

UNIVERSIDAD DEL PAPALOAPAN

Campus Tuxtepec

División de Estudios de Posgrado

**Estudio de la relación cuantitativa entre la estructura
química y la actividad (QSAR) de inhibidores selectivos
del parólogo GSK3 β**

TESIS

Para obtener el grado de

Maestro en Ciencias Químicas

Presenta:

Marco Antonio Hipólito Gutiérrez

Director de tesis:

Dr. Guillermo Ramírez Galicia

Codirector de tesis:

Dr. Octavio Silva García

San Juan Bautista Tuxtepec, Oaxaca, México 2022.



UNIVERSIDAD DEL PAPAALOAPAN

DIVISIÓN DE ESTUDIOS DE POSGRADO

OFICIO	DEP-MCQ/2022/11
ASUNTO	Autorización de impresión de tesis

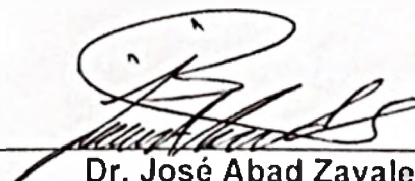
San Juan Bautista Tuxtepec, Oaxaca, México a 28 de febrero de 2022

M. EN C. HÉCTOR LÓPEZ ARJONA
VICE-RECTOR ACADÉMICO
UNIVERSIDAD DEL PAPAALOAPAN

Sirva la presente para informarle que el jurado del examen para obtener el grado de Maestro en Ciencias Químicas del C. Marco Antonio Hipólito Gutiérrez, con número de Matrícula 18130004, ha autorizado la impresión del manuscrito que lleva por título "Estudio de la relación cuantitativa entre la estructura química y la actividad (QSAR) de inhibidores selectivos del parálogo GSK3b", para su posterior presentación y defensa por parte del sustentante.

De antemano agradezco su atención, sin más que agregar, quedo a sus órdenes

Atentamente
terra uberrima, mens aperta
Bou Lo-tama, chí jí jú



Dr. José Abad Zavaleta

Jefe de la División de Estudios de Posgrado



C c p C. Marco Antonio Hipólito Gutiérrez – Estudiante
C c p Dr. Guillermo Ramirez Galicia – Director de Tesis
C c p Dr. Erick Adrian Juarez Arellano – Coordinador del Posgrado en Ciencias Químicas
C c p M.E. Yesenia Barrientos Arenal – Jefa de Servicios Escolares
C c p Archivo

CAMPUS TUXTEPEC
C. Circuito central No. 200, Col. Parque Industrial.
C.P. 38301, Tuxtepec, Oax.
Tel. 01(287)8759240

www.unpa.edu.mx

CAMPUS LOMA BONITA
Av. Ferrocarril S/N, Ciudad universitaria.
C.P. 68400, Loma Bonita, Oax.
Tel. 01(281)8729230



UNIVERSIDAD DEL PAPALOAPAN

DIVISIÓN DE ESTUDIOS DE POSGRADO

OFICIO	DEP/2022/270
ASUNTO	Jurado para examen de grado

San Juan Bautista Tuxtepec, Oaxaca, a 11 de enero de 2022

C. MARCO ANTONIO HIPOLITO GUTIERREZ
MAESTRIA EN CIENCIAS QUÍMICAS
UNIVERSIDAD DEL PAPALOAPAN

Por este medio le informo que el jurado de su examen para obtener el grado de Maestría en Ciencias Químicas estará integrado por los siguientes investigadores.

Dr. Guillermo Ramírez Galicia	Universidad del Papaloapan
Dr. Francisco Noé Mendoza Ambrosio	Universidad del Papaloapan
Dr. Emanuel Hernández Núñez	CINVESTAV
Dra. Roxana Martínez Pascual	Universidad del Papaloapan
Dr. Enrique Villalobos Amador	Universidad del Papaloapan

Sin más por el momento, le envío saludos cordiales.

Atentamente

terra uberrima. mens aperta
Bou Lo-tama, chi ji jú

Dr. José Abad Zavaleta
Jefe de la División de Estudios
de Posgrado



M. en C. Héctor López Arjona
Vice-Rector Académico
Vo. Bo.



C.c.p. Dr. Guillermo Ramírez Galicia – Director de tesis
C.c.p. Dr. Aurelio Hernández Ramírez – Coordinador de la Maestría en Ciencias Químicas
C.c.p. M. E. Yesenia Barrientos Arenal – Jefa del Departamento de Servicios Escolares

CAMPUS TUXTEPEC
C. Circunferencial No. 200, Col. Parque Industrial,
C.P. 38301, Tuxtepec, Oax.
Tel. 01(287)8759230

www.unpa.edu.mx

CAMPUS LOMA BONITA
Av. Ferrocarril S/N, Ciudad Universitaria,
C.P. 68400, Loma Bonita, Oax.
Tel. 01(281)8729230

Dedicatoria.

Esta tesis está dedicada a:

A mis padres Marco A.H.G. y Crisanta G.P. quienes con su amor y esfuerzo me han permitido llegar a cumplir un sueño más, gracias por inculcar en mí el ejemplo de esfuerzo y dedicación, de no temer las adversidades porque Dios está conmigo siempre.

A mi hermano Eduardo D.H.G., mi esposa Nancy G.P.H e hija Mía N.H.P, por su apoyo incondicional, durante todo este proceso, por estar conmigo en todo momento gracias. A toda mi familia porque con sus oraciones, consejos y palabras de aliento hicieron de mí una mejor persona y de una u otra forma me acompañan en todos mis sueños y metas.

Agradecimientos.

Quiero expresar mi gratitud a Dios, quien con su bendición llena siempre mi vida y a toda mi familia por estar siempre presentes.

Mi profundo agradecimiento a todas las autoridades y personal de la unidad educativa, por confiar en mí, abrirme las puertas y permitirme realizar todo el proceso dentro de su establecimiento educativo.

De igual manera mis agradecimientos a la Universidad del Papaloapan, a toda la Facultad de Química, a mis profesores en especial a la Dr. Guillermo Ramírez Galicia, Dr. Octavio Silva García y Dr. Francisco Noé Mendoza Ambrosio quienes con la enseñanza de sus valiosos conocimientos hicieron que pueda crecer día a día como profesional, gracias a cada uno de ustedes por su paciencia, dedicación, apoyo incondicional y amistad.

Finalmente quiero expresar mi más grande y sincero agradecimiento al Dr. Emanuel Hernández Núñez del CINVESTAV, quien me apoyó y me permitió realizar este viaje.

Índice.

Lista de abreviaturas.....	1
Resumen.	2
Abstract.....	3
Introducción.	4
Capítulo 1.- Antecedentes.....	6
1. La proteína glucógeno sintasa cinasa (GSK3).....	6
1.1. Generalidades de la GSK3.....	6
1.2. Roles Biológicos de GSK3.	8
1.3. GSK3 y su relación con enfermedades.	11
1.3.1. Enfermedad de Alzheimer.	12
1.3.2. GSK3 en la síntesis de péptidos A β	13
1.3.3. GSK3 en la fosforilación de Tau.	14
1.4. GSK3 en el desarrollo de varios tipos de cáncer.....	16
1.5. Estrategias para inhibir la función de GSK3.	18
1.6. Los parálogos de la proteína GSK3.	21
1.7. El paradigma de la inhibición de la proteína GSK3.....	22
1.8. Inhibición selectiva del parálogo GSK3 β	23
1.9. Características cristalográficas de GSK3 β	24
1.10. Mecanismo de fosforilación de GSK3 β	26
1.11. Factibilidad de inhibidores selectivos de la proteína GSK3 β	28
Capítulo 2.- Química computacional.	30
2.1. ¿La química computacional es una alternativa?.....	30
2.2. Química computacional.....	31
2.3. Modelado molecular.....	34
2.4. Dinámica molecular.....	35
2.5. Relación cuantitativa estructura actividad (QSAR).	36
2.6. Ecuación QSAR.	37
2.6.1. El coeficiente de Pearson.	38
2.7. Actividad biológica.	38
2.8. Descriptores.....	39
2.9. Redes neuronales.	40
2.10. Validación del modelo.....	41
2.11. Método de aplicabilidad de dominio.	45
2.12. Aproximación por apalancamiento (LA).	47

2.13. Aproximación por estandarización (SA).....	48
Capítulo 3.- Justificación.....	50
Capítulo 4.- Hipótesis.....	50
Capítulo 5.- Objetivos.....	51
5.1. Generales.....	51
5.2. Específicos.....	51
Capítulo 6.- Metodología.....	52
6.1. Dinámica molecular GSK3 β	52
6.2. Búsqueda de moléculas reportadas en la base de datos ChemBL.....	53
6.3. Representación de moléculas inhibidoras.....	54
6.4. Optimización geométrica de los inhibidores.....	54
6.5. Acoplamiento molecular.....	55
6.6. Cálculo de descriptores cuánticos.....	55
6.7. Modelo QSAR.....	55
6.7.1. Regresión lineal múltiple.....	56
6.7.2. Análisis de aplicabilidad del dominio.....	56
6.7.3. Valores atípicos.....	57
6.7.4. Análisis MAE.....	57
6.8. Comparación de los parálogos de la proteína GSK3.....	58
6.9. Propuesta de moléculas nuevas.....	58
Capítulo 7.- Resultados.....	60
7.1. Dinámica molecular.....	60
7.2. Radio de giro.....	62
7.3. Búsqueda de moléculas reportadas en la base de datos ChemBL.....	63
7.4. Síntesis <i>in silico</i> de los inhibidores.....	67
7.5. Optimización geométrica de los inhibidores.....	68
7.6. Cálculos descriptores cuánticos.....	69
7.7. Acoplamiento molecular.....	70
7.8. Diferencias de los parálogos de la proteína GSK3.....	73
7.9. Cálculo de Descriptores.....	77
7.10. Preselección de descriptores.....	80
7.11. Generar el modelo QSAR.....	81
7.11.1. Entrenamiento y predicción.....	81
7.11.2. Regresión lineal múltiple (RLM).....	81
7.11.3. Análisis AIC y FIC.....	85
7.11.4. Aplicabilidad del dominio.....	87

7.11.5. Análisis MAE y ANN.....	88
7.12. Inhibidores preferentemente beta.	94
Capítulo 8.- Descriptores y propuesta de inhibidores selectivos.....	98
Capítulo 9.- Conclusiones.	99
Aportaciones.....	100
Referencias.....	101
Anexo.....	108

Índice de figuras.

Figura 1.- Enzima GSK3.	7
Figura 2.-Fosforilación de Serinas por GSK3.	8
Figura 3.- Red neuronal.	9
Figura 4.-GSK3 "Regulador Maestro".....	11
Figura 5.-GSK3, relación con enfermedades.....	11
Figura 6.-Representación de la enfermedad del Alzheimer.	12
Figura 7.-Degradación neuronal por hiperfosforilación de Tau.	15
Figura 8.-Cavidades analizadas de GSK3.....	20
Figura 9.-Parálogos de GSK3.	22
Figura 10.-GSK3 β representación listones.	25
Figura 11.-GSK3 β , dominios de fosforilación.	27
Figura 12.-GSK3 β proceso de fosforilación.....	27
Figura 13.-Representación de Red neuronal.....	41
Figura 14.- Error medio absoluto.....	45
Figura 15.-Métodos para determinar la aplicabilidad del dominio	47
Figura 16.-Análisis de la aplicabilidad de dominio.	56
Figura 17.-Pasos de la metodología.....	59
Figura 18.-GSK3 β Cristalizada.....	60
Figura 19.-GSK3 β cambios fisiológicos.....	61
Figura 20.-Radio de giro.	62
Figura 21.-Clasificación de los inhibidores.	64
Figura 22.-Síntesis in sílico.	68
Figura 23.-Optimización geométrica.....	69

Figura 24.-Edición de caja para Docking.....	70
Figura 25.-Grafica del acoplamiento molecular número 81.	71
Figura 26.-inhibición nula.	72
Figura 27.-Acoplamiento molecular número 3, aminoácidos del sitio catalítico.	73
Figura 28.-Aminoácidos de GSK3 α y GSK3 β	74
Figura 29.-Comparación de GSK3 β nativa y GSK3 β modificada.....	75
Figura 30.-Diferencias de GSK3 β modifica y nativa.	76
Figura 31.-Disimilitud de GSK3 β nativa y modificada.	77
Figura 32.-Gráfica de regresiones lineales múltiples realizadas por SSPS.	84
Figura 33.-Gráfica de dispersión con 13 descriptores.	85
Figura 34.-Correlación 12 descriptores.	86
Figura 35.-Aplicabilidad del dominio con 12 descriptores.....	88
Figura 36.-Gráfica de correlación ANN preferente beta.....	90
Figura 37.- Gráfica correlación ANN preferente beta y alfa.	90
Figura 38.- Gráfica correlación ANN preferente beta y sin preferencia.....	91
Figura 39.- Gráfica correlación ANN preferente beta β con menor residuo.	92
Figura 40.- Gráfica correlación ANN preferente beta y alfa con menor residuo.	93
Figura 41.- Gráfica correlación ANN preferente beta y sin preferencia con menor residuo.	94
Figura 42.- Interacción de moléculas preferente β con proteína GSK3 β	95
Figura 43.-Interacción de moléculas preferentes α con proteína GSK3 β	96
Figura 44.-Interacción de moléculas sin preferencia con proteína GSK3 β	97
Figura 45.- Base estructural para inhibidores de GSK3 β	98

Índice de tablas.

Tabla 1.-Moléculas inhibidoras con preferencia por GSK3 β	28
Tabla 2.-Inhibidores preferente GSK3 β	65
Tabla 3.-Inhibidores preferente GSK3 α	66
Tabla 4.-Inhibidores sin preferencia.	67
Tabla 5.-Descriptores calculados.	78

Tabla 6.-Descriptores comunes.....	79
Tabla 7.-Descriptores con valores constantes.....	80
Tabla 8.-Resultado de SSPS.	82
Tabla 9.- Listas SSPS obtenidas.....	84
Tabla 10.-Calculo de AIC y FIT.	86
Tabla 11.-Resultados obtenidos con el cálculo AIC y FIT.	87
Tabla 12.-Análisis MAE.....	89
Tabla 13.-Resultados obtenidos con el análisis MAE.	90
Tabla 14.- Resultados obtenidos preferentemente beta más alfa.....	91
Tabla 15.-Resultados obtenidos preferentemente beta más sin preferencia ...	91
Tabla 16.-Resultados obtenidos preferentemente beta con menores residuos.	92
Tabla 17.-Resultados obtenidos preferentemente beta más preferencia alfa con menor residuo.	93
Tabla 18.-Resultados obtenidos preferentemente beta más sin preferencia con menor residuo.	94
Tabla 19.-Descriptores del modelo.....	98
Tabla 20.-Inhibidores selectivos por la proteína GSK3 β	108

Lista de abreviaturas.

Sigla	Significado.
GSK3	Glycogen Synthase Kinase 3 (Glucógeno sintasa cinasa).
EA	Enfermedad del Alzheimer.
Ser/Thr	Serina/treonina.
GSK3 α	Glycogen Synthase Kinase 3 Alfa (Glucógeno sintasa cinasa Alfa).
GSK3 β	Glycogen Synthase Kinase 3 Beta (Glucógeno sintasa cinasa Beta).
PK	Proteína quinasa
FA	Factor activador A.
Mg	Magnesio
ATP	Adenosín trifosfato.
A β	Beta amiloide.
TPKI	Tau protein kinase I (proteína tau quinasa I).
APP	Amyloid Precursor Protein (proteína amiloide precursora).
RNA	Ribonucleic acid (Ácido ribonucleico).
WT	Wild Type (Ratones Silvestres).
LMA	Leucemia mieloide aguda.
DFT	Density Functional Theory (teoría del funcional de la densidad).
HF	Hartree-Fock.
QSAR	Quantitative Structure-Activity Relationship (Relación cuantitativa estructura actividad).
RLM	Multiple Linear Regression (regresión lineal múltiple).
In silico	En silicio (Realizado por Computadora).
Rg	Radio de Giro.
Ki	Inhibitor constant (Contante inhibitoria).
HOMO	Highest Occupied Molecular Orbital (Orbital molecular ocupado de más alta energía).
LUMO	Lowest Unoccupied Molecular Orbital (Orbital molecular no ocupado de más baja energía)
η	Dureza.
χ	Electronegatividad.
ω	Electrofilicidad.

Resumen.

La proteína glucógeno sintasa quinasa 3 (GSK3), es la quinasa más activa en la mayoría de las células, el cual desempeña un papel en una serie de procesos fisiológicos que van desde el metabolismo del glucógeno hasta la transcripción de genes, conservada y ampliamente codificada por dos genes que generan dos proteínas relacionadas GSK3 α y GSK3 β , los cuales comparten el 98% de similitud de identidad de aminoácidos dentro de su dominio catalítico, es activa en las células en condiciones de reposo y se regula principalmente mediante la inhibición o desviación de su actividad.

La desregulación de las vías de señalización que involucran a GSK3 está asociada con la proliferación de numerosos tipos de patologías, se determinó a través de la metodología QSAR, las características principales de los inhibidores reportadas en bases de datos que permitan una inhibición selectiva, debido a que es de beneficio una inhibición selectiva de la proteína GSK3 β .

Con todo lo obtenido se ofrece una base de datos con moléculas con selectividad por la proteína GSK3 β así como la propuesta de una estructura química que permitirá el diseño de nuevos prospectos a fármacos con mayor potencia de inhibición donde la base de estudio fue la constante de inhibición (KI).

Abstract.

The protein glycogen synthase kinase 3 (GSK3), is the most active kinase in most cells, playing a role in several physiological processes ranging from glycogen metabolism to gene transcription, conserved and widely encoded by two genes that generate two proteins related GSK3 α and GSK3 β , which share 98% similarity of amino acid identity within their catalytic domain, are activated in cells under repose conditions and are regulated mainly by inhibiting or diverting their activity.

The deregulation of signaling pathways involving GSK3 is associated with the proliferation of numerous types of pathologies, was determined through the QSAR methodology, the main characteristics of the inhibitors reported in databases that allow a selective inhibition, due to the benefit of a selective inhibition of the GSK3 β protein.

With everything obtained, a database with molecules with selectivity for the GSK3 β protein is offered, as well as the proposal of a chemical structure that will allow the design of new prospects for drugs with greater inhibition power where the basis of study was the inhibition constant (KI).

Introducción.

La glucógeno sintasa cinasa 3 (GSK3) es una enzima presente en organismos eucariotas que inicialmente se asoció al metabolismo del glucógeno, sin embargo, se ha caracterizado su participación en diferentes funciones celulares como la apoptosis, la división celular y la inflamación. A la fecha, la GSK3 es la enzima cinasa con el mayor número de sustratos identificados y se expresa en todos los tejidos de organismos metazoarios. La diversidad de sustratos que posee la enzima es en parte la razón de su importancia en la regulación de múltiples procesos celulares. Presenta la peculiaridad de ser una enzima constitutivamente activa y cuya regulación se lleva a cabo principalmente por inhibición de su función; lo que le confiere la capacidad de limitar la activación innecesaria de los procesos que regula en células no estimuladas.¹

Poco después de su caracterización, se descubrió que a falta de una regulación adecuada de GSK3, ésta se activa promoviendo el desarrollo de diversas enfermedades como diabetes, cáncer, inflamación crónica y la enfermedad del Alzheimer. Esta característica promueve el interés en su investigación y se especuló que su inhibición podría ser beneficiosa en distintos contextos de estas enfermedades.²

Con este objetivo se ha logrado, en una primera etapa, la identificación de más de 2000 moléculas de diferente naturaleza con la capacidad de inhibir la actividad enzimática de GSK3 y su efecto de revertir los efectos nocivos de la enfermedad de Alzheimer, además reduce la viabilidad de masas tumorales de pulmón, páncreas, colon, neuroblastoma y próstata; además favorece la absorción de glucosa y reduce los índices glucémicos en sangre con lo que se reduce el desarrollo de diabetes.³

No obstante, aunque parece un panorama prometedor, la inhibición de GSK3 como terapia presenta dificultades, debido a la amplia gama de sustratos que son regulados por esta enzima, su inhibición afecta no solo a los tejidos enfermos. Por ejemplo, se ha descrito que la inhibición crónica de GSK3 en ratones favorece el desarrollo de leucemia múltiple y favorece cambios histológicos en diversos tejidos haciendo pensar que la terapia basada en la inhibición de GSK3 podría no ser factible.⁴ Se ha observado que los efectos

nocivos que acompañan a la inhibición de la GSK3 se deben en parte a la pérdida de regulación negativa en la vía Wnt/ β -catenina debido a la transducción de señales responsable de controlar la proliferación y regulación celular.⁵

Un aspecto no explorado hasta la fecha es que las funciones enzimáticas de GSK3 son llevadas a cabo por 2 proteínas parálogas llamadas GSK3 α y GSK3 β .²

Se ha obtenido evidencia experimental mediante inhibición genética para reprimir la actividad de un parálogo a la vez, esto indica que los efectos benéficos de la inhibición de GSK3 se presentan siempre y cuando se inhiba a una sola proteína a la vez, sin embargo, estas proteínas presentan una similitud de 98% en su dominio catalítico lo que ha dificultado la identificación de moléculas que tengan la capacidad de inhibir preferentemente a una sola proteína. Por lo que de las más de 2000 moléculas conocidas como inhibidores de GSK3 no existe una que se emplee en la clínica como parte de un tratamiento farmacológico.⁶

Se ha reportado que existen moléculas que tienen preferencia inhibitoria sobre el parálogo GSK3 β , por ejemplo, la molécula 4-[4-(4-fluorofenil)-2-[4-(metilsulfanil) Fenil]-1H-imidazol-5-il] piridina, con identificación ChEMBL17370 en la base de datos ChEMBL. Esta molécula presenta una actividad inhibitoria 20 veces superior sobre GSK3 β que en GSK3 α . Por otra parte, se ha descrito recientemente que las proteínas presentan diferencias conformacionales que podrían explicar la preferencia inhibitoria de ciertas moléculas por un parálogo a pesar de que las proteínas sean tan similares en los residuos de aminoácidos que las forman.⁷

Es por esto, que la propuesta de este trabajo es analizar mediante una estrategia computacional, las estructuras químicas de las moléculas disponibles en bases de datos públicas, por ejemplo PubChem, ChEMBL, ChemSpider entre otros, que tengan una preferencia inhibitoria por GSK3 β para determinar los mecanismos de acoplamiento con la enzima y conocer las características de estos inhibidores con el objetivo de sintetizar nuevas moléculas más potentes y selectivas que permitan la terapia enfocada en GSK3 β .

Capítulo 1.- Antecedentes.

1. La proteína glucógeno sintasa cinasa (GSK3).

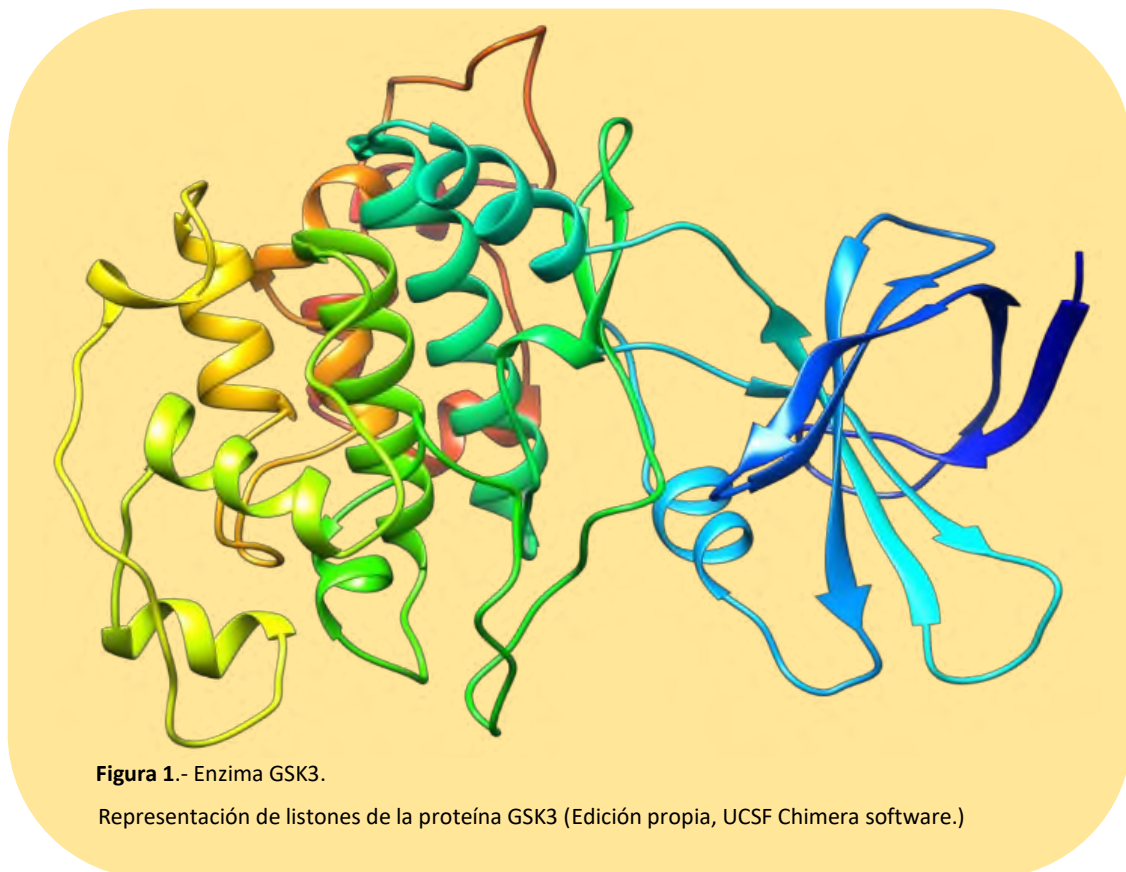
1.1. Generalidades de la GSK3.

La GSK3 es una proteína de serina/treonina (Ser/Thr) cinasa, ha estado implicada durante años en la regulación de la inflamación y se ha abordado en una gran cantidad de informes científicos, utilizándose en una gran variedad de modelos y de enfoques experimentales (enfermedades). La GSK3 está involucrada en la regulación de una plétora de eventos intracelulares moleculares, orgánicos celulares y pato-fisiológicos.⁸ Se ha observado que la GSK3 está conservada evolutivamente en organismos eucariotes desde mamíferos, insectos, hongos, helmintos e incluso plantas. Esta enzima regula diferentes procesos celulares como la división celular, la diferenciación, el metabolismo del glucógeno, la inmunidad y la apoptosis. Hasta la fecha, se han encontrado más de 100 sustratos de la enzima GSK3 por evidencia experimental, pero los datos de bioinformática predicen más de 520 sustratos.⁹ Aunque es muy probable que este número sea una sobreestimación, la proteína GSK3 es probablemente la enzima cinasa más activa en las células eucariotes y explica en parte los diversos procesos celulares donde participa.¹⁰

El primer miembro que se identificó de esta proteína fue en extractos de proteínas del músculo esquelético de conejo como parte de una vía metabólica capaz de fosforilar la proteína glucógeno sintasa en un mecanismo independiente de otras dos enzimas: la proteína quinasa cíclica dependiente de Adenosín monofosfato cíclico (AMPc) GSK1 y fosforilasa cinasa 2 (GSK2). La proteína glucógeno sintasa cinasa fue la tercera enzima descrita que fosforila la glucógeno sintasa de donde obtuvo su nombre.⁷ La GSK3 es una enzima particular porque en las células en reposo es constitutivamente activa, lo que significa que la regulación de su actividad se lleva a cabo por inactivación o desvío de su actividad en lugar de activarse con un estímulo.⁷ Otra característica interesante de esta proteína es que muestra preferencia por sustratos pefosforilados, otras cinasas y el resultado de su actividad conduce en general a la pérdida de función del sustrato.¹¹

Notablemente, la proteína puede ser inhibida por su propio dominio *N-terminal* después de la fosforilación en el residuo Serina 9 (Ser9) porque este dominio

modificado constituye un pseudo sustrato a su propio dominio catalítico y compite con los sustratos por la unión.¹² En este contexto de acción, para que un sustrato de GSK3 lleve a cabo su actividad, se necesitan dos eventos: 1) la activación por una cinasa específica y 2) la inactivación de GSK3 por la misma cinasa o una alternativa. Por ejemplo, el factor transcripcional que es un elemento de la respuesta a la unión de AMPc (CREB) es fosforilado, por la proteína quinasa (PK), también conocida como Akt, en el residuo Serina-133 (Ser133) para estimular su actividad transcripcional.¹³ Sin embargo, pCREB-Ser133 es un sustrato de GSK3 que lo fosforila en el residuo Serina-129 (Ser129), dando como resultado una reducción de la actividad transcripcional de CREB. En la Figura 1 podemos visualizar a la proteína GSK3.



Por otra parte, durante la síntesis de los polipéptidos de la proteína, las moléculas se auto fosforilan en el residuo de Tirosina-216 (Tyr-216), esta modificación potencia su actividad cinasa al facilitar el acceso de los sustratos al dominio catalítico, aunque estrictamente no es necesaria para que la enzima fosforile sustratos. Curiosamente, no se ha encontrado que la proteína fosforile

residuos de Tirosina (Tyr) en otras proteínas, esto significa que GSK3 es una Tyr cinasa en su etapa de síntesis, pero cuando las cadenas de polipéptidos están completamente maduras, su actividad cinasa evoluciona a ser serina/treonina (Ser/Thr) cinasa.¹⁴

Otra característica interesante de esta proteína es el hecho de que puede fosforilar sustratos de manera múltiple y en forma seriada en el sitio de consenso S/TXXXS/T(p), si la secuencia en el sustrato se repite, la enzima puede fosforilar cada S/T ubicada de 3 a 4 residuos después del último residuo fosforilado en la dirección amino terminal creando su propio sustrato cebado. Por ejemplo, en la glucógeno sintasa, pueden fosforilar los residuos Ser 641, 645, 649 y 653. En la Figura 2 se muestra este proceso.¹⁵

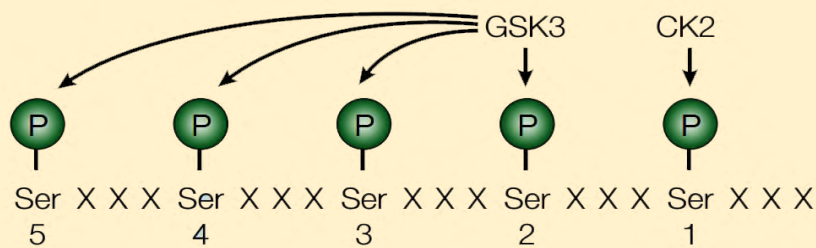


Figura 2.-Fosforilación de Serinas por GSK3.

Reconocimiento de Serinas por GSK3 para su fosforilación Ser/XXX/Ser.¹

1.2. Roles Biológicos de GSK3.

Aunque se limitó inicialmente a las vías metabólicas del glucógeno, un descubrimiento interesante mostró que la proteína factor activador A (FA) de la proteína fosfatasa dependiente de magnesio-adenosín trifosfato, (Mg-ATP) también está relacionada con la fosforilación de neurofilamentos en las células neuronales; por lo tanto, GSK3 se clasificó como una cinasa de múltiples sustratos.^{16,17} Dada su participación como represor de varias vías de señalización celular, como la apoptosis, la insulina, la fosfoinositol 3-cinasa (PI3K), Wnt/ β -catenina, entre otros, esta proteína está involucrada esencialmente en todos los procesos principales de la célula.¹⁸ Las funciones de esta proteína en las células son variadas y dependen del tipo celular que se esté analizando. Por ejemplo, en las células neuronales la proteína GSK3 regula la

formación de las sinapsis nerviosas mediante la desestabilización de la estructura microtubular en un mecanismo que involucra la fosforilación de la proteína tau.⁹ Donde Tau es una proteína estructural cuya principal función es la de estabilizar los microtúbulos de actina que forman la estructura básica del citoesqueleto neuronal. Al ser fosforilada la proteína tau por GSK3, esta se desensambla de los microtúbulos perdiendo de esta manera estabilidad, esto elimina las conexiones neuronales, lo que evita poder formar otras nuevas, dando lugar así a la plasticidad neuronal.³ En la Figura 3 podemos ver el ejemplo de una red neuronal.

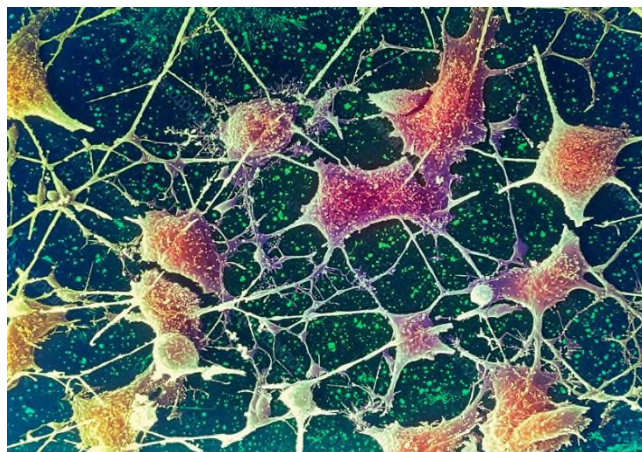


Figura 3.- Red neuronal.

Micrografía electrónica de barrido en color (SEM) **Autor:** Steve G.

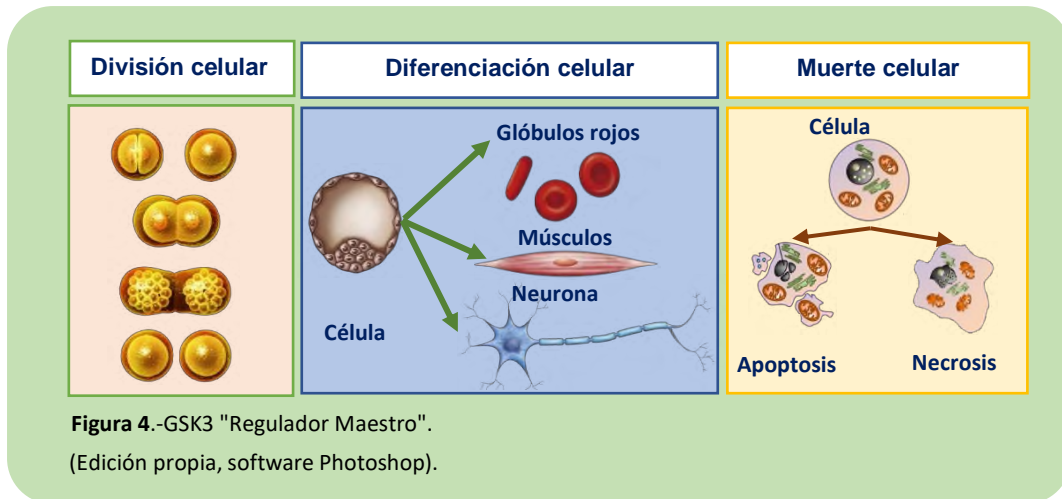
Otro ejemplo de las funciones de la proteína GSK3 es la regulación de la respuesta inflamatoria en los leucocitos. Durante un proceso infeccioso, que puede ser viral, bacteriano, fúngico o parasitario. Los microorganismos patógenos son reconocidos por una familia de receptores membranales llamados receptores tipo Toll (TLR) los cuales representan una parte importante en el vínculo entre la respuesta inmune innata y adaptativa.¹⁹

Este reconocimiento conduce a la activación de la vía de transducción PI3K/Akt (la cual es una vía de señalización intracelular importante para regular el ciclo celular), al ser activada regula negativamente la función cinasa de GSK3. De manera interesante, mientras GSK3 esté activa (actividad constitutiva) se favorece la actividad del factor transcripcional, factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas (NF- κ B) y la expresión de

citocinas proinflamatorias. Sin embargo, cuando la estimulación de los leucocitos es continua por parte de los microorganismos, la actividad de GSK3 al ser inhibida conduce a la activación del factor transcripcional CREB y a la expresión de citocinas antiinflamatorias. Esto significa que la capacidad de los leucocitos de regular de manera positiva o negativa el proceso inflamatorio depende del estado de actividad de GSK3.²

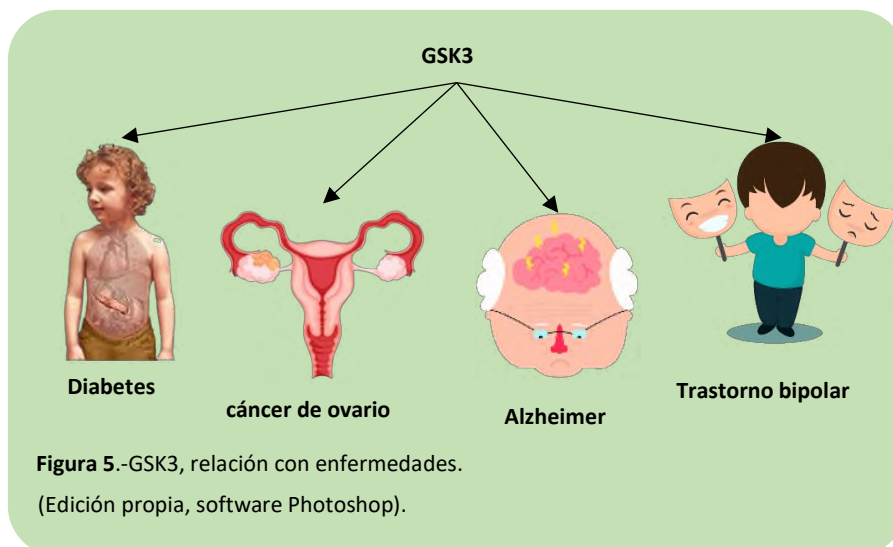
Por lo tanto, la función de la proteína en células madre es vital para regular su diferenciación, durante el desarrollo embrionario y fetal, la función de GSK3 es vital para permitir una correcta diferenciación y polarización de las células embrionarias.¹ Uno de los primeros órganos desarrollados durante el crecimiento embrionario es el corazón.²⁰ Los factores de transcripción que regulan este proceso están bien identificados y corresponden a genes que son regulados por la vía Wnt/ β -catenina, de la cual GSK3 es el principal regulador negativo.⁶ La importancia de GSK3 en el desarrollo embrionario ha quedado de manifiesto en los experimentos llevados a cabo en la mosca de la fruta en los cuales, la delección de la proteína Shaggy (homóloga de GSK3) provoca la muerte embrionaria debido a la incapacidad de las células madre de diferenciarse en diversos tejidos incluido el corazón.²¹ Este mismo efecto ha sido observado en mamíferos en los que como resultado de la ausencia de función de GSK3 se pierde la capacidad de las células madre de diferenciarse y reproducirse.¹⁴

En resumen, la proteína GSK3 se ha reconocido como un “regulador maestro” implicado en funciones celulares de diversa índole, lo que demuestra su importancia en la fisiología celular y que es dependiente del tipo de células en estudio. Cabe mencionar que los ejemplos presentados no son únicos ya que esta proteína también está involucrada en la regulación metabólica, procesos de muerte celular por apoptosis, autofagia, regulador del ciclo celular, diferenciación celular entre otros. ²² En la Figura 4 se observa un cuadro en el que contiene una serie de procesos donde la proteína GSK3 está implicada, lo cual representa la gran versatilidad.



1.3. GSK3 y su relación con enfermedades.

Debido a su implicación en diversas vías de señalización celular, después de su descubrimiento se reportó que la desregulación de la proteína GSK3 participa de manera activa en el desarrollo de varias enfermedades. Dada su actividad constitutiva, los mecanismos que conducen a una falta de regulación negativa sobre la proteína, provocan una “ganancia de función” que se ha descrito que está presente en enfermedades como Alzheimer, diabetes, trastorno bipolar, cáncer, entre otros.²² En la Figura 5, se ilustran algunos ejemplos.



1.3.1. Enfermedad de Alzheimer.

La enfermedad de Alzheimer (EA) es una de las principales causas de la demencia neurodegenerativa en los seres humanos. Se caracteriza por la presencia constante de dos lesiones fisiopatológicas distintivas en el tejido cerebral: las placas seniles en el espacio extracelular formadas por la polimerización de los péptidos β -amiloides ($A\beta$) de 40 y 42 aminoácidos, y las "madejas" neurofibrilares, es decir, conglomerados anormales de proteínas compuestos por pequeñas fibrillas entrelazadas dentro de las neuronas. Una característica principal de la EA y de manera más precisa en el compartimento intracelular de las células neuronales es que están estructuradas de formas hiperfosforiladas por la proteína tau. Por lo tanto, estas lesiones tisulares impiden la función neuronal adecuada y, en última instancia, afectan las conexiones de las células neuronales es decir señalización neuronal, lo que provoca la degeneración o muerte de las neuronas, lo que finalmente termina con la memoria y otras funciones mentales importantes.²³ En la Figura 6 observamos las diferencias de un cerebro sano a uno con la enfermedad el Alzheimer, provocada por la falta de regulación de la proteína GSK3.

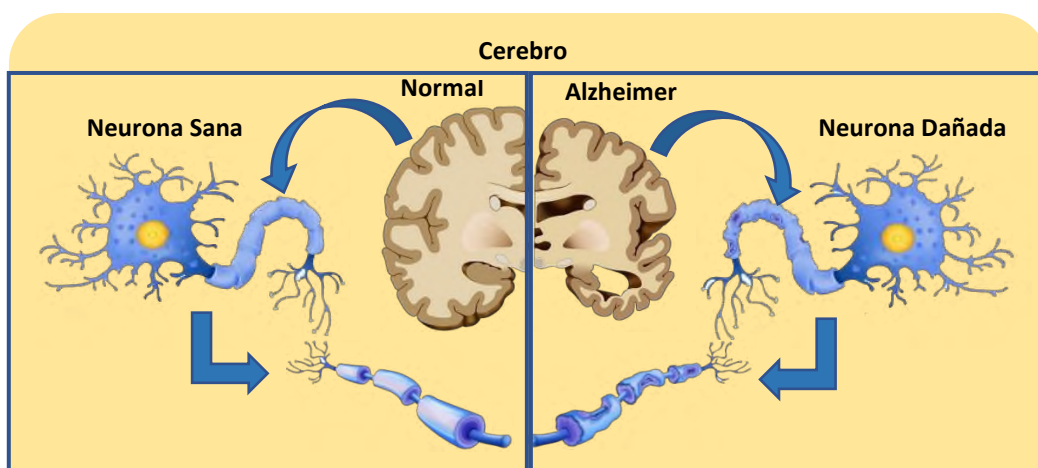


Figura 6.-Representación de la enfermedad del Alzheimer.
(Edición propia, software Photoshop).

En estudios de estimulación con péptidos $A\beta$ se detectó que estos promueven la actividad de la proteína GSK3, lo que provoca una sobreactividad de la proteína GSK3, misma que propicia el proceso fisiopatológico de la proteína tau, indicador

principal de la EA en experimentos llevados a cabo en neuronas del hipocampo de ratas.⁷ En contraste, la adición de oligonucleótidos “anti sentido” para reducir la expresión de TPKI previene la muerte neuronal y reduce la abundancia de los péptidos A β evitando la EA.²⁴ También se ha descrito que el cloruro de litio (LiCl), inhibidor de GSK3, puede impedir la síntesis de péptidos A β , lo que pone de relieve la relación entre la actividad de GSK3 con la síntesis de péptidos A β y viceversa.²⁵ Por otra parte, mediante técnicas de inmunohistoquímica dirigida para detectar a GSK3 en el tejido cerebral de los pacientes con EA, se ha observado que la proteína muestra una señal de co-localización con la proteína tau hiperfosforilada y que la expresión de GSK3 en pacientes con EA se incrementa hasta un 200% lo que establece su relación con la enfermedad.²⁴

1.3.2. GSK3 en la síntesis de péptidos A β .

Podríamos destacar que la EA está relacionada con la síntesis de los péptidos A β , la síntesis de los péptidos A β que componen los depósitos de la placa senil, se derivan de la proteólisis de la proteína precursora del péptido amiloide (APP), en una secuencia coordinada catalizada por la proteasa beta-secretasa del aspartilo 1 y 2 (BACE1/2) para generar el precursor C-99 y el posterior procesamiento por el corte proteolítico dependiente de la α -secretasa, β -secretasa y γ -secretasa dependiente de presenilina.²⁶ Como consecuencia, se realizaron estudios experimentales con la finalidad de reducir la síntesis de A β , principal indicador de la EA, se efectuaron experimentos con las células CHO de hámster, es decir ovario de hámster chino que expresan APP recombinante, con la finalidad de atenuar la expresión de GSK3, lo que condujo a una reducción de la síntesis de péptidos A β 40/42. Este efecto también se reprodujo en células neuronales de ratones y en células que expresan de forma estable una forma mutante del APP que reproduce características de la EA.²⁷ En este estudio, el pretratamiento de las células con LiCl no afectó a la expresión relativa del APP, lo que indicó que la interferencia de la síntesis de los péptidos A β se encuentra en el paso postranscripcional durante el procesamiento del APP. En estudios con las células de neuroblastoma humano SH-SY5Y, la transfección con RNA de silenciamiento contra GSK3, redujo el procesamiento de APP al fragmento intermedio C-99 al disminuir la expresión de BACE1 lo que condujo a una

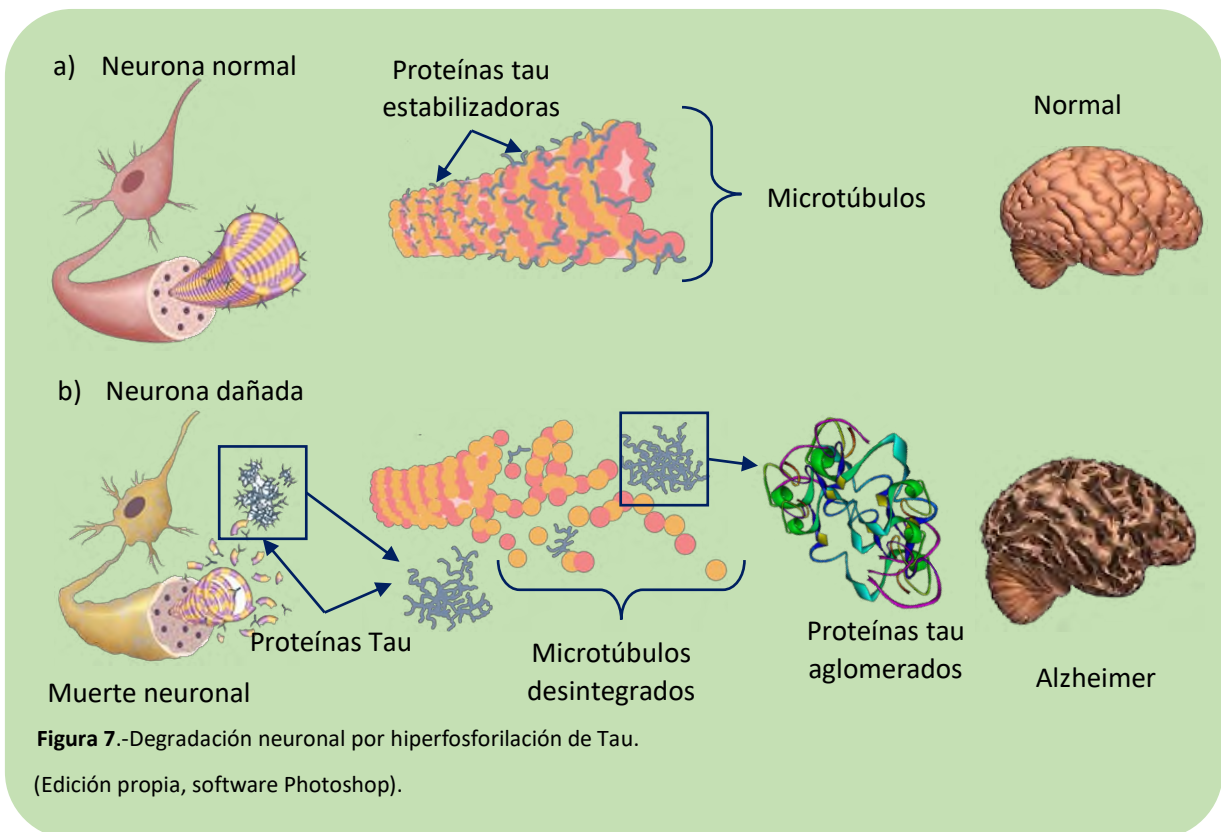
reducción de la síntesis de péptidos A β 40/42, indicando que la regulación en el procesamiento del APP por parte de GSK3 puede tener lugar en diferentes etapas.²⁸

En un escenario diferente, en un ensayo de detección para encontrar proteínas capaces de interactuar con el péptido A β 42 (forma dañina de A β), a partir de más de 9,000 proteínas, GSK3 surgió como el principal candidato. Esta interacción fue posteriormente validada por otras técnicas más precisas como resonancia plasmónica superficial, termoforesis, ensayos de extracción y microscopía confocal en cultivos primarios de células neuronales de ratones. en un ensayo *in vitro* con GSK3 y tau purificado, la adición de los péptidos A β 42 estimuló la fosforilación dependiente de GSK3 sobre tau en el residuo Ser396, lo que indica que GSK3 es un vínculo entre las dos principales fisiopatologías que se encuentran en el tejido de los pacientes con EA.²⁹

En un modelo diferente, ratones transgénicos recién nacidos que expresaron la proteína precursora del péptido amiloide mutante PDAPP+/-,³⁰ o la combinación de mutantes de APP y tau PS19+/-; PDAPP+/- (modelos de la EA)³¹ la inhibición en la expresión de GSK3 evitó la síntesis de péptidos A β , la deposición de placa senil en ratones a los 11 meses de edad y mejoró las habilidades de memoria en los ratones a los 17 meses de edad.^{32,27} En conjunto, estos resultados indicaron que la inhibición específica de GSK3 podría ser beneficiosa para reprimir el procesamiento de APP para producir péptido A β .

1.3.3. GSK3 en la fosforilación de Tau.

La proteína tau forma parte del citoesqueleto en los microtúbulos de los axones de las células neuronales, formando un enlace entre los heterodímeros de la tubulina para estabilizar las fibras neuronales.³³ Se ha demostrado que la hiperfosforilación de tau conduce al desmontaje de las fibras del microtúbulo y a la pérdida de la integridad del axón, generando la muerte neuronal. En la Figura 7a se muestra la representación de la proteína tau en donde su función es estabilizar a las fibras del microtúbulo. En la Figura 7b, las proteínas tau hiperfosforiladas por GSK3 evitan la estabilización de las fibras microtubulares provocando daño a la neruona, lo que provoca su muerte.³⁴



Las subunidades monoméricas de la proteína Tau tienden a auto agregarse y forman la estructura básica, madejas neurofibrilares dentro de las células neuronales que constituyen un sello distintivo de la EA. GSK3 puede fosforilar a la proteína Tau en 26 de sus 45 residuos fosforilables identificados; por lo tanto, es la cinasa con la mayor capacidad para fosforilar Tau identificada hasta ahora.³²

En estudios con ratones transgénicos, se ha reportado que la inhibición de GSK3 puede impedir la auto agregación de Tau y su fosforilación en Ser202, Thr205, Ser262 y Thr231.³⁵ En ratones GSK3-KO (ratón knockout, ratón modificado en el que uno o varios genes están inactivados), la forma fosforilada de Tau-Ser396 se reduce en comparación con los ratones WT (Wild Type, ratones silvestres). Se han obtenido resultados similares en otras cepas de ratones GSK3-KO de la línea transgénica doble GSK3-KO/tau. P301L (modelo de EA), lo que demuestra que en los animales deficientes de GSK3, los niveles de fosfo-Tau-Ser396/404, Ser199 y Thr231 están disminuidos en comparación con los ratones de control WT, P301L.³⁶ En un enfoque diferente, en un ensayo de monitoreo llevado a cabo en células, la sobreexpresión de 352 cinasas humanas una a la vez, en

células de neuroblastoma humano SK-N-AS demostró que GSK3 es la cinasa que promueve el mayor aumento de la fosforilación de Tau sobre Ser396/404, Thr231, Ser235, y Ser202.³⁷ En el mismo estudio se reportó que GSK3 promueve la síntesis e interactúa con la estructura patológica del péptido A β que a su vez estimula la fosforilación de GSK3 sobre Tau.

En conjunto, estos datos destacan la función de GSK3 como un importante vínculo entre las dos marcas histopatológicas de la EA, revelan su papel prominente y señalan su inhibición específica como candidato para investigar su tratamiento, enfatizando la necesidad de una terapia basada en esta proteína.

1.4. GSK3 en el desarrollo de varios tipos de cáncer.

Esencialmente, el cáncer es una enfermedad que se caracteriza por la pérdida por parte de las células de su capacidad de regular la división, la diferenciación y la apoptosis, lo que produce un desarrollo anormal de masas celulares que poco a poco sustituyen el parénquima tisular y cuando afecta a órganos vitales y conduce a la muerte. Dado que la capacidad de las células para regular su crecimiento depende de la actividad de vías de transducción de señales, como la vía Wnt/ β -catenina y la vía fosfoinositol 3-quinasa o fosfoinositida-3-quinasa (PI3K) de las cuales, la actividad de GSK3 es un regulador negativo, se pensó que su función era la de reprimir el desarrollo celular. No obstante, muchos estudios apoyan la noción opuesta, ya que se ha informado que la inhibición de la GSK3 con moléculas químicas exhibe actividades antiproliferativas y propapoptóticas en diversos tipos de cáncer, como se ha examinado en colon, hígado y otros órganos.³⁸ Además, GSK3 participa también como un regulador de la apoptosis al controlar la función de la proteína p53 (conocida como el guardián del genoma), y regula tanto la vía intrínseca como extrínseca de la apoptosis al controlar factores clave como el modulador de apoptosis regulado positivamente por p53 (PUMA) Noxa y BAX, las cuales pertenecen a la subfamilia BNIP3. Estas proteínas forman complejos de homodimerización estables que se localizan en la membrana externa de las mitocondrias después del estrés celular.³⁹

La aparición y desarrollo de cáncer en diferentes tejidos depende de distintos mecanismos intrínsecos a cada tejido, por ejemplo, el cáncer cérvico uterino se asocia a infecciones por el virus del papiloma humano que al infectar células del epitelio intestinal introduce como parte de su mecanismo de replicación dos oncogenes que sirven para expresar las proteínas E6 y E7, esto conduce a la pérdida de función de 2 proteínas, es decir que la proteína p53 (Sustrato de GSK3) y la proteína del retinoblastoma que regula positivamente la apoptosis, no pueden suprimir los tumores y regula negativamente el ciclo celular dando como resultado finalmente la proliferación acelerada de las células epiteliales, en este sentido cáncer.²² Por su parte el cáncer de colon se produce principalmente por la presencia de mutaciones somáticas del tejido que promueven la falta de función de la proteína supresora tumoral Adenomatous polyposis coli (APC) lo que provoca un exceso de función de la vía de transducción Wnt/ β -catenina (sustrato de GSK3), esto resulta en un desarrollo celular anormal dado que la actividad de esta vía promueve la replicación celular.²² En este sentido, GSK3 puede regular también el desarrollo de diferentes tipos de células cancerosas mediante diferentes mecanismos, no obstante, se ha reconocido la función de GSK3 como un requisito importante en diversos tipos de cáncer para permitir la supervivencia y desarrollo celular como el cáncer de pulmón, la leucemia mieloide aguda (LMA), el glioma, el adenocarcinoma del conducto pancreático, las células cancerosas de la próstata y del epitelio cérvico uterino.¹⁸ Un claro ejemplo de la participación de GSK3 como promotor de desarrollo celular anormal proviene de la caracterización de diversas líneas celulares de leucemia mieloide. Inicialmente para conocer que proteínas se podrían encontrar relacionadas al desarrollo de esta enfermedad se encontró que la expresión y actividad de GSK3 está incrementada en las líneas celulares de leucemia múltiple U-266, OPM-2, RPMI8226, e INA-6, por otro lado, un segundo reporte independiente muestra que las líneas de leucemia mieloide HL-60, U937, THP-1, y NB4, muestran el mismo comportamiento.⁹ Aunque esta relación no es una prueba de que el incremento de función o expresión de la enzima es la causa de la enfermedad, esta evidencia se ha establecido mediante la inhibición de la expresión de GSK3 en progenitores mieloides mediante una estrategia que permite eliminar la expresión del gen únicamente en la médula ósea. Este estudio reportó que es suficiente en este tipo celular con la pérdida de función

de GSK3 para provocar diferenciación de las células en la médula ósea.⁴⁰ En otro panorama, se ha analizado la función de más de 1000 proteínas en las células de leucemia mieloide aguda HL-60 mediante la inhibición de expresión de una proteína a la vez, se reportó que la represión de GSK3 fue el ensayo que más significativamente provocó el arresto del ciclo celular y provocó diferenciación de las células LMA. Para confirmar este hallazgo, en el mismo estudio se llevó a cabo el análisis de una biblioteca de más de 3000 compuestos para descubrir cuáles de estos podrían provocar arresto del ciclo celular o incluso apoptosis de las células LMA. De manera interesante, los cinco compuestos que tuvieron el mayor efecto como represores del ciclo celular son moléculas que funcionan como inhibidores de GSK3.⁹ De esta manera se identificó a GSK3 como un regulador central de la proliferación de células de LMA y definen su potencial como diana terapéutica contra esta enfermedad.

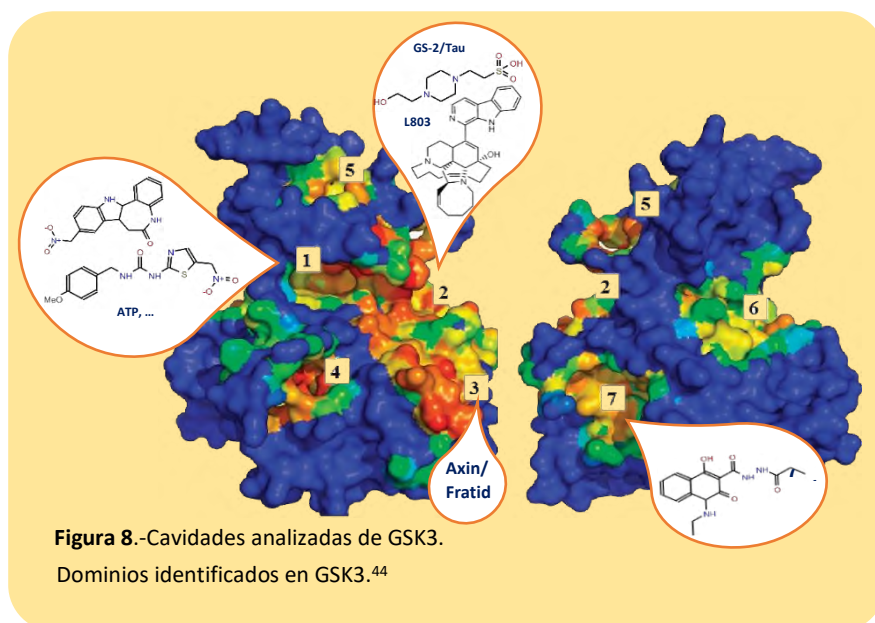
En resumen, la inhibición de GSK3 es potencialmente una estrategia prometedora como tratamiento contra diversas enfermedades sin limitarse a las citadas en este trabajo ya que, además, se ha identificado su importancia en el desarrollo de diabetes, padecimientos cardíacos, enfermedades autoinmunes, entre otras.

1.5. Estrategias para inhibir la función de GSK3.

La función de GSK3 se ha analizado a través de inhibir su función cinasa por diferentes métodos, por ejemplo, mediante la generación de modelos transgénicos, principalmente ratones, a los que se les producen mutaciones con el objetivo de eliminar la expresión de la proteína o mediante la introducción de moléculas de ARN “anti-sentido” (llamado ARN de silenciamiento) también para reducir la expresión de la proteína en modelos celulares. Sin embargo, aunque estas estrategias han generado significativamente parte de la información de GSK3 que se conoce hasta la fecha, y son las metodologías que permiten la interrupción de la función de GSK3 sin afectar a ninguna otra proteína,¹⁸ No obstante, no parecen ser factibles para generar una terapia que tenga como blanco a GSK3 en el tratamiento de enfermedades debido a que los mecanismos para controlar la introducción de mutaciones genéticas aún no se han

perfeccionado.⁴¹ Como consecuencia, debido al potencial benéfico descubierto para la inhibición de GSK3, diversas compañías farmacéuticas, se han dado a la tarea de investigar la generación de moléculas con este potencial. De acuerdo con la base de datos ChemBL,⁴² existen más de 2000 moléculas con actividad inhibitoria sobre GSK3,⁴³ entre las que se encuentran moléculas de tipo peptídico, sales metálicas, flavonoides de plantas, moléculas heterocíclicas e incluso moléculas aisladas de esponjas marinas,⁴³ sin embargo, a pesar de que estas moléculas demuestran ser efectivas en ensayos *in vitro*, muchas de ellas fallan en modelos animales debido a su baja solubilidad o escasa permeabilidad celular.⁴⁴ Aunque estas moléculas pueden actuar mediante su acoplamiento a distintos dominios de la proteína, el dominio de unión a Adenosín trifosfato (ATP) es el más estudiado y contra el que más inhibidores sintéticos se han diseñado.⁴⁵ Un ejemplo de este tipo de inhibidor (ATP-competitivo) el cual se ha utilizado para revertir la inflamación cerebral y para activar la redes neuronales, en estudios preclínicos del síndrome de Rett, es el inhibidor SB216763, (3- (2,4-Diclorofenil) -4- (1-metil-1H-indol-3-il) -1H-pirrol-2,5-diona) que puede inhibir la actividad de GSK3 con una constante de inhibición (KI) de 9.10 nM.⁴⁴ Por su parte el péptido L803-mts (un péptido es molécula formada por la unión de varios aminoácidos mediante enlaces peptídicos) inhibe la fosforilación de GSK3 sobre β -catenina a través de la interrupción de la interacción de GSK3 con la proteína axina, con lo que se impide la formación de un complejo multi proteico encargado de fosforilar a β -catenina para que sea degradada. De manera interesante, el mecanismo de las sales de litio que se han empleado para el tratamiento de los desórdenes bipolares durante más de 50 años se mantuvo desconocido durante un largo periodo de tiempo; sin embargo, en la década de los 90s se reportó que un inhibidor no selectivo de GSK3 cuyo mecanismo de inhibición involucra la sustitución del cofactor iónico Mg^{2+} . En la Figura 8 observamos todas las cavidades estudiadas de la proteína GSK3 numeradas de 1 al 7 y con globos que contienen algunos inhibidores que fueron utilizados y otros mencionados como L803-mts que es un inhibidor competitivo. El cuadro marcado con 1 es el principal sitio de unión del ATP mientras que los otros son solo sitios de unión indicados con GS-2/TAU y Axin/Fratid. Estas siete cavidades fueron analizadas y estudiadas con 25 estructuras con formato PDB, aunque existen más formatos que pueden utilizarse para este estudio, para esta investigación donde se

realizaron estudios de sus cavidades se reportó que el formato utilizado fue PDB. De forma independiente cabe mencionar que no todos los inhibidores son competitivos del ATP, de las 7 cavidades estudiadas sólo tres de ellas (las cavidades 1, 2 y 3) son sitios de unión ATP conocidos, y las cuatro cavidades restantes pueden interactuar diferentes ligandos.



Aprovechando esta diversidad de mecanismos, varias empresas farmacéuticas inspiradas en los hallazgos bioquímicos de GSK3 se han dedicado a la búsqueda de moléculas inhibitoras como tratamiento de varias enfermedades, no obstante, ningún inhibidor específico de GSK3 es usado en la clínica y solo unos pocos han alcanzado la etapa de ensayo clínico. Esta dificultad radica en la cantidad de sustratos que tiene la enzima y que al inhibirse provoca múltiples efectos nocivos lo que ha limitado la visión de GSK3 como blanco farmacológico.⁴ Por ejemplo, aunque la inhibición de GSK3 es benéfica en los contextos de enfermedad que ya se mencionaron (sección 1.3), en otros órganos puede ser perjudicial. Una serie de análisis llevados a cabo en perros indican que el uso de dos fármacos de manera independiente como inhibidores de GSK3, producen alteraciones histológicas en la vesícula y los conductos biliares que evolucionan hasta provocar colecistitis crónica.⁴⁶ Otro estudio independiente hecho en ratones ha reportado que la inhibición de GSK3 mediante el uso de

fármacos provoca en la médula ósea la hiperproliferación de progenitores mieloides lo que conduce a la aparición de síndromes mielodisplásicos.⁴⁷

No obstante, aunque la terapia dirigida para inhibir a GSK3 no ha sido efectiva a la fecha, un aspecto poco estudiado de esta enzima es su función en mamíferos es llevada a cabo por dos proteínas parálogas llamadas GSK3 α y GSK3 β las cuales muestran redundancia funcional como actividades específicas.²²

1.6. Los parálogos de la proteína GSK3.

Como se mencionó previamente, el primer miembro de la familia GSK3 se identificó en extractos de proteínas de músculo esquelético de conejo como parte de una vía metabólica capaz de fosforilar a la proteína glucógeno sintasa (Sección 1.1), más de una década después de su identificación, se aisló e identificó una proteína con una actividad cinasa similar y una identidad de secuencia de aminoácidos altamente conservada también del músculo esquelético de conejo y debido a que las dos proteínas están codificadas en genes distintos se definieron como dos parálogos llamados GSK3 α y GSK3 β que comparten sustratos diversos como la GS, c-jun y β -catenina.⁹ A menudo denominadas isoformas, en humanos estas proteínas parálogas están codificadas en los genes *GSK3A* ubicados en el cromosoma 19q13.2 que codifica un polipéptido de 51 kDa de 483 aminoácidos y *GSK3B* que codifica un polipéptido de 47 kDa de 420 aminoácidos y está ubicado en el cromosoma 3q13.33. Las proteínas comparten ~98% de identidad de aminoácidos dentro de su dominio catalítico, pero solo 36% en los 76 residuos de aminoácidos C-terminales.⁴⁸ GSK3 α también posee una extensión de Glicinas en el extremo N-*terminal* adicional que falta en GSK3 β y está relacionada con formaciones complejas de proteínas y localización subcelular.⁹ Dentro del dominio catalítico, la diferencia entre las dos isoformas es la presencia de un residuo de ácido aspártico en la posición 133 en GSK3 β equivalente a un ácido glutámico en la posición 196 en GSK3 α .⁹ Se observa en la Figura 9 la comparación de los parálogos de GSK3, en la figura señalado con un círculo en amarillo el aminoácido diferencial dentro de sus sitios catalíticos o de inhibición y, adicionalmente se incluye la secuencia de aminoácidos en la parte inferior de la imagen (en verde los aminoácidos emparentados, en azul los aminoácidos

relevantes para su inhibición, de color blanco son aquellos aminoácidos por los cuales se diferencian las proteínas, en rojo con azul los aminoácidos diferenciales del sitio activo).

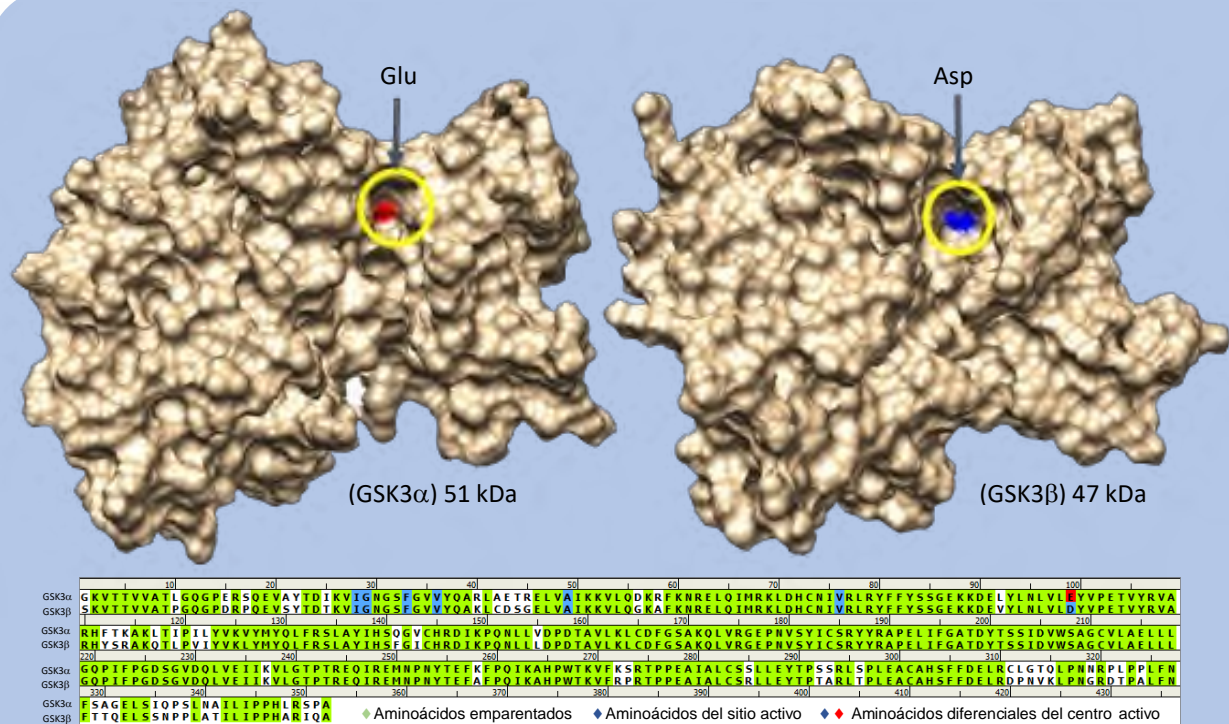


Figura 9.-Parálogos de GSK3.

Comparación de la secuencia de aminoácidos, Edición propia, software Photoshop y Chimera).

1.7. El paradigma de la inhibición de la proteína GSK3.

Aun cuando los parálogos de la proteína GSK3 comparten virtualmente los mismos sustratos, no son fisiológicamente equivalentes. En definitiva, queda de manifiesto por experimentos llevados a cabo en ratones en los que la delección de la proteína GSK3 α resulta en un fenotipo que muestra sensibilización a la insulina y cambios conductuales, mientras que la delección de GSK3 β resulta en la muerte embrionaria.⁴⁹ Además, se ha reportado que la expresión de GSK3 β de origen murino en células de la mosca de fruta mutantes carente de la proteína Shaggy (homóloga de GSK3), es capaz de rescatar la función de esta proteína y permitir el desarrollo de los embriones mientras que la expresión de GSK3 α es incapaz de suplir esta función. Lo que muestra la importancia que GSK3 β tiene en la fisiología celular.⁵⁰ Sin embargo, otras funciones de los parálogos de GSK3

si muestran redundancia como es el caso de la regulación negativa de la vía Wnt. Se ha reportado que la delección de los parálogos de GSK3 por separado no promueve la hiper-activación de la vía Wnt y para que se produzca su activación, se requiere que ambos parálogos estén ausentes o se reduzca su actividad mediante moléculas inhibitoras.⁵⁰ Es en parte por este mecanismo que se considera que la terapia de GSK3 como blanco farmacológico ha fallado sistemáticamente, sin embargo, debido a que los parálogos son redundantes en este contexto, se ha propuesto que la inhibición de solo un parálogo en un momento dado, no conlleva al desarrollo de los efectos nocivos que son llevados por los inhibidores que son incapaces de distinguir entre los parálogos.⁵¹

1.8. Inhibición selectiva del parálogo GSK3 β .

Se ha reportado que la inhibición de esta enzima provoca la diferenciación de células de la leucemia mieloide aguda e induce la apoptosis de células de cáncer pulmonar, hepático, cerebral o de mama.⁵² Se ha demostrado que favorece la supervivencia de ratones modelo de accidentes cardíacos y reduce los síntomas y desarrollo de depósitos fibrilares de la proteína tau y del péptido β -amiloide, lo que resulta en alargamiento del tiempo de aparición de demencia y síntomas de neurodegeneración en varios modelos experimentales de la enfermedad de Alzheimer en ratones.⁵³ Además, reduce la inflamación de tejidos expuestos a patógenos, lo que potencia la capacidad del sistema inmune para controlar y resolver estímulos lesivos. En lo que respecta a la diabetes mellitus tipo 2 (DM-II), la inhibición de la proteína GSK3 β favorece la captación de glucosa y sensibiliza las células hepáticas y del músculo esquelético ante la insulina para sintetizar glucógeno, disminuyendo de esta manera los efectos nocivos del exceso de glucosa en el torrente sanguíneo.⁵⁴ No obstante, y a pesar de los potenciales beneficios que resultan de la inhibición de la enzima GSK3 β , a la fecha no se ha podido llevar a la práctica clínica debido a que no se conocen a detalle los mecanismos celulares que son afectados por la inhibición de cada parálogo de GSK3 por separado. Además, no se han conducido investigaciones que describan el uso de inhibidores químicos específicos o selectivos de cada proteína ya que la mayor parte de la información que se ha generado de la función de cada parálogo se ha obtenido mediante la aplicación de técnicas de

edición genética en cultivos celulares o modelos animales. La proteína GSK3 también lleva a cabo funciones de andamiaje en la formación de complejos supra proteínicos como parte de sus mecanismos de regulación en diversas funciones celulares. Esto significa que las estrategias de edición genética podrían pasar por alto las funciones no enzimáticas de estas proteínas y explicar las diferencias encontradas en la regulación celular en comparación con la inhibición química que se caracteriza por la afectación a las funciones de fosfotransferencia de los parálogos de GSK3, pero no de sus funciones de andamiaje.⁹

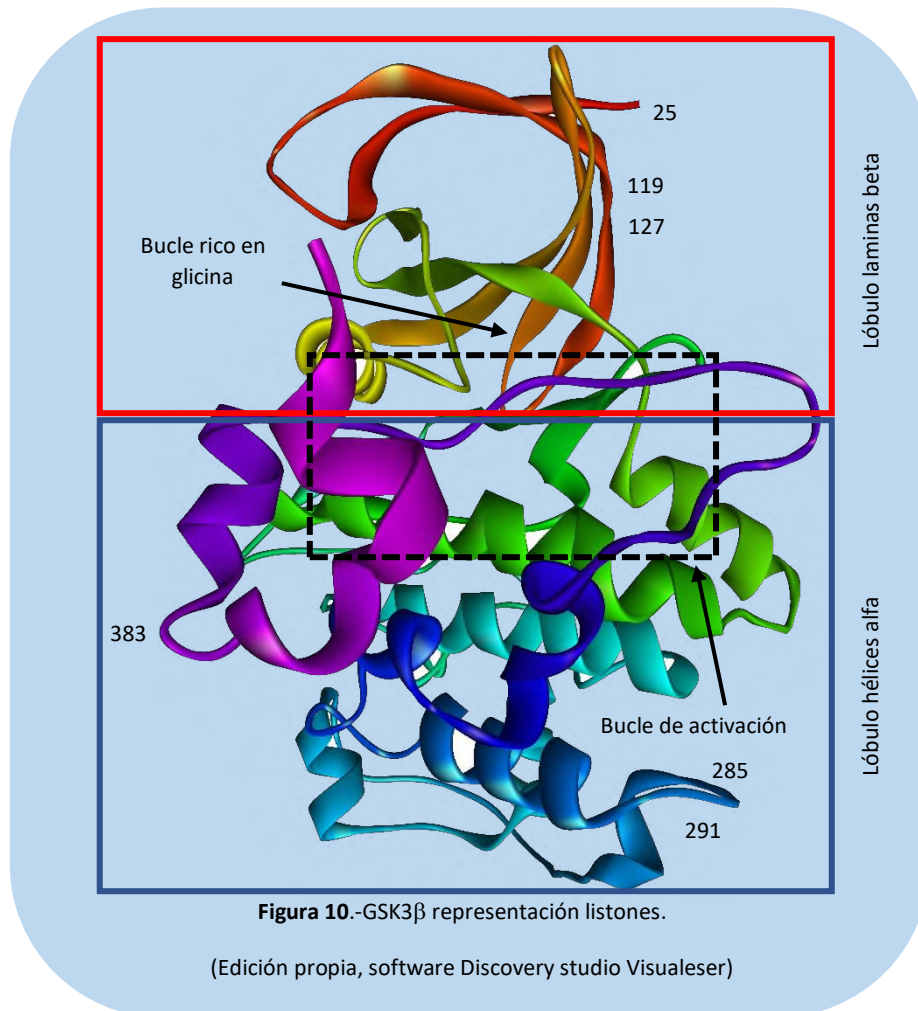
Por otra parte, debido a la regulación negativa que los parálogos de GSK3 ejercen sobre la vía de regulación en la proliferación celular Wnt/ β -catenina, es fácil predecir que la represión de la actividad cinasa simultánea de ambos parálogos conlleva a la estimulación de esta vía y al desarrollo celular anormal. Sin embargo, se sabe que la actividad de un solo parólogo de la enzima GSK3, α o β , es suficiente para reprimir este efecto y restaurar la represión sobre la vía.⁵⁵

Por lo tanto, el efecto terapéutico reportado como resultado de la inhibición de GSK3 se podría llevar a cabo siempre y cuando se logre la inhibición selectiva del parólogo GSK3 β . De manera interesante se han reportado diversos inhibidores químicos que presentan preferencia por una u otra de estas proteínas, sin embargo, no se han caracterizado las singularidades estructurales que determinan su selectividad para este efecto.⁵⁵

1.9. Características cristalográficas de GSK3 β .

Estudios cristalográficos han mostrado que la estructura tridimensional de la GSK3 β es común a la mayoría de las cinasas, con un lóbulo N-*terminal* pequeño (residuos 25-138) constituido principalmente por láminas- β , y un lóbulo C-*terminal* grande (residuos 344-382) formado esencialmente por hélices- α . El sitio de unión con el Adenosín Trifosfato (ATP) se encuentra entre los dos lóbulos (ver Figura 10) el cual suele ser conservado entre las diferentes cinasas, bordeado por un bucle rico en glicina cabe mencionar que esta zona se le conoce como región bisagra. El bucle de activación (residuos 200-226) se encuentra en la superficie del lugar de unión al sustrato. Esta enzima posee dos lugares de

fosforilación que presentan influencia en su actividad catalítica: Ser9, con efecto inhibitorio, y Tyr216, localizado en el bucle de activación, que incrementa la actividad catalítica.⁵⁶



1.10. Mecanismo de fosforilación de GSK3 β .

El mecanismo de fosforilación de GSK3 β es un proceso que lleva a cabo esta proteína, su actividad cinasa consiste en transferir grupos fosfatos a sus sustratos. Esta actividad requiere del ATP, con ayuda de cofactores (iones metálicos como Mg²⁺) para promover la catálisis de la reacción bioquímica, la principal función de este cofactor es desestabilizar la molécula del ATP. Todo este proceso inicia cuando la proteína GSK3 β toma una molécula de ATP y esta se posiciona en el “surco” ubicado en el centro de la proteína formada por los lóbulos de hélices alfa y láminas beta, lo cual permite al bucle rico en glicinas que se posiciona en la parte superior de la molécula de ATP asegurándola “como un cinturón” por glicinas Gly63, Gly65 y Gly68 (ver Figura 11), luego existe un residuo de lisina (Lys183) que se une a los fosfatos alfa y beta del ATP, manteniéndolo en su lugar. Un ácido glutámico (Glu97) forma un puente de sal con esa lisina para mantener la forma de la cinasa estable. En otro lado de la proteína tenemos un dominio formado por tres aminoácidos el ácido glutámico, fenilalanina y glicina (dominio DFG) que une los iones metálicos que necesita para transferir el fosfato, también tenemos un bucle de activación el cual determina si la cinasa reconoce tirosinas o serinas, (bucle G) que está formada por una secuencia de aminoácidos Glicina, Asparagina, Glicina, Serina, Fenilalanina y Glicina (GNGSFG) y por último tenemos el dominio (HRD) formado por los aminoácidos Histidina179, Arginina180 y Ácido Aspartico181, éste último es un aminoácido particular pues tiene la función de transferir el fosfato gamma liberado del ATP a un sustrato Serina o Treonina.⁵⁷ En relación con lo anterior observamos en la Figura 11, la cual contiene los principales dominios que permiten la función de fosforilación de la proteína GSK3 β a partir de la molécula ATP indicados con los diferentes colores así como los respectivos nombres mencionados anteriormente. Mientras que en la Figura 12 se presenta la ubicación de estos dominios dentro de la proteína GSK3, en la cual se indica el sitio catalítico, es decir, el lugar donde la molécula del ATP será posicionada por la proteína para poder liberar los fosfatos.

GSK3 β y dominios en el proceso de la fosforilación

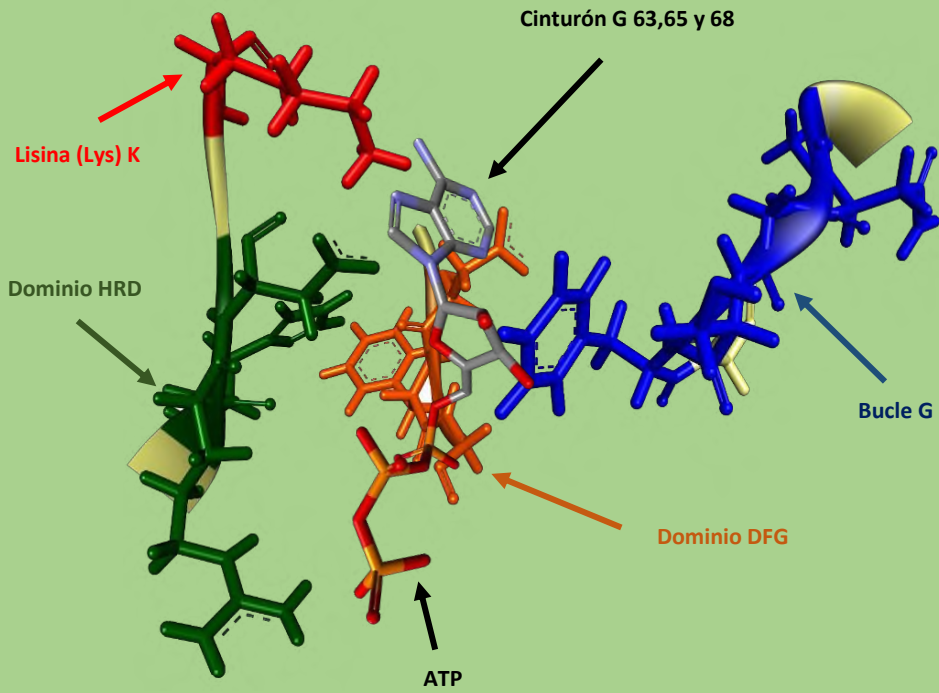


Figura 12.-GSK3 β proceso de fosforilación.

(Edición propia, software Discovery studio Visualeser)

GSK3 β dominios encargados de la fosforilación

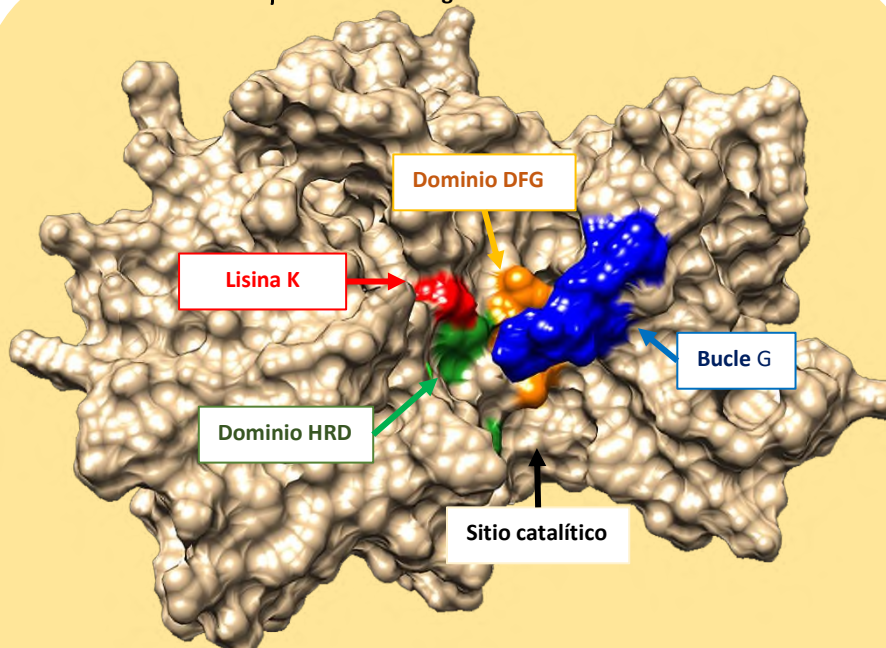


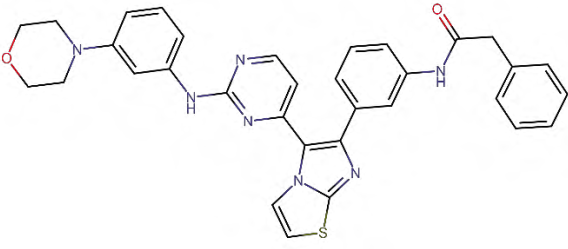
Figura 11.-Gsk3 β , dominios de fosforilación.

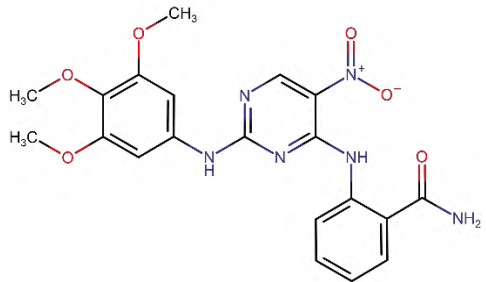
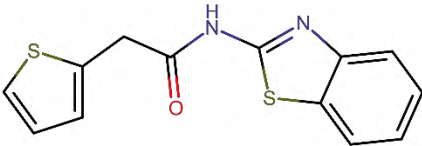
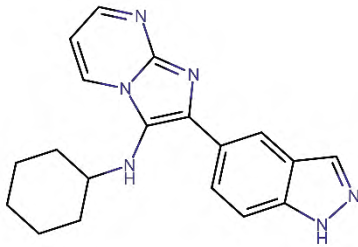
(Edición propia, software Chimera)

1.11. Factibilidad de inhibidores selectivos de la proteína GSK3 β .

A pesar de que la enzima GSK3 β ha sido ampliamente estudiada, la idea de generar inhibidores selectivos de este parólogo representa un desafío debido a la similitud con GSK3 α (Sección 1.6). No obstante, una profunda revisión en diferentes bases de datos ha revelado que existen moléculas con selectividad inhibitoria por GSK3 β sobre GSK3 α , Tabla 1, en ella se muestra algunos de los ejemplos tomados de la base de datos ChEMBL, aunque estas no se han caracterizado en la búsqueda de firmas químicas que permita su análisis. En este sentido cabe mencionar que no hay reportes de estudios que indiquen las diferencias principales entre las moléculas con preferencia a la proteína GSK3 α y las que tienen selectividad por la proteína GSK3 β , en lo que se basan principalmente son a las constantes de inhibición media IC₅₀, constante de inhibición (K_i) entre otras.

Tabla 1.-Moléculas inhibidoras con preferencia por GSK3 β .

Inhibidores preferentes GSK3 β	Moléculas
<p>Código de reconocimiento ChEMBL1090360</p> <p>Constante de inhibición (K_i) Unidades Nanomolar (nM) GSK3α= 3981.07 nM GSK3β=794.33 nM</p> <p>Nombre IUPAC= N - [3- [5- [2- (3- morfolin-4-ilanilino) pirimidin-4-il] imidazo [2,1-b] [1,3] tiazol-6-il] fenil] - 2- fenilacetamida</p>	

<p>CHE Código de reconocimiento CHEMBL1241473</p> <p>Constante de inhibición (Ki) Unidades Nanomolar (nM) GSK3 α = 3,162 nm GSK3 β = 6.31 nm</p> <p>Nombre IUPAC= 2 - [[5-nitro-2-(3,4,5-trimetoxianilino) pirimidin-4-il] amino] benzamida</p>	
<p>Código de reconocimiento CHEMBL1421720</p> <p>Constante de inhibición (Ki) Unidades Nanomolar (nM) GSK3 α = 3981.07 nm GSK3 β = 501.19 nm</p> <p>Nombre IUPAC= N- (1,3-benzotiazol-2-il) -2-tiofen-2-ilacetamida</p>	
<p>Código de reconocimiento CHEMBL1682552</p> <p>Constante de inhibición (Ki) Unidades Nanomolar (nM) GSK3 α = 63.1 nm GSK3 β = 10 nm</p> <p>Nombre IUPAC= n-ciclohexil-2- (1h-indazol-5-il) imidazo [1,2-a] pirimidin-3-amina</p>	

Capítulo 2.- Química computacional.

2.1. ¿La química computacional es una alternativa?

Una de las dificultades que representa el análisis de los inhibidores, es diferenciar las características que permitan la inhibición selectiva de un parálogo. Por lo anterior, determinar estas actividades por medio de pruebas biológicas o físicas, puede ser difícil o inclusive puede demorarse varios años de investigación. Una alternativa que es bien aceptada por muchos investigadores y que ha cobrado fuerza en tiempos actuales, es la química computacional, la cual es una rama de la química que utiliza teorías, conceptos y modelos de la química teórica, basados en tratamientos físicos de la materia provenientes de la física clásica, cuántica y la mecánica estadística, incorporados en softwares científicos especialmente diseñado para calcular la estructura y/o las propiedades estáticas, dinámicas de moléculas y agregados moleculares en estado gaseoso, en solución y de cuerpos sólidos. Mientras sus resultados complementan la información que puede obtenerse en experimentos químicos, pueden, en otros casos, predecir fenómenos químicos no observados a la fecha, orientando el diseño de nueva actividad experimental o sustituyendo la ausencia de otro conocimiento empírico en problemas donde el diseño experimental tiene asociado un alto costo económico y/o el experimento empírico resulta impracticable en los términos actuales. La química computacional es ampliamente utilizada desde hace varias décadas en el diseño de nuevos medicamentos y materiales.

Podríamos destacar que la química computacional ha cobrado fuerza por las aportaciones que ha logrado, algunos ejemplos, son las moléculas que se han descubierto para el tratamiento de diferentes tipos de patologías tales como el cáncer, la diabetes, el Alzheimer y/u otras enfermedades de carácter crónico, de esta forma con ayuda de la química computacional, podemos incluso indicar que se han conseguido crear inhibidores enzimáticos competitivos frente a diferentes sustrato, de los cuales existen bibliotecas virtuales con más de 20,000 moléculas. Algunas de estas bases de datos son ChEMBL, Genbank, Pubchem entre otros. Los inhibidores de acetilcolinesterasa utilizados para el tratamiento del Alzheimer son un claro ejemplo de porque la química computacional es una alternativa para el descubrimiento de nuevas moléculas con actividad biológica.⁵⁸

Por lo cual esta investigación fue realizada con el apoyo de la química computacional debido a su reproducibilidad y por las grandes aproximaciones que tiene, los resultados pueden e inclusive son similares a los obtenidos en un laboratorio y por el gran aporte que representó en esta investigación, esta alternativa proporcionó información relevante.

2.2. Química computacional.

La química computacional es una disciplina relativamente joven, dado que antes de que las computadoras fueran de acceso general, los cálculos teóricos eran resueltos por aproximaciones analíticas, las cuales carecían de precisión para ser predicciones fiables. Fue hasta inicios de 1960 cuando las computadoras estuvieron disponibles para la comunidad, gracias al aumento y desarrollo de tecnología en el área informática, la química teórica ha evolucionado a un aspecto computacional, de ahí el nombre de química computacional.⁵⁹ La química computacional consiste en la predicción de las propiedades de sistemas atómicos y moleculares mediante cálculos complejos con ordenador basados en las leyes fundamentales de la mecánica cuántica.⁶⁰ Esta metodología se centra en la aplicación de cálculo para resolver ecuaciones y predecir propiedades a nivel molecular, por lo que es utilizada como herramienta de aprendizaje y de respaldo en investigaciones experimentales. El objetivo principal de la química computacional es modelar todo tipo de propiedades moleculares de sistemas químicos utilizando la fisicoquímica, la física molecular y la física cuántica, empleando una gran variedad de técnicas teóricas en constante desarrollo.⁵⁹ Además, se encarga de investigar la estructura de la molécula, detallando sus propiedades intrínsecas, llamados descriptores. Las cualidades principales de estos estudios son la interpretación, caracterización y predicción de las propiedades geométricas como son la distancia y ángulos de enlace, funciones de onda como los orbitales naturales, coeficientes, mapas de contorno de orbitales moleculares, y energías como de potencial de ionización, afinidad electrónica, energía electrónica y energía de interacción entre otras características estructurales.⁶¹

Para poder realizar los cálculos es necesario utilizar teorías que se basan en leyes de la mecánica cuántica, los principales métodos para calcular la estructura electrónica se menciona a continuación.⁶²

***Los métodos semiempíricos**, estos métodos son químico-cuánticos, es decir, obtienen la energía de los orbitales moleculares a partir del uso de datos experimentales en la solución de la ecuación de Schrödinger.

Los métodos *ab initio, El término *Ab initio* proviene del latín y significa “desde el principio”. Se da este nombre a los métodos computacionales desarrollados de principios teóricos (tales como la ecuación de Schrödinger), sin incluir ningún tipo de información experimental. Las aproximaciones usadas son usualmente matemáticas.

$$\hat{H}\Psi = E\Psi$$

Donde E es la energía del sistema y Ψ es la función de onda, \hat{H} es el operador hamiltoniano que está definido por:

$$\hat{H} = -\sum_{i=1}^n \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{k=1}^n \frac{\hbar^2}{2m_k} \nabla_k^2 + \sum_{k<l}^n \frac{Z_k Z_l e^2}{r_{kl}} - \sum_k^N \sum_{i=1}^n \frac{Z_k e^2}{r_{ki}} + \sum_{i<j}^n \frac{e^2}{r_{ij}}$$

Siendo ∇^2 el operador Laplaciano $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$; los subíndices i y j corresponden a los electrones, los subíndices k y l corresponden a los núcleos, m_e es la masa del electrón, e es la carga del electrón, m_k corresponde a la masa del núcleo k-ésimo, Z_k y Z_l son los números atómicos de los diferentes núcleos presentes en la molécula, r es la distancia entre las diferentes partículas y \hbar es la constante de Planck reducida o constante de Dirac, la cual se define como:

$$\hbar = \frac{h}{2\pi}$$

Las aproximaciones matemáticas usadas en la solución de la ecuación de Schrödinger se puede catalogar en dos tipos, las que se aplican a la función de onda y las que se aplican al operador, en donde la función de onda introduce funciones base para expandir los orbitales moleculares de acuerdo con el tipo de sistema molecular de estudio, asimismo el del operador busca utilizar formas más simples de resolver una ecuación diferencial. El cálculo más utilizado de los métodos *ab initio* es el cálculo Hartree-Fock (HF). El método de HF es una forma aproximada de las ecuaciones de mecánica cuántica para fermiones, utilizada

en física, química y, es el equivalente a la aproximación de orbitales moleculares. Este esquema de cálculo es un procedimiento iterativo para calcular la mejor solución a la ecuación de Schrödinger independiente del tiempo, para moléculas aisladas, tanto en su estado fundamental como en estados excitados. La interacción de un único electrón en un problema de muchos cuerpos con el resto de los electrones del sistema se aproxima promediándolo como una interacción entre dos cuerpos (tras aplicar la aproximación de Born-Oppenheimer). De esta forma, se puede obtener una aproximación a la energía total de la molécula. Como consecuencia, calcula la energía de intercambio de forma exacta.⁶³

*La teoría del funcional de la densidad, (o DFT, por sus siglas en inglés, *Density Functional Theory*), se basa en la determinación de la energía de un estado electrónico a partir de la densidad electrónica. Consiste en la introducción de la correlación electrónica usando funcionales, tomando en cuenta los teoremas de Hohenberg y Khon (1964). Es uno de los métodos más utilizados en los cálculos cuánticos de la estructura electrónica de la materia, tanto en la física de la materia condensada como en la química cuántica.⁶⁴

La DFT aplicada a sistemas electrónicos, es un procedimiento de variaciones alternativo a la solución de la ecuación de Schrödinger, donde el funcional de la energía electrónica es minimizado con respecto a la densidad electrónica. Es uno de los métodos más utilizados en los cálculos cuánticos de la estructura electrónica de la materia. Los métodos tradicionales dentro de las teorías de la estructura electrónica de la materia, en particular la teoría de Hartree-Fock y los derivados de este formalismo, se basan en una función de onda multi electrónica. Si bien esta resolución de la ecuación de Schrödinger permite describir de forma exacta el comportamiento de los sistemas muy pequeños, su capacidad de predicción se ve limitada por el hecho de que sus ecuaciones son demasiado complejas de resolver numéricamente y analíticamente.⁶⁴

La DFT reformula el problema para ser capaz de obtener, por ejemplo, la energía y la distribución electrónica del estado fundamental, trabajando con el funcional de la densidad electrónica en vez de la función de ondas. Una ventaja es que la densidad es una magnitud mucho más simple que la función de ondas y por lo

tanto más fácil de calcular y en la práctica son accesibles en sistemas mucho más complejos: la función de ondas de un sistema de N-electrones depende de $3N$ -variables, mientras que la densidad electrónica sólo depende de 3 variables. Una desventaja es que, salvo los casos más simples, no se conoce de manera exacta el funcional que relaciona esta densidad con la energía del sistema. En la práctica, se usan funcionales que se han comprobado que dan buenos resultados. Originalmente, la DFT se desarrolló en el marco de la teoría cuántica no relativista (ecuación de Schrödinger independiente del tiempo) y de la aproximación de Born-Oppenheimer. La teoría fue extendida posteriormente al dominio de la mecánica cuántica dependiente del tiempo, y se habla de la TD-DFT o Teoría del Funcional de la Densidad Dependiente del Tiempo y del dominio relativista. Entre otras cosas, esto permite calcular estados excitados como, por ejemplo: cuerpos que consisten en moléculas, macromoléculas, nanopartículas o micropartículas, protones o partículas cargadas en movimiento.⁶⁵

2.3. Modelado molecular.

El modelado molecular presenta mayor relevancia en los proyectos de investigación desde la entrega del Premio Nobel de Química en 1998 a Walter Kohn por su desarrollo de la "Teoría de la densidad funcional" a John A. Pople por su desarrollo de "métodos computacionales en la química cuántica". La aplicación del modelado molecular en química fue reconocida en 2013 al ser otorgado un nuevo Premio Nobel, esta vez a Martin Karplus, Michael Levitt y Arieh Warshel por el desarrollo de "modelos multiescala de sistemas químicos complejos".

El modelado molecular avanzado es una herramienta sobre todo predictiva, que permite acercarse a la realidad y a la posibilidad de establecer si una reacción va a ocurrir o no; diseñar nuevos materiales o fármacos; conocer cómo responden determinadas proteínas a contaminantes o a fármacos; determinar las interacciones fármaco-receptor, entre otras. Es una herramienta de gran utilidad, pues permite interpretar de manera racional, el proceso de interacción entre una molécula y su diana, y proporciona datos fiables que permiten predecir

el comportamiento del sistema. Actualmente, estos modelos tienen tal poder predictivo que se pueden hacer experimentos de química en ordenador llegándose a comparar con los de un laboratorio convencional, algunos ejemplos de lo anterior expuesto pueden ser los análisis de toxicidad, la constante de inhibición, reacciones químicas y otros, lo que permite complementar la parte experimental al fin de disminuir los tiempos y costos de inversión. El modelado molecular es la herramienta que permite aproximar las estructuras obtenidas por computadora con las observaciones experimentales, puede definirse como una descripción simplificada o idealizada de un sistema molecular o proceso entre moléculas, ideada para facilitar cálculos y predicciones.⁶⁶

2.4. Dinámica molecular.

La dinámica molecular (MD, Molecular Dynamics) es una simulación molecular que permite analizar el comportamiento de un sistema (físico, químico o biológico) a través del tiempo. Es un método utilizado para generar las trayectorias de un sistema compuesto de n-partículas por integración numérica directa de las ecuaciones de movimiento de Newton, con especificaciones de un potencial de interacción interatómico de condiciones iniciales y de frontera adecuadas. Constituye un método de modelado y simulación a nivel atómico.⁶⁷

Por medio de MD, se pueden calcular diferentes propiedades fisicoquímicas del sistema como la energía libre, entropía, solubilidad, viscosidad, presión, temperaturas, y en sistemas biológicos, permite medir la fuerza de interacción entre posibles fármacos y sus dianas biomoleculares o receptores, e incluso, describir el comportamiento de una proteína y moléculas complejas bajo ciertas condiciones, por mencionar algunas de sus capacidades. Este tipo de estudios han mostrado buena correlación con resultados experimentales, y se pueden realizar con equipos de cómputo ciertamente convencionales, pero que cumplan ciertas especificaciones técnicas.⁶⁷

2.5. Relación cuantitativa estructura actividad (QSAR).

La relación cuantitativa estructura- actividad es el proceso por el cual la estructura química se correlaciona cuantitativamente con un proceso bien definido, como la actividad biológica (unión fármaco-receptor) o la reactividad química (afinidad de una sustancia por otra para producir una reacción), es decir, la suposición es que moléculas semejantes tendrán una actividad similar. Por ejemplo, la actividad biológica se puede expresar cuantitativamente como la concentración requerida de una sustancia para dar una cierta respuesta biológica. Además, cuando las propiedades fisicoquímicas o las estructuras se expresan mediante números, podemos construir una relación matemática, o relación cuantitativa estructura-actividad, entre las dos. La expresión matemática puede entonces usarse para predecir la respuesta de otras estructuras químicas. Este método asigna parámetros a cada grupo químico, de forma que al modificar la estructura química puede valorarse la contribución de cada grupo funcional a la actividad del fármaco o de la toxicidad en cuestión y a partir de ahí, cómo variará la actividad de esa sustancia.⁶⁸

Como parámetros relacionados con la actividad biológica pueden usarse la afinidad al sitio activo de la enzima, las constantes cinéticas, las constantes de inhibición (K_i) o valores de concentración inhibitoria media (IC_{50}), y los parámetros fármaco cinéticos.⁶⁹

El problema de predecir la actividad biológica de compuestos químicos y el diseño de nuevos compuestos con una actividad determinada es uno de los retos más importantes en la química moderna. La manera de resolver este problema es determinar el mecanismo de interacción de las moléculas activas con el bio-receptor, sin embargo, la complejidad es enorme. La aproximación empírica al problema consiste en comparar todos los compuestos químicos activos con los inactivos que tengan estructuras similares y determinar posiblemente las similitudes y diferencias entre ellos a través de descriptores. De esta forma se podrán determinar las principales características de las moléculas responsables de la actividad. En este tipo de aproximaciones se relaciona la actividad biológica con algunas propiedades fisicoquímicas, así como la presencia o ausencia de ciertas características estructurales. Por tanto, el objetivo de un análisis QSAR es la predicción y diseño de compuestos biológicamente activos.⁷⁰

2.6. Ecuación QSAR.

Procedimiento de la obtención de la ecuación QSAR.

1.-Se debe elegir un conjunto de estructuras con actividad biológica, ya conocida o bien definida.

2.-Se procede a la elección de los parámetros (propiedades y descriptores) que pueden formar parte de la ecuación. Los dos criterios primordiales en esta elección son que los compuestos elegidos deben dar una buena dispersión de valores en los parámetros, y además los descriptores no deben estar correlacionados.

3.- Determinar el grado de correlación mediante el coeficiente de Pearson definido por la ecuación.

$$r^2 = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{[\sum_{i=1}^n (x_i - \langle x \rangle)^2][\sum_{i=1}^n (y_i - \langle y \rangle)^2]}}$$

Donde:

r^2 =Coeficiente de Pearson.

n = número de muestras.

x = suma de valores en x_i .

y = suma de valores en y_i .

$\langle x \rangle$ = Media de valores en x .

$\langle y \rangle$ =Media de valores en y .

4.- Obtener la ecuación QSAR, este método se realiza al determinar mediante una regresión lineal múltiple (RLM) la mejor combinación de coeficientes en la ecuación QSAR. El uso de un elevado número de descriptores plantea problemas fundamentales, pues la existencia de muchísimas combinaciones incrementa el grado de incertidumbre o, por consiguiente, el tiempo de cálculo, además que los descriptores en general no son ortogonales, para resolver estos problemas se han desarrollado nuevas metodologías como el análisis de componentes principales "Principal Component Analysis" (PCA) y regresión de

mínimos cuadrados parciales "Nonlinear Partial Least-Squares" (NIPALS) que transforman los descriptores no ortogonales en factores principales ortogonales, aunque existen formas matemáticas para calcular la ortogonalidad. Otras alternativas son la utilización de modelos no lineales como las redes neuronales y los algoritmos genéticos (conjunto de unidades que permiten la entrada de información, produciendo valores de salida, decisiones). Esta metodología se podrá utilizar para predecir la actividad de otros compuestos pertenecientes a la misma familia que el conjunto de referencia utilizado para generar la ecuación. Aquellos compuestos que muestran una mayor actividad serán por tanto los candidatos iniciales para ser sintetizados.⁶⁸

2.6.1. El coeficiente de Pearson.

El coeficiente de correlación de Pearson es una medida de dependencia lineal entre dos variables aleatorias cuantitativas, la correlación de Pearson es independiente de la escala de medida de las variables, podemos definir el coeficiente de correlación de Pearson como un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas y continuas. Vinculado a esto, cuando existe un valor igual a cero significa que no hay correlación, si se obtiene un valor igual a 1 existe una correlación perfecta, por ello, entre más cercano el valor a 1 mayor será la correlación.⁷¹

2.7. Actividad biológica.

La actividad biológica es el efecto que tiene un determinado compuesto sobre las células o dicho de otra forma son los efectos benéficos o adversos de un fármaco (mezcla química compleja) determinado sobre un cuerpo vivo. La determinación de las propiedades biológicas de los compuestos y el estudio de su perfil toxicológico son fundamentales para el descubrimiento de nuevos fármacos (no probados), es decir que es una predicción teórica de la modulación de la actividad biológica, optimizando su paso por el proceso a prueba y error, o de síntesis orgánica. Para llevar a cabo un estudio de tipo QSAR se necesitan básicamente tres tipos de información:

- Estructura molecular de diferentes compuestos que tengan un mismo mecanismo de acción (farmacodinamia), y que por lo tanto se considera como ligandos ante una diana biomolecular común.
- Datos de actividad biológica de cada uno de los ligandos incluidos en el estudio (por ejemplo: Dosis efectiva Media (DE₅₀), Dosis letal media (DL₅₀), Constante de inhibición (K_i), concentración inhibitoria media (IC₅₀), etc.
- Propiedades fisicoquímicas (descriptores numéricos) de los ligandos calculados por medios computacionales a partir de la estructura molecular generada *in silico*.⁶⁸

2.8. Descriptores.

Los descriptores son propiedades fisicoquímicas que se calculan a partir de una estructura virtual por métodos computacionales y/o de manera experimental obtenida y evaluada en laboratorios, son un reflejo cuantitativo o describen numéricamente a cada una de las moléculas. Estos pueden ser de diferentes tipos, y ello depende de la complejidad de información que requiera el cálculo. Así algunos descriptores son bastante “globales” describiendo toda la molécula y por ende son muy simples (como la masa molecular) y otros pueden resultar muy complejos (como la variabilidad topológica, distribución de carga eléctrica, superficie polar, por mencionar algunos).

Algunos de los principales descriptores se mencionan a continuación:

- **Generales.** Punto de fusión, punto de ebullición, presión de vapor, constantes de disociación, energía de activación, calor de reacción, potencial de reducción.
- **Hidrofóbicos.** Fundamentalmente, el coeficiente de reparto *n-octanol/agua* (log P), se pueden incluir el coeficiente cromatográfico de la fase inversa, la solubilidad en agua. La importancia del coeficiente de partición ha propiciado el desarrollo de software para su determinación (por ejemplo, CLogP, LogKow, ACDLogP).
- **Fisicoquímicos.** Se incluyen la constante de Hammett (σ) y variantes como la constante de Taft para sustituyentes polares (σ^*), así como las

constantes de campo (R) y de resonancia (F) que representan el componente inductivo y resonante de la constante de Hammett.

- **Electrónicos.** Potencial de ionización, la constante dieléctrica, el momento dipolar y el enlace por puente de hidrogeno.
- **Químico-Cuánticos.** Parámetros derivados de cálculos cuánticos, cargas parciales, densidades electrónicas, reactividad, polarizabilidad electrónica y energía de los orbitales HOMO (orbital molecular ocupado de más alta energía) y LUMO (orbital molecular no ocupado de más baja energía).
- **Estéricos.** Las interacciones polares resultantes de la correlación de van der Waals son descritas adecuadamente mediante la refractividad molar (MR). Parámetros puramente estéricos son: la constante de Taft de sustituyentes estéricos (Es), el volumen molecular, la forma molecular, el área superficial molecular y los radios de van der Waals. Se incluyen también los parámetros de Verloop (L y B) que describen la 'longitud' y 'amplitud' de los sustituyentes.
- **Estructurales.** Se incluyen los fragmentos de átomos y enlaces, las subestructuras, el entorno de la subestructura, el número de átomos en un elemento estructural determinado, el número de anillos (en compuestos policíclicos). También se utilizan los índices de conectividad molecular (%) como ejemplo de parámetros topológicos.⁶⁹

2.9. Redes neuronales.

Las redes neuronales artificiales son un modelo computacional evolucionado, el cual consiste en un conjunto de unidades, llamadas neuronas artificiales, conectadas entre sí para transmitirse señales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones) produciendo unos valores de salida.⁷²

Cada neurona está conectada con otras a través de unos enlaces. En estos enlaces el valor de salida de la neurona anterior es multiplicado por un valor de peso. Estos pesos en los enlaces pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Del mismo modo, a la salida de la neurona, puede existir una función limitadora o umbral, que modifica el valor resultado o impone un límite que no se debe sobrepasar antes de propagarse a

otra neurona. Una red neuronal combina diversas capas de procesamiento y utiliza elementos simples que operan en paralelo, y están inspiradas en los sistemas nerviosos biológicos. Consta de una capa de entrada, una o varias capas ocultas y una capa de salida. Las capas están interconectadas mediante nodos, o neuronas; cada capa utiliza la salida de la capa anterior como entrada, Figura 13 (Una red neuronal artificial es un grupo interconectado de nodos similar a la vasta red de neuronas en un cerebro biológico). Cada nodo circular representa una neurona artificial y cada flecha representa una conexión desde la salida de una neurona a la entrada de otra.⁷²

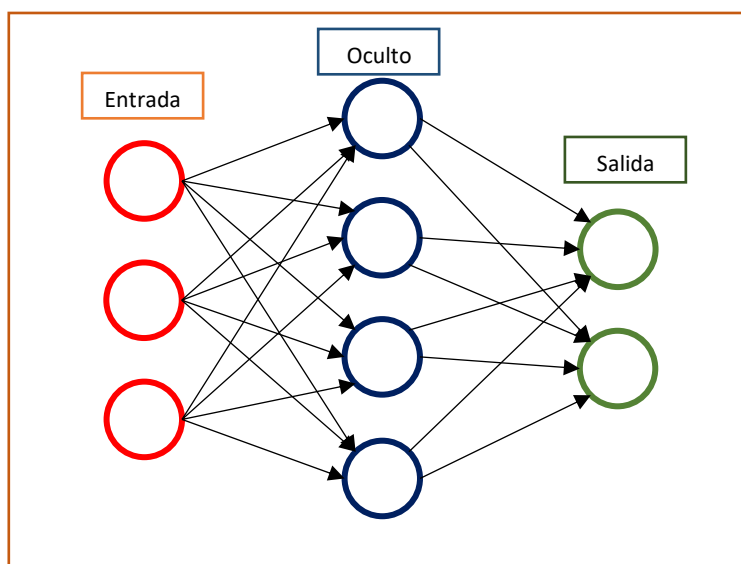


Figura 13.-Representación de Red neuronal. Edición propia.

2.10. Validación del modelo.

La validación del modelo QSAR es el concepto más importante para el desarrollo y aplicación de los mismos, por lo tanto, para decir que un modelo es descriptivo y/o predictivo se tiene que validar.⁷³

Por lo anterior expuesto, el proceso de validación confirma la fiabilidad de los modelos junto con la aceptabilidad de cada paso durante el desarrollo como la evaluación de los datos de entrada (modelo descriptivo), diversidad del conjunto de datos, y la capacidad de predicción (modelo predictivo).⁷⁴

Para poder asegurar la capacidad de predicción del modelo de entrenamiento es necesario un proceso de validación, que consiste en aplicar el modelo en un conjunto de compuestos conocidos (conjunto de predicción) que no se utilizaron en la construcción de este mismo, la validación puede ser externa o interna. En el primer caso se utiliza un conjunto de moléculas independientes, incluyendo los datos experimentales y un conjunto de descriptores moleculares, En el caso realizar una validación interna, es necesario realizar una división aleatoria de todo el conjunto de moléculas en subconjuntos. En general, se pueden clasificar en tres subconjuntos:

- Subconjunto de Entrenamiento: Son los datos con los que se construye el modelo y en el caso de las regresiones no lineales, como lo son las redes neuronales, se utiliza para “entrenar” al modelo.
- Subconjunto de Validación: Es el conjunto de datos que se reserva para comprobar el modelo.
- Subconjunto de Predicción: Entrega el error real cometido con el modelo seleccionado.⁶²

Durante el proceso de selección del mejor modelo, el ajuste se realiza sobre los datos del conjunto de entrenamiento y el error de predicción para dichos modelos es obtenido mediante el uso de los datos de validación. Finalmente, una vez que termina el proceso y se tiene seleccionado un modelo, se pueden utilizar los datos de prueba para evaluar la manera en que el modelo seleccionado sea útil para los datos que no jugaron ningún papel en la generación. El número de compuestos que se deben asignar a cada conjunto en una distribución típica puede ser de 50% para el conjunto de entrenamiento y 25% para el conjunto de validación y el conjunto de prueba, respectivamente. Sin embargo, estos porcentajes pueden variar de acuerdo al número de compuestos que se encuentren el conjunto inicial, ya que depende del número de compuestos con el que se trabaje.⁶²

Vinculado a esto para evaluar los modelos es necesario calcular diferentes parámetros estadísticos como el criterio de información de Akaike (AIC) la función de aptitud de Kubinyi (FIT) y el error medio absoluto (MAE). Estos permiten evaluar la gran cantidad de datos que se utilizan para construir y validar

el modelo mediante el análisis del valor de cada parámetro. Los parámetros AIC y FIT sirven para seleccionar el mejor modelo dentro de un conjunto obtenidos con los mismos datos.⁷⁵

El parámetro FIT determina si el modelo mejora con la adición de nuevas variables, y se calcula de acuerdo con la siguiente ecuación:

$$FIT = \frac{R^2(n - k - 1)}{(n + k^2)(1 - R^2)}$$

Donde:

n =número de compuestos en el conjunto de entrenamiento.

k =número de descriptores utilizados en el modelo.

R^2 =El cuadrado del coeficiente de correlación.

El criterio de FIT presenta una baja sensibilidad a los cambios en los valores de k , siempre que sean números pequeños, y una sensibilidad sustancialmente incrementada para un mayor número de descriptores. En consecuencia, el mejor modelo presentará el valor más alto de esta función.⁷⁴

El criterio AIC toma en cuenta la bondad de ajuste estadístico y el número de parámetros que se deben estimar para lograr ese grado de ajuste. Este criterio nos ofrece una medida relativa de la pérdida de información de un modelo y se calcula utilizando la siguiente ecuación.

$$AIC = (RSS) \left(\frac{(n + k)}{(n - k)^2} \right)$$

Donde:

RSS = la suma del cuadrado de los residuos.

n = número de compuestos en el conjunto de entrenamiento.

k = número de descriptores en el modelo.

De acuerdo con este criterio el modelo que produce el valor AIC mínimo debe considerarse potencialmente el más útil.⁷⁴

El uso del error medio absoluto (MAE) evita llegar a conclusiones engañosas sobre la predicción de un modelo elegido exclusivamente por parámetros de validación basados en R^2 , que están determinados por atributos tales como el intervalo de respuesta y la distribución de datos alrededor del conjunto de entrenamiento o predicción. Los criterios basados en el MAE sirven para evaluar el rendimiento de los modelos con un mayor grado de confianza, por lo cual un valor de MAE igual a cero indica que la predicción es perfecta, el valor de MAE aumenta a medida que los datos predichos se alejan de los datos observados.

La fórmula

$$MAE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}$$

donde:

n = número de compuestos.

y_i =es la predicción.

\hat{y}_i =el valor verdadero.

El conjunto de predicción no refleja una distribución uniforme de los residuos de predicción entre los datos, por lo tanto, un estudio sobre la distribución de los errores es razonable. Los errores pueden indicar la presencia (o ausencia) de un número considerable de predicciones atípicas en el conjunto de predicción que de otro modo permanecerían desapercibidas por una cantidad relativamente grande de puntos de datos donde la media el error todavía puede tener un valor bajo. Los errores de predicción en un conjunto pueden estar presentes en un pequeño número, con altos valores de residuo, o un mayor número con valores de residuos moderados. Se considera una predicción de conjuntos donde la mayoría de los compuestos están bien predichos, y sólo algunos de ellos están mal predichos. Se puede tener otro conjunto donde el rendimiento predictivo del modelo es moderado para la mayoría de los compuestos. Ahora, usando una métrica basada en errores, se puede mostrar que el valor de la métrica es igual o cerca de sí en los dos casos, pero no da ninguna idea sobre la distribución de los errores, por lo tanto, el mejor modelo es aquel en donde la mayoría de las predicciones son buenas, el 95% de ellas, basándose en una distribución normal

estándar y una probabilidad de cometer errores, de 0.005 de los compuestos en el conjunto de predicción, ya que puede haber un número limitado de predicciones con valores atípicos que pueden afectar a los valores de las métricas basadas en errores calculados desde el conjunto de entrenamiento. Por ende, se revisa continuamente dos aspectos, los errores de predicción, que confiere a la dispersión de errores, y en segundo la comparación de los valores de las métricas basados en error después de una pequeña fracción definida (5%) de compuestos con alto contenido valor residual para ese conjunto de predicción, MAE se considera un parámetro más simple y más directo en la cuantificación de los errores de predicción.⁶²

Ejemplo del análisis MAE, en la Figura 14 se localizan dos puntos que muestran el error absoluto medio y para el error cuadrático medio. Donde tenemos la comparación de dos observaciones en la que $X_1=2$ y $X_2=6$.

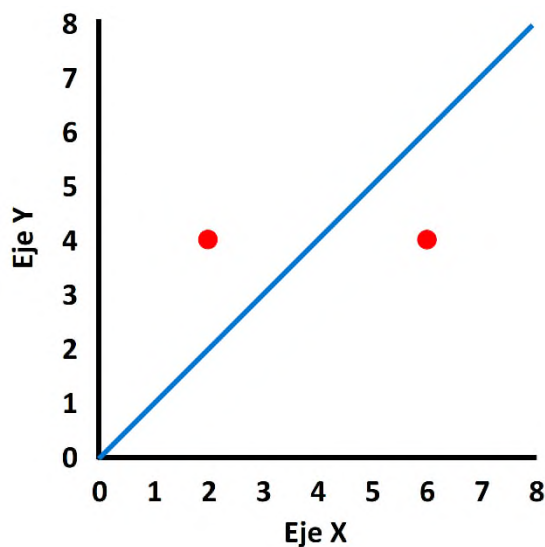


Figura 14.- Error medio absoluto.

2.11. Método de aplicabilidad de dominio.

La aplicabilidad de dominio (AD), permite identificar a los compuestos que se encuentran en la región campo fisicoquímico, donde tienen características similares dentro del modelo matemático, por lo tanto, si algún compuesto del conjunto de entrenamiento presenta diferencias en la región estructural o tuviera un mecanismo de acción diferente al que tuvieron los demás compuestos del

conjunto, podrían ser eliminados del conjunto de entrenamiento, para volverlo más representativo.⁷⁶

Existen dos tipos de valores atípicos, los X- y los Y-valor atípico, los primeros son los valores que se encuentran fuera de la AD y por lo tanto conducen a resultados erróneos con respecto a análisis estadístico, es decir, no pertenecen directamente a la estadística, mientras que los segundos se caracterizan por presentar un valor de residuo alto y por lo tanto están fuera de la estadística, sin embargo, se pueden integrar a ella por medio de un método no lineal.⁷⁶

Tipos de métodos disponibles para determinar la AD.

Existen varios enfoques para determinar la EA de los modelos QSAR, los enfoques más comúnmente empleados para estimar las regiones de interpolación en un espacio multivariado incluyen los siguientes.

- Rangos en el espacio del descriptor.
- Métodos geométricos.
- Métodos basados en la distancia.
- Distribución de densidad de probabilidad.
- Rango de la variable de respuesta.

Los primeros cuatro enfoques se basan en la metodología utilizada para la caracterización del espacio de interpolación en el espacio del descriptor del modelo. Por el contrario, el último depende únicamente del espacio de respuesta de las moléculas del conjunto de entrenamiento. Los métodos existentes para determinar la AD se muestran en la Figura 15. Entre los métodos existentes, ninguno puede considerarse el mejor universalmente, cada método tiene sus propios méritos y defectos. De las publicaciones QSAR de la última década, se puede ver que el enfoque de apalancamiento (gráfico de Williams), DModX y diferentes enfoques de evaluación de similitud, se han empleado en gran medida para identificar los valores atípicos y los compuestos que residen fuera de la AD. Estos métodos dependen en cierta medida del uso de herramientas de software estadístico especializadas para el cálculo de diferentes parámetros para la identificación de valores atípicos y la determinación de la AD.⁷³

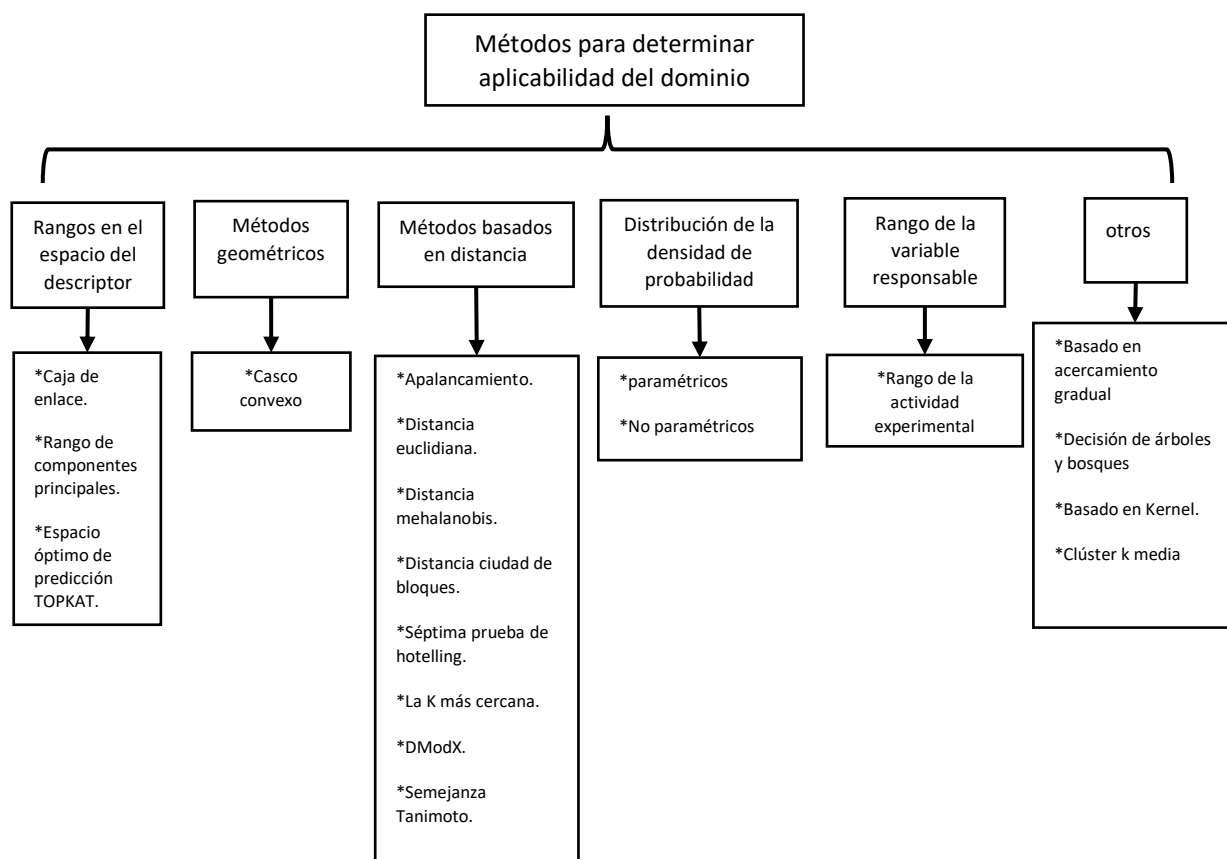


Figura 15.-Métodos para determinar la aplicabilidad del dominio

El método de apalancamiento (LA por sus siglas en inglés) y la aproximación por estandarización (SA, por sus siglas en inglés) son métodos basados en las distancias que existen entre descriptores, además, los valores atípicos se identifican fácil y rápidamente.⁷³

2.12. Aproximación por apalancamiento (LA).

La aproximación por apalancamiento (LA) es una forma de determinar la AD, estimando la distancia de un compuesto al centroide de los compuestos de entrenamiento. Es adecuada para evaluar explícitamente el grado de influencia que un compuesto de entrenamiento tiene en el modelo. En particular, identifica rápidamente los compuestos con menos influencia en la construcción de modelos y los de alta influencia que aumentan el campo químico, físico o biológico que están cubiertos por el modelo, por medio de la condición $h_i > h^*$, en

donde h_i es el valor de apalancamiento calculado para cada compuesto y h^* es el valor calculado para el conjunto de entrenamiento, si se cumple la condición, el compuesto i -ésimo se considera como un X-valor atípico, $x_{núcleo}$ se puede definir como el centro de los centroides de cada descriptor.⁷⁷

$$x_{núcleo} = (x^T \cdot x)^{-1}$$

La distancia de un compuesto sobre el modelo del espacio de variables originales se calcula.

$$h_i = (x_i^T)(x_{núcleo})(x_i)$$

Donde:

x_i =vector de i -descriptores del compuesto considerado.

x =matriz de los valores de descriptores del modelo, derivada de la serie de entrenamiento.

El apalancamiento se calcula con:

$$h^* = \frac{3(g + 1)}{n}$$

Donde:

n = número de conjuntos de entrenamiento.

g = número de descriptores utilizados en el desarrollo del modelo.⁷⁷

2.13. Aproximación por estandarización (SA).

Esta aproximación se basa en que el comportamiento de los descriptores de los compuestos en el conjunto de entrenamiento debe seguir una distribución normal. De acuerdo con esta distribución, el 99,7% de la población permanecerá dentro del intervalo de la media ± 3 desviación estándar ($\pm 3\sigma$). Por lo tanto, la media $\pm 3s$ representa la zona en donde se encuentran la mayoría de los compuestos del conjunto de entrenamiento. Cualquier compuesto fuera de esta zona es diferente a la mayoría de los compuestos. Para realizar el análisis SA,

primero se normaliza por descriptor, los i-descriptores de los k-compuestos presentes en el modelo, de acuerdo con la siguiente fórmula.

$$S_{Ki} = \frac{|X_{Ki} - \bar{X}_i|}{\sigma_i}$$

Donde:

S_{Ki} =el i-ésimo descriptor estandarizado por cada k-ésimo compuesto.

X_{Ki} =es el i-ésimo descriptor original del k-ésimo compuesto.

\bar{X}_i =valor medio de cada serie de descriptores en el conjunto de entrenamiento.

σ_i =desviación estándar de cada serie de descriptores en el conjunto entrenamiento.

El valor de SA se obtiene a partir de los valores estandarizados por los descriptores. Se inicia calculando la media y la desviación estándar de los descriptores estandarizados sobre cada compuesto, y se realiza la suma de la media y el valor de 1.28 veces la desviación estándar que indica una confiabilidad de 90 %.

$$SA = (\bar{S}_k + 1.28)(\sigma_k)$$

Donde:

\bar{S}_k =valor medio de los i-ésimos descriptores estandarizados por cada k-ésimo compuesto.

σ_k =desviación estándar del componente.

Si el valor de SA de un compuesto se encuentra por debajo de 3, se puede considerar que no es un X-valor atípico, contrario será un X-valor atípico. ⁷⁷

Capítulo 3.- Justificación.

El potencial de la inhibición de GSK3 como una terapia contra varias enfermedades está ampliamente descrito en la literatura. Sin embargo, existen igual número de reportes que indican efectos nocivos derivados de inhibir a la enzima GSK3 mediante moléculas que reduzcan la actividad cinasa de sus dos parálogos debido a su similitud. Los experimentos de edición genética indican que, inhibir la actividad de GSK3 β , pero no de GSK3 α , reproduce los efectos benéficos de la inhibición de GSK3 en enfermedades como Alzheimer y cáncer. Por lo tanto, identificar moléculas y sus características que permitan la inhibición selectiva de GSK3 β es una estrategia poco explorada pero potencialmente fructífera en la farmacología. En este sentido, la química computacional como herramienta tiene la característica de poder analizar el comportamiento de cientos o miles de moléculas con el objeto de predecir el comportamiento de un sistema de inhibidor/enzima y en consecuencia poder determinar las características comunes de un grupo variado de moléculas. Aplicando esta estrategia a los inhibidores conocidos con actividad inhibidora selectiva por GSK3 β , permitirán en una primera etapa conocer parte del mecanismo de inhibición de estas moléculas y generar modelos predictivos. En una segunda etapa se podrán analizar otras moléculas ya conocidas, pero no consideradas inhibidores de GSK3 β , con el objetivo de identificar nuevos inhibidores más potentes o selectivos.

Capítulo 4.- Hipótesis.

Los inhibidores selectivos de GSK3 β se pueden diferenciar por sus características estructurales, electrónicas y fisicoquímicas a partir de una estrategia computacional.

Capítulo 5.- Objetivos.

5.1. Generales.

Establecer la relación cuantitativa estructura actividad (QSAR) que determine las características estructurales, electrónicas y fisicoquímicas de inhibidores selectivos de la actividad del parólogo GSK3 β .

5.2. Específicos.

- Realizar la dinámica molecular de la proteína GSK3 β .
- Diseñar la síntesis *in silico* y optimizar una serie de moléculas estables inhibidoras del parólogo GSK3 β .
- Establecer las interacciones entre el parálogos GSK3 β y los inhibidores a través de acoplamiento molecular (Docking).
- Calcular los descriptores estructurales, eléctricos y fisicoquímicos de las moléculas inhibidoras.
- Generar un modelo tipo QSAR utilizando los descriptores anteriores.
- Diferenciar las características de las moléculas inhibidoras del parólogo GSK3 β .
- Como perspectiva se diseñará *in silico* nuevas moléculas inhibidoras de la proteína GSK3 β a partir del modelo QSAR obtenido.

Capítulo 6.- Metodología.

6.1. Dinámica molecular GSK3 β .

La metodología corresponde en primera instancia al tratamiento de la proteína GSK3 β , obtenida de la base de datos Protein Data Bank (<https://www.rcsb.org/>), la cual cuenta con un identificador de objeto digital DOI:10.2210/pdb1H8F/pdb. Esta plataforma provee más de 150000 proteínas los cuales han sido utilizadas para diferentes métodos experimentales como cristalografía de rayos X, espectroscopia de resonancia magnética nuclear, microscopía electrónica y otros. Para este caso la proteína de estudio cuenta con aproximadamente 70 proteínas de GSK3 β , es importante destacar que estas 70 proteínas no son la misma debido a que varias de estas son homólogas de otras especies como plantas, animales, hongos, entre otros; cabe destacar que el número de éstas va en aumento, haciendo un análisis en esta base de datos se encontró que 15 de estas proteínas son específicamente de humano. Se seleccionó a la proteína con código pdb/1H8F; es necesario comentar que, en la elección de la proteína para este estudio se priorizó la función de evaluar inhibidores competitivos de ATP, ya que los demás son GSK3 β , pero éstas son de diferentes estudios para inhibir otros complejos de proteínas, evaluaciones toxicológicas, diferentes tipos de patologías, de las cuales no destaca para evaluar inhibidores competitivos del ATP que es su principal sustrato. Una vez seleccionada la proteína, la obtenemos en formato PDB en estado cristalizado debido a que ha sido utilizado para análisis de rayos x. La edición de la proteína se realizó con el software Chimera 1.13.1, para aislar uno de los dos monómeros que se encuentran en el archivo original, debido que ésta cuenta con dos proteínas de GSK3 β , moléculas de agua e inhibidores, además se eliminaron los cofactores para aislar a la proteína y realizar la dinámica molecular con software Gromacs MD suite, en el cual se asignaron los parámetros y condiciones necesarias para simular las propiedades fisicoquímicas y termodinámicas en un ambiente fisiológico.⁷⁸

La biblioteca virtual Protein Data Bank es reconocida a nivel mundial por ser una base de datos de la estructura tridimensional de las proteínas y ácidos nucleicos, los cuales son enviados por biólogos y bioquímicos de todo el mundo. Están bajo el dominio público y pueden ser usados libremente.

6.2. Búsqueda de moléculas reportadas en la base de datos ChemBL.

En relación con búsqueda, en la base de datos se realizó el mismo procedimiento para los inhibidores de la proteína GSK3 β , existen varias bibliotecas virtuales especializadas en moléculas para inhibir proteínas, uso para fármacos y otras cuestiones como lo son PubChem, ChEMBL, ChemSpider, etc. En relación a lo mencionado nos centramos en el uso de la base de datos Chembl, la cual contiene moléculas bioactivas con propiedades de tipo farmacéuticas.

Una vez identificada la base de datos, se realizó la búsqueda de inhibidores de la proteína GSK3, en la página oficial ([//ebi.ac.uk/chembl/](http://ebi.ac.uk/chembl/)), las consideraciones tomadas para la selección de los inhibidores es que presentaran constantes de inhibición para GSK3 de los dos parálogos (alfa y beta), tomando como punto de partida las moléculas reportadas en la especie *Homo sapiens* (humano). Además de considerar la constante inhibitoria (K_i) y la concentración inhibitoria media máxima (IC_{50}). Por consiguiente, se formaron tres grupos de moléculas utilizando los valores K_i y de IC_{50} , y como condiciones de elección se tomó en cuenta lo siguiente: el primer grupo con preferencia al parálogo alfa, el segundo con preferencia al parálogo beta y finalmente el tercero, que no presentó preferencia por ninguno de los parálogos. Este procedimiento se realizó para optimizar la clasificación y facilitar el manejo de la base de datos de más de 2000 moléculas, para ello fue necesario definir un coeficiente relativo de inhibición, el cuál es el resultado de dividir el valor de K_i reportado para el parálogo beta sobre el valor reportado para el parálogo alfa, si el valor del coeficiente es menor a 1 (en el intervalo de 0.1 a 0.7) significa que el inhibidor es selectivo por el parálogo beta y si es mayor a 1 (mayor a 1.2) significa que hay selectividad hacia el parálogo alfa y si el valor es igual a 1 indica que el inhibidor no tiene selectividad. Este mismo procedimiento se realizó con las IC_{50} , para que un inhibidor se considere selectivo debe tener una selectividad por el parálogo GSK3 β de 2.5 órdenes de magnitud debido a que es el valor mínimo reportado para impedir la activación de la WNT por un inhibidor selectivo, de esta manera se obtuvieron los tres grupos de inhibidores.

6.3. Representación de moléculas inhibidoras.

Una vez analizadas y recopiladas las moléculas, las estructuras moleculares se esbozaron *in silico* a través del programa GaussView3.0. En algunos casos, la estructura tridimensional se encontró en la base datos y tan solo requirió obtenerlas en la plataforma (pocos casos), debido a que en su mayoría las moléculas son representadas en dos dimensiones (planos), no aportan la suficiente información y debido a la falta profundidad fue necesario su esbozo para obtener la información a través de los diferentes programas, los cuales fueron presentados, esto permitió obtener información relevante. Otras cuestiones por la que no hay más datos al respecto de los inhibidores es por qué están en fase experimental debido a esto no reportan información más reciente al respecto.

6.4. Optimización geométrica de los inhibidores.

No obstante, fue necesario llevar a cabo un ajuste de las geometrías de estas moléculas debido a que los átomos dentro de la configuración molecular no tienen un equilibrio espacial respecto a sí mismo, por lo cual, las geometrías se optimizaron. Para obtener los ángulos de torsión finales de los confórmeros en cada mínimo, los cálculos de estas geometrías se realizaron sin restricción en los ángulos diedros, los cálculos se llevaron a cabo usando el programa Gaussian03. Además, se utilizó el funcional híbrido B3LYP, que es más rápido que la mayoría de los métodos de Post Hartree-Fock y produce resultados comparables a los obtenidos en un laboratorio. También es bastante robusto para un método DFT sumándose a éste el conjunto base 6-31G* para partir de una geometría de mínima energía. El conjunto de bases 6-31G* (definido para los átomos H a Zn) es un conjunto de bases polarizadas de valencia doble-zeta que suma al conjunto 6-31G cinco funciones de polarización cartesiana-gaussiana de tipo *d* en cada uno de los átomos de Li a Ca y diez funciones de polarización gaussianas cartesianas de tipo *f* en cada uno de los átomos Sc a Zn.

6.5. Acoplamiento molecular.

El acoplamiento molecular se realizó utilizando a la enzima GSK3 β con receptor y a los diferentes ligandos (inhibidor) en un formato de archivo llamado *.pdbqt, que es un formato de banco de datos de proteína modificado que contiene cargas atómicas, definiciones de tipo de átomo y la información topológica (enlaces rotables) obtenidas por la optimización e igual fue necesario realizar exhaustivas revisiones debido a que los errores humanos se hacen presentes.

6.6. Cálculo de descriptores cuánticos.

Una vez seleccionados e identificados los conformeros de menor energía obtenidos por el acoplamiento molecular, se realizó el cambio de coordenadas de *.GJF a *.MOL con el programa Babel versión 1.3 con el fin de obtener el formato adecuado para el cálculo de los descriptores por medio del programa PaDEL Descriptor.

El número de descriptores a considerar estará en función de las herramientas computacionales de cálculo con las que se cuente y del número de moléculas incluidas en el estudio.

6.7. Modelo QSAR.

Con base a los descriptores obtenidos se construyó un modelo que permitirá determinar la K_i , la cual se expresó como:

Actividad = Suma de sus propiedades fisicoquímicas y/o propiedades estructurales.

Se realizó un alineamiento de la proteína-inhibidor donde las constantes (descriptores) fueron las constantes dadas por la ecuación de regresión multilínea, este modelo se utilizó para predecir inhibidores de la proteína GSK3 β .

Claramente debemos tomar en cuenta que los descriptores pueden ser considerablemente numerosos por lo cual para generar el modelo QSAR requirió de un análisis el cual se excluyeron los descriptores que tenían el 70% de los valores constantes.

Por ende, el conjunto de moléculas bajo estudio se dividió en dos subconjuntos, uno de entrenamiento y otro de predicción. Para el subconjunto de entrenamiento se asignó un porcentaje de 66.1%, mientras que al subconjunto de predicción estuvo constituido por el 33.9 % de los compuestos iniciales.

6.7.1. Regresión lineal múltiple.

Una vez seleccionados los compuestos que formarán al conjunto de entrenamiento, se utilizó el programa SPSS estadístico para Windows para calcular la regresión lineal múltiple bajo la metodología “paso por paso”. Se utilizó como variable dependiente al Log (Ki) experimental y como variables independientes los descriptores seleccionados en el “pool de descriptores”, incluyendo descriptores cuánticos, fisicoquímicos, topológicos, polares, estéricos y los índices topológicos.

6.7.2. Análisis de aplicabilidad del dominio.

La AD se realizó sobre los subconjuntos de entrenamiento y predicción, este análisis se llevó a cabo utilizando dos metodologías: por estandarización (SA) y por apalancamiento (LA).

La AD global se determinó al trazar los valores obtenidos de SA y LA para los dos subconjuntos. El campo físico se localiza tomando los valores límites de 3 para el SA y de h^* para el LA, formándose 4 regiones, en la Figura 16 muestra la propuesta para el análisis de la aplicabilidad de dominio. La línea azul representa el valor límite del LA y la línea roja señala el límite del SA.

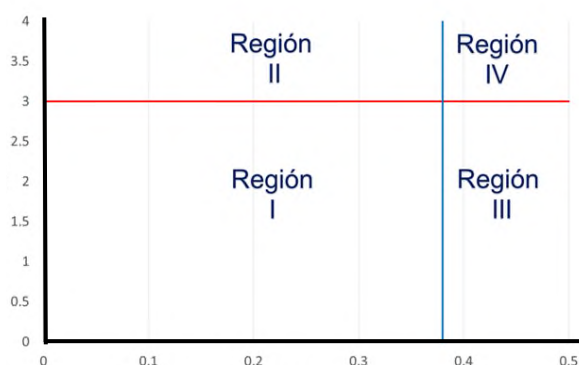


Figura 16.-Análisis de la aplicabilidad de dominio.

Para entender la función se identificó que la región I están los valores pertenecientes al campo físico/químico/biológico, en la región II se localizan los valores atípicos identificados por SA, en la región III se encuentran los valores atípicos identificados por LA y, finalmente, en la región IV se encuentran los valores atípicos identificados por ambas metodologías.

La selección de los compuestos para cada subconjunto se realizó por medio de la AD, repitiendo el análisis hasta que el subconjunto de entrenamiento no presente valores atípicos.

6.7.3. Valores atípicos.

Se sabe que estadísticamente la existencia de varios criterios para identificar valores atípicos dentro de un conjunto, los valores atípicos son puntos que se desvían significativamente de un modelo, es decir, no se ajustan al modelo, afectando los parámetros estadísticos del mismo. Generalmente, la identificación de valores atípicos busca un mejoramiento cuantitativo del modelo. Por lo cual existen dos tipos de valores atípicos los X y los Y. Los Y-valores atípicos presentan un valor de residuo alto y por lo tanto están fuera de la estadística, sin embargo, se pueden integrar a ella por medio de un método no lineal como redes neuronales. Los procedimientos de RLM, AD y cálculos de parámetros estadísticos se deben repetir hasta que estos parámetros indiquen que la calidad del modelo no mejora, y los valores atípicos tipo X, no pueden ser integrados al modelo ya que son compuestos que no comparten características con todo el conjunto por lo cual deben ser eliminados del modelo para que este pueda mejorar para no generar resultados erróneos.

6.7.4. Análisis MAE.

El análisis MAE requiere la comprobación de los subconjuntos de entrenamiento, predicción y que al menos tengan 10 compuestos para que el resultado sea válido estadísticamente por lo cual MAE se calcula utilizando la fórmula descrita en la sección 2.10 con el 100 % de los compuestos para cada subconjunto. Se calcula $MAE+3\sigma$, omitiendo el 5% de los compuestos con alto contenido valor residual para cada subconjunto utilizando la fórmula de MAE, ya con los valores

de MAE, se calcula el intervalo de Log (Ki) para el conjunto de entrenamiento, una vez establecido el intervalo se utilizan las siguientes condiciones para determinar si la calidad de los subconjuntos es buena, es decir:

$$si MAE \leq 0.1 * I \text{ y } MAE + 3\sigma \leq 0.2 * I$$

Donde el intervalo I del Log (Ki) se cumple entonces la predicción del modelo es buena, en dado caso no se cumpla las condiciones, se analizarán las condiciones y se continua con la siguiente, en caso contrario se da por terminada la metodología.

$$si MAE \leq 0.15 * I \text{ y } MAE + 3\sigma \leq 0.25 * I$$

Entonces la predicción del modelo será considerada regular, de lo contrario será una mala predicción.

6.8. Comparación de los parálogos de la proteína GSK3.

El propósito de esta actividad es poder identificar las características importantes de la proteína GSK3 respecto a su parálogo, que en definitiva, la comparación de las proteínas fue muy importante debido a que se reconocieron disimilitudes importantes que pueden influir en la selectividad de inhibidores por GSK3 β , el procedimiento para este análisis partió plenamente de que hay una modificación en la proteína GSK3 β , esta modificación es plenamente sobre el aminoácido diferencial dentro del sitio catalítico mencionado en la sección 1.6, el cual permitió comparar la GSK3 β original con la GSK3 β modificada; cabe mencionar que al contrastar las dos estructuralmente se visualizaron cambios notables dentro del sitio catalítico, los cuales se mencionan con más detalles en las secciones siguientes, lo encontrado marca la diferencia para predecir la selectividad de las moléculas.

6.9. Propuesta de moléculas nuevas.

Como perspectiva, con base en la estrategia de farmacóforo (unidad central molecular responsable de la actividad biológica), se diseñaron moléculas y se realizaron propuestas de inhibidores que tengan actividad ante la proteína

GSK3 β (inhibición) cumpliendo así con nuestras hipótesis de que existe selectividad de los inhibidores sobre la proteína GSK3 β y permitir realizar otros futuros estudios con base a esta propuesta.

En la Figura 17 se ilustra la metodología empleada en forma de diagrama de flujo.

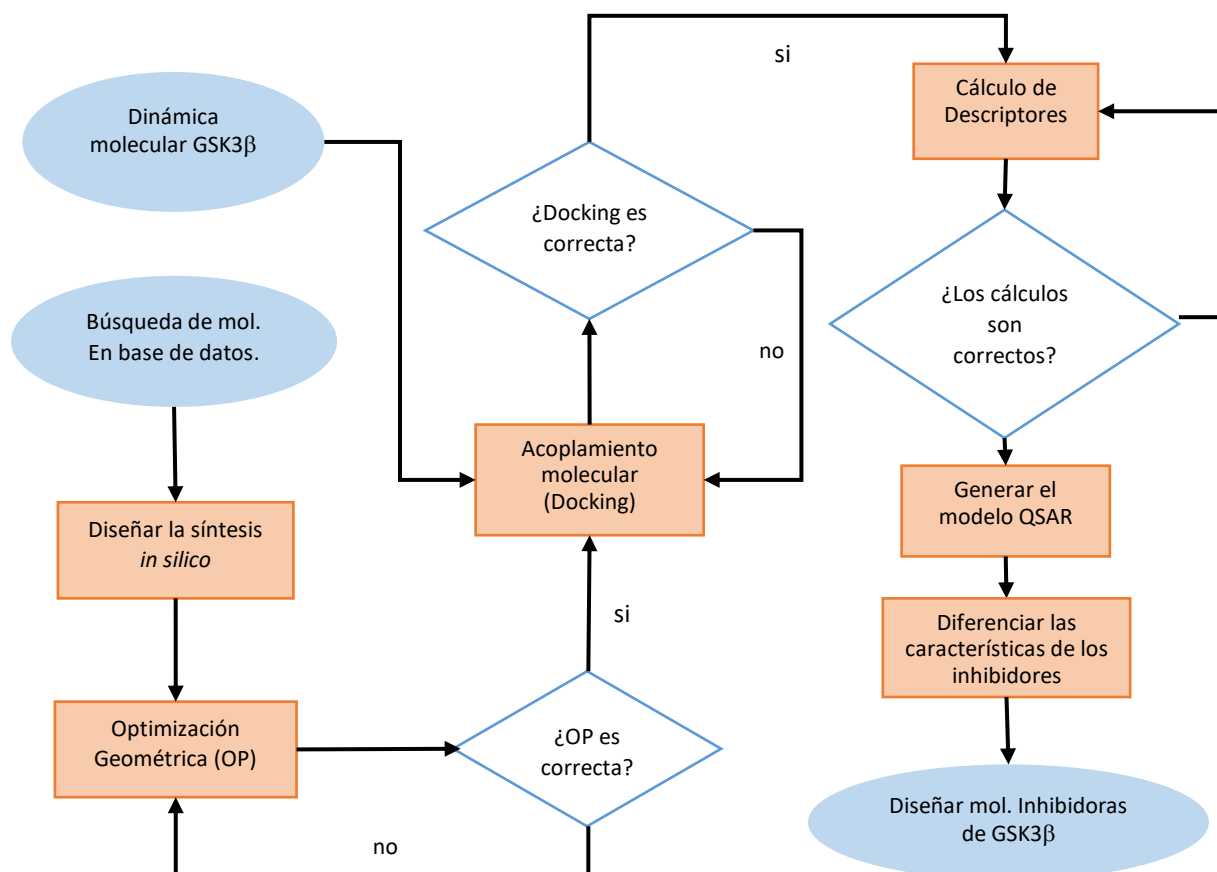


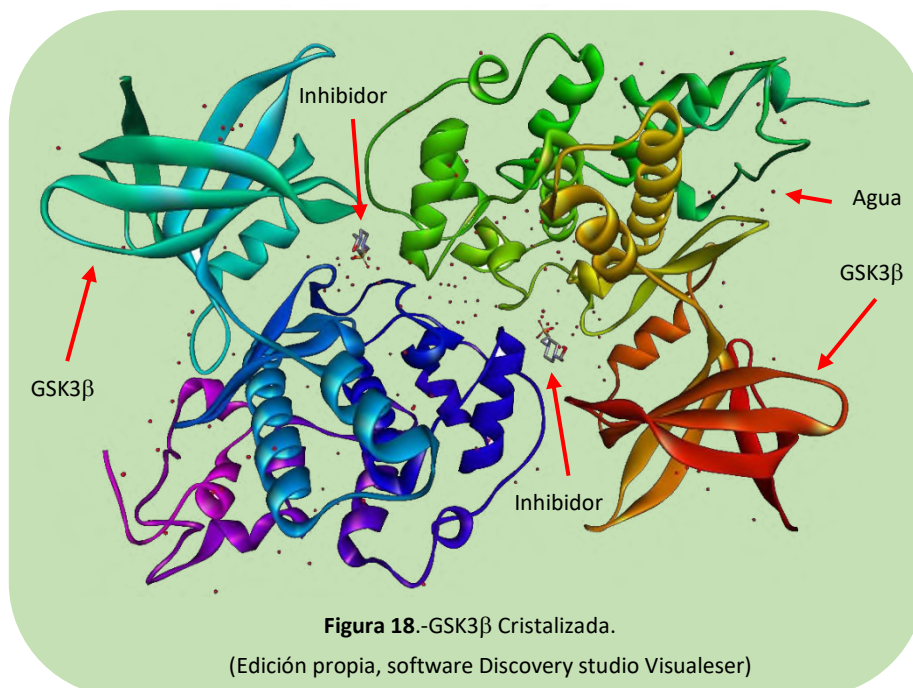
Figura 17.-Pasos de la metodología.

Capítulo 7.- Resultados.

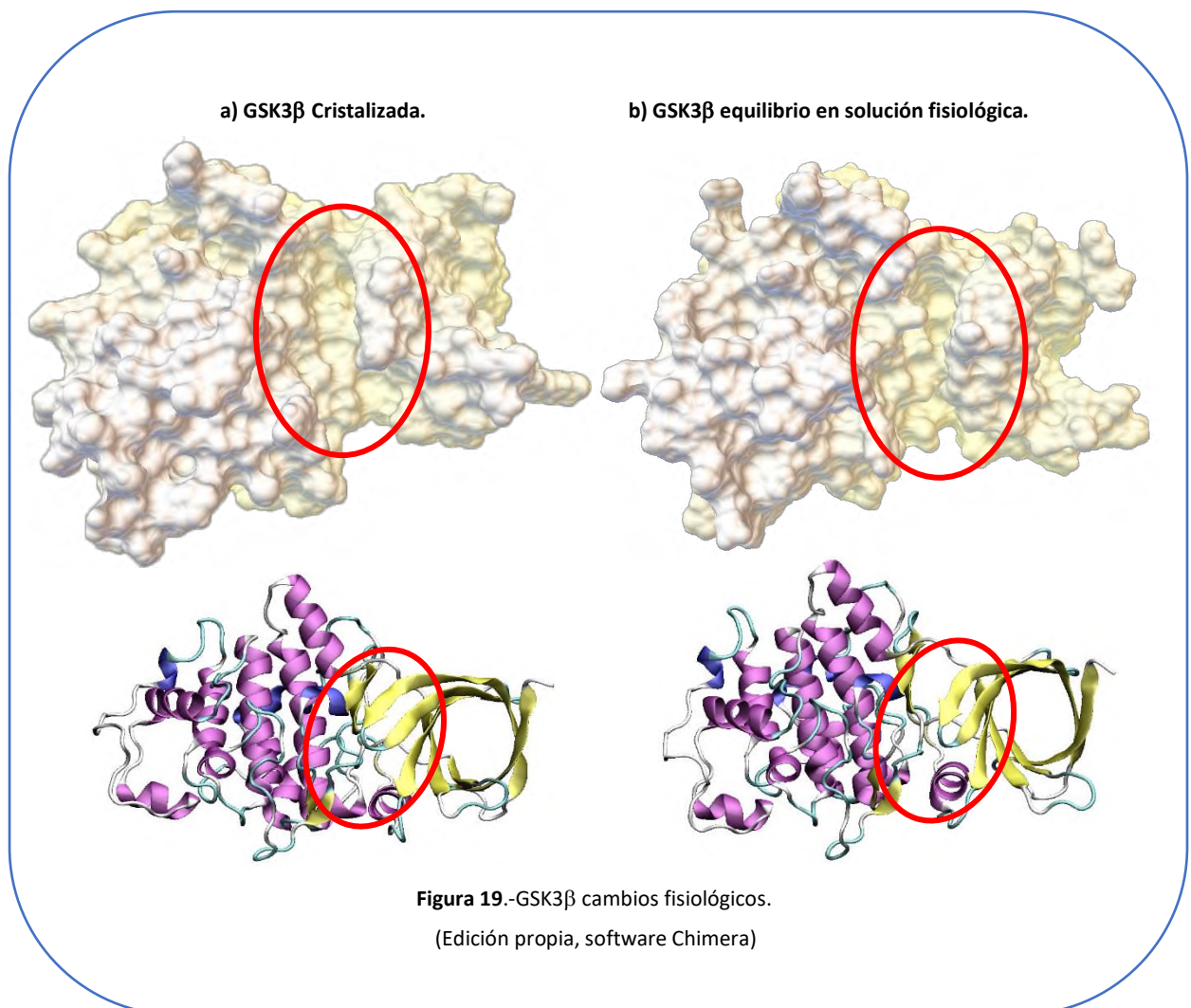
7.1. Dinámica molecular.

La proteína GSK3 β se seleccionó de la base de datos protein data bank, el código se verificó 1H8F, se observó que esta proteína estaba formada por dos macromoléculas es decir de forma dímera, particularmente la proteína GSK3 β no requiere de otra proteína para llevar a cabo su actividad cinasa, por tal motivo, se eliminó una proteína de GSK3 β , se eliminaron moléculas de agua, inhibidores, y/u otras moléculas incluidas en el archivo con formato PDB. Es necesario resaltar, que la proteína es específicamente en Humano (*homo sapiens*), con la particularidad de determinar inhibidores competitivos del ATP. Los cálculos de dinámica molecular iniciaron con la preparación de las condiciones fisicoquímicas y termodinámicas. Las condiciones en las que se encuentra esta proteína son pH=7.5, temperatura de -173.15 °C y presión de una atmósfera.

La proteína GSK3 β con la cual se llevó a cabo la dinámica molecular puede observarse en la Figura 18, en ella se indica con flechas rojas los componentes que se eliminaron. Las condiciones termodinámicas y fisicoquímicas utilizadas en el cálculo con el Software GROMACS fueron: temperatura a 37 °C, pH de 7.5 y presión de una atmósfera, es decir se le proporcionaron las condiciones fisiológicas.



Los resultados de la dinámica molecular consisten en cambios estructurales, estos cambios son visibles en la estructura de la proteína GSK3 β , dentro de este marco, es de relevancia el sitio catalítico, debido a que nuestro objetivo es encontrar inhibidores selectivos. En la Figura 19a se representa la conformación inicial de la proteína, se observa el sitio catalítico estrecho, debido a las condiciones y al inhibidor, objeto de estudio para otras investigaciones, por esto mismo nos referimos a que la proteína esta tensionada o vagamente dicho estresada. En la Figura 19b se presenta el mismo sitio catalítico más amplio, la proteína llega a un equilibrio estructural o termodinámico, en este sentido las moléculas que conforman a la proteína estaban tensionadas por sus energías, ángulos y/u otros, estos cambiaron por las condiciones asignadas, la relajación se da cuando se cambian las condiciones a fisiológicas, los cambios son notorios visualmente.



7.2. Radio de giro.

En relación con lo anterior, los cambios notables visualmente en la proteína se observan cuantitativamente por medio del análisis del radio de giro. Estos cambios conformacionales se miden en un periodo de tiempo y distancia, considerando un tiempo de 200 ns y los cambios conformacionales por medio de los movimientos generados en nanómetros (nm). En la Figura 20 se observa la gráfica del radio de giro en función del tiempo de la dinámica molecular, en la cual se observa que en el tiempo de 0 a 50 ns, la proteína tiene movimientos realmente sobresalientes; después de los 50 hasta los 200 ns no se observan cambios significativos, es decir que la proteína obtiene una media del radio de giro con movimientos prácticamente constantes, por lo cual la proteína tiene un equilibrio. Se obtiene una media del radio de giro aproximada indicada con la línea roja, concluyendo que el sitio catalítico se abre y cierra hasta llegar a un equilibrio finalizando en un punto donde este no tenga cambios significativos en la estructura, alcanzando su equilibrio o estado de relajación.

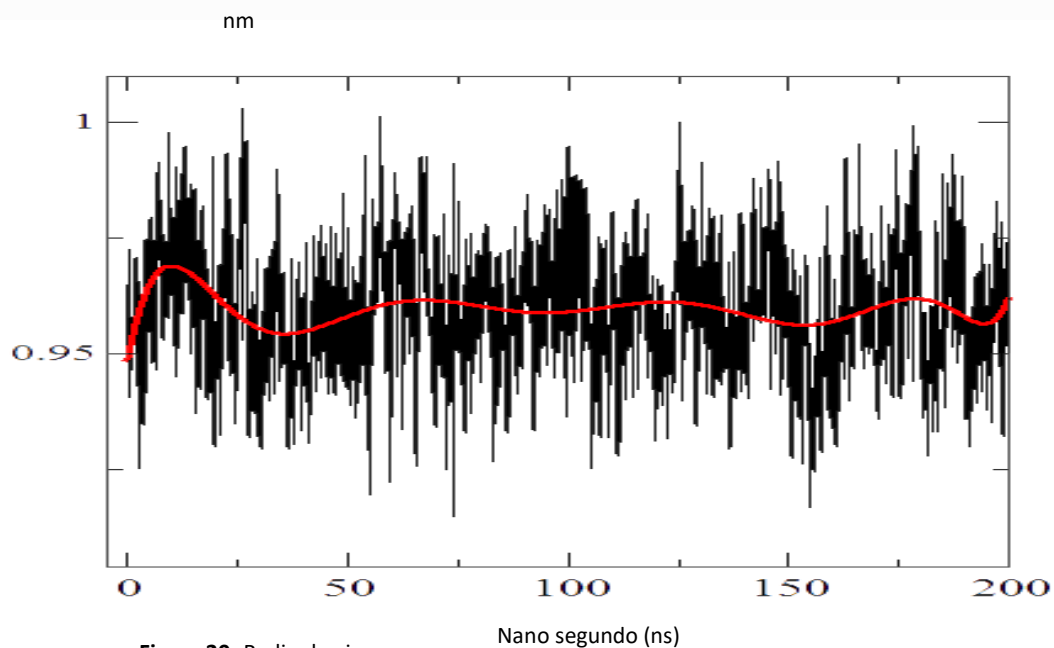


Figura 20.-Radio de giro.

(Edición propia, software photoshop)

7.3. Búsqueda de moléculas reportadas en la base de datos ChemBL.

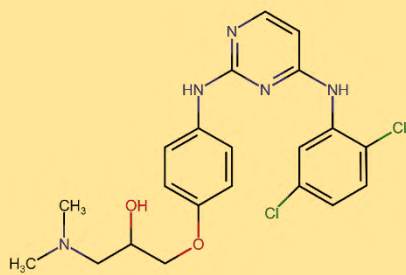
El conjunto de moléculas obtenidas de la base de datos de ChemBL se analizaron de acuerdo a su preferencia, esto mismo fue simplificado por la propia base de datos, de las cuales se encontraron más de 2000 moléculas reportadas para la proteína GSK3. Este número de moléculas se redujo al observar la concentración inhibitoria media máxima (IC_{50}), sin embargo, debido al reducido número de moléculas, ésta no sobrepasó más de 100 moléculas para los parálogos de GSK3, se decidió cambiar por la constante de inhibición (K_i), lo que permitió una lista con más de 1400 moléculas reportadas, la lista se redujo a 250 moléculas, debido a que deben estar reportadas las moléculas para ambos parálogos de GSK3, lo que proporcionó 150 moléculas, las que presentan preferencia por el parálogo beta, 50 moléculas con preferencia al parálogo alfa y 50 moléculas sin preferencia.

La selección de estas moléculas se llevó a cabo de la manera siguiente: se procedió a realizar la relación entre los valores del $\log K_i$ de los parálogos α/β simplificando la manipulación de los datos, el resultado permitió clasificar a las moléculas con preferencias a un parálogo, si el valor de la relación es igual a uno significa que las moléculas no tienen preferencia por algún parálogo; si el valor de la relación es mayor a 1.2 las moléculas tienen preferencias al parálogo beta; finalmente, que si los valores de la relación son menores a 0.7, la preferencia es para el parálogo alfa.

En la Figura 21 se muestran las operaciones realizadas para la clasificación de los parálogos de la proteína GSK3 β ya antes mencionada, el total de moléculas obtenidas con preferencia al parálogo GSK3 β son 150 indicadas en la (Tabla 2) con preferencia GSK3 α hay un total de 50 moléculas (Tabla 3). En la (Tabla 4) se ilustran las 50 moléculas que no tienen preferencia. El total de moléculas obtenidas en este análisis es de 250.

$$\frac{\alpha}{\beta} = 1 \text{ Inespecífico}$$

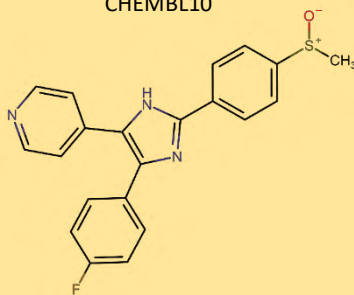
CHEMBL101311



$$\frac{3981.07\text{nM}}{3981.07\text{nM}} = 1$$

$$\frac{\alpha}{\beta} > \text{Favorable } \beta$$

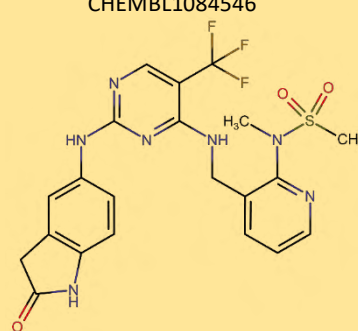
CHEMBL10



$$\frac{3981.07\text{nM}}{1584.89\text{nM}} = 2.51$$

$$\frac{\alpha}{\beta} < 1 \text{ Favorable } \alpha$$

CHEMBL1084546



$$\frac{79.43\text{nM}}{100\text{nM}} = 0.79$$

Figura 21.-Clasificación de los inhibidores.

(Edición propia, software photoshop)

Tabla 2.-Inhibidores preferente GSK3β.

N°	Código Chem	Ki	clave	α/β	N°	Código Chem	Ki	clave	α/β
1	CHEMBL10	1584.89	molchem001	2.5	51	CHEMBL1979855	1258.93	molchem053	6.3
2	CHEMBL1241473	6.31	molchem003	501.1	52	CHEMBL1979883	2511.89	molchem054	3.1
3	CHEMBL1242373	3162.28	molchem004	2.5	53	CHEMBL1980142	6.31	molchem055	3.1
4	CHEMBL127898	2511.89	molchem005	3.1	54	CHEMBL1980144	2511.89	molchem056	3.1
5	CHEMBL1421720	501.19	molchem006	7.9	55	CHEMBL1980167	316.23	molchem057	2.5
6	CHEMBL1516890	1995.26	molchem007	3.9	56	CHEMBL1980391	19.95	molchem058	2.5
7	CHEMBL1624529	31.62	molchem008	251.2	57	CHEMBL1981045	630.96	molchem059	12.5
8	CHEMBL1682345	39	molchem009	2.5	58	CHEMBL1982135	1995.26	molchem060	3.9
9	CHEMBL1682552	10	molchem010	6.31	59	CHEMBL1982563	1000	molchem061	6.3
10	CHEMBL1725279	316.23	molchem011	3.9	60	CHEMBL1982924	501.19	molchem062	2.5
11	CHEMBL17370	158.49	molchem012	19.9	61	CHEMBL1983006	1584.89	molchem063	3.1
12	CHEMBL1964687	1584.89	molchem013	5.0	62	CHEMBL1983025	1584.89	molchem064	5.0
13	CHEMBL1964692	1995.26	molchem014	3.9	63	CHEMBL1983449	158.49	molchem065	3.9
14	CHEMBL1965589	3162.28	molchem015	2.5	64	CHEMBL1984296	1584.89	molchem066	2.5
15	CHEMBL1965988	3162.28	molchem016	2.5	65	CHEMBL1984402	1258.93	molchem067	6.3
16	CHEMBL1966722	501.19	molchem017	3.1	66	CHEMBL1984432	501.19	molchem068	15.8
17	CHEMBL1967302	100	molchem018	6.3	67	CHEMBL1984700	2511.89	molchem069	3.1
18	CHEMBL1967887	2511.89	molchem019	2.5	68	CHEMBL1985681	31.62	molchem070	2.5
19	CHEMBL1968406	1000	molchem020	7.9	69	CHEMBL1986186	2511.89	molchem071	2.5
20	CHEMBL1969049	100	molchem021	2.5	70	CHEMBL1986328	630.96	molchem072	9.9
21	CHEMBL1969372	630.96	molchem022	3.9	71	CHEMBL1986781	79.43	molchem073	7.9
22	CHEMBL1969506	2511.89	molchem023	3.1	72	CHEMBL1987009	1995.26	molchem074	2.5
23	CHEMBL1969561	15.85	molchem024	2.5	73	CHEMBL1987430	6.31	molchem075	794.2
24	CHEMBL1969664	100	molchem025	25.1	74	CHEMBL1987679	630.96	molchem076	3.9
25	CHEMBL1969735	3162.28	molchem026	2.5	75	CHEMBL1987793	2511.89	molchem077	3.1
26	CHEMBL1969755	2511.89	molchem027	3.1	76	CHEMBL1988141	100	molchem078	3.1
27	CHEMBL1970083	1000	molchem028	7.9	77	CHEMBL1988387	199.53	molchem079	39.8
28	CHEMBL1970903	10	molchem029	12.5	78	CHEMBL1988581	79.43	molchem080	5.0
29	CHEMBL1971149	1995.26	molchem030	3.9	79	CHEMBL1989646	63.1	molchem081	2.5
30	CHEMBL1971186	199.53	molchem031	19.9	80	CHEMBL1989805	251.19	molchem082	6.3
31	CHEMBL1971245	1584.89	molchem032	5.0	81	CHEMBL1990254	63.1	molchem083	3.9
32	CHEMBL1972220	1258.93	molchem033	3.1	82	CHEMBL1990288	630.96	molchem084	5.0
33	CHEMBL1972346	1000	molchem034	3.1	83	CHEMBL1990885	2511.89	molchem085	2.5
34	CHEMBL1972355	398.11	molchem035	2.5	84	CHEMBL1990904	398.11	molchem086	2.5
35	CHEMBL1972934	15848.93	molchem036	2.5	85	CHEMBL1990912	1000	molchem087	7.9
36	CHEMBL1973098	794.33	molchem037	2.5	86	CHEMBL1991063	31.62	molchem088	3.1
37	CHEMBL1973211	794.33	molchem038	9.9	87	CHEMBL1991143	3162.28	molchem089	2.5
38	CHEMBL1973359	398.11	molchem039	2.5	88	CHEMBL1991188	39.81	molchem090	2.5
39	CHEMBL1974803	3162.28	molchem040	2.5	89	CHEMBL1991285	1258.93	molchem091	5.0
40	CHEMBL1974870	15.85	molchem041	3.1	90	CHEMBL1991429	50.12	molchem092	5.0
41	CHEMBL1975233	1995.26	molchem042	3.9	91	CHEMBL1991734	1258.93	molchem093	3.1
42	CHEMBL1975900	63.1	molchem043	3.1	92	CHEMBL1992042	398.11	molchem094	5.0
43	CHEMBL1976240	79.43	molchem044	19.9	93	CHEMBL1992796	3162.28	molchem095	2.5
44	CHEMBL1976732	630.96	molchem045	12.5	94	CHEMBL1994321	1995.26	molchem096	3.9
45	CHEMBL1977135	199.53	molchem046	6.3	95	CHEMBL1994693	1995.26	molchem097	3.9
46	CHEMBL1977148	1995.26	molchem047	3.9	96	CHEMBL1995172	5011.87	molchem098	3.9
47	CHEMBL1978200	1258.93	molchem048	6.3	97	CHEMBL1995927	501.19	molchem099	15.8
48	CHEMBL1978371	501.19	molchem049	7.9	98	CHEMBL1996831	794.33	molchem100	9.9
49	CHEMBL1979577	1258.93	molchem051	6.3	99	CHEMBL1997023	398.11	molchem101	6.3
50	CHEMBL1979718	630.96	molchem052	2.5	100	CHEMBL1997193	316.23	molchem102	102

N°	Código Chem	Ki	clave	α/β	N°	Código Chem	Ki	clave	α/β
101	CHEMBL1997349	794.33	molchem103	9.9	126	CHEMBL2006188	1258.93	molchem128	3.1
102	CHEMBL1998585	1000	molchem104	6.3	127	CHEMBL2006263	100	molchem129	3.1
103	CHEMBL1999126	1584.89	molchem105	5.0	128	CHEMBL2006580	1258.93	molchem130	6.3
104	CHEMBL1999428	158.49	molchem106	2.5	129	CHEMBL2006933	125.89	molchem131	3.1
105	CHEMBL1999496	398.11	molchem107	2.5	130	CHEMBL2007002	398.11	molchem132	19.9
106	CHEMBL1999590	100	molchem108	6.0	131	CHEMBL2007064	1584.89	molchem133	5.1
107	CHEMBL1999918	6.31	molchem109	3.9	132	CHEMBL210887	1995.26	molchem134	3.9
108	CHEMBL2000335	316.23	molchem110	25.1	133	CHEMBL213505	1995.26	molchem135	3.9
109	CHEMBL2000408	3162.28	molchem111	2.5	134	CHEMBL215152	1995.26	molchem136	2.5
110	CHEMBL2000429	2511.89	molchem112	3.1	135	CHEMBL223367	158.49	molchem137	63.0
111	CHEMBL2000481	1995.26	molchem113	3.9	136	CHEMBL223460	794.33	molchem138	12.5
112	CHEMBL2000508	1584.89	molchem114	5.1	137	CHEMBL243088	12.59	molchem139	630.8
113	CHEMBL2001257	3162.28	molchem115	2.5	138	CHEMBL296586	19.95	molchem140	398.1
114	CHEMBL2001751	63.1	molchem116	6.0	139	CHEMBL302449	15.85	molchem141	3.9
115	CHEMBL2002240	1258.93	molchem117	6.3	140	CHEMBL316264	630.96	molchem142	12.5
116	CHEMBL2002446	2511.89	molchem118	3.1	141	CHEMBL336961	10	molchem143	501.2
117	CHEMBL2002450	31622.78	molchem119	3.1	142	CHEMBL360847	794.33	molchem144	2.5
118	CHEMBL2002723	1258.93	molchem120	2.5	143	CHEMBL371206	1995.26	molchem145	3.9
119	CHEMBL2002726	63.1	molchem121	5.0	144	CHEMBL373598	251.19	molchem146	39.8
120	CHEMBL2003768	501.19	molchem122	6.3	145	CHEMBL413779	199.53	molchem147	39.8
121	CHEMBL2004365	1000	molchem123	7.9	146	CHEMBL462120	2511.89	molchem148	3.1
122	CHEMBL2004544	630.96	molchem124	2.5	147	CHEMBL474432	1584.89	molchem149	2.5
123	CHEMBL2004716	3162.28	molchem125	2.5	148	CHEMBL514499	1995.26	molchem150	3.9
124	CHEMBL2004771	3162.28	molchem126	2.5	149	CHEMBL539474	50.12	molchem152	3.9
125	CHEMBL2004934	125.89	molchem127	3.9	150	CHEMBL546797	1258.93	molchem153	2.5

Tabla 3.-Inhibidores preferente GSK3 α .

N°	Código Chem	Ki	clave	α/β	N°	Código Chem	Ki	clave	α/β
1	CHEMBL1419458	1258.93	alfa1	0.314	26	CHEMBL1989029	2511.89	alfa26	0.398
2	CHEMBL1964644	79.43	alfa2	0.398	27	CHEMBL1991180	100	alfa27	0.398
3	CHEMBL1965495	316.23	alfa3	0.199	28	CHEMBL1991846	630.96	alfa28	0.125
4	CHEMBL1967564	19.95	alfa4	0.125	29	CHEMBL1992732	6.31	alfa29	0.158
5	CHEMBL1968070	25.12	alfa5	0.251	30	CHEMBL1992922	398.11	alfa30	0.251
6	CHEMBL1969126	199.53	alfa6	0.158	31	CHEMBL1993661	158.49	alfa31	0.199
7	CHEMBL1970913	50.12	alfa7	0.398	32	CHEMBL1994538	6.31	alfa32	0.398
8	CHEMBL1971029	501.19	alfa8	0.158	33	CHEMBL1994669	12.59	alfa33	0.002
9	CHEMBL1971801	199.53	alfa9	0.125	34	CHEMBL1996510	39.81	alfa34	0.009
10	CHEMBL1972276	2511.89	alfa10	0.398	35	CHEMBL1997129	630.96	alfa35	0.199
11	CHEMBL1973516	501.19	alfa11	0.316	36	CHEMBL1997534	19.95	alfa36	0.398
12	CHEMBL1973711	100	alfa12	0.398	37	CHEMBL1999718	6.31	alfa37	0.001
13	CHEMBL1974935	398.11	alfa13	0.316	38	CHEMBL2000104	15.85	alfa38	0.398
14	CHEMBL1975256	50.12	alfa14	0.316	39	CHEMBL2001987	12.59	alfa39	0.001
15	CHEMBL1975787	19.95	alfa15	0.099	40	CHEMBL2002599	10	alfa40	0.001
16	CHEMBL1976090	199.53	alfa16	0.125	41	CHEMBL2002613	12.59	alfa41	0.316
17	CHEMBL1977138	251.19	alfa17	0.398	42	CHEMBL2003817	25.12	alfa42	0.398
18	CHEMBL1981720	39.81	alfa18	0.005	43	CHEMBL2004515	10000	alfa43	0.398
19	CHEMBL1981744	15.85	alfa19	0.001	44	CHEMBL2005886	794.33	alfa44	0.251
20	CHEMBL1982880	2511.89	alfa20	0.398	45	CHEMBL2205426	21	alfa45	0.205
21	CHEMBL1985074	158.49	alfa21	0.398	46	CHEMBL248757	911	alfa46	0.129
22	CHEMBL1985580	2511.89	alfa22	0.398	47	CHEMBL274064	2511.89	alfa47	0.199
23	CHEMBL1987054	50.12	alfa23	0.398	48	CHEMBL437747	125.89	alfa48	0.050
24	CHEMBL1988153	6.31	alfa24	0.001	49	CHEMBL443962	794.33	alfa49	0.251
25	CHEMBL1988173	158.49	alfa25	0.199	50	CHEMBL583144	1000	alfa50	0.251

Tabla 4.-Inhibidores sin preferencia.									
N°	Código Chem	Ki	clave	α/β	N°	Código Chem	Ki	clave	α/β
1	CHEMBL1081312	1995,26	SINP1	1	26	CHEMBL1984548	7.943	SINP26	1
2	CHEMBL1233887	1000	SINP2	1	27	CHEMBL2001641	12589,25	SINP27	1
3	CHEMBL1825138	3981,07	SINP3	1	28	CHEMBL482967	10000	SINP28	1
4	CHEMBL191003	158,49	SINP4	1	29	CHEMBL1992536	63,1	SINP29	1
5	CHEMBL1964290	25,12	SINP5	1	30	CHEMBL2000393	5011,87	SINP30	1
6	CHEMBL1964399	6309,57	SINP6	1	31	CHEMBL1983630	15848,93	SINP31	1
7	CHEMBL1965660	6,31	SINP7	1	32	CHEMBL2005449	50,12	SINP32	1
8	CHEMBL1970340	50118,72	SINP8	1	33	CHEMBL2007375	39,81	SINP33	1
9	CHEMBL1971519	19952,62	SINP9	1	34	CHEMBL1987815	39810,72	SINP34	1
10	CHEMBL1972119	10	SINP10	1	35	CHEMBL1989043	3162,28	SINP35	1
11	CHEMBL1973720	398,11	SINP11	1	36	CHEMBL1989471	251,19	SINP36	1
12	CHEMBL1974416	794,33	SINP12	1	37	CHEMBL1991782	79432,82	SINP37	1
13	CHEMBL1975121	630,96	SINP13	1	38	CHEMBL1997846	1258,93	SINP38	1
14	CHEMBL1975212	39810,72	SINP14	1	39	CHEMBL1999414	3162,28	SINP39	1
15	CHEMBL1975647	5.012	SINP15	1	40	CHEMBL1999778	2511,89	SINP40	1
16	CHEMBL1975921	63095,73	SINP16	1	41	CHEMBL2000029	125,89	SINP41	1
17	CHEMBL1975923	1995,26	SINP17	1	42	CHEMBL2000832	251,19	SINP42	1
18	CHEMBL1977041	1000	SINP18	1	43	CHEMBL2003456	2511,89	SINP43	1
19	CHEMBL1977931	63095,73	SINP19	1	44	CHEMBL2003785	50118,72	SINP44	1
20	CHEMBL1983111	398,11	SINP20	1	45	CHEMBL2004513	3162,28	SINP45	1
21	CHEMBL1983575	794,33	SINP21	1	46	CHEMBL2007296	1258,93	SINP46	1
22	CHEMBL1986143	1258,93	SINP22	1	47	CHEMBL207995	1584,89	SINP47	1
23	CHEMBL1983715	1584,89	SINP23	1	48	CHEMBL246970	630,96	SINP48	1
24	CHEMBL1983980	630,96	SINP24	1	49	CHEMBL340384	125,89	SINP49	1
25	CHEMBL1984500	79432,82	SINP25	1	50	CHEMBL560813	3162,28	SINP50	1

7.4. Síntesis *in silico* de los inhibidores.

La síntesis *in silico* de las 250 moléculas se realizó con el programa GaussView 3.07 obteniéndose la estructura tridimensional en el formato *.gjf. Posteriormente se realizó la optimización geométrica de cada una de las moléculas para obtener una geometría estable. En la Figura 22 se muestra el ejemplo de construcción de una molécula completa partiendo de la figura bidimensional, posteriormente pasa por la estructura tridimensional, se indican en rojo a los átomos de oxígenos, en azul a los átomos de nitrógenos, en gris los átomos carbonos y por último en blanco los átomos de hidrógenos. El proceso realizado para la síntesis *in silico* fue usando la clave de verificación CHEMBL1241473 y se renombró a molchem003, la primera parte se tomó de mol=molécula y chem= chemical, el número 003 corresponde al orden de aparición de las moléculas con el fin de poder identificar rápidamente a las moléculas.

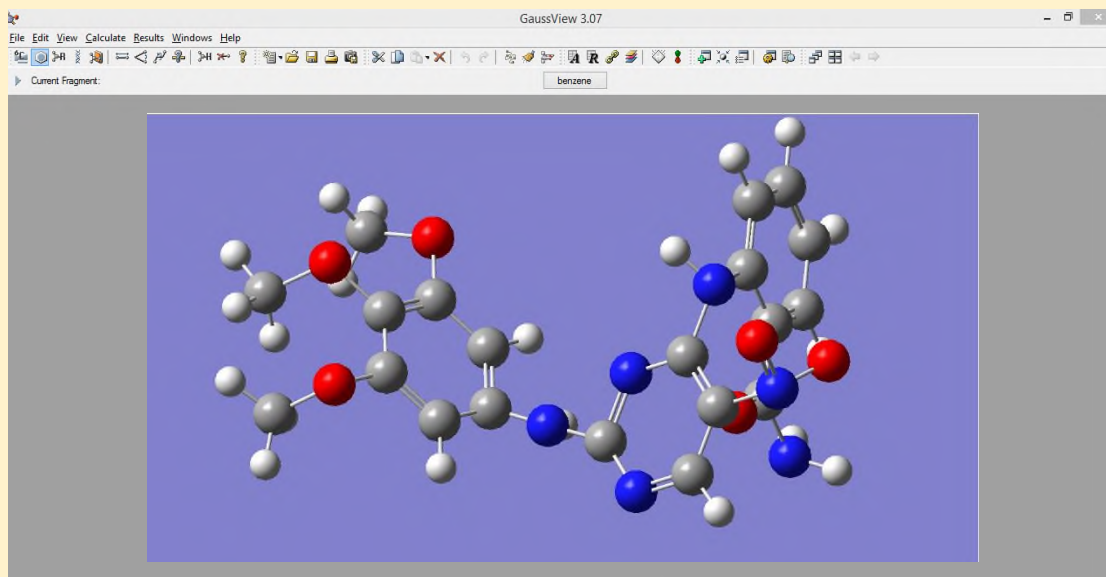
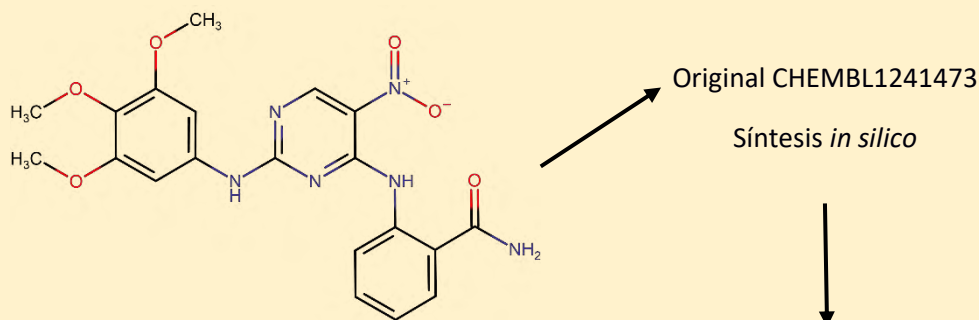


Figura 22.-Síntesis in silico.

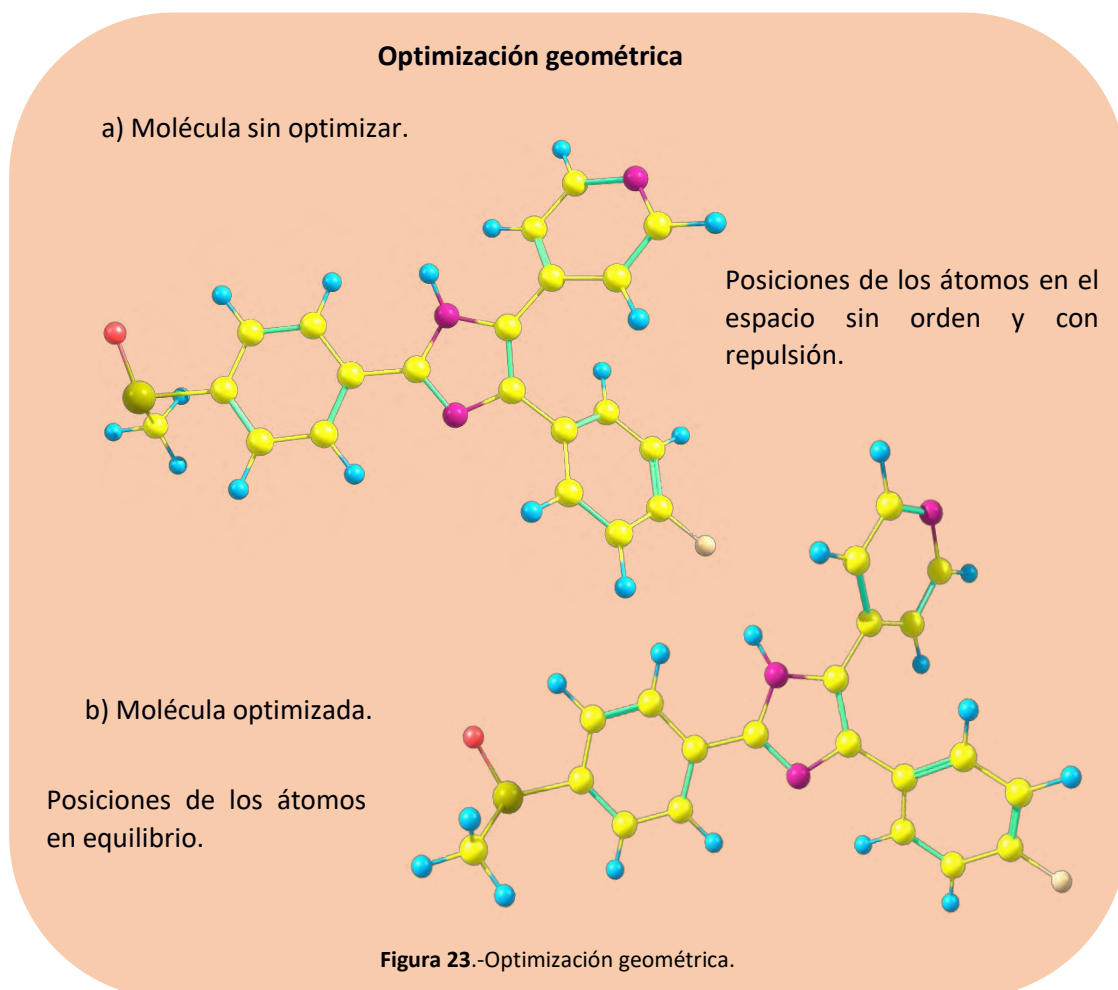
(Edición propia, software Photoshop)

7.5. Optimización geométrica de los inhibidores.

La optimización geométrica utilizada permitió localizar la conformación estable de cada una de las moléculas por medio el funcional B3LYP y la base 6-31G* utilizando el paquete Gaussian 03.

El programa permite a través de cálculos cuánticos ajustar las coordenadas de los átomos de las moléculas para obtener una geometría tridimensional, a partir de sus propiedades electrónicas lo que permite proponer una estructura molecular en equilibrio con respecto a sí misma, cabe mencionar que antes de realizar el cálculo las moléculas tenían un formato GJF y su salida es en un formato OUT. En el ejemplo Figura 23 se tienen dos veces a la misma molécula, donde los colores indicados en amarillo representan a los carbonos, rojo los

oxígenos, azul hidrógenos y morado los nitrógenos; en primera instancia en el a) tenemos la primera forma de la molécula antes de hacer la optimización y la segunda, en el b), después de una optimización geométrica mostrando cambios en su estructura.



7.6. Cálculos descriptores cuánticos.

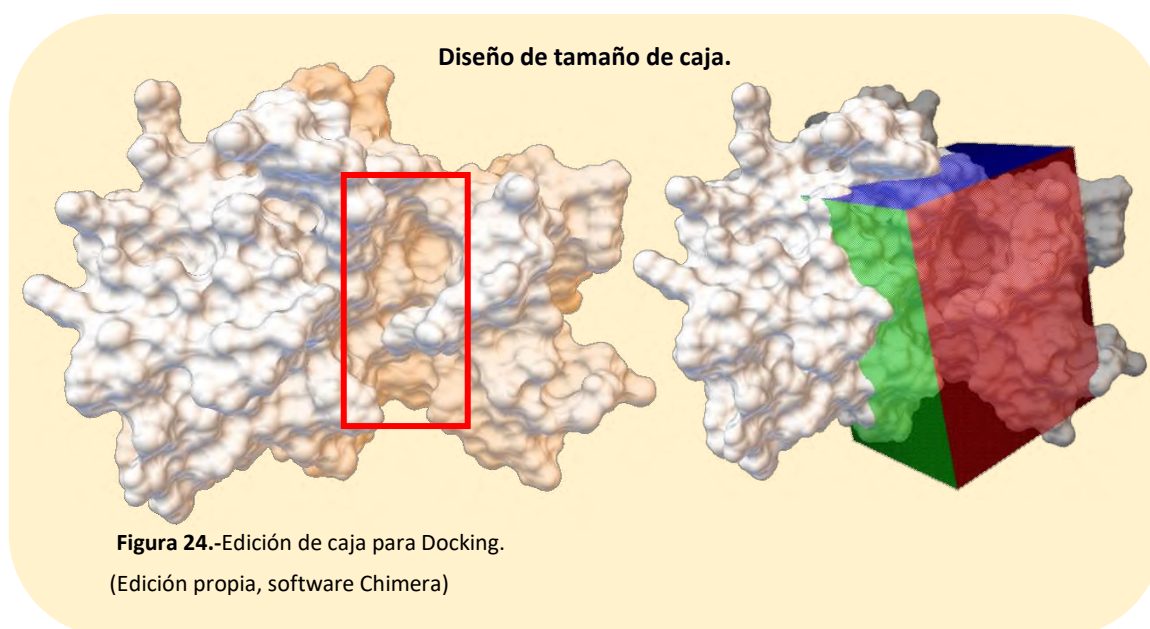
Posteriormente, se obtuvieron los siguientes descriptores cuánticos: Logaritmo de (K_i) junto con la energía Hartree-Fock, del cual se calculó su logaritmo $\log(-E_c)$; el momento dipolar en las tres direcciones cartesianas (D_x , D_y y D_z); el momento dipolar total (DT) este se calculó con la suma de los cuadrados de las tres direcciones cartesianas; las energías de los orbitales moleculares frontera (HOMO y LUMO); la dureza (η), la cual se calculó con la diferencia del valor de HOMO menos el valor de LUMO entre dos; la electronegatividad (χ) se calculó

con la suma del valor de HOMO más el valor de LUMO entre dos, y la Electrofilicidad (ω), la cual se obtuvo al calcular la electronegatividad elevado al cuadrado entre el producto de dos por el valor de la dureza. Estos datos se extraen de los resultados de la optimización geométrica de cada molécula, en formato OUT el cual contiene los datos mencionados para realizar cada uno de los cálculos.

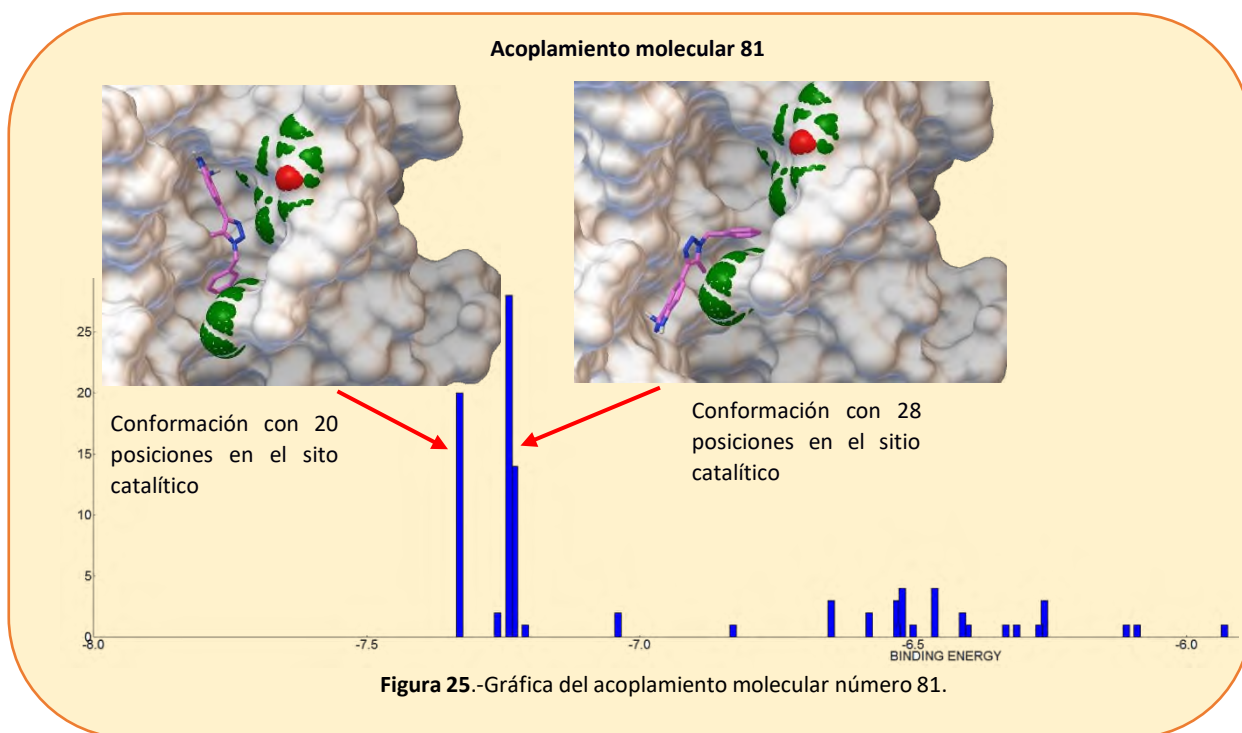
7.7. Acoplamiento molecular

El acoplamiento molecular se realizó a través del software AutoDockTools-1.5.6, se usó a la proteína GSK3 β obtenida por los cálculos de dinámica molecular y las moléculas optimizadas geoméricamente a través de Gaussian 03.

Como parte del acoplamiento molecular, se simula la recepción de los inhibidores que compiten con la molécula de ATP. Se utilizó un tamaño de caja de 50 Å en X, 66 Å en Y y 84 Å en Z, esta caja incluye al sitio activo del inhibidor. El acoplamiento molecular se realizó para cada una de las 250 moléculas obtenidas, obteniendo 100 conformaciones dentro de esta caja como se muestra en la Figura 24, en la imagen de la izquierda, con el cuadro rojo se visualiza el sitio de inhibición de la proteína y la imagen de la derecha, se muestra el tamaño de la caja en los inhibidores interactuaron con GSK3 β , donde X está representado por el color rojo Y con color verde y es Z con azul.

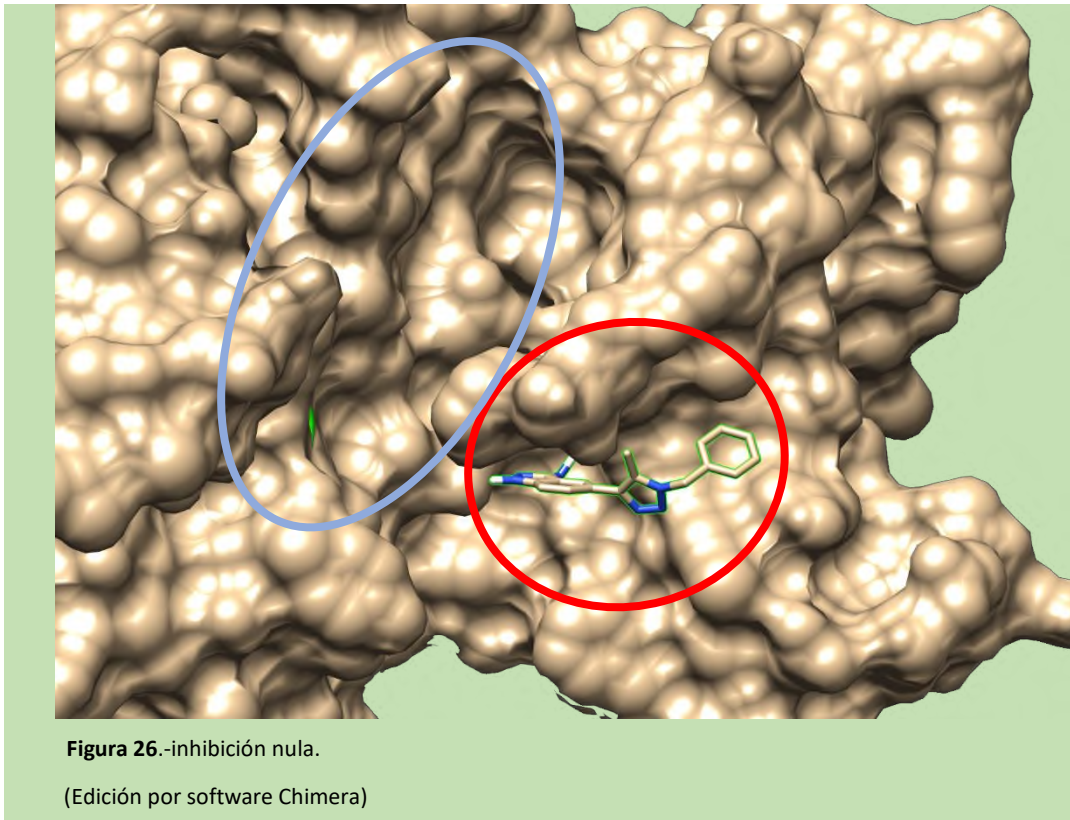


El acoplamiento molecular se llevó a cabo de manera dirigida, ya que el sitio de inhibición en la proteína es conocido y los inhibidores utilizados han sido reportados como competidores de ATP. En la Figura 25 se muestran los resultados del análisis del acoplamiento molecular de la molécula 81, la gráfica de barras representa a 100 posiciones adoptadas por el inhibidor dentro del sitio. Para el ejemplo, la gráfica de barras se observa que el inhibidor presentó 20 posiciones con una energía de -7.33 kcal/mol (primera barra a la izquierda). Cada una de las barras indica el número de posiciones obtenidas en el sitio. Una segunda barra presenta mayor número de conformaciones (tercera barra de izquierda a derecha) con 28 conformaciones dentro del sitio catalítico con una energía de -7.24 kcal/mol.

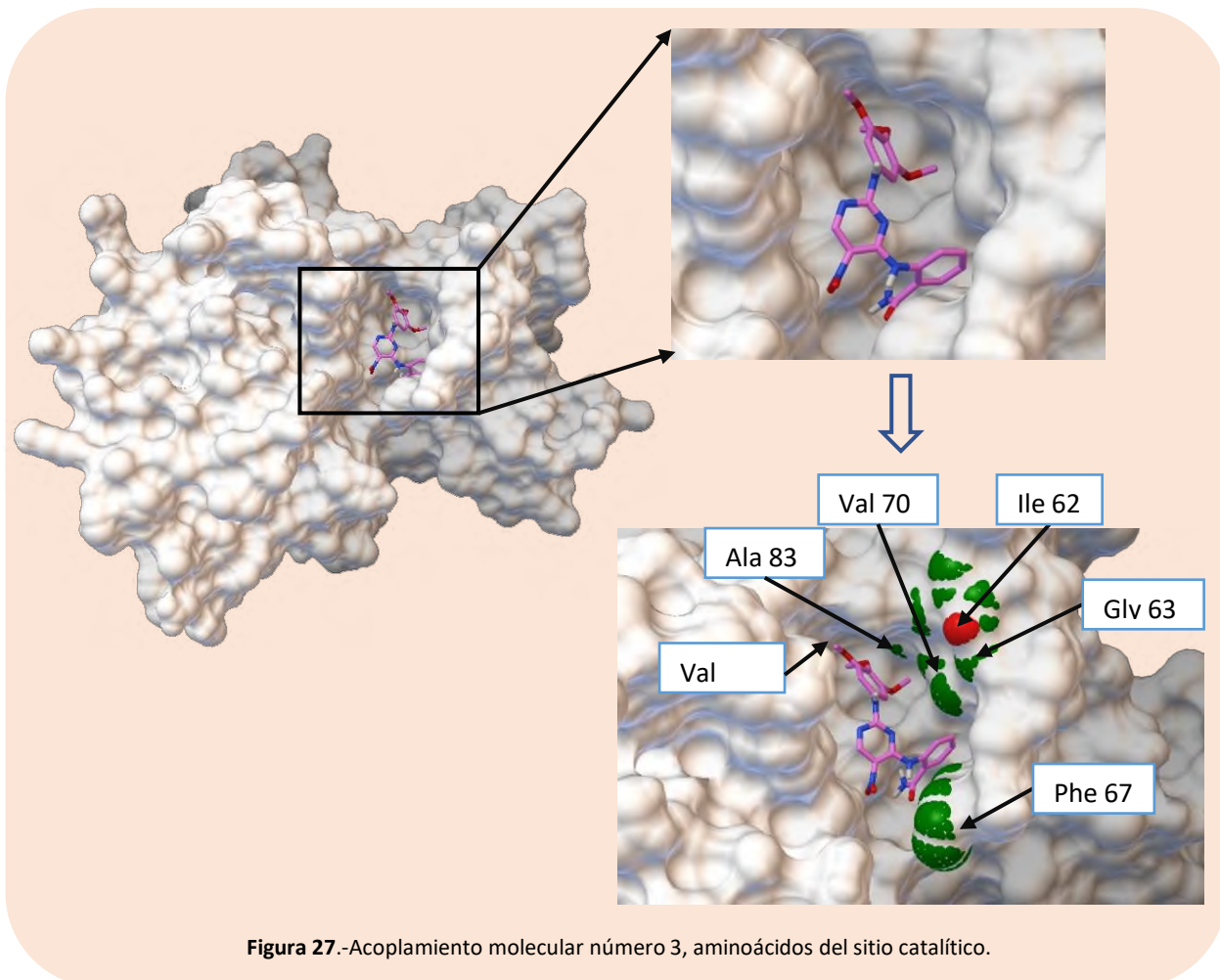


Es importante destacar que el sitio de unión es de gran importancia debido a que buscamos competidores del ATP, en caso de no estar en el sitio activo se tiene que rechazar dichas conformaciones, de igual manera, aunque existan moléculas que tengan una mayoría de conformaciones y éstas no interactuaron con el sitio de unión también fueron rechazadas pues es importante que las moléculas se posicionarán sobre el surco donde se ha indicado que esta el sitio de inhibición. Un ejemplo del rechazo de una conformación mayoritaria se

observa en la Figura 26, donde el inhibidor esta fuera del sitio catalítico indicado con el círculo rojo, en cual se observa en la parte exterior del sitio y sobresaliendo en la parte inferior de la proteína, siguiendo la idea de lo antes mencionado, esta conformación de esta proteína se rechaza, claramente marcado con el óvalo azul se presenta el lugar en el que debería ubicarse la molécula para lograr la inhibición de la GSK3 β .



Un segundo ejemplo se presenta en la Figura 27, en donde se muestra el acoplamiento molecular del inhibidor 3 (Tabla 2), en la figura se señaló con un recuadrado negro la interacción de proteína-ligando, los aminoácidos relevantes se marcaron con flechas y colores estos son de gran importancia para un acoplamiento molecular adecuado, los cuales fueron Isoleucina 62, Glicina 63, Fenilalanina 67, Valina 70, Alanina 83 y Valina110.



7.8. Diferencias de los parálogos de la proteína GSK3.

La serie de aminoácidos que permiten la inhibición de los parálogos se encontraron en la bibliografía “base estructural de la inhibición de GSK3 por fosforilación N-terminal y por receptor Wnt LRP6”. En las proteínas GSK3 α y GSK3 β encontramos las mismas, series de aminoácidos diferenciados únicamente por un ácido glutámico en alfa y un ácido aspártico en beta, por este motivo podemos asumir que la presencia de estos aminoácidos puede influir en la selectividad de los inhibidores, una teoría que cobra fuerza debido a la presencia de inhibidores con preferencias para cada uno de los parálogos, así como la existencia inhibidores sin preferencia. En la Figura 28 se muestran la lista de todos los aminoácidos importantes presentes dentro del sitio catalítico, con colores diferentes, pero respetando los mismos colores para ambos parálogos, para que se aprecien las similitudes.

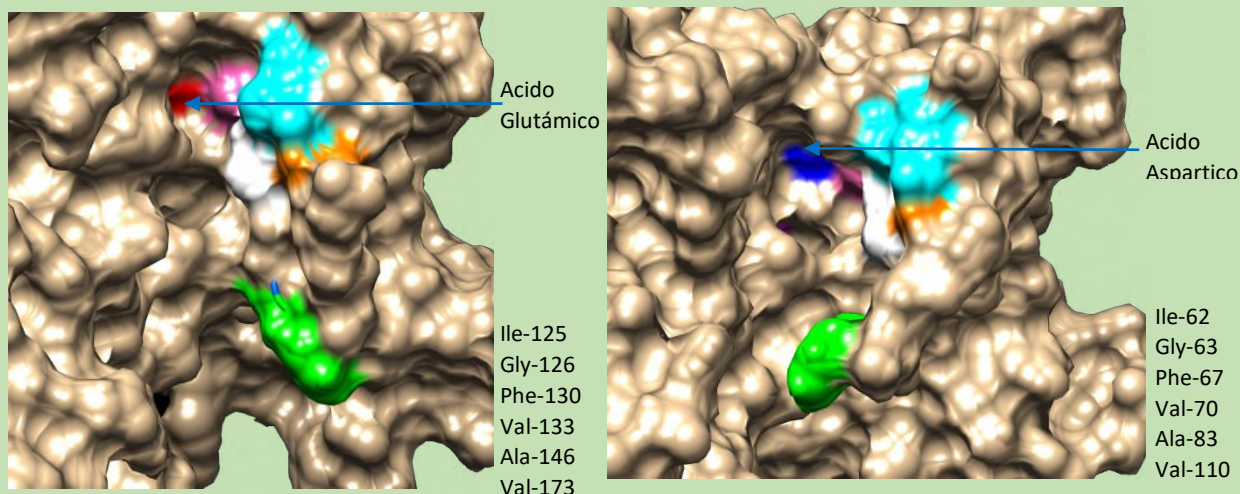
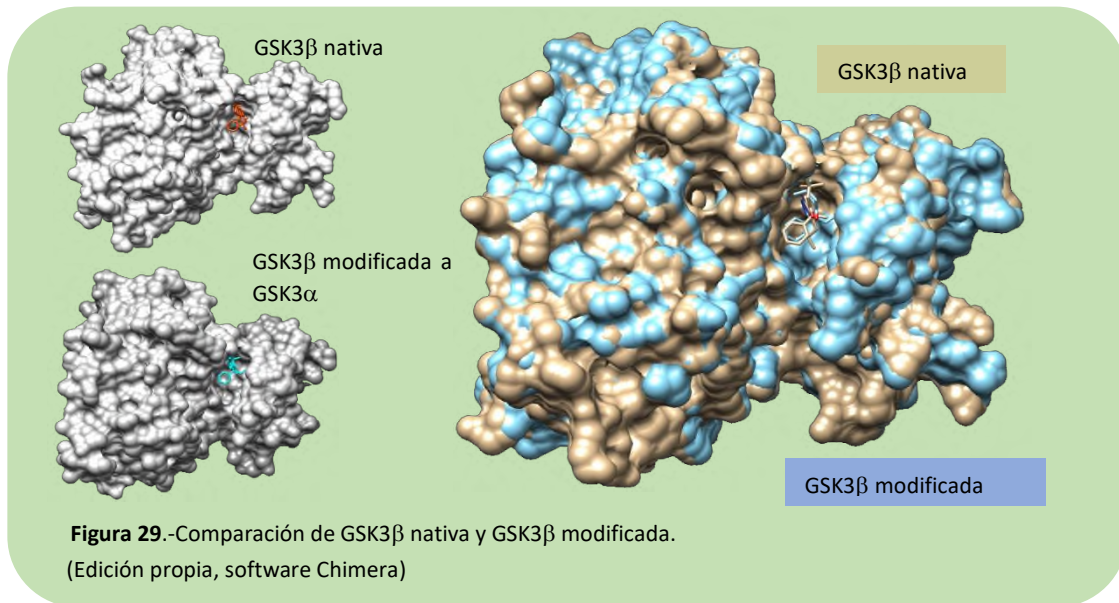


Figura 28.-Aminoácidos de GSK3 α y GSK3 β .

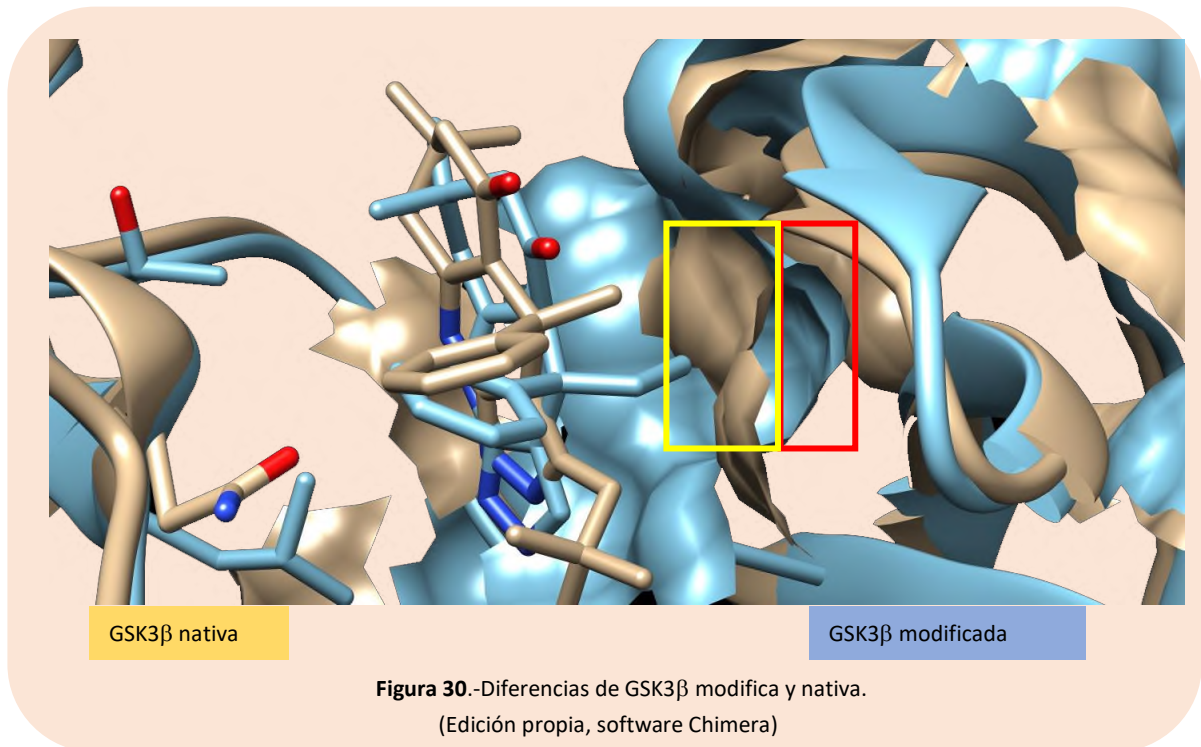
(Edición por software Chimera)

Debido a estas diferencias, se realizaron comparaciones de las proteínas con la estructura nativa de GSK3 β y otra modificada de beta donde el ácido aspártico es sustituido por el ácido glutámico para simular a la proteína GSK3 α , se observa una diferencia significativa en la cual se puede visualizar una especie de cavidad generada por la sustitución, por el momento se puede decir que es la diferencia visual más relevante dentro del sitio catalítico, al realizar este cambio, la conformación estructural de la proteína modificada es afectada. Comparando estas estructuras con su inhibidor se visualiza que tienen un posicionamiento que permite la interacción con la proteína, esto claramente se observa en la Figura 29, en ella podemos visualizar que al realizar una superposición de las proteínas con los inhibidores hay cambios significativos en la estructura de la proteína modificada, para poder entender la figura tenemos en la parte izquierda a la proteína GSK3 β (proteína nativa) mientras que la GSK3 β modificada, se ha cambiado el ácido aspártico por glutámico, al compararse dichas proteínas tenemos que el color dorado representa a la proteína nativa mientras que el azul representa a la proteína modificada.



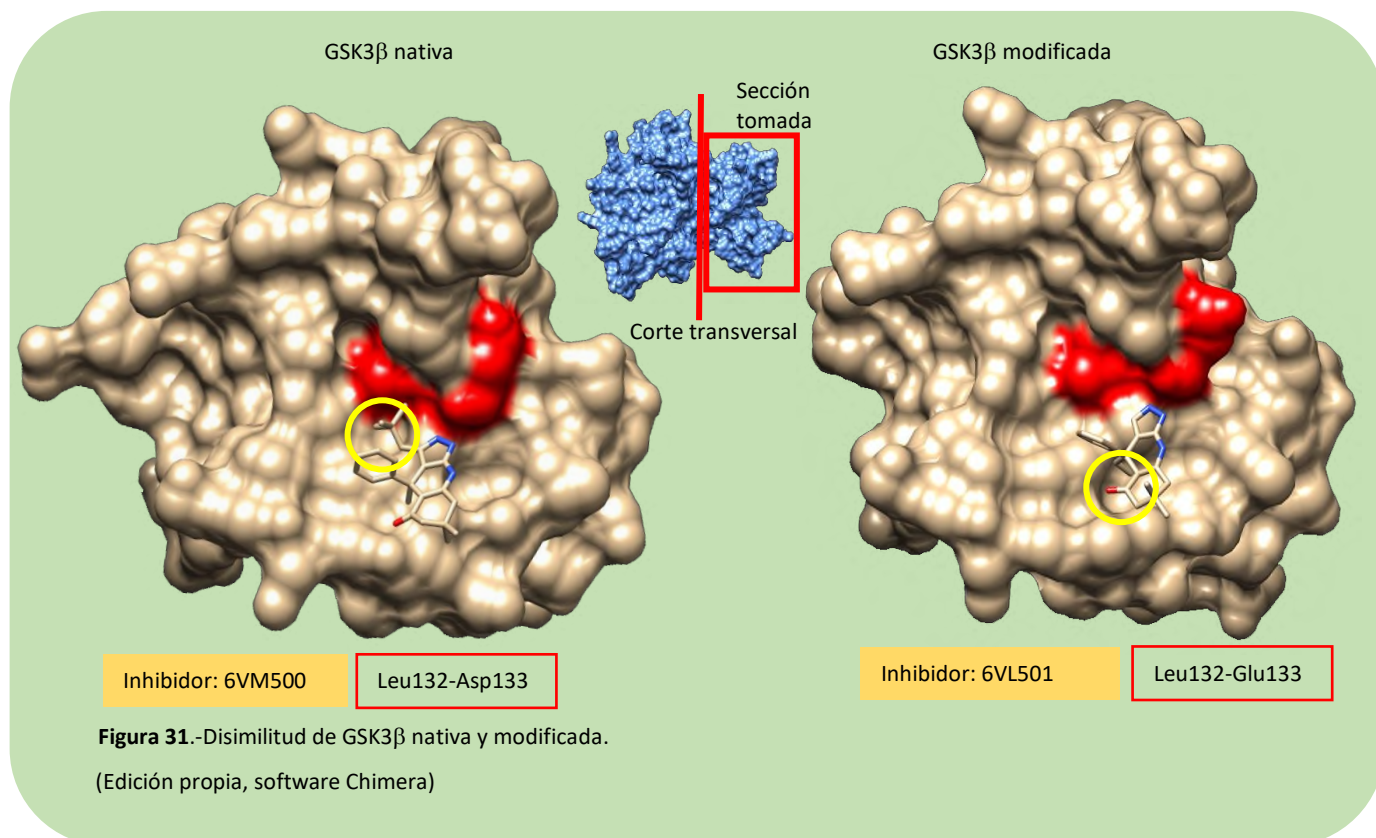
En la comparación entre la proteína de GSK3β nativa con la GSK3β modificada se detectó la formación de una cavidad que es solo visible en la proteína modificada, la cual representa a GSK3α, esto propicia que el inhibidor mantenga una afinidad dentro de esta cavidad, esto puede conferirle selectividad por GSK3α mientras que la GSK3β mantiene una sección completamente plana en el que no se genera una cavidad Figura 30, la formación de la cavidad en la GSK3β modificada se muestra con un recuadro rojo y en la GSK3β nativa no se genera esa cavidad mostrada en el recuadro amarillo, donde la proteína GSK3β nativa está representada en color dorado mientras que la modificada en color azul.

Podríamos destacar que este resultado es de vital importancia, de este modo podemos indicar que la selectividad de los inhibidores puede ser causada específicamente por la diferencia de un aminoácido dentro del sitio de inhibición, debido a que permite una distinta conformación estructural lo que admite que ciertos inhibidores tengan una interacción selectiva para los diferentes parálogos.



De este modo con la finalidad de demostrar que realmente sí existen diferencias importantes estructuralmente de un parólogo a otro, se realizó un corte transversal de las proteínas en donde se observa que la GSK3β tiene dentro de su sitio activo una cavidad ubicada en la parte central de la proteína, caso diferente con GSKβ modificada. Esto es fundamental sobre la selectividad de los inhibidores. La presencia del ácido glutámico en el parólogo alfa y el ácido aspártico en el parólogo beta les confiere una conformación estructural diferente lo cual permitió identificar que los inhibidores analizados e indicados como 6VM500 en la proteína nativa y 6VL502 en la proteína modificada sean específicos en estos ejemplos, por ende realizar una inhibición selectiva con inhibidores diseñados *in silico* puede ser alcanzable a pesar de tener el ~98 de similitud del sitio activo, cabe mencionar que en la Figura 31 se muestran estos cambios por medio de un círculo color amarillo, en él se marcan cavidades con el cual interactúan los inhibidores mencionados, los aminoácidos se marcaron con rojo junto con el aminoácido más cercano a este debido que permite ubicar las posiciones de la mejor manera en este ejemplo visual. Se puede observar a partir de los aminoácidos indicados en rojo dentro de la proteína, Leu132-Asp133 de origen de la proteína nativa, en el caso de la proteína modifica tenemos a

132Leu-133Glu indicados en rojo, la diferencias entre las distancias que se tiene con respecto al inhibidor ante estos aminoácidos son distintos.



7.9. Cálculo de Descriptores.

Los cálculos de los descriptores se realizaron a través del software PaDEL (página oficial: [//www.yapcsoft.com/dd/padeldescriptor/](http://www.yapcsoft.com/dd/padeldescriptor/)). PaDEL-Descriptor es un software para calcular descriptores moleculares y huellas dactilares. Estos descriptores y huellas digitales se calculan principalmente mediante el kit de desarrollo químico, el cual incluye descriptores de estado electro topológico de tipo átomo, volumen de McGowan, descriptores de relación de energía libre lineal molecular, recuentos de anillos, recuento de subestructuras químicas identificadas por Laggner y huellas dactilares binarias y recuento de subestructuras químicas identificadas por Klekota y Roth. Padel permite el cálculo de aproximadamente de 1875 descriptores, este programa requiere del complemento Java JRE versión 6, el programa PaDEL reconoce los archivos con extensión Mol, SMILES y MDL por lo que fue necesario realizar la conversión de

formato, para lo cual se utilizó el programa Babel. En una carpeta se almacenaron las 150 moléculas para que el programa permitiera la selección de las moléculas y se pudieran realizar los cálculos de los descriptores, el programa proporcionó un archivo en formato CVS, compatible con el programa Excel, el cual contiene las 150 moléculas en las conformaciones seleccionadas. En la Tabla 5 se observa, a manera de ejemplo, una fracción de la tabla donde se indica en la fila 1 de cada columna el nombre del descriptor, en la parte izquierda el número de molécula y los valores obtenidos de cada una de ellas, contando un total de 1800 descriptores.

Tabla 5.-Descriptores calculados.										
Molécula	ALogp2	AMR	apol	naAromAtom	nAromBond	nAtom	nH	nC	nN	nO
molchem11	0,21187609	2,5646	32,976344	14	16	25	8	14	2	1
molchem109	3,254416	25,5791	39,373551	15	17	28	7	14	2	1
molchem130	0,20629764	19,2737	35,636344	15	16	26	8	12	3	1
molchem58	0,44262409	32,7395	48,954102	17	17	39	14	18	4	1
molchem144	0,78074896	67,1262	74,756204	18	20	63	28	28	4	3
molchem07	0,29333056	32,6497	44,023895	12	12	36	15	17	3	1
molchem142	0,290521	5,045	48,282309	23	26	38	13	20	3	0
molchem67	3,29967225	43,16	36,452309	6	6	31	13	13	3	2
molchem143	0,01731856	23,8898	40,21993	15	17	30	10	15	3	1
molchem94	0,28376929	36,7784	69,935239	22	23	56	23	29	2	1
molchem131	0,05410276	16,0548	35,44993	14	16	27	10	13	2	1
molchem12	0,56580484	17,6627	54,385688	23	26	42	16	21	3	0
molchem98	0,85488516	16,5001	27,517551	11	12	20	7	10	2	0

Cabe mencionar que PaDEL consta de un componente de biblioteca y un componente de interfaz, el componente de biblioteca permite que se integre fácilmente en el software de relación de actividad de estructura cuantitativa para proporcionar la función de cálculo del descriptor, mientras que el componente de interfaz permite que se utilice como un software independiente, el software utiliza un patrón maestro / trabajador para aprovechar los múltiples núcleos de CPU que están presentes en la mayoría de las computadoras modernas para acelerar los cálculos de descriptores moleculares.

El programa contiene una serie de familias de descriptores divididas en subgrupos de todos los descriptores, un ejemplo de esto se puede ver en la Tabla 6, la cual contiene los descriptores más comunes, estos datos son

proporcionados por el propio programa y la información fue obtenida desde la página oficial Padel Software. Es posible seleccionar las familias de descriptores para el cálculo, en nuestro trabajo se utilizó el programa en los valores predeterminados por lo cual nos proporcionó más de 1800 descriptores por cada molécula.

Tabla 6.-Descriptores comunes.	
Familia	Descriptores
Constitucional	Grupos funcionales, peso molecular, recuentos simples, por ejemplo (número de átomos, enlaces, anillos)
Topológico	Pares de átomos, Índice de balaban, BCUT, Índices de contenido de información, Índices de forma Kappa, índices de conectividad de Kier y Hal, Índice de flexibilidad de Kier, Índices de forma de Kier, la caminata molecular cuenta, Índices randic, Índice de Wiener
Geométrico	Índice de gravitación, área de superficie molecular, volumen molecular, índices de sombra, superficie molecular accesible al disolvente
Electrostático	Área de superficie polar cargada, índices de carga topológica de Gálvez, capacidades de enlace de hidrógeno, cargos parciales máximos y mínimos, polarizabilidades moleculares
Huellas dactilares	Luz de día, teclas MDL, unidad
Hidrofóbico	Índices de aromaticidad, constante de sustituyente de Hansch, registro D, registro P
Estérico	Parámetro estérico de Charton, refractividad molar, paracaidista, parámetro taft steric
Cuánticos	Cargas, energías HOMO y LUMO, densidades de electrones orbitales, Super deslocalizabilidades, polarizabilidades átomo-átomo, polarizabilidades moleculares, momentos dipolares e índices de polaridad, energías.
Combinación	3D - MoRSE, Índices de estado electro topológico

7.11. Generar el modelo QSAR.

7.11.1. Entrenamiento y predicción.

Los 150 compuestos con preferencia por el parólogo beta se ordenaron de menor a mayor con respecto los valores de Log (K_i), los cuales se dividieron de forma azarosa en 2 subconjuntos, el de entrenamiento (66.1 %) y el de predicción (33.9%).

La división en subconjuntos de entrenamiento y predicción se realizó por medio del programa Octave Software, utilizando un script de separación aleatoria, utilizando los valores de log (K_i) como parámetro de separación. El conjunto original contó con 182 moléculas, originalmente son 150 de los cuales tomamos las conformaciones con base al acoplamiento molecular proporcionando las 182 conformaciones para la base de datos y se dividió en 9 subintervalos, en el primer subintervalo se encuentra 12 moléculas, en el segundo 6 moléculas y con 8, 16, 14, 21, 28, 43 y 32 moléculas en cada subintervalo. Cada uno de los subintervalos se dividió en 2 subconjuntos de forma azarosa. La distribución total de estos subconjuntos finalizó con 122 moléculas en el subconjunto de entrenamiento y 60 compuestos en el subconjunto de predicción. Este proceso se repite para encontrar n combinaciones diferentes.

7.11.2. Regresión lineal múltiple (RLM).

Se realizó una selección de los descriptores a través de la metodología paso a paso por medio del programa Octave para obtener una serie de regresiones lineales, el valor de R² de estas regresiones obtenidas oscilaron entre 0.40 a 0.86 como valor máximo obtenido de 182 moléculas con 921 descriptores en el cual el proceso consistía en ir eliminando descriptores hasta obtener un aumento en la regresión lineal, aunque fue posible mejorar la regresión lineal, la metodología de paso a paso por medio de Octave no fue suficiente para obtener una regresión que sobre pase del 0.86.

Adicionalmente, se realizaron las regresiones lineales múltiples con el programa SPSS, obteniéndose listas con los descriptores y un R², el programa propone listas con los cuales podemos seleccionar los mejores resultados de descriptores para obtener un R² =0.99, como por ejemplo la Tabla 8 muestra la lista de

descriptores y R^2 , el cual fue de importancia para este análisis con la finalidad de ir tomando una lista apta para obtener un regresión lineal que permita caracterizar a los descriptores más importantes para el modelo QSAR, es decir se genera un “Pool” de descriptores.

Tabla 8.-Resultado de SSPS.			
Resumen del modelo			
Modelo	R	R cuadrado	R cuadrado ajustado
BCUTp-1h	,351 ^a	0,123	0,118
ATSC6i	,449 ^b	0,201	0,192
MATS7e	,522 ^c	0,272	0,260
AATS8m	,563 ^d	0,317	0,301
BCUTw-1l	,604 ^e	0,365	0,347
minssCH2	,635 ^f	0,403	0,382
ATSC5c	,662 ^g	0,438	0,415
GATS4c	,683 ^h	0,466	0,441
hmax	,701 ⁱ	0,491	0,464
Poses	,724 ^j	0,525	0,497
SdssC	,734 ^k	0,539	0,508
JGI9	,746 ^l	0,557	0,525
MAXDN	,758 ^m	0,575	0,542
MATS2s	,768 ⁿ	0,590	0,555
BCUTc-1h	,780 ^o	0,609	0,573
MATS1e	,788 ^p	0,620	0,583
piPC10	,796 ^q	0,634	0,595
VE1_D	,803 ^r	0,644	0,604
ZMIC2	,811 ^s	0,657	0,616
ZMIC4	,818 ^t	0,670	0,628
AATSC4c	,826 ^u	0,681	0,639
AATSC8i	,833 ^v	0,694	0,651
VE3_Dzs	,843 ^w	0,711	0,668
AATS3m	,848 ^x	0,720	0,676
ZMIC5	,854 ^y	0,729	0,684
AATS2s	,860 ^z	0,740	0,695

De igual forma con las listas obtenidas por el programa SPSS se amplió estas, con los descriptores de posición, se incluyó el número de poses dentro del sitio catalítico, antes mencionado en la Sección 7.7. La importancia de que las moléculas estén ubicadas en el sitio catalítico, por estas razones el log (Ki) y las posiciones permitieron que los datos fueran simples de manipular debido a que aseguramos que exista la interacción de cada molécula con el sitio catalítico, por ende, obtuvimos datos que van 1 pose hasta 80 posiciones. La lista con 182 moléculas (lista inicial), se fue reduciendo al ir eliminando el número de posiciones, es decir se eliminaron todas las moléculas con n-posiciones y se calculó la regresión lineal múltiple con SSPS, esto nos proporcionó 11 listas partiendo con todas las posiciones hasta eliminar 11 posiciones obteniéndose como resultado la Tabla 9.

En la tabla 9 se resume la lista de descriptores (ver la Tabla 8) con un valor R^2 . Además, se nombró cada lista de acuerdo con la posición eliminada (sinp#). En la columna 1 se muestra el nombre de la lista, el valor de regresión R^2 obtenida por el programa SSPS, en la tercera fila tenemos el número de moléculas con pose eliminada, en la cuarta fila de la columna tenemos el número de descriptores que tiene esa lista y consecuentemente tenemos el número de moléculas totales utilizadas para regresión lineal múltiple, finalizando con las últimas dos filas de la primera columna con el número de moléculas que están en el entrenamiento y la predicción.

Con estos resultados se seleccionó una de las listas para obtener nuestro modelo QSAR, el cual fue utilizado durante todo este proceso para mejorar la RLM donde fue necesario realizar una gráfica para determinar que la lista era la adecuada para nuestro modelo. Si observamos Tabla 9 (en color amarillo) la columna con título Sip7 es decir , sin poses 7 está marcada por ser nuestra mejor opción debido a que esta contiene 13 descriptores con un valor de R^2 de 0,808 y que en comparación de las demás listas, éstas no sobrepasan el valor de R^2 igual a 0.780, sin embargo, debemos de tener en cuenta que no podemos tener un número excesivo de descriptores en nuestro modelo o muy pocos debido a que podemos tener un modelo que no pueda establecer una regla con poder predictivo y confiable (ecuación lineal).

Tabla 9.- Listas SSPS obtenidas.												
Nombre	sinp0	sinp1	sinp2	sinp3	sinp4	sinp5	sinp6	sinp7	sinp8	sinp9	sinp10	sinp11
R2	0,628	0,636	0,768	0,437	0,644	0,739	0,724	0,808	0,732	0,542	0,857	0,656
sin pose	0	1	2	3	4	5	6	7	8	9	10	11
# descriptores	13	12	20	8	12	17	13	13	11	8	19	11
# de moléculas	182	179	176	167	162	154	148	130	125	116	110	103
entrenamiento	122	120	118	112	109	103	99	88	83	78	75	70
prediccion	60	59	58	55	53	51	49	42	42	38	35	33

Para visualizar un poco mejor el motivo de la selección y complementando la Tabla 9, se graficó el número de poses con respecto al valor de R^2 obtenido por SSPS, por tal motivo en la Figura 32 se visualiza con una línea naranja el cambio obtenido al ir eliminando posiciones, por tal motivo se tuvo una duda el cual nos proporcionó la siguiente pregunta ¿cuál puede ser nuestro mejor selección para mejorar el modelo?, para responder esto podemos observar que al eliminar 10 posiciones nos provee con un valor de R^2 mayor, pero no se consideró para generar nuestro modelo por el siguiente motivo, en la Tabla 9 observamos que sinp10 nos proporciona un valor de $R^2=0,857$ con 19 descriptores, mientras que sinp7 provee un valor de $R^2=0.808$ con 13 descriptores, hay una diferencia de 6 descriptores, si por 10 posiciones tenemos un valor de $R^2=0.85$, con 7 tenemos un valor de $R^2=0.80$, esto indica que eliminando 7 posiciones puede proveernos de un mejor modelo para nuestro R^2 , por lo cual, no necesitamos muchos descriptores si con menos se logra una mejor regresión, por tal motivo sinp7 es la mejor selección.

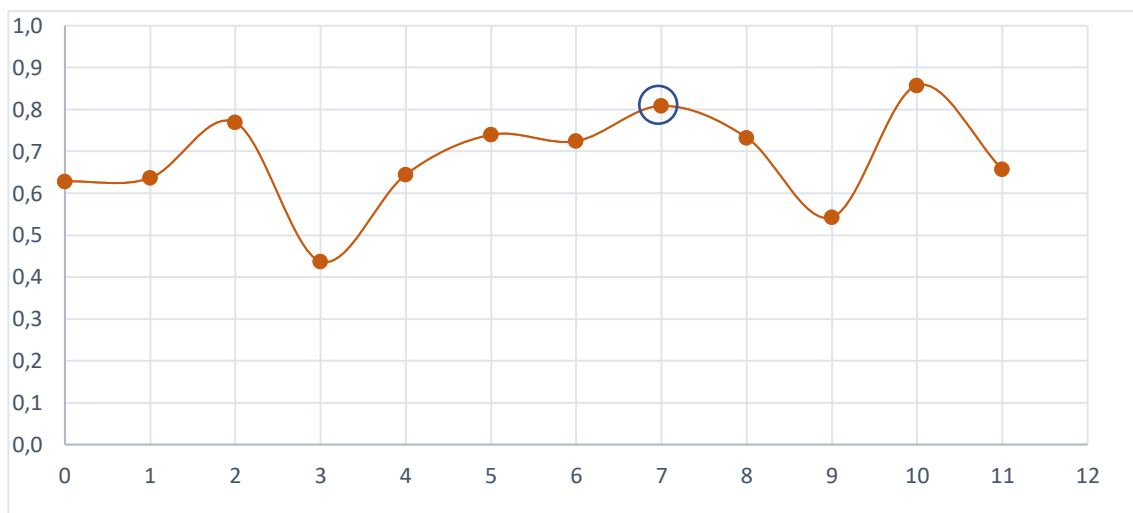


Figura 32.-Gráfica de regresiones lineales múltiples realizadas por SSPS.

Con respecto a lo anterior, si utilizamos los 13 descriptores con las 130 moléculas para generar el modelo de R^2 con los valores de SSPS, obtenemos una correlación como se muestra en la Figura 33, a pesar de tener un R^2 de 0.80 no es buena, este modelo provee una dispersión grande de puntos el cual visualizamos que el entrenamiento y la predicción, esto indica que no puede ser un buen modelo para ser predictivo para inhibidores selectivos, para la proteína GKS3 β .

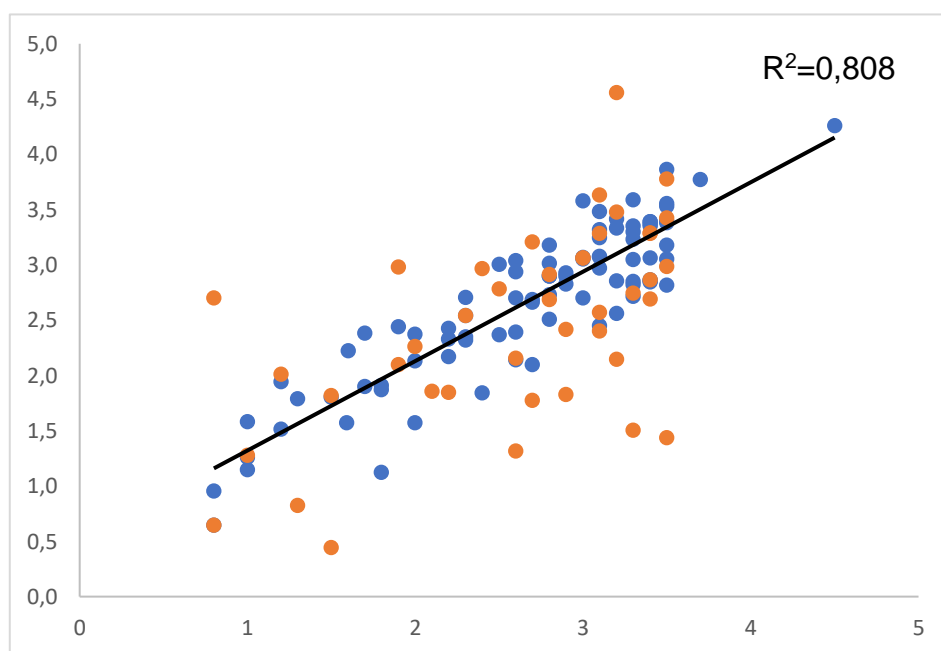


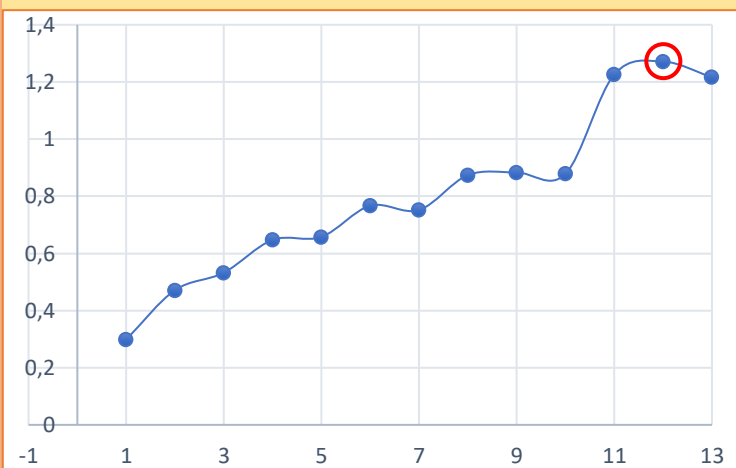
Figura 33.-Gráfica de dispersión con 13 descriptores. • Entrenamiento • predicción

7.11.3. Análisis AIC y FIT.

El análisis de los parámetros AIC y FIT permite establecer la pertinencia de adicionar un nuevo descriptor a la metodología de la RLM paso a paso, (Tabla 10). Los criterios para la adición de un nuevo descriptor al modelo son favorables si el parámetro FIT aumenta y el parámetro AIC disminuye. Tomando en cuenta estos criterios, se observa que, hasta la adición del duodécimo descriptor, el parámetro FIT aumenta, mientras que el parámetro AIC disminuye. Estos resultados permiten establecer que el número de descriptores en el modelo debe de ser doce como máximo, por lo tanto, se contó para este punto con 130 moléculas de las cuales 88 en entrenamiento y 42 predicción.

Tabla 10.-Cálculo de AIC y FIT.

Modelo	Descriptores	AIC	FIT
1	MATS7e	0,478	0,299
2	ATSC6i	0,414	0,471
3	VE2_D	0,387	0,533
4	minssCH2	0,345	0,647
5	GATS2i	0,328	0,658
6	AATSC0i	0,288	0,767
7	Energía Doc.	0,273	0,752
8	SRW5	0,233	0,874
9	GATS2c	0,215	0,883
10	GATS4c	0,199	0,878
11	AATSC4c	0,143	1,226
12	VE3_Dzs	0,127	1,271
13	maxHBd	0,120	1,216



El segundo aspecto, se trazó el nuevo modelo correspondiente a 12 descriptores propuesto por el cálculo AIC y FIT, obteniéndose la siguiente gráfica donde la correlación es de 0.7972 (Figura 34), donde los puntos naranjas corresponden a las moléculas de entrenamiento y azul a las de predicción. En la Tabla 11 se muestran los datos obtenidos de la correlación respecto a la gráfica de la Figura 34.

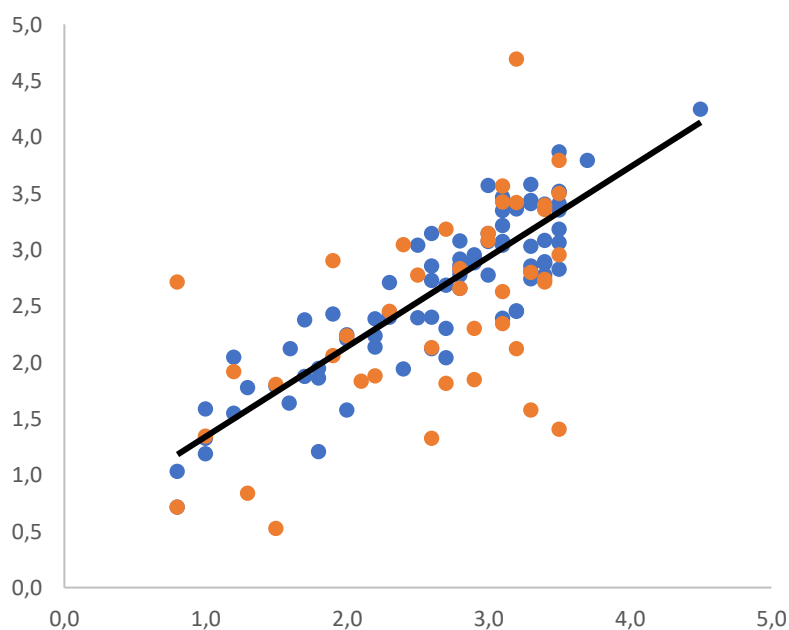


Figura 34.-Correlación 12 descriptores.

Tabla 11.-Resultados obtenidos con el cálculo AIC y FIT.					
12 descriptores	n	Pendiente	Desviación Es.	R	R ²
Entrenamiento	88	0,7972	0,7389	0,8928	0,7972
Predicción	42	0,6434	0,8182	0,4649	0,2161

7.11.4. Aplicabilidad del dominio.

La Aplicabilidad del Dominio (AD) del modelo tiene por objetivo la identificación de los compuestos que contienen valores atípicos. En los resultados obtenidos, se observó que en los conjuntos de entrenamiento y predicción existen varios de estos valores, en la Figura 34 podemos observar la gráfica de aplicabilidad de dominio para los dos subconjuntos, en azul el conjunto de entrenamiento, y en naranja el conjunto de predicción, de los cuales podemos observar varios de los puntos seleccionados con círculos rojos, esto indica que todos estos que están marcados son valores atípicos donde el valor límite para aproximación por estandarización (SA) es igual a 2.00 (línea punteada roja) y para el apalancamiento (LA) se sitúa el valor de 0.26 (línea punteada en azul), eliminando esos valores atípicos obtendremos una correlación óptima para nuestro modelo QSAR, es necesario repetir esta operación con la finalidad de evitar tener valores atípicos, no obstante, debemos mencionar que este cálculo se realizó sobre 12 descriptores lo indicamos en la sección 7.11.3. La aplicabilidad de dominio (AD) del modelo se estableció de acuerdo con metodología descrita.

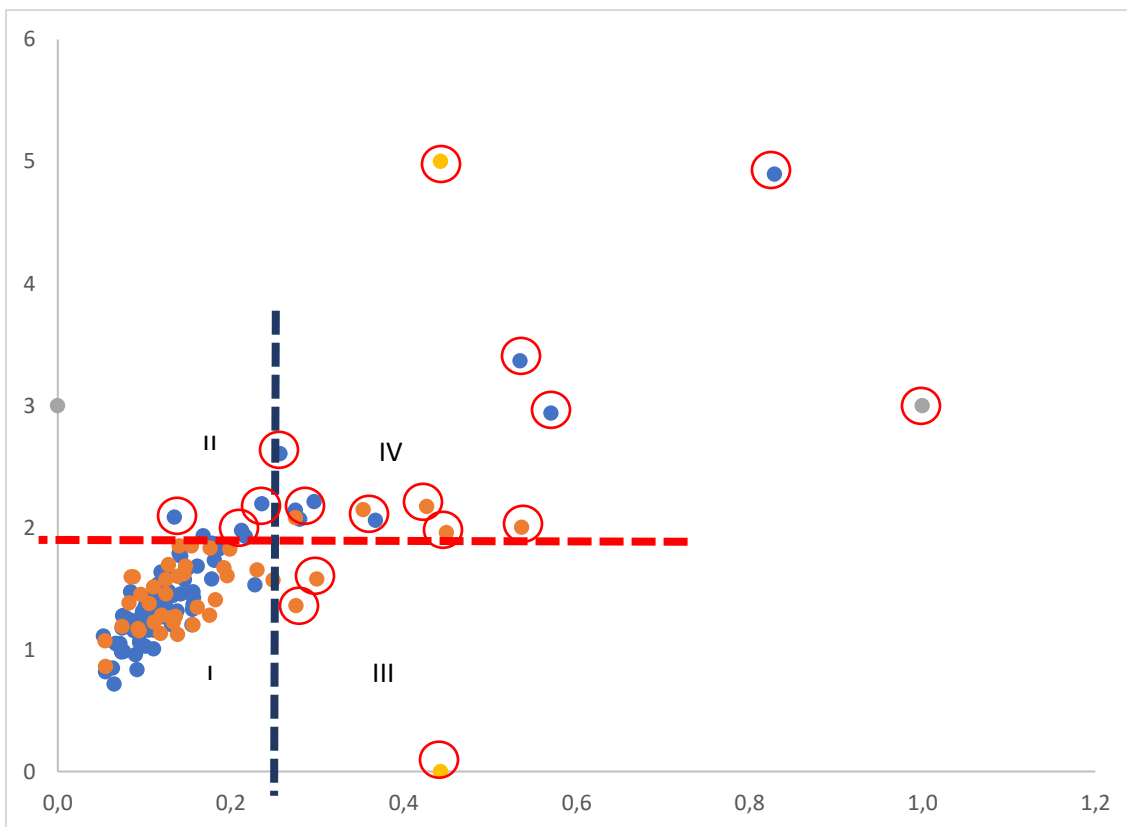


Figura 35.-Aplicabilidad del dominio con 12 descriptores. • Entrenamiento • predicción

7.11.5. Análisis MAE y ANN.

Se realizó el análisis del error medio absoluto (MAE), para conocer la calidad del modelo debido a que era necesario comparar cuáles eran las moléculas con mayor residuo, como se discutió en la sección 7.11.4, con la aplicabilidad del dominio encontramos varios valores atípicos con análisis MAE, es decir, se omitieron el %5 que corresponde con aquellos que no beneficiaron al modelo.

En el intervalo de los Log (Ki) utilizado para calcular MAE (mencionado en la metodología), se observó el valor MAE del subconjunto de entrenamiento con 88 compuestos, MAE + 3σ medido antes y después de omitir 5% de datos con los residuos más altos, se muestra los resultados R², MAE, σ, MAE +3σ tabla 12, donde denota sobre la calidad del modelo utilizando la metodología de ANN.

Donde con la metodología usada fue con ANN y con el cual mejoró la calidad del conjunto de predicción, es decir, ir identificando el modelo con ANN presente residuos pequeños.

Relacionado con lo anterior con solo el 95% de las moléculas se pudo generar un modelo con modelo ANN sin el valor atípico.

Tabla 12.-Análisis MAE.					
	ANN 1		ANN 2		
Porcentaje de datos	100%	95%	100%	95%	
R ²	0,9782	0,9859	0,9964	0,9967	Entrenamiento
n	88	86	86	86	
MAE	0,0838	0,0868	0,0364	0,0369	
σ	0,0786	0,0549	0,0283	0,0273	
MAE+3 σ	0,3197	0,2516	0,1213	0,1187	
Calidad	verdadero	verdadero	verdadero	verdadero	
R ²	0,2260	0,3821	0,7320	0,7804	
n	42	40	31	30	
MAE	0,5043	0,4643	0,3006	0,3662	
σ	0,4928	0,4241	0,2657	0,2411	
MAE+3 σ	1,9826	1,7365	1,0977	1,0024	
Calidad	Falso	Falso	Falso	Falso	

En la Figura 36 se observa una gráfica con una mejor correlación en la que se muestra un R²=0,99 lo que indica un RLM bastante adecuado, se visualiza con color azul el conjunto entrenamiento y de color naranja el conjunto de predicción. Se puede observar la línea de tendencia ideal marcada con una línea negra, y anexando a la información el total de moléculas con las que contamos para este cálculo fue de 126 moléculas con 86 moléculas para entrenamiento y 40 para la predicción, resultado total de eliminar a los valores atípicos, anexando la tabla 13 vemos los resultados obtenidos por el análisis MAE.

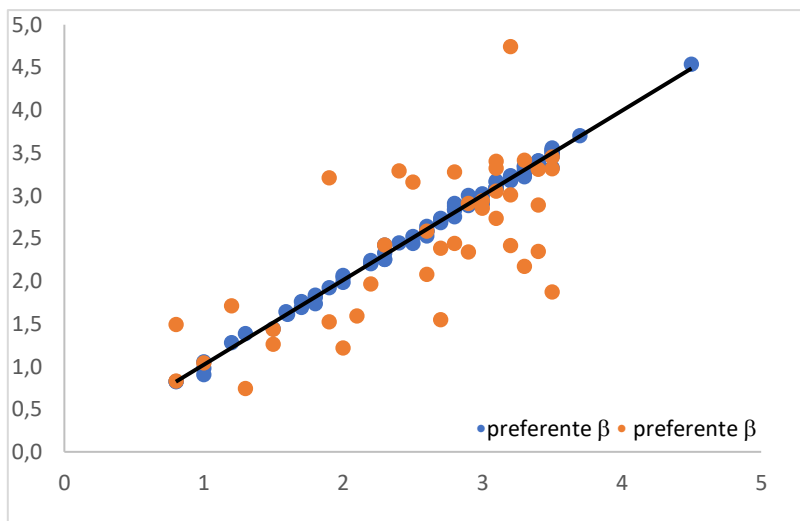


Figura 36.-Gráfica de correlación ANN preferente beta.

Tabla 13.-Resultados obtenidos con el análisis MAE.

12 descriptores	n	Pendiente	Desviación Es.	R ²
Entrenamiento β	86	0,9910	0,7091	0,9959
Predicción β	40	0,8002	0.8409	0,5277

utilizando este mismo modelo generado por ANN, anexamos los valores obtenidos de preferencia alfa (ver Figura 37), en el evaluamos cuáles de las moléculas con preferencia alfa tiene selectividad por la GSK3β, en la correlación se indican los puntos con color verde los preferentemente alfa y azul preferente beta.

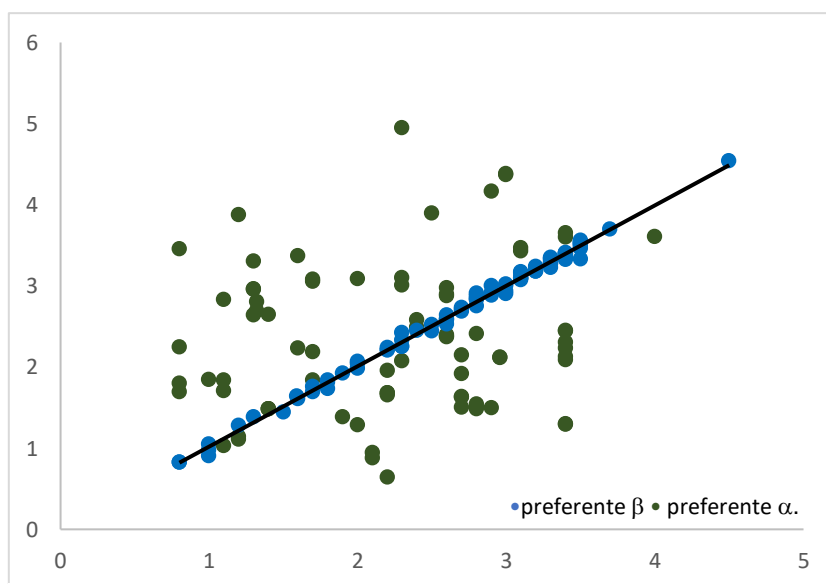


Figura 37.- Gráfica correlación ANN preferente beta y alfa.

A pesar de haber preferencia por algunas moléculas notamos claramente que hay otras que no la tienen por ende su dispersión es muy alta esto lo comprobamos al visualizar los datos de la Tabla 14.

Tabla 14.- Resultados obtenidos preferentemente beta más alfa.				
12 descriptores	n	Pendiente	Desviación Es.	R ²
Entrenamiento β	86	0,9910	0,7091	0,9959
Prediccion α	79	0,9096	1.9748	0,0240

Figura 38 se observan a las moléculas sin preferencia indicados con puntos rojos y preferente beta con los puntos azules, visualizándose así, las moléculas que tiene preferencia por beta, algo notable es que su dispersión de puntos de sin preferencia es grande, esto se demuestra en la Tabla 15 donde se indican los resultados obtenidos.

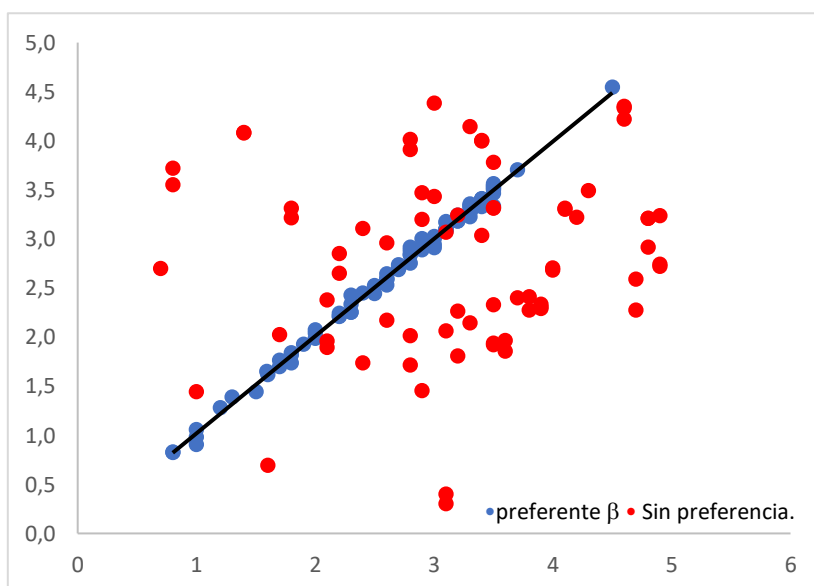


Figura 38.- Gráfica correlación ANN preferente beta y sin preferencia.

Tabla 15.-Resultados obtenidos preferentemente beta más sin preferencia				
12 descriptores	n	Pendiente	Desviación Es.	R ²
Entrenamiento β	86	0,9910	0,7091	0,9959
Prediccion sin preferencia	71	0,1026	1.0217	0.0145

Con base a lo comentado y omitiendo a las moléculas con mayor residuo de preferente beta, alfa y sin preferencia se obtuvo lo siguiente.

A continuación, se muestra en la Figura 39 la correlación con solo las moléculas que tienen preferencia hacia la proteína GSK3 β , con base al modelo generado en ANN. Se aplicó el análisis MAE para omitir a las moléculas con mayor residuo, donde para el entrenamiento contamos con 86 moléculas indicados en la imagen con puntos color azul y la predicción con los puntos color naranja siendo este de 28 moléculas, con un total de 114 moléculas con preferencia beta.

Claramente es visible el cambio, con solo concentrar las moléculas con preferencia beta, las cuales son notorias al estar cerca de la línea de tendencia en la gráfica línea negra punteada, más los valores obtenidos (Tabla 16) el cual notamos claramente los cambios sobre el valor de R².

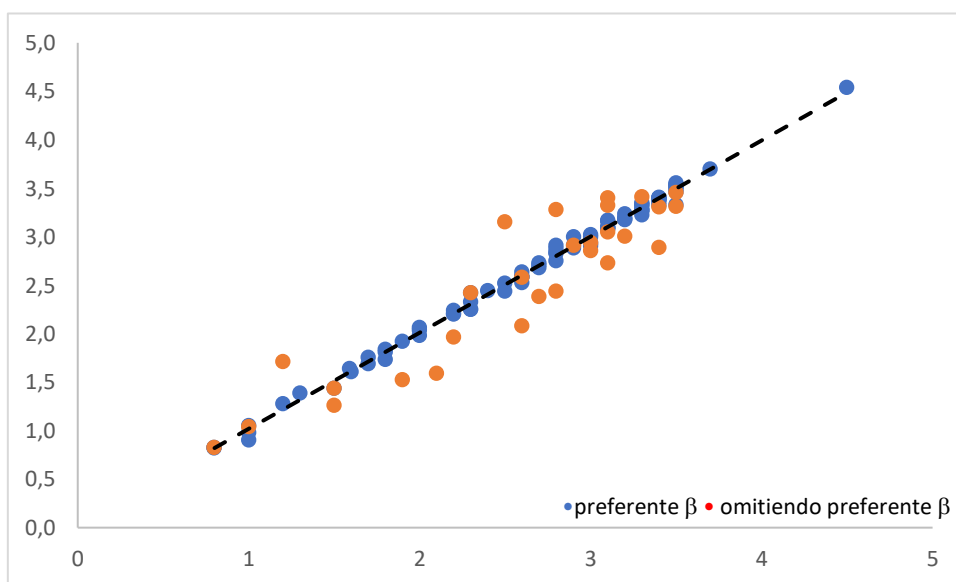


Figura 39.- Gráfica correlación ANN preferente beta β con menor residuo.

Tabla 16.-Resultados obtenidos preferentemente beta con menores residuos.				
12 descriptores	n	Pendiente	Desviación Es.	R ²
Entrenamiento β	86	0,9910	0,7091	0,9959
Predicción β	28	0,9513	0.7838	0.8577

Cuando se realizó la misma operación para preferente alfa, se obtuvo una dispersión y una mejor correlación visible en R^2 ver tabla 17

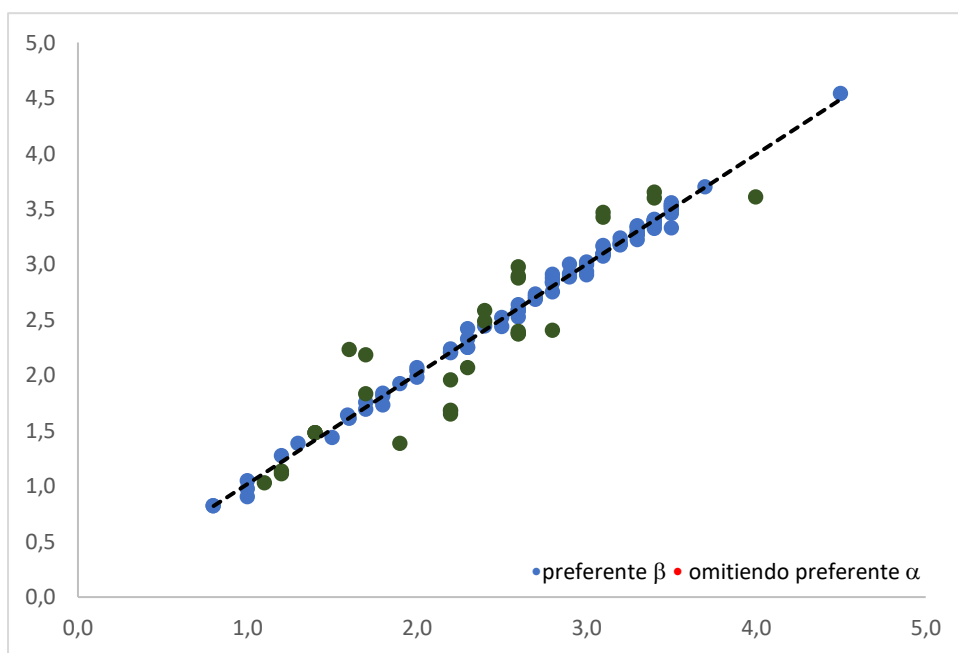


Figura 40.- Gráfica correlación ANN preferente beta y alfa con menor residuo.

Tabla 17.-Resultados obtenidos preferentemente beta más preferencia alfa con menor residuo.

12 descriptores	n	Pendiente	Desviación Es.	R^2
Entrenamiento β	86	0,9910	0,7091	0,9959
Prediccion α	29	0,9922	0.7604	0.8300

Por último, tenemos a las moléculas sin preferencia que como hemos observado fueron cambiando sus valores en R^2 , al ir omitiendo a las moléculas con los residuos altos, lo mismo sucedió con los de sin preferencia (ver Figura 41). Los resultados obtenidos en la Tabla 18, donde en este caso tenemos en el entrenamiento un total de 86 moléculas que es la base para predecir la selectividad para GSK3 β , con respecto a la predicación este se redujo aún total de 21 moléculas con selectividad a GSK3 β .

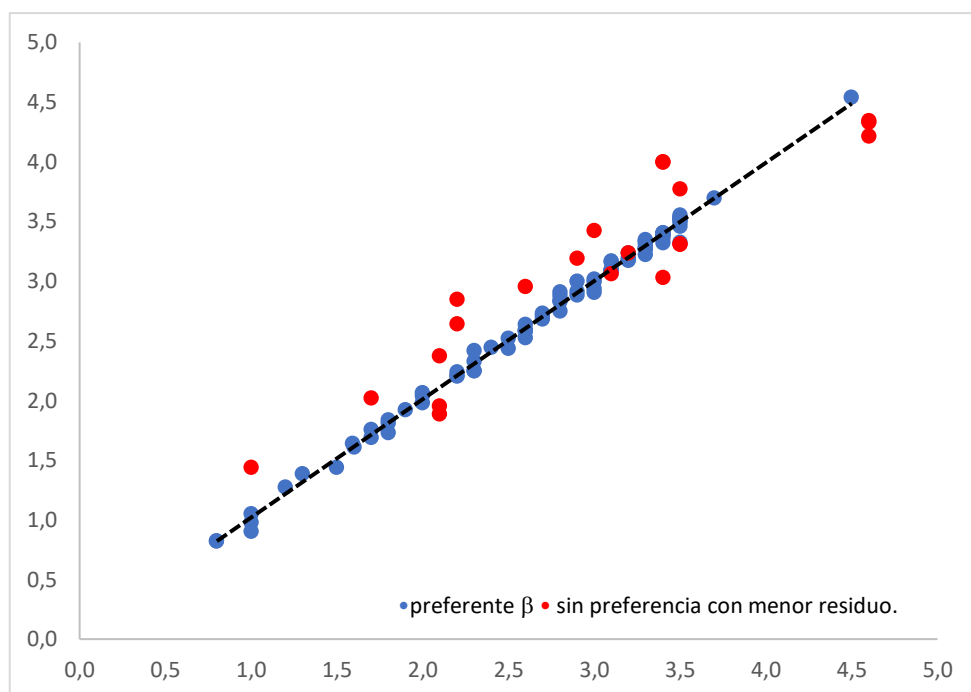


Figura 41.- Gráfica correlación ANN preferente beta y sin preferencia con menor residuo.

Tabla 18.-Resultados obtenidos preferentemente beta más sin preferencia con menor residuo.

12 descriptores	n	Pendiente	Desviación Es.	R2
Entrenamiento β	86	0,9910	0,7091	0,9959
Predicción sin preferencia	21	0,8165	0.8946	0.8746

7.12. Inhibidores preferentemente beta.

Con base lo anterior se analizó cada una de estas moléculas que dentro de nuestro modelo indican tener selectividad por la proteína GSK3 β vistos en la sección 7.11.5. Con la finalidad de encontrar las mayores interacciones y ver cuáles son las importantes para llevar a cabo una inhibición selectiva, por ende, encontramos que muchas moléculas comparten interacciones de tipo Van der Waals, enlace de hidrógeno, halógenos, carbono hidrógeno, pi-sigma, pi-cación. En la Figura 42 se indican los colores, se muestran las interacciones, lo sobresaliente, es que la mayoría de las interacciones observadas con las GSK3 β son de tipo de Van der Waals, enlace hidrógeno y estas son más frecuentes en anillos aromáticos, a manera de ejemplo visualizamos la Figura 42 tenemos a molchem33 el cual corresponde a la molécula CHEMBL1972220 y

molchem47 que corresponde a CHEMBL1977148 de la página oficial ChEMBL, los colores representa a cada átomo, los azules representan a los átomos de nitrógeno, los grises a los de carbono, azul al flúor, rojo al oxígeno, amarillo representa al azufre y por último los átomos de color blanco representan a los hidrógenos, estos ejemplos son de las moléculas que tienen preferencia a GSK3 β en la predicción.

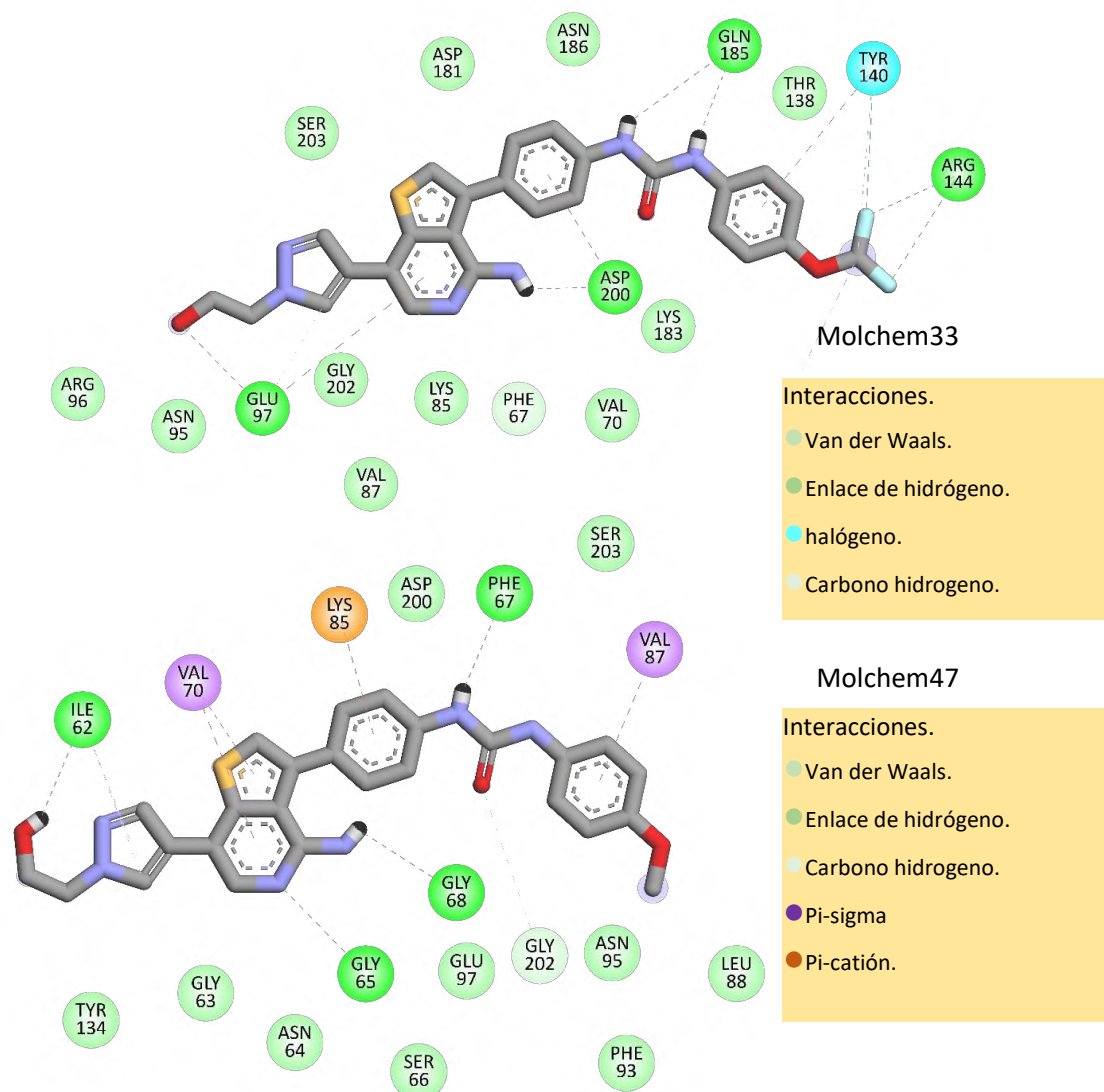


Figura 42.- Interacción de moléculas preferente β con proteína GSK3 β

Posteriormente, el análisis de las moléculas con preferencia por GSK3 α , podemos visualizar en el ejemplo de la Figura 43, se muestran 2 moléculas que tienen interacción con la proteína GSK3 β , lo importante a destacar es que tienen las mismas interacciones, es decir los enlaces de Van der Waals, enlace

hidrogeno, también muestran algunas interacciones de enlace Pi, lo relevante es que las interacciones son sobre los anillos aromáticos, se observa como en el ejemplo anterior, que al menos uno de sus anillos contiene ya sea Nitrógeno o Azufre ,es decir, que al menos debe existir un heterociclo dentro de la molécula.

Asimismo, las moléculas seleccionadas corresponden a alfa 2 el cual corresponde CHEMBL1964644 y alfa5big que corresponde a CHEMBL1968070 página oficial ChEMBL, cabe mencionar que los colores descritos sobre el ejemplo anterior representan a los mismos átomos mientras que el átomo en color verde representa al cloro en la molécula con nombre alfa5big.

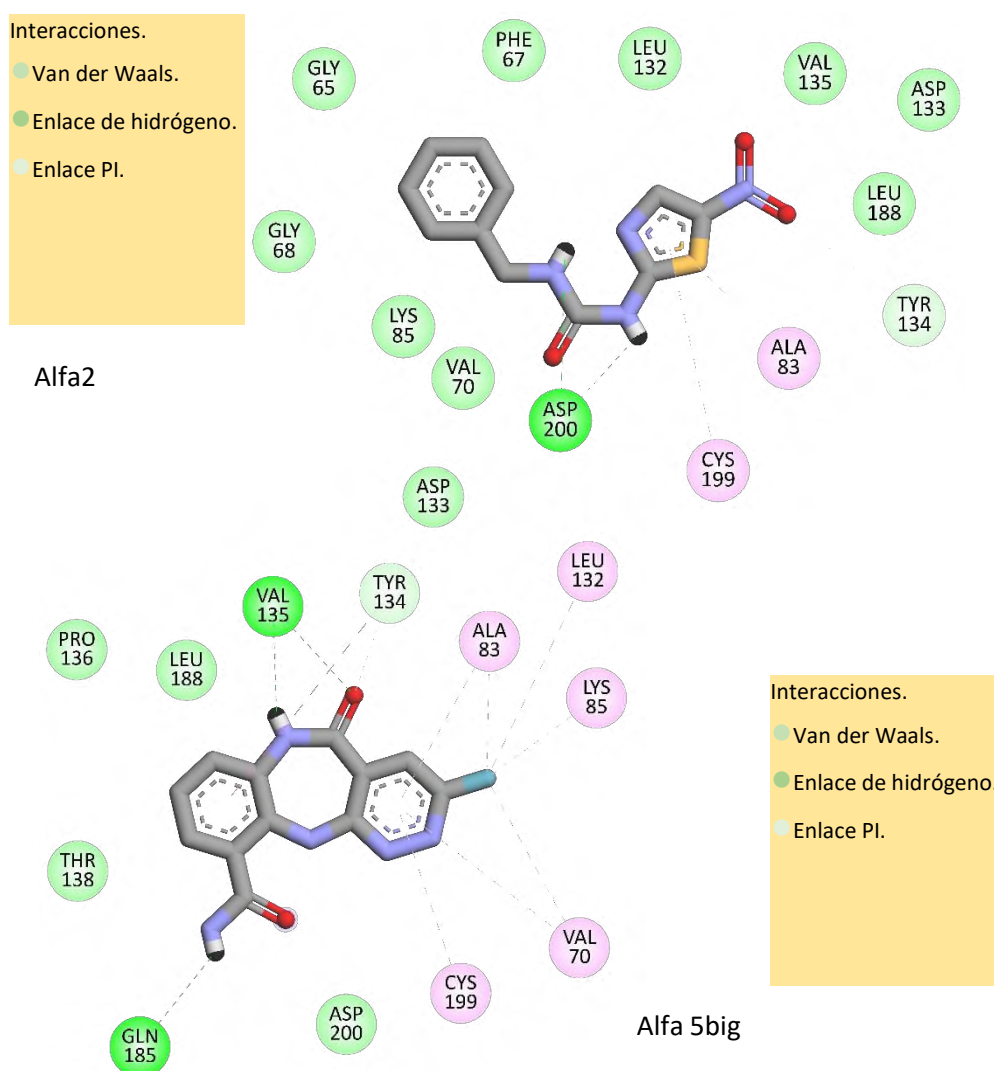


Figura 43.-Interacción de moléculas preferentes α con proteína GSK3 β .

Finalmente se realizó el análisis con las moléculas que no tienen preferencia con ningún parálogo Figura 44, donde Sinp4a es CHEMBL191003 y sinp40big es CHEMBL1999778, página oficial de ChEMBL, es la misma predisposición vista en las moléculas con preferencia beta y alfa, las cuales son las interacciones de tipo Van der Waals, enlace hidrogeno, con algún halógeno , en nuestro ejemplo es por el átomo de flúor indicada con color azul cielo, en particular la mayoría de las interacciones corresponden plenamente con los anillos aromáticos, claro sin omitir las interacciones que existen con los oxígenos, nitrógenos que acompañan a los grupos aromáticos, la tendencia con mayor relevancia es sobre los anillos aromáticos, donde al menos uno de sus átomos esta sustituido, ya sea por nitrógeno o el azufre por ende esta es de una gran aportación, pues el conocer las interacciones, permite conocer las moléculas que tienen una inhibición selectiva por la proteína GSK3β.

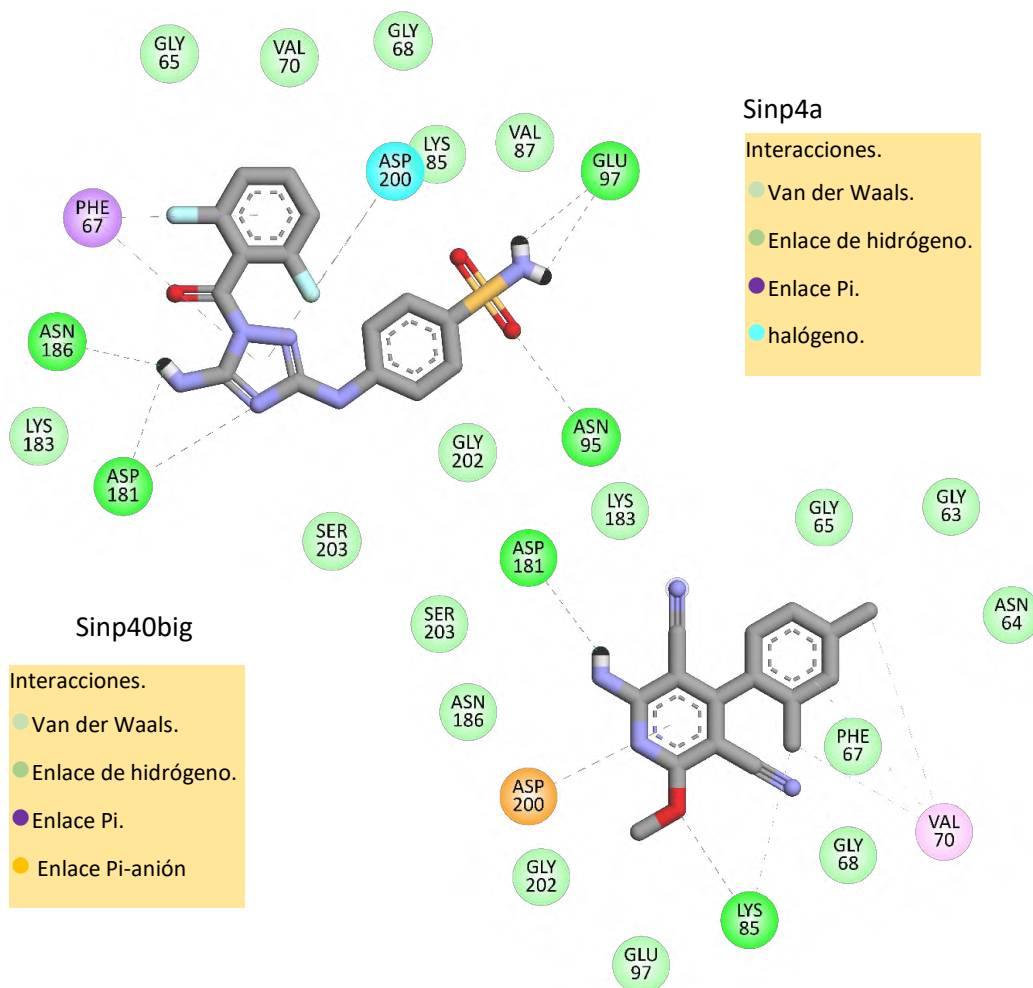


Figura 44.-Interacción de moléculas sin preferencia con proteína GSK3β.

Capítulo 8.- Descriptores y propuesta de inhibidores selectivos.

Los descriptores principales para la inhibición selectiva están dados en la Tabla 18 en la que se muestran a los descriptores del modelo QSAR y el tipo de descriptor al que corresponde, cabe mencionar que con lo visto en la Sección 7.12, se puede comentar que las posibles moléculas con actividad selectiva por la proteína GSK3 β deben contener dos anillos aromáticos (ver Figura 45), tenemos una de las posibles propuestas, partiendo de la base de una molécula, 1,3-difenilurea donde podemos intentar sustituir las posiciones 1, 3, 5, 6 o 12, 14, 16 y 15 con un átomo de nitrógeno o azufre, aunque solo uno de los anillos debe tener esta sustitución y de igual manera podemos anexar radicales por sustitución en las posiciones meta y para, que corresponden a los números 3, 5, 6 o 14, 15 y 16, las combinaciones pueden ser variadas pero existen posibilidades de que se generen moléculas que tengan mayor potencial como inhibidor de la proteína GSK3 β .

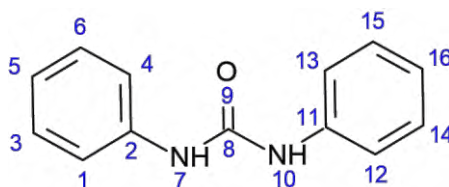


Figura 45.- Base estructural para inhibidores de GSK3 β .

Tabla 19.-Descriptores del modelo.	
Descriptor	Tipo de descriptor
MATS7e	Autocorrelación 2D
ATSC6i	Autocorrelación 2D
VE2_D	Descriptores basados en matrices 2D
minssCH2	Estado electro topológico de tipo átomo
GATS2i	Autocorrelación 2D
AATSC0i	Autocorrelación 2D
Energía Docking	Cuántico
SRW5	Recuento de recorridos de retorno automático del pedido 05
GATS2c	Autocorrelación 2D
GATS4c	Autocorrelación 2D
AATSC4c	Autocorrelación 2D
VE3_Dzs	Descriptores basados en matrices 2D

Capítulo 9.- Conclusiones.

Se puede obtener un modelo teórico descriptivo y predictivo, a partir de la correlación de descriptores cuánticos, fisicoquímicos, topológicos entre otros, como el Log Ki de moléculas obtenidas de base de datos como son ChEMBL, PubChem, ChEMSPIDER entre otros, utilizando la metodología de QSAR con métodos combinados como aplicabilidad de dominio, MAE y ANN para determinar y obtener un mejor modelo.

Se logró identificar el mínimo global para cada una de las moléculas y se obtuvieron los descriptores moleculares de cada uno de ellos.

Se logró elaborar y plantear el objetivo propuesto de la creación de nuevos inhibidores selectivos del parálogo GSK3 β con el cual podemos diseñar moléculas que tengan una mayor actividad inhibitoria,

Se identificaron las interacciones relevantes de los inhibidores con selectividad por GSK3 β del conjunto de moléculas que presentaron ser competitivos por la proteína del ATP y las que tienen de acuerdo con el modelo mayor selectividad por GSK3 β .

Se logró la identificación de los aminoácidos importantes del sitio catalítico que permiten la inhibición de la proteína GSK3 β , así como las diferencias relevantes que existen entre los parálogos de la proteína GSK3 con sus respectivos aminoácidos.

Se demostró que a pesar de contar con el 98% de similitud dentro del sitio catalítico diferenciado únicamente por un aminoácido dentro del sitio catalítico puede haber inhibidores selectivos con mayor actividad sobre la proteína GSK3.

El potencial de la proteína GSK3 β como terapia para múltiples patologías puede e inclusive ser mayor de lo que se cree, así lo demuestran las investigaciones sobre la misma.

Aportaciones.

Una biblioteca virtual con 250 moléculas de las cuales 150 moléculas son preferentemente inhibitorias para las proteínas GSK3 β , 50 sin preferencia y 50 con preferencia por GSK3 α .

250 moléculas optimizadas geométricamente.

Dinámica molecular de la proteína GSK3 β en condiciones fisiológicas.

250 acoplamientos moleculares (Docking)

La obtención del Modelo QSAR, así como un “pool” de descriptores de aproximadamente 900.

Una lista de moléculas que tienen preferencia por la proteína GSK3 β con base al modelo obtenido, estas moléculas son de la base de datos ChemBL y se pueden observar en el apartado anexo al final de este documento.

Referencias.

1. Cohen P, Frame S. The renaissance of GSK3. *Nat Rev Mol Cell Biol.* 2001;2(10):769–776. doi:10.1038/35096075
2. Mulholland DJ, Dedhar S, Wu H, Nelson CC. PTEN and GSK3 β : Key regulators of progression to androgen-independent prostate cancer. *Oncogene.* 2006;25(3):329–337. doi:10.1038/sj.onc.1209020
3. Hooper C, Killick R, Lovestone S. The GSK3 hypothesis of Alzheimer's disease. *J Neurochem.* 2008;104(6):1433–1439. doi:10.1111/j.1471-4159.2007.05194.x
4. Finkelman HE, Martinez A. GSK-3 Inhibitors: Preclinical and Clinical Focus on CNS. *Front Mol Neurosci.* 2011;4. doi:10.3389/fnmol.2011.00032
5. Ochoa Hernández AB, Juárez Vázquez CI, Rosales Reynoso MA, Barros Núñez P. La vía de señalización Wnt- β -catenina y su relación con cáncer. *Acad Mex Cirugía, AC.* 2012;80(4):389–398. <https://www.redalyc.org/pdf/662/66224459015.pdf>
6. Chen X, Wang R, Liu X, et al. A Chemical-Genetic Approach Reveals the Distinct Roles of GSK3 α and GSK3 β in Regulating Embryonic Stem Cell Fate. *Dev Cell.* 2017;43(5):563-576.e4. doi:10.1016/j.devcel.2017.11.007
7. Jope RS, Yuskaitis CJ, Beurel E. Glycogen synthase kinase-3 (GSK3): Inflammation, diseases, and therapeutics. *Neurochem Res.* 2007;32(4–5):577–595. doi:10.1007/s11064-006-9128-5
8. Linding R, Jensen LJ, Ostheimer GJ, et al. Systematic discovery of in vivo phosphorylation networks. *Sci Direct.* 2007;129(7):1415–1426. doi:10.1016/j.cell.2007.05.052
9. Beurel E, Grieco SF, Jope RS. Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. *Pharmacol Ther.* 2015;148:114–131. doi:10.1016/j.pharmthera.2014.11.016
10. Phillips GN, Lattman EE, Cummins P, Lee KY, Cohen C. Crystal structure and molecular interactions of tropomyosin. *Nature.* 1979;278(5703):413–417. doi:10.1038/278413a0
11. Dajani R, Fraser E, Roe SM, et al. Crystal structure of glycogen synthase kinase 3 β : Structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell.* 2001;105(6):721–732. doi:10.1016/S0092-8674(01)00374-9
12. Du K, Montminy M. CREB is a regulatory target for the protein kinase Akt/PKB. *J Biol Chem.* 1998;273(49):32377–32379. doi:10.1074/jbc.273.49.32377
13. Kato S, Ding J, Du K. Differential activation of CREB by Akt1 and Akt2. *Biochem Biophys Res Commun.* 2007;354(4):1061–1066. doi:10.1016/j.bbrc.2007.01.094
14. Cole A, Frame S, Cohen P. Further evidence that the tyrosine

- phosphorylation of glycogen synthase kinase-3 (GSK3) in mammalian cells is an autophosphorylation event. *Biochem J.* 2004;377(1):249–255. doi:10.1042/BJ20031259
15. MacAulay K, Doble BW, Patel S, et al. Glycogen Synthase Kinase 3 α -Specific Regulation of Murine Hepatic Glycogen Metabolism. *Cell Metab.* 2007;6(4):329–337. doi:10.1016/j.cmet.2007.08.013
 16. Hemmings BA, Yellowlees D, C. Kernohan J, Cohen P. Purification of Glycogen Synthase Kinase 3 from Rabbit Skeletal Muscle. Copurification with the Activating Factor (FA) of the (Mg-ATP) Dependent Protein Phosphatase. *Eur J Biochem.* 1981;119(3):443–451. doi:10.1111/j.1432-1033.1981.tb05628.x
 17. Yang S-D, Yu J-S, Shiah S-G, Huang J-J. Protein Kinase FA/Glycogen Synthase Kinase-3 α After Heparin Potentiation Phosphorylates τ on Sites Abnormally Phosphorylated in Alzheimer's Disease Brain. *J Neurochem.* 2002;63(4):1416–1425. doi:10.1046/j.1471-4159.1994.63041416.x
 18. McCubrey JA, Steelman LS, Bertrand FE, et al. Multifaceted roles of GSK-3 and Wnt/ β -catenin in hematopoiesis and leukemogenesis: Opportunities for therapeutic intervention. *Leukemia.* 2014;28(1):15–33. doi:10.1038/leu.2013.184
 19. Martin M, Rehani K, Jope RS, Michalek SM. Toll-like receptor - Mediated cytokine production is differentially regulated by glycogen synthase kinase 3. *Nat Immunol.* 2005;6(8):777–784. doi:10.1038/ni1221
 20. Mora A, Sakamoto K, McManus EJ, Alessi DR. Role of the PDK1-PKB-GSK3 pathway in regulating glycogen synthase and glucose uptake in the heart. *Sch Life Sci.* 2005;579(17):3632–3638. doi:10.1016/j.febslet.2005.05.040
 21. Pronobis MI, Rusan NM, Peifer M. A novel GSK3-regulated APC:Axin interaction regulates Wnt signaling by driving a catalytic cycle of efficient β catenin destruction. *Elife.* 2015;4(September 2015). doi:10.7554/eLife.08022
 22. Cortés Vieyra R, Silva García O, Oviedo Boyso J, et al. The Glycogen Synthase Kinase 3 α and β Isoforms Differentially Regulates Interleukin-12p40 Expression in Endothelial Cells Stimulated with Peptidoglycan from *Staphylococcus aureus*. Hsieh Y-H, ed. *PLoS One.* 2015;10(7):e0132867. doi:10.1371/journal.pone.0132867
 23. Wang J, Gu BJ, Masters CL, Wang YJ. A systemic view of Alzheimer disease - Insights from amyloid- β metabolism beyond the brain. *Nat Rev Neurol.* 2017;13(10):612–623. doi:10.1038/nrneurol.2017.111
 24. Takashima A, Noguchi K, Sato K, Hoshino T, Imahori K. Tau Protein kinase I is essential for amyloid β -protein-induced neurotoxicity. *Proc Natl Acad Sci U S A.* 1993;90(16):7789–7793. doi:10.1073/pnas.90.16.7789
 25. Noble W, Planel E, Zehr C, et al. Inhibition of glycogen synthase kinase-3 by lithium correlates with reduced tauopathy and degeneration in vivo. *Proc Natl Acad Sci U S A.* 2005;102(19):6990–6995.

doi:10.1073/pnas.0500466102

26. Murphy MP, Levine H. Alzheimer's disease and the amyloid- β peptide. *J Alzheimer's Dis.* 2010;19(1):311–323. doi:10.3233/JAD-2010-1221
27. Phiel CJ, Wilson CA, Lee VMY, Klein PS. GSK-3 α regulates production of Alzheimer's disease amyloid- β peptides. *Nature.* 2003;423(6938):435–439. doi:10.1038/nature01640
28. Philip T.T. L, Yili W, Zou H, et al. Inhibition of GSK3 β -mediated BACE1 expression reduces Alzheimer-associated phenotypes. *J Clin Invest.* 2013;123(1):224–235. doi:10.1172/JCI64516
29. Dunning CJ, McGauran G, Willén K, Gouras GK, O'Connell DJ, Linse S. Direct High Affinity Interaction between A β 42 and GSK3 α Stimulates Hyperphosphorylation of Tau. A New Molecular Link in Alzheimer's Disease? *ACS Chem Neurosci.* 2016;7(2):161–170. doi:10.1021/acscchemneuro.5b00262
30. Games D, Adams D, Alessandrini R, et al. Alzheimer-type neuropathology in transgenic mice overexpressing V717F β -amyloid precursor protein. *Nature.* 1995;373(6514):523–527. doi:10.1038/373523a0
31. Yoshiyama Y, Higuchi M, Zhang B, et al. Synapse Loss and Microglial Activation Precede Tangles in a P301S Tauopathy Mouse Model. *Sci Direct.* 2007;53(3):337–351. doi:10.1016/j.neuron.2007.01.010
32. Hurtado DE, Molina-Porcel L, Carroll JC, et al. Selectively silencing GSK-3 isoforms reduces plaques and tangles in mouse models of Alzheimer's disease. *J Neurosci.* 2012;32(21):7392–7402. doi:10.1523/JNEUROSCI.0889-12.2012
33. Kadavath H, Hofele R V., Biernat J, et al. Tau stabilizes microtubules by binding at the interface between tubulin heterodimers. *Proc Natl Acad Sci U S A.* 2015;112(24):7501–7506. doi:10.1073/pnas.1504081112
34. Hanger DP, Anderton BH, Noble W. Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. *Trends Mol Med.* 2009;15(3):112–119. doi:10.1016/j.molmed.2009.01.003
35. Maurin H, Lechat B, Dewachter I, et al. Neurological characterization of mice deficient in GSK3 α highlight pleiotropic physiological functions in cognition and pathological activity as Tau kinase. *Mol Brain.* 2013;6(1):27. doi:10.1186/1756-6606-6-27
36. Cavallini A, Brewerton S, Bell A, et al. An unbiased approach to identifying tau kinases that phosphorylate tau at sites associated with Alzheimer disease. *J Biol Chem.* 2013;288(32):23331–23347. doi:10.1074/jbc.M113.463984
37. Maqbool M, Hoda N. GSK3 Inhibitors in the Therapeutic Development of Diabetes, Cancer and Neurodegeneration: Past, Present and Future. *Curr Pharm Des.* 2017;23(29). doi:10.2174/1381612823666170714141450
38. Elmore S. Apoptosis: A Review of Programmed Cell Death. *Toxicol Pathol.* 2007;35(4):495–516. doi:10.1080/01926230701320337

39. Simabuco FM, Morale MG, Pavan ICB, Morelli AP, Silva FR, Tamura RE. p53 and metabolism: From mechanism to therapeutics. *Oncotarget*. 2018;9(34):23780–23823. doi:10.18632/oncotarget.25267
40. Gupta K, Stefan T, Ignatz-Hoover J, et al. GSK-3 inhibition sensitizes acute myeloid leukemia cells to 1,25D-mediated differentiation. *Cancer Res*. 2016;76(9):2743–2753. doi:10.1158/0008-5472.CAN-15-2290
41. Davies M, Nowotka M, Papadatos G, et al. ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Res*. 2015;43(W1):W612–W620. doi:10.1093/nar/gkv352
42. Martinez A, Gil C, Perez DI. Glycogen synthase kinase 3 inhibitors in the next horizon for Alzheimer's disease treatment. *Int J Alzheimers Dis*. 2011;2011. doi:10.4061/2011/280502
43. Hamann M, Alonso D, Martín-Aparicio E, et al. Glycogen synthase kinase-3 (GSK-3) inhibitory activity and structure-activity relationship (SAR) studies of the manzamine alkaloids. Potential for Alzheimer's disease. *J Nat Prod*. 2007;70(9):1397–1405. doi:10.1021/np060092r
44. Palomo V, Soteras I, Perez DI, et al. Exploring the binding sites of glycogen synthase kinase 3. identification and characterization of allosteric modulation cavities. *J Med Chem*. 2011;54(24):8461–8470. doi:10.1021/jm200996g
45. Jorge Torres O de la C, Domenech SG. La inhibición de gsk3b reduce la inflamación cerebral y activa las redes neuronales en un modelo preclínico del síndrome de Rett -. *Genética Médica News*. doi:10.1038/85899.
46. Hoffmeister L, Diekmann M, Brand K, Huber R. GSK3: A Kinase Balancing Promotion and Resolution of Inflammation. *Cells*. 2020;9(4):820. doi:10.3390/cells9040820
47. Acikgoz E, Güler G, Camlar M, Oktem G, Aktug H. Glycogen synthase kinase-3 inhibition in glioblastoma multiforme cells induces apoptosis, cell cycle arrest and changing biomolecular structure. *Mol Biomol Spectrosc*. 2019;209:150–164. doi:10.1016/j.saa.2018.10.036
48. Piazzini M, Bavelloni A, Faenza I, Blalock W. Glycogen synthase kinase (GSK)-3 and the double-strand RNA-dependent kinase, PKR: When two kinases for the common good turn bad. *Biochim Biophys Acta - Mol Cell Res*. 2020;1867(10):118769. doi:10.1016/j.bbamcr.2020.118769
49. Beilin OK, Woodgett JR. GSK-3: Functional Insights from Cell Biology and Animal Models. *Front Mol Neurosci*. 2011;4. doi:10.3389/fnmol.2011.00040
50. Chien AJ, Conrad WH, Moon RT. A wnt survival guide: From flies to human disease. *J Invest Dermatol*. 2009;129(7):1614–1627. doi:10.1038/jid.2008.445
51. Gómez-Sintes R. GSK-3 mouse models to study neuronal apoptosis and neurodegeneration. *Front Mol Neurosci*. 2011;4. doi:10.3389/fnmol.2011.00045

52. Webster M-T, Rozycka M, Sara E, et al. Sequence variants of the axin gene in breast, colon, and other cancers: An analysis of mutations that interfere with GSK3 binding. *Natl Cent Biotechnol information*. 2000;28(4):443–453. doi:10.1002/1098-2264(200008)28:4<443::AID-GCC10>3.0.CO;2-D
53. Hernandez F, Lucas JJ, Avila J. GSK3 and tau: Two convergence points in Alzheimer's disease. *J Alzheimer's Dis*. 2013;33(1). doi:10.3233/JAD-2012-129025
54. Wang L, Wang Y, Zhang C, et al. Inhibiting glycogen synthase kinase 3 reverses obesity-induced white adipose tissue inflammation by regulating apoptosis inhibitor of macrophage/CD5L-mediated macrophage migration. *Arterioscler Thromb Vasc Biol*. 2018;38(9):2103–2116. doi:10.1161/ATVBAHA.118.311363
55. Polakis P, Hart M, Rubinfeld B. Defects in the regulation of β -catenin in colorectal cancer. *Adv Exp Med Biol*. 2000;470:23–32. doi:10.1007/978-1-4615-4149-3_3
56. López Martos AM. Glucógeno sintasa quinasa 3 (GSK3) como diana de fármacos anti-Alzheimer. Published online 2015. <https://eprints.ucm.es/48770/1/AURORA MARIA LOPEZ MARTOS %281%29.pdf>
57. Steinberg SF. Post-translational modifications at the ATP-positioning G-loop that regulate protein kinase activity. *Pharmacol Res*. 2018;135:181–187. doi:10.1016/j.phrs.2018.07.009
58. Smith SJ, Sutcliffe BT. *The Development of Computational Chemistry in the United Kingdom*. John Wiley & Sons, Ltd; 2007. doi:10.1002/9780470125878.ch5
59. Valles-Sánchez A, Rosales-Marines L, Serrato-Villegas LE, Farías-Cepeda L. *Métodos y Usos de la Química Computacional Computational Chemistry Methods and its Applications*. Vol 6.; 2014.
60. Alemán C, Muñoz-Guerra S. Aplicaciones de los Métodos Computacionales al Estudio de la Estructura y Propiedades de Polímeros On the Applicability of Computational Methods to Study the Structure and Properties of Polymers. *Polímeros Ciência e Tecnol*. 2003;13(2):250–264.
61. Vázquez MIN, Chiñas EM, Martínez FMC, Ruvalcaba RM. *Algunos Aspectos Basicos de la Química Computacional*. UNAM; 2006.
62. Cruz Regalado PI (Universidad del P. Modelo Teórico Experimental de la Relación Cuantitativa entre la Estructura Química y el Tiempo de Retención (QSRR) para algunos compuestos volátiles del café. Published online 2018.
63. Frank Jensen. *Introduction to Computational Chemistry*. 2a ed. (Ltd JW& S, ed.); 2007.
64. Fiolhais C, Nogueira F, Marques MAL. *A Primer in Density Functional Theory*. (Springer, ed.); 2003.
65. S. Sholl D, A. Steckel J. *Density Functional Theory: A Practical Introduction*

- David Sholl, Janice A Steckel - Google Libros. Published 2009. Consultado enero 19, 2020. https://books.google.com.mx/books?hl=es&lr=&id=_f994dmAdv0C&oi=fnd&pg=PT7&dq=.+Density+functional+theory:+a+practical+introduction:+John+Wiley+%26+Sons.&ots=hbDM8jnwra&sig=-OqTDHCoty3G_YLBgzWn4jO7REU&redir_esc=y#v=onepage&q=.Density+functional+theory%3A
66. Sánchez Montero JM. Metodologías de modelado molecular en el diseño, síntesis y explicación racional de resultados. *Real Acad Nac Farm.* 2016;2(2):168–184. Consultado enero 19, 2020. <https://www.analesranf.com/index.php/aranf/article/view/1708>
 67. Lozano Ponte J, Scior T. ¿Qué sabe Ud. Acerca de ... Dinámica Molecular? *Revista mexicana de ciencias farmacéuticas.* Published 2014. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-01952014000100010&lng=es&tlng=es.
 68. Lozano Ponte J, Scior T. ¿Qué sabe Ud. acerca de ... QSAR? *Revista mexicana de ciencias farmacéuticas.* Published 2012. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-01952012000200010&lng=es&tlng=es.
 69. Andrés Bort JM, Alvariño Herrero JM, Borondo Rodríguez F, Caballol Lorenzo R, Frau Munar J. *Química teórica y computacional.*; 2001.
 70. Caballol R, Sola M. Métodos de la química cuántica. En: Andrés J, Bertrán J, eds. *Química teórica y computacional.* Publicacions de la Universitat Jaume I; 2001:544.
 71. S JD. Correlación. *Rev Chil Anest.* Published online 2014:150–153. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=http%3A%2F%2Fwww.sachile.cl%2Fupfiles%2Frevistas%2F54e63a1a778ff_15_correlacion-2-2014_edit.pdf&clen=325699&chunk=true
 72. Escalona Arranz JC, Padrón Alexander JG, Carrasco Vela R. *Introducción al diseño racional de fármacos.* DR.C.Raúl.; 2008. https://books.google.com.mx/books?hl=es&lr=&id=5eDzDwAAQBAJ&oi=fnd&pg=PP4&dq=Introducción+al+diseño+racional+de+fármacos.Escalona+Arranz,+Julio+César+Padrón+Alexander,+Juan+García+Carrasco+Vela,+Ramón&ots=Je5aiaBNSr&sig=B8EQ_JmNpqRLPSa0tIPPvoLuitY&redir_esc=y
 73. Roy K, Kar S, Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst.* 2015;145:22–29. doi:10.1016/J.CHEMOLAB.2015.04.013
 74. Nicolotti O (University of B, Gadaleta DG (University of B, Felice G (University of B, Catto M (University of B, Carotti A (University of B. Applicability Domain for QSAR Models: Where Theory Meets Reality. *Int J Quant Struct Relationships.* 2016;1. <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fwww.researchgate.net%2Fprofile%2FMarco->

Catto%2Fpublication%2F291955223_Applicability_Domain_for_QSAR_Models%2Flinks%2F56ab2fa608aeadd1bdcc1841%2FApplicability-Domain-for-

75. Anderson MJ, Legendre P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J Stat Comput Simul.* 2007;62(3):271–303. doi:10.1080/00949659908811936
76. Liu H, Shah S, Jiang W. On-line outlier detection and data cleaning. *Comput Chem Eng.* 2004;28(9):1635–1647. doi:10.1016/J.COMPCHEMENG.2004.01.009
77. Tukur S, Shallangwa GA, Ibrahim A. Theoretical QSAR modelling and molecular docking studies of some 4-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme inhibitors potentially used as herbicides. *Heliyon.* 2019;5(11). doi:10.1016/J.HELİYON.2019.E02859
78. Gajula MN., Kumar A, Ijaq J. Protocol for Molecular Dynamics Simulations of Proteins. *BIO-PROTOCOL.* 2016;6(23). doi:10.21769/bioprotoc.2051.

Anexo.

Tabla 20.-Inhibidores selectivos por la proteína GSK3β.				
CHEMBL1980142	CHEMBL546797	CHEMBL1682552	CHEMBL1990912	CHEMBL1974935
CHEMBL1965988	CHEMBL1979577	CHEMBL1999496	CHEMBL2006580	CHEMBL1992922
CHEMBL336961	CHEMBL2002240	CHEMBL2007002	CHEMBL1991734	CHEMBL1997129
CHEMBL1978200	CHEMBL1985681	CHEMBL462120	CHEMBL1972220	CHEMBL1419458
CHEMBL1970903	CHEMBL1971186	CHEMBL1973359	CHEMBL1968406	CHEMBL1989029
CHEMBL1974870	CHEMBL1987430	CHEMBL1982924	CHEMBL1964687	CHEMBL2004515
CHEMBL296586	CHEMBL2000508	CHEMBL1978371	CHEMBL1977148	CHEMBL1972119
CHEMBL1979855	CHEMBL1984296	CHEMBL2003768	CHEMBL1986186	CHEMBL2005449
CHEMBL1682345	CHEMBL1983025	CHEMBL316264	CHEMBL2000429	CHEMBL2000029
CHEMBL1991188	CHEMBL1999126	CHEMBL302449	CHEMBL1990904	CHEMBL340384
CHEMBL539474	CHEMBL1516890	CHEMBL1624529	CHEMBL1974803	CHEMBL191003
CHEMBL1991429	CHEMBL1972934	CHEMBL1986328	CHEMBL1980167	CHEMBL1983111
CHEMBL1990254	CHEMBL1975233	CHEMBL1979718	CHEMBL1972355	CHEMBL1983575
CHEMBL1989646	CHEMBL210887	CHEMBL360847	CHEMBL1965589	CHEMBL1977041
CHEMBL2002726	CHEMBL1982135	CHEMBL1973211	CHEMBL2004716	CHEMBL1986143
CHEMBL1988581	CHEMBL1987009	CHEMBL1996831	CHEMBL2006933	CHEMBL207995
CHEMBL1999590	CHEMBL1994321	CHEMBL1972346	CHEMBL223367	CHEMBL1999778
CHEMBL1988141	CHEMBL514499	CHEMBL1976240	CHEMBL2001987	CHEMBL2003456
CHEMBL1969049	CHEMBL1990885	CHEMBL1982563	CHEMBL1981744	CHEMBL1999414
CHEMBL17370	CHEMBL1979883	CHEMBL1998585	CHEMBL1968070	CHEMBL560813
CHEMBL1999428	CHEMBL1976732	CHEMBL1984402	CHEMBL1996510	CHEMBL1975212
CHEMBL1983449	CHEMBL1969755	CHEMBL1242373	CHEMBL1970913	CHEMBL1987815
CHEMBL413779	CHEMBL2002446	CHEMBL1995172	CHEMBL1987054	
CHEMBL1969372	CHEMBL1984700	CHEMBL2002450	CHEMBL1964644	
CHEMBL1997349	CHEMBL1991143	CHEMBL1997193	CHEMBL1985074	
CHEMBL1988387	CHEMBL2004771	CHEMBL1992042	CHEMBL1988173	
CHEMBL373598	CHEMBL1969735	CHEMBL1421720	CHEMBL1971801	
CHEMBL1725279	CHEMBL2001257	CHEMBL1990288	CHEMBL1977138	