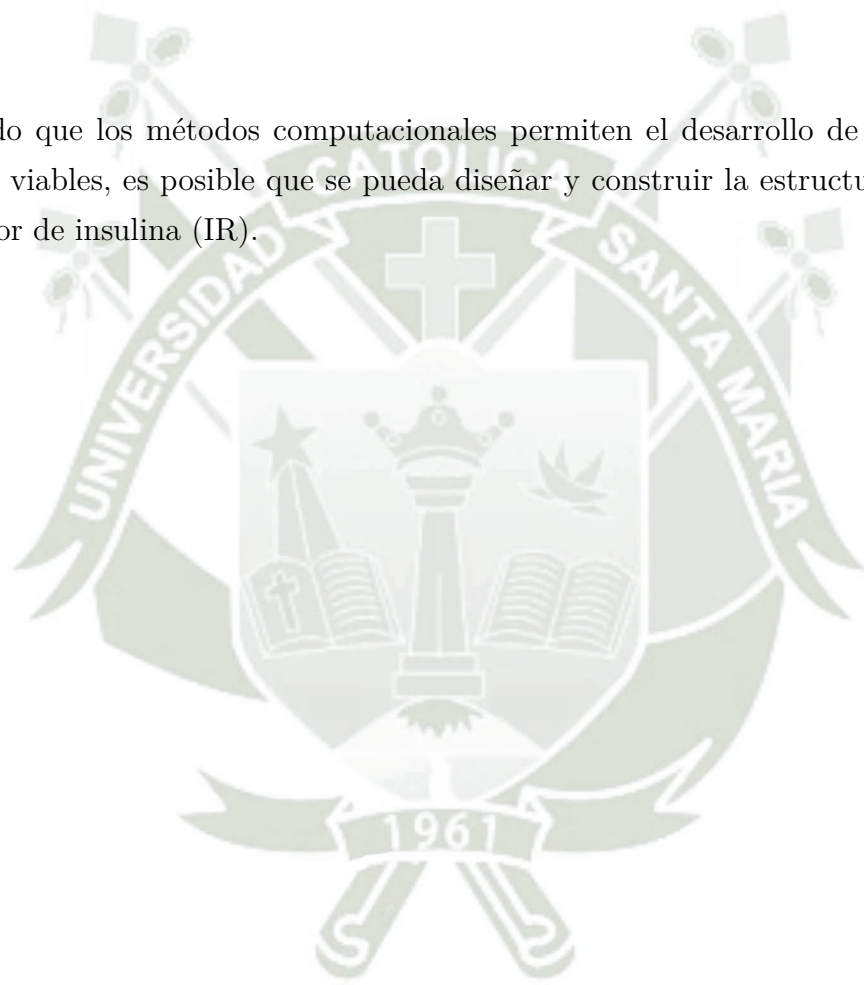
del receptor de insulina humana Isoforma B usando métodos computacionales como la predicción por homología y predicción *de novo*. Lo que hace que mi hipótesis sea dado que los métodos computacionales permiten el desarrollo de estructuras terciarias viables es posible que se pueda diseñar y construir la estructura completa del Receptor de Insulina (IR).

Siendo así, la metodología a usar consistió en obtener la secuencia FASTA completa del receptor de insulina humano Isoforma B para luego determinar los segmentos sin estructura secundaria o terciaria dilucidada mediante la ayuda de la base de datos Protein Data Bank. Una vez realizado esto se escogió el mejor método computacional para construir la estructura de cada fragmento y optimizar la misma. Los fragmentos construidos se aunaron a los ya existentes y se procedió con la construcción del receptor completo como dímero declarando los enlaces disulfuro correspondientes. A la estructura construida se le añadieron los glicanos más abundantes y se procedió con la optimización para finalmente obtener la estructura completa.



Hipótesis

Dado que los métodos computacionales permiten el desarrollo de estructuras terciarias viables, es posible que se pueda diseñar y construir la estructura completa del receptor de insulina (IR).



Objetivos

Objetivo General

- Diseñar y construir la estructura completa del receptor de insulina (IR) mediante métodos computacionales.

Objetivos Específicos

- Obtener la secuencia FASTA completa del IR.
- Construir el segmento transmembrana utilizando métodos computacionales.
- Desarrollar la estructura tridimensional completa del IR.
- Estabilizar la estructura completa del receptor insulínico IR, en un ensamble NPT.
- Evaluar y reportar la estructura terciaria completa del IR.

Capítulo 1

Marco Teórico

1.1. Receptor de insulina

El receptor de insulina (IR), como lo indican Scapin et al.⁴, es un dímero formado por dos cadenas heterodiméricas y está representado por la abreviación $(\alpha\beta)_2$. Se le considera así porque cada monómero está conformado por dos subunidades, una de estas es la α que se encuentra localizada en el ectodominio, y la otra es la subunidad β que tiene 194 residuos de su lado amino-terminal en el ectodominio, seguido por el segmento transmembrana de 23 aa y dominio intracelular o citoplasmático de 403 aa, este último conformado por una enzima tirosina quinasa⁵.

Precisamente por la presencia de esta última es que el IR es considerado como parte de la familia de los receptores tirosina quinasa (RTK, Receptor tyrosine kinases), los cuales se dividen a su vez en 19 subfamilias con diferentes funciones pero estructuras similares. Todos los RTK van a presentar en el lado N-terminal un ectodominio muy diverso en tamaño y composición entre subfamilias, un único dominio transmembrana y un dominio quinasa intracelular muy conservado que termina en una región C-terminal larga sin estructura determinada¹.

El gen que se encarga de la producción del IR monomérico se encuentra en el cromosoma 19, exactamente en las bandas p13.2 y p13.3⁶ y presenta 22 exones. Este es capaz de traducirse a una proteína de 1382 y de 1370 aminoácidos según como se exprese el exon 11 generando así 2 isoformas que serán descritas posteriormente. Cuando la proteína es expresada existirá al inicio del lado amino terminal una secuencia señal de

27 aminoácidos, lo que hace que la proteína tenga que pasar procesos de glicosilación y formación de enlaces disulfuros antes de llegar a la superficie de la célula. Cuando esta pierde la secuencia señal por un proceso de rotura proteolítica y queda insertada en la membrana se dice que se cuenta con un receptor maduro⁷. El precursor de la proteína se sintetiza en el Retículo Endoplasmático y en este mismo organelo se formará el dímero de esta. El procesamiento posterior, como la adición de glicanos se dará tanto en el ER como en el Aparato de Golgi. Por último, la rotura que dividirá ambas subunidades se dará en la Red del Trans Golgi⁸.

Se mencionó brevemente que el receptor insulínico cuenta con 2 isoformas, denominadas A o corta y B o larga. Esto es causado por un empalme alternativo del exón 11, haciendo que la isoforma B cuente con 12 residuos de más al final del extremo C-terminal de la subunidad *alpha*⁹ en la posición 717 del receptor maduro. Esta pequeña inserción causa efectos en la afinidad del receptor, haciendo que la isoforma A tenga mayor afinidad por el factor de crecimiento insulínico (IGF) tipo I y II¹⁰. En cuanto a la insulina, la isoforma B presentara una mayor afinidad por esta sin llegar a ser muy significativa la diferencia. No obstante, se considera a esta última como la más específica y por ende, la que se encuentra mas involucrada en los efectos metabólicos¹¹. Por otro lado, en algunos estudios se ha podido observar la presencia de receptores de insulina híbridos que forman un dímero con una cadena heterodimérica de cada isoforma¹² o incluso, con una cadena del receptor de IGF-1¹³.

Asimismo, cabe resaltar que no todos los tejidos van a contar con la misma proporción de IR-A e IR-B sino que habrá una mayor expresión del primero en células indiferenciadas, células que contribuyen en el crecimiento de tejidos, principalmente tejidos fetales, y en el cerebro, donde además se ha observado que en las neuronas solo se puede encontrar esta isoforma. El IR-B estará presente en mayor cantidad en células que ya han pasado por la división celular y se encuentren diferenciadas, ubicándose principalmente en tejidos periféricos^{12,14}

1.1.1. Estructura

La subunidad α , presenta aproximadamente 735 aminoácidos en su isoforma larga con un peso de 135 kDa, mientras que la subunidad β tiene alrededor de 620 aminoácidos y 95 kDa.^{5,15} El número total de aminoácidos se confirmó en 1990 al realizarse la amplificación por reacción en cadena de la polimerasa y secuenciación parcial del receptor insulínico humano.⁶ Dicha secuenciación es la que se mantiene como base en

la actualidad y ha sufrido diferentes anotaciones por otros autores¹⁶⁻¹⁹ a lo largo de todos estos años, siendo su última actualización en la base de datos del NCBI (National Center for Biotechnology Information) en Junio del 2020 y puede ser encontrada con el código de acceso P06213.4.

Es importante notar que las subunidades del receptor de insulina actúan como enzimas alostéricas, ya que las subunidades α van a inhibir la actividad de la enzima tirosina quinasa que se encuentra en la subunidad β . Así, cuando la insulina se acerca a su receptor para activar la vía de la señalización, esta necesita unirse a las dos cadenas del receptor en la superficie, razón por la cual las subunidades α se encuentran unidas mediante un puente disulfuro formando un centro de unión para una sola molécula de insulina. Dicha interacción ocasiona un cambio conformacional en la subunidad β haciendo que la enzima sufra un proceso de activación y por lo tanto, sea capaz de inducir a la autofosforilación de 6 o 7 residuos tirosina de la misma subunidad pero del heterodímero contrario. Este proceso conlleva a que el receptor sea capaz de fosforilar a los sustratos dentro de la célula, conocidos como IRS, que tienen residuos específicos de tirosina, marcando así el inicio de la señalización intracelular.²⁰⁻²².

Este receptor resalta sobre los demás receptores que pertenecen a la subfamilia de RTK porque existen naturalmente en forma dimérica y presentan una característica forma de U invertida cuando no hay ligandos cerca. En dicho estado, los segmentos transmembrana y próximos a esta se encontrarán separados por aproximadamente 11.5 nm. Por el contrario, en presencia de algún ligando que es capaz de provocar un cambio conformacional al unirse al receptor, los segmentos se juntarán y formarán una estructura similar a una T, indicando que el receptor se encuentra activo². Esta diferencia en la estructura se puede observar en la Figura 1.1.

Ectodominio

Solamente en la región extracelular se pueden encontrar 7 dominios. Comenzando desde la HIS1 en el receptor maduro o lado N-terminal se tiene al Dominio Repetido Rico en Leucina (L1) desde el primer residuo hasta el 119, seguidamente está la Región Rica en Cisteína (CR), otro Dominio Repetido Rico en Leucina (L2) desde el residuo 312 a 428 y 2 Dominios de Fibronectina Tipo III (FnIII-1 y FnIII-2). Aquí la FnIII-2 quedará, en cierta forma, dividida en 2 partes, una pertenecerá a la subunidad α , la cual es denominada FnIII-2a y la otra formará parte de la subunidad β (FnIII-2b), no obstante ambos dominios quedarán uno al lado del otro permitiendo la unión

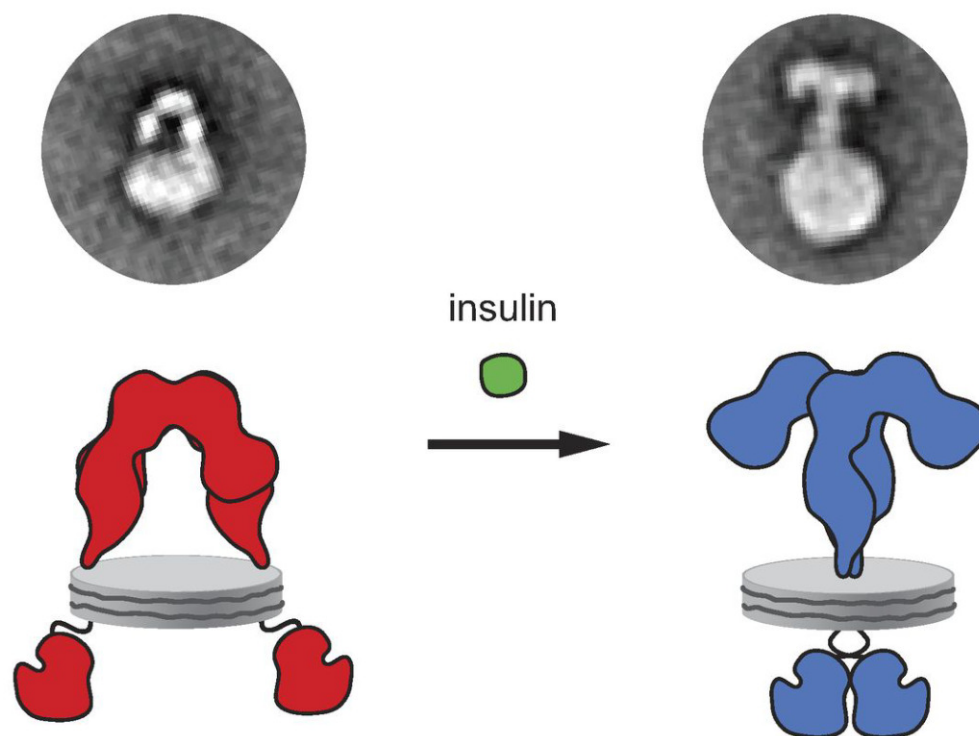


Figura 1.1: Cambio conformacional del receptor de insulina en presencia de un ligando².

entre las subunidades. El dominio de inserción (ID) de 120 aa también constará de dos partes, una porción de este se encontrará terminando la subunidad α ($ID\alpha$) después del FnIII-2a y el otro comenzará la subunidad β ($ID\beta$) antes del FnIII-2b. Finalmente, el ectodominio terminará en el FnIII-3²³⁻²⁵. Todo esto puede ser observado en la Figura 1.2.

Adicionalmente, en un estudio realizado por Sparrow et al.⁷ se observó que la CYS884 estaba presente como un tiol enterrado.

Segmento transmembrana

La secuencia del segmento transmembrana de este receptor presenta una diferencia muy grande respecto a la de otros miembros de la misma familia²⁸. Sin embargo, no existen muchos estudios enfocados en la determinación de su estructura. Se ha predicho que esta es una única α hélice que atraviesa la membrana, lo cual fue comprobado en una investigación realizada por Li et al.²⁹ donde además se observó que el segmento comprendido entre los residuos 942 y 948 ubicados cerca a la membrana, en el lado extracelular, tenía la tendencia de formar una α -hélice pequeña. Dicha estructura se

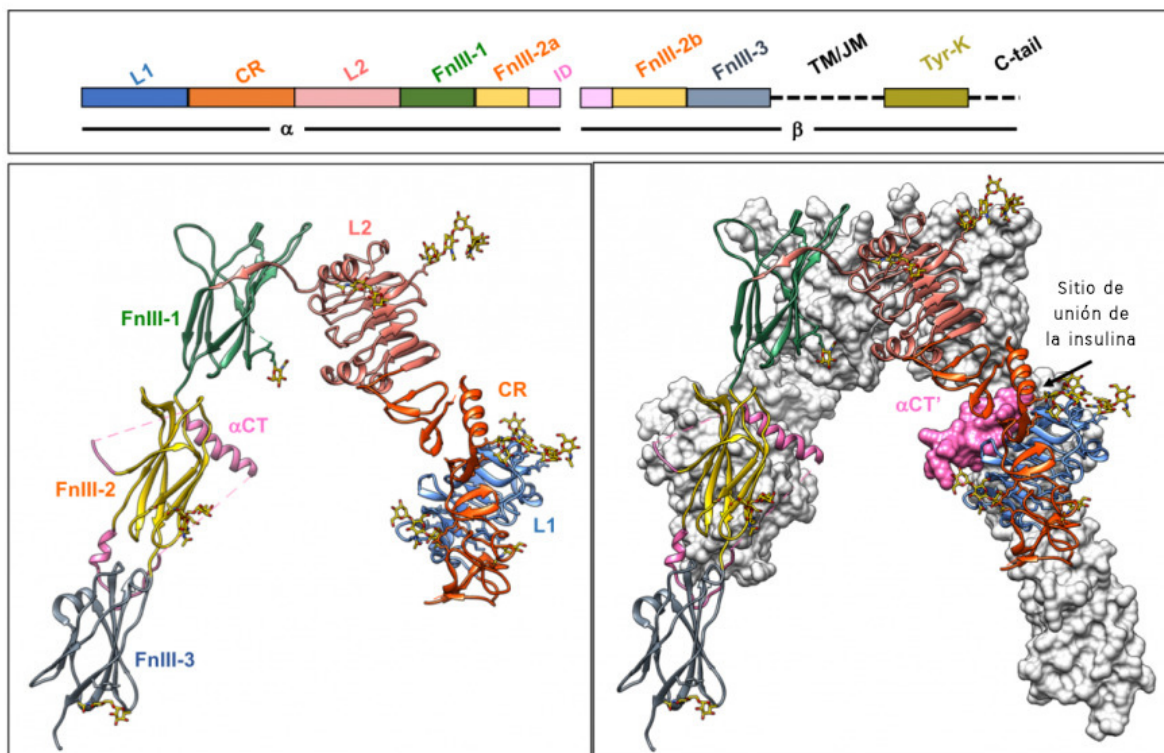


Figura 1.2: Dominios en el receptor de insulina. Los dominios de la estructura PDB ID 4zxb²⁶ se han coloreado para coincidir con la representación lineal. Adicionalmente, se muestran dos representaciones de el ectodominio del IR uno mostrando una representación de superficie donde se ha resaltado en rosado el dominio que participa en la unión de la insulina. Obtenido de: <https://pdb101.rcsb.org/>²⁷

puede visualizar en la Figura 1.3.

Componente intracelular

El componente intracelular va a estar compuesto mayoritariamente por el dominio tirosina quinasa, es decir, una enzima. Esta consta de dos lóbulos, el más grande formado por α -hélices y el que le sigue, en la porción amino terminal, contiene una hoja β de 5 hebras antiparalelas y una α -hélice³⁰. Como ya se ha mencionado anteriormente, no hay una estructura descrita para las secuencia de aminoácidos que se encuentra después del último lóbulo y en la que termina la subunidad β . Una imagen del dominio tirosina quinasa correspondiente al PDB ID 4xlv³¹ puede ser observado en la Figura 1.3.

Puentes disulfuro

Los puentes o enlaces disulfuros se forman por la condensación entre grupos sulfhidrilo. En una proteína usualmente se forman por la presencia de residuos cisteína

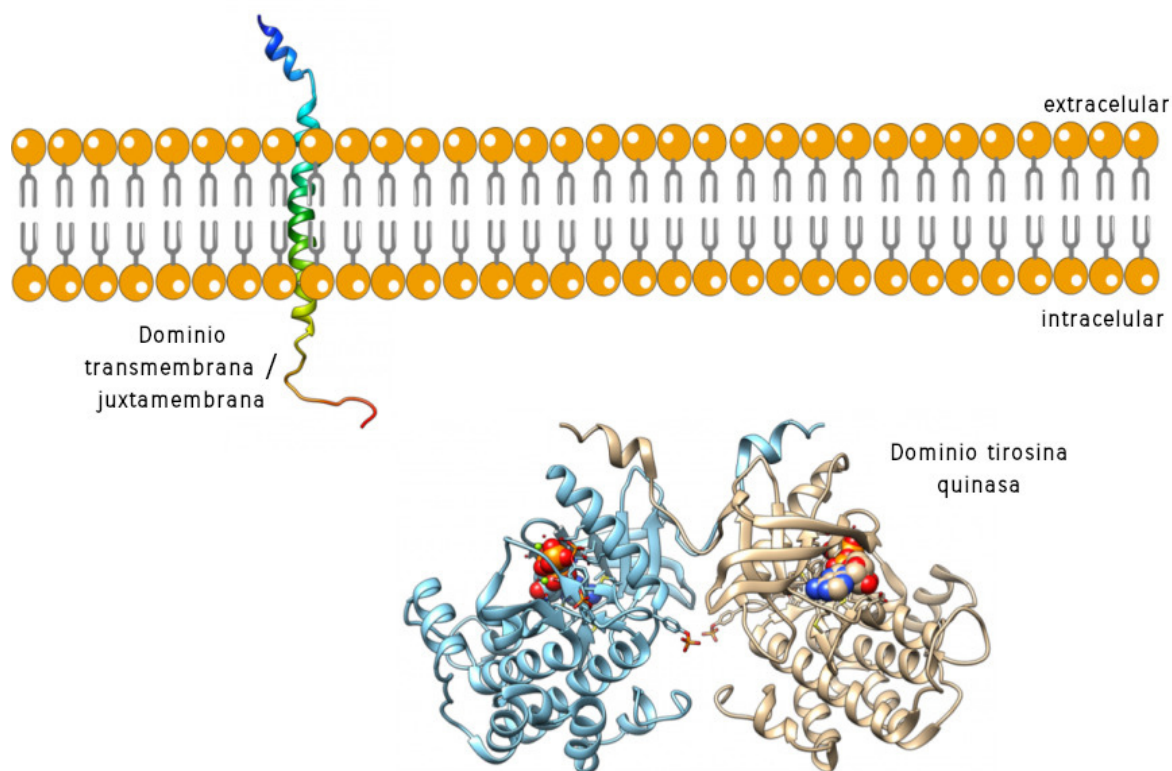


Figura 1.3: Segmento transmembrana y componente intracelular. Obtenido de: <https://pdb101.rcsb.org/>²⁷

ubicados a una distancia cercana³², de aproximadamente 2.04 \AA ³³, un ejemplo de este tipo de enlace lo podemos observar en la Figura 1.4. La finalidad de su formación se suele asociar con un rol estructural, sin embargo, algunos de estos enlaces también pueden estar involucrados en procesos funcionales³⁴. De forma que la rotura de estos enlaces puede ocasionar que se pierda la conformación nativa de la proteína, su función o ambas opciones³³.

A lo largo de los años se han ido describiendo los enlaces disulfuros presentes en el IR, principalmente aquellos que se encargan de unir las subunidades de cada heterodímero y los que permiten la formación del receptor dimérico. En el enlace entre subunidades están involucradas las cisteínas CYS647 (subunidad α) y CYS860 (subunidad β) si hablamos de la isoforma A, y CYS647 (subunidad α) con CYS872 en la isoforma B (subunidad β). Mientras que las cisteínas encargadas de formar el dímero serán CYS524 de una subunidad α con el mismo residuo del otro heterodímero y al menos una de las cisteínas del triplete CYS682, CYS683 y CYS685 con una del otro heterodímero²⁴. No obstante, en un estudio realizado por Lu³⁵ en 1996 se observó una preferencia por la formación del enlace disulfuro entre los residuos CYS682.

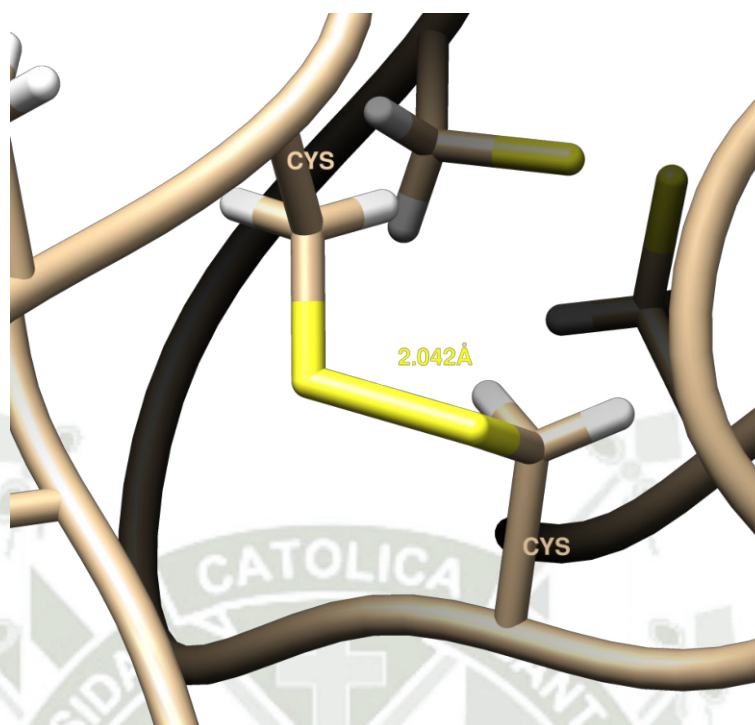


Figura 1.4: Puente disulfuro entre cisteínas.

Se ha identificado que existen aproximadamente 5 veces más residuos cisteína en el ectodominio que en el segmento intracelular⁷ por lo que es lógico asumir que es probable que se formen más puentes disulfuros en la porción que se encuentra fuera de la célula.

Una clasificación que se le da a los puentes disulfuros está ligada a qué tan difícil es su hidrólisis, siendo los de Clase I aquellos que se rompen fácilmente en condiciones reductoras débiles y de Clase II los que requieren una condición reductora fuerte para hacerlo. Usando esta clasificación se ha identificado que el enlace entre las CYS524 pertenece a la Clase I, al igual que el de otras cisteínas en la subunidad α y el enlace entre subunidades α y β a la Clase II^{7,35}.

Glicanos asociados

Los glicanos son moléculas que van a estar conformadas por varias cadenas de sacáridos o carbohidratos por lo que son considerados como polisacáridos. Estos estarán unidos entre sí por enlaces covalentes conocidos como enlaces glicosídicos. Pueden presentarse unidos a otras moléculas que no pertenecen a la categoría de carbohidratos, como las proteínas, en cuyo caso dicha molécula es considerada como un aglicano y el conjunto de glicano + proteína es llamado glicoproteína. En las plantas y algunos microorganismos se observan enlazados covalentemente a moléculas orgánicas de pequeño

tamaño donde se les conoce como glicósidos. No obstante, también pueden encontrarse de forma libre³⁶. Es importante hacer notar que glicanos y glucanos no se refieren a lo mismo, ya que el primer término hace referencia a la unión de un mismo tipo o diferentes tipos de sacáridos, mientras que los glucanos son definidos como la unión exclusiva de solo unidades de glucosa³⁷. De esta forma, se puede decir que todos los glucanos pertenecen a los glicanos pero no todos los glicanos son específicamente glucanos.

Entonces, ¿cómo afectan los glicanos a la estructura de una proteína? Pues primero se debe entender que esta última suele pasar por procesos post-transduccionales, es decir, modificaciones que se le hacen a la estructura ya formada para que sea funcionalmente activa (madura) en el lugar de la célula en el que se deba encontrar. Una de estas modificaciones es la glicosilación, donde se va añadir de forma covalente glicósidos por el enlace al átomo de hidrógeno de los residuos asparagina (N-glicosilación) o por enlace al oxígeno de los residuos de serina y treonina (O-glicosilación) del polipéptido mediante el uso de enzimas. Dado que el glicósido añadido puede llegar a tener una estructura muy compleja y grande por las diferentes ramificaciones de azúcares, este puede influir en la forma en la que la proteína se dobla sobre sí misma (plegamiento) ayudando a obtener la estructura tridimensional nativa^{38,39}. Si bien en algunos casos puede no afectar a la estructura 3D directamente, la glicosilación también cumple un rol muy importante en los procesos de modulación y reconocimiento, formando parte de la protección contra proteasas, almacenamiento, degradación, señalización intracelular, entre otros procesos⁴⁰.

Específicamente en el Receptor de Insulina, los N-glicanos han demostrado participar en la correcta formación y plegamiento de la proteína en su estado dimérico. También, contribuyen a que se puedan dividir las subunidades de cada monómero, y que una vez completadas las modificaciones post-traduccionales, la proteína pueda ser llevada a la membrana celular. Asimismo, parece ser que así se observen mutaciones en los sitios de glicosilación y mientras exista al menos un sitio glicosilado en cada uno de los dominios L1, CR y L2, el receptor puede tener un procesamiento normal y garantizar la unión al ligando⁸.

La posición de los glicanos respecto a una estructura suele determinarse por cristalografía de rayos X, sin embargo, por su naturaleza flexible, este tipo de análisis no siempre indica la presencia de dichas moléculas y por ello tienen que predecirse con programas computacionales³⁸. Hasta el momento se ha podido identificar que el IR tiene un ectodominio muy glicosilado con aproximadamente 19 sitios por cada monómero

para que se realice N-glicosilaciones y 17 de estos se han confirmado que están unidos a glicanos. También, solo se han observado glicosilaciones del tipo N en la subunidad α y tanto N- como O-glicosilaciones en la subunidad restante, de estas últimas se observaron al menos 6 en el IDb del ectodominio. Se ha descrito que los glicósidos asociados a las N-glicosilaciones son del tipo $\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$ y a las O-glicosilaciones de tipo GalNAc.Gal.NeuAc o GalNAc.Gal . Asimismo, se menciona que las glicosilaciones en ASN397 y ASN418 son realmente importantes y necesarias si se quiere contar con un receptor correctamente sintetizado y procesado^{8,23,41}. Finalmente, también se ha indicado que los sitios de glicosilación desde el dominio L1 hasta L2 se encuentran del lado externo a diferencia de los sitios en los dominios FnIII donde se encuentran en la cara interior²⁴.

1.2. Bioinformática y Dinámica Molecular

En palabras sencillas, la bioinformática resulta ser la aplicación de una computadora al estudio de las Ciencias de la Vida, junto con las demás áreas que siempre van de la mano con la informática, como las matemáticas y ciencias de la información. Sin embargo, este campo es tan amplio que no se le puede asociar un objetivo general determinado y básicamente este va a depender de lo que se trabaje. Algunos de los objetivos más comunes involucran la representación de sistemas experimentales mediante simulación o dinámica molecular, análisis y almacenamiento de información biológica en bases de datos, predicción de estructuras proteicas, entre otros. No obstante, una de las áreas con mayor estudio es la industria farmacéutica porque mediante técnicas computacionales es posible identificar las moléculas, proteínas o sitios dentro de una secuencia genómica que deben ser el objetivo de un tratamiento determinado. De forma que se pueden desarrollar fármacos específicos^{42,43}.

En cuanto a las simulaciones de Dinámica Molecular, estas tienen como objetivo predecir el lugar en el que se debe encontrar una partícula o átomo al haber transcurrido un tiempo determinado. Para ello, va a ser necesario el uso de campos de fuerza que contienen descripciones de las interacciones entre los átomos de un sistema, indicando datos como tamaños de enlace, ángulos, descripción de diedros, factores de corrección entre otros valores⁴⁴.

Los datos contenidos en el campo de fuerza provienen en su mayoría de cálculos mecánico cuánticos realizados con diferentes moléculas e incluso de algunos procedimientos

experimentales. Esta información va a permitir que se pueda calcular las fuerzas que actúan sobre cada átomo como consecuencia de su interacción con los átomos que lo rodean, para lo cual utiliza modelos físicos como la Ley de Newton del movimiento. De forma que, tras realizar los cálculos correspondientes, los átomos se desplacen hacia una nueva posición donde se volverá a realizar el cálculo, tanto de fuerzas como de velocidad, y estos se moverán nuevamente buscando la posición más adecuada dentro del sistema según el tipo de análisis que se haya especificado inicialmente. La información del movimiento de los átomos es almacenada constantemente en un archivo de trayectorias, el cual puede incorporarse a un programa de visualización junto con la estructura inicial y observarse en una película el movimiento que ha tenido está a lo largo del tiempo especificado⁴⁴.

Entonces, ¿cuál es la importancia de las simulaciones de Dinámica Molecular? Pues podría decirse que es la información adicional y necesaria que le brinda a los trabajos experimentales. Veámoslo así, muchas veces se busca entender la función de una determinada molécula, digamos una proteína, en algún sistema, como el cuerpo humano y para ello se realizan determinaciones experimentales pero tras varios ensayos sigue siendo difícil determinar o describir la función de esta. Esto se debe a que la función suele estar muy relacionada con la dinámica conformacional de la proteína a través de varios mecanismos, algunos de ellos muy complejos, por lo que se requiere conocer tanto su estructura tridimensional como la dinámica de sus átomos a través del tiempo y dentro de un ambiente controlado⁴⁵, ya que componentes ajenos ella también pueden afectar su comportamiento.

El primer requisito se puede obtener fácilmente en el laboratorio mediante técnicas de cristalografía con Rayos X, Resonancia Magnética Nuclear (NMR), Microscopía Electrónica (cryo-EM) y Transferencia de Energía de Resonancia de Förster (FRET)⁴⁴, con las cuales no siempre se obtienen estructuras completas y detalladas, pero siguen representando la mejor forma de determinación de estructuras tridimensionales. En cuanto al segundo requisito, este es más difícil de obtener en el laboratorio, ya que observar el movimiento, y por lo tanto, determinar la función de ciertas macromoléculas, es un proceso que aún está en estudio, siendo la espectroscopía de Resonancia Magnética Nuclear una de las más trabajadas⁴⁶. Esta técnica ha permitido la observación de la dinámica de algunos complejos con resolución atómica⁴⁷. No obstante, uno de los problemas asociados a esta técnica es la sensibilidad de detección que para ser mejorada requiere de modificaciones al procedimiento convencional⁴⁸, lo que se puede traducir en mayores gastos, más pruebas y por supuesto, más tiempos de ensayo. Esto

hace que sea necesario buscar otras alternativas menos complejas para la evaluación de la dinámica de una proteína y es ahí donde la bioinformática, en especial la dinámica molecular, se hace tan necesaria.

Las ventajas que ofrece la Dinámica Molecular son abundantes, dentro de ellas están^{44,45,49}:

- Simular sistemas experimentales, ya sea al vacío o dentro de un solvente declarado de forma implícita o explícita, con condiciones controladas como temperatura, pH, presión, entre otros.
- Estudiar y entender las propiedades estructurales, motoras y fisicoquímicas de una molécula o biomolécula dentro de un sistema.
- Permitir que una proteína alcance su estado nativo mediante la pérdida de la energía potencial e incentivo del plegamiento.
- Construir y evaluar sistemas en presencia de una membrana lipídica.
- Examinar la dinámica de una estructura al estar solvatada, incrustada en una membrana o en presencia de un ligando.
- Evaluar el proceso de transición entre dos estados conocidos.
- Predecir la reacción de biomoléculas frente a cambios o perturbaciones del sistema como mutaciones, fosforilación, adición de hidrógenos o cuando se añade o retira un ligando cercano.
- Disminuir el número de ensayos experimentales al descartar con dinámica molecular las condiciones que definitivamente no serían favorables para lograr un objetivo en un ensayo práctico.

A pesar de las considerables ventajas que ofrece esta técnica, aun existen limitaciones en su aplicación, las cuales están ligadas principalmente a la falta de computadoras con la capacidad para soportar y acelerar los cálculos requeridos en grandes sistemas. Los avances tecnológicos que continúen surgiendo en años venideros definitivamente contribuirán al desarrollo de esta área en particular, viéndose ya una mejora desde que es posible adicionar a computadoras de uso diario Unidades de Procesamiento Gráfico (GPUs) que aceleran considerablemente los cálculos y reducen el tiempo de

evaluación. No obstante, siempre será más recomendable contar con súpercomputadoras que permitan el cálculo de varios sistemas de forma paralela y en tiempos relativamente cortos⁴⁴.

1.2.1. Selección del campo de fuerza

La dinámica molecular es muy valiosa por los beneficios que se han descrito anteriormente, no obstante, la capacidad de obtenerlos va a estar directamente relacionada con la selección del mejor campo de fuerza para el objetivo que se quiere conseguir.

Como ya se ha indicado anteriormente, los campos de fuerza contienen información detallada de sistemas que ya han sido estudiados a profundidad por diversos investigadores y han encontrado que ciertos parámetros se mantienen constantes. La mayoría de estos siguen formas parecidas de trabajo pero la diferencia se presentará por la función de energía potencial utilizada y los parámetros que considera dentro de dicha función⁵⁰. Entonces, ¿cuál se debe escoger?. Esta pregunta no tiene una única respuesta ya que cada campo de fuerza contará con una serie de fortalezas y debilidades según las moléculas a las que ha dirigido su estudio. En otras palabras, algunos campos de fuerza serán más adecuados para proteínas, otros para líquidos, fármacos, lípidos y demás. Así, para la molécula que uno desee estudiar existirá un campo de fuerza mejor desarrollado, con resultados bastante cercanos a los que se podría obtener de forma experimental. A veces, se puede observar que hay más de un campo de fuerza adecuado para el objetivo que se está persiguiendo, en ese caso la elección podría centrarse en aquel que es más sencillo de manipular o entender, por si dada la situación, se necesita realizar algún cambio en él.

Algunos de los campos de fuerza más utilizados son AMBER, CHARMM y OPLS⁴⁴. El primero ha sido mejorado para trabajar con ácidos nucleicos⁵¹ y proteínas⁵². Pero también se puede utilizar en el cálculo de carbohidratos, lípidos, moléculas pequeñas y agua⁵³. CHARMM es probablemente el campo de fuerza más conocido, su fortaleza reside en el amplio número de biomoléculas con las que puede trabajar, especialmente proteínas. No obstante, aun presenta limitaciones frente a la presencia de ligandos de gran tamaño o complejidad⁵⁴. Por último, OPLS en su versión *all atom*, está optimizado para el estudio de líquidos orgánicos e interacciones en fase gas⁵⁵. Sin embargo, se ha reparametrizado para trabajar con biomoléculas como nucleótidos⁵⁶ y es recomendable su uso en sistemas que busquen la descripción de moléculas orgánicas o biomoléculas en agua⁵⁷.

1.2.2. Diseño del sistema de trabajo

Cuando ya se ha escogido el campo de fuerza con el que se trabajará, es necesario determinar las propiedades que se evaluarán. Esto debido a que se debe incluir en el sistema todas las variables que puedan contribuir a la obtención de la propiedad que se quiere analizar o en todo caso, aquellas que influyan directamente con la estructura trabajada. Por ejemplo, si yo quiero analizar la velocidad de difusión de una molécula en un líquido, requiero que mi sistema tenga un solvente como requisito primordial. Tanto este como mi molécula, y por lo tanto, el valor de difusión, se verán afectados por un cambio de temperatura del sistema. Esto hace necesario realizar la evaluación a una temperatura conocida, variable que también debo describir en mi sistema. De igual forma, cuando se trabaja con estructuras proteicas, el comportamiento de una proteína va a ser diferente en medio acuoso que en vacío, puede desnaturalizarse a altas temperaturas o incluso perder funcionalidad con un pH diferente al óptimo. Razón por la cual todas las variables que influyen en un ensayo experimental o en estado fisiológico tengan que ser descritas obligadamente en el sistema trabajado mediante dinámica molecular.

Ensamblajes

Con la finalidad de poder incluir lo ya mencionado anteriormente y así aproximar los resultados obtenidos con dinámica molecular a los valores experimentales, además de poder calcular funciones termodinámicas, se han desarrollado ensamblajes estructurales^{58,59}. Dentro de los ensamblajes más utilizados para sistemas cerrados se encuentran:

- **Ensamble Microcanónico:**

Este ensamble es considerado como la base para los demás que serán descritos a continuación, e incluso, para el Ensamble Gran Canónico desarrollado para sistemas abiertos⁶⁰. La razón por la que es clasificado más como un tipo de análisis básico o predecesor se debe a que solamente considera en el sistema un número de partículas (N), volumen (V) y energía (E) constante, siendo definido también como ensamble NVE. Esto hace que no se puedan ingresar valores para la temperatura o presión. Por lo que la interpretación de estos valores se realiza una vez terminada la simulación y a través de los valores energéticos obtenidos, pudiendo considerarse el análisis como ineficiente si es que ya se contaba con las condiciones de laboratorio en las que el sistema debió ser trabajado y al final, la

dinámica no se realizó con estos. Por otro lado, su alta relación con la energía del sistema permite asociarlo con el cálculo del potencial termodinámico de entropía, y su simplicidad lo ha clasificado como uno de los esquemas más sencillos para realizar simulaciones de dinámica molecular⁶¹.

- **Ensamble Canónico:**

Este ensamble permite describir en el sistema un número constante de partículas (N), volumen (V) y temperatura (T), razón por la que también se conoce como Ensamble NVT. Está relacionado con el potencial termodinámico de la Energía libre de Helmholtz⁶¹ y a diferencia del ensamble anterior, este reconoce que la molécula en estudio se encuentra en un sistema que tiene una fuente de calor constante, llamado comúnmente baño térmico⁶². Esto se consigue adicionando un termostato, donde el más común es el de Nosé-Hoover o en caso contrario, usando la dinámica de Langevin que es capaz de describir el movimiento Browniano⁶³.

Así, es posible considerar que las moléculas del sistema presentan un movimiento constante pudiendo difundirse en el medio en que se encuentran (movimiento Browniano)⁶⁴, y además, dado que los alrededores ocupan la mayor parte del sistema, pueden intercambiar con la molécula de interés cantidades insignificantes de energía. Si analizamos esta última descripción, es fácil darse cuenta por qué el Ensamble Microcanónico es considerado como la base de este. Analizando el sistema en equilibrio como un todo, es decir, molécula más alrededores sin incluir el termostato, podemos observar que la energía del sistema se va a mantener relativamente constante y por lo tanto, el sistema estará siendo tratado como un Ensamble Microcanónico⁶². Por último, si bien este ensamble es reconocido como importante para alcanzar el equilibrio de un sistema cuando se realiza mecánica estadística⁶⁵, hubo discrepancias sobre si este era adecuado para trabajar cuando el sistema contenía interacciones de largo alcance. No obstante el estudio realizado por Baldovin⁶⁶ demostró que también es adecuado el uso del ensamble NVT en este tipo de sistemas, desde el punto de vista de las ciencias físicas.

- **Ensamble Isotérmico-Isobárico:**

Como lo indica su nombre, este ensamble incluye en su descripción parámetros fijos de temperatura (T) y presión (P), además del número de partículas (N), siendo llamado en muchos artículos como ensamble NPT. Este está relacionado con el potencial termodinámico Energía libre de Gibbs⁶¹ y surgió al tomarse en cuenta que una molécula se encuentra mayormente bajo una presión y temperatura constante ya sea en estado natural o en un ensayo de laboratorio, más que

contenida en un volumen fijo. Por ello, el uso de un ensamble que incluya ambos parámetros en una dinámica sería el más adecuado para comparar resultados teóricos y prácticos.

Entonces, en un ensamble NPT, además de incluirse un termostato se añade un barostato, y para este último se recomienda usar los barostatos de Andersen Parrinello-Rahman y Martyna-Tuckerman-Klein. A diferencia de los anteriores ensambles, trabajar con este implica que el sistema pueda verse afectado por el principio de *Le Chatelier*^{63,67}, donde se describe que cuando se ejerce presión sobre un sistema, este reducirá su volumen debido a que las moléculas dispersas por el estado en que se encuentran comenzarán a juntarse. Esto aplicado a un sistema construido computacionalmente podría resultar en una ligera compresión, disminuyendo el tamaño de la caja del sistema y permitiendo una mayor interacción entre moléculas. Incluso, en casos donde la presión es muy alta, las moléculas de solvente ingresan dentro de la estructura de la proteína contribuyendo a su desnaturalización⁶⁸.

La elección del mejor ensamble va a depender, como ya se indicó, de las propiedades que se quieran analizar o el problema que se busca tratar, así como el tiempo dado para la dinámica. Cuando se realizan tiempos muy cortos de simulación las variables de presión, volumen y temperatura no presentan un efecto muy significativo sobre las estructuras evaluadas. Sin embargo, con tiempos mayores tendrá que evaluarse el parámetro más importante a tomar en cuenta y con él escoger el mejor ensamble para la simulación. Usualmente el ensamble NPT es el que da resultados más semejantes a los obtenidos en un laboratorio, ya que en estos se suele controlar la presión. Asimismo, las biomoléculas pueden verse más afectadas por la presión del sistema, esto asociado principalmente a un ligero cambio en las distancias interatómicas⁶⁸, que por el volumen de solvente que las rodea.

Potencial de Interacción

La dinámica molecular requiere de la resolución de la ecuación de movimiento para poder determinar las nuevas posiciones de los átomos involucrados en la dinámica. Para que esta pueda ser resuelta y las fuerzas puedan ser determinadas se requiere conocer el potencial de interacción gobernante, que se relaciona con la energía potencial de las partículas del sistema⁶⁹. Los potenciales de interacción no son más que expresiones matemáticas que tratan de describir el potencial que se forma naturalmente en una

interacción entre moléculas. Su uso dentro de un cálculo de Dinámica Molecular es esencial para obtener valores de propiedades cercanos a los experimentales, por lo que muchos de ellos se han desarrollado mediante ajustes de los parámetros dentro de la expresión. Para obtener mejores resultados, estos pequeños cambios se realizan hasta que los valores obtenidos mediante el cálculo usando ese potencial sean significativamente cercanos a los que se obtienen en un laboratorio⁷⁰.

Los potenciales de interacción se pueden clasificar en dos tipos:

- Corto alcance o de repulsión que buscan representar las fuerzas dipolo-dipolo y requieren que los átomos estén muy cerca⁷¹. En este tipo de interacción las partículas no presentan carga eléctrica⁷².
- Largo alcance o de atracción que describen tres tipos diferentes de interacción. La primera es por energía electrostática y como su nombre lo indica, está presente cuando dos átomos cargados interactúan o cuando existe un momento dipolar permanente. Luego, está la energía de inducción donde la carga de un sistema cambiará por fenómenos de polarización asociados a moléculas vecinas. Y por último, la energía de dispersión⁷⁰ que aparece cuando hay nubes electrónicas polarizables en el sistema⁷³.

La elección del potencial de interacción a utilizarse no debe realizarse únicamente por el alcance que este es capaz de analizar. Es muy importante tomar en cuenta el tipo de resultados que se quieran obtener pues la selección de un potencial muy complejo involucra un costo computacional elevado, incrementando notoriamente los tiempos de simulación. Por ello, para la optimización de sistemas sencillos o análisis superficiales no se requiere que el potencial describa una gran cantidad de detalles, mientras que para la obtención de propiedades cuantitativas reales va a ser necesario contar con un potencial que incluya en su análisis una amplia gama de parámetros.

Uno de los potenciales más usados es el Potencial de Lennard Jones, el cuál también será utilizado en la presente tesis. Este se destaca porque en su ecuación toma en consideración a las fuerzas atractivas y repulsivas⁷⁴ que actúan para modificar la distancia entre los centros de dos átomos no enlazados y no iónicos. En otras palabras, permite evaluar las interacciones de largo y corto alcance. La explicación para expresar de esta forma el potencial de interacción es muy fácil de entender. Cuando dos átomos están muy lejos las probabilidades de que lleguen a interactuar son prácticamente

nulas, por lo que se considera que la energía potencial de unión es cero. Sin embargo, las mismas condiciones del sistema pueden llevar a que estos dos átomos comiencen a acercarse, lo que eventualmente llevaría a que las posibilidades de interacción aumenten y con ello la energía potencial de unión disminuya hasta alcanzar un valor negativo. Esto sucede principalmente, cuando los átomos están tan cerca que son capaces de formar un enlace entre ellos. No obstante, el movimiento de los átomos no se detiene ahí, sino que estos pueden seguir acercándose buscando una distancia con la que logren alcanzar el mínimo de energía potencial conocido como estado de equilibrio.

En el caso de que estos átomos se acerquen mucho más de la distancia con la que se logra el equilibrio, el valor de la energía potencial incrementará dado que se produce un efecto repulsivo entre los átomos haciendo que se vuelvan a separar para llegar al equilibrio nuevamente. Esto sucede como consecuencia de que al estar muy juntos los átomos, su nube electrónica se ve afectada por la del otro, viéndose forzados a evitar la necesidad de ocupar los orbitales del contrario mediante la repulsión. Si bien el modelo es muy utilizado por ser bastante asertivo en muchos casos, también se ha demostrado que la parte de la ecuación encargada de describir la repulsión entre átomos requiere de un mayor refinamiento para obtener mejores resultados⁷⁵.

Algoritmos de Integración

Como ya se ha mencionado anteriormente la mecánica molecular busca resolver las ecuaciones de Newton para poder determinar la nueva posición de un átomo tras haberse movilizado en el sistema bajo un campo de fuerza y condiciones determinadas. Esta trayectoria del átomo puede ser determinada mediante diversos tipos de integración. Para seleccionar el modelo más adecuado se debe considerar uno que sea capaz de aprovechar la velocidad de la computadora y brinde un resultado con buena precisión para así poder determinar el movimiento de cientos o incluso miles de moléculas que conformen el sistema, tratando de que se asemeje en lo más posible a la realidad. Existen diversos métodos de integración, de los cuales los más usados son los de diferencia finita que permiten integrar ecuaciones diferenciales de segundo orden. Dentro de este tipo de integración se encuentra el modelo de Verlet, método Leap/frog, método de velocidad de Verlet, método Beeman, método Gear y muchos más⁷⁶.

Parámetros de los archivos

Lista de vecinos o lista de Verlet: La lista de vecinos de Verlet es un algoritmo que permite generar un nuevo radio de análisis, el cual es mayor al radio de corte, de forma que todos los átomos que se encuentren dentro de este nuevo radio son llamados vecinos y así el sistema calcula cada cierto tiempo una nueva lista de vecinos para los pares de partículas que se encuentran dentro del radio de corte pero cerca del borde de este.⁷⁷ Así no es necesario identificar y realizar los cálculos de interacción de todas las partículas dentro del radio de corte en cada paso.⁷⁸ Esta información que se actualiza cada cierto tiempo, este último puede ser indicado por el usuario, puede ser almacenada y reutilizada por muchos pasos hasta que alguna partícula sobrepase más de la mitad del radio de la lista de vecinos. Si bien, esta técnica acelera la velocidad del cálculo de interacciones no enlazantes en sistemas con pocos átomos o de desplazamiento lento, en sistemas muy grandes presenta la desventaja de requerir mucha memoria y dificultar la implementación en paralelo, lo cual reduce eficiencia computacional.

1.2.3. Predicción de estructuras de proteínas

En la actualidad, se conoce una gran variedad de secuencias proteicas gracias al desarrollo de la secuenciación del ADN genómico. Sin embargo, solo se ha podido obtener la estructura tridimensional de 57430 proteínas de secuencias humanas hasta el 11 de Junio del 2022, según el portal del *Protein Data Bank*^{a79}. Para la mayoría de casos, estas estructuras son descritas utilizando técnicas de cristalografía por difracción de rayos X, Resonancia Magnética Nuclear o Microscopía Electrónica, las cuales no siempre son suficientes para determinar la estructura completa de una proteína. Por esta razón existe una escasez de estructuras tridimensionales lo que hace necesario desarrollar y evaluar nuevas técnicas que permitan la predicción de las mismas⁸⁰.

En el mismo servidor del PDB se puede observar que en cuanto a proteínas humanas solo existe el registro de un modelo clasificado como teórico, obtenido por la combinación de NMR con metodologías de interacción para obtener la estructura de un complejo proteico⁸¹. Sin embargo, es posible apreciar que técnicas teóricas similares también se han usado para predecir otras 6 estructuras más, asociadas a microorganismos. Esto demuestra que las metodologías computacionales están empezando a ser tomadas en consideración para la obtención de estructuras tridimensionales cuando las técnicas experimentales no son suficientes para lograrlo.

^aDisponible en: <https://www.rcsb.org/>

Para entender como funcionan estas metodologías computacionales para la predicción de estructuras, primero es importante conocer en qué se fundamentan. Es aquí donde se hace necesario conocer cómo están compuestas las proteínas y cómo contribuye su composición a la estructura que optará finalmente. Para empezar, las proteínas se forman gracias a largas cadenas de aminoácidos unidos entre sí por enlaces peptídicos. Estos se caracterizan por presentar en un extremo un grupo amino y en el otro un grupo carbonilo, permitiendo así la formación del enlace peptídico entre aminoácidos. Sin embargo, la característica que diferencia a cada uno de ellos e influye notoriamente en la estructura de la proteína es la cadena secundaria que presentan, única en cada uno de los 20 aminoácidos. La presencia de dicha cadena le otorga propiedades al aminoácido, permitiendo su clasificación en no polares o polares, con carga negativa o positiva, ácidos, aromáticos, entre otras diferentes clasificaciones. La propiedad otorgada es importante porque le permitirá formar enlaces, como puentes de hidrógeno, enlaces iónicos, puentes disulfuro, además de otros, con los aminoácidos a los que no se encuentra unido por un enlace peptídico. Asimismo, dependiendo del lugar que ocupe el aminoácido dentro de la secuencia, contribuirá al plegamiento de la macromolécula y con ello a la formación de estructuras secundarias⁸².

Ya que la estructura primaria de una proteína, es decir, su secuencia de aminoácidos, es una de las condiciones que permite definir su estructura tridimensional, debería ser factible el poder modelar dicha proteína mediante el conocimiento de su secuencia. Es por esta razón que se han desarrollado diferentes métodos teóricos que se basan en principios fisicoquímicos o en métodos empíricos con la finalidad de predecir la estructura a partir de moldes ya conocidos. No obstante, uno de los mayores inconvenientes es la complejidad del área, ya que en muchos de los casos requiere de la resolución de cálculos grandes. Esto hace que exista la necesidad de contar con recursos computacionales elevados si es que se desea obtener resultados con mayor calidad y exactitud, razón por la cual su desarrollo estuvo limitado hasta la actualidad y solo se utilizaba como última instancia cuando la determinación de la estructura era muy difícil de realizar por otros métodos debido a su forma y plegamiento.⁸³

1.2.4. Predicción por comparación

Esta técnica provee modelos de proteínas confiables y precisos ya que se basa en la observación general de que las secuencias relacionadas evolutivamente tienen estructuras tridimensionales similares. Por lo tanto, un modelo de una proteína de interés, conocido como *target* u objetivo, puede ser construido desde proteínas relacionadas de

estructura conocida, a estas se les conoce como *template* o molde, que comparte una secuencia similar estadísticamente significativa. Este tipo de modelamiento involucra numerosos pasos consecutivos que usualmente se repiten una y otra vez hasta lograr un modelo satisfactorio. Para que esto pueda ser alcanzado, se necesita alinear la secuencia del target al template, identificar regiones estructuralmente conservadas, modelar las cadenas laterales, refinar y evaluar el modelo resultante. A pesar de que cada paso puede introducir errores que afecten el modelado de la estructura, el uso adecuado de la información estructural de los moldes disponibles y las correcciones realizadas por alineamiento son los mayores determinantes para obtener un modelo final de calidad.⁸⁴

1.2.5. Predicción *de novo*

Esta técnica también es conocida como modelamiento *ab initio*, modelamiento basado en física o modelamiento libre. Se basa en la realización de una búsqueda guiada por la función de energía diseñada, lo cual va a generar un número de conformaciones posibles dentro de las cuáles se encontrará la estructura final. La calidad de dichos modelos va a depender principalmente de 3 factores según *Lee et. al*⁸⁵:

- Contar con una función de energía precisa donde la estructura de la proteína nativa va a corresponder al estado termodinámicamente más estable respecto a las demás estructuras predichas.
- Utilizar un método eficiente de búsqueda que pueda identificar rápidamente los estados de baja energía a través de una búsqueda conformacional.
- Identificar una estrategia que pueda seleccionar modelos cercanos a los nativos de una gran variedad de estructuras posibles.

Análisis de estructuras proteicas

Para garantizar que la conformación estructural obtenida es adecuada y lo bastante próxima a una estructura nativa, se requiere la realización de diferentes análisis computacionales que permitan descartar a las estructuras no posibles, así como mejorar aquellas que son plausibles.

Algunas de estas técnicas son:

- Gráfico de Ramachandran: El estudio de la estructura secundaria de una proteína permite determinar si una estructura terciaria puede ocurrir naturalmente. Por lo tanto, se ha reconocido que la conformación del esqueleto de una proteína cerca a un residuo aminoacídico dado puede ser especificado en gran parte por dos ángulos dihedros conocidos como ψ y ϕ . Es así que Ramachandran y sus colaboradores⁸⁶ dedujeron que las cadenas principales de péptidos se encuentran en solo algunas regiones del espacio de configuración del ángulo diédrico (regiones permitidas o favorecidas). Esto debido a los choques estéricos que se producen entre los átomos de los residuos, principalmente por las cadenas laterales. De forma que aquellos ángulos que provoquen choques estéricos en la estructura harán que el residuo al que pertenecen aparezca en una región no permitida del gráfico de Ramachandran.⁸⁷ Esto conlleva a que este gráfico sea considerado como un muy buen indicativo de que la estructura aun no se encuentra correctamente optimizada cuando se observan muchos residuos de la proteína en las zonas no permitidas.
- Desviación cuadrática media (RMSD): Se emplea comúnmente para evaluar la estabilidad de los cambios conformacionales a través del tiempo, siendo efectivo en el análisis de las trayectorias de dinámicas moleculares. Es un buen indicador para determinar cuándo la estructura ya está en el equilibrio, comprobar si es que la minimización o dinámica se ha realizado eficientemente e incluso nos facilita el poder agrupar conformaciones similares. Por otro lado, esta gráfica es usada en las predicciones de estructuras proteicas para evaluar la coincidencia entre la estructura modelada y la obtenida por técnicas de rayos X o NMR, así como en los cálculos de acoplamiento donde se puede escoger las conformaciones más parecidas a las nativas partiendo de muchas conformaciones acopladas. Asimismo, el RMSD permite la comparación entre la estructura de la proteína plegada nativa y su estructura total o parcialmente desplegada, permitiendo identificar la existencia de un cambio conformacional. Por último, ayuda a identificar los cambios estructurales producidos después de una o varias modificaciones a una proteína.⁸⁸ En resumen, según *Carugo y Pongor*⁸⁹ el cálculo de RMSD es la desviación cuadrática media de la distancia calculada entre átomos equivalentes de dos estructuras y está definido como:

$$RMSD = \sqrt{\frac{\sum_i d_f^2}{n}} \quad (1.1)$$

donde d es la distancia entre cada uno de los n pares de átomos en dos estructuras

superpuestas. Cuando el resultado de la fórmula es 0 o cercano a 0 las estructuras son idénticas o muy parecidas entre sí, mientras que un valor muy lejano de dicho número indicaría que hay notorias diferencias en la posición de los átomos de una estructura respecto a la otra.

- **Fluctuación Cuadrática Media (RMSF):** Este análisis permite observar qué regiones, residuos o átomos de la proteína se mueven durante una simulación, permitiendo conocer las zonas que aun no se encuentran estables o incluso evaluar si es que algún ligando o residuo cercano es capaz de afectar la flexibilidad de la proteína en determinada zona⁹⁰. La obtención de esta propiedad da un valor que puede ser comparado al factor-B obtenido mediante rayos X, el cual es indicador de la existencia de un desorden estático y uno causado por las vibraciones que genera la temperatura en un sistema. Se puede calcular fácilmente tras haber realizado una dinámica con un sistema que incluya una proteína. Normalmente, el análisis solo se lleva a cabo midiendo la fluctuación de los carbonos alfa de la proteína^{91,92}, por lo tanto el cálculo de este valor se realiza según la siguiente ecuación:

$$RMSF = \sqrt{\frac{1}{N} \sum_J (x_i(j) - \langle x_i \rangle)^2} \quad (1.2)$$

donde x es la posición del residuo (átomo C-alfa) i en la trayectoria de la dinámica j y $\langle \rangle$ describe el promedio sobre toda la trayectoria⁹³.

- **Radio de Giro (Rg):** Este parámetro está directamente relacionado con el tamaño de la proteína a analizar. De forma que el valor obtenido de la resolución de la ecuación debajo descrita sufre cambios cuando la forma o el tamaño del objeto en estudio se ve alterado^{94,95}.

$$R_g^2 = N^{-1} \sum_{i=1}^N (r_i - r_g)^2 \quad (1.3)$$

donde N es el número de residuos de aminoácidos, mientras que r_i y r_j son todas las posiciones del centro de masa del i -ésimo residuo y de la proteína, respectivamente. Usualmente, un menor valor del radio de giro suele ser indicador de una proteína plegada. A lo que un mayor puede significar que la proteína está perdiendo su conformación o desnaturalizándose⁹⁶.

- **Interacciones de enlace de hidrógeno:** Este análisis permite identificar el número de enlaces de hidrógeno que existen en la proteína así como visualizar cómo este

número varía con el tiempo de la dinámica. Dado que los puentes de hidrógeno contribuyen con la formación de estructuras secundarias en las proteínas, es importante tomar en cuenta la variación del número de estos enlaces durante una simulación, ya que un menor valor podría indicar la pérdida de la conformación estructural de la proteína que se está evaluando.^{97,98}.



Capítulo 2

Metodología

2.1. Lugar de ejecución

Este proyecto de tesis se trabajó en el Centro de Investigación en Ingeniería Molecular (CIIM) ubicado en las instalaciones de la Universidad Católica de Santa María, Arequipa, Perú.

2.2. Detalles Computacionales

2.2.1. Hardware

Se hizo uso de una estación de trabajo (*WorkStation*), con un procesador E7 de 3.1 GHz, con una memoria RAM de 64Gb, 3 discos duros de 1, 4 y 6TB y dos tarjetas aceleradoras de video GeForce GTX 980.

2.2.2. Softwares

Gromacs 2019.2

GROMACS^a es un software de código abierto que permite realizar simulaciones de dinámica molecular y cálculos de minimización de energía desde una terminal. Este

^aDisponible en:<http://gromacs.org>

se usa para trabajar principalmente con proteínas o biomoléculas, aunque también ha sido optimizado para trabajar con sistemas no biológicos. Las principales ventajas que ofrece este programa son la variedad de campos de fuerza que pueden ser utilizados y el alto rendimiento que otorga sus algoritmos de paralelización y la aceleración de CPU-GPU integrados al programa⁹⁹.

UCSF Chimera Alpha Linux

El programa UCSF Chimera^b es un visualizador de estructuras construido en C++ y Python, que además es de acceso libre. Se encuentra dividido en un núcleo y extensiones. El primero se encarga de realizar tareas sencillas, como seleccionar, mover, y juntar modelos, así como permitir diferentes visualizaciones gráficas. Mientras que el segundo permite extender las capacidades que tiene Chimera para otro tipo de necesidades. Su ventaja está en el amplio número de archivos de estructura que puede leer y su interfaz sencilla de manejar¹⁰⁰.

VMD

VMD^c es un programa de código abierto construido en C++ que permite visualizar gráficos moleculares, principalmente biomoléculas, otorgándoles diferentes tipos de representación, para que puedan ser guardados como imágenes de alta calidad. Además, ha sido diseñado especialmente para poder mostrar una animación de las trayectorias generadas por simulaciones de dinámica molecular, facilitando la evaluación de los sistemas de estudio. Este programa también cuenta con la posibilidad de trabajar acoplado a programas de simulación de dinámica molecular y visualizar las estructuras en el proceso¹⁰¹.

MEGA X

El programa libre de Análisis Genético Evolutivo Molecular^d (MEGA, por sus siglas en inglés) presenta herramientas que permiten realizar alineamientos de secuencias, inferir árboles evolutivos, estimar distancias genéticas y de diversidad, entre otras aplicaciones. Sus mayores ventajas se ven evidenciadas en la última versión, X, que ha sido optimizada para mejorar el consumo de memoria RAM y acelerar el rendimiento del programa. MEGA X está diseñado para funcionar en los tres principales sistemas

^bDisponible en: <https://www.cgl.ucsf.edu/chimera/>

^cDisponible en: <https://www.ks.uiuc.edu/Research/vmd/>

^dDisponible en: <https://www.megasoftware.net/>

operativos del mercado, Windows, Linux y Mac, además de contar con una interfaz sencilla y moderna organizando los diferentes análisis en grupos lógicos¹⁰².

PyMOL

PyMOL^e es un visualizador molecular basado en una fundación de código abierto, desarrollado por el grupo de trabajo Schrödinger, LLC. Puede ser instalado en cualquier sistema operativo y cuenta con una interfaz gráfica moderna y sencilla que permite la obtención de imágenes de alta calidad. Además, tiene muchas funciones para mejorar la edición y visualización de estructuras en 3D¹⁰³.

DoGlycans

DoGlycans es un software que permite construir y modificar carbohidratos y sus polímeros, glicoproteínas y glicolípidos. Su ventaja reside en que además de generar la estructura final también brinda la topología de la misma para que esta pueda ser trabajada con Gromacs. Por el momento, el programa solo puede trabajar con las unidades de azúcar que se encuentran definidas en su propio campo de fuerza GLYCAM y para el caso de la glicosilación de proteínas, las topologías que genera solo son compatibles con los campos de fuerza de AMBER y OPLS. En este trabajo de investigación solo se ha utilizado la glicosilación de proteínas pero es importante indicar que la topología generada para glicolípidos solo se puede usar con OPLS y para los polímeros de carbohidratos solo se puede usar con GLYCAM¹⁰⁴.

2.2.3. Servidores Web

NCBI

La página del Centro Nacional para la Información Biotecnológica (NCBI, por sus siglas en inglés) contiene recursos para obtener información y datos genómicos, genéticos y biomédicos, con fuentes que incluyen las bases de datos de GenBank y PubMed¹⁰⁵. Dentro de la información contenida en este portal (ubicado en <https://www.ncbi.nlm.nih.gov/>) se puede encontrar desde secuencias genéticas, dominios, estructuras proteicas y análisis taxonómicos hasta libros, artículos científicos, tutoriales y cursos relacionados a la biología molecular, entre otras áreas. Toda esta información

^eDisponible en: <https://pymol.org/2/>

tiene como propósito la difusión del conocimiento dentro de la comunidad científica.

UniProt

UniProt presenta una amplia colección de secuencias y datos para más de 120 millones de proteínas. Este servidor está construido con información recolectada de investigaciones reconocidas y confiables, ya sea a través de bases de datos biológicas o por extracción de textos. Además, en sus publicaciones más recientes se han incluido nuevas formas de visualización de datos para las diferentes proteínas, dentro de las que se encuentran las opciones de ubicación celular, estructura molecular e interacciones¹⁰⁶. Se puede acceder a este servidor desde: <https://www.uniprot.org/>

Protein Data Bank

El Banco de Datos de Proteínas (PDB, por sus siglas en inglés) contiene una gran cantidad de estructuras proteicas, que siguen incrementándose cada día, obtenidas por métodos de rayos-X, NMR, microscopía crioelectrónica y modelamiento teórico. De esta forma, el objetivo principal de este servidor es facilitar el uso y análisis de estructuras⁷⁹, ofreciendo no solo archivos con definición estructural, sino también secuencias FASTA y artículos relacionados. Es importante mencionar que estas estructuras se encuentran clasificadas por organismo, resolución, fecha de publicación, método de obtención, entre otras características, que facilitan la búsqueda de proteínas de interés. Para visitar dicha base de datos se puede ingresar a: <https://www.rcsb.org/>

SWISS-MODEL

El servidor de SWISS-MODEL, <https://swissmodel.expasy.org/>, tiene como principal función el facilitar la obtención de estructuras tridimensionales de proteínas mediante la modelación por homología a partir de su secuencia aminoacídica. Para ello, consta de diferentes tipos de formas para realizar el modelamiento. Algunas de estas son permitir que el servidor busque las secuencias a usar como molde o subir en formato .pdb una estructura seleccionada previamente para que Swiss-Model la use como tal. Por otro lado, el procedimiento que realiza este servidor hasta obtener la estructura final se basa en seleccionar un molde, realizar el alineamiento eliminando moldes no adecuados, construir el modelo, reconstruir las cadenas secundarias y realizar un minimización de energía con el campo de fuerza GROMOS, para finalmente reportar los modelos diseñados con sus respectivas evaluaciones y calificaciones¹⁰⁷.

PEP-FOLD3

El servidor de PEP-FOLD3 permite la determinación estructural de pequeños péptidos que consten de 5 a 50 aminoácidos mediante la modelación *de novo*. Para la cual primero utiliza un método probabilístico conocido como alfabeto estructural, genera la estructura tridimensional y al final, calcula la mejor descripción conformacional usando algoritmos estándar *Hidden Markov Models*. Este programa solo requiere de la secuencia aminoacídica de la proteína, ya sea libre en solución o interactuando con una proteína, para ofrecer en pocos minutos las estructuras de los mejores modelos ordenados por calificación y con la posibilidad de poder visualizarlos en 3D desde la web¹⁰⁸. El ingreso al servidor se realiza mediante el siguiente link: <https://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3>.

Ab Initio Domain Assembly Server (AIDA)

El servidor de AIDA permite el ensamblaje de proteínas con dominios continuos y discontinuos sin imponer un límite de cantidad, para lo cual solo es necesario ingresar la secuencia fasta y las estructuras a unir. También puede utilizarse para el modelamiento y ensamblaje de proteínas multidominio donde se puede comenzar directamente desde la secuencia fasta. Por otro lado, es posible definir en este servidor de forma opcional las restricciones de distancia que deben haber entre los dominios¹⁰⁹. Este servidor mantiene rígidas las estructuras ingresadas durante la minimización pero las zonas de unión totalmente flexibles, haciendo posible la determinación de la mejor posición de los dominios y la correcta unión entre estos, dando como resultado una sola estructura considerada como la correcta. AIDA se encuentra alojado en: <http://aida.godziklab.org/>.

3DRefine

3DRefine, alojado en <http://sysbio.rnet.missouri.edu/3Drefine/>, es un servidor que permite optimizar las estructuras de una sola cadena. Para ello utiliza un protocolo iterativo que realiza optimizaciones de los enlaces de hidrógeno y una minimización de energía de la estructura ingresada. Este proceso se realiza 5 veces después de haber ingresado la estructura inicial, de forma que se puedan obtener 5 modelos diferentes con sus respectivas calificaciones métricas que incluyen, por defecto, GDT-TS, GDT-HA Y RMSD, más dos opciones que se pueden especificar al momento de ingresar la estructura, las cuales son RWplus y Molprobit¹¹⁰. Todos los modelos finales pueden ser visualizados desde la web y descargados según se desee.

CHARMM-GUI

El servidor de CHARMM-GUI (<http://www.charmm-gui.org/>) permite la generación de archivos que pueden ser utilizados en dinámica molecular usando el campo de fuerza CHARMM. Su ventaja reside en ofrecerle al usuario la capacidad de editar sus estructuras de forma sencilla mediante una interfaz gráfica, disminuyendo así los errores triviales ya que cualquier error en el diseño del sistema podrá ser corregido regresando al paso anterior, y también, permitiendo la obtención de resultados con mejor calidad. Los nuevos sistemas generados se pueden descargar a la máquina del usuario y así comenzar con la dinámica requerida¹¹¹. En la presente investigación se usó específicamente la opción que permite la construcción de membranas a través de 6 pasos, los cuales consisten en ingresar la estructura a trabajar, orientarla, determinar el tamaño del sistema, generar los componentes elegidos por el usuario, para finalmente, ensamblarlos y brindar la estructura lista para ser equilibrada con un script diseñado por el mismo programa¹¹².

Quark Online

Este servidor, al que se puede ingresar mediante el enlace <https://zhanglab.ccmb.med.umich.edu/QUARK/>, funciona de forma similar a PEP-FOLD3 permitiendo la determinación de modelos tridimensionales usando una predicción *ab initio*. No obstante, este permite ingresar secuencias de hasta 200 aa y utiliza un método que se basa en extraer fragmentos pequeños de hasta 20 aa de proteínas no relacionadas y mejorar su selección mediante características importantes como el alineamiento de secuencias y la concordancia entre puentes disulfuro¹¹³. Por otro lado, para poder realizar la predicción es necesario registrarse en el servidor y contar con la secuencia FASTA de la estructura. Cuando los resultados estén listos, las estructuras finales son enviadas al correo registrado.

GlyProt

Este servidor web <http://www.glycosciences.de/glyprot/> permite la adición de N-glicanos a estructuras proteicas en formato PDB. Su ventaja principal es la interfaz amigable del programa que facilita grandemente la tarea de agregar glicanos. El programa también evalúa la factibilidad de que un glicano pueda ser agregado a una secuencia Asn-X-Ser/Thr mediante una evaluación previa de la unión del glicano *Man₃N* a la proteína a trabajar y de esta forma descarta las conformaciones no viables, que son to-

das aquellas que se solapan con la estructura proteica. Una vez el programa determina qué regiones de la proteína pueden contener glicanos, ofrece la opción de adicionarlos usando una estructura típica, una estructura específica que se encuentre en la base de datos del programa o construir una desde cero usando la nomenclatura IUPAC. Al finalizar, el programa brinda una estructura en formato PDB que ya contiene los glicanos añadidos en las regiones indicadas e incluso es posible visualizar algunos parámetros fisicoquímicos de la proteína no glicosilada y la que se ha glicosilado que demuestran los cambios ocurridos por efecto de adicionar los glicanos¹¹⁴.

MolProbity

MolProbity <http://molprobity.biochem.duke.edu/> es un servidor web que permite realizar diferentes análisis de validación a una estructura a partir de sus coordenadas, ya sea subiéndola al servidor o ingresando su código PDB o de la base de datos NDB. Dentro de las funciones que presenta se encuentra la adición de hidrógenos, análisis de contacto de todos los átomos, generación del diagrama de Ramachandran, análisis de rotámeros, análisis de geometría covalente y análisis de ácidos nucleicos, además de muchas otras. Por otro lado, MolProbity brinda un puntaje que provee información sobre la calidad de la proteína y se obtiene al combinar los puntajes de choque, el porcentaje de átomos en zonas no favorecidas de la gráfica de Ramachandran y el porcentaje de rotámeros defectuosos de las cadenas laterales, donde un puntaje menor que el de una resolución cristalográfica de la misma estructura indica un modelo de mejor calidad¹¹⁵.

2.3. Métodos

2.3.1. Obtención de las secuencias de interés

El primer paso para poder diseñar la estructura de una proteína es contar con su secuencia aminoacídica completa. Por ello, se realizó la búsqueda de la secuencia del Receptor de Insulina Humana en las bases de datos del NCBI y Uniprot.

2.3.2. Construcción de segmentos

Se realizó la búsqueda de las estructuras tridimensionales del ectodominio y segmento intracelular más recientes en la base de datos del *Protein Data Bank*. Estas

estructuras se descargaron y abrieron con el visualizador Chimera para eliminar aguas, iones, ligandos y estructuras no deseadas, dejando solo la estructura de un monómero. Luego, mediante el uso de un script^f se obtuvo la secuencia FASTA correspondiente a dichas estructuras.

Las secuencias obtenidas se alinearon con la secuencia del IR en Mega X usando el algoritmo MUSCLE (Align >Edit/Build Alignment >Create a new alignment >Protein) y se compararon para escoger los mejores modelos. Una vez seleccionados, se anotó las secuencias de los segmentos que faltaban ser modelados. Estas se ingresaron a un servidor de predicción *de novo* de estructuras, para cuya elección se tomó en consideración el tamaño de la secuencia a predecir según el siguiente criterio:

- Secuencia menor a 50 aa: PEP-FOLD 3. El servidor brindará una numerosa lista de modelos ordenados por el puntaje que se haya seleccionado al momento de ingresar la secuencia. El más utilizado, y el que se usó en esta investigación, es el sOPEP. Este puntaje se refiere a la energía obtenida con una versión modificada del “*Optimized Potential for Efficient Structure Prediction (OPEP)*”¹¹⁶ que se ha unido a un algoritmo para ayudar a disminuir los problemas estéricos¹¹⁷, por lo que un menor valor será indicativo de un modelo estéricamente mejor desarrollado.
- Secuencia mayor a 50 aa pero menor de 200 aa: Quark Online. Este servidor ofrece 5 modelos, sin embargo, su desventaja reside en que no tiene una calificación para escoger entre los modelos y la elección dependerá en su totalidad del usuario.

Los segmentos obtenidos se unieron a sus respectivos dominios usando dos metodologías según el tamaño de los segmentos a unir y la posibilidad de que se originara un cambio conformacional notable.

Para los segmentos pequeños o con poco riesgo de alteración de las posiciones originales se utilizó el servidor de AIDA que permite el ensamblaje de dos o más estructuras proteicas de forma sencilla. Para ello, en el punto 1 de la página se ingresó la secuencia FASTA de los segmentos a unir. Luego, los archivos .pdb con los modelos correspondientes se comprimieron en un solo archivo .zip y se subieron al servidor en el punto 2. La opción 3 se dejó desmarcada y no se subió ningún archivo al punto 4 ya que no es

^fDesarrollado por Pierre Poulain y obtenido de: <https://cupnet.net/pdb2fasta/>

necesario. Finalmente, se dio click en *Submit* y se obtuvo un único modelo en formato .pdb. La interfaz del servidor descrito se puede observar en la Figura 2.1.



AIDA: Ab Initio Domain Assembly Server Godzik Lab

Please be advised that we have recently moved our server to a new location. We tried our best to fully test all functions on the website, but if something does not work for you please let us know.

Using knowledge-based ab initio potential, AIDA attempts to find the best domain arrangements while maintaining chain connectivity. The structures themselves are kept rigid during the energy minimization simulation, treating only the link regions as completely flexible. The AIDA server is able to work with proteins containing discontinuous domains by fixing the relative positions of the discontinuous domains. The server also supports structure assembly from sequence only. To do this uncovered sequence domains are iteratively split to align with the PDB template found by the FFAS-3D fold recognition program. Furthermore, AIDA also supports restraint-guided simulation producing a final assembly satisfying user-specified inter-domain distance restraints.

- 1). Please paste the full amino-acid sequence in fasta format. (sample input)
 . 10 . 20 . 30 . 40 . 50 . 60 . 70 . 80 . 90 . 100 . 110 . 120 . 130 . 140 .
- 2). Please upload compressed domain models (**pmodel1.pdb, pmodel2.pdb...pmodeln.pdb**) in one file (*tar.bz2, *.tar.gz, *.zip, *.bz2 or *.gz) (optional). (sample input)
Coordinates of few residues in the two ends of each domain structure could be deleted to make the assembly more flexible. **No gap is allowed in the middle of each domain structure.**
Numbering of each residue in domain structures is its position in the full-length sequence. Download the [package](#) and run "renumber" program to handle it.
- 3). check this box only if you upload **three** domain models and **pmodel1.pdb+pmodel3.pdb** constitute a **discontinuous** domain. [explanation](#)
- 4). Please upload Cα distance restraints to assist assembly (optional). (sample input)
Seleccionar archivo No se eligió archivo
- 5). Email: (optional, where result will be sent to)

Figura 2.1: Interfaz de AIDA.
Obtenida de: <http://aida.godziklab.org/>

Posteriormente, esta estructura debe ser refinada para evitar superposiciones o errores estéricos, lo cual se realizó en 3DRefine. En este servidor se ingresó la estructura ensamblada y además se marcó las opciones de realizar análisis con MolProbity y RW-Plus. Una vez procesado el trabajo, el sistema ofrecerá 5 modelos en formato .pdb con sus respectivas calificaciones. La selección del modelo adecuado dependerá enteramente del objetivo que se quiera conseguir, si se desea un modelo físicamente más realista se debe observar los valores de MolProbity donde un valor más pequeño será el mejor. Por otro lado, si se quiere modelos con mejor calidad, valores bajos de 3D^{RefineScore} y RWplus garantizarán esto. Por último, los valores de GDT-TS, GDT-HA y RMSD son indicadores de cuanto difieren las estructuras refinadas respecto a la original, de forma que para las dos primeras clasificaciones valores altos indican un refinamiento conservativo, mientras que valores altos de RMSD son indicadores de un refinamiento agresivo.

En cuanto a las estructuras grandes o con posiciones que deben mantenerse para,

por ejemplo, evitar rotura de enlaces, se utilizó el programa UCSF Chimera ya que este cuenta con la opción de unir modelos de forma manual. Para conseguirlo, las estructuras a unir se abrieron en una misma sesión y se acercó una, seleccionándola primero, hasta la posición de enlace con la otra usando la opción *Tools >Movement >Movement Mouse Mode >Move Selection*. Una vez posicionadas, se seleccionaron los aminoácidos que van a ser unidos para, posteriormente, mostrar sus átomos y ocultar la estructura de cinta en *Actions >Atoms/Bonds >Show* y *Actions >Ribbon >Hide*, respectivamente. A continuación, se seleccionaron los átomos que van a formar parte del enlace amino y se unieron usando la opción de *Join Models >C-N peptide bond* dentro de *Tools >Structure Editing >Build Structure*. En esta opción si los átomos seleccionados pueden ser utilizados para formar un enlace amino el botón *Apply* se encontrará activo y se procederá con la unión de los modelos, como se puede observar en la Figura 2.2. Finalmente, la estructura se guardó en formato .pdb ingresando a *File >Save PDB*. Cabe indicar, que esta misma metodología se utilizó para unir los segmentos que conforman el monómero del IR. Como dato adicional, puede darse el caso que la estructura añadida se encuentra en una posición no adecuada para unir el siguiente segmento o se solapa con otras cadenas, esto se puede corregir seleccionando la parte cuya estructura se quiere mover y usando la herramienta de *Mouse Selection* mencionada anteriormente. Para este tipo de modificación hay que tener cuidado ya que se pueden formar ángulos no viables naturalmente, no obstante, es posible corregir algunos de estos errores durante la posterior optimización.

La adición de segmentos puede causar una numeración incorrecta de los residuos en el archivo .pdb, por ello es importante que antes de realizar la construcción del dímero se haga una renumeración, esto se consigue utilizando el comando de Gromacs que se observa a continuación:

```
gmx editconf -f "nombre_del_monomero".pdb -o "nombre_del_monomero".pdb  
-resnr 1
```

Dicho comando indica que la renumeración debe comenzar desde el número 1, este puede ser cambiado al que se desee según la descripción de la estructura trabajada.

En caso el monómero cuente con más de una subunidad, como en este trabajo, es importante dejarlo establecido en el archivo .pdb, poniéndole una letra diferente a cada una y colocando la palabra TER al final de las cadenas involucradas.

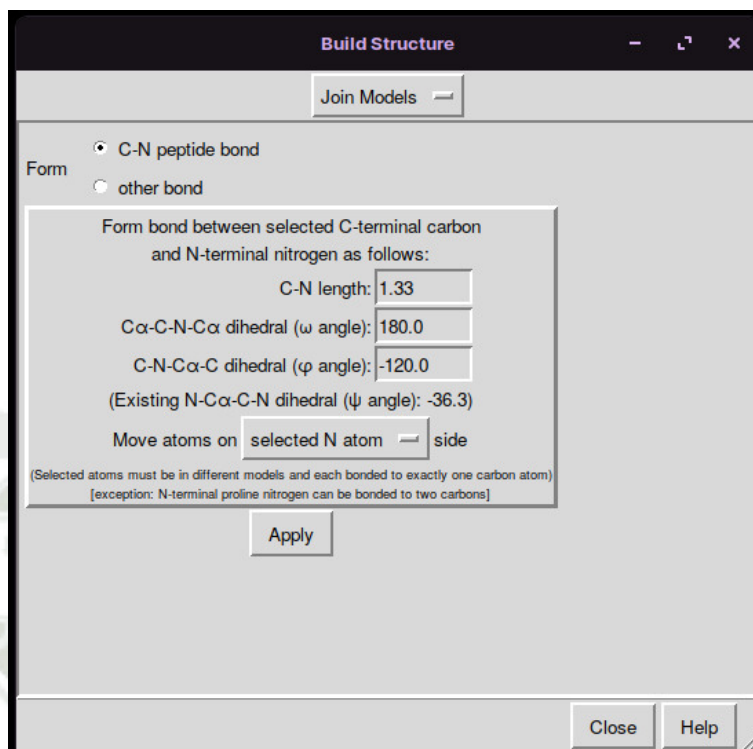


Figura 2.2: Cuadro de diálogo de la opción Build Structure en UCSF Chimera¹⁰⁰.

2.3.3. Optimización con una bicapa lipídica

Uno de los segmentos, el que atraviesa la membrana celular, se optimizó usando una bicapa lipídica. Esta se construyó haciendo uso del servidor CHARMM-GUI en la sección de construcción de membranas[§]. Aquí lo primero a realizar fue subir el archivo .pdb del segmento objetivo ya refinado en *Protein/Membrane System* > *Seleccionar archivo* y marcar la opción de *PDB* en *PDB Format*. Luego, se le dio click a *Next Step: Select Model/Chain*. En el siguiente paso se puede seleccionar los modelos que se desean trabajar, dado que solo se estaba trabajando con un modelo no fue necesario modificar nada en esta sección y se prosiguió con el siguiente paso *Next Step: Manipulate PDB*. En este paso tampoco se marcó ninguna opción pues solo se buscaba optimizar la estructura mas no realizar cambios en ella. Se continuó con el siguiente paso *Next Step: Generate PDB and Orient Molecule*.

En este paso se le permite al usuario escoger entre 4 opciones para alinear la proteína respecto a la membrana como se puede observar en la Figura 2.3. La primera opción utiliza la orientación que tiene la proteína en el archivo subido, en la segunda se va a alinear el eje principal de la estructura a lo largo del eje Z, en la tercera se usa 2 aa como

[§]Dirección para la construcción de membranas: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

inicio y final del segmento que debe ser alineado en Z, por ende, el que estará contenido en la membrana, y la última, usa un servidor asociado para predecir cuál es la mejor orientación de la proteína respecto a la bicapa. Dado que se conocían los residuos que corresponden al segmento transmembrana, se escogió la tercera opción *Align a Vector (Two Atoms) Along Z* para la presente investigación. El resto de las opciones se dejaron como estaban y se continuó con *Next Step: Calculate Cross-Sectional Area*.

En el siguiente paso, lo primero a realizar fue revisar la estructura que se generó, para ello se le dio click a *(view structure* ubicado al lado de *Oriented PDB: step2_orient.pdb*. De esta forma se puede comprobar que la estructura esté ubicada correctamente, además en la parte inferior, debajo de *Calculated Cross Sectional Area*, se podrá observar una gráfica en la que se indica el área que estará dentro de la membrana, así como la extracelular e intracelular. De no haberse obtenido la estructura deseada se puede regresar al paso anterior para realizar las modificaciones correspondientes.

Una vez se aseguró que el segmento deseado se encontraba dentro de la membrana se continuó con las opciones de *System Size Determination Options* en la que se pueden añadir los lípidos que se deseen en la membrana. En este caso se escogieron las siguientes opciones dentro de *Heterogeneous Lipid*:

- *1.Box Type: Rectangular*
- *2.Length of Z based on: Water thickness*
- *3.Length of XY based on: Numbers of lipid components*

Por otro lado, el servidor ofrece una amplia gama de lípidos para construir la membrana y su selección va a depender del usuario, en este caso se decidió trabajar con los dipalmitoilfosfatidilcolina (DPPC) en ambos lados de la bicapa, como se puede observar en la Figura 2.4. El número de lípidos va a variar según el sistema trabajado y podrá ser establecido basándose en trabajos previos. En este trabajo de tesis se comenzó agregando como base 128 moléculas lipídicas y, según la información brindada por *Show the system info* en la parte superior de la tabla, fue necesario agregar lípidos en la zona requerida, ya que se recomienda mantener una área similar en ambos lados de la membrana. Si esta no muestra ninguna advertencia se puede continuar con el siguiente paso *Next Step: Determine the System Size*.

Aquí también fue necesario evaluar la estructura obtenida, para ello se le dio click

Computed Energy:

Please beware of that the computed energy is CHARMM single-point energy and is displayed to make sure all the coordinates are defined.

ENER ENR:	Eval#	ENERgy	Delta-E	GRMS	DIHEdral	IMPRopers
ENER INTERN:	BONds	ANGLes	UREY-b	PMF2D	PRIMO	ASP
ENER CROSS:	CMAPs	PMF1D	HBONds	PRIMO	ASP	USER
ENER EXTERN:	VDWaals	ELEC				
ENER>	0	2419.93770	0.00000	228.99513		
ENER INTERN>		266.84295	219.82544	38.30766	617.26843	0.07437
ENER CROSS>		17.81321	0.00000	0.00000	0.00000	
ENER EXTERN>		1508.88353	-248.27789	0.00000	0.00000	0.00000

Orientation Options: ?

- Use PDB Orientation This option is suggested for an oriented structure from <http://opm.phar.umich.edu>
- Align the First Principal Axis Along Z This option is suggested for small helical bundle or homo-oligomer.
- Align a Vector (Two Atoms) Along Z This option is suggested for an irregular, hetero-oligomer.
- Use PPM Server This option submits an input structure on http://opm.phar.umich.edu/ppm_server
It may take some minutes depending on protein size.

Positioning Options:

- Rotate Molecule respect to the X axis Degree
- Rotate Molecule respect to the Y axis Degree
- Translate Molecule along Z axis Angstrom
- Flip Molecule along the Z axis

Area Calculation Options:

- Generate Pore Water and Measure Pore Size


Next Step: 
Calculate Cross-Sectional Area

Figura 2.3: Configuración de la orientación de la proteína en la membrana.
Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

a *view structure* ubicado al lado de *step3_packing.pdb* en *Packing Simulation*. Con la estructura analizada se continuó en *System Building Options* y se escogió la única opción activa para el sistema que era *Replacement method*. Seguidamente, en *Component Building Options* se marcó ambas opciones de *Include Ions >Add neutralizing ions* e *Ion Placing Method >Distance* y se continuó con *Next Step: Build Components*.

El paso 4 verifica que no haya problemas de interacción entre los lípidos y la membrana evaluando si hay colas de lípidos que traspasan la superficie de la proteína o si pasan a través de los grupos cíclicos. Si este es el caso, se podrá leer la información en *Protein surface penetration* y *Lipid ring penetration*. De no haber problemas se puede continuar con el siguiente paso en *Next Step: Assemble Components*, esto guiará a una sección donde se puede visualizar el sistema en el apartado de *Component PDB*. Si no

Lipid Type	Charge [e]	Tail Info. [sn1/sn2]	Images	# of Lipid on Upperleaflet	# of Lipid on Lowerleaflet	Surface Area
► Sterols						
► PA (phosphatidic acid) Lipids						
▼ PC (phosphatidylcholine) Lipids						
DDPC	0	10:0 / 10:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="64.1"/>
DCPC	0	11:0 / 11:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="64.1"/>
DLPC	0	12:0 / 12:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="64.1"/>
DMPC	0	14:0 / 14:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="64.1"/>
DPPC	0	16:0 / 16:0	[Image]	<input type="text" value="130"/>	<input type="text" value="128"/>	<input type="text" value="63.0"/>
DSPC	0	18:0 / 18:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="65.6"/>
PSPC	0	16:0 / 18:0	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="68.3"/>
POPC	0	16:0 / 18:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="68.3"/>
PLPC	0	16:0 / 18:2	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="65.2"/>
PEPC	0	16:0 / 22:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="66.5"/>
SOPC	0	18:0 / 18:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="66.2"/>
SLPC	0	18:0 / 18:2	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="65.2"/>
DRPC	0	14:1 / 14:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="67.0"/>
DYPC	0	16:1 / 16:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="67.0"/>
YOPC	0	16:1 / 18:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="67.0"/>
DOPC	0	18:1 / 18:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="69.7"/>
OEPC	0	18:1 / 22:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="66.5"/>
DUPC	0	18:2 / 18:2	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="70.0"/>
DGPC	0	20:1 / 20:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="66.7"/>
DEPC	0	22:1 / 22:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="66.5"/>
DNPC	0	24:1 / 24:1	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="62.3"/>
DLiPC	0	18:2 / 18:2	[Image]	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="65.2"/>

Figura 2.4: Configuración de la distribución de lípidos en la bicapa lipídica.
Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

hay cambios a realizar se puede continuar con el siguiente paso, *Next Step: Assemble Components* en el que se van a ensamblar todos los componentes desarrollados hasta el momento y así generar el archivo que se desea para optimizar y equilibrar la estructura.

En el paso 5, se escogieron los parámetros del campo de fuerza con el que se trabajará el equilibrio de la estructura. En *Force Field Options* se seleccionó CHARMM36, *Input Generation Options >GROMACS* y en *Equilibration Options* se marcó las opciones *Generate grid information for PME FFT automatically* y *NPT ensemble*. Finalmente, se especificó la temperatura promedio del cuerpo humano que es 309.65 K. Realizado esto, se continuó con *Next Step: Generate Equilibration and Dynamics Inputs*.

Se descargó los archivos generados en la opción *download.tgz*. Este archivo al ser descomprimido contará con todos los elementos necesarios para realizar la optimización

y equilibrio NPT del segmento ingresado dentro de la carpeta *gromacs*. Esta carpeta contiene también un script de nombre *README* que facilitará el procedimiento, ya que a este solo se le necesita dar permisos de ejecución y ejecutarlo para que la optimización se realice de forma automática. Los permisos se concedieron desde una terminal y luego se ejecutó el comando:

```
./README
```

Dado que todos los procesos son iniciados de forma automática, solo se necesita esperar que terminen para lo cual se debe estar visualizando la terminal abierta. Una vez terminado el procedimiento, se generó un archivo *.gro* con el número del paso final de la dinámica. Este archivo se puede convertir a formato *.pdb* utilizando el comando:

```
gmx editconf -f "archivo_final".gro -o "cualquier_nombre".pdb
```

Finalmente, el archivo *.pdb* se abrió con el visualizador UCSF Chimera, donde en *Select >Residue* se seleccionó la opción *all nonstandard* y luego *Actions >Atoms/Bonds >Delete* para eliminar la membrana, las aguas y iones agregados con CHARMM y así guardar la estructura optimizada y limpia.

2.3.4. Descripción de enlaces

Es necesario verificar que los enlaces ya existentes en la estructura se hayan conservado a pesar de las adiciones realizadas. Los más importantes en este trabajo son los enlaces disulfuros y pueden ser visualizados en Chimera colocando el comando **disp:cys** en *Favorites >Command line* con la estructura previamente abierta.

En el archivo *.pdb* de las estructuras originales utilizadas como modelo se puede visualizar una sección que tiene la descripción *SSBOND* donde se indican los enlaces disulfuros descritos en dicha estructura. Estos deben ser reconocidos por Gromacs para que puedan conservarse unidos, por lo que la comprobación de esto se realizó utilizando el comando que permite generar el archivo de topología de la proteína ingresada, el cual es el siguiente:

```
gmx pdb2gmx -f "monomero_renumerado".pdb -o "monomero_renumerado".gro  
-ss -merge all -chainsep ter
```

Usando la instrucción *-ss* del comando, el programa preguntará si se quiere unir determinados azufres para formar el puente disulfuro, *-merge all* unirá todas las cadenas en una sola topología y *-chainsep ter* indicará que en cada TER termina una cadena. Con esto se comprobó que los enlaces descritos en los .pdb iniciales fueran reconocidos por Gromacs y se observó si existían nuevos enlaces como resultado de las estructuras añadidas.

Los enlaces que no fueron reconocidos por el programa tuvieron que ser descritos usando Chimera. Para ello, en la estructura con los átomos de las cisteínas activas, se seleccionó 2 átomos de azufre que estuvieran involucrados en un enlace, luego se ingresó al menú de *Build Structure* y se seleccionó la opción de *Adjust Bonds*. Aquí, se cambió la opción de *reasonable* por *all possible* y se le dio click a *Add*. Seguidamente, se seleccionó el nuevo enlace creado y se ajustó su tamaño a 2Å para que pudiera ser reconocido por Gromacs. Este procedimiento se repitió con todos los enlaces de la estructura que no se encontraban declarados, tanto en el monómero como en el dímero posteriormente desarrollado. Es importante mencionar que esta técnica solo se puede realizar si es que los átomos iniciales no están muy separados, de lo contrario podrían crearse problemas en la simulación.

2.3.5. Construcción del dímero

Para la construcción del dímero se creó una copia del archivo que contenía la estructura del monómero desarrollada, esta se abrió en Chimera y en *Tools >Structure Editing >Change Chain IDs* y se cambió las cadenas *A* y *B* por *C* y *D*, y luego, se guardó la estructura.

A continuación, en el programa Pymol se abrió los dos monómeros y la estructura dimérica del modelo inicial usado para construir el ectodominio. Usando la línea de comandos de Pymol se ingresó el comando observado a continuación, donde *monomero1* es el nombre del archivo con el monómero construido, *monomero2* la copia del *monomero1* y, *dimero*, el ectodominio usado como modelo inicial. *Chain A* y *C* corresponden a la identificación de la subunidad α en el archivo original.

```
align "monomero1" in "dimero", chain A  
align "monomero2" in "dimero", chain C
```

Se eliminó la estructura del dímero inicial y se guardó el dímero construido en *File >Export Molecule >PDB Options*, marcando las opciones de *Write segment identifier (segi) column* y *Retain atoms ids*, al presionar *Save* se cambió el tipo de archivo a PDB (*.pdb *.pdb.gz).

2.3.6. Adición de glicanos

La adición de glicanos a la estructura del IR se realizó en dos partes. Primero, se utilizó el servidor de GlyProt para evaluar que los sitios de glicosilación descritos se encontraran disponibles, así como para obtener la descripción estructural de los glicanos. Y luego se utilizó el programa DoGlycans para obtener la topología del sistema y así poder realizar la simulación.

2.3.7. Optimización de la estructura completa del IR

Se construyó el sistema de trabajo con Gromacs, colocando a la proteína en el centro de una caja triclinica con un volumen total de 11325.77 nm^3 . Esta se solvató utilizando el modelo de agua TIP4P¹¹⁸ y se añadieron los iones sodio (Na^+) correspondientes para neutralizar el sistema. Una vez construido se procedió con la minimización donde se usó el algoritmo "steepest descent".

Antes de realizar la simulación de dinámica molecular se realizaron dos equilibrios del sistema. El primero fue usando un ensamble NVT (número de partículas, volumen y temperatura constante) a una temperatura de referencia de 309.65 K alcanzada con el termostato *v-rescale* durante 50 ps, teniendo como constante de tiempo 0.1 ps. Luego, se utilizó un ensamble NPT (número de partículas, presión y temperatura constante) a la misma temperatura antes indicada y durante el mismo tiempo, usando un acoplamiento de presión isotrópico Parrinello-Rahman con una constante de tiempo de 2 ps y presión de referencia de 1 bar. Para ambos casos se restringió la zona correspondiente a la sección transmembrana utilizando una constante de fuerza de $10000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

Finalmente, se realizó una dinámica de 100 ns a partir de la estructura equilibrada usando un ensamble NPT a la misma temperatura que se trabajó los equilibrios

anteriores y 1 bar de presión, usando un acoplamiento de presión isotrópico Berendsen con una constante de tiempo de 1 ps y compresibilidad isotérmica de $4.5 \times 10^{-5} \text{ bar}^{-1}$. Aquí también se mantuvo la restricción de la región transmembrana. Además las interacciones electrostáticas se calcularon con el método de Ewald de malla de partículas (PME)¹¹⁹ usando 1.2 nm para el corte electrostático de corto alcance y de van der Waals. Asimismo, se aplicaron condiciones periódicas de contorno en las tres dimensiones.

Tanto la simulación de dinámica molecular como los equilibrios se realizaron con un intervalo de tiempo de integración de 1 fs y la información de salida de las energías se guardaron cada 500 ps, mientras que la información de las trayectorias se guardaron cada 500 ps para los equilibrios y cada 15000 ps para la dinámica molecular.

Cabe mencionar que todas las imágenes de la estructura del IR visualizadas en el Capítulo 3 se obtuvieron con el visualizador de UCSF Chimera.

2.3.8. Evaluación de la estructura terciaria optimizada

La evaluación de la estructura se realizó utilizando el software de Gromacs con el cual se obtuvo las gráficas y valores de la Desviación Cuadrática Media (RMSD), Fluctuación Cuadrática Media (RMSF), Radio de Giro y Puentes de Hidrógeno. Adicionalmente, se usó el servidor de MolProbity¹¹⁵ para obtener el diagrama de Ramachandran.

2.4. Diagrama de flujo de actividades

El diagrama de flujo de actividades del presente trabajo de investigación puede ser observado en la siguiente página.

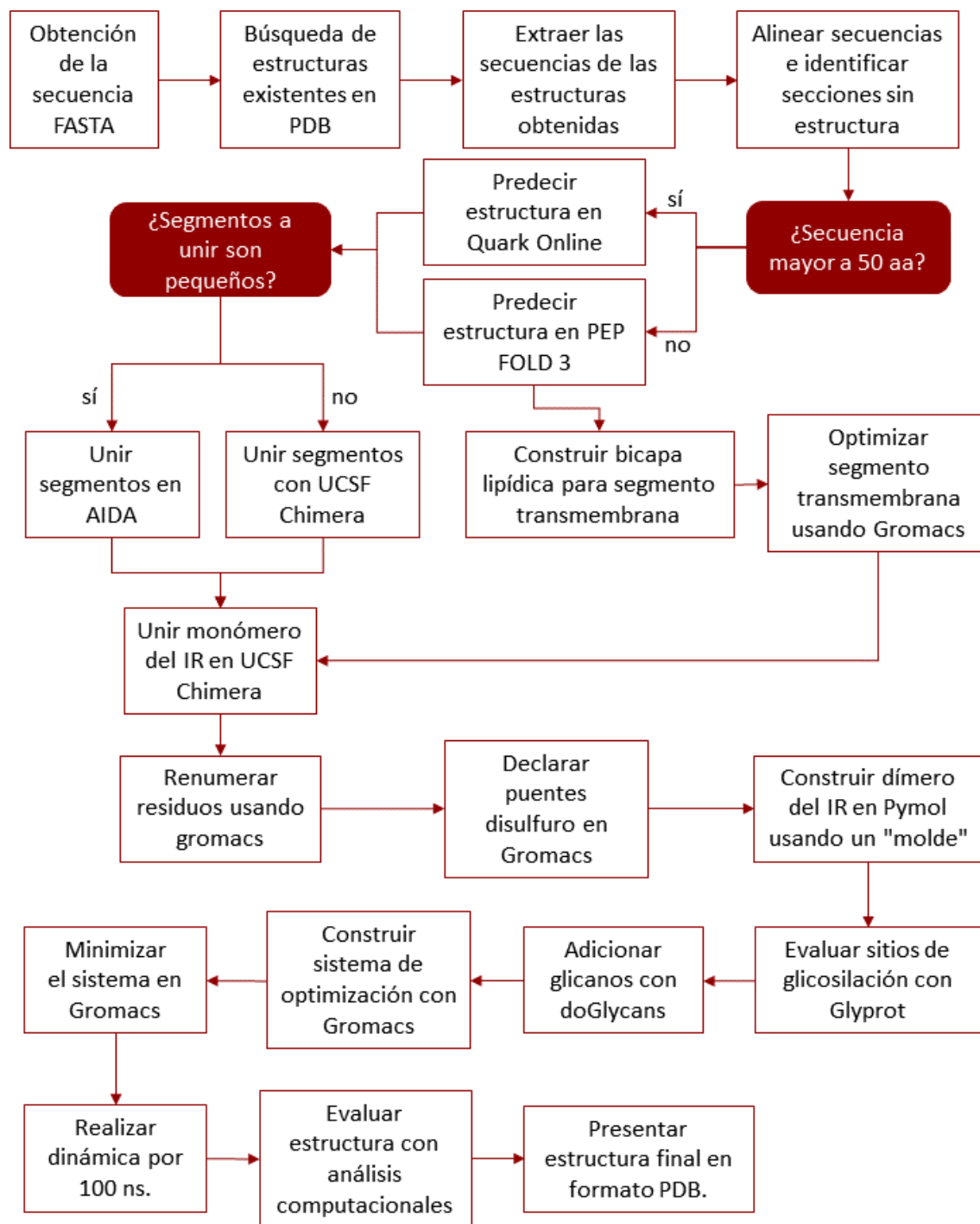


Figura 2.5: Diagrama de flujo de actividades

Capítulo 3

Resultados y Discusión

Una búsqueda en la base de datos del NCBI de *Insulin Receptor* en *protein*, dio como resultado una secuencia completa de 1382 aa con código de acceso P06213-4 y última fecha de actualización el 17 de Junio del 2020. De este resultado se extrajo la secuencia FASTA de la isoforma B o larga y en UniProt se encontró la secuencia de 1370 aa de la isoforma A o corta con código de acceso P06213-2. El alineamiento de ambas secuencias, realizado en ClustalW¹²⁰, se puede observar en la Tabla 3.1 donde de color rojo se observa los 12 aa de más que presenta la isoforma B.

De ambas secuencias, se decidió construir la estructura de la isoforma larga, ya que esta presenta una mayor relación con los problemas metabólicos, especialmente en la Diabetes Tipo II.

3.0.1. Búsqueda y selección de dominios existentes

Se buscaron en la base de datos del PDB las estructuras más recientes que se obtuvieran al buscar en *Advanced Search: Deposition >Structure Title >insulin receptor ectodomain or extracellular region* e *Deposition >Structure Title >Insulin Receptor and kinase*, filtrando la búsqueda por solo estructuras pertenecientes a *Homo sapiens* y que estén clasificadas como *Protein* en *Polymer Entity Type*. Descartando los resultados no relevantes y las estructuras muy incompletas o antiguas, la búsqueda dio 6 modelos para el ectodominio, cuyos ID fueron 4ZXB²⁶, 6CE9⁴, 6SOF¹²¹, 6CEB⁴, 6PXV¹²² y 5KQV¹²³, y 10 resultados para el segmento intracelular con los ID 3EKK¹²⁴, 4XLV³¹, 3EKN¹²⁵, 3BU6¹²⁶, 3BU3¹²⁶, 1GAG¹²⁷, 4IBM¹²⁸, 3BU5¹²⁶, 1RQQ¹²⁹ y 5HHW¹³⁰. Los

sp P06213.4	MATGRRGAAAAPLLVAVAALLGAAGHLYPGEVCPGMDIRNNLTRLHELENCVIEGHL	60
sp P06213-2	MATGRRGAAAAPLLVAVAALLGAAGHLYPGEVCPGMDIRNNLTRLHELENCVIEGHL	60
sp P06213.4	QILLMFKTRPEDFRDLSFPKLLIMITDYLLLFRVYGLESKDLFPNLTVIRGSRLFFNYAL	120
sp P06213-2	QILLMFKTRPEDFRDLSFPKLLIMITDYLLLFRVYGLESKDLFPNLTVIRGSRLFFNYAL	120
sp P06213.4	VIFEMVHLKELGLYNLMNITRGSVRIEKNNELCYLATIDWSRILDSVEDNYIVLNKDDNE	180
sp P06213-2	VIFEMVHLKELGLYNLMNITRGSVRIEKNNELCYLATIDWSRILDSVEDNYIVLNKDDNE	180
sp P06213.4	ECGDIICPGTAKGKTNCPATVINGQFVERCWTHSHCQKVCPTICKSHGCTAEGLCCHSECL	240
sp P06213-2	ECGDIICPGTAKGKTNCPATVINGQFVERCWTHSHCQKVCPTICKSHGCTAEGLCCHSECL	240
sp P06213.4	GNC SQDDPTKCVACRNFYLDGRCVETCPPYYHFQDWRCVNF SFCQDLHHKCKNSRRQG	300
sp P06213-2	GNC SQDDPTKCVACRNFYLDGRCVETCPPYYHFQDWRCVNF SFCQDLHHKCKNSRRQG	300
sp P06213.4	CHQYVIHNNKCIPECP SGYTMNSSNLCTPCLGPCPKVCHLEGEKTIDSVTSAQELRGC	360
sp P06213-2	CHQYVIHNNKCIPECP SGYTMNSSNLCTPCLGPCPKVCHLEGEKTIDSVTSAQELRGC	360
sp P06213.4	TVINGSLIINIRGGNNLAAELEANLGLIEEISGYLKIRRSYALVLSLFFRKLRLIRGETL	420
sp P06213-2	TVINGSLIINIRGGNNLAAELEANLGLIEEISGYLKIRRSYALVLSLFFRKLRLIRGETL	420
sp P06213.4	EIGNYSFYALDNQNLRLQWDWSKHNLTITQKGLFFHYNPKLCLSEIHKMEEVSGTKGRQE	480
sp P06213-2	EIGNYSFYALDNQNLRLQWDWSKHNLTITQKGLFFHYNPKLCLSEIHKMEEVSGTKGRQE	480
sp P06213.4	RNDIALKLTNGDQASCENELLKFSYIRTSFDKILLRWEPYWPP6DFRDLGFMFLFYKEAPYQ	540
sp P06213-2	RNDIALKLTNGDQASCENELLKFSYIRTSFDKILLRWEPYWPP6DFRDLGFMFLFYKEAPYQ	540
sp P06213.4	NVTEFDGQDACGSNSWTVVDIDPPLRSNDPKSQNHPGWLMRGLKLPWTQYAFVKT LVTFS	600
sp P06213-2	NVTEFDGQDACGSNSWTVVDIDPPLRSNDPKSQNHPGWLMRGLKLPWTQYAFVKT LVTFS	600
sp P06213.4	DERRTYGAKSDIIVYQTDATNPSVPLDPI SVSNSSSQIILKWKPPSDPNGNITHYL VFW	660
sp P06213-2	DERRTYGAKSDIIVYQTDATNPSVPLDPI SVSNSSSQIILKWKPPSDPNGNITHYL VFW	660
sp P06213.4	RQAEDSEL FELDYCLKGLKLP SRTWSP PFESEDSQKHNQSEYEDSAGECCSCP KTD SQIL	720
sp P06213-2	RQAEDSEL FELDYCLKGLKLP SRTWSP PFESEDSQKHNQSEYEDSAGECCSCP KTD SQIL	720
sp P06213.4	KELEESSFRKTFEDYLHNVV FVP RKTS SGT GAEDP RPSRKRRSLGDVGNVTVAAPTVA AF	780
sp P06213-2	KELEESSFRKTFEDYLHNVV FVP - - - - - RPSRKRRSLGDVGNVTVAAPTVA AF	780
sp P06213.4	PNTSSTSVPTSPEEHRPFEKVVNKE S LVISGLRHFTGYRIELQACNQDTPEERCSVAAYV	840
sp P06213-2	PNTSSTSVPTSPEEHRPFEKVVNKE S LVISGLRHFTGYRIELQACNQDTPEERCSVAAYV	840
sp P06213.4	SARTMPEAKADDIVGPVTHEIFENNVVHLMWQEPKEPNGLIVLYEVSRYRYGDEELHLCV	900
sp P06213-2	SARTMPEAKADDIVGPVTHEIFENNVVHLMWQEPKEPNGLIVLYEVSRYRYGDEELHLCV	900
sp P06213.4	SRKHFALERGCRLRGLSPGNYSVIRATSLAGNGSWTEPTYFYVTDYLDVPSNIAKIIIG	960
sp P06213-2	SRKHFALERGCRLRGLSPGNYSVIRATSLAGNGSWTEPTYFYVTDYLDVPSNIAKIIIG	960
sp P06213.4	PLIFVFLFSVVI GSIYLF L RKRQPDG PLGPLYASSNPEYLSASDVFPCS VYVPDEWEVSR	1020
sp P06213-2	PLIFVFLFSVVI GSIYLF L RKRQPDG PLGPLYASSNPEYLSASDVFPCS VYVPDEWEVSR	1020
sp P06213.4	EKITLLRELQGGSFGMVYEGNARDIIKGEAETRAVKT VNESASLRRIEFLNEASVMKG	1080
sp P06213-2	EKITLLRELQGGSFGMVYEGNARDIIKGEAETRAVKT VNESASLRRIEFLNEASVMKG	1080
sp P06213.4	FTCHHVRL LGV VSKGQPTLVVMELMAHGD LKSYLRS LRPEAENNPGRPPPTLQEMIQMA	1140
sp P06213-2	FTCHHVRL LGV VSKGQPTLVVMELMAHGD LKSYLRS LRPEAENNPGRPPPTLQEMIQMA	1140
sp P06213.4	AEIADGMAYLNAKKFVHRDLAARNCMVAHFTVKIGDFGMTRDIYETDYRKGKGLLPV	1200
sp P06213-2	AEIADGMAYLNAKKFVHRDLAARNCMVAHFTVKIGDFGMTRDIYETDYRKGKGLLPV	1200
sp P06213.4	RWMAPE SLKDG VFTTS SDMW SFGV LWEITSLAEQPYQGLSNEQV LKFVMDGGYLDQPDN	1260
sp P06213-2	RWMAPE SLKDG VFTTS SDMW SFGV LWEITSLAEQPYQGLSNEQV LKFVMDGGYLDQPDN	1260
sp P06213.4	CPEVRTDLMRMCWQFNPKMRPTFLEIVNLLKDDLHPSFPEV SFFHSEENKAPES E LEME	1320
sp P06213-2	CPEVRTDLMRMCWQFNPKMRPTFLEIVNLLKDDLHPSFPEV SFFHSEENKAPES E LEME	1320
sp P06213.4	FEDMENVPLDRSSHQREEAGGRDGGSLGFKRSYEEHIPYTHMNGGKKNRILTLPRSN	1380
sp P06213-2	FEDMENVPLDRSSHQREEAGGRDGGSLGFKRSYEEHIPYTHMNGGKKNRILTLPRSN	1380
sp P06213.4	PS	1382
sp P06213-2	PS	1370
	*	

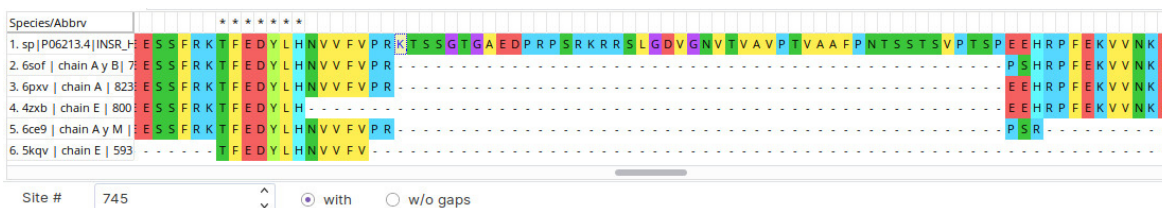
Tabla 3.1: Alineamiento de las secuencias FASTA de las 2 isoformas del IR.

archivos .pdb de cada estructura se descargaron y se eliminaron de ellos las cadenas pertenecientes a componentes no requeridos para la presente investigación, como inhibidores, glicósidos, aguas, entre otras moléculas, dejando finalmente la estructura como un monómero. Una vez realizado esto se obtuvo la secuencia FASTA de cada estructura utilizando un script.

Las secuencias del ectodominio se alinearon en Mega X con la secuencia completa del IR, obteniéndose así que la estructura 6SOF era la mejor para construir el ectodominio ya que contaba con solo un *gap* o rotura en el alineamiento, que correspondía desde el residuo LYS745 hasta el GLU794, contando los aminoácidos que no pudieron alinearse con la secuencia completa del IR. Esto es posible observar en la Figura 3.1a, en la cual también se puede apreciar que dicho segmento no ha sido descrito en las otras estructuras seleccionadas. Asimismo, se identificó que el modelo escogido tenía el ectodominio resuelto hasta el residuo LYS956 (Figura 3.1c), contando así con la mejor descripción del dominio mencionado hasta el momento.

Por otro lado, si se tuviera que elegir otras estructuras en lugar de la primera, las que presentan el código 6CE9 y 5KQV serían descartadas por encontrarse muy incompletas, quedando así la estructura 6PXV como segunda opción y 4ZXB como tercera. Otro detalle importante a destacar aquí es que al momento de realizar el dímero tenemos que asegurarnos de que las cisteínas involucradas en la formación del puente disulfuro que une ambas cadenas estén cerca, por lo que fue importante revisar dentro del archivo .pdb de cada estructura los enlaces descritos en la sección de *SSBOND* y así revisar que existieran enlaces entre CYS524 con CYS524 y entre CYS682 con CYS682, CYS683 o CYS685 de la otra subunidad alpha. Esto se observó en la estructura seleccionada mas no en las que quedaron en segundo y tercer lugar.

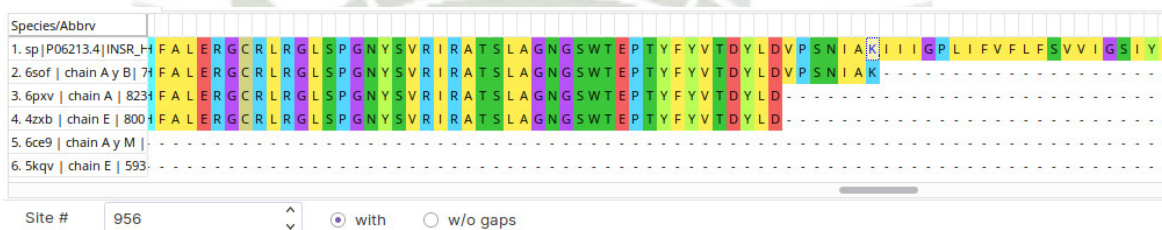
En cuanto al segmento intracelular, el alineamiento de 10 secuencias solución con la secuencia completa del IR demostró que la enzima tirosina quinasa se encuentra desarrollada casi en su totalidad. Una visión completa del alineamiento permitió descartar 3BU5, 3EKK y 3EKN debido a que presentaban zonas incompletas a lo largo de la estructura. Por otro lado, el inicio de estas secuencias plasmado en la Figura 3.2a, está mejor descrito por la estructura 5XLV. No obstante presenta un cambio en el aminoácido SER1008 donde en lugar de observarse una cisteína se ve una serina. La siguiente secuencia con buen desarrollo inicial es la 6HHW, sin embargo, esta presenta otra mutación dentro de la secuencia en la posición ASN1159.



(a) Único *gap* de la estructura 6SOF.



(b) Inicio del ectodominio



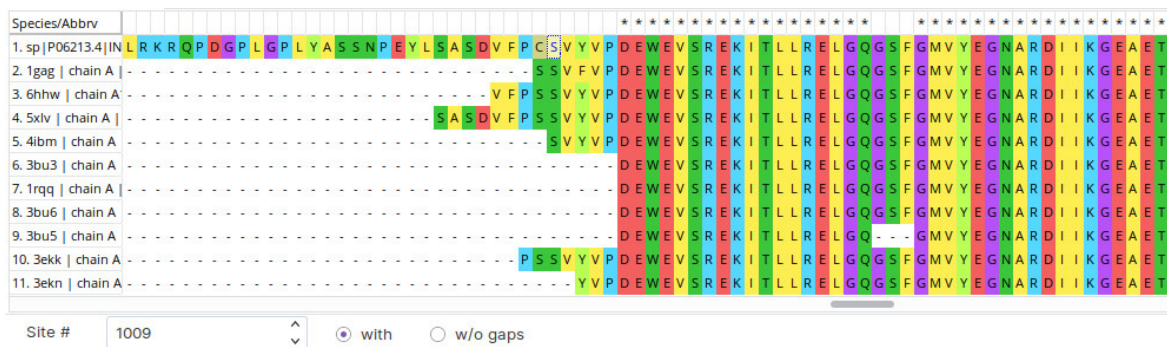
(c) Final del ectodominio

Figura 3.1: Alineamiento de las secuencias del ectodominio con IR completa.

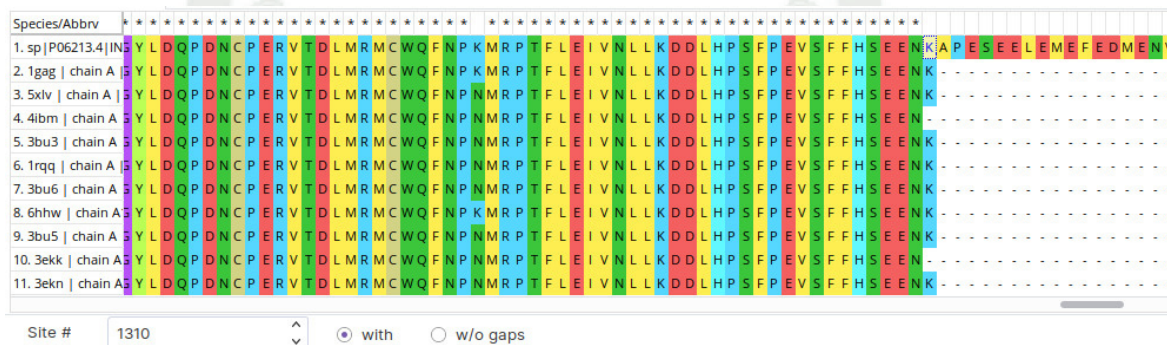
La secuencia 1GAG presenta también una mutación al inicio de la estructura, sin embargo ese pequeño segmento puede descartarse y empezar desde el residuo VAL1012. En cuanto a las secuencias 4IBM, 3BU3, 1RQQ y 3BU6, todas presentan una mutación en el residuo ASN1278 donde hay una asparagina en lugar de una lisina. Por último, si se observa el final de las secuencias podemos ver que casi todas, a excepción de 4IBM y 3EKK, describen la estructura del segmento intracelular hasta el residuo LYS1310 de la secuencia completa del IR (Figura 3.2b). El análisis permitió escoger el modelo 1GAG como el mejor para realizar la construcción de la estructura del IR. Si se requiriera de otro modelo, se utilizaría el 6HHW en segunda instancia.

Adicionalmente, se buscó *Deposition >Structure Title >Insulin Receptor AND transmembrane domain* donde se obtuvo una sola estructura con ID 2MFR²⁹ que representa un segmento del dominio transmembrana y jxtamembrana.

De este primer análisis de los modelos existentes para la construcción del IR también se notó que ninguna estructura del ectodominio describe los 27 primeros aminoácidos de la secuencia completa de IR. Esto se debe a que dicho segmento representa una



(a) Inicio del segmento intracelular



(b) Final del segmento intracelular

Figura 3.2: Alineamiento de las secuencias del segmento intracelular.

estructura señal la cual se escinde cuando el Receptor de Insulina llega a la membrana de la célula convirtiéndose este en un receptor maduro. Por ello, en la mayoría de publicaciones se toma al residuo HIS28 de la secuencia completa del IR como la HIS1. Con efecto de poder comparar la secuencia completa del IR con las estructuras halladas, hasta este momento se ha considerado los 27 primeros aminoácidos en la numeración. Sin embargo, de ahora en adelante, se suprimirá dicho segmento tomando en consideración a la histidina como el primer aminoácido de la estructura del IR.

3.0.2. Construcción del segmento transmembrana y juxtamembrana

Primero se buscó desarrollar la estructura del segmento medio que atraviesa la membrana, para ello se halló la secuencia entre los dos modelos escogidos. Como se vio previamente el modelo de ectodominio seleccionado (6SOF) termina en el residuo LYS929. Sin embargo, al comparar su secuencia con la de la estructura semi desarrollada del segmento transmembrana sin los 27 primeros aminoácidos podemos observar en la Figura 3.3 una superposición desde el residuo THR913 hasta el LYS929 y luego la estructura se extiende hasta el residuo LEU961. Dado que esta última se ha diseñado enfocada en



Figura 3.3: Alineamiento del segmento transmembrana con el IR y ectodominio seleccionado.

dicha zona, se decidió utilizar esa parte de la estructura para el diseño de la IR completa. Por ello se seleccionó la secuencia desde el aminoácido THR913 hasta el TYR984, ya que el segmento intracelular escogido comienza en el residuo VAL985. Dicha secuencia fue la siguiente:

TYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIIGSIYLFRLKRQPDGPLGPLYASSNPEYLSASDVFPSCSVY

Luego, la secuencia hallada del segmento transmembrana y juxtamembrana se ingresó al servidor de Swiss-Model, con la finalidad de que el servidor completara la estructura de los aminoácidos que faltaran. El resultado del servidor se puede ver en la Figura 3.4, como es posible observar, este también utilizó la estructura escogida anteriormente dado que es la única disponible en la actualidad. No obstante, faltaron dilucidar los últimos 23 aminoácidos que se querían hallar. Se descargó la estructura de este análisis pues el archivo original ofrecía muchos modelos para una misma secuencia, en cambio el servidor nos da un solo modelo ya optimizado facilitando el proceso de elegir alguno del archivo inicial.

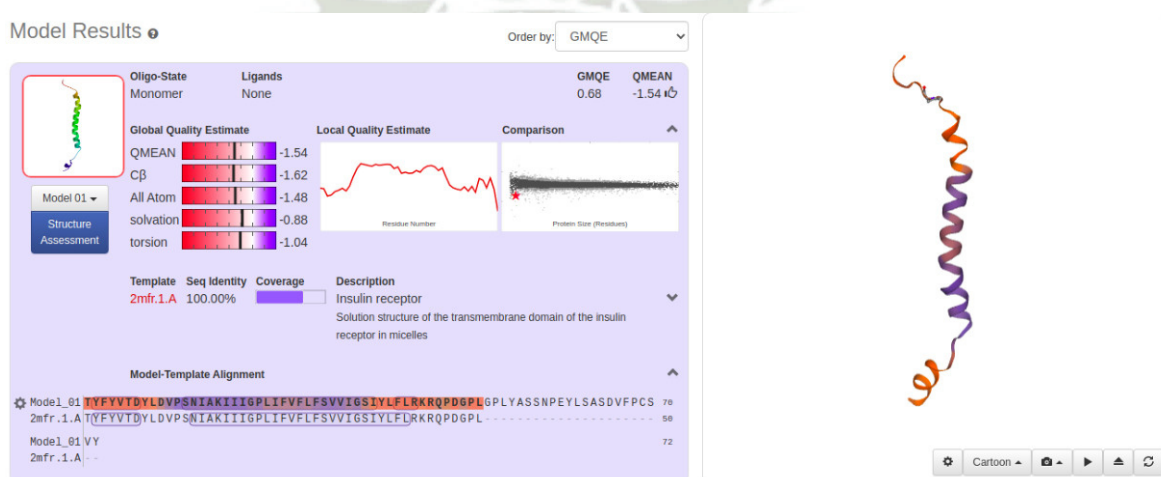


Figura 3.4: Modelo obtenido con Swiss-Model

La secuencia de aminoácidos no dilucidada se ingresó a PEP-FOLD3 para determinar

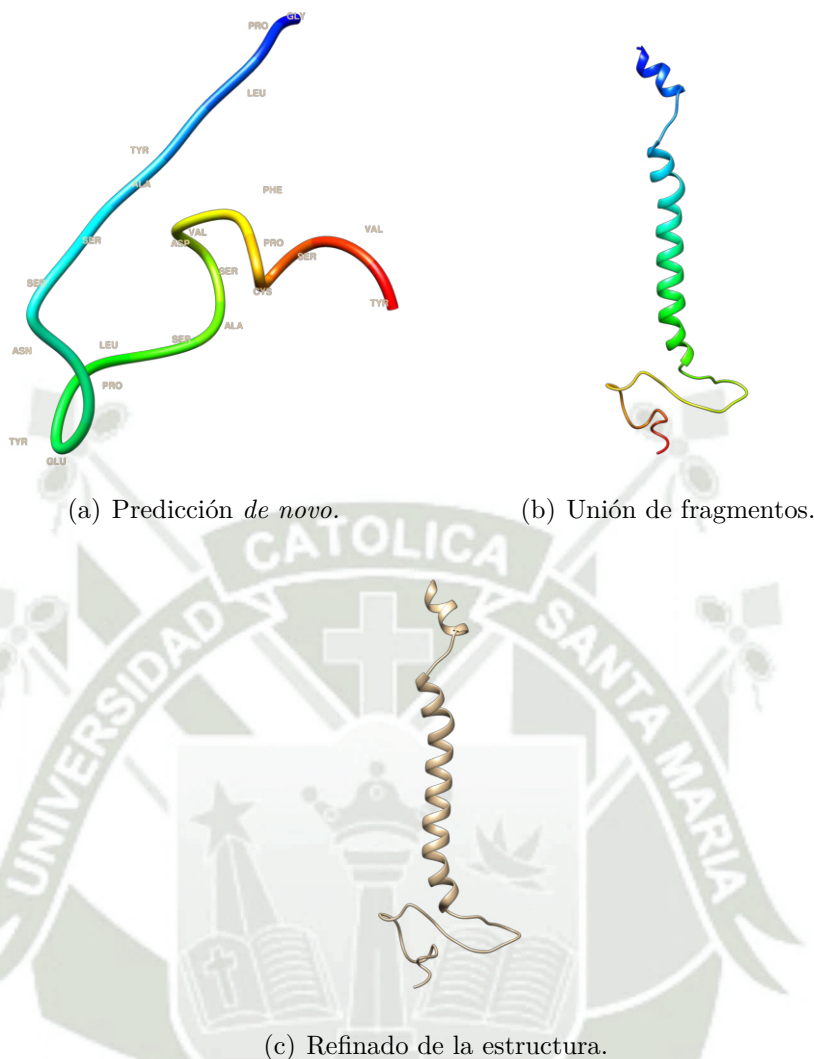


Figura 3.5: Construcción del dominio transmembrana y juxtamembrana. Gráficos visualizados con UCSF Chimera¹⁰⁰.

su estructura mediante una metodología *de novo*, este servidor dio como resultado 70 modelos de los cuales se escogió el primero pues contenía la puntuación sOPEP más negativa, lo cual significa que su energía potencial es menor. Esta estructura que se observa en la Figura 3.5a, se ingresó junto con el segmento de Swiss-Model a AIDA para unirlos. Luego, la estructura final (Figura 3.5b) se refinó en 3DRefine, dándonos finalmente la estructura que se observa en la Figura 3.5c. En este último servidor se tomó en cuenta los puntajes brindados, observando primero el puntaje de MolProbity donde un valor más pequeño es indicativo de un modelo físicamente más realista y en segundo lugar, los puntajes dados por $3D^{refine}Score$ y RWPlus, los cuales están relacionados a la energía potencial del modelo, el primero respecto al campo de fuerza de $3D^{refine}$ y el segundo según el potencial estadístico de RWplus. Para ambos casos, un valor menor es indicador de una mejor calidad del modelo.

Model #	3D ^{refine} Score	GDT-TS	GDT-HA	RMSD (Å)	MolProbity ▼	RWPlus	Refined Model	View Model
2	5121.09	1.0000	1.0000	0.204	2.068	-8872.271026	Refined Model 2	<input type="button" value="View"/>
1	6015.88	1.0000	1.0000	0.151	2.068	-8872.120699	Refined Model 1	<input type="button" value="View"/>
5	4814.04	1.0000	0.9931	0.294	2.114	-8872.099468	Refined Model 5	<input type="button" value="View"/>
3	4945.21	1.0000	0.9931	0.245	2.114	-8815.609713	Refined Model 3	<input type="button" value="View"/>
4	4863.76	1.0000	0.9931	0.267	2.183	-8823.674515	Refined Model 4	<input type="button" value="View"/>

Figura 3.6: Calificación de los modelos en 3DRefine.

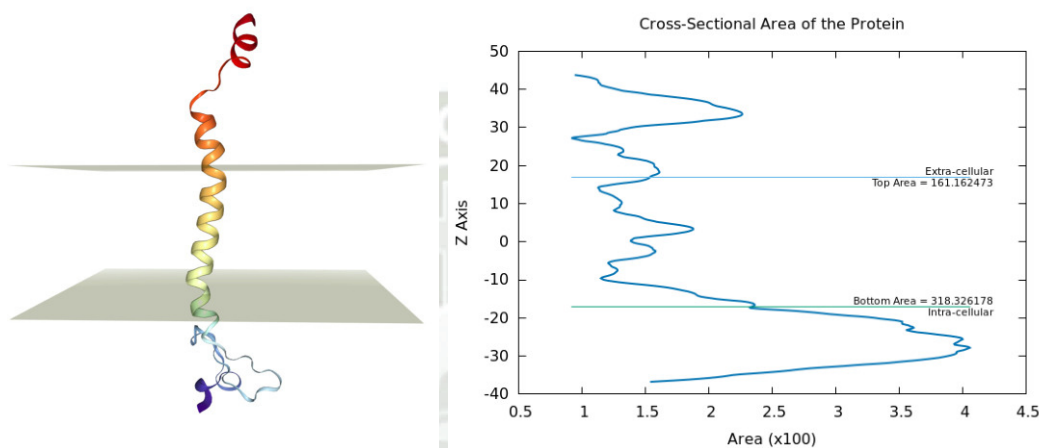


Figura 3.7: Alineamiento del segmento transmembrana en la bicapa lipídica.
Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

Los valores obtenidos para la estructura diseñada se pueden visualizar en la Figura 3.6, donde se puede notar que el menor valor de MolProbity se obtuvo con los modelos 1 y 2, no obstante, los valores de 3D^{refine}Score y RWPlus del modelo 2 fueron más bajos, haciendo que este se escogiera como el mejor modelo.

3.0.3. Optimización del segmento transmembrana y juxtamembrana

La estructura construida se ingresó al servidor de CHARMM-GUI para poder rodearla de una bicapa lipídica y optimizarla. Para alinear la proteína en la membrana se tomó en cuenta que el segmento transmembrana se encuentra entre los residuos ILE930 y LEU952. En la estructura armada hasta ese momento comprendía los residuos ILE18 hasta LEU40, por ello estos se seleccionaron como los aminoácidos para alinear la proteína en el eje Z. El alineamiento generado, así como la gráfica que identifica la zona que será embebida en la membrana se puede visualizar en la Figura 3.7.

Calculated XY System Size:		
	Upperleaflet	Lowerleaflet
Protein Area	161.162473	318.326178
Lipid Area	8190	8064
# of Lipids	130	128
Total Area	8351.162473	8382.326178
Protein X Extent	13.18	
Protein Y Extent	17.89	
Average Area	8366.74	
A	91.47	
B	91.47	

Figura 3.8: Tamaño del sistema que incluye el segmento transmembrana y juxtamembrana embebido en una bicapa lipídica.

Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

Para el agregado de lípidos en el siguiente paso, se tomó en consideración trabajos previos realizados con membranas en el CIIM, Centro de Investigación que pertenece a la Universidad Católica de Santa María. Por ello, se comenzó teniendo como base 128 lípidos DPPC para cada lado de la membrana y luego se revisó el tamaño del sistema calculado por el servidor. Este va a estar representado en una tabla que indica el área de la proteína, área lipídica, número de lípidos y área total, además de la extensión de la proteína en X y en Y, así como el área promedio. Inicialmente, al ingresar 128 DPPC en ambos lados, se observó una advertencia que indicaba que la parte superior de la bicapa lipídica podía contar con más lípidos, lo cual se demostraba al observar los valores de área total, donde el valor del *Upperleaflet* o capa superior tenía una área por debajo del *Lowerleaflet* o capa inferior. Es así que se aumentó el número de lípidos en la capa superior hasta un valor de 130, alcanzando un área total similar en ambos lados y aceptado por el servidor para continuar. Toda la información del tamaño del sistema construido se puede visualizar en la Figura 3.8.

En el siguiente paso se visualizó la estructura para asegurarnos que los lípidos estuvieran colocados correctamente, rodeando la proteína (Figura 3.9a). Por otro lado, en el siguiente paso no se obtuvo ningún error de malos contactos entre las colas de los lípidos y la proteína, por lo que inmediatamente se pasó al ensamblaje de componentes.

	1 cycle	2 cycle	3 cycle	4 cycle	5 cycle	6 cycle
BB	10.0	5.0	2.5	1.0	0.5	0.1
SC	5.0	2.5	1.0	0.5	0.1	0.0
wforce	2.5	2.5	1.0	0.5	0.1	0.0
tforce	2.5	2.5	1.0	0.5	0.1	0.0
mforce	2.5	2.5	1.0	0.5	0.1	0.0
ion	10.0	0.0	0.0	0.0	0.0	0.0

Figura 3.10: Parámetros para la optimización y equilibrado de la proteína en la membrana. Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

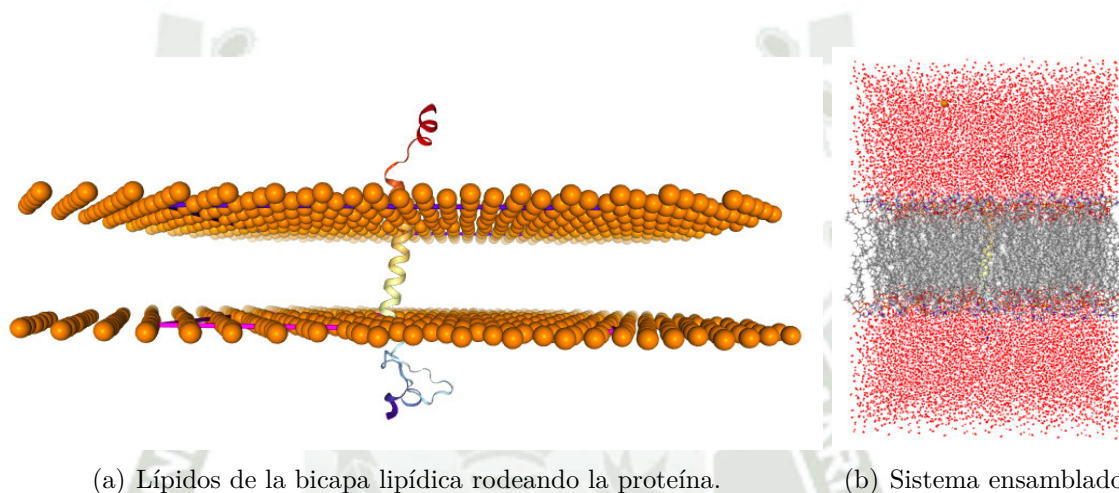


Figura 3.9: Construcción de la bicapa lipídica. Obtenido de: <http://www.charmm-gui.org/?doc=input/membrane.bilayer>

En el paso 5 se visualizó el sistema ensamblado detallado en la Figura 3.9b, donde se puede ver en el centro de la figura la membrana con la proteína embebida, las moléculas de agua rodeando la membrana representadas por el color rojo y el ion agregado para neutralizar la carga del sistema en color anaranjado. Finalmente, en el paso 6 se indicó que el sistema se ha preparado para realizar un esquema de equilibrado de 5 ciclos con la configuración observada en la Figura 3.10.

3.0.4. Construcción del ectodominio

Como se pudo observar en el alineamiento de la secuencia del ectodominio con el IR, existía un segmento de esta estructura cuya ubicación espacial en fase cristalina no había sido determinada. Dado que contaba con menos de 50 aa se decidió obtener el segmento usando PEP-FOLD3. Su secuencia se puede observar a continuación:

KTSSGTGAEDPRPSRKRRSLGDVGNVTVAVPTVAAFPNTSSTSVPPTSPEE

La estructura seleccionada de las obtenidas con PEP-FOLD3 se puede observar en la Figura 3.11a. Nuevamente se seleccionó aquella que contenía el valor sOPEP más negativo, para este caso el modelo 1. El archivo .pdb de este modelo se dividió en 2 archivos diferentes. En el primero se colocó la información de los 18 aminoácidos iniciales, ya que en el residuo ARG735 culminaba la primera cadena o subunidad α , la estructura de esta secuencia se puede observar en la Figura 3.11b. En el otro archivo se colocó los aminoácidos restantes de la estructura predicha, que eran los que iniciaban la subunidad β , Figura 3.11c. Luego, la estructura 6SOF limpia se abrió en UCSF Chimera y se procedió a añadir los segmentos faltantes que se encontraban en los archivos creados. Una vez ensamblados se guardó todo el ectodominio en formato .pdb, este puede ser observado en la Figura 3.11d, donde los segmentos de color azul y rojo corresponden a las partes añadidas.

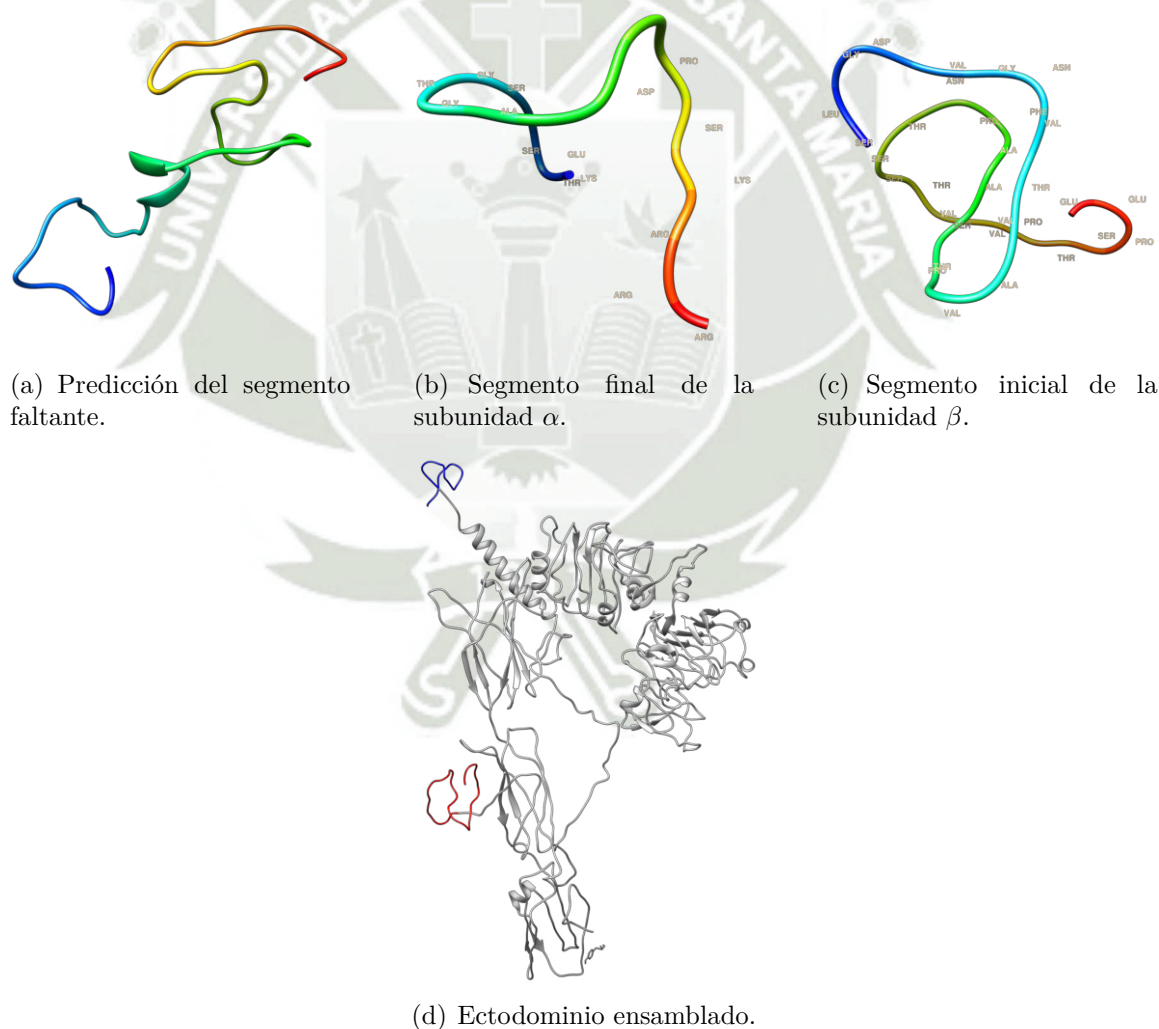


Figura 3.11: Construcción del ectodominio. Gráficos visualizados con UCSF Chimera¹⁰⁰.

La razón por la que el ectodominio se ensambló usando Chimera y no AIDA, es debido al gran tamaño de este y la necesidad de conservar las posiciones de la forma original para evitar el movimiento de los puentes disulfuros que ya se encuentran descritos en la estructura. Este visualizador que también permite la modificación de estructuras, da la opción de poder escoger los segmentos que se quieren mover y cuáles deben conservar su posición, permitiendo lograr el objetivo antes mencionado.

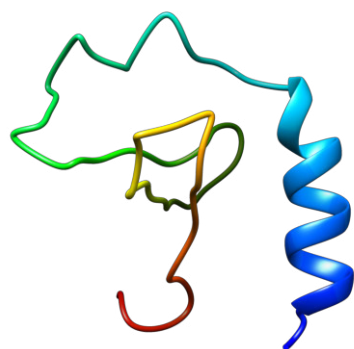
Luego, al final de esta estructura se le eliminó los aminoácidos que también presentaba el segmento transmembrana ensamblado, puesto que en esta última se utilizó como modelo una estructura que partió de una investigación enfocada en dilucidar dicho segmento, lo que llevó a pensar que esta estructura podría ser más acertada y adecuada para el modelado.

3.0.5. Construcción del segmento intracelular

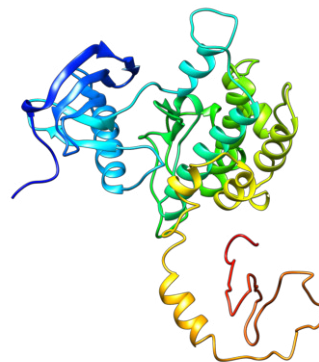
La secuencia del segmento intracelular también se tuvo que modelar usando una predicción *de novo*, sin embargo, dado que esta secuencia contaba con 72 aa, no se pudo utilizar PEP-FOLD3 para la predicción que solo permite un máximo de 50 aa. En su lugar se utilizó el servidor de Quark Online que acepta hasta 200 aa. La secuencia ingresada a QUARK online se puede observar a continuación:

APESSELEMEFEDMENVPLDRSSHQREEAGGRDGGSSLGFKRSYEEHIPYTHMNGGKKNRILTLPRSNPS

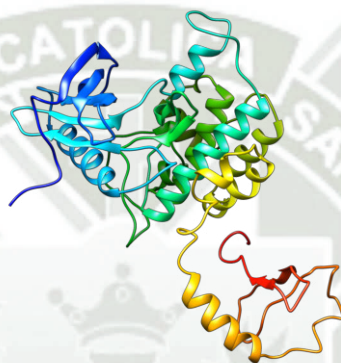
El servidor generó 5 modelos, de los cuales se escogió la primera estructura, observada en la Figura 3.12a para unirla al modelo que fue escogido anteriormente utilizando AIDA, lo cual nos dio la estructura observada en la Figura 3.12b. Por último, después del refinamiento en 3Drefine se seleccionó el modelo 4 por presentar el segundo valor de MolProbity, indicador de la posibilidad de un modelo realista, más bajo. Se escogió este y no el que tenía el valor más bajo debido a que este último presentaba la energía potencial más alta de todos los modelos lo cual disminuía su calidad, mientras que el modelo escogido tenía el segundo valor más bajo de energía potencial.



(a) Predicción del segmento faltante.



(b) Unión de segmentos del componente intracelular



(c) Segmento refinado

Figura 3.12: Construcción del segmento intracelular. Gráficos visualizados con UCSF Chimera¹⁰⁰.

3.0.6. Ensamblado de la estructura terciaria

Construcción del monómero

Como ya se contaba con la estructura de las tres secciones que conforman el monómero del Receptor de Insulina, se procedió a unirlos para tener una única estructura formada por dos subunidades. Para ello se comenzó uniendo el ectodominio con el segmento transmembrana usando *Join Models* de Chimera y de la misma forma se agregó el componente intracelular. En ese proceso fue necesario utilizar la opción de *Move selection* y así poder modificar las estructuras solapadas, provocado por la unión de los segmentos. Esta estructura final se guardó y se procedió con la renumeración de los residuos usando Gromacs.

Una vez realizado esto se utilizó Gromacs para obtener información sobre los puentes disulfuro y así declarar los faltantes. En la estructura utilizada como molde se había

descrito una unión intracatenaria entre CYS682 y CYS685, sin embargo, en el estudio realizado por Lu y Guidotti³⁵ se identificó que los enlaces disulfuro involucrados en la formación de dímeros pertenecen a los residuos CYS524 y CYS682, por lo que este último debería permanecer libre y así poder enlazarlo posteriormente con el mismo residuo del otro monómero. No obstante, en dicha investigación se menciona que también es posible observar el enlace descrito en dicha estructura en algunos casos donde el CYS682 no se encuentra disponible.

En esta investigación se prefirió tomar CYS682 para formar el enlace intercatenario y establecer un enlace entre CYS683 y CYS685 de cada monómero. Este procedimiento se realizó utilizando Chimera en *Tools > Structure Editing > Model/Refine Loops* donde se escogió la cadena A y luego se seleccionó la opción de *Chimera selection region*, aquí los residuos seleccionados en la estructura correspondieron a los ubicados entre GLU681 Y LYS687. El software utiliza el servidor de Modeller¹³¹ para predecir la estructura que puede tener el *loop* y brinda el número de modelos que se ha indicado previamente, para este caso 5. De todos los modelos brindados se escogió el modelo de 4 debido a que con dicha estructura los residuos CYS683 Y CYS685 se encontraban más cerca y dejaba expuesto al CYS682. Además, con la opción de *Rotamers* dentro de *Structure Editing* se corrigió, usando la librería de Dunbrack 2010¹³², el ángulo en el que se encontraría el azufre para así mejorar su posición respecto al otro. La diferencia entre la localización de dichos residuos en la estructura inicial y la modelada se puede observar en la Figura 3.13. Con esta modificación realizada se volvió a utilizar el comando de Gromacs para observar las cisteínas que se unían y se comprobó que el nuevo enlace era reconocido.

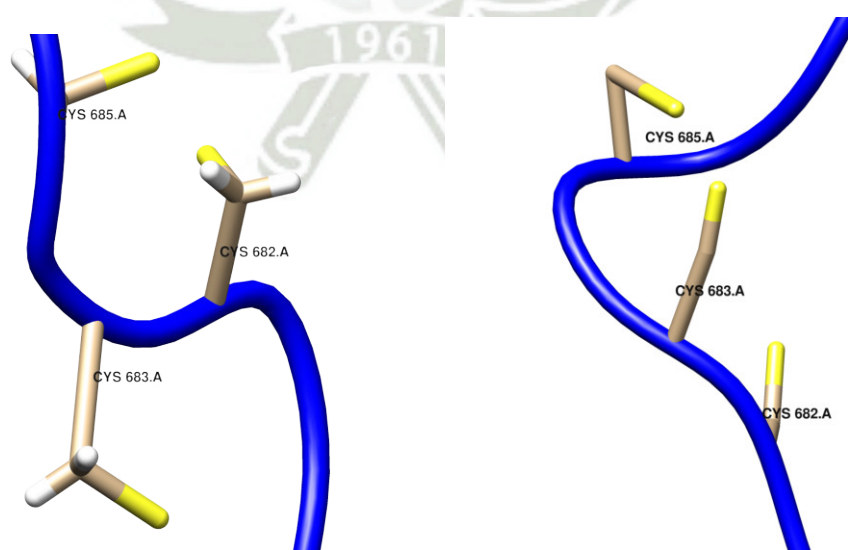


Figura 3.13: Modelado de enlaces disulfuro.

A la izquierda, estructura inicial y a la derecha, estructura después del modelado.

Por otro lado, el mismo comando que se utilizó para observar los puentes disulfuro indicó que había errores con 2 aa ya que su declaración se encontraba incompleta en el archivo de estructura, estos residuos correspondían a ILE388 y ARG717. Para corregir dicho error se utilizó el comando `swapaa` en el visualizador de Chimera, colocándolo en la línea de comandos. Este comando requiere indicar el aminoácido por el que se va a reemplazar la selección y la posición del residuo a modificar, de forma que el comando para el primer residuo erróneo quedaría como se ve a continuación:

```
swapaa ile #0:388.a
```

Dicho comando indica que se quiere reemplazar el aminoácido 388 de la cadena A y del modelo 0 por una isoleucina. El mismo procedimiento se siguió para corregir la arginina faltante. La estructura final se guardó y se volvió a ingresar en Gromacs el comando para obtener la topología.

Se observó un nuevo problema con la estructura, algunos residuos de histidina tenían un estado de protonación diferente al considerado por defecto para Gromacs. Esto se debía principalmente a que se utilizó diferentes modelos para construir el monómero de forma que los residuos HIS787, HIS832, HIS1057, HIS1058, HIS1081, HIS1130, HIS1142, HIS1268, HIS1278, HIS1307, HIS1331 y HIS1336, pertenecientes al segmento intracelular, presentaban un hidrógeno (HD1) unido a un nitrógeno ND1, mientras que Gromacs considera por defecto la presencia de un HE2 unido a un NE2. Esta diferencia es apreciable en la Figura 3.14 si comparamos la estructura del residuo HIS768 e HIS787. Para evitar tener que colocar manualmente los estados de protonación para dichos aminoácidos se utilizó de nuevo el comando `swapaa` reemplazando los residuos mencionados por histidina que al ser reemplazada no contiene hidrógenos. Estos son agregados después por Gromacs al momento de generar la topología usando la configuración por defecto.

Construcción del dímero

El dímero del Receptor de Insulina se construyó utilizando Pymol mediante el alineamiento de dos monómeros a la estructura dimérica 6SOF usada como modelo inicial, esto se puede visualizar en la Figura 3.15, donde de color celeste se observan los monómeros y de color rosado el modelo 6SOF. La estructura obtenida se guardó en formato .pdb y se abrió con Chimera para poder declarar los puentes disulfuro intercatenarios faltantes.

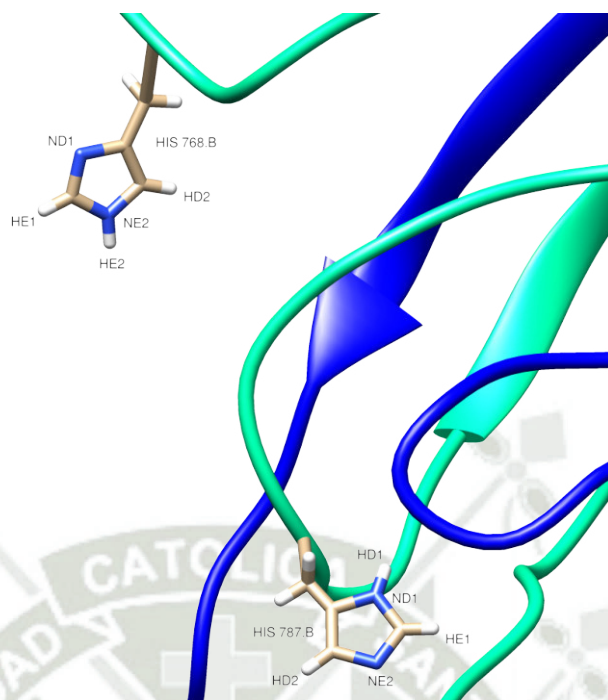


Figura 3.14: Diferencia en la protonación de los residuos de histidina.

Los enlaces intercatenarios declarados se pueden observar en la Figura 3.16a, así como los intracatenarios CYS683 - CYS685 que se formaron anteriormente. La estructura del dímero final con todos los puentes disulfuros encontrados en la literatura descritos se puede observar en la Figura 3.16b. En dicha figura se observa de color azul y rosado las subunidades α de cada monómero identificadas con nombre de cadena A y C, respectivamente. Mientras que de color celeste y morado se aprecian las subunidades β con identificación B y D.

Adición de glicanos

Se subió la estructura del IR en formato .pdb y luego se visualizó la información brindada por el programa. Gracias a la investigación realizada por Sparrow⁸, se conocía los sitios de interés para la glicosilación en el ectodominio, así como los glicósidos que debían añadirse en cada uno de esos sitios, esta información puede observarse en la Tabla 3.2. Por lo tanto, se comprobó que todos los sitios descritos en el artículo se encontraran reconocidos y disponibles para la adición de un glicano con el servidor. Una vez hecho esto, se procedió a seleccionar los glicanos que debían unirse a los sitios. La ventaja principal de este servidor es la función de búsqueda de glicósidos que te permite ingresar las cantidades de cada componente que conforma el glicano a agregar, así como su peso total. Esto se puede observar en la Figura 3.17. La búsqueda puede dar diferentes estructuras que cumplan con los criterios ingresados por lo que se ha

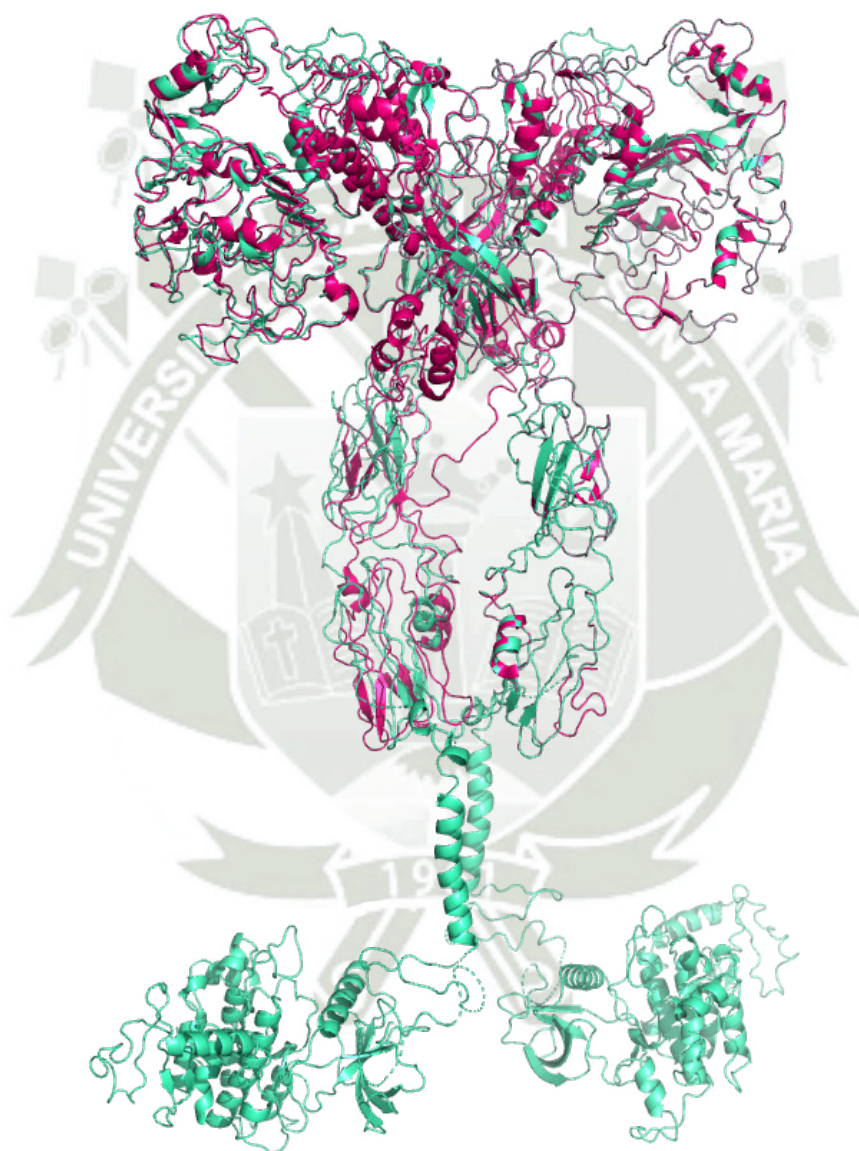
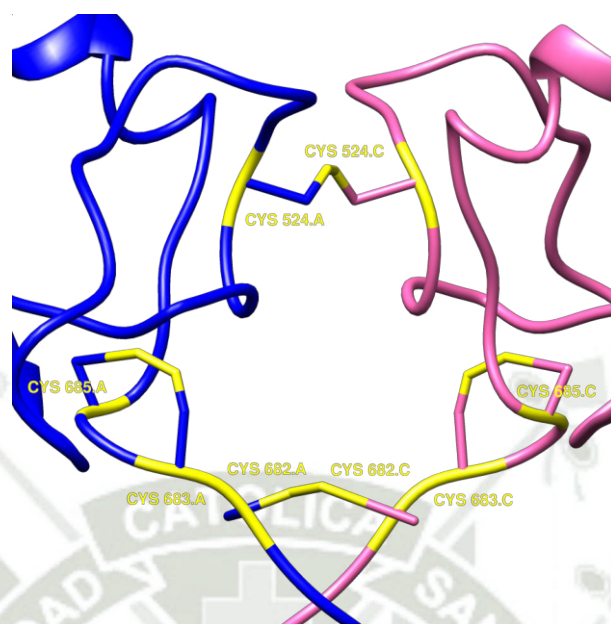
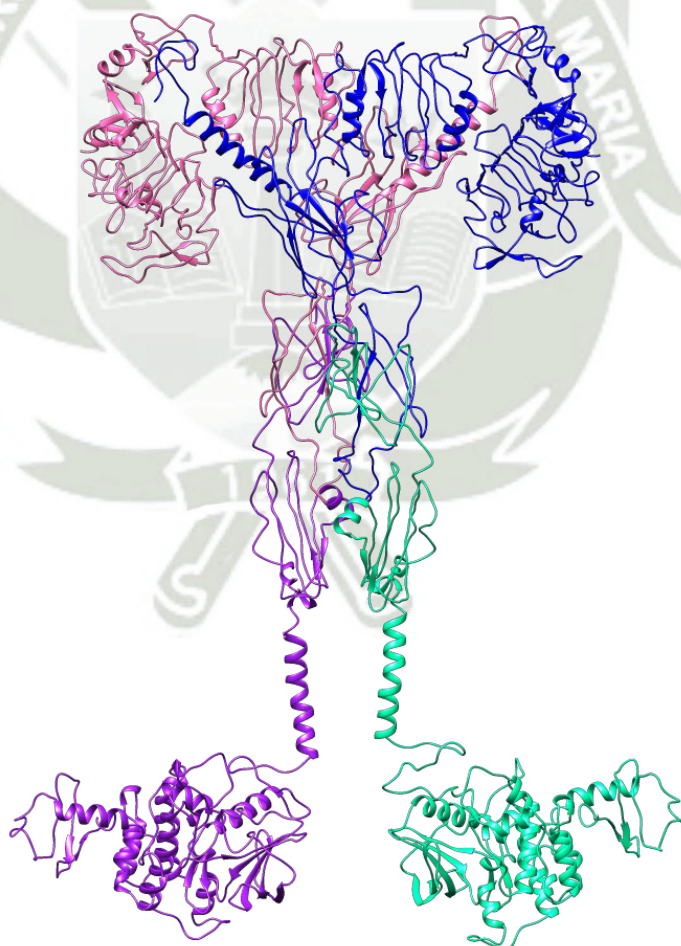


Figura 3.15: Alineamiento de monómeros al ectodominio 6SOF.



(a) Enlaces disulfuro intercatenarios.



(b) Dímero de IR.

Figura 3.16: Construcción del dímero del Receptor de Insulina.

Glycom-DB with 1021 Entries					
Antenna ? ▾		No of Monosaccharides ? ▾		Core Fuc ? ▾	
GlcNAc ? ▾	Manp ? ▾	Fucp ? ▾	Galp ? ▾	Neup5Ac ? ▾	Xylp ? ▾
Glc ? ▾	Galp6N ? ▾	Neup5Gc ? ▾	GalpNAc ? ▾	GlcN ? ▾	Sulfate ? ▾
Galf ? ▾	P ? ▾	Neup4Ac5Ac ? ▾	Manp3Me ? ▾	Galp3Me ? ▾	Neup5Ac9Ac ? ▾
Galp4Me ? ▾	Neup ? ▾				
Mono isotopic mass ? ▾			Average isotopic mass ? ▾		
Tolerance			Tolerance		
Query Database			reset		

Figura 3.17: Búsqueda de glicanos en Glyprot.

de seleccionar la más adecuada para el usuario. Cuando ya se han seleccionado todos los glicósidos a añadir se le da click a *Build Structure!* y se obtendrá la estructura en formato .pdb con los glicanos añadidos. Si bien, ya se cuenta con la estructura hasta este momento, surge el problema de que la estructura de los glicanos no se encuentra descrita en los campos de fuerza usados por Gromacs por lo que no se podrá obtener la topología de forma manual. Esto hace que se requiera de un programa capaz de generar la topología a utilizar por Gromacs, y para el caso de la presente tesis, configurado para ser usado con el campo de fuerza OPLS-aa.

Residuo	Composición del Glicósido
16	GlcNAc2Man5
25	GlcNAc4Man3Gal2Fuc
111	GlcNAc2Man9
215	GlcNAc2Man6
255	GlcNAc4Man3Gal2Fuc
295	GlcNAc4Man3Gal2Fuc
418	GlcNAc4Man3Gal2Fuc
514	GlcNAc2Man5
606	GlcNAc4Man3Gal2Fuc
624	GlcNAc4Man3Gal2Fuc
671	GlcNAc6Man3Gal4Fuc
742	GlcNAc4Man3Gal2Fuc
755	GlcNAc4Man3Gal2Fuc
893	GlcNAc4Man3Fuc

Tabla 3.2: Glicósidos más abundantes en cada residuo glicosilado del IR.

Es aquí donde se necesita del segundo paso. Se volvió a construir la estructura del IR glicosilado pero usando el programa doGlycans que es capaz de generar la topología pertinente para trabajar en Gromacs. No obstante, este programa es más complejo de manejar que el anteriormente mencionado pues la estructura de los glicanos debe ser

descrita en un archivo llamado *sequence_file.seq* usando el formato GLYCAM. En este archivo también se tiene que especificar la posición en la que se añadirá el glicano y los ángulos que deberán presentar los átomos de este para no causar problemas estéricos. Es aquí donde sirve la información obtenida con el primer servidor. Gracias a este se contaba con las estructuras de los glicanos conseguidos a través del servicio de búsqueda de Glyprot. Además, se pudo obtener los ángulos que debían formar los enlaces en la estructura del glicano al abrir la estructura descargada del primer servidor con UCSF Chimera y usando la opción de *Structure Measurements* localizada en *Tools >Structure Analysis >Angles / Torsions*.

Este programa pidió además, generar una primera topología de la estructura proteica sin incluir los glicanos y usando la opción *gmx pdb2gmx* de Gromacs. El archivo generado fue modificado comentando todas las líneas que no fueran parte de [*moleculetype*], [*atoms*], [*bonds*], [*angles*] y [*dihedrals*]. Realizado esto, se cambió la extensión del archivo de .top a .itp. Luego, se procedió a utilizar el programa desde la terminal, como la indica su manual, donde se requiere añadir el archivo .pdb con la estructura de la IR no glicosilada, el archivo .itp de la topología previamente creado y el *sequece_file.itp* con la información de los glicanos. Además, se especificó que se trabajaría con el campo de fuerza OPLS (-opls). Tras realizar esto, se obtuvo la estructura .pdb glicosilada y el archivo de topología con la información de los glicanos. A este último se le modificó el tipo de átomo y la carga de los sitios donde se adicionó el glicano según el manual de doGlycans.

Adicionalmente, se tuvo que corregir la carga de algunos glicanos para obtener un valor entero que pudiera ser neutralizado. Finalmente, en el archivo .top se utilizó el comando *#include* para indicar todos los archivos que debían ser leídos junto con la topología y que vienen por defecto al instalar el programa, más el .itp obtenido después de la adición de glicanos. En este archivo también se describió que se adicionaría un archivo de restricción *posre.itp* y se describió los iones que se utilizarían para neutralizar el sistema, ya que estos no se encontraban declarados en el campo de fuerza usado por el programa. Con esta modificación final, ya se podía utilizar la topología para realizar la optimización del sistema.

3.0.7. Optimización del Receptor de Insulina

La primera generación de la topología mostró la formación de 40 enlaces disulfuros, de estos 19 enlaces eran intracatenarios por cada monómero y 2 pertenecían a los

intercatenarios. Los enlaces declarados se pueden observar a continuación:

Intracatenarios:

- CYS-8 SG-12993 y CYS-26 SG-13139
- CYS-126 SG-13969 y CYS-155 SG-14210
- CYS-159 SG-14236 y CYS-182 SG-14401
- CYS-169 SG-14302 y CYS-188 SG-14454
- CYS-192 SG-14485 y CYS-201 SG-14548
- CYS-196 SG-14513 y CYS-207 SG-14587
- CYS-208 SG-14593 y CYS-216 SG-14650
- CYS-212 SG-14624 y CYS-225 SG-14717
- CYS-228 SG-14735 y CYS-237 SG-14814
- CYS-241 SG-14843 y CYS-253 SG-14957
- CYS-259 SG-15006 y CYS-284 SG-15217

- CYS-266 SG-15066 y CYS-274 SG-15130
- CYS-288 SG-15247 y CYS-301 SG-15341
- CYS-304 SG-15361 y CYS-308 SG-15386
- CYS-312 SG-15415 y CYS-333 SG-15573
- CYS-435 SG-16405 y CYS-468 SG-16657
- CYS-647 SG-18135 y CYS-872 SG-20703
- CYS-683 SG-18420 y CYS-685 SG-18432
- CYS-798 SG-19561 y CYS-807 SG-19696

Intercatenarios:

- CYS-524 SG-4213 y CYS-524 SG-17143
- CYS-682 SG-5484 y CYS-682 SG-18414

Era importante que estos fueran declarados para evitar que Gromacs coloque hidrógenos en los átomos de azufre que forman enlaces y así se reconozca la unión entre cisteínas hasta el final de la dinámica.

La construcción del sistema de trabajo constó, primero, de la modificación del tamaño y forma de la caja, seleccionándose como adecuada la forma triclínica que permite determinar el tamaño de la caja en los ejes x , y y z . Esto se realizó con la finalidad de aprovechar el espacio pues la estructura del receptor de insulina presenta una forma alargada en el eje y . De colocarse una estructura de cubo, habría mucho espacio sobrante alrededor de la proteína que, posteriormente, sería llenado con moléculas de agua, incrementándose en gran medida el número de átomos del sistema. Dicha construcción es poco favorable porque aumenta el tiempo de cálculo del sistema, que ya de por sí es muy grande, e implica un alto costo computacional.

Después, se procedió con la solvatación donde se añadieron 329705 moléculas de solvente TIP4P. Posterior a esto se realizó la neutralización del sistema, reemplazando 58 moléculas de agua por iones Na^+ . El sistema construido se puede observar en la

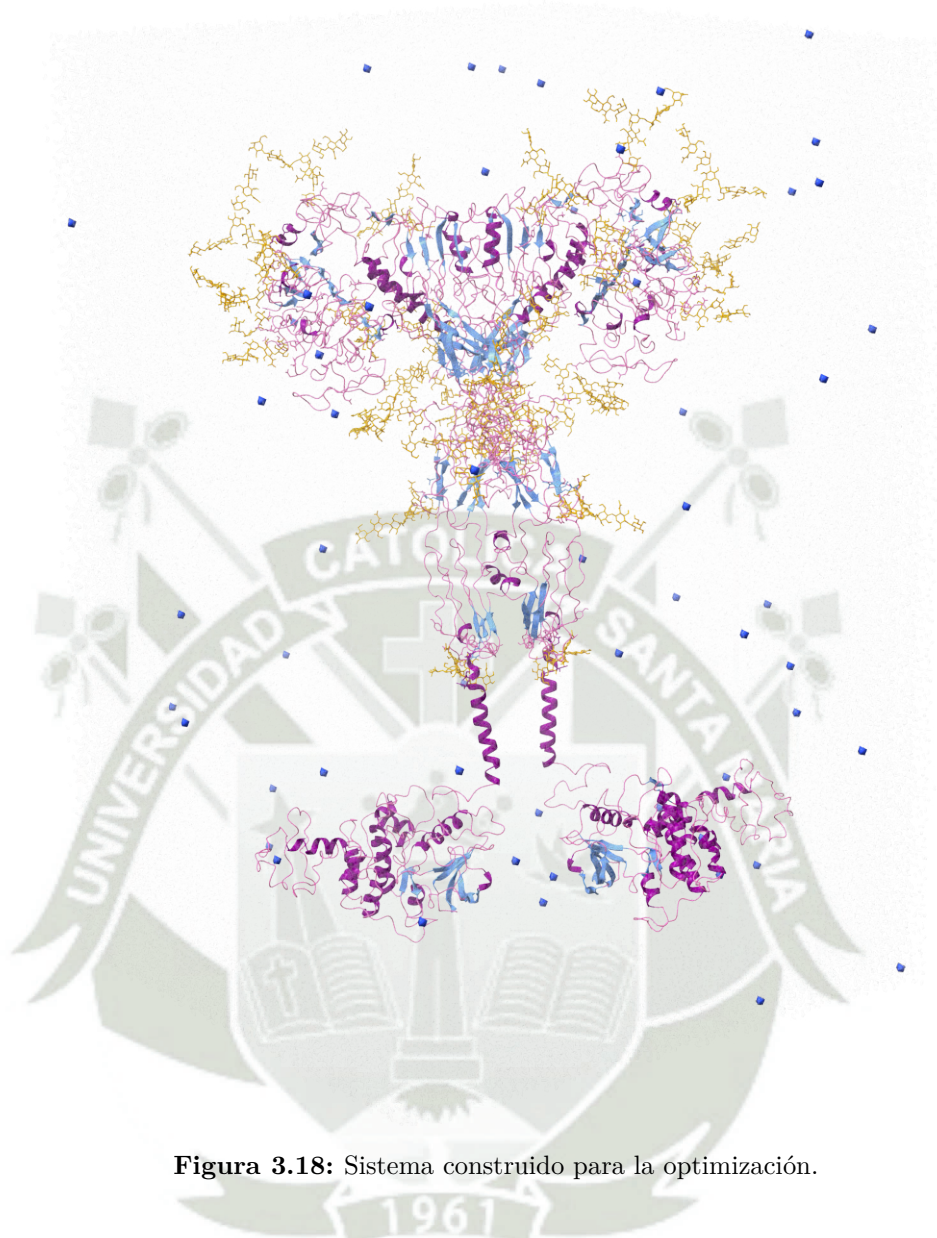


Figura 3.18: Sistema construido para la optimización.

Figura 3.18. Las moléculas de agua se visualizan de color gris, la proteína está coloreada según su estructura secundaria y los glicanos se muestran en color anaranjado dentro de la caja. Asimismo, los iones añadidos se observan como esferas de color morado.

Se realizaron 5 minimizaciones al sistema para reducir la energía potencial y evitar solapamientos mediante la reorientación de cadenas secundarias. Con esta metodología se busca alcanzar un mínimo local de energía que se acerque conformacionalmente a la estructura nativa¹³³. Sin embargo, como se ha indicado, la minimización no garantiza encontrar un mínimo global, que es el principal objetivo de las predicciones estructurales basadas en métodos de minimización. Para ello, será necesario realizar una o varias dinámicas moleculares utilizando parámetros adicionales como temperatura y presión que le permitan a la estructura encontrar la mejor conformación en condiciones simi-

lares a las naturales. En definitiva, la razón por la que se hace necesario la búsqueda de un mínimo global es debido a que en investigaciones previas se ha logrado confirmar que las estructuras nativas presentan equilibrio termodinámico y una energía libre mínima¹³⁴.

En el presente trabajo se realizó dos pasos de equilibrio previos a la dinámica molecular. El primero fue usando un ensamble NVT que permitió que la proteína, glicósidos, iones y el solvente encontraran las posiciones más adecuadas para interactuar entre sí a la temperatura requerida de 36.5 grados Celsius. Mientras que el segundo equilibrio se utilizó para incluir el parámetro de presión al sistema y así alcanzar una densidad adecuada y presión estable después de que la temperatura se había estabilizado.

Por último, se realizó la dinámica molecular de 100 ns considerando un ensamble NPT. Mediante la restricción de posición de los átomos pertenecientes a la región transmembrana, se trató de simular lo que ocurriría si la proteína estuviera embebida en una bicapa lipídica. Esta no se agregó debido a la cantidad de átomos que ya presentaba el sistema pues, de colocarla, el costo computacional sería mucho más elevado y requeriría de mayor tiempo máquina.

Adicionalmente, se realizó la evaluación de los valores de temperatura, presión y energía potencial, cinética y total del sistema para garantizar que estos se encontraran dentro de los rangos correctos. Para lograr esto, los valores de las propiedades mencionadas se guardaron cada 0.5 ps durante los 100 ns de la simulación. Como se visualiza en la Figura 3.19, las propiedades físicas, así como las energías se mantuvieron a valores constantes durante la simulación obteniéndose una temperatura promedio de 309.649 ± 0.00045 K y una presión de 1.11961 ± 0.0071 bar.

Por último, la dinámica molecular se evaluó mediante las gráficas de RMSD, RMSF, radio de giro y puentes de hidrógeno. La estructura final del IR se minimizó nuevamente para reportar la conformación con menor energía potencial y así obtener mejores resultados en el diagrama de Ramachandran. Esta se puede observar en la Figura 3.20

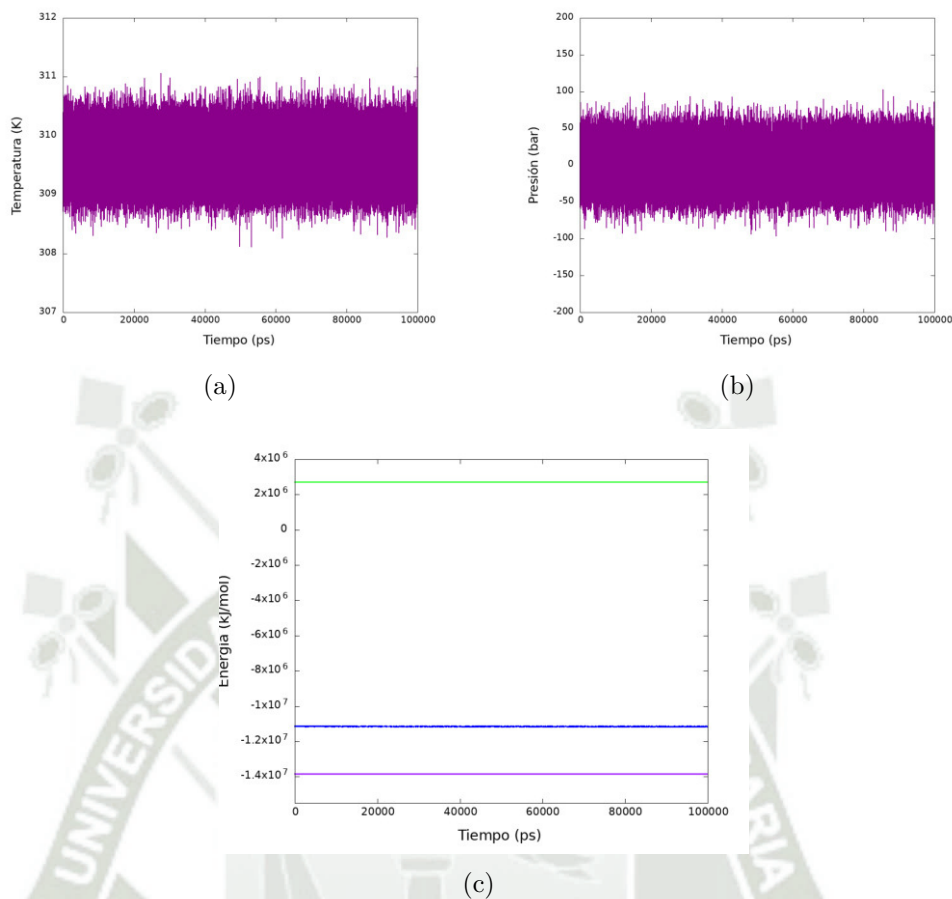


Figura 3.19: Parámetros registrados en la dinámica molecular. Fluctuación de la a) temperatura, b) presión y c) energía potencial (morado), cinética (verde), total (azul) durante la simulación.

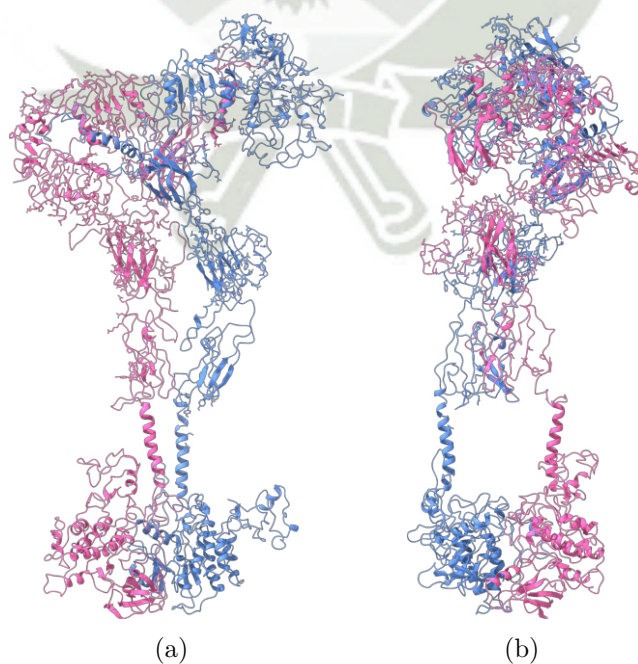


Figura 3.20: Estructura del Receptor de Insulina optimizada.

3.0.8. Análisis de la estructura optimizada del Receptor de Insulina

Raíz de la Desviación Media Cuadrática (RMSD)

El primer análisis que se realizó a la estructura fue la Raíz de la Desviación Media Cuadrática para poder determinar si la estructura final ya se encontraba en una posición estable a lo largo del tiempo. Para ello se utilizó el comando de Gromacs `gmx rms` y se comparó la estructura de la proteína respecto a sí misma, excluyendo al solvente, iones y glicósidos. Estos últimos se omitieron debido a que a lo largo de la simulación presentarán grandes fluctuaciones, principalmente por ser de pequeño tamaño, pudiendo influir notoriamente en la evaluación del RMSD. Además, mi objetivo principal era obtener la estructura del Receptor de Insulina optimizada, por lo que sus alrededores no requieren ser parte del análisis estructural.

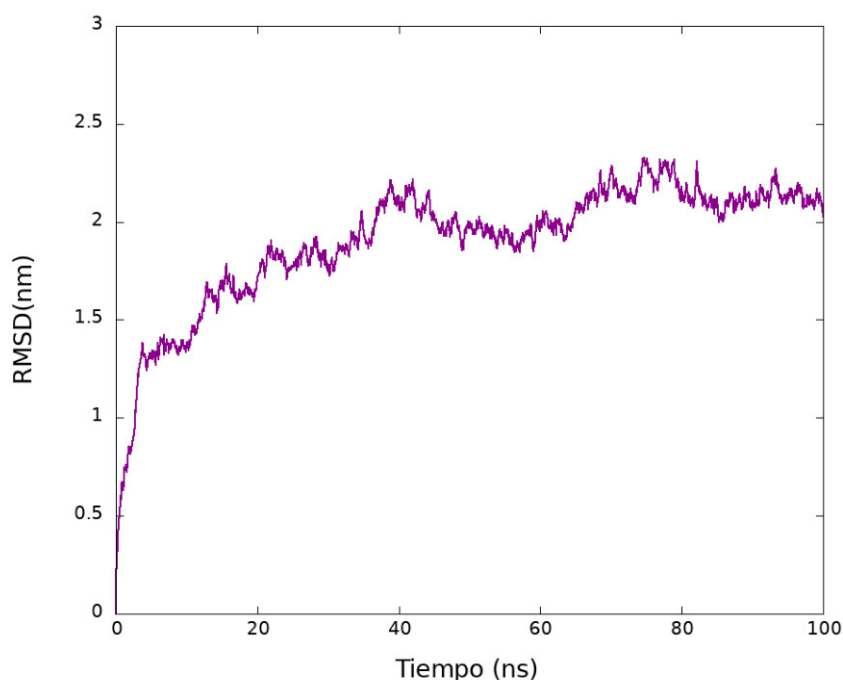


Figura 3.21: RMSD del Receptor de Insulina tras 100 ns.

Como se puede observar en la Figura 3.21, a partir de los 80 ns la gráfica comienza a adquirir una tendencia horizontal, indicativo de que la estructura ha dejado de sufrir cambios significativos a lo largo del tiempo. Sin embargo, 20 ns no son suficientes para afirmar que la estructura ya ha alcanzado una estabilidad estructural o un mínimo global. Para poder garantizarlo, sería necesario incrementar el tiempo de la dinámica

y volver a obtener el RMSD.

Raíz de la Fluctuación Media Cuadrática (RMSF)

La Raíz de la Fluctuación Media Cuadrática nos permite observar qué átomos o residuos presentan mayor movimiento durante la simulación. De esta forma se puede evaluar con mayor precisión aquellas zonas de la estructura que están causando energías muy altas o problemas durante la simulación.

Como podemos observar en la Figura 3.22 hay picos muy altos, cuya desviación de posición supera los 1.5 nm. Estos corresponden a los residuos SER2115, VAL2116, PRO2117, THR2118, SER2119, PRO2120, GLU2121, PRO2325, GLU2326 y TYR2327, que pertenecen a solo uno de los monómeros de la estructura. Dado que la gráfica dificulta la visualización de la posición exacta de estos residuos, y por ende, las zonas con mucho movimiento, se realizó un mapa de calor. Este facilita la ubicación de dichos residuos al crear una superficie coloreada sobre la estructura para diferenciar las zonas que presentan movimiento de las que no.

Para el presente proyecto se diferenció la superficie en tres zonas:

- Zona roja: residuos de fluctuación alta que superan los 1.5 nm.
- Zona anaranjada - amarilla: residuos de fluctuación intermedia que se encuentran entre 1 nm y 1.5 nm de fluctuación.
- Zona verde: residuos estables y de poca fluctuación.

Según la Figura 3.23 podemos ver que los residuos con alto movimiento pertenecen a regiones sin estructura secundaria o también llamadas "loops". Ambos grupos de residuos corresponden a predicciones realizadas con métodos ab-initio. Sin embargo, como se mencionó anteriormente, esto solo sucede con un lado del dímero, lo que llevaría a pensar que no es la predicción en sí la causante de la inestabilidad, dado que ambos monómeros son idénticos. En su lugar, esto podría asociarse al comportamiento de la proteína al interactuar con los alrededores pues en la Figura 3.20 ya vemos un ligero plegamiento hacia un lado de esta.

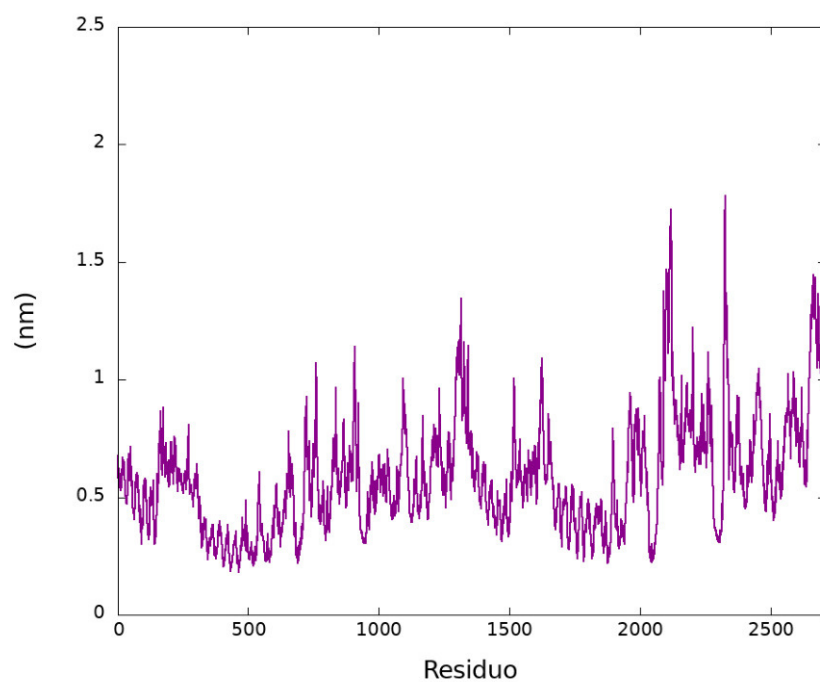


Figura 3.22: RMSF del Receptor de Insulina tras 100 ns.

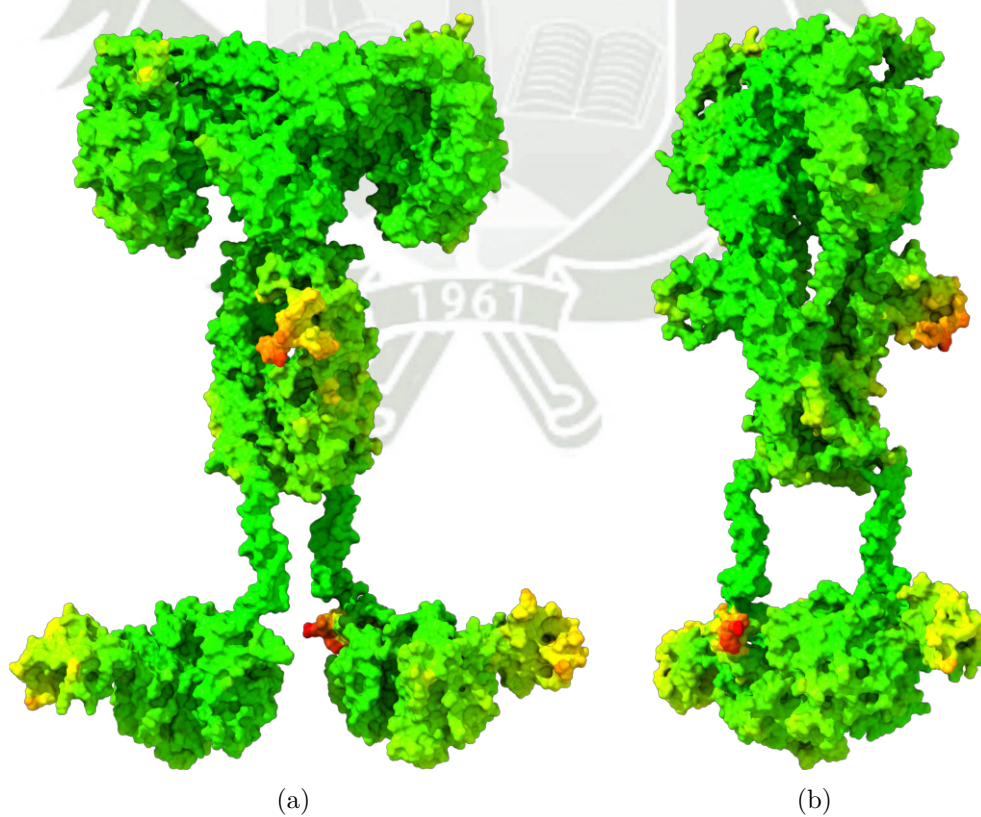


Figura 3.23: Análisis del factor b usando superficies. a) Proteína visualizada de frente.
b) Proteína visualizada de lado.

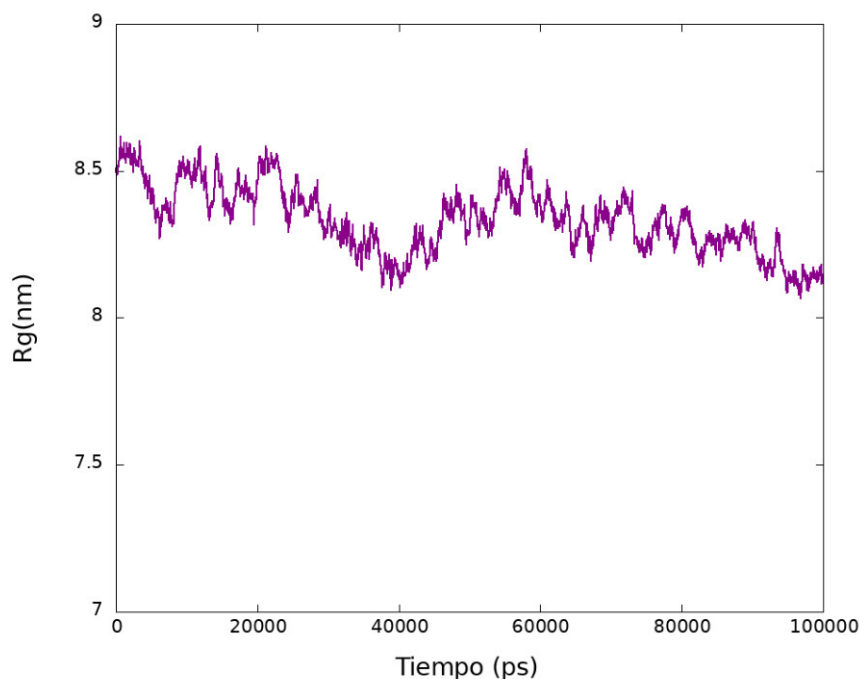


Figura 3.24: Radio de giro del Receptor de Insulina tras 100 ns.

Radio de giro

Este análisis permite analizar el grado de compactación de la proteína. Valores bajos indican una estructura compacta o plegada, mientras que valores muy altos podrían llegar a indicar una posible desnaturalización de la proteína. La gráfica obtenida en este trabajo (Figura 3.24) muestra que la estructura aún no se encuentra estable estructuralmente, lo que una vez más indica que se debe extender el tiempo de simulación. Asimismo, si se compara el primer valor del radio de giro con el último obtenido al final de la simulación se puede ver claramente que hay una compactación de la proteína. Lo cual ya se había resaltado anteriormente. Esto puede ser un indicativo de la formación de nuevos enlaces durante la simulación.

Enlaces de Hidrógeno

La cantidad de enlaces de hidrógenos a lo largo de la simulación puede ayudar a obtener una mejor comprensión de lo que sucede en la gráfica de radio de giro como sucede en esta evaluación. Al observar la Figura 3.25 se puede ver que el número de enlaces de hidrógeno continúa variando, pues aún la estructura no se encuentra estable, lo cual concuerda perfectamente con los demás análisis. Sin embargo, este no es el detalle más importante. La gráfica muestra un claro aumento del número de enlaces

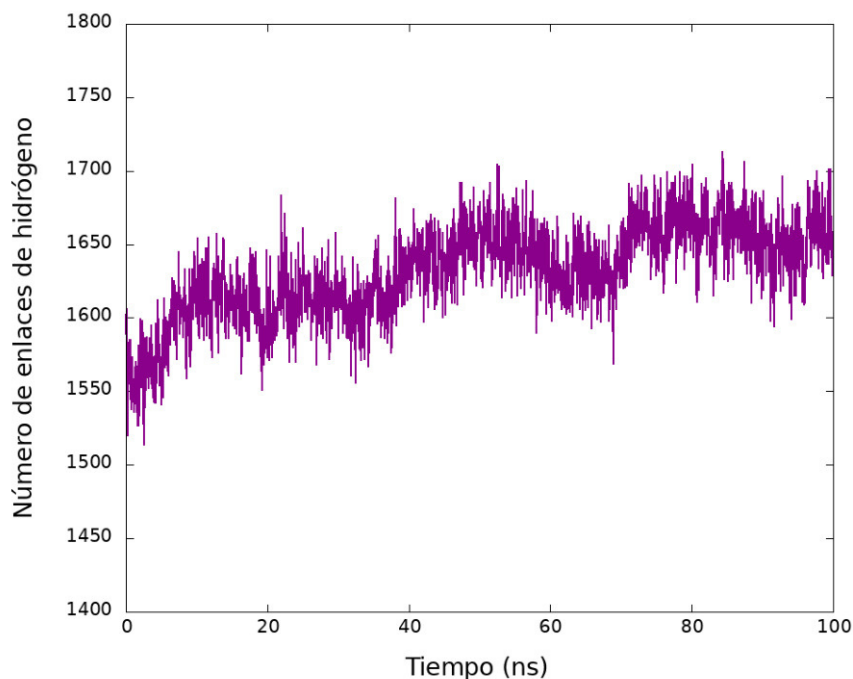


Figura 3.25: Enlaces de hidrógenos presentes durante la simulación del Receptor de Insulina.

en los últimos 40 ns de la simulación. Esto podría explicar por qué el radio de giro comienza a disminuir considerablemente desde los 60 ns aproximadamente. En otras palabras, la formación de nuevos enlaces contribuye al plegamiento de la proteína, lo cual influye notoriamente en la evaluación del radio de giro.

Diagrama de Ramachandran

Las gráficas de Ramachandran¹³⁵ son una herramienta útil para poder evaluar los residuos de la proteína que podrían estar involucrados en choques estéricos. A muy grandes rasgos, este diagrama, que evalúa los ángulos de torsión, tiene 3 grandes zonas que se consideran como regiones permitidas¹³⁶ y donde deberían ubicarse la mayor cantidad de residuos de la estructura. Las gráficas de Ramachandran se pueden clasificar en 4 tipos básicos según la esteoquímica del aminoácido, estos son: genérico donde no se evalúa a los aminoácidos glicina y prolina, glicina, prolina y pre-prolina¹³⁷. Adicionalmente, MolProbity¹³⁸ incluye en su análisis un Ramachandran de Isoleucina y Valina debido a sus características únicas al tener un $C\beta$ ramificado.

Visualizando la gráfica genérica se pudo observar que la mayor parte de los residuos se encontraban en las zonas permitidas. Sin embargo, 39 aminoácidos de los 2710 que componen la estructura del IR se encuentran fuera de estas regiones.

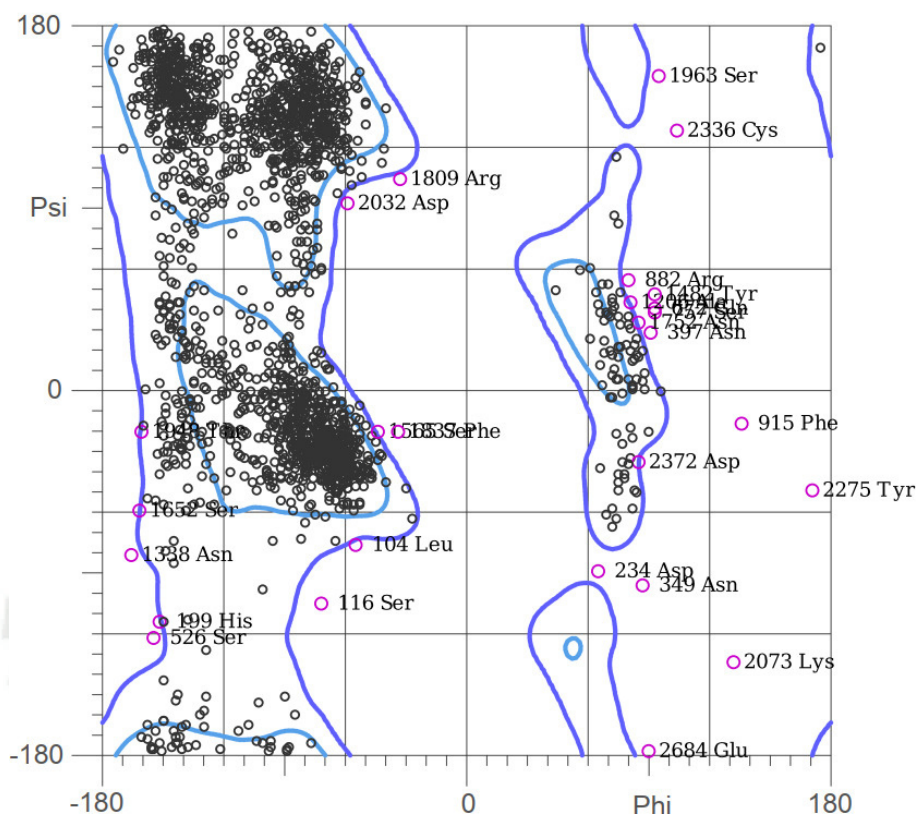


Figura 3.26: Gráfica de Ramachandran Genérica del Receptor de Insulina después de la dinámica de 100 ns.

En porcentajes, 86.9 % de los residuos estaban en las regiones favorecidas, mientras que 98.6 % se encontraban en las regiones permitidas. Como lo explican *Hollingsworth y Karplus*¹³⁶, la diferencia entre estas dos áreas hace referencia al tamaño que tendrán las esferas que representan los átomos, pues conceptualmente para realizar este tipo de análisis se considera que cada átomo es una esfera impenetrable. De forma que, las regiones favorecidas son aquellas donde no hay choques o colisiones al asignar radios estándares a las esferas que representan los átomos dentro de un péptido. Mientras que las permitidas o "de puente" son las regiones en las que no ocurren choques cuando se le asigna a las esferas radios más pequeños a los estándares. Estos radios pequeños representan los menores valores que podrían ser considerados posibles.

En la Figura 3.27 se pueden observar en color rojo los residuos que presentan problemas estéricos según la gráfica de Ramachandran. Estos se encuentran distribuidos a lo largo de toda la estructura, pero donde hay un mayor número de problemas estéricos es en el ectodominio. Por esta razón se profundizó en el análisis de esta región, la cual presenta muy pocos fragmentos reconstruidos.

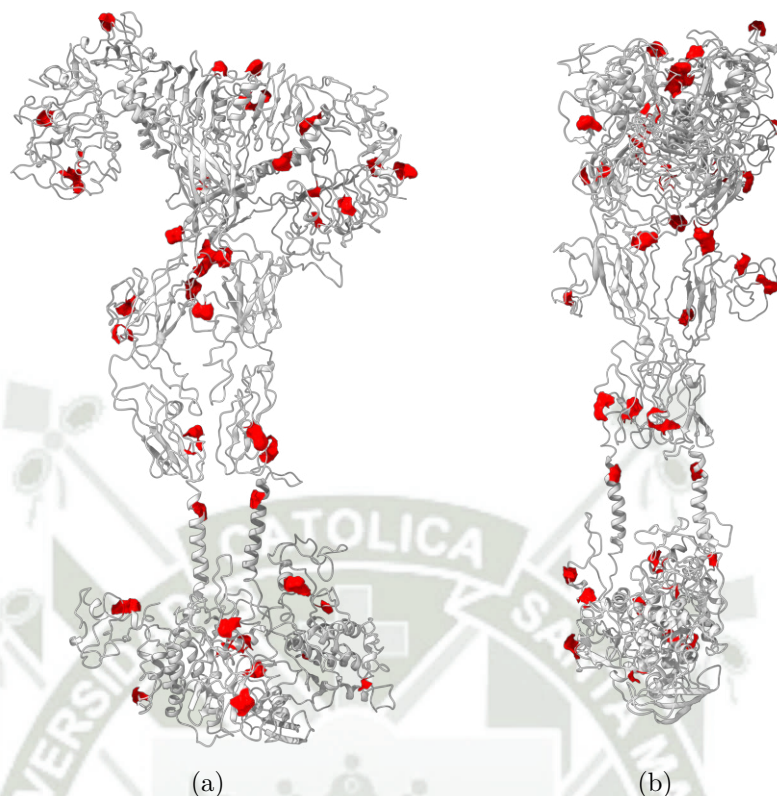


Figura 3.27: Estructura del IR resaltando en rojo los residuos ubicados en zonas no permitidas de la Gráfica de Ramachandran genérica. a) Visualización desde el frente b) Visualización desde uno de los lados.

El ectodominio descargado del Protein Data Bank fue el 6SOF, como ya se mencionó anteriormente. Esta estructura, antes de dejarse en su forma monomérica para la construcción del IR, presentaba además 4 insulinas localizadas cerca a sus sitios de unión con el ectodominio. Dichas hormonas pueden observarse de color morado en la Figura 3.28.a. Con la finalidad de poder observar si es que los residuos cercanos a las insulinas eran los que estaban teniendo problemas estéricos en la estructura construida, se resaltó de color rojo los residuos ubicados a menos de 5\AA de radio¹³⁹ de las hormonas. Esto se comparó con el ectodominio obtenido después de la dinámica, cuyos sitios con problemas estéricos, según la gráfica de Ramachandran, se encontraban resaltados también en rojo (Figura 3.28.b).

Observando la comparación se pudo comprobar que algunos de los residuos que están lo suficientemente cerca para interactuar con el ligando están en zonas no permitidas del Diagrama de Ramachandran después de la simulación, especialmente en la parte central del ectodominio. Esto llevó a pensar que los sitios de unión de la insulina podrían tener una influencia en los resultados obtenidos en la Gráfica de Ramachandran.

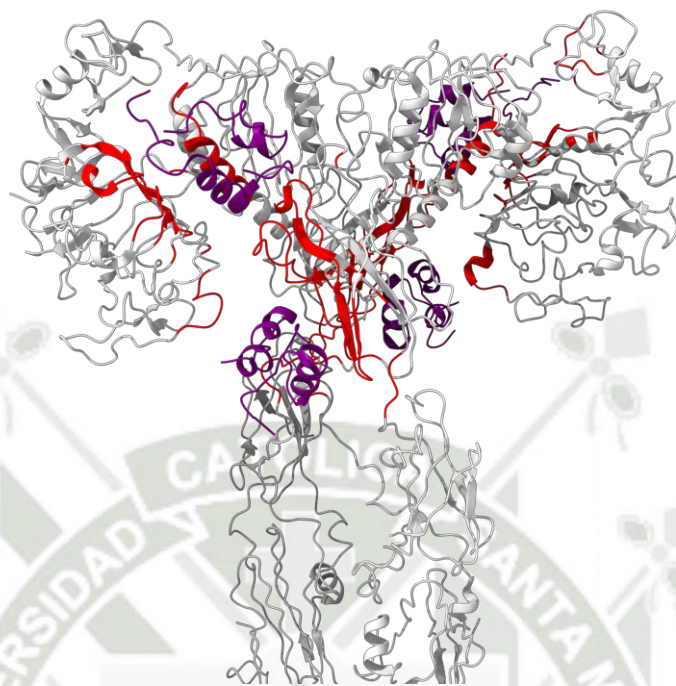
Una búsqueda de literatura permitió comprobar esta hipótesis, ya que un estudio de *Gunasekaran y Nussinov*¹⁴⁰ demostró que algunas veces los residuos de una proteína que están involucrados en determinadas funciones muestran conformaciones no permitidas en la gráfica de Ramachandran. De hecho, en su estudio se observó que había un mayor registro de residuos pertenecientes a sitios de unión en zonas no permitidas de la gráfica y que al añadir los ligandos a la simulación muchos de estos pasaban a tener una conformación permitida, sobretodo en aquellas proteínas que presentaban grandes cambios conformacionales.

La última observación mencionada es importante recalcarla dado que el Receptor de Insulina, como ya se ha mencionado previamente, tiene dos formas conformacionales: una activa cuando la insulina está unida a este y otra inactiva cuando no hay presencia de hormonas. La construcción del receptor se ha hecho usando un ectodominio "activo" pero en la simulación no se consideró la adición de las insulinas por la cantidad de átomos que ya tenía el sistema. Por lo tanto, esta falta de ligandos también podría estar retrasando la estabilidad de la estructura y el comportamiento de los residuos cercanos al sitio de unión.

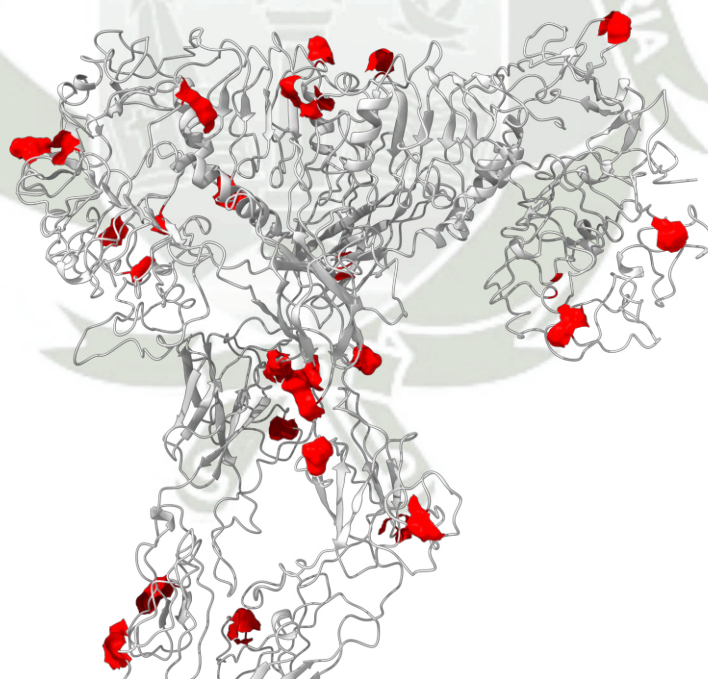
La gráfica de Ramachandran de la glicina difiere de la gráfica genérica en que abarca mayor área de zonas permitidas. Esto se debe a que la glicina no presenta una cadena secundaria, por lo tanto, no tiene $C\beta$ lo que reduce significativamente el número de choques estéricos dentro de la estructura¹³⁷.

Observando la Figura 3.29 se pudo notar que de todas las glicinas presentes en la estructura solo 4 se encontraban en zonas no permitidas. Sin embargo, estos residuos se encontraban en casi la misma posición, muy cerca a una zona permitida. Probablemente, esto se deba a que la estructura aún no ha encontrado una conformación estable en el mínimo de energía. Pues, tomando en consideración que esta gráfica representa la medición de los ángulos ϕ y ψ , se podría decir que dichos residuos se están comportando de forma similar en el sistema.

En el caso de los residuos Gly933 y Gly2288, es de esperar un comportamiento similar dado que ambos se encuentran en la misma posición dentro de la estructura solo que en heteromonómeros diferentes. Además, son parte de la región transmembrana, cuyas posiciones fueron restringidas al inicio de la dinámica. Razón por la cual, posiblemente, no han podido encontrar la mejor conformación. Los otros dos residuos no se corresponden entre sí como es el caso de los ya mencionados, sin embargo, ambos se



(a)



(b)

Figura 3.28: Comparación del ectodominio del IR usado como molde y el obtenido después de la simulación a) Ectodominio 6SOF, en morado se observan las insulinas en sus sitios de unión y en rojo las regiones que interactúan con la hormona a menos de 5Å, b)Ectodominio después de la simulación, en rojo se observan los residuos fuera de las zonas permitidas de la gráfica de Ramachandran.

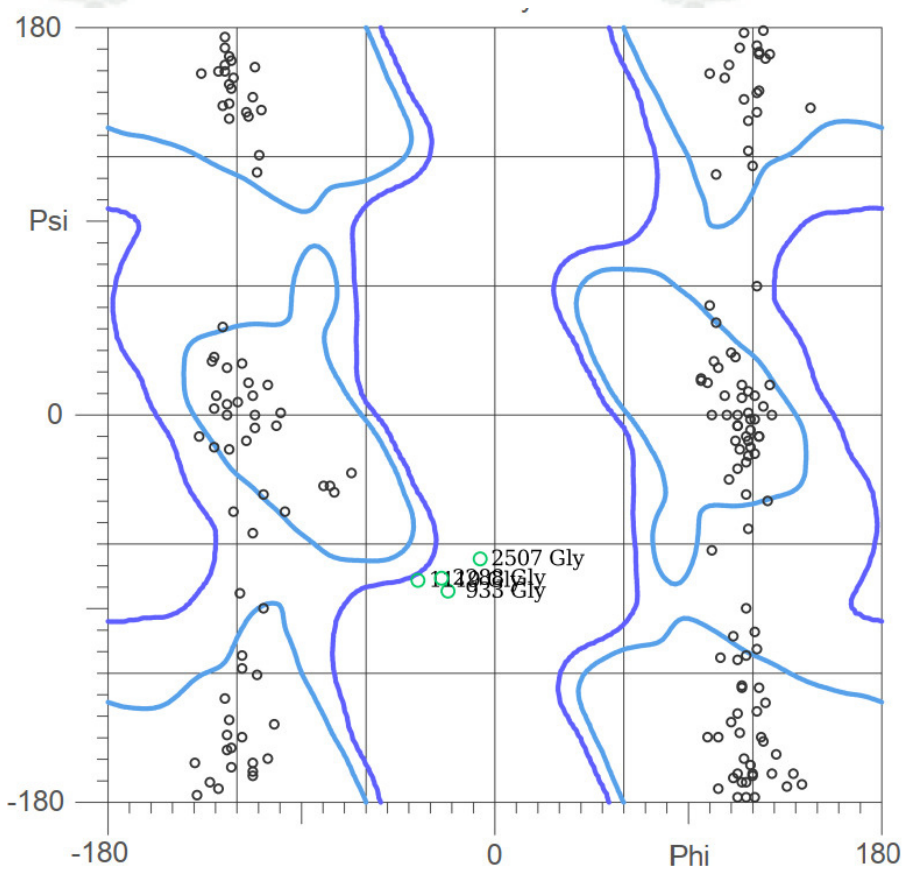


Figura 3.29: Gráfica de Ramachandran de Glicina del Receptor de Insulina después de la dinámica de 100 ns.

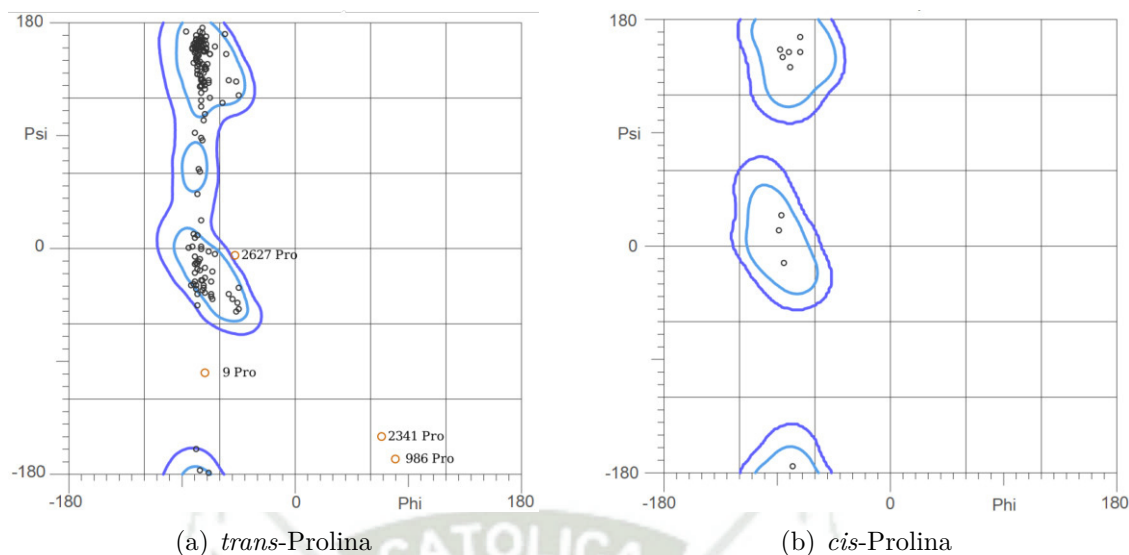


Figura 3.30: Gráficas de Ramachandran de Prolina del Receptor de Insulina después de la dinámica de 100 ns.

encuentran en la región intracelular. Uno de ellos, es parte de una alfa hélice y el otro pertenece a un *loop*.

La gráfica de Ramachandran de la prolina, a diferencia de la gráfica de la glicina, es mucho más restrictiva. Esto se debe a la presencia de un anillo, donde el $C\delta$ está unido covalentemente al nitrógeno que es parte del enlace peptídico, impidiendo la rotación libre alrededor de ϕ^{141} . Además, dependiendo del carbono de la prolina al que se una el siguiente aminoácido se formarán dos posibles conformaciones, *trans*-Prolina o *cis*-Prolina, siendo esta última la menos estable por lo que su abundancia suele ser menor en una estructura proteica¹⁴².

Observando la Figura 3.30 notamos la clara diferencia en abundancia de los residuos *trans*-Prolina respecto a *cis*-Prolina, lo cual concuerda con lo ya mencionado anteriormente. Asimismo, solo 4 residuos de *trans*-Prolina se encontraron en zonas no permitidas de la gráfica de Ramachandran. De estos, los residuos Pro986 y Pro2341, ocupaban la misma posición en la estructura pero en diferentes heteromonómeros, por ello el comportamiento de sus ángulos es similar en la gráfica. En este caso estos pertenecían a *loops* en la región intracelular. Los otros 2 residuos no guardaban relación de posición, uno de ellos se encontraba en el ectodominio y el otro a la región intracelular. Dado que la estructura aun no se encuentra en el mínimo global de energía es posible asumir que este es el por qué de los residuos en zonas no favorables.

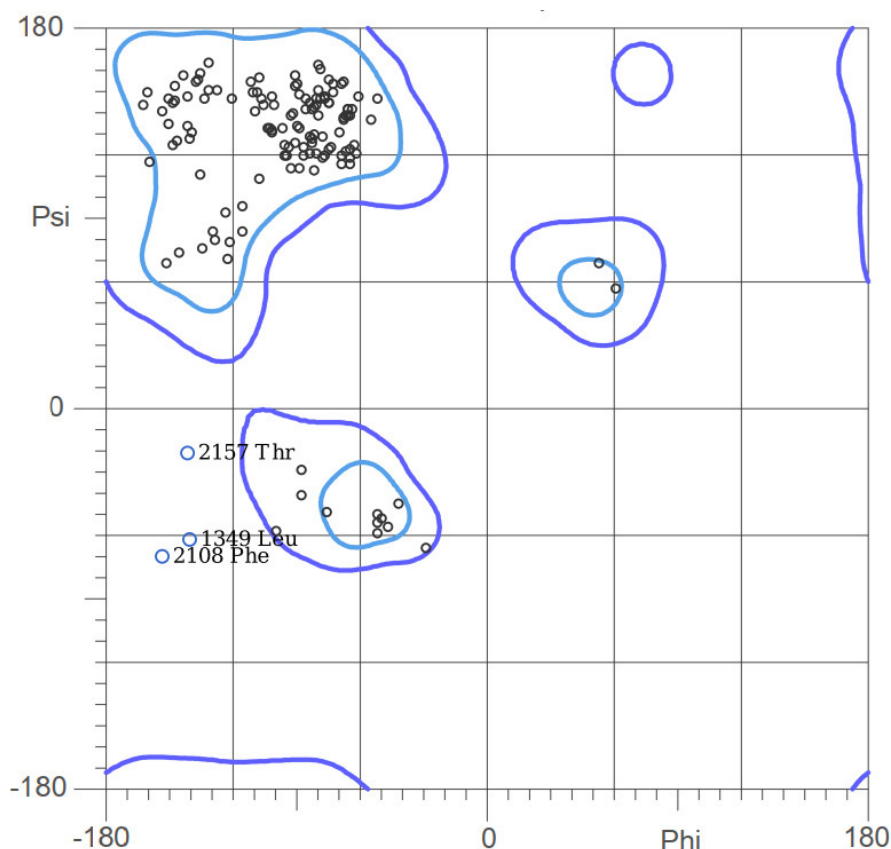


Figura 3.31: Gráfica de Ramachandran de pre-Prolina del Receptor de Insulina después de la dinámica de 100 ns.

La Gráfica de Ramachandran de pre-prolina también es muy restrictiva, abarcando una área menor de conformaciones permitidas. Esto se debe, principalmente, a la interacción del grupo CH_2 de la prolina ($C\gamma$) con el residuo que le antecede, ya que este grupo puede producir choques estéricos con la pre-prolina¹³⁷.

En la gráfica obtenida se pudo apreciar que tres residuos se encontraban fuera de las zonas permitidas. Sin embargo, dada su cercanía a una región aceptable, es posible que con un mayor tiempo de dinámica logren posicionarse en la zona permitida. No obstante, los tres residuos pertenecen a zonas sin estructura secundaria, conocidos como *loops*, los cuales se mueven constantemente aun cuando la proteína presenta una estructura estable¹⁴³, por lo que este factor también podría estar influyendo en los resultados obtenidos.

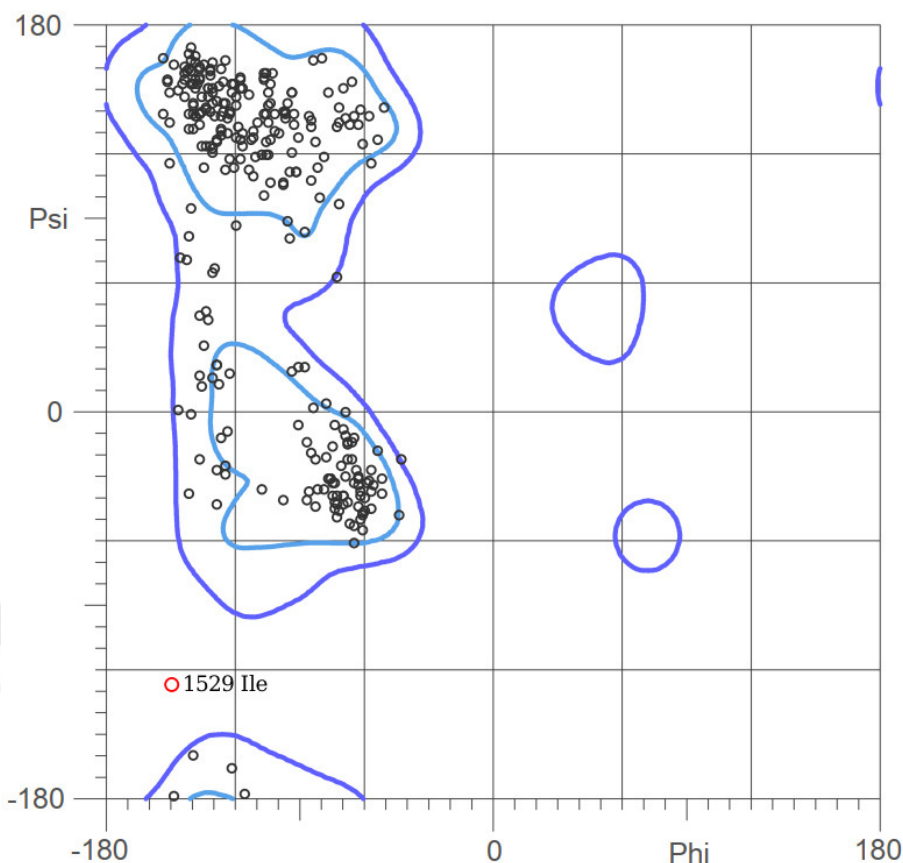


Figura 3.32: Gráfica de Ramachandran de Isoleucina y Valina del Receptor de Insulina después de la dinámica de 100 ns.

Por último, se tiene la gráfica de Ramachandran de Isoleucina y Valina, la cual presenta una mayor restricción respecto al ángulo Phi, el cual suele causar menos problemas estéricos cuando se encuentra entre los -180° y 0° , como se puede ver en la Figura 3.32. Esto se debe principalmente a que los aminoácidos mencionados presentan un $C\beta$ ramificado¹³⁸, el cual puede causar diversos choques estéricos con los átomos contiguos.

En la estructura construida se puede apreciar que solo un residuo de isoleucina no está en ninguna de las áreas permitidas, este corresponde también a un *loop* pero esta vez en el ectodominio. Nuevamente, la principal causa para esto podría ser que la estructura no se encuentra optimizada completamente.

Analizando todos los resultados obtenidos de la gráfica de Ramachandran podemos concluir que hay varios residuos pertenecientes a *loops* fuera de las zonas permitidas. Esto puede estar relacionado con la alta flexibilidad de los mismos que permiten colocar en las posiciones adecuadas a los residuos catalíticos¹⁴⁰ y ayudar con la estabilidad

de la estructura. Sin embargo, también se ha comprobado que la presencia de *loops* de pequeño tamaño entre 2 estructuras secundarias podría llevar a observar residuos no permitidos¹⁴⁴. Esto particularmente, se observa en los extremos del ectodominio e interiores del segmento intracelular.

Finalmente, una comparación de estas gráficas con el mapa de calor obtenido por los cálculos de RMSF, demostraron que la mayoría de los residuos localizados en regiones no permitidas tienen una fluctuación térmica pequeña. A excepción de una de las cadenas predichas y añadida al inicio de la subunidad β que presenta grandes fluctuaciones y los extremos del segmento intracelular que tienen residuos de fluctuación moderada. En estos casos sí se podría atribuir las configuraciones inusuales a la falta de estabilidad que presenta la proteína en dichas zonas.

Movimiento e influencia de los glicósidos

Se sabe que los glicósidos son de gran importancia para mejorar la estabilidad de la proteína, asegurar su bioactividad e incluso protegerla de la degradación por enzimas proteolíticas¹⁴⁵. Debido a su influencia en la estabilidad es que estos necesitan ser considerados como parte de las dinámicas moleculares que involucran proteínas glicosiladas. Este es el caso del receptor de insulina, proteína altamente glicosilada, cuyos glicósidos favorecen su correcto plegamiento y funcionamiento del mismo. Es así que, en el presente análisis se buscó evaluar si es que estos sufrieron cambios visibles que pudieran afectar a la estructura de la proteína.

En la Figura 3.33 se puede apreciar en color anaranjado los glicósidos unidos a la proteína, la cual se había curvado hacia un lado. La observación de la estructura demostró que la mayoría de los glicósidos se comportaron de forma similar durante la simulación, ya que sus posiciones en ambos lados del monómero son similares. Sin embargo, hubo dos cambios importantes que pudieron haber tenido alguna repercusión en la conformación adoptada por la estructura.

El primero de ellos es muy visible, como se puede ver en la parte central del ectodominio del monómero derecho (monómero 2) parece haber una gran acumulación de glicanos en un mismo sitio. Esto podría estar generando efectos de repulsión en los residuos cercanos haciendo que la estructura se curve hacia el lado opuesto. Este fenómeno se puede observar con mayor claridad en la Figura 3.34.

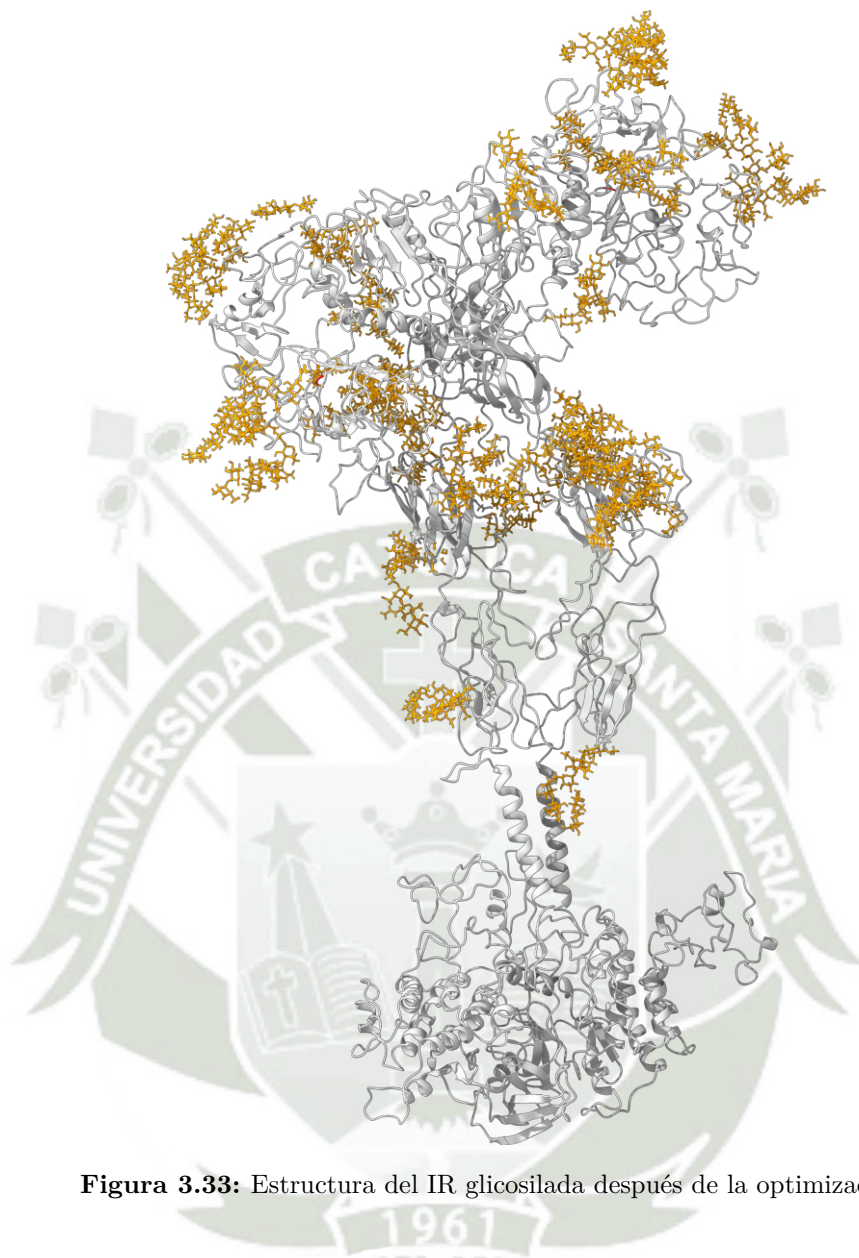


Figura 3.33: Estructura del IR glicosilada después de la optimización.

De forma que, las imágenes *a* y *b* muestran la conformación final del monómero 1 (representado siempre del lado izquierdo de la proteína). Mientras que las representaciones *c* y *d* corresponden al monómero 2. Para este análisis el último monómero mencionado se rotó hasta estar en la misma posición que el monómero 1 con la finalidad de facilitar la observación de las diferencias entre ambos.

Cuando se observa dichas estructuras desde los lados, se puede apreciar que los glicanos asociados a los residuos ASN606 y ASN624 (coloreados en rojo) exhiben comportamientos ligeramente diferentes. En el caso del monómero 1 es posible apreciar que el glicano asociado a la posición ASN624 (ubicado en el lado superior derecho de las Figuras 3.34(a) y 3.34(b)) se encuentra dirigido hacia arriba, lo cual no sucede en el

monómero 2 donde el glicano se encuentra dirigido hacia el exterior derecho de la estructura. De esta forma se combina con los glicanos cercanos que tienen una orientación similar formando la acumulación que se puede apreciar en la Figura 3.33.



Figura 3.34: Análisis de la influencia de los glicósidos en la estructura del IR.

Algo similar ocurre con el glicano de la posición ASN606 (ubicado en el lado inferior izquierdo de las Figuras 3.34(a) y 3.34(b)), en el monómero 1 este se encuentra hacia afuera de la estructura mientras que en el monómero 2 se encuentra hacia el interior de la misma contribuyendo a la acumulación antes mencionada. Como se puede ver, el monómero 2 podría verse influenciado por estos dos glicanos que forman acumulaciones en la estructura pudiendo ser la causa de que la misma se doble hacia el lado contrario.

El análisis de la estructura completa del receptor de insulina construido sugiere que la estructura requiere de un mayor tiempo de dinámica para poder determinar si las regiones de mucho movimiento corresponden a regiones activas de la proteína o a la falta de estabilidad de la misma. No obstante, ha sido posible obtener una estructura completa del receptor de insulina parcialmente optimizado en formato .pdb, el cual puede ser utilizado para futuras dinámicas moleculares.



Conclusiones

- PRIMERA. Se logró identificar la secuencia completa de 1382 aa en la base de datos NCBI (National Center of Biotechnology Information) del Receptor de Insulina Humana Isoforma B, así como la secuencia de 1370 aa de la Isoforma A en la base de datos Uniprot. De ambas secuencias, se decidió trabajar con la isoforma B por su prevalencia en problemas metabólicos. Dicha secuencia de aminoácidos se descargó con extensión *.fasta.
- SEGUNDA. El segmento transmembrana y juxtamembrana comprendido por aproximadamente 72 aminoácidos fue construido mediante la unión en el servidor de AIDA de la estructura con código PDB 2MFR y un fragmento de 23 aminoácidos obtenido por predicción *de novo* usando el servidor PEP-FOLD3. Este fragmento fue optimizado en el interior de una bicapa lipídica construida en la plataforma CHARMM-GUI.
- TERCERA. La estructura tridimensional completa del IR se obtuvo mediante la unión de los fragmentos completos construidos del ectodominio, segmento transmembrana y segmento intracelular. Para ello se usó los programas de AIDA y UCSF Chimera para la construcción del monómero y Pymol para la construcción del dímero. Para este último paso se requirió de la alineación de los monómeros a un *molde* del ectodominio (código PDB 6SOF) que permitió la correcta ubicación de los mismos.
- CUARTA. El receptor de insulina fue equilibrado y posteriormente optimizado usando GROMACS, encontrándose glicosilado en un sistema con solvente (TIP4P) dentro de una caja triclinica y neutralizado con iones Na^+ . La dinámica molecular se realizó por 100 ns considerando un ensamble NPT a una temperatura de 36.5 °C y 1 bar de presión.
- QUINTA. La estructura resultante de la dinámica molecular se evaluó mediante las gráficas de RMSD, RMSF, radio de giro, puentes de hidrógeno y gráfica de

Ramachandran. Concluyéndose que la estructura requiere un mayor tiempo de dinámica molecular para confirmar que la estructura se encuentre realmente estable y en su estado nativo o si el gran tamaño que presenta influye en los resultados observados. Esto principalmente debido a que se observaron regiones con mucha fluctuación y residuos en regiones no permitidas de la gráfica de Ramachandran, a pesar de que las gráficas de RMSD, radio de giro y enlaces de hidrógeno ya comenzaban a estabilizarse. No obstante, la estructura obtenida en formato .pdb puede ser utilizada para futuras dinámicas.



Recomendaciones

Sería adecuado realizar una dinámica molecular siguiendo los mismos parámetros utilizados en el presente trabajo de investigación pero con un mayor tiempo de dinámica de aproximadamente 500ns.

Asimismo, considero que se obtendrían mejores resultados al trabajar con una bicapa lipídica en la sección del segmento transmembrana en lugar de restringir la sección. Esto con la finalidad de observar un comportamiento más real de la proteína en el tiempo.

Por último, se recomienda considerar el pH del sistema para la adición de los hidrógenos faltantes en la estructura y así poder evaluar si estos tienen influencia directa en el plegamiento de la proteína durante la dinámica. Adicional a ello, también puede ser relevante considerar la adición de los O-glicanos en la estructura ya que estos no han sido considerados en la presente investigación.

Referencias Bibliográficas

- [1] Trenker R, Jura N. Receptor tyrosine kinase activation: From the ligand perspective. *Current Opinion in Cell Biology*. 2020;63:174–185.
- [2] Gutmann T, Kim KH, Grzybek M, Walz T, Coskun U. Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *Journal of Cell Biology*. 2018;217(5):1643–1649.
- [3] Diwanji D, Thaker T, Jura N. More than the sum of the parts: Toward full-length receptor tyrosine kinase structures. *IUBMB life*. 2019;71(6):706–720.
- [4] Scapin G, Dandey VP, Zhang Z, Prosis W, Hruza A, Kelly T, et al. Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis. *Nature*. 2018;556(7699):122–125.
- [5] Kasuga K, Kaneko H, Nishizawa M, Onodera O, Ikeuchi T. Generation of intracellular domain of insulin receptor tyrosine kinase by γ -secretase. *Biochemical and biophysical research communications*. 2007;360(1):90–96.
- [6] Seino S, Seino M, Bell GI. Human insulin-receptor gene: partial sequence and amplification of exons by polymerase chain reaction. *Diabetes*. 1990;39(1):123–128.
- [7] Sparrow LG, McKern NM, Gorman JJ, Strike PM, Robinson CP, Bentley JD, et al. The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. *Journal of Biological Chemistry*. 1997;272(47):29460–29467.
- [8] Sparrow LG, Lawrence MC, Gorman JJ, Strike PM, Robinson CP, McKern NM, et al. N-linked glycans of the human insulin receptor and their distribution over the crystal structure. *Proteins: Structure, Function, and Bioinformatics*. 2008;71(1):426–439.

- [9] Denley A, Wallace J, Cosgrove L, Forbes B. The insulin receptor isoform exon 11- (IR-A) in cancer and other diseases: a review. *Hormone and Metabolic Research*. 2003;35(11/12):778–785.
- [10] De Meyts P, Sajid W, Palsgaard J, Theede AM, Gauguin L, Aladdin H, et al. Insulin and IGF-I receptor structure and binding mechanism. In: *Mechanisms of insulin action*. Springer; 2007. p. 1–32.
- [11] Vigneri R, Goldfine I, Frittitta L. Insulin, insulin receptors, and cancer. *Journal of endocrinological investigation*. 2016;39(12):1365–1376.
- [12] Cignarelli A, Genchi VA, Perrini S, Natalicchio A, Laviola L, Giorgino F. Insulin and insulin receptors in adipose tissue development. *International journal of molecular sciences*. 2019;20(3):759.
- [13] Vella V, Nicolosi ML, Giuliano M, Morrione A, Malaguarnera R, Belfiore A. Insulin receptor isoform A modulates metabolic reprogramming of breast cancer cells in response to IGF2 and insulin stimulation. *Cells*. 2019;8(9):1017.
- [14] Pomytkin I, Costa-Nunes JP, Kasatkin V, Veniaminova E, Demchenko A, Lyundup A, et al. Insulin receptor in the brain: Mechanisms of activation and the role in the CNS pathology and treatment. *CNS neuroscience & therapeutics*. 2018;24(9):763–774.
- [15] Pimentel E. *Hormones Growth Factors & Oncogenes*. CRC press; 1987.
- [16] Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, et al. The DNA sequence and biology of human chromosome 19. *Nature*. 2004;428(6982):529.
- [17] Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, et al. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome research*. 2004;14(10B):2121–2127.
- [18] Hosoe J, Kadowaki H, Miya F, Aizu K, Kawamura T, Miyata I, et al. Structural basis and genotype-phenotype correlations of INSR mutations causing severe insulin resistance. *Diabetes*. 2017;66(10):2713–2723.
- [19] Højlund K, Hansen T, Lajer M, Henriksen JE, Levin K, Lindholm J, et al. A novel syndrome of autosomal-dominant hyperinsulinemic hypoglycemia linked to a mutation in the human insulin receptor gene. *Diabetes*. 2004;53(6):1592–1598.

- [20] Fernández PL, et al. Velázquez, farmacología básica y clínica. Panamericana; 2015.
- [21] Berg JM, Stryer L, Tymoczko JL. Bioquímica. Reverté; 2007.
- [22] Anaya COM, Ariza IDS. Acción insulínica y resistencia a la insulina: aspectos moleculares. Revista de la Facultad de Medicina. 2005;53(4):235–243.
- [23] Belfiore A, Malaguarnera R, Vella V, Lawrence MC, Sciacca L, Frasca F, et al. Insulin receptor isoforms in physiology and disease: an updated view. Endocrine reviews. 2017;38(5):379–431.
- [24] McKern NM, Lawrence MC, Streltsov VA, Lou MZ, Adams TE, Lovrecz GO, et al. Structure of the insulin receptor ectodomain reveals a folded-over conformation. Nature. 2006;443(7108):218–221.
- [25] Macaulay SL, Polites M, Hewish DR, Ward CW. Cysteine-524 is not the only residue involved in the formation of disulphide-bonded dimers of the insulin receptor. Biochemical Journal. 1994;303(2):575–581.
- [26] Croll TI, Smith BJ, Margetts MB, Whittaker J, Weiss MA, Ward CW, et al. Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. Structure. 2016;24(3):469–476.
- [27] Zardecki C, Dutta S, Goodsell DS, Lowe R, Voigt M, Burley SK. PDB-101: Educational resources supporting molecular explorations through biology and medicine. Protein Science. 2022;31(1):129–140.
- [28] Kuznetsov AS, Zamaletdinov MF, Bershatsky YV, Urban AS, Bocharova OV, Bennisroune A, et al. Dimeric states of transmembrane domains of insulin and IGF-1R receptors: Structures and possible role in activation. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2020;p. 183417.
- [29] Li Q, Wong YL, Kang C. Solution structure of the transmembrane domain of the insulin receptor in detergent micelles. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2014;1838(5):1313–1321.
- [30] Hubbard SR. Structure and mechanism of the insulin receptor tyrosine kinase. In: Handbook of Cell Signaling. Elsevier; 2010. p. 307–313.
- [31] Cabail MZ, Li S, Lemmon E, Bowen ME, Hubbard SR, Miller WT. The insulin and IGF1 receptor kinase domains are functional dimers in the activated state. Nature communications. 2015;6(1):1–8.

- [32] Ocampo R, Ríos LA, Betancur LA, Ocampo DM. Curso práctico de química orgánica. Enfocado a biología y alimentos. Universidad de Caldas; 2008.
- [33] Wiedemann C, Kumar A, Lang A, Ohlenschläger O. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Frontiers in Chemistry*. 2020;8:280.
- [34] Pijning AE, Hogg P. Classification of Protein Disulphide Bonds. In: *Functional Disulphide Bonds*. Springer; 2019. p. 1–8.
- [35] Lu K, Guidotti G. Identification of the cysteine residues involved in the class I disulfide bonds of the human insulin receptor: properties of insulin receptor monomers. *Molecular biology of the cell*. 1996;7(5):679–691.
- [36] Raju TS. *Co-and Post-translational Modifications of Therapeutic Antibodies and Proteins*. John Wiley & Sons; 2019.
- [37] Arias EB. *Biomoléculas*. vol. 51. Universidad de Salamanca; 2013.
- [38] Besançon C, Guillot A, Blaise S, Dauchez M, Belloy N, PrévotEAU-Jonquet J, et al. New visualization of dynamical flexibility of N-Glycans: Umbrella Visualization in UnityMol. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2018. p. 291–298.
- [39] Sparrow LG, Gorman JJ, Strike PM, Robinson CP, McKern NM, Epa VC, et al. The location and characterisation of the O-linked glycans of the human insulin receptor. *Proteins: Structure, Function, and Bioinformatics*. 2007;66(2):261–265.
- [40] Varki A. Biological roles of glycans. *Glycobiology*. 2017;27(1):3–49.
- [41] de Freitas-Junior JCM, Andrade-da Costa J, Silva MC, Pinho SS. Glycans as regulatory elements of the insulin/IGF system: impact in cancer progression. *International journal of molecular sciences*. 2017;18(9):1921.
- [42] Lahoz-Beltrá R. *Bioinformática: Simulación, vida artificial e inteligencia artificial*. Ediciones Díaz de Santos; 2004.
- [43] Cañedo Andalia R, Arencibia Jorge R. *Bioinformática: en busca de los secretos moleculares de la vida*. Acimed. 2004;12(6):1–1.
- [44] Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. 2018;99(6):1129–1143.

- [45] Karplus M, Kuriyan J. Molecular dynamics and protein function. Proceedings of the National Academy of Sciences. 2005;102(19):6679–6685.
- [46] Wand AJ. Dynamic activation of protein function: a view emerging from NMR spectroscopy. Nature structural biology. 2001;8(11):926–931.
- [47] Delaforge E, Kragelj J, Tengo L, Palencia A, Milles S, Bouvignies G, et al. Deciphering the dynamic interaction profile of an intrinsically disordered protein by NMR exchange spectroscopy. Journal of the American Chemical Society. 2018;140(3):1148–1158.
- [48] Rogawski R, McDermott AE. New NMR tools for protein structure and function: Spin tags for dynamic nuclear polarization solid state NMR. Archives of biochemistry and biophysics. 2017;628:102–113.
- [49] Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature structural biology. 2002;9(9):646–652.
- [50] MacKerell Jr AD. Empirical force fields for biological macromolecules: overview and issues. Journal of computational chemistry. 2004;25(13):1584–1604.
- [51] Galindo-Murillo R, Robertson JC, Zgarbova M, Sponer J, Otyepka M, Jurečka P, et al. Assessing the current state of amber force field modifications for DNA. Journal of chemical theory and computation. 2016;12(8):4114–4127.
- [52] Wang LP, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, et al. Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. The Journal of Physical Chemistry B. 2017;121(16):4023–4039.
- [53] Lee J, Hitzengerger M, Rieger M, Kern NR, Zacharias M, Im W. CHARMM-GUI supports the Amber force fields. The Journal of Chemical Physics. 2020;153(3):035103.
- [54] Kim S, Lee J, Jo S, Brooks III CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. Journal of computational chemistry. 2017;38(21):1879–1886.
- [55] Sonibare K, Rathnayaka L, Zhang L. Comparison of CHARMM and OPLS-aa forcefield predictions for components in one model asphalt mixture. Construction and Building Materials. 2020;236:117577.

- [56] Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved treatment of nucleosides and nucleotides in the OPLS-AA force field. *Chemical physics letters*. 2017;683:276–280.
- [57] Liang L, Chen EY, Shen JW, Wang Q. Molecular modelling of translocation of biomolecules in carbon nanotubes: method, mechanism and application. *Molecular Simulation*. 2016;42(10):827–835.
- [58] Bottaro S, Bussi G, Kennedy SD, Turner DH, Lindorff-Larsen K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science Advances*. 2018;4(5).
- [59] Hill TL. *An introduction to statistical thermodynamics*. Courier Corporation; 1986.
- [60] Hill TL. *Statistical Mechanics: Principles and Selected Applications*. Dover Publications, Inc.; 1987.
- [61] Tobias DJ, Martyna GJ, Klein ML. Molecular dynamics simulations of a protein in the canonical ensemble. *The Journal of Physical Chemistry*. 1993;97(49):12959–12966.
- [62] Tuckerman M. *Statistical mechanics: theory and molecular simulation*. Oxford university press; 2010.
- [63] Gao X, Fang J, Wang H. Sampling the isothermal-isobaric ensemble by Langevin dynamics. *The Journal of chemical physics*. 2016;144(12):124113.
- [64] Karatzas I, Shreve SE. Brownian motion. In: *Brownian Motion and Stochastic Calculus*. Springer; 2014. p. 47.
- [65] Cheng YC, Qian H, Zhu Y. Asymptotic Behavior of a Sequence of Conditional Probability Distributions and the Canonical Ensemble. *arXiv preprint arXiv:191211137*. 2019;.
- [66] Baldovin M. Physical interpretation of the canonical ensemble for long-range interacting systems in the absence of ensemble equivalence. *Physical Review E*. 2018;98(1):012121.
- [67] Yamauchi M, Mori Y, Okumura H. Molecular simulations by generalized-ensemble algorithms in isothermal-isobaric ensemble. *Biophysical reviews*. 2019;p. 1–13.

- [68] Paci E. High pressure simulations of biomolecules. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. 2002;1595(1-2):185–200.
- [69] Martín JMA. Láminas delgadas y recubrimientos: preparación, propiedades y aplicaciones. vol. 11. Editorial CSIC-CSIC Press; 2003.
- [70] Bort JA. Química teórica y computacional. vol. 2. Publicacions de la Universitat Jaume I; 2001.
- [71] Bockris JO, Reddy AK. *Electroquímica moderna*. II. Reverté; 1979.
- [72] Kitamura T. Chapter 10 - The Liquid-Glass Transition in Multi-Component Liquids. In: Kitamura T, editor. *Liquid Glass Transition*. Oxford: Elsevier; 2013. p. 291–331. Available from: <http://www.sciencedirect.com/science/article/pii/B9780124071773000106>.
- [73] Dobson JF, Gould T. Calculation of dispersion energies. *Journal of Physics: Condensed Matter*. 2012;24(7):073201.
- [74] Maza J, Mosqueira J, Veira JA. Física del estado sólido: ejercicios resueltos. vol. 11. Univ Santiago de Compostela; 2013.
- [75] Atkins P, de Paula J. *Physical chemistry*. WH Freeman and Company New York; 2006.
- [76] Gao Z. Research on Numerical Integration Algorithm in Molecular Dynamics Simulation. In: *Proceedings of the 2017 2nd International Conference on Machinery, Electronics and Control Simulation (MECS 2017)*. Atlantis Press; 2016. Available from: <https://doi.org/10.2991/mecs-17.2017.42>.
- [77] Luo J, Liu L. Optimisation of data locality in energy calculations for large-scale molecular dynamics simulations. *Molecular Simulation*. 2017;43(4):284–290.
- [78] Gonnet P. Pseudo-Verlet lists: a new, compact neighbour list representation. *Molecular Simulation*. 2013;39(9):721–727.
- [79] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic acids research*. 2000;28(1):235–242.
- [80] Voet D, Voet JG, Pratt CW. *Fundamentos de Bioquímica-: A Vida em Nível Molecular*. Artmed Editora; 2014.

- [81] Dominguez C, Bonvin AM, Winkler GS, van Schaik FM, Timmers HTM, Boelens R. Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure*. 2004;12(4):633–644.
- [82] O'Connor CM, Adams JU, Fairman J. *Essentials of cell biology*. Cambridge, MA: NPG Education. 2010;1:54.
- [83] Voet D, Voet JG. *Bioquímica*. Ed. Médica Panamericana; 2006.
- [84] Ginalski K. Comparative modeling for protein structure prediction. *Current opinion in structural biology*. 2006;16(2):172–177.
- [85] Lee J, Freddolino PL, Zhang Y. Ab initio protein structure prediction. In: *From protein structure to function with bioinformatics*. Springer; 2017. p. 3–35.
- [86] Ramachandran Gt, Sasisekharan V. Conformation of polypeptides and proteins. *Advances in protein chemistry*. 1968;23:283–437.
- [87] LASKOWSKI RA, FURNHAM N, THORNTON JM. The Ramachandran plot and protein structure validation. In: *Biomolecular forms and functions: a celebration of 50 years of the ramachandran map*. World Scientific; 2013. p. 62–75.
- [88] Sargsyan K, Grauffel C, Lim C. How molecular size impacts RMSD applications in molecular dynamics simulations. *Journal of chemical theory and computation*. 2017;13(4):1518–1524.
- [89] Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein science*. 2001;10(7):1470–1473.
- [90] Islam R, Parves MR, Paul AS, Uddin N, Rahman MS, Mamun AA, et al. A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*. 2021;39(9):3213–3224.
- [91] Sun Z, Liu Q, Qu G, Feng Y, Reetz MT. Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chemical reviews*. 2019;119(3):1626–1665.
- [92] Fuglebakk E, Echave J, Reuter N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics*. 2012 07;28(19):2431–2440. Available from: <https://doi.org/10.1093/bioinformatics/bts445>.

- [93] Vardhan S, Sahoo SK. In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19. *Computers in biology and medicine*. 2020;124:103936.
- [94] He L, Niemeyer B. A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration. *Biotechnology progress*. 2003;19(2):544–548.
- [95] Tyn MT, Gusek TW. Prediction of diffusion coefficients of proteins. *Biotechnology and bioengineering*. 1990;35(4):327–338.
- [96] Yamamoto E, Akimoto T, Mitsutake A, Metzler R. Universal relation between instantaneous diffusivity and radius of gyration of proteins in aqueous solution. *Physical review letters*. 2021;126(12):128101.
- [97] Taylor B, Sztain-Pedone T. Hydrogen Bond Analysis Tutorial [Online]; 2021. [Accessed 19th August 2022]. /urlhttps://ctlee.github.io/BioChemCoRe-2018/h-bond/#: :text=Hydrogen %20bond %20analysis %20identifies %20the,) %20(see %20Figure %20
- [98] Newberry RW, Raines RT. Secondary forces in protein folding. *ACS chemical biology*. 2019;14(8):1677–1686.
- [99] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1:19–25.
- [100] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004;25(13):1605–1612.
- [101] Humphrey W, Dalke A, Schulten K, et al. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996;14(1):33–38.
- [102] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*. 2018;35(6):1547–1549.
- [103] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8; 2015.
- [104] Danne R, Poojari C, Martinez-Seara H, Rissanen S, Lolicato F, Róg T, et al. doGlycans—tools for preparing carbohydrate structures for atomistic simulations of glycoproteins, glycolipids, and carbohydrate polymers for GROMACS. *Journal of chemical information and modeling*. 2017;57(10):2401–2406.

- [105] Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Research*. 2018;46(D 1):D8–D13.
- [106] Consortium U. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*. 2019;47(D1):D506–D515.
- [107] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research*. 2003;31(13):3381–3385.
- [108] Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic acids research*. 2016;44(W1):W449–W454.
- [109] Xu D, Jaroszewski L, Li Z, Godzik A. AIDA: ab initio domain assembly server. *Nucleic acids research*. 2014;42(W1):W308–W313.
- [110] Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic acids research*. 2016;44(W1):W406–W409.
- [111] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*. 2008;29(11):1859–1865.
- [112] Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of computational chemistry*. 2014;35(27):1997–2004.
- [113] Xu D, Zhang Y. Toward optimal fragment generations for ab initio protein structure assembly. *Proteins: Structure, Function, and Bioinformatics*. 2013;81(2):229–239.
- [114] Bohne-Lang A, von der Lieth CW. GlyProt: in silico glycosylation of proteins. *Nucleic acids research*. 2005;33(suppl_2):W214–W219.
- [115] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*. 2010;66(1):12–21.
- [116] Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. *Proteins: Structure, Function, and Bioinformatics*. 2007;69(2):394–408.

- [117] Maupetit J, Derreumaux P, Tufféry P. A fast method for large-scale De Novo peptide and miniprotein structure prediction. *Journal of computational chemistry*. 2010;31(4):726–738.
- [118] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983;79(2):926–935.
- [119] Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*. 1993;98(12):10089–10092.
- [120] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*. 2019;47(W1):W636–W641.
- [121] Gutmann T, Schäfer IB, Poojari C, Brankatschk B, Vattulainen I, Strauss M, et al. Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *Journal of Cell Biology*. 2020;219(1).
- [122] Uchikawa E, Choi E, Shang G, Yu H, Bai Xc. Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *Elife*. 2019;8:e48630.
- [123] Menting JG, Whittaker J, Margetts MB, Whittaker LJ, Kong GKW, Smith BJ, et al. How insulin engages its primary binding site on the insulin receptor. *Nature*. 2013;493(7431):241–245.
- [124] Chamberlain SD, Wilson JW, Deanda F, Patnaik S, Redman AM, Yang B, et al. Discovery of 4, 6-bis-anilino-1H-pyrrolo [2, 3-d] pyrimidines: Potent inhibitors of the IGF-1R receptor tyrosine kinase. *Bioorganic & medicinal chemistry letters*. 2009;19(2):469–473.
- [125] Chamberlain SD, Redman AM, Wilson JW, Deanda F, Shotwell JB, Gerding R, et al. Optimization of 4, 6-bis-anilino-1H-pyrrolo [2, 3-d] pyrimidine IGF-1R tyrosine kinase inhibitors towards JNK selectivity. *Bioorganic & medicinal chemistry letters*. 2009;19(2):360–364.
- [126] Wu J, Tseng YD, Xu CF, Neubert TA, White MF, Hubbard SR. Structural and biochemical characterization of the KRLB region in insulin receptor substrate-2. *Nature structural & molecular biology*. 2008;15(3):251.

- [127] Parang K, Till JH, Ablooglu AJ, Kohanski RA, Hubbard SR, Cole PA. Mechanism-based design of a protein kinase inhibitor. *Nature structural biology*. 2001;8(1):37–41.
- [128] Anastassiadis T, Duong-Ly KC, Deacon SW, Lafontant A, Ma H, Devarajan K, et al. A highly selective dual insulin receptor (IR)/insulin-like growth factor 1 receptor (IGF-1R) inhibitor derived from an extracellular signal-regulated kinase (ERK) inhibitor. *Journal of Biological Chemistry*. 2013;288(39):28068–28077.
- [129] Hu J, Liu J, Ghirlando R, Saltiel AR, Hubbard SR. Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Molecular cell*. 2003;12(6):1379–1389.
- [130] Stauffer F, Cowan-Jacob SW, Scheufler C, Furet P. Identification of a 5-[3-phenyl-(2-cyclic-ether)-methylether]-4-aminopyrrolo [2, 3-d] pyrimidine series of IGF-1R inhibitors. *Bioorganic & Medicinal Chemistry Letters*. 2016;26(8):2065–2067.
- [131] Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*. 1993;234(3):779–815.
- [132] Shapovalov MV, Dunbrack Jr RL. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 2011;19(6):844–858.
- [133] Roy K, Kar S, Das RN. Chapter 5 - Computational Chemistry. In: Roy K, Kar S, Das RN, editors. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Boston: Academic Press; 2015. p. 151–189. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128015056000053>.
- [134] Bhasin M, Raghava GPS. 8 - Computational Methods in Genome Research. In: Arora DK, Berka RM, Singh GB, editors. *Applied Mycology and Biotechnology*. vol. 6 of *Applied Mycology and Biotechnology*. Elsevier; 2006. p. 179–207. Available from: <https://www.sciencedirect.com/science/article/pii/S1874533406800110>.
- [135] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*. 1963;7(1):95–99. Available from: <https://www.sciencedirect.com/science/article/pii/S0022283663800236>.

- [136] Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *BioMol Concepts*. 2010;1(3-4):271–283. Available from: <https://doi.org/10.1515/bmc.2010.022>.
- [137] Ho BK, Brasseur R. The Ramachandran plots of glycine and pre-proline. *BMC structural biology*. 2005;5(1):1–11.
- [138] Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 2018;27(1):293–315.
- [139] Diago LA, Morell P, Aguilera L, Moreno E. Setting up a large set of protein-ligand PDB complexes for the development and validation of knowledge-based docking algorithms. *BMC bioinformatics*. 2007;8(1):1–15.
- [140] Gunasekaran K, Nussinov R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *Journal of molecular biology*. 2007;365(1):257–273.
- [141] Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by $C\alpha$ geometry: ϕ , ψ and $C\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics*. 2003;50(3):437–450.
- [142] Wedemeyer WJ, Welker E, Scheraga HA. Proline cis- trans isomerization and protein folding. *Biochemistry*. 2002;41(50):14637–14644.
- [143] Shehu A, Kaviraki LE. Modeling Structures and Motions of Loops in Protein Molecules. *Entropy*. 2012;14(2):252–290. Available from: <https://www.mdpi.com/1099-4300/14/2/252>.
- [144] Pal D, Chakrabarti P. On residues in the disallowed region of the Ramachandran map. *Biopolymers: Original Research on Biomolecules*. 2002;63(3):195–206.
- [145] Zhang L, Luo S, Zhang B. Glycan analysis of therapeutic glycoproteins. *mAbs*. 2016;8(2):205–215. PMID: 26599345.