



**FACULTAD DE CIENCIAS
PROGRAMA DE LICENCIATURA EN CIENCIAS MENCIÓN
BIOLOGÍA O QUÍMICA**

Relaciones de estructura-actividad cuantitativas tridimensionales (3D-QSAR) CoMFA/CoMSIA y diseño de compuestos de estructura (5-imidazol-2-il-4-fenilpirimidin-2-il)[2- (2-piridilamino)etil]amina como inhibidores potentes y selectivos de la enzima glucógeno sintasa quinasa 3 (GSK3) con capacidad antidiabetes tipo II en modelos animales

**TESIS PARA OPTAR AL GRADO DE
LICENCIADO EN CIENCIAS MENCIÓN QUÍMICA**

David Josué Cabezas González

PROFESOR GUÍA

Dr. Jaime Mella Raipán

2018

DAVID JOSUÉ CABEZAS GONZÁLEZ

Trabajo de titulación presentado en conformidad a los requisitos para optar al grado académico de Licenciatura en Ciencias mención Química

COMISIÓN DE TÍTULO

Dr. Jaime Mella Raipán

Director

Universidad de Valparaíso

Dr. Jorge Soto Delgado

Miembro de la comisión

Universidad Andrés Bello

Dr. Gonzalo Recabarren Gajardo

Miembro de la comisión

P. Universidad Católica de Chile

ÍNDICE

ÍNDICE DE TABLAS.....	8
ÍNDICE DE FIGURAS.....	9
RESUMEN.....	12
1. INTRODUCCIÓN.....	13
1.1. GSK3.....	13
1.1.1. ¿Qué es GSK-3?.....	13
1.1.1.1. Estructura enzimática e isoforma.....	13
1.1.1.2. Historia.....	16
1.1.1.3. Función y distribución a nivel celular.....	16
1.1.2. Importancia terapéutica de GSK3.....	18
1.1.2.1. Enfermedad de Alzheimer.....	18
1.1.2.2. Enfermedades psiquiátricas.....	19
1.1.2.3. Cáncer.....	19
1.1.2.4. Diabetes.....	20
1.1.3. Moléculas inhibidoras de GSK3.....	27
1.2. QSAR.....	30
1.2.1. Introducción.....	30
1.2.2. QSAR tradicional (2D-QSAR).....	31
1.2.3. Descriptores fisicoquímicos.....	31
1.2.3.1. Lipofilia.....	32
1.2.3.2. Efectos electrónicos.....	34
1.2.3.3. Factores estéricos.....	37
1.2.3.4. Factor estérico de Taft (E_s).....	37
1.2.3.5. Refractividad molar (RM).....	38
1.2.3.6. Ecuación de Hansch.....	39
1.2.3.7. Análisis de Free-Wilson.....	40
1.2.4. 3D-QSAR.....	41
1.2.5. Análisis comparativo de campos moleculares, CoMFA.....	41
1.2.5.1. Campo estérico y electrostático.....	41
1.2.5.2. Relación de forma y distribución electrónica con la actividad biológica.....	46
1.2.5.3. Potencial hidrofóbico.....	48

1.2.5.4. Ventajas del 3D-QSAR sobre el QSAR tradicional.....	48
1.2.5.5. Problemas potenciales de 3D-QSAR	50
1.2.5.6. Análisis Comparativo de Índices de Similitud Molecular, CoMSIA	51
2. HIPÓTESIS	53
3. OBJETIVOS.....	54
3.1. Objetivo General	54
3.2. Objetivos Específicos.....	54
4. MATERIALES Y MÉTODOS.....	55
4.1. Compilación de la base de datos.....	55
4.1.1. Generación de las moléculas	55
4.1.2. Realización del análisis de minimización de energía.....	78
4.2. Alineamiento molecular	79
4.3. Análisis QSAR	80
4.3.1. Cálculo de los campos CoMFA y CoMSIA	80
4.3.2. Selección de los mejores confórmeros	81
4.3.3. Validación interna y análisis de mínimos cuadrados parciales (PLS)	83
4.3.4. Validación externa.....	83
4.3.5. Dominio de aplicabilidad (DA).....	84
4.4. Obtención de mapas de contornos y predicciones.....	85
5. RESULTADOS Y DISCUSIÓN.....	86
5.1. Resultados estadísticos	86
5.2. Análisis de mapas de contornos.....	98
5.2.1. CoMFA.....	98
5.2.2. CoMSIA	103
5.3. Valores atípicos (“Outliers”)	112
5.4. Dominio de Aplicabilidad	114
5.5. Predicciones	114
6. CONCLUSIONES.....	118
7. PROYECCIONES.....	119
8. REFERENCIAS	120
9. ANEXO	132

AGRADECIMIENTOS

Primero que todo, quiero agradecer a Dios por darme la oportunidad de llegar a esta instancia y acompañarme en todo momento.

Al profesor Jaime Mella por abrirme las puertas a su laboratorio desde un inicio cuando era mechón y le mostré mi interés en el área de la Química Medicinal. Un buen profesor guía y motivador tanto en el tema académico como espiritual. Y por supuesto, un buen rival en Street Fighter.

Agradecer a mis padres que siempre me alentaron a nunca rendirme en mis estudios y por demostrarme el orgullo que sienten por mí y su amor. A Martín Cabezas, mi hermano pequeño, y a mi primo Benjamín Maturana, que es como mi hermano del medio, por darme muchas alegrías y buenos ratos en los momentos que debía viajar a Peumo y distraerme un poco de las actividades universitarias. También a la congregación en la que me crie en mi comuna Peumo, y a mi Pastor Ricardo Reyes por siempre apoyarme en muchas situaciones.

A mi amada novia, Muriel Orellana Escobar, por ser una parte fundamental en todo este proceso. Me ha dado de su apoyo durante todo este tiempo y también es una fuente de motivación para no rendirme y cumplir mis metas y nuestros respectivos sueños. Y también por ayudarme a ordenar mis ideas y contribuir en la búsqueda de nuevas moléculas. Para mí, ella es un gran regalo que Dios me entregó, la mejor novia, la mejor amiga, y compañera.

A mis amigos de la universidad, en especial a Juan José Alvear (el Juan), ya que fue el primer amigo que tuve en la universidad que pasó a ser como un hermano más, y que incluso su familia me veía como uno más. Nos apoyamos en muchas ocasiones y espero

sigamos así por mucho tiempo. También a Gustavo Vásquez (Karl) que, aunque no compartimos muchas asignaturas, siempre estaba dispuesto para conversar y en poco tiempo demostró ser un gran amigo. Por supuesto también agradezco a Rolando Carrizo (el rolo), Daniel Alvares (Dinodan), Felipe González (feli), Gabriel Rodrigues (Gabo), Michael Zamora (el mike), y muchos más que han estado en distintos momentos compartiendo alegrías, rateando en la biblioteca (no todo era estudio) y en juntas diversas.

A mis compañeros de la mención Química con los cuales compartí momentos de estrés durante los ramos de la mención, nos apoyábamos para que ninguno quedara atrás, y luchábamos por salir adelante como un buen equipo.

También a mis amigos y amigas de Peumo y alrededores como: Alan Díaz, José Donoso, Rodrigo Reyes, Natanael Reyes, Ivan Reyes, Sergio Salazar, Guillermo Rojas, Yulissa Rubio, Angelo Valenzuela, Fernanda Rubio, Barbara Tobar, Yan Díaz, Katy Díaz, Felipe Tobar, y a mis amigos del colegio Carlos Carvajal, Tomás González, y a Paulo Miranda.

Finalmente, a todos y todas las personas que de una forma u otra han estado presentes en distintas etapas de mi vida. Dios me ha recompensado con grandes personas y amigos que, al pasar de los años, siguen estando presentes ya sea de forma cercana o distante.

A todos, muchas gracias.

ÍNDICE DE TABLAS

Tabla 1: Estructura y valores de IC_{50} del conjunto de moléculas analizadas ⁹⁹	56
Tabla 2: Parámetros estadísticos y combinaciones de campos para CoMFA y CoMSIA. .	87
Tabla 3: Resumen de los parámetros de validación externa para el CoMFA y el CoMSIA.	89
Tabla 4: Valores de pIC_{50} experimentales, predictivos y valores residuales para los modelos CoMFA y CoMSIA.	89
Tabla 5: Valores de q^2 y r^2_{ncv} después de varias pruebas de la aleatorización-Y.	97
Tabla 6: Estructuras de los compuestos predichos a partir de los resultados obtenidos en los mapas de contornos CoMFA y CoMSIA.	115
Tabla S1: Fármacos inhibidores de GSK-3 reportados por Meijer et al ⁷⁰	132
Tabla S2: Valores de π para diversos sustituyentes en el anillo aromático.....	133
Tabla S3: Valor de σ para diversos sustituyentes en el anillo aromático. Se muestran valores para el sustituyente en la posición meta- (σ_m) y en la posición para- (σ_p).	133
Tabla S4: Valores de E_s para distintos sustituyentes.....	134

ÍNDICE DE FIGURAS

Figura 1: Cebado de la fosforilación de la glucógeno sintasa por la GSK-3. Al residuo Ser o Thr que se ha fosforilado en el sustrato se le asignó el índice 0; los residuos del lado amino-terminal de este residuo se numeran -1, -2, y así sucesivamente; los residuos sobre el lado carboxi-terminal se numeran +1, +2, etc ⁶	14
Figura 2: Estructura GSK-3 destacando su zona de pseudosustrato ⁶	14
Figura 3: Representación esquemática de GSK-3 α y GSK-3 β de mamíferos. Los sitios de fosforilación de serina y tirosina están indicados con puntas de flechas azules. Se resalta el N-terminal rico en glicina, exclusivo de GSK-3 α , y el dominio de quinasa conservado en ambas isoformas ¹	15
Figura 4: Esquema de la insulina; el péptido C está representado por la línea continua a partir de los dos puntos negros.....	22
Figura 5: Ruta desde la insulina a la GSK-3 y la glucógeno sintasa. La inactivación de la GSK-3 permite que la fosfoproteína fosfatasa 1 (PP1) desfosforile la glucógeno sintasa, convirtiéndola en su forma activa. De esta forma, la insulina estimula la síntesis del glucógeno.....	23
Figura 6: Estructura de sulfonilureas, metformina y troglitazona (tiazolidindionas).	24
Figura 7: Estructura de algunos de los inhibidores farmacológicos más estudiados de la glucógeno sintasa quinasa 3 (GSK-3). Los átomos que interactúan con la cadena lateral de Val135 y Asp133 de la GSK-3 se muestran en rojo (aceptores de hidrógeno) y azul (donantes de hidrógeno) ⁷⁰	28
Figura 8: Compuestos preclínicos utilizados como patrón para la generación de nuevos derivados. En el lado izquierdo se encuentra CHIR-99021 y en el lado derecho se encuentra CHIR-98014.....	29
Figura 9: Curva parabólica de $\log(1/C)$ versus $\log P$	33
Figura 10: Ionización del ácido benzoico.	35
Figura 11: Posición de equilibrio dependiendo del sustituyente X.....	36
Figura 12: Identificación de la conformación activa y farmacóforo. El farmacóforo en este ejemplo contempla los átomos de oxígeno, el átomo de nitrógeno, y el centro del anillo aromático.....	42
Figura 13: (a) Representación de un farmacóforo con sus propiedades químicas y distancias entre sus diferentes regiones. HBA, aceptor de puente de hidrogeno; RA, anillo aromático; HY, zona hidrofóbica; ZB enlace de cinc. (b) Una molécula solapada con el farmacóforo.	42
Figura 14: Representación del farmacóforo (poliedro de puntos rojos) colocado al interior del enrejado virtual. Cada vértice representa un punto en el espacio.....	44
Figura 15: Ejemplo de una serie de análogos en estudio alineados todos respecto a un farmacóforo de referencia.	44

Figura 16: Proceso CoMFA. Las “diferencias de campo estérico” son las sumas de raíz de las diferencias cuadradas entre los valores de celda dentro de un par de filas en la tabla QSAR ⁹⁴	45
Figura 17: Se muestran dos mapas de contornos para una misma molécula donde se señala mediante color verde los sitios donde un sustituyente voluminoso mejoraría la actividad biológica, y en amarillo donde un sustituyente voluminoso perjudica la actividad biológica. El color rojo indica que un sustituyente con potencial electrostático negativo incrementa la actividad biológica, mientras que en las zonas azules un sustituyente con potencial electrostático positivo beneficia la actividad.	47
Figura 18: Alineamiento del set de moléculas utilizados en la metodología 3D-QSAR. ...	80
Figura 19: Histograma de la frecuencia de distribución de datos.	82
Figura 20: Gráfica de regresión del modelo CoMFA.	93
Figura 21: Gráfica de regresión del modelo CoMSIA.	93
Figura 22: Gráfico de residuales del modelo CoMFA.	94
Figura 23: Gráfico de residuales del modelo CoMSIA.	95
Figura 24: Gráfico relación actividad experimental versus actividad predicha del modelo CoMFA.	96
Figura 25: Gráfico relación actividad experimental versus actividad predicha del modelo CoMSIA.	96
Figura 26: Esquema general de las moléculas utilizadas para el estudio QSAR y en la discusión de los mapas de contornos.	98
Figura 27: Mapa de contornos estéricos del modelo CoMFA. Los poliedros verdes indican que es favorable para la actividad biológica la presencia de compuestos voluminosos, mientras que los poliedros amarillos indican que no es favorable el uso de sustituyentes voluminosos. En A) está la molécula más activa (130) y en B) la menos activa (41).	100
Figura 28: Mapa de contornos electrostáticos del modelo CoMFA. Los poliedros azules indican que es favorable para la actividad biológica la presencia de sustituyentes electropositivos, mientras que los poliedros rojos indican que es favorable la presencia de sustituyentes electronegativos. En A) está la molécula más activa (130), y en B) la menos activa (41).	102
Figura 29: Mapa de contornos estéricos del modelo CoMSIA. Los poliedros verdes indican que es favorable para la actividad biológica la presencia de sustituyentes voluminosos, mientras que los poliedros amarillos indican que no es favorable el uso de sustituyentes voluminosos. En A) está la molécula más activa (130) y en B) la menos activa (41).	104
Figura 30: Mapas de contornos hidrofóbico. El poliedro amarillo indica que es favorable para la actividad biológica el uso de sustituyentes hidrofóbicos, mientras que el	

	poliedro gris indica que es favorable el uso de sustituyentes hidrofílicos. En A) está la molécula más activa (130) y en B) la menos activa (41).	106
Figura 31:	Mapas de contornos dador de puentes de hidrógeno. El poliedro cian indica que es favorable para la actividad biológica el uso de sustituyentes dadores de puentes de hidrógeno, mientras que el poliedro purpura indica que no es favorable el uso de sustituyentes dadores de puentes de hidrógeno. En A) está la molécula más activa (130) y en B) la menos activa (41).	108
Figura 32:	Mapas de contornos aceptor de puentes de hidrógeno. El poliedro magenta indica que es favorable para la actividad biológica el uso de sustituyentes aceptores de puentes de hidrógeno, mientras que el poliedro rojo indica que no es favorable la presencia de sustituyentes aceptores de puentes de hidrógeno. En A) está la molécula más activa (130) y en B) la menos activa (41).	111
Figura 33:	Principales relaciones estructura-actividad encontradas en este estudio.	118

RESUMEN

La Glucógeno sintasa quinasa 3 (GSK-3) es una enzima de tipo quinasa que originalmente fue estudiada por inhibir la glucógeno sintasa, enzima clave en la síntesis del glucógeno. Actualmente la GSK-3 está relacionada con el tratamiento de diversas enfermedades, como trastornos psiquiátricos, enfermedad de Alzheimer y cáncer. Por otra parte, existen estudios que demuestran que la insulina inhibe a GSK-3 para mantener activa la glucógeno sintasa. Por lo que inhibir a la GSK-3 tendría un efecto pro-insulínico indirecto en relación con la síntesis del glucógeno y, por lo tanto, sería una útil alternativa para el tratamiento de la diabetes tipo II.

En una reciente publicación, Allan et al. sintetizaron 135 compuestos de estructura (5-imidazol-2-il-4-fenilpirimidin-2-il)[2-(2-piridilamino)etil]amina como inhibidores de GSK-3 con la finalidad de obtener una nueva familia de compuestos aminopirimidínicos con actividad antidiabética y antiobesidad. En el presente trabajo de tesis, llevamos a cabo un extenso estudio 3D-QSAR en el set de 135 moléculas con el fin de encontrar útiles relaciones de estructura-actividad y de esta forma disponer de una base para el diseño racional de nuevos compuestos con valores de actividad biológica promisorios para ser utilizados como inhibidores de GSK-3 con potencial eficacia terapéutica en diabetes tipo II. Los resultados fueron significativos. Se obtuvieron valores de q^2 para CoMFA-SE y CoMSIA-SHDA de 0.563 y 0.598 respectivamente, y valores de r^2_{test} de 0.823 y 0.925 respectivamente. Estos valores fueron validados estadísticamente y a partir de esto se generaron los mapas de contornos y una propuesta de 10 nuevos compuestos con valores de actividad inhibitoria $pIC_{50} > 9.00$.

1. INTRODUCCIÓN

1.1. GSK3

1.1.1. ¿Qué es GSK-3?

1.1.1.1. Estructura enzimática e isoforma

La glucógeno sintasa quinasa 3 (GSK-3) es una enzima serina/treonina quinasa funcional que se encuentra en todas las células eucariontes. Funciona como un regulador clave en numerosas vías de señalización, incluidas las respuestas celulares a Wnt, receptor de tirosina quinasa y receptores acoplados a proteína G¹. Esta enzima presenta más de un centenar de proteínas que se han identificado como sustratos de GSK-3, y se ha realizado un gran número de estudios bioinformáticos²⁻⁴. Originalmente, esta enzima fue identificada por su papel como regulador negativo de la síntesis del glucógeno⁵. Como se muestra en la **Figura 1**, la glucógeno sintasa quinasa 3 se asocia primeramente con su sustrato (en este caso la glucógeno sintasa) mediante la interacción entre tres residuos cargados positivamente (Arg⁹⁶, Arg¹⁸⁰, Lys²⁰⁵) y un residuo de fosfoserina en la posición +4 del sustrato⁶. En la **Figura 2**, se aprecia que la estructura de la GSK-3 tiene un residuo Ser cerca de su extremo amino que puede ser fosforilado. Esto produce una región “pseudosustrato” en esta enzima, la cual se pliega dando un sitio cebador que hace que el sitio activo sea inaccesible a otra proteína sustrato, inhibiendo la GSK-3 hasta que el grupo fosforilo de cebado de su región pseudosustrato es eliminado por una fosfoproteína fosfatasa (PP1 en el músculo, otra fosfatasa en el hígado)⁶.

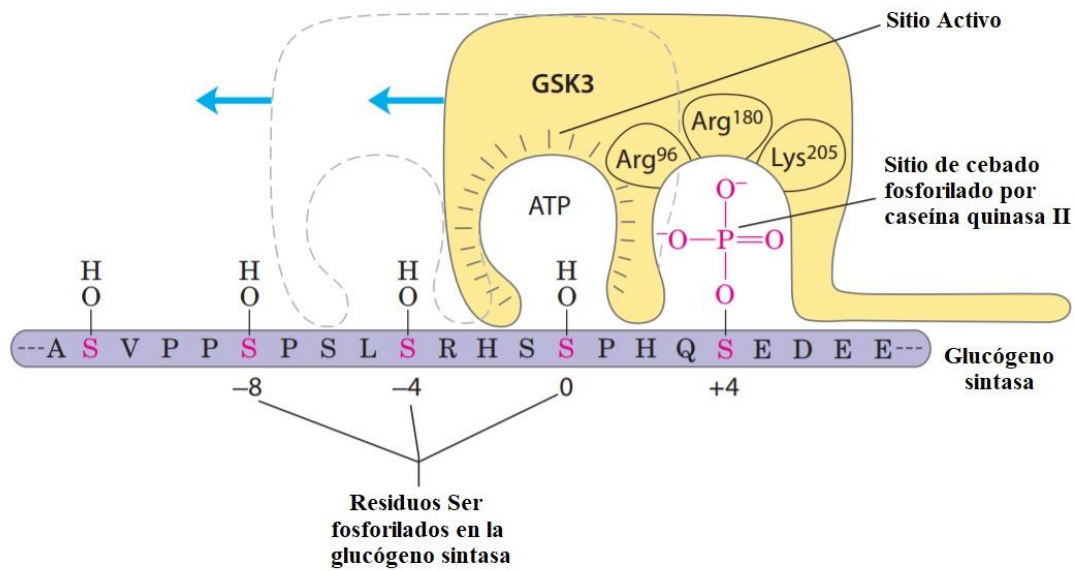


Figura 1: Cebado de la fosforilación de la glucógeno sintasa por la GSK-3. Al residuo Ser o Thr que se ha fosforilado en el sustrato se le asignó el índice 0; los residuos del lado amino-terminal de este residuo se numeran -1, -2, y así sucesivamente; los residuos sobre el lado carboxi-terminal se numeran +1, +2, etc⁶.

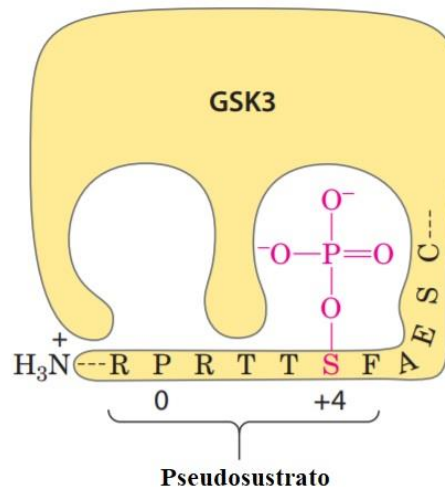


Figura 2: Estructura GSK-3 destacando su zona de pseudosustrato⁶.

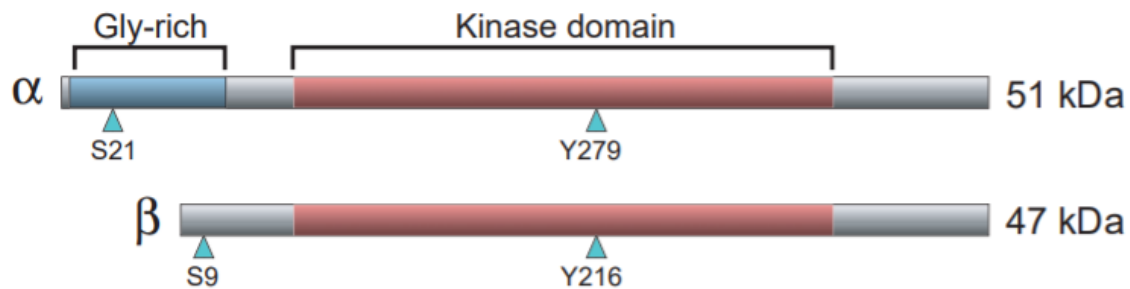


Figura 3: Representación esquemática de GSK-3 α y GSK-3 β de mamíferos. Los sitios de fosforilación de serina y tirosina están indicados con puntas de flechas azules. Se resalta el *N*-terminal rico en glicina, exclusivo de GSK-3 α , y el dominio de quinasa conservado en ambas isoformas¹.

En los mamíferos, GSK-3 existe en 2 formas, GSK-3 α y GSK-3 β , estas comparten un dominio catalítico altamente conservado pero difieren en el hecho de ser codificados por genes distintos, en humanos son llamados GSK-3A y GSK-3B⁷. La GSK-3 α tiene una masa de 51 kDa, mientras que la GSK-3 β es una enzima de 47 kDa. La diferencia de tamaño se debe a una extensión rica en glicina en el *N*-terminal de GSK-3 α . Aunque son altamente homólogos en su dominio quinasa (98% de similitud), los dos productos genéticos comparten solo el 36% de identidad en los últimos 76 residuos C-terminales⁸ (**Figura 3**). Los homólogos de GSK-3 existen en todas las células eucariontes estudiadas hasta la fecha¹. Aunque GSK-3 α y GSK-3 β son estructuralmente idénticos, no lo son funcionalmente. Esto se hizo evidente después de la ablación de la isoforma GSK-3 β en ratones, lo que dio como resultado un fenotipo embrionario letal⁹. Los embriones que llevan delaciones homocigóticas del exón 2 de GSK-3 β mueren alrededor de los 16 días embrionarios debido a la degeneración hepática masiva causada por la apoptosis de hepatocitos. La incapacidad de GSK-3 α para rescatar a los ratones con nulo GSK-3 β indica que la degeneración del fenotipo hepático surge

específicamente por la pérdida de la isoforma beta. Este fenotipo es notablemente similar al de los animales que carecen de quinasas tipo 2 como ReIA o I-kB¹⁰.

1.1.1.2. Historia

La glucógeno sintasa quinasa 3 ha sorprendido a los investigadores de transducción de señales desde su detección en el músculo esquelético hace más de 25 años¹¹. La GSK-3 recibió el nombre debido a la base de su función, la cual fue descrita en primera instancia como una quinasa que inactiva a la glucógeno sintasa¹². Se consideró de interés porque tiene características no convencionales para que una quinasa sea constitutivamente activa¹³. Después de la estimulación del receptor de insulina, la fosfoinositido-3-quinasa y la Akt son activadas, lo que conduce a la fosforilación y la inactivación de la GSK-3¹².

1.1.1.3. Función y distribución a nivel celular

Se ha considerado tradicionalmente que GSK-3 es principalmente una proteína citosólica. Sin embargo, GSK-3 también está presente en las mitocondrias y el núcleo, así como en otros compartimentos subcelulares, donde sus niveles y/o su estado de activación puede regularse mediante actividades de señalización localizadas. La GSK-3 mitocondrial puede ser especialmente importante en el estrés oxidativo y ciertas condiciones apoptóticas¹⁴. Sin embargo, se sabe poco sobre los mecanismos que regulan los niveles y actividades de GSK-3 en la mitocondria, estos temas pueden proporcionar información importante sobre posibles funciones patológicas de GSK-3 mitocondrial en enfermedades que implican la disfunción mitocondrial¹³. En contraste con sus acciones mitocondriales, se han identificado muchos efectos nucleares de GSK-3, particularmente aquellos que involucran la regulación y expresión del gen. GSK-3 ingresa al núcleo a través de un transportador nuclear (CRM1)

por secuencias de aminoácidos básicos (lisina y arginina) de GSK-3 β ¹⁵. Los niveles nucleares de GSK-3 β aumenta rápidamente en respuesta a algunos estímulos apoptóticos¹⁶. Al ser distinta la región *N*-terminal de la GSK-3 α , genera la exclusión nuclear que puede llegar a ser contra-activada mediante señalización de calcio¹⁷.

Recientemente, se han detectado efectos moduladores significativos de GSK-3 en proteínas involucradas en la epigenética. La evidencia es particularmente fuerte ya que GSK-3 contribuye a la regulación de la modificación de histonas, las que juegan un papel central en la epigenética. GSK-3 puede fosforilar directamente la histona 1.5¹⁷ y promueve la fosforilación de la histona 3, aunque el mecanismo no fue identificado¹⁸. Además de los compartimentos clásicos nucleares y mitocondriales, GSK-3 también está compartimentado en otras estructuras subcelulares, como en los “conos de crecimiento”¹⁹⁻²¹, y en complejos proteicos.

Poco se sabe sobre los mecanismos que regulan el tráfico de GSK-3 dentro de los dominios citosólicos, lo que probablemente tenga consecuencias importantes para las funciones de GSK-3, aunque la recuperación de fluorescencia después de un análisis de “Photobleaching” (FRAP) indicó que el 84% de GSK-3 β citosólico era móvil en fibroblastos embrionarios de ratón¹⁵. La localización de GSK-3 obviamente es un tema de importancia para la regulación de su acción, pero aún quedan muchos estudios a realizar sobre los mecanismos que regulan el tráfico celular de GSK-3 y las diversas consecuencias funcionales¹³.

1.1.2. Importancia terapéutica de GSK3

1.1.2.1. Enfermedad de Alzheimer

Las mutaciones en los genes app, ps-1 y ps-2 resultan de la aparición de la enfermedad del Alzheimer familiar (FAD). Aunque en muchos casos, esas mutaciones resultan en un aumento de la cantidad de péptidos beta amiloide, no hay una correlación clara entre esa cantidad y el momento del inicio de la enfermedad. Por lo tanto, otros factores pueden explicar las mutaciones de estos genes, como la actividad de la GSK-3²². En FAD, las mutaciones en app o en los genes de preinsulina facilitan la producción de péptidos beta amiloides que podrían interferir con las vías de la insulina o Wnt²³⁻²⁵ y, por otro lado, promovería un aumento de la actividad de GSK-3^{26,27}. Por otra parte, las mutaciones en ps-1 pueden afectar a la interacción de esta proteína con la subunidad reguladora de PI3K, disminuyendo la actividad de esta quinasa y la actividad de GSK-3²⁸. En la enfermedad del Alzheimer esporádico (SAD), la presencia de diferentes isoformas de ApoE podrían interferir, de una manera diferente, con un correceptor de Wnt²⁹⁻³¹ y podría afectar a la actividad de GSK-3. También se ha observado un polimorfismo de GSK-3 asociado con la enfermedad del Alzheimer³². Por lo tanto, hay una convergencia del papel de las diferentes proteínas relacionadas con FAD y SAD con un aumento de la actividad de GSK-3²². Finalmente, también se ha visto que un aumento en la actividad de GSK-3 daría como resultado la fosforilación de otros sustratos que pueden ser tóxicos para las neuronas corticales que se degenerarían, como se ha encontrado en pacientes con la enfermedad del Alzheimer en etapas posteriores de la enfermedad y que podrían estar relacionados con el

inicio de la demencia, después de afectar los lóbulos frontal y parietal. Es por esto que se sugiere que GSK-3 juega un papel importante en la enfermedad del Alzheimer²².

1.1.2.2. Enfermedades psiquiátricas

Alrededor del 10% de los casos de demencia son causados por una mutación en el gen tau. Y existen estudios que relacionan esta mutación con la actividad de GSK-3³². De todas formas, es probable que las enfermedades psiquiátricas sean una de las principales clases que son susceptibles a la estrategia terapéutica inhibidora de GSK-3. Pacientes con trastornos bipolares, o los llamados maníaco-depresivos, ya se están tratando con Litio de manera eficiente. Esto se debe a que el Litio es un inhibidor de GSK-3^{33,34}. Aunque el mecanismo de acción terapéutico del Litio queda aún por esclarecerse definitivamente, existe evidencia sustancial que indica que la inhibición de GSK-3 es un componente importante en la capacidad de estabilización del estado del ánimo^{35,36}.

1.1.2.3. Cáncer

Existe un gran interés en GSK-3 como posible objetivo terapéutico en distintos tipos de cáncer³⁷⁻⁴⁰. Sin embargo, la aplicación en el cáncer se complica por algunos hallazgos que indican que la GSK-3 puede actuar como un supresor tumoral o puede promover la proliferación celular en diferentes tipos de cáncer. La acción de supresión tumoral está ejemplificada por la fosforilación de β -catenina mediante GSK-3 y su posterior degradación, un coactivador transcripcional que a menudo promueve la proliferación celular⁴¹. Hay evidencia de que la quinasa GSK-3 puede actuar como supresor tumoral en el cáncer de piel^{42,43}, y el cáncer de mamas⁴⁴⁻⁴⁶, y la represión de la transcripción de ARN polimerasa 1 por GSK-3 contribuye a su acción supresora tumoral⁴⁷. Por otra parte, las acciones de

formación de tumores de GSK-3 están indicadas por el aumento de los niveles de GSK-3, como en el cáncer de colon⁴⁸ y el cáncer pancreático^{49,50} y en la leucemia de linaje mixto⁵¹, mientras que existen datos contradictorios para el cáncer de próstata^{52,53}.

1.1.2.4. Diabetes

Se le denomina diabetes mellitus a un conjunto de síndromes caracterizados por la presencia de hiperglucemia crónica a la que generalmente se le asocian variablemente un conjunto de complicaciones vasculares, como también complicaciones sistémicas. La diabetes se presenta agrupada en diversos síntomas. La diabetes tipo 1, antiguamente llamada dependiente de insulina, en la cual se aprecia, en el momento de ser diagnosticada, la virtual desaparición de las células β del páncreas (las encargadas de producir insulina) debido a un proceso autoinmune. Este tipo de diabetes suele aparecer por lo general antes de los 30 años, pero puede manifestarse a cualquier edad. En el caso de la diabetes tipo 2, es la de mayor frecuencia y, en general, la que más tarda en aparecer, por lo que se le suele asociar a la obesidad. Previamente fue denominada diabetes no dependiente de insulina. En este caso, el páncreas segrega cantidades muy variables de insulina, de modo que su concentración plasmática puede ser normal o incluso superior a la normal, pero relativamente insuficiente para mantener niveles normales de glucemia. Esto indica que en la diabetes tipo 2 puede generarse una resistencia a la acción de la insulina, es decir, una excesiva producción hepática de glucosa y una deficiente utilización periférica a pesar de estar presente la insulina⁵⁴.

La insulina es un polipéptido de 51 aminoácidos (5.8 kDa) sintetizado preferentemente por las células β del páncreas. Consta de dos cadenas, la A, con 21 aminoácidos, y la B con 30 aminoácidos, unidas entre sí por dos puentes disulfuro (**Figura**

4); la cadena A presenta además otro puente disulfuro entre sus aminoácidos 6 y 11. La liberación de la insulina en el páncreas está sometida a múltiples factores de regulación, químicos, nerviosos y hormonales, pero, como es de esperarse, son las modificaciones de los principales sustratos energéticos (glucosa, aminoácidos, ácidos grasos libres, cuerpos cetónicos) los que inducen modificaciones inmediatas en la respuesta⁵⁴. La insulina se fija a receptores específicos de membrana situados en las células insulino-sensibles. Estos receptores pertenecen al grupo de los receptores con actividad tirosina-quinasa. Los receptores insulínicos son proteínas que constan de 2 subunidades, α y β , cuyos pesos moleculares son 135 y 90 kDa respectivamente. La insulina se fija selectivamente a la subunidad α ; como consecuencia de esta fijación, la subunidad β se autofosforila. Además de fosforilarse, este receptor fosforila a otras proteínas citosólicas llamadas sustratos 1 y 2 del receptor insulínico (IRS1, IRS2) en sus residuos de triosina⁵⁴.

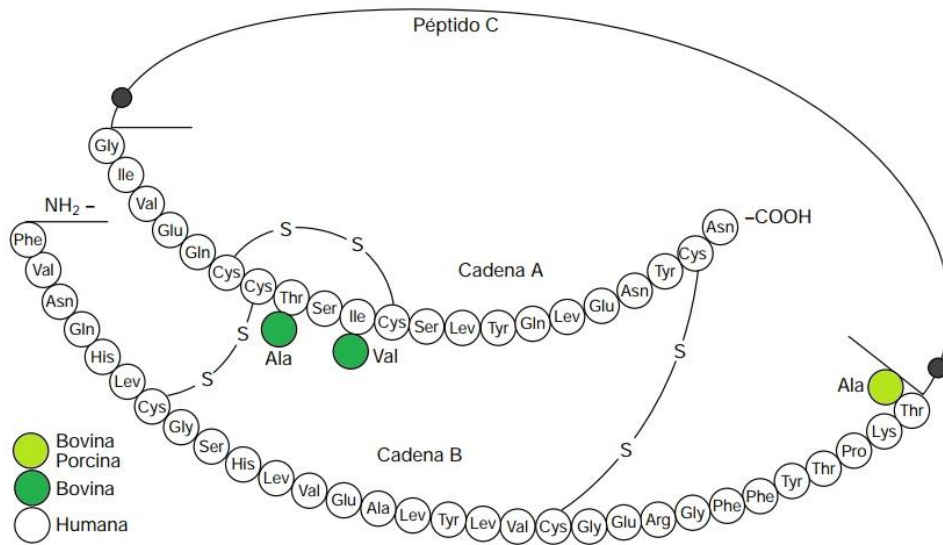


Figura 4: Esquema de la insulina; el péptido C está representado por la línea continua a partir de los dos puntos negros.

La ruta de señalización de la insulina se ramifica en IRS1. La enzima fosfoinositol 3-quinasa (PI3K) se une a IRS1 a través del dominio SH2 de PI3K. Activándose de esta forma, PI3K convierte el lípido de la membrana fosfatidiliositol 4,5-bifosfato (PIP₂) en fosfatidilinositol 3,4,5-trifosfato (PIP₃). La cabeza de PIP₃ genera múltiples cargas que sobresalen en el lado citoplasmático, el cual es inicio para una segunda rama de señalización. Una proteína quinasa (PDK1), que es activada por la unión a PIP₃, activa a una segunda proteína quinasa denominada proteína quinasa B (PKB) la cual es fosforilada. De esta forma, la PKB fosforila residuos Ser o Thr de sus proteínas dianas, una de las cuales es la glucógeno sintasa quinasa 3⁶, tanto a su forma α como β ⁵⁵. Al estar activa, es decir fosforilada, la GSK-3 fosforila a la glucógeno sintasa, inactivandola (**Figura 5**)⁶.

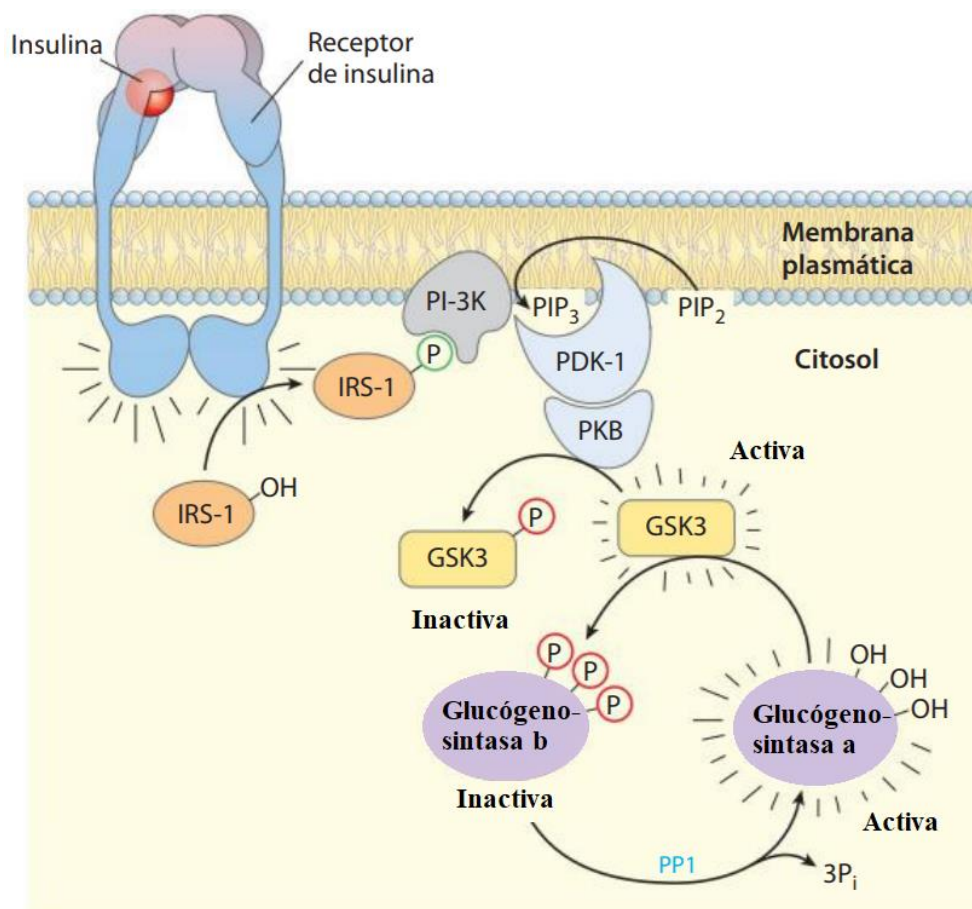


Figura 5: Ruta desde la insulina a la GSK-3 y la glucógeno sintasa. La inactivación de la GSK-3 permite que la fosfoproteína fosfatasa 1 (PP1) desfosforile la glucógeno sintasa, convirtiéndola en su forma activa. De esta forma, la insulina estimula la síntesis del glucógeno.

Para esta enfermedad existen diversas alternativas de tratamiento. Algunos agentes orales como sulfonilureas, metformina y tiazolidindionas (**Figura 6**) en combinación con ejercicio y dieta pueden ayudar a controlar la diabetes medida por la cantidad de hemoglobina glicosilada circulante (HbA1c), pero todavía hay un 30-40% de pacientes que no responden a dicha terapia⁵⁶.

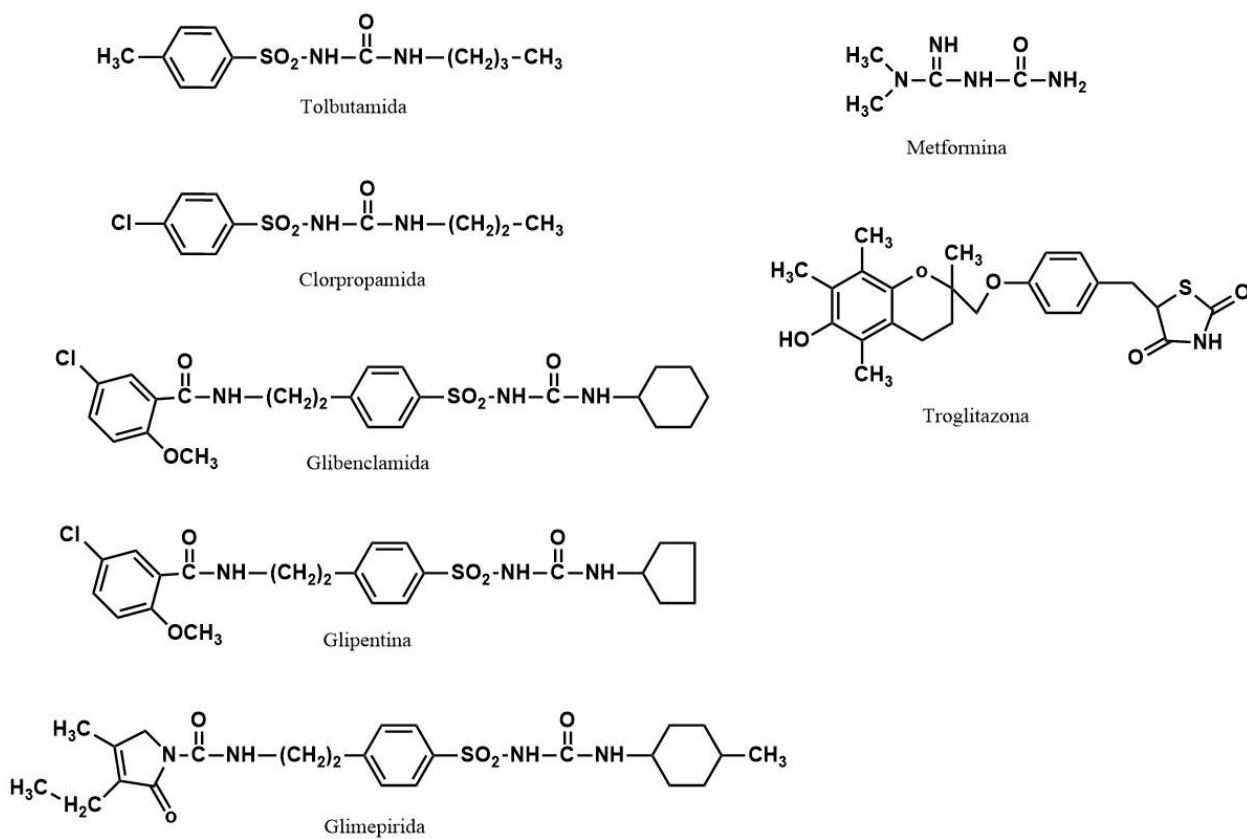


Figura 6: Estructura de sulfonilureas, metformina y troglitazona (tiazolidindionas).

Las sulfonilureas, a corto plazo, provocan la liberación de insulina en las células β del páncreas, ya que aumenta su sensibilidad a la glucosa. Para esto, las sulfonilureas actúan con gran afinidad sobre receptores asociados a los canales de K^+ sensibles a ATP, fijándose de manera específica a la proteína SUR1 adjunta a dicho canal. A largo plazo, la tolerancia a la glucosa mejora, pero los niveles plasmáticos de insulina, tanto basales como después de glucosa, no permanecen altos, sino que pueden ir descendiendo. El efecto clave es la reducción de los niveles plasmáticos de glucosa. Este descenso en los niveles de glucemia disminuye la glucotoxicidad a la que son tan sensibles las células β del páncreas. Puede

generar hipoglicemia muy intensa e incluso mortal, y mantenida aunque se le trate con soluciones de glucosa⁵⁴.

La metformina es el único derivado biguanídico aceptado. No provoca liberación de insulina. Entre algunas de sus acciones se encuentran: el aumento del metabolismo de la glucosa en los tejidos, en particular de la glucólisis anaeróbica, reducción de la gluconeogénesis hepática e inhibición de la absorción de glucosa, aminoácidos, y otros compuestos a nivel intestinal. A nivel subcelular las biguanidas se fijan a la membrana mitocondrial, donde podrían alterar los sistemas de transporte. Se ha comprobado en adipocitos y en células musculares que la metformina aumenta la translocación de transportadores GLUT4 desde la membrana microsómica a la membrana plasmática provocada por la insulina. La metformina puede sustituir a las sulfonilureas cuando no son toleradas o han fracasado.

Los tiazolidinedionas son fármacos hipoglucemiantes que se caracterizan por sensibilizar o incrementar la acción de la insulina sin que aumente su secreción, por lo que son útiles en situaciones en que se desarrolla resistencia a la insulina. Este fármaco se fija de manera directa y actúa sobre uno de los subtipos del receptor nuclear del proliferado activado de peroxisomas (PPAR γ). Este receptor pertenece a la superfamilia de receptores nucleares cuya activación desencadena la transcripción del ADN. La principal consecuencia de activar PPAR γ es el incremento de la transcripción de genes de enzimas que normalmente son inducidos por la insulina e interviene en el metabolismo hidrocarbonado y lipídico. Por consiguiente, estos fármacos exigen la presencia de insulina ya que, en definitiva, van a facilitar o incrementar su acción al disponer de un sustrato enzimático más abundante. Puede

producir molestias gastrointestinales en diversos caracteres⁵⁴. Existe una necesidad de generar nuevos y mejores agentes antihiperlipidémicos, especialmente a aquellos que se dirigen a disfunciones metabólicas subyacentes que avanzan progresivamente la enfermedad⁵⁶.

El análisis de distintos modelos diabéticos ha proporcionado evidencia del papel de GSK-3 para esta enfermedad. En un modelo murino de obesidad inducida por la dieta en ratones C57Black6 que se vuelven insensibles a la insulina, se informó que la actividad de GSK-3 en adipocitos era el doble que la de animales alimentados con una dieta de control⁵⁷. También se han observado niveles más altos de actividad de GSK-3 en ratones ob/ob en comparación con animales delgados⁵⁸. También se ha informado que las mediciones de actividad de GSK-3 en el músculo esquelético son elevadas en comparación con actividades de pacientes no diabéticos^{59,60}.

Por supuesto, tales datos son relativos y los cambios de los niveles y actividad de GSK-3 pueden simplemente reflejar un efecto consecuente debido a la adaptación fisiológica de los tejidos a los niveles elevados de glucosa e insulina en la sangre que tipifican a la diabetes tipo 2. La pregunta crítica es si la modulación específica de la actividad de GSK-3 puede revertir los efectos de insensibilidad a la insulina. En este sentido, las propiedades inusuales de GSK-3 ofrecen una ventaja significativa. Es mucho más fácil desarrollar una molécula pequeña que inhiba una enzima, a desarrollar una molécula que la active. Si la incapacidad de la insulina para inhibir a la GSK-3 es un elemento importante para la diabetes tipo 2, entonces un inhibidor sintético de GSK-3 debería ser prometedor, ya que se estaría restaurando uno de los principales efectos de la acción de la insulina¹¹.

1.1.3. Moléculas inhibidoras de GSK3

50 años después del descubrimiento de las propiedades únicas del litio en la enfermedad de depresión maníaca o bipolaridad se identificó a la quinasa GSK-3 como una de sus dianas principales^{34,61}. Desde entonces, el Litio se ha utilizado ampliamente como un inhibidor farmacológico de GSK-3, a pesar de las concentraciones milimolares que se requieren para afectar a la GSK-3 en células vivas. En la actualidad se han descubierto más de 30 inhibidores, algunos con valores de IC₅₀ en el rango nanomolar (**Tabla S1** del anexo y **Figura 7**)⁶²⁻⁶⁵. La selectividad de la mayoría de los inhibidores disponibles de GSK-3 es poco conocida y, esencialmente, se basa en su evaluación con paneles limitados de quinasas^{66,67} y la purificación de posibles dianas en inhibidores inmovilizados mediante cromatografía de afinidad^{68,69}.

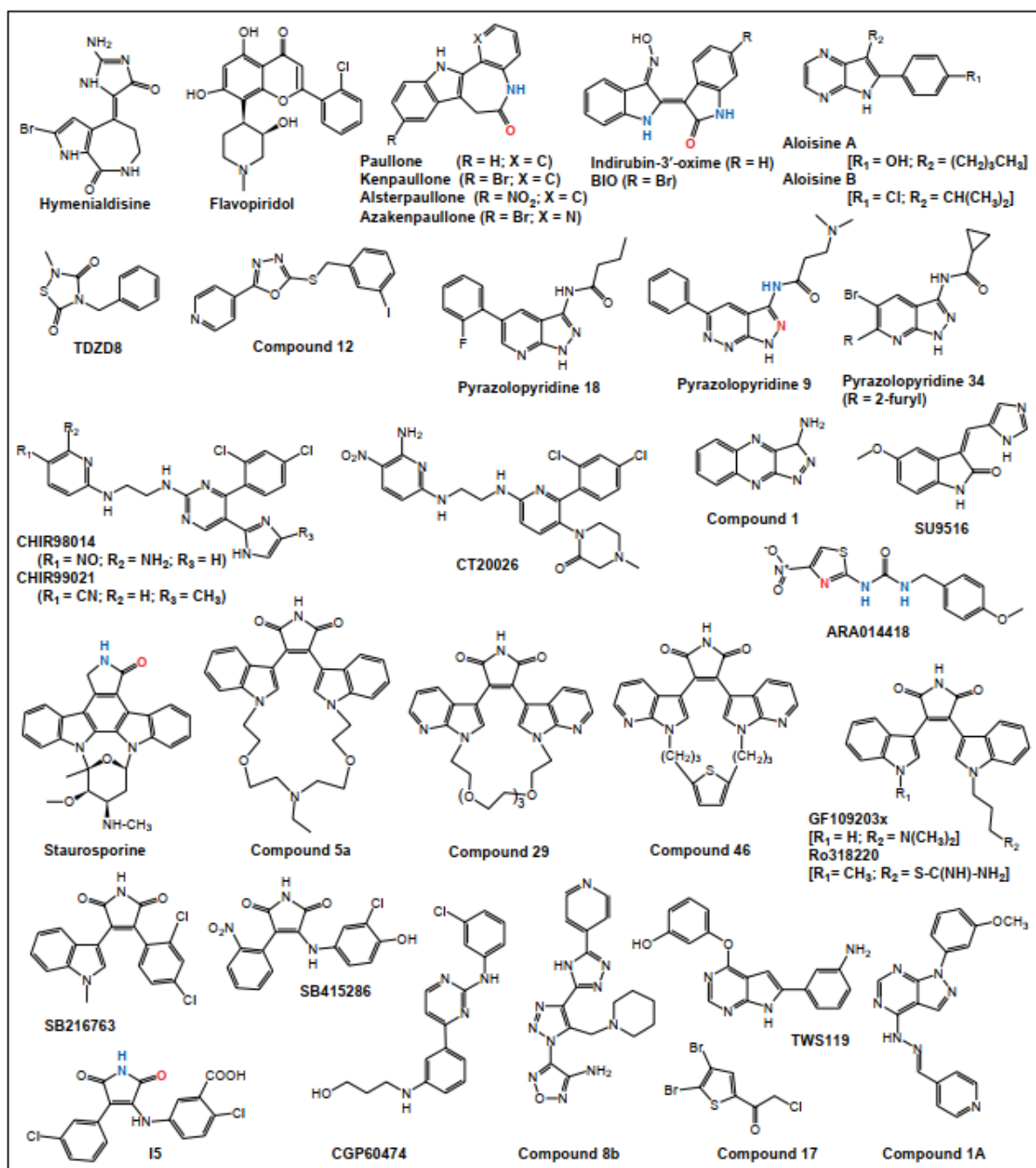


Figura 7: Estructura de algunos de los inhibidores farmacológicos más estudiados de la glucógeno sintasa quinasa 3 (GSK-3). Los átomos que interactúan con la cadena lateral de Val135 y Asp133 de la GSK-3 se muestran en rojo (aceptores de hidrógeno) y azul (donantes de hidrógeno)⁷⁰.

Los inhibidores farmacológicos de GSK-3^{71,72} se ha demostrado que imitan la actividad de la glucógeno sintasa inducida por la insulina, la síntesis del glucógeno, la supresión de la gluconeogénesis y el aumento de la captación de la glucosa en diversos cultivos celulares. Los efectos similares a la insulina también se observan *in vivo*^{71,73,74}. La inhibición farmacológica de GSK-3 potencia la acción de la insulina en el músculo esquelético de ratas resistentes a la insulina⁷³. En conjunto, estos resultados indican que la inhibición de GSK-3 podría mejorar la utilización de glucosa y la sensibilidad a la insulina terapéutica⁷⁰. Con el fin de responder definitivamente preguntas sobre el rol de GSK-3 en diabetes y el potencial de efectos secundarios relacionados con otros roles reguladores de GSK-3, Allan et al. sintetizó un set de moléculas del tipo aminopirimidínico utilizando el compuesto preclínico avanzado 1 (también conocido como CHIR-). En este trabajo nos enfocamos a el análisis de los derivados de CHIR-99021 y del compuesto CHIR-98014 (Figura 8)

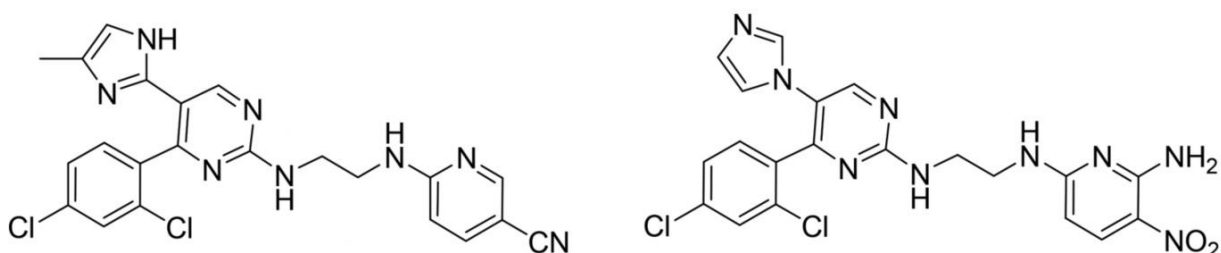


Figura 8: Compuestos preclínicos utilizados como patrón para la generación de nuevos derivados. En el lado izquierdo se encuentra CHIR-99021 y en el lado derecho se encuentra CHIR-98014.

1.2. QSAR

1.2.1. Introducción

Durante más de 2000 años, la tradición médica hipocrática pesó sobre el desarrollo de una medicina moderna y un enfoque renovado del tratamiento de enfermedades. La base para el uso de drogas se mantuvo basada en teorías empíricas vinculadas al equilibrio de los “humores” del cuerpo que consiste en sanguíneo, melancólico, flemático y colérico. La salud y la enfermedad se consideraban una cuestión de equilibrio o desequilibrio con los alimentos y las hierbas clasificadas según su capacidad para afectar la homeostasis natural. Antes del siglo XIX, la farmacia seguía siendo una ciencia empírica, guiada por la medicina tradicional, y heredada de los “ancestros”. Numerosos medicamentos, la mayoría preparados con extractos de plantas, a veces eficaces, estaban disponibles. Pero ninguno de ellos podía responder a la definición química de lo que llamamos hoy a una droga, excepto las drogas de origen mineral. La tecnología de fabricación de drogas era, en el mejor de los casos, tosca: tinturas, cataplasmas, sopas e infusiones con extractos de hierbas recién molidas o secas o productos de origen animal como huesos, grasas o incluso a base de alcohol. Al formular el principio de conservación de la masa, dio una clara diferenciación entre elementos y compuestos, algo muy importante para la química farmacéutica. Muchos científicos integraron avances prácticos de una nueva generación de experimentadores, las que generaron el impacto del desarrollo tanto de la química industrial como la química medicinal. El papel de la química en la fabricación de nuevos medicamentos es esencial⁷⁵.

Es por esto por lo que la química medicinal es un área aplicada de la ciencia que con un enfoque multidisciplinario aborda el problema del descubrimiento de nuevos compuestos

biológicamente activos con aplicaciones terapéuticas. Los compuestos se diseñan para una diana biológica específica, esperando generar o suprimir una respuesta fisiológica clave en el tratamiento de alguna patología. Una de las estrategias claves en el moderno diseño racional de nuevos compuestos son los estudios QSAR que se tratan en la sección siguiente.

1.2.2. QSAR tradicional (2D-QSAR)

Los estudios de relación estructura actividad cuantitativa (QSAR, por sus siglas en inglés “*quantitative structure-activity relationships*”) buscan identificar y cuantificar las propiedades fisicoquímicas de una molécula y evaluar cuál de estas propiedades tiene un efecto en la actividad biológica. De este modo, se podría generar una ecuación que correlacione las propiedades fisicoquímicas de las moléculas con la actividad biológica de cada compuesto elucidando así que propiedades (y en qué grado) juegan un rol importante en la farmacodinámica y/o farmacocinética de los compuestos. Una vez establecida esta ecuación se podrá realizar una predicción de la actividad biológica de un nuevo derivado, análogo a los utilizados para desarrollar el modelo. Esto genera una gran ventaja a la hora de investigar y desarrollar nuevos fármacos, ya que no sería necesario sintetizar una gran cantidad de moléculas para ser evaluadas en ensayos biológicos, sino que permite priorizar a síntesis de compuestos análogos que presenten una alta actividad predicha. De este modo se ahorraría tiempo y recursos valiosos⁷⁶.

1.2.3. Descriptores fisicoquímicos

Los descriptores fisicoquímicos más utilizados en la metodología QSAR son la lipofilia, las propiedades electrónicas, y los efectos estéricos. Las propiedades lipofílicas pueden ser fácilmente cuantificadas para la moléculas completa como para un sustituyente

en particular. Mientras que las propiedades electrónicas y estéricas son más complejas de cuantificar para una molécula entera, por lo que en estos casos se trabaja habitualmente con sustituyentes individuales.

1.2.3.1. Lipofilia

La lipofilia de una molécula es importante cuando la molécula es una droga que debe traspasar una membrana celular y también cuando interacciona con un receptor⁷⁷.

El carácter lipofílico de una droga puede ser medido de forma experimental mediante la distribución relativa en una mezcla de *n*-octanol/agua. Las moléculas hidrófobas se van a disolver en la fracción *n*-octanol, mientras que las moléculas hidrofílicas se disolverán en la fracción acuosa. Esta distribución relativa se conoce como coeficiente de partición (*P*) y se obtiene de la siguiente ecuación:

$$P = \frac{\text{Concentración de la droga en octanol}}{\text{Concentración de la droga en la fase acuosa}} \quad \text{ec(1)}$$

De la ecuación 1 se observa que los compuestos hidrofóbicos tendrán un valor alto de *P*. Variando los sustituyentes de una molécula se producirán análogos con distintos valores de *P*. Se puede graficar la actividad biológica de estos análogos en función de *P* y así comprobar si existe algún tipo de relación en el carácter lipofílico de la droga y su actividad biológica⁷⁸. Las actividades biológicas se expresan comúnmente como $1/C$, donde *C* es la concentración requerida para alcanzar cierto nivel de actividad biológica deseada. Por lo que una droga con un nivel alto de actividad biológica requerirá de menor concentración.

Esta relación se puede graficar, utilizando $\log(1/C)$ versus $\log P$. Experimentalmente, se ha comprobado que la actividad biológica aumenta al incrementar P hasta llegar a un valor máximo, luego si se sigue incrementando P la actividad biológica comenzará a disminuir obteniéndose el gráfico de parábola (**Figura 9**). En aquellos casos en los que el coeficiente de partición es el único factor que influencia la actividad biológica, la curva parabólica se puede expresar con la siguiente ecuación:

$$\log\left(\frac{1}{C}\right) = -k_1(\log P)^2 + k_2 \log P + k_3 \quad \text{ec(2)}$$

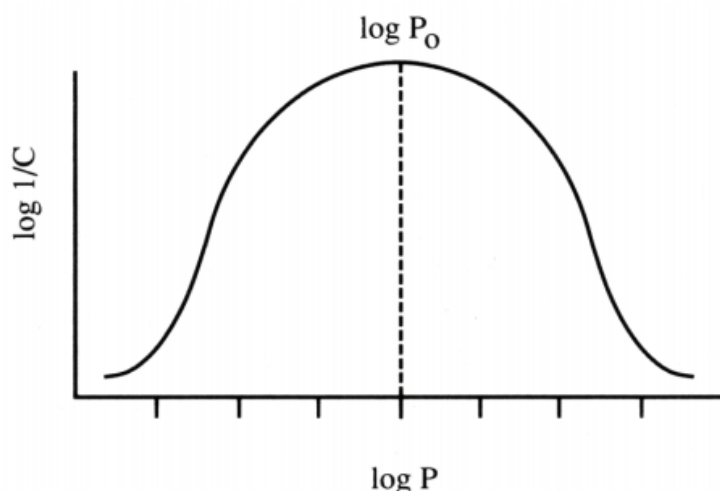


Figura 9: Curva parabólica de $\log(1/C)$ versus $\log P$.

La parábola se produce debido a que, al incrementar la lipofilia de una molécula, esta cruza de forma sencilla por la membrana celular, hasta llegar al máximo de lipofilia generando la disminución de la actividad biológica. Este fenómeno ocurre debido a que la molécula tiende a quedar retenida en la membrana celular en vez de atravesarla fácilmente, lo que se debe a que la membrana celular está conformada por una bicapa fosfolipídica⁷⁹.

Los coeficientes de partición pueden ser calculados conociendo la contribución de varios sustituyentes que aportan a la lipofilia. Dicha contribución se conoce como constante de hidrofobia del sustituyente (π) y es una medida de cuán hidrófobo es un sustituyente en comparación con el hidrógeno. El valor se puede obtener midiendo el coeficiente de partición experimentalmente para un compuesto estándar como el benceno con y sin sustituyente (X). La constante de hidrofobia para el sustituyente (X) se obtiene de la ecuación 3:

$$\pi_X = \log P_X - \log P_H \quad \text{ec(3)}$$

En la ecuación 3, P_H es el coeficiente de partición para el compuesto de referencia y P_X es el coeficiente de partición para el compuesto con el sustituyente X. Un valor positivo de π indica que el sustituyente es más hidrófobo que el hidrógeno. Por lo que un valor negativo de π indica que el sustituyente es menos hidrófobo.

El coeficiente de partición P es una medida de la hidrofobia global de una droga, por lo que es una medida importante de que tan eficiente es la transportación de una droga hacia la diana biológica. Por su parte, la constante π mide la hidrofobia de una región específica del esqueleto de la droga y al estar presente en la ecuación QSAR, puede enfatizar cualquier interacción hidrófoba importante que incluya esa región de la molécula con el sitio de unión. En la **Tabla S2** del anexo se muestran valores de π para sustituyentes comunes en compuestos bioactivos.

1.2.3.2. Efectos electrónicos

Las propiedades electrónicas de un sustituyente tienen efecto en la polaridad de una molécula o en su ionización. Esto a su vez afecta en el sentido de que tan fácilmente puede

atravesar una membrana celular o que tan fuertemente puede una droga unirse a su diana biológica (interacciones dipolares, iónicas o de puentes de hidrógeno).

Para los sustituyentes de un anillo aromático se utiliza una medida denominada *constante de sustituyente de Hammett* (σ). Esta es una medida de la habilidad de un sustituyente para atraer electrones hacia sí mismo y donarlos, y ha sido determinada midiendo la disociación de una serie de ácidos benzoicos sustituidos comparados con el ácido benzoico sin sustituir⁸⁰.

Como el ácido benzoico es un ácido débil, sólo se disocia parcialmente en agua (**Figura 10**). Se alcanza un equilibrio entre las formas ionizadas y no ionizadas del ácido, donde las propiedades relativas de dichas especies se conocen como la constante de disociación K_H (el subíndice H significa que el anillo aromático no presenta sustituyentes).

$$K_H = \frac{[\text{PhCO}_2^-][\text{H}^+]}{[\text{PhCO}_2\text{H}]} \quad \text{ec(4)}$$

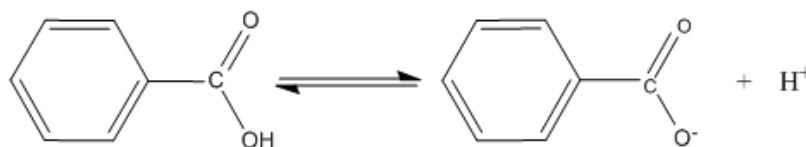


Figura 10: Ionización del ácido benzoico.

Como existe un sustituyente presente en el anillo aromático, el equilibrio se ve afectado. Los grupos que atraen electrones, como el grupo nitrilo, dan como resultado que el anillo aromático atraiga con más fuerza los electrones y de esa manera se estabiliza el ion carboxilato. Así, el equilibrio se verá desplazado hacia la forma ionizada y el ácido benzoico

sustituido será un ácido más fuerte que aquel sin sustituyentes, teniendo un valor de K_X (X representa el sustituyente en el anillo aromático) (**Figura 11**).

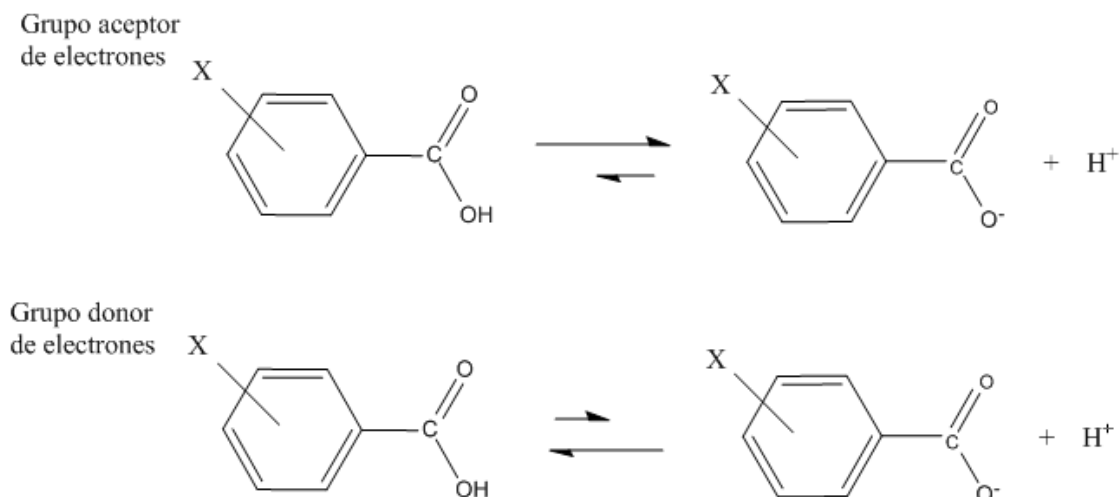


Figura 11: Posición de equilibrio dependiendo del sustituyente X.

La constante de sustituyente de Hammett (σ_X) para un sustituyente en particular (X) está definido por la ecuación 5:

$$\sigma_X = \log K_X - \log K_H \quad \text{ec(5)}$$

Los ácidos benzoicos con sustituyentes atrayentes de electrones tendrán valores mayores de K_X que el ácido benzoico no sustituido, por ende, el valor de σ_X será positivo. Sustituyentes tales como Cl, CN, o CF_3 presentan valores positivos de σ_X (**Tabla S3** anexo).

Por el contrario, los ácidos benzoicos que tengan sustituyentes dadores de carga tendrán valores más pequeños de K_X que el ácido benzoico sin sustituir. Por lo que el valor de σ_X será negativo para un sustituyente donante electrónico.

Cabe destacar que la constante de Hammett toma en consideración los efectos inductivos y los de resonancia⁸¹. Por lo que el valor de σ para un sustituyente en particular dependerá de su posición *meta* o *para*⁸².

La constante de Hammett no puede ser medida para sustituyentes en posición *orto*-debido a que los sutituyentes en esta posición tienen considerable efecto estérico, además de electrónico.

1.2.3.3. Factores estéricos

La facilidad con la que una droga se aproxima y une a su respectivo receptor se verá influida por el tamaño y forma de la molécula. Un sustituyente voluminoso puede dificultar la interacción de una droga con un sitio de unión óptima y así mejorar la actividad biológica. Las propiedades estéricas son más difíciles de cuantificar que las propiedades hidrófobas o electrónicas^{83,84}.

1.2.3.4. Factor estérico de Taft (E_s)

Se han realizado intentos por cuantificar los efectos estéricos del sustituyente mediante el factor estérico de Taft (E_s), cuyo valor puede obtenerse comparando el grado de hidrólisis de un éster alifático sustituido con el grado de hidrólisis de un éster tradicional bajo condiciones ácidas⁸⁵.

$$E_s = \log k_X - \log k_0 \quad \text{ec(6)}$$

En la ecuación anterior k_X es la constante de hidrólisis de un éster alifático con el sustituyente X, mientras que k_0 es la constante de hidrólisis del éster de referencia. Si el

sustituyente X es más voluminoso que el átomo de hidrógeno en el éster de referencia, el grado de hidrólisis será menor, por lo que E_s tendrá valor negativo. Por lo que un valor negativo de E_s implica que el sustituyente aumenta los efectos estéricos sobre la molécula. Por el contrario, un valor positivo de E_s implica que el sustituyente produce menos efectos estéricos sobre la molécula respecto al éster de referencia.

Los valores de E_s , sin embargo, al ser una medida de un efecto estérico intramolecular, no dan cuenta de la mejor forma sobre la influencia estérica de las sustituciones o sobre el volumen de una molécula, pues las drogas interactúan con el sitio de unión de forma intermolecular⁸⁶. En la tabla S4 del anexo se muestran valores E_s para sustituyentes comunes en compuestos bioactivos.

1.2.3.5. Refractividad molar (RM)

Otra medida del factor estérico es proporcionada por un parámetro conocido como refractividad molar (MR). Esta es una medida del volumen ocupado por un átomo o grupo de átomos y su resistencia a la polarización⁸⁷. Se obtiene según la siguiente ecuación:

$$MR = \frac{(n^2-1)}{(n^2+2)} \times \frac{MW}{d} \quad \text{ec(7)}$$

En la ecuación 7, n es el índice de refracción, MW es la masa molar, y d es la densidad. El término MW/d define un volumen, y el término $(n^2-1)/(n^2+2)$ provee un factor de corrección definiendo cuán fácilmente el sustituyente o molécula puede ser polarizado. Esto es significativo si el sustituyente tiene electrones π o pares de electrones sin enlazar.

1.2.3.6. Ecuación de Hansch

Si la actividad biológica de un compuesto está relacionada únicamente con una sola propiedad fisicoquímica, se puede establecer una ecuación simple que relacione la actividad con dicha propiedad fisicoquímica. Sin embargo, la actividad biológica de la mayoría de las drogas depende de una combinación de propiedades fisicoquímicas, por lo que se necesita una ecuación que relacione la actividad biológica con un número de parámetros diferentes. Este tipo de ecuaciones son llamadas ecuaciones de Hansch y relacionan la actividad biológica con las propiedades fisicoquímicas vistas anteriormente ($\log P$ o π , σ , y factores estéricos)⁸⁸. Cuando el rango de valores para la lipofilia es pequeño la ecuación será lineal (ecuación 8).

$$\log \frac{1}{C} = k_1 \log P + k_2 \sigma + k_3 E_S + k_4 \quad \text{ec(8)}$$

Si se utiliza un rango amplio de $\log P$, la ecuación será parabólica (debido a como la lipofilia afecta la actividad biológica) (ecuación 9).

$$\log \frac{1}{C} = -k_1 (\log P)^2 + k_2 \log P + k_3 \sigma + k_4 E_S + k_5 \quad \text{ec(9)}$$

Las constantes k_n son determinadas mediante regresión lineal múltiple de modo que se obtenga la ecuación con el mejor ajuste.

1.2.3.7 Análisis de Free-Wilson

En el análisis de Free-Wilson, se mide la actividad biológica de un compuesto y luego se compara con la actividad de una serie de análogos sustituidos. Entonces se deriva una ecuación que relaciona la actividad biológica con la presencia o ausencia de un sustituyente en particular⁸⁹.

$$\text{Actividad} = k_1X_1 + k_2X_2 + k_3X_3 + \dots + k_nX_n + Z \quad \text{ec(10)}$$

En la ecuación 10, X_n está definida como una variable indicadora, tomando valores de 1 o 0, dependiendo respectivamente si el sustituyente (n) está presente o no. La contribución a la actividad que realiza cada sustituyente está determinada por el valor de k_n . La actividad promedio de la estructura en estudio está representada por Z ⁹⁰.

Debido a que este análisis sólo toma en consideración el efecto global de un sustituyente en la actividad biológica en vez de sus propiedades fisicoquímicas, no se requiere de tablas ni constantes fisicoquímicas, por lo que el método sólo requiere de medidas experimentales de la actividad biológica. Esto resulta ser útil cuando se intenta cuantificar el efecto de sustituyentes inusuales que no se encuentran tabulados.

La desventaja de este análisis es la gran cantidad de análogos que se deben sintetizar y testear para que la ecuación sea significativa.

Otra desventaja es la dificultad para racionalizar los resultados y explicar por qué un sustituyente en una posición particular es bueno o malo para la actividad⁹¹. Finalmente, los

efectos de diferentes sustituyentes pueden no ser aditivos. Puede haber interacciones intramoleculares que afecten la actividad.

1.2.4. 3D-QSAR

En la metodología 3D-QSAR las propiedades tridimensionales de la molécula son consideradas como un todo en lugar de tomar en cuenta los sustituyentes individuales. En esta técnica se asume que las características más importantes de una molécula son su tamaño, forma y sus propiedades electrónicas (campos electrostáticos).

Existen diversas aproximaciones a la técnica 3D-QSAR, pero el método que ha captado mayor interés fue desarrollado por la compañía Tripos y se le conoce como CoMFA (*Comparative Molecular Field Analysis*)⁹². También, en este trabajo, se utilizó el método CoMSIA (*Comparative, Molecular Similary Index Analysis*).

1.2.5. Análisis comparativo de campos moleculares, CoMFA

Se basa en el supuesto de que las interacciones entre la droga y el receptor son del tipo no covalentes y que los cambios en la actividad biológica están correlacionados con cambios en los campos estéricos y/o electrostáticos de la droga⁹².

1.2.5.1. Campo estérico y electrostático

Para poder definir los campos estéricos y electrostáticos, una molécula se construye utilizando software de modelación molecular. Si se tienen diversas conformaciones estables para una molécula, entonces la conformación activa debe ser identificada y la energía minimizada. Luego se identifica el farmacóforo (**Figura 12** y **Figura 13**), que corresponde al conjunto de rasgos estéricos y electrónicos incluyendo la disposición espacial de ciertos

átomos que son necesarios para asegurar las óptimas interacciones intermoleculares con un blanco biológico específico y así gatillar o bloquear, según sea el caso, la respuesta biológica⁹³.

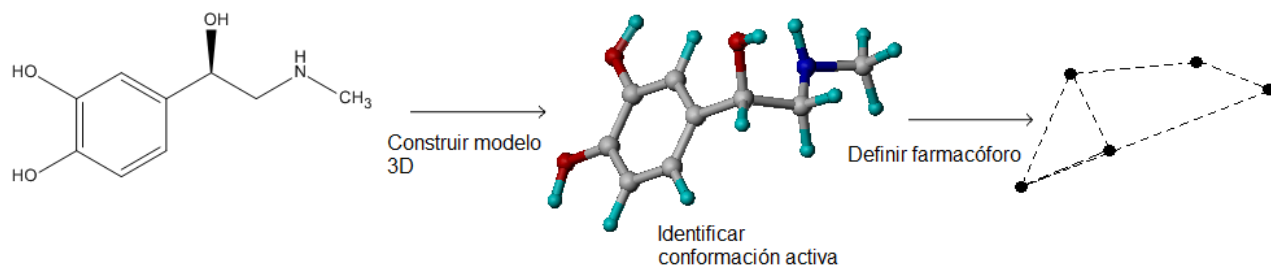


Figura 12: Identificación de la conformación activa y farmacóforo. El farmacóforo en este ejemplo contempla los átomos de oxígeno, el átomo de nitrógeno, y el centro del anillo aromático.

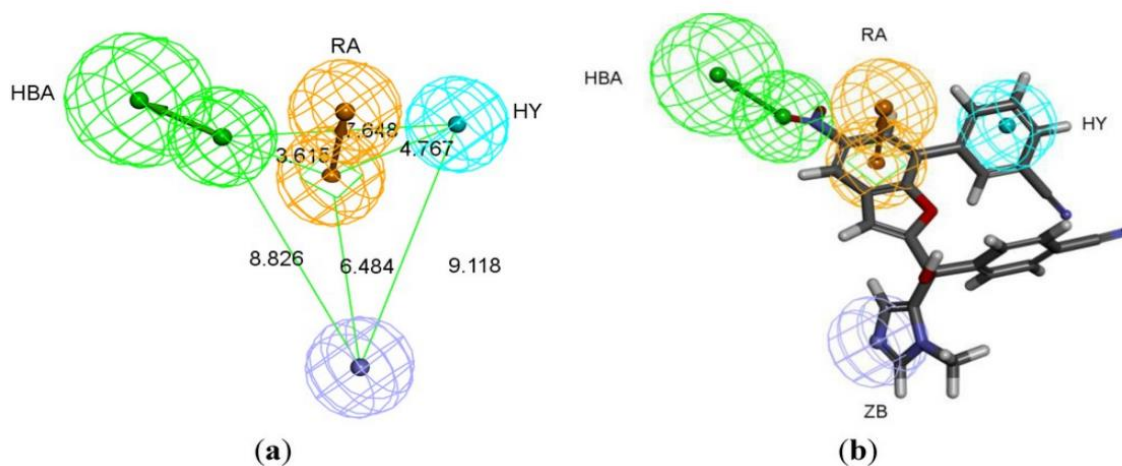


Figura 13: (a) Representación de un farmacóforo con sus propiedades químicas y distancias entre sus diferentes regiones. HBA, aceptor de puente de hidrogeno; RA, anillo aromático; HY, zona hidrofóbica; ZB enlace de cinc. (b) Una molécula solapada con el farmacóforo.

El siguiente paso es colocar el farmacóforo en un enrejado virtual (**Figura 14**). La posición del farmacóforo se mantiene constante y actúa como un punto de referencia cuando se posicionan diferentes moléculas en el enrejado. Por cada molécula que se estudie, su conformación activa y su farmacóforo son identificados y luego la molécula es colocada en el enrejado de tal manera que su farmacóforo calce con el farmacóforo de referencia. Para obtener los mejores resultados al superponer todas las moléculas en estudio sus farmacóforos respectivos deben calzar lo mejor posible con el de referencia (**Figura 15**).

Una vez que se ha colocado una molécula en el enrejado, se mide el campo estérico y electrostático alrededor de ella. Esta medida se realiza posicionando un átomo sonda como un protón o un carbocatión con hibridación sp^3 en cada uno de los puntos del enrejado virtual, en un reticulado cúbico espaciado de 2\AA a 18\AA por lado, cuyas coordenadas X, Y, Z más bajas son -4, -12, -8\AA , con un punto de corte (valor máximo) de 30kcal/mol para puntos reticulares cuya interacción estérica total con todos los átomos de cadena lateral es mayor que el punto de corte⁹⁴.

En lo que concierne al campo estérico, este aumentará a medida que el átomo sonda se acerca a la molécula. Por otra parte, en lo que respecta al campo electrostático, habrá una atracción entre la carga positiva del carbocatión y las regiones ricas en electrones de la molécula, y habrá repulsión entre el carbocatión y las regiones deficientes en electrones de la molécula.

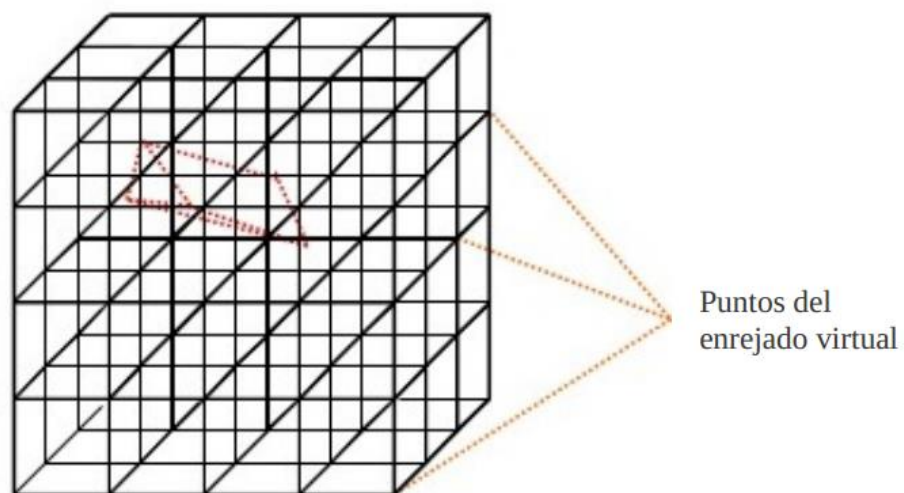


Figura 14: Representación del farmacóforo (poliedro de puntos rojos) colocado al interior del enrejado virtual. Cada vértice representa un punto en el espacio.

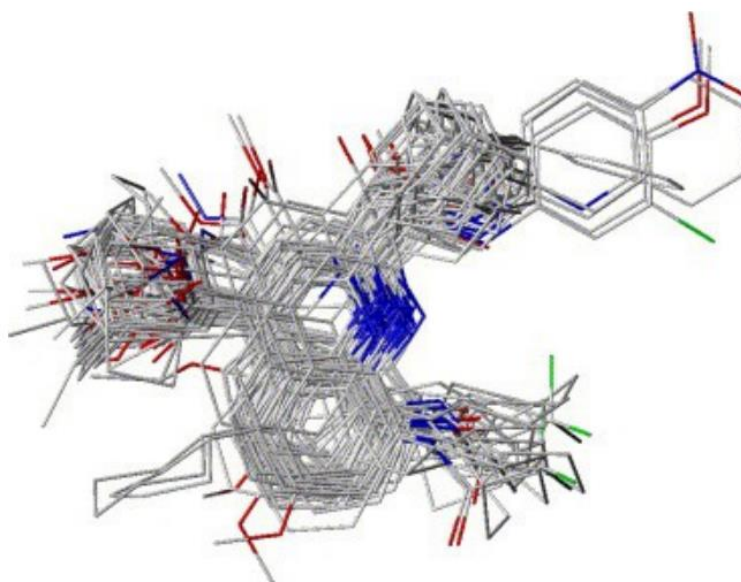


Figura 15: Ejemplo de una serie de análogos en estudio alineados todos respecto a un farmacóforo de referencia.

El valor de los campos estéricos y electrostáticos en cada punto del enrejado son tabulados, y esto se repite para cada molécula en estudio (**Figura 16**). Un valor en particular de la energía estérica se elige para definir la forma de la molécula, y los puntos del enrejado que tengan ese valor de energía se conectan por líneas de contorno para definir el campo estérico. Para medir el campo electrostático se lleva a cabo un proceso similar. Las regiones ricas y pobres son definidas para cada molécula por líneas de contorno apropiadas.

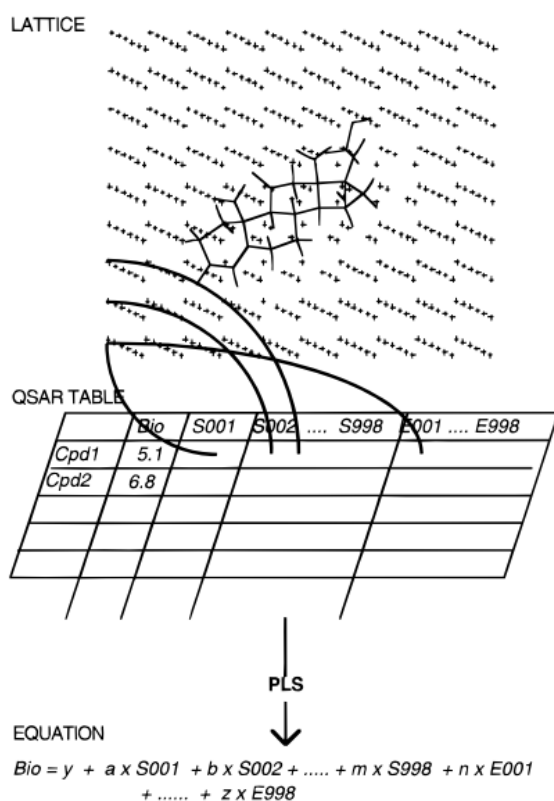


Figura 16: Proceso CoMFA. Las “diferencias de campo estérico” son las sumas de raíz de las diferencias cuadradas entre los valores de celda dentro de un par de filas en la tabla QSAR⁹⁴.

1.2.5.2. Relación de forma y distribución electrónica con la actividad biológica

En 3D-QSAR, las variables para cada molécula son las interacciones estéricas y electrónicas calculadas en miles de puntos del enrejado virtual. Si se tuvieran 100 moléculas en estudio, el número de variables sobrepasa en gran manera el número de estructuras, por lo que no es posible relacionarlas con la actividad biológica por medio del análisis de regresión lineal estándar. Se debe seguir un procedimiento estadístico diferente, que se denomina mínimos cuadrados parciales o PLS (del inglés *partial least squares*)(como aparece en la Figura 16). Es un proceso estadístico computacional, el cual es repetido iterativamente para intentar encontrar la mejor ecuación que relacione la actividad biológica con las diversas variables. Como parte de este proceso el número de variables se reduce a medida que se filtran aquellas que claramente no están relacionadas con la actividad biológica. Una característica importante de este análisis es que una estructura es deliberadamente dejada fuera (método LOO, del inglés *Leave-One-Out*) mientras que con el resto de las moléculas se formula una regresión parcial. Cuando una ecuación ha sido definida, ésta es probada con la estructura que fue previamente excluida. A esto se le llama validación cruzada, y permite evaluar cuán bien la fórmula predice la actividad biológica para la molécula que fue excluida⁹⁵. Los resultados son empleados en una nueva ronda de cálculos, pero esta vez se incluye en los cálculos la estructura que fue retirada y una estructura distinta se excluye. Esto conduce a una nueva fórmula mejorada la cual es nuevamente probada con el compuesto dejado fuera. Este proceso continúa hasta que la validación cruzada se ha llevado a cabo contra todas las estructuras.

Al término del proceso se obtiene la ecuación óptima. El poder de predicción de la ecuación final está cuantificado por el coeficiente de correlación de validación cruzada r^2 , que usualmente se escribe como q^2 . En contraste con el QSAR normal (2D-QSAR) donde r^2 debe ser mayor que 0.8, los valores de q^2 mayores que 0.5 son considerados significativos. Es más útil mostrar una representación gráfica que muestre cuáles regiones alrededor de la molécula son importantes para la actividad biológica por motivos estéricos o electrónicos. De esta manera, un mapa estérico muestra una serie de contornos coloreados indicando las interacciones estéricas benéficas y perjudiciales alrededor de una molécula representativa del conjunto de moléculas probadas. Un mapa similar de contornos se crea para ilustrar las interacciones electrostáticas benéficas y perjudiciales (**Figura 17**).

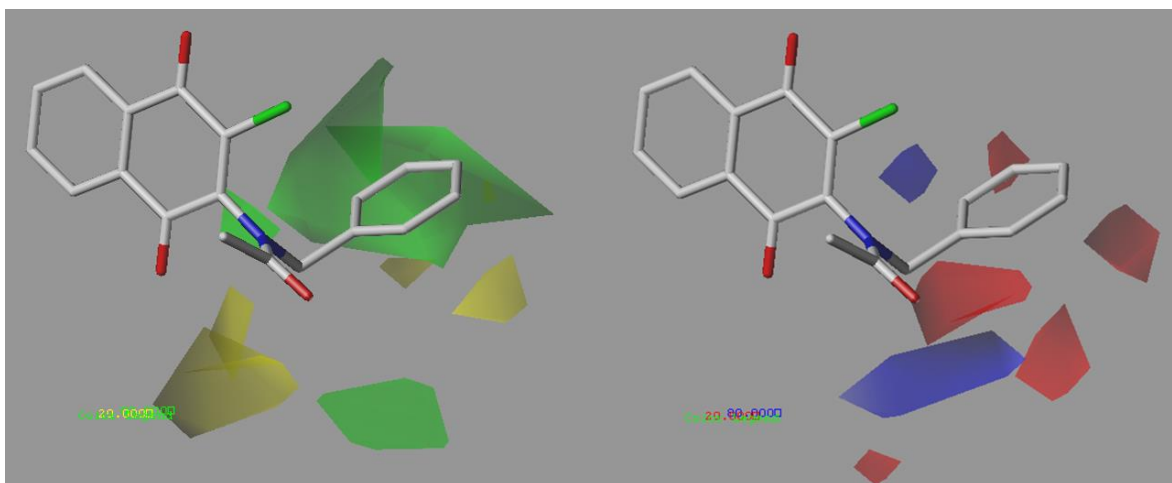


Figura 17: Se muestran dos mapas de contornos para una misma molécula donde se señala mediante color verde los sitios donde un sustituyente voluminoso mejoraría la actividad biológica, y en amarillo donde un sustituyente voluminoso perjudica la actividad biológica. El color rojo indica que un sustituyente con potencial electrostático negativo incrementa la actividad biológica, mientras que en las zonas azules un sustituyente con potencial electrostático positivo beneficia la actividad.

1.2.5.3. Potencial hidrofóbico

En lo referente a QSAR tradicional (2D-QSAR), el factor hidrofóbico es muy importante y la mayoría de las ecuaciones incluyen un valor de $\log P$. Sorpresivamente, el cálculo del factor hidrofóbico en un estudio de 3D-QSAR no es tan crucial y la mayoría de los estudios exitosos se han basado únicamente en factores estéricos y electrostáticos. Sin embargo, estudios de 3D-QSAR pueden llevarse a cabo utilizando el factor hidrofóbico si el átomo sonda es reemplazado por una molécula de agua para así calcular el potencial hidrofóbico en cada punto del enrejado.

1.2.5.4. Ventajas del 3D-QSAR sobre el QSAR tradicional

La técnica de QSAR tradicional (2D-QSAR) presenta los siguientes problemas:

- Sólo se pueden estudiar moléculas con estructuras similares.
- La validez de los descriptores numéricos está en tela de juicio debido a que estos descriptores se obtienen midiendo velocidades de reacción y constantes de equilibrio en reacciones modelo y luego son tabuladas. Sin embargo, separar una propiedad de otra no es siempre posible en mediciones experimentales. Por ejemplo, la medición del factor estérico de Taft no es puramente una medida del factor estérico porque las velocidades de reacción empleadas para definirlo son también afectadas por factores electrónicos.
- Los descriptores tabulados podrían no incluir datos para sustituyentes inusuales.
- Se requiere sintetizar un rango de moléculas donde se varían los sustituyentes para probar una propiedad en particular.

- Las ecuaciones de 2D-QSAR no sugieren directamente que nuevos compuestos sintetizar.

Los problemas anteriores se evitan realizando estudios 3D-QSAR como CoMFA, que además posee las siguientes ventajas:

- Las interacciones favorables y desfavorables son representadas gráficamente por contornos tridimensionales alrededor de una molécula representativa. Una imagen es mucho más fácil de interpretar que una fórmula matemática.
- En 3D-QSAR las propiedades de las moléculas son calculadas individualmente por un programa. No existe dependencia de factores experimentales o tabulados. No hay necesidad de restringir el estudio sólo a moléculas de estructura similar. Mientras se esté seguro de que todos los compuestos en estudio compartan el mismo farmacóforo e interaccionan del mismo modo con la diana biológica, pueden ser analizados en un estudio de 3D-QSAR.
- La representación gráfica de las interacciones benéficas y perjudiciales permiten diseñar nuevas estructuras. Como ejemplo, si un mapa de contorno muestra un efecto estérico favorable en una ubicación particular, esto implica que el sitio de unión de la droga tiene espacio para extender más la molécula en esa ubicación. Esto puede conducir a interacciones más favorables entre la droga y su receptor.
- Tanto el QSAR tradicional como el 3D pueden ser empleados sin la necesidad de conocer la estructura de la diana biológica.

1.2.5.5. Problemas potenciales de 3D-QSAR

Hay que tener en cuenta ciertas consideraciones al momento de realizar un 3D-QSAR para evitar equivocaciones. Algunas de estas son:

- Se debe procurar que cada molécula este en su conformación activa cuando se procede al alineamiento.
- Cada molécula debe estar debidamente alineada con respecto a las otras, de manera tal que sus farmacóforos coincidan.

Es posible conocer la conformación activa en estructuras rígidas como los esteroides, pero es más difícil con moléculas flexibles que tienen la capacidad de rotar sus enlaces. Por lo que resulta de utilidad tener un análogo cuya conformación esté restringida y sea biológicamente activo para que actúe como guía para la probable conformación activa. Entonces se pueden construir en el computador moléculas más flexibles con la conformación que calce lo más posible con la del análogo más rígido. Si la estructura del sitio de unión se conoce, esto puede ser útil para decidir la conformación activa probable para las moléculas y como deben ser alineadas previo al estudio 3D-QSAR.

Se debe asegurar que todos los compuestos en estudio interactúen con la diana farmacológica de manera similar. Por ejemplo, un estudio de todos los posibles inhibidores de la acetilcolinesterasa está condenado al fracaso. La gran diversidad de estructuras involucradas hace imposible alinearlas para generar un farmacóforo tridimensional. Además, los diferentes inhibidores no interactúan con la enzima del mismo modo. Estudios cristalográficos de rayos X de complejos enzima-inhibidor muestran que los inhibidores

tacrina, edrofonio, y decametonio tienen todas diferentes orientaciones al unirse al sitio activo.

1.2.5.6. Análisis Comparativo de Índices de Similitud Molecular, CoMSIA

El otro método empleado en la técnica de 3D-QSAR es el CoMSIA. En el método de CoMFA tradicional se emplean únicamente potenciales de Lennard-Jones y Coulomb. Hay evidencia de que estos potenciales sólo describen las contribuciones energéticas a la constante de unión ligando-receptor⁹⁶. Las influencias entrópicas no son consideradas en el CoMFA. Por tanto, de manera que sea posible incluir las contribuciones entrópicas, se requiere algún tipo de campo que considere las contribuciones hidrofóbicas a la actividad.

Los campos usados habitualmente en CoMFA presentan algunos problemas adicionales. Por ejemplo, el potencial de Lennard-Jones es extremadamente alto en la superficie de las moléculas. Como consecuencia, la energía potencial expresada en los puntos de enrejado en las proximidades a la superficie cambia dramáticamente. De este modo, tan sólo pequeños desplazamientos de las moléculas o cambios conformacionales mínimos pueden resultar en fuertes variaciones de estos descriptores.

Adicionalmente, los potenciales de Lennard-Jones y Coulomb presentan singularidades en las posiciones atómicas. Para evitar valores energéticos inaceptablemente grandes, los cálculos de los potenciales están normalmente restringidos a las regiones exteriores de la molécula, y se hace necesario definir algunos valores de corte energético arbitrarios. Estos valores de corte y otros parámetros arbitrarios pueden llevar a la pérdida de información sobre un determinado campo.

Para superar los problemas mencionados anteriormente surge el método CoMSIA. Usando una sonda común, se calculan los índices de similitudes para cada una de las moléculas alineadas en los puntos del enrejado espaciados regularmente. En principio, cualquier propiedad fisicoquímica puede ser considerada en esta aproximación para calcular un “campo” de índices de similitudes. Sin embargo, normalmente se emplean sólo propiedades estéricas, electrostáticas e hidrofóbicas⁹⁷. Los índices obtenidos son evaluados en un análisis de PLS según el protocolo usual del método CoMFA⁹⁸, obteniendo finalmente gráficos de contornos análogos a los de CoMFA.

En este trabajo, utilizando las herramientas de la química medicinal, la metodología 3D-QSAR CoMFA/CoMSIA, y los componentes aminopirimidínicos propuestos por Allan et al.⁹⁹, generamos una propuesta de nuevos derivados de dichos compuestos para su optimización como moléculas inhibidoras de la glucógeno sintasa quinasa 3, como propuesta a un posible tratamiento para la diabetes tipo 2.

2. HIPÓTESIS

Postulamos que es posible hallar correlaciones válidas entre los potenciales estérico, electrostático, hidrofóbico, dador y aceptor de puente de hidrógeno y la capacidad inhibitoria GSK3 de derivados de estructura general (5-imidazol-2-il-4-fenilpirimidin-2-il) [2-(2piridilamino) etil] amina que permitan explicar la capacidad antidiabética de la base de datos estudiada.

3. OBJETIVOS

3.1. Objetivo General

Elucidar la relación estructura-actividad basándose en la metodología 3D-QSAR para diseñar nuevas estructuras que funcionen como inhibidores de la enzima quinasa glucógeno sintasa quinasa 3 (GSK3) con potencial actividad antidiabética.

3.2. Objetivos Específicos

1. Compilar una base de datos de compuestos heterocíclicos con actividad en GSK3
2. Realizar el alineamiento de las moléculas
3. Llevar a cabo un análisis estadístico 3D-QSAR y validarlo por análisis de Tropsha.
4. Obtener y analizar mapas de contorno para la propuesta de nuevas moléculas promisorias

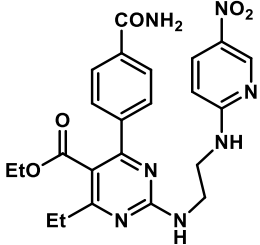
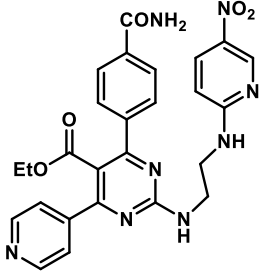
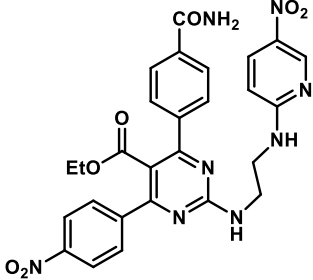
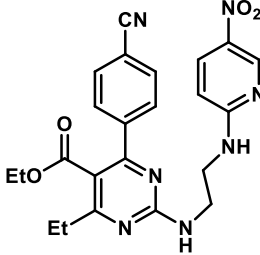
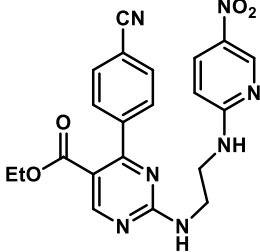
4. MATERIALES Y MÉTODOS

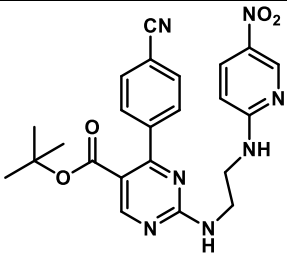
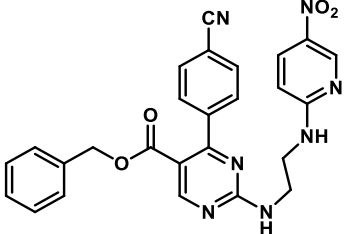
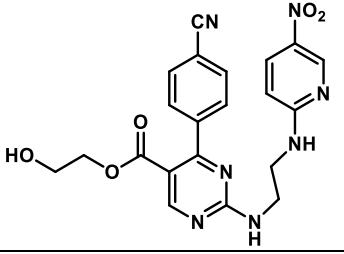
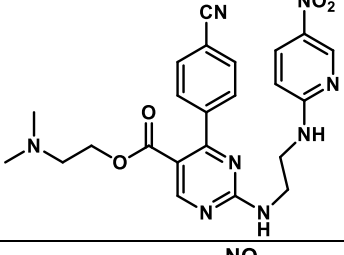
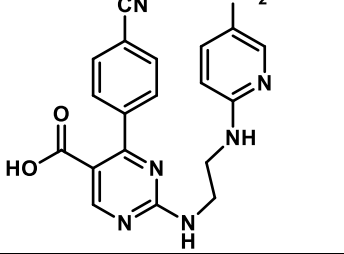
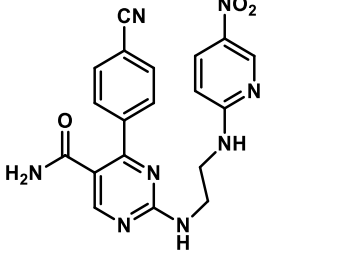
4.1. Compilación de la base de datos

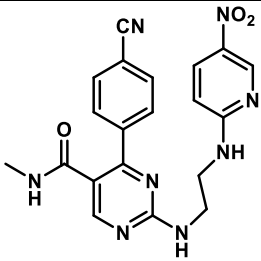
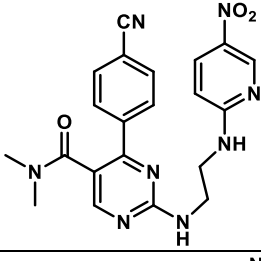
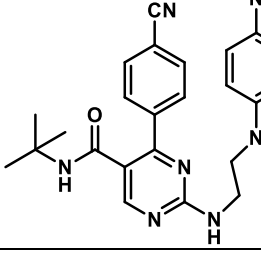
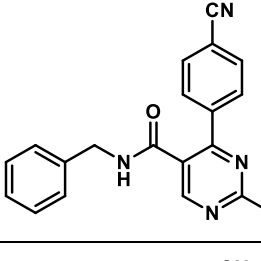
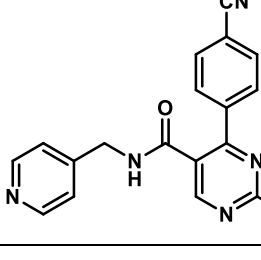
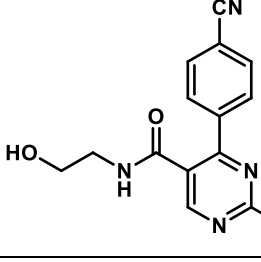
4.1.1. Generación de las moléculas

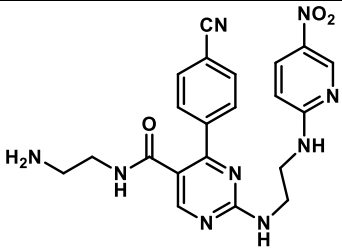
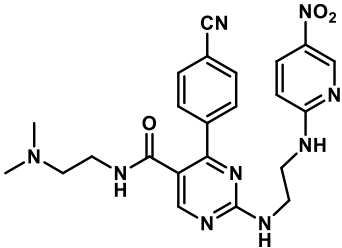
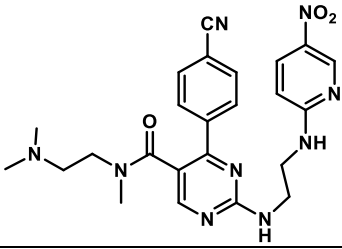
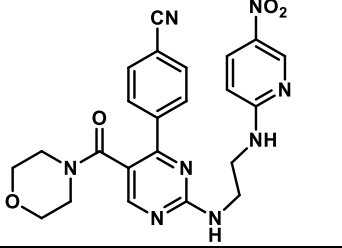
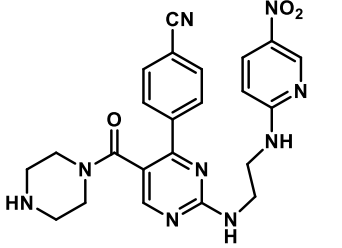
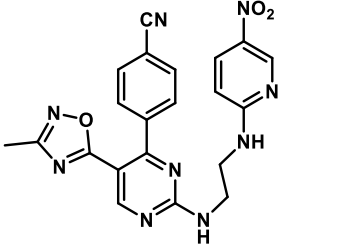
El primer paso fue diseñar en el programa ChemDraw, en su versión 15.1, el set de 132 moléculas en 2D según lo reportado en el trabajo de Allan et al⁹⁹, donde se respetó el orden de síntesis del set de moléculas y el orden de adición de nuevos sustituyentes. A continuación, se exportaron las 132 moléculas diseñadas en ChemDraw a un área tridimensional con el programa Chem3D en su versión 15.1, para luego ser ingresadas en formato mol2 al programa SYBYL-X en su versión 1.2. Antes de ingresar las moléculas al programa SYBYL-X, se calcularon los conformeros de mínima energía con distintos métodos de cálculo para comparar las conformaciones y seleccionar así el método más apropiado. Las moléculas en 2D se encuentran en la **Tabla 1**.

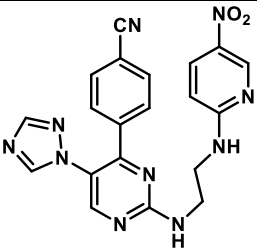
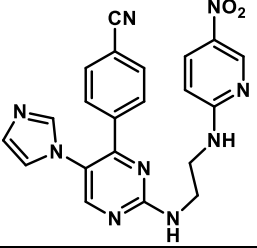
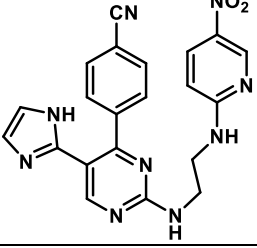
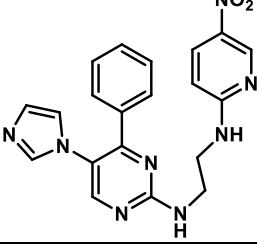
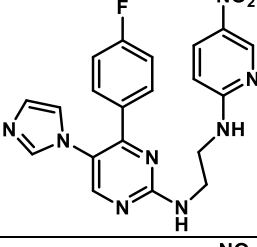
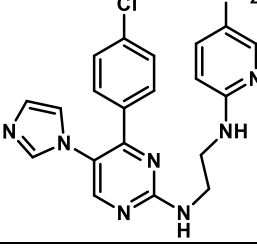
Tabla 1: Estructura y valores de IC₅₀ del conjunto de moléculas analizadas⁹⁹.

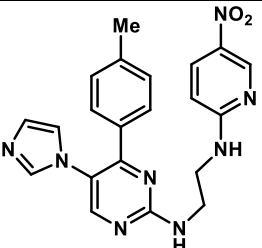
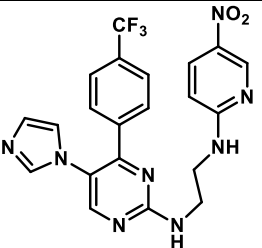
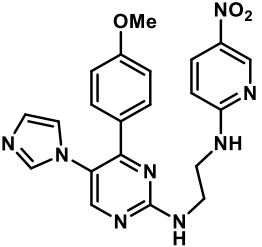
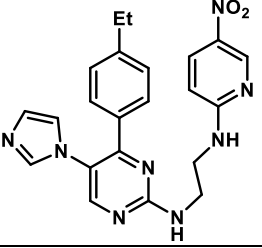
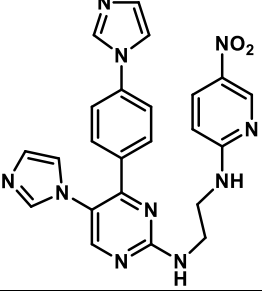
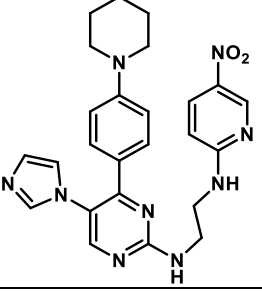
Compuesto Paper	Compuesto QSAR	Estructura	IC ₅₀ (μM)	pIC ₅₀
11a	001		0.12	6.921
11b	002		0.14	6.854
11c	003		0.15	6.824
11d	004		0.04	7.398
16	005		0.052	7.284

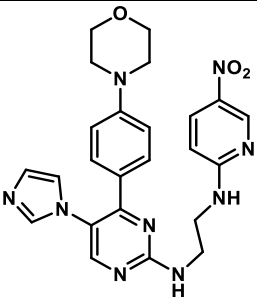
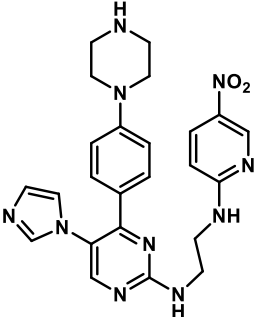
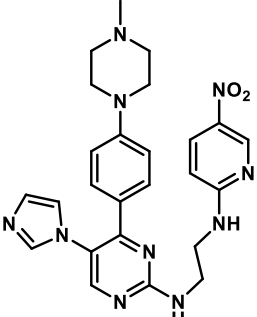
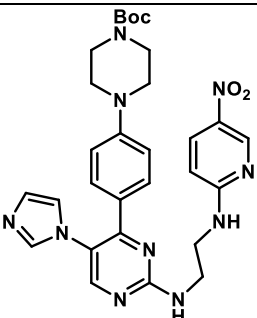
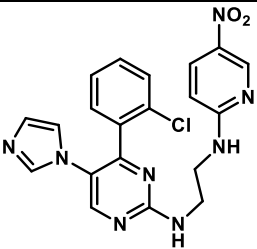
18a	006		0.025	7.602
18b	007		0.190	6.721
18c	008		0.051	7.292
18d	009		0.050	7.301
17	010		0.035	7.456
19a	011		0.029	7.538

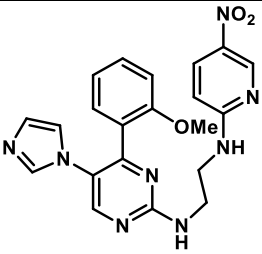
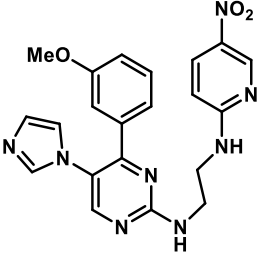
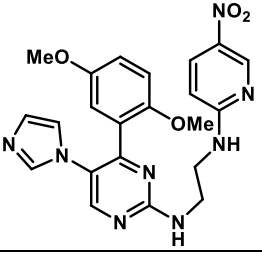
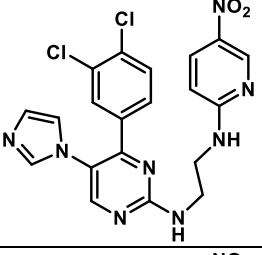
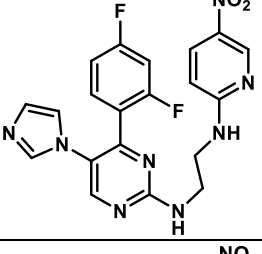
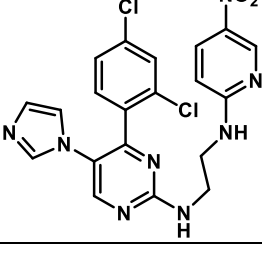
19b	012		0.033	7.481
19c	013		0.061	7.215
19d	014		0.110	6.959
19e	015		0.075	7.125
19f	016		0.047	7.328
19g	017		0.029	7.538

19h	018		0.153	6.815
19i	019		0.071	7.149
19j	020		0.140	6.854
19k	021		0.161	6.793
19l	022		0.417	6.380
20	023		0.033	7.481

26	024		0.019	7.721
25a	025		0.011	7.959
35a	026		0.019	7.721
25b	027		0.197	6.706
25c	028		0.200	6.699
25d	029		0.053	7.276

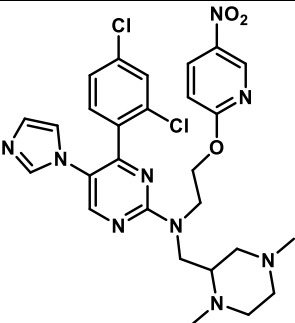
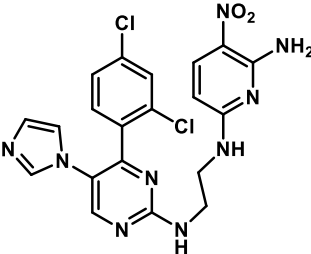
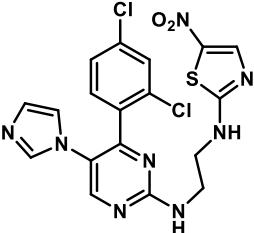
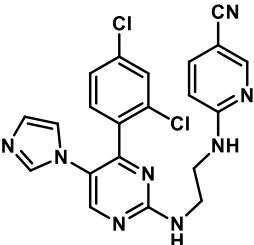
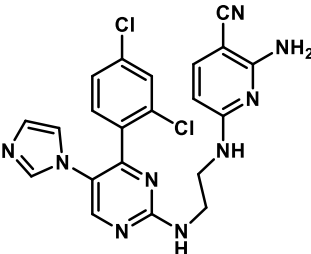
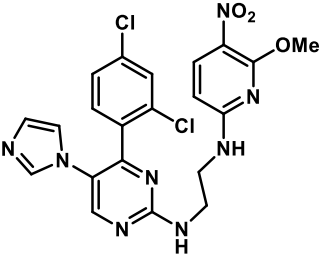
25e	030		0.034	7.469
25f	031		0.040	7.398
25g	032		0.068	7.167
25h	033		0.009	8.046
25i	034		0.124	6.907
25j	035		0.777	6.110

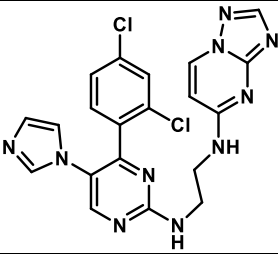
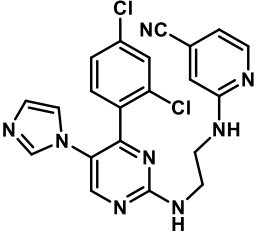
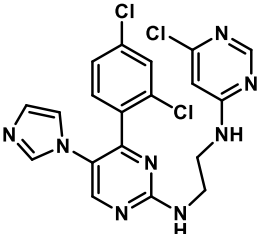
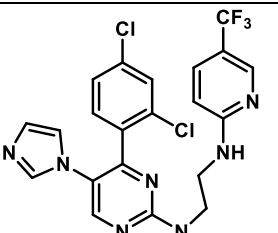
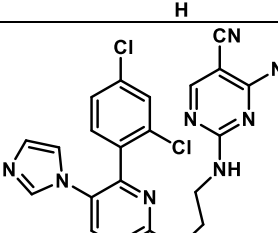
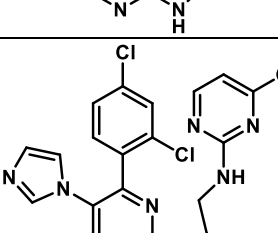
25k	036		0.029	7.538
25l	037		0.166	6.780
25m	038		1.117	2.952
25n	039		1.930	2.714
25o	040		0.024	7.620

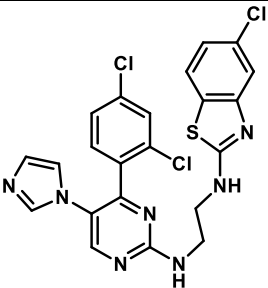
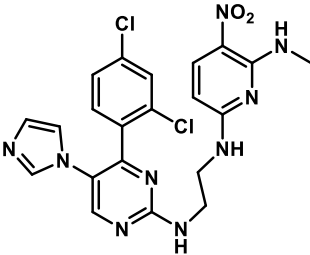
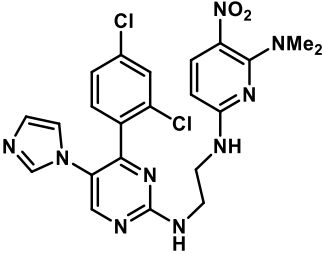
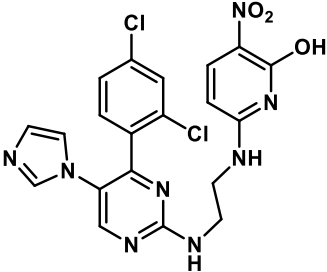
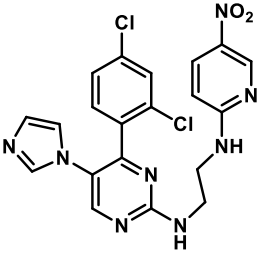
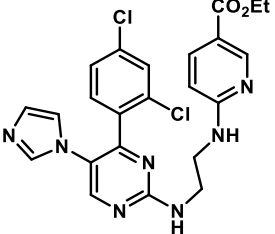
25p	041		0.884	6.054
25q	042		0.347	6.460
25r	043		4.870	2.312
25s	044		0.646	6.190
25t	045		0.031	7.509
25u	046		0.006	8.222

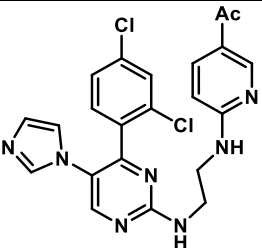
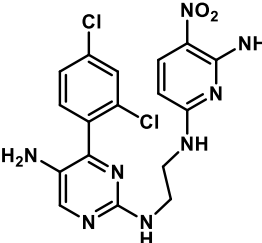
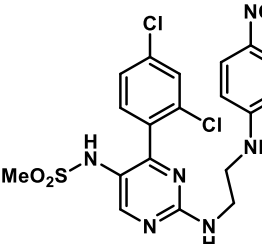
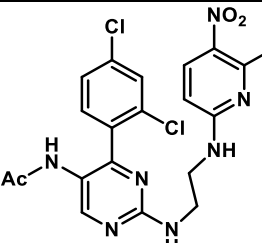
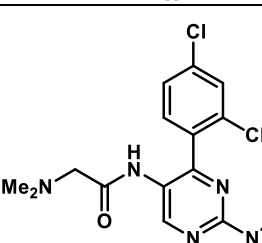
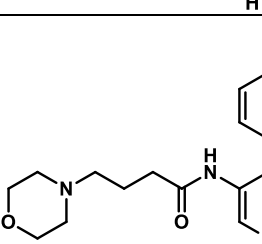
25v	047		0.038	7.420
25w	048		0.027	7.569
38	049		2.829	2.548
39	050		0.022	7.658
46a	051		0.175	6.757
46b	052		0.035	7.456

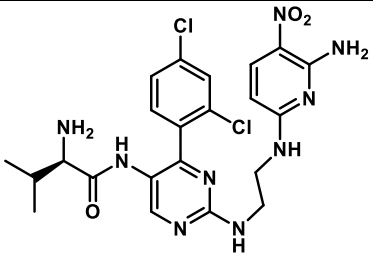
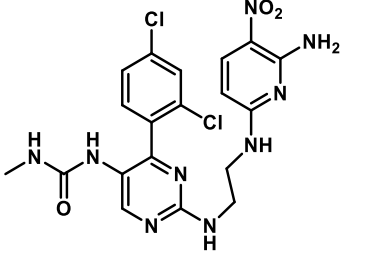
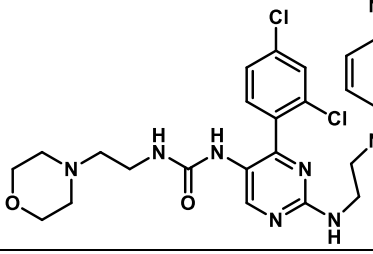
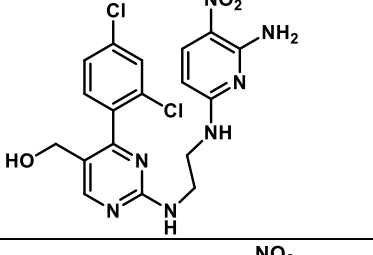
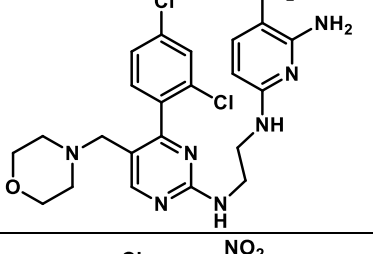
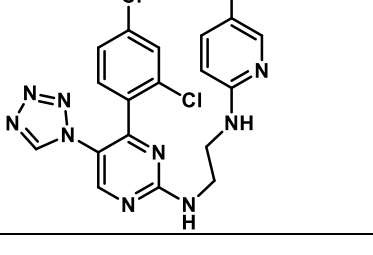
46c	053		0.026	7.585
46d	054		0.027	7.569
46e	055		0.015	7.824
46f	056		0.032	7.495
46g	057		0.113	6.947

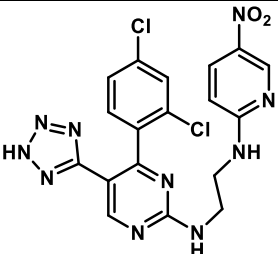
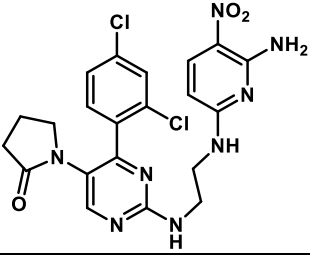
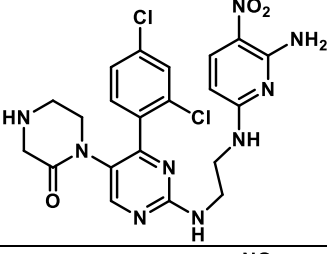
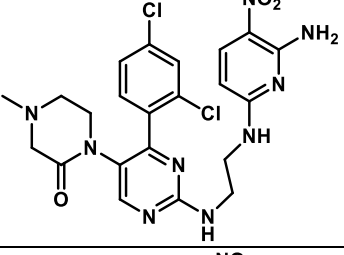
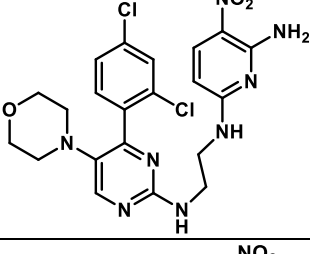
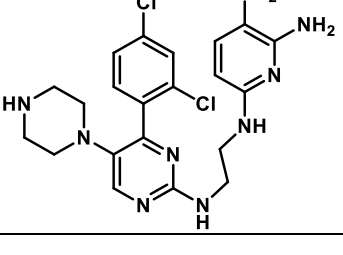
46h	058		0.121	6.917
2	059		0.001	9.000
51a	060		0.024	7.620
51b	061		0.025	7.602
51c	062		0.056	7.252
51d	063		0.081	7.092

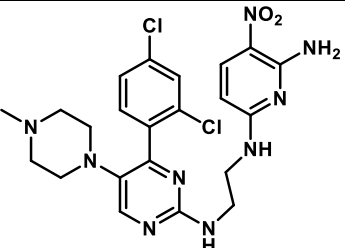
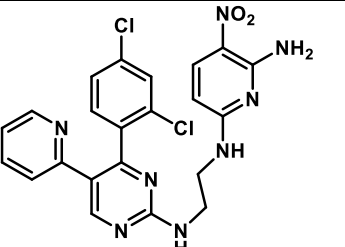
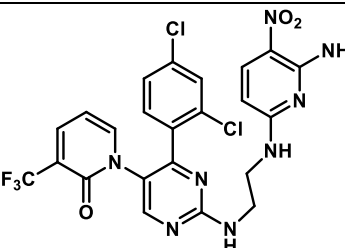
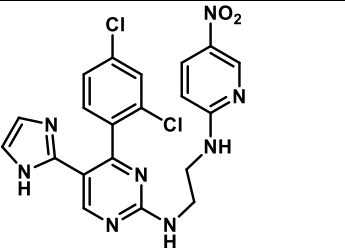
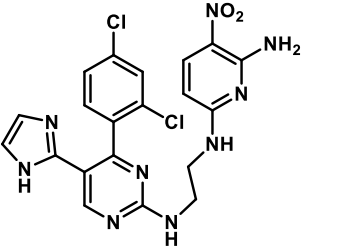
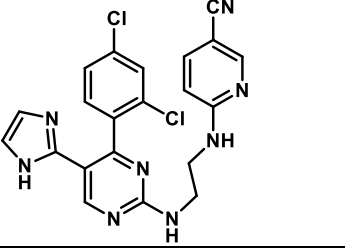
51e	064		0.108	6.967
51f	065		0.200	6.699
51g	066		0.201	6.697
51h	067		0.223	6.652
51i	068		0.236	6.627
51j	069		0.381	6.419

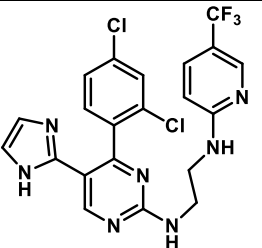
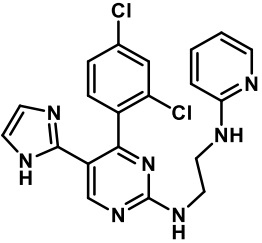
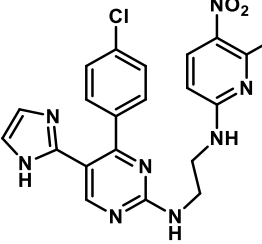
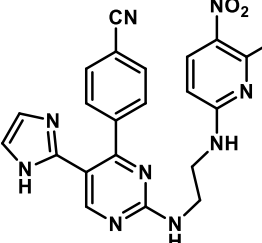
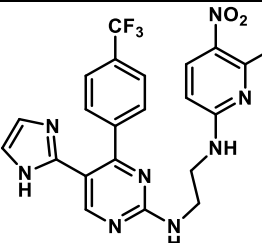
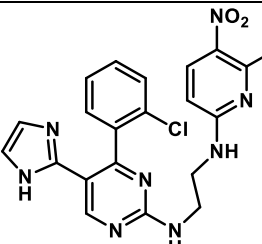
51k	070		0.527	6.278
51l	071		0.576	6.240
51m	072		1.000	3.00
51n	073		1.000	3.00
51o	074		3.370	2.472
51p	075		3.650	2.438

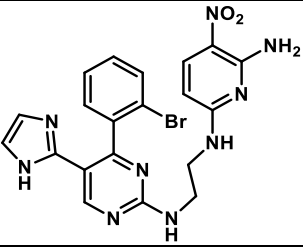
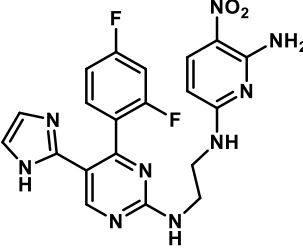
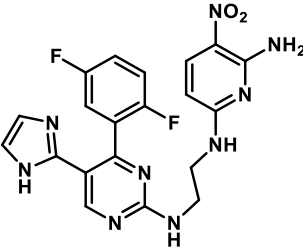
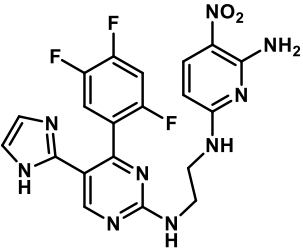
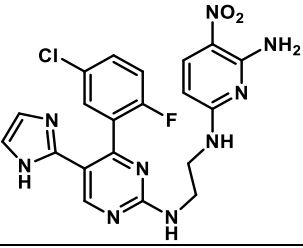
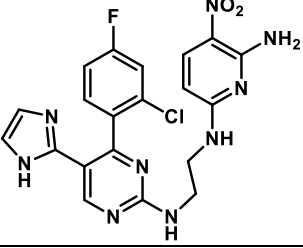
51q	076		5.950	2.225
60a	077		0.010	8.000
60b	078		0.009	8.046
60c	079		0.006	8.222
60d	080		0.025	7.602
60e	081		0.024	7.620

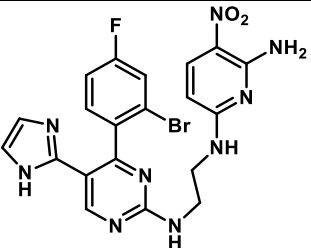
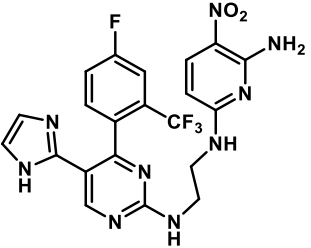
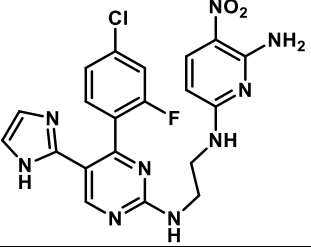
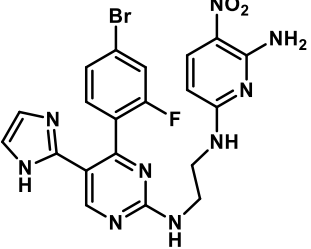
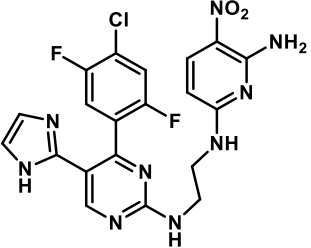
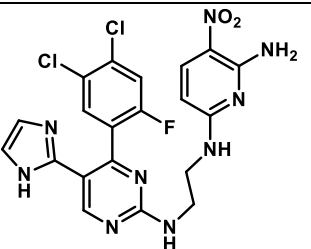
60f	082		0.014	7.854
60g	083		0.027	7.569
60h	084		0.016	7.796
60i	085		0.007	8.155
60j	086		0.017	7.770
60k	087		0.015	7.824

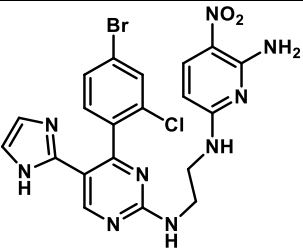
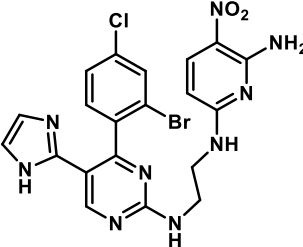
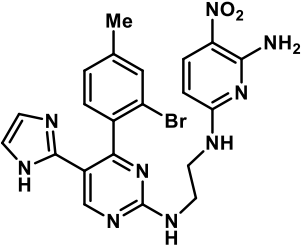
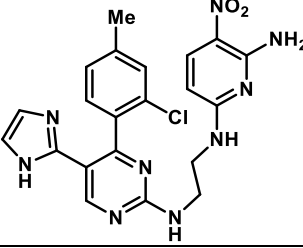
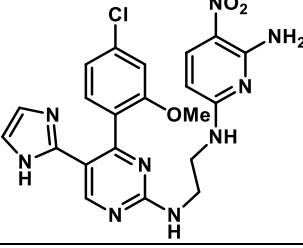
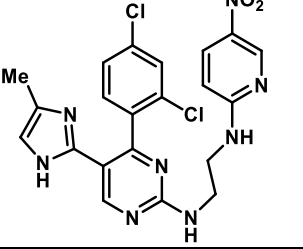
60l	088		0.031	7.509
60m	089		0.006	8.222
60n	090		0.015	7.824
60o	091		0.004	8.398
60p	092		0.007	8.155
60q	093		0.018	7.745

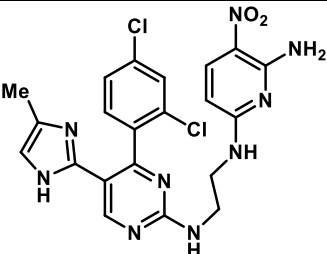
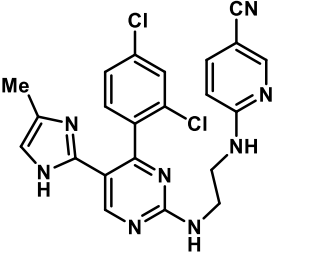
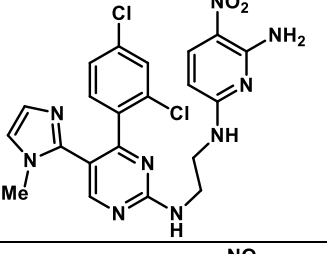
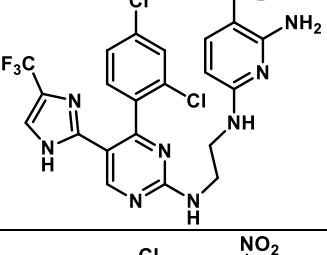
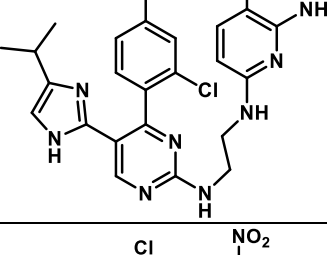
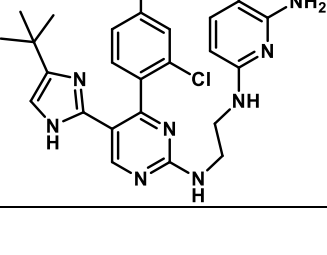
60r	094		0.006	8.222
60s	095		0.007	8.155
60t	096		0.004	8.398
36a	097		0.004	8.398
36b	098		0.001	9.000
36c	099		0.004	8.398

36d	100		0.035	7.456
61a	101		2.856	2.544
61b	102		0.031	7.509
35b	103		0.007	8.155
61c	104		0.051	7.292
61d	105		0.012	7.921

61e	106		0.007	8.155
61f	107		0.009	8.046
61g	108		0.047	7.328
61h	109		0.038	7.420
61i	110		0.032	7.495
61j	111		0.007	8.155

61k	112		0.004	8.398
61l	113		0.013	7.886
61m	114		0.005	8.301
61n	115		0.005	8.301
61o	116		0.019	7.721
61p	117		0.013	7.886

61q	118		0.004	8.398
61r	119		0.006	8.222
61s	120		0.004	8.398
61t	121		0.019	7.721
61u	122		0.249	6.604
37a	123		0.008	8.097

37b	124		0.002	8.699
1	125		0.005	8.301
62a	126		0.001	9.000
62b	127		0.034	7.469
62c	128		0.021	7.678
62d	129		0.027	7.569

62e	130		0.001	9.000
62f	131		0.003	8.523
62g	132		0.022	7.658

4.1.2. Realización del análisis de minimización de energía

Luego de diseñar el set de moléculas y llevarlas a un área tridimensional, fue necesario para el análisis 3D-QSAR el utilizar técnicas de minimización de energía. La primera prueba fue realizada a todas las moléculas, cada compuesto diseñado fue sometido a una optimización geométrica preliminar utilizando el campo de mecánica molecularo MM2 implementado en ChemBio3D. Por otro lado, se estudió un total de 4 moléculas cuyos valores de actividad biológica eran superiores al resto ($pIC_{50} = 9.00$, compuestos **59**, **98**, **126** y **130**). Todas con valor de pIC_{50} igual a 9. Este set de moléculas, después del cálculo MM2, fueron refinados por un método de cálculo semiempírico AM1 en el programa GAUSSIAN en su versión 5.0. Luego fueron sometidas a un cálculo *ab initio* (mecánico cuántico) con el método

de la teoría funcional de la densidad (DFT) usando como set de base B3LYP/6-311G ++ añadiendo orbitales d y p. Esto con el fin de generar una distribución de carga no homogénea, obteniendo como resultado la conformación geométrica más aceptada según los principios cuánticos. Luego de esto, todas las moléculas fueron ingresadas al programa SYBYL-X en su versión 1.2 donde, al resto de moléculas que no pasaron por el programa GAUSSVIEW, se les realizó el método de minimización de Powell¹⁰⁰ para ajustar los ángulos de enlace.

4.2. Alineamiento molecular

El alineamiento, y todo el resto del estudio 3D-QSAR, se llevó a cabo en el programa SYBYL-X en su versión 1.2¹⁰¹. utilizando la opción “Align Compounds” y “Fit Atom”. Esto con el fin de que, el resto de las moléculas que no pasaron por el análisis DFT, tengan la misma geometría que las moléculas templado. Por lo que en la opción “Fit “Atom” se selecciona una molécula molde, que en este caso fue la molécula **130** (optimizada por DFT) y átomo por átomo se fue acomodando cada molécula a la estructura obtenida en el DFT. Esto se realizó tomando como farmacóforo la zona de la pirimidina central (PC), la zona del conector, y la zona del lado derecho (LD). Dando como resultado lo obtenido en la **Figura 18**. Finalizando este procedimiento se le asignó cargas MMFF94 a cada átomo de cada molécula del set.

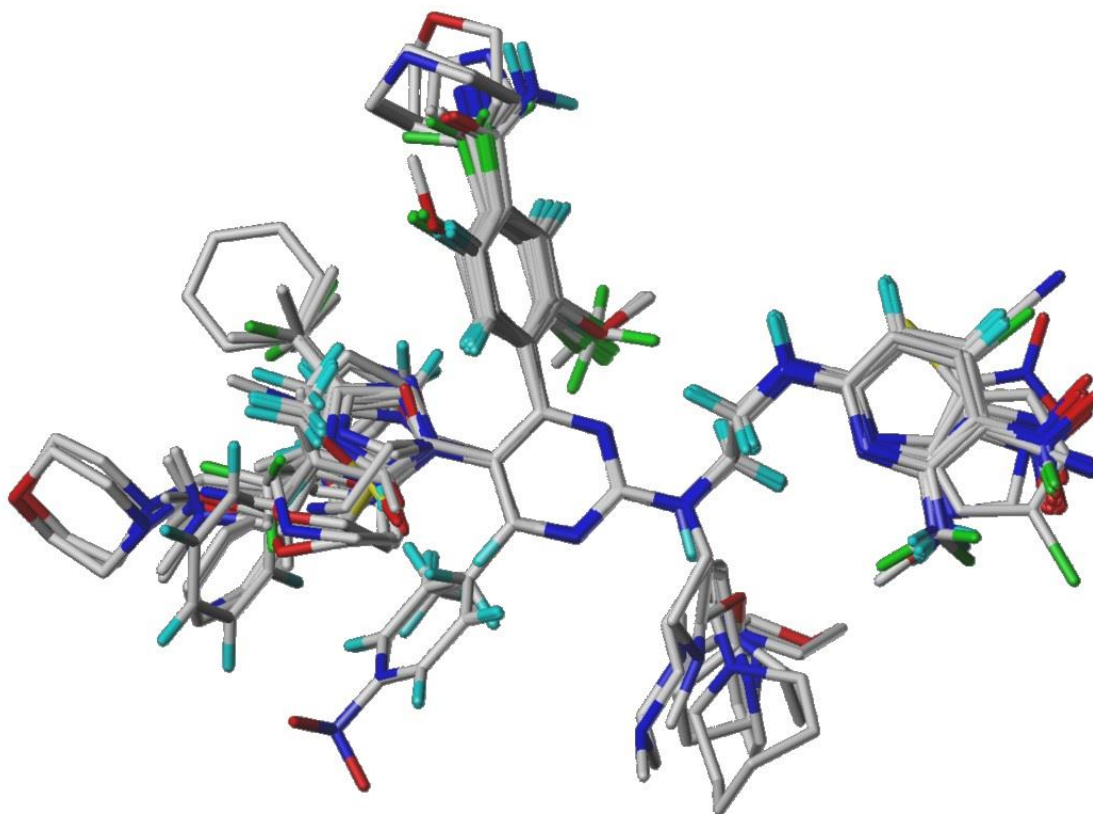


Figura 18: Alineamiento del set de moléculas utilizados en la metodología 3D-QSAR.

4.3. Análisis QSAR

4.3.1. Cálculo de los campos CoMFA y CoMSIA

Para derivar los campos descriptores del CoMFA y CoMSIA, las moléculas del conjunto de entrenamiento (“Training Set”) alineadas se colocaron en una celosía cúbica tridimensional con un espaciado cuadrículado de 2\AA en las direcciones X, Y, Y, y Z de manera que todo el conjunto de moléculas se incluyera en él. Las energías de campo estérico y electrostático del CoMFA se calcularon utilizando un átomo de sonda de carbono sp^3 con un radio de Van der Waals de $1,52\text{\AA}$ y una carga de $+1.0$. Los valores de corte para los

campos estéricos y electrostáticos se establecieron en 30.0 kcal/mol. Para el análisis CoMSIA, los ajustes estándar (sonda con carga +1.0, radio 1Å, hidrofobia +1.0, donación de puentes de hidrógen +1.0 y aceptores de puentes de hidrógeno +1.0⁹⁷) se utilizaron para calcular cinco campos diferentes: estérico, electrostático, hidrofóbico, donante y aceptor de puentes de hidrógeno. La dependencia de la distancia de tipo gaussiano se utilizó para medir la atenuación relativa de la posición del campo de cada átomo en la red y condujo a un muestreo mucho más suave de los campos alrededor de las moléculas en comparación con el CoMFA. El valor por defecto se ajustó a 0.3 para el factor de atenuación α .

4.3.2. Selección de los mejores confórmeros

Se realizaron estudios de CoMFA y CoMSIA en un conjunto de 122 derivados de (5-imidazol-2-il-4-fenilpirimidin-2-il)[2- (2-piridilamino)etil]amina reportados por Allan et al⁶⁵.(Tabla 1). Los derivados mostraron una potente actividad inhibitoria de la enzima glucógeno sintasa quinasa 3 (GSK3). Los valores de IC₅₀ se transformaron a pIC₅₀ (-log IC₅₀). De las 132 moléculas iniciales, se descartaron 10 moléculas que se encontraban entre los rangos de 2.22 y 3.00 en pIC₅₀, mientras que el resto de las 122 moléculas van de los 6.05 y 9.00, por lo que se tomó la decisión de trabajar sólo con 122 moléculas que presentaban mejores valores de actividad biológica. Las moléculas descartadas fueron: **38, 39, 43, 49, 72, 73, 74, 75, 76**, y la molécula **101**. Se evaluaron varias combinaciones de conjuntos de entrenamiento (Training set) y conjuntos de prueba (Test set). Los compuestos se dividieron manual y aleatoriamente en Training set (86 compuestos, 70.5%) y Test set (36 compuestos, 29.5%), asegurando que ambos sets contuviesen compuestos estructuralmente diversos con actividad alta, media y baja, y una distribución uniforme para evitar posibles problemas

durante la validación externa. En la búsqueda secuencial de los mejores modelos se evaluaron varios grupos de Test set. Para la construcción de los modelos finales, se seleccionó el Test set que generó el valor más alto de r^2 en cada caso (CoMFA y CoMSIA). En la **Figura 19** se muestra la distribución de los valores de pIC_{50} para el conjunto completo (Whole Set), el Training set y el Test set. En los tres casos, la actividad biológica siguió una distribución gaussiana. El rango de actividades biológicas va de 6.05 a 9.00.

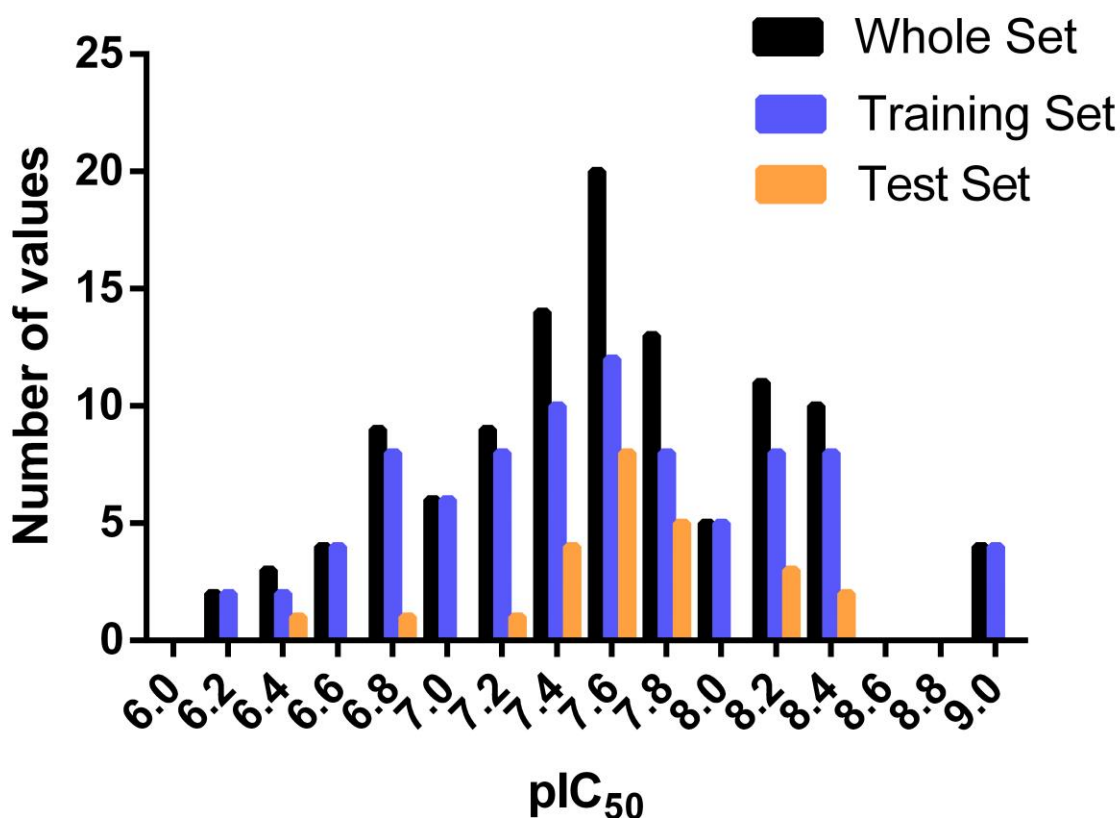


Figura 19: Histograma de la frecuencia de distribución de datos.

4.3.3. Validación interna y análisis de mínimos cuadrados parciales (PLS)

El análisis PLS se utilizó para construir una correlación lineal entre los descriptores CoMFA y CoMSIA (variables independientes) y los valores de actividad (valores dependientes¹⁰²). Para seleccionar el mejor modelo, el análisis de validación cruzada se realizó utilizando el método “dejar-uno-afuera” (en inglés LOO) (y distancia de muestras PLS [en inglés SAMPLS]), que genera el cuadrado del coeficiente de validación cruzada (q^2) y el número óptimo de componentes (N). La validación no cruzada se realizó con un valor de filtro de columna (column filter) de 2.0 para acelerar el análisis y reducir el ruido. El valor de q^2 , que es una medida de la calidad interna de los modelos, se obtuvo de acuerdo a la siguiente ecuación (ecuación 11):

$$q^2 = 1 - \frac{\sum(y_i - y_{\text{pred}})^2}{\sum(y_i - \bar{y})^2} \quad \text{ec(11)}$$

Donde y_i , \bar{y} , e y_{pred} son los valores experimentales promedio y predictivo de actividad biológica en el Training set, respectivamente.

4.3.4. Validación externa

Los modelos fueron sometidos a criterios de validación externa de acuerdo al test de validación propuesto por Golbraikh y Tropsha^{103,104}, que considera predictivo un modelo QSAR, si se cumplen las siguientes condiciones (ecuaciones 12-16):

$$q^2 > 0.5 \quad \text{ec(12)}$$

$$r_{\text{test}}^2 > 0.6 \quad \text{ec(13)}$$

$$\frac{(r_{test}^2 - r_0^2)}{r_{test}^2} < 0.1 \text{ or } \frac{(r_{test}^2 - r_0'^2)}{r_{test}^2} < 0.1 \quad \text{ec(14)}$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \quad \text{ec(15)}$$

$$|r_0^2 - r_0'^2| < 0.3 \quad \text{ec(16)}$$

Se ha demostrado¹⁰³ que todos los criterios anteriores son realmente necesarios para evaluar adecuadamente la capacidad predictiva de un modelo QSAR.

Adicionalmente, el poder predictivo externo de los modelos se examinó calculando el valor de r_m^2 métrico como se muestra a continuación¹⁰⁵ (ecuación 17):

$$r_m^2 = r_{test}^2 (1 - \sqrt{|r^2 - r_0^2|}) \quad \text{ec(17)}$$

Donde r_{test}^2 y r_0^2 son los coeficientes de correlación al cuadrado entre las actividades observadas y predichas del Test set con y sin la intercepción (0,0), respectivamente. Para una validación externa significativa, el valor de r_m^2 debe ser mayor a 0.5.

4.3.5. Dominio de aplicabilidad (DA)

El DA fue evaluado en base al método de estandarización reportado por Roy et al.¹⁰⁶. Primero, se estandariza (S_{ik}) cada descriptor “i” para cada compuesto “k”. Todos los compuestos deben tener un valor $[Si]_{\text{máx}(k)} \leq 3$. En el caso de que $[Si]_{\text{máx}(k)} > 3$ y su valor $[Si]_{\text{mín}(k)} < 3$, entonces el parámetro $S_{\text{new}(k)}$ debe ser calculado y debe cumplir la condición: $S_{\text{new}(k)} = \bar{S}_k + 1.28 * \sigma_{S_k}$, donde \bar{S}_k es la medida de los valores S_{ik} para el compuesto k y σ_{S_k} , es la desviación estándar para dichos valores. El software está disponible en el sitio web de los autores: <http://dtclab.webs.com/software-tools> y http://teqip.jdvu.ac.in/QSAR_Tools/.

4.4. Obtención de mapas de contornos y predicciones

Luego de obtener los mapas de contorno y las validaciones estadísticas internas y externas, se procedió a generar el diseño de nuevos compuestos de la familia (5-imidazol-2-il-4-fenilpirimidin-2-il)[2- (2-piridilamino)etil]amina en base a lo obtenido en los mapas de contornos de los modelos CoMFA y CoMSIA, donde finalmente se les calculó su valor de pIC_{50} predictivo y se seleccionaron los 10 mejores compuestos predichos.

5. RESULTADOS Y DISCUSIÓN

5.1. Resultados estadísticos

En la **Tabla 2** se presenta un resumen de los resultados estadísticos de los modelos CoMFA y CoMSIA. La búsqueda de nuevos modelos fue realizada a través de combinaciones sucesivas de campo. El primer parámetro con el cual se evaluó la solidez estadística del modelo QSAR fue el valor de q^2 . El q^2 es el coeficiente de validación cruzada obtenido en el cálculo “Leave-one-out”(LOO); es un indicador de capacidad predictiva interna de un modelo QSAR. Idealmente q^2 debe tener un valor mayor a 0.5. Para CoMFA, el modelo que consideró las contribuciones de campo estérico y electrostático (CoMFA-SE) presentó un valor de 0.563, mientras que el modelo CoMSIA estérico + hidrofóbico + dador de puentes de hidrógeno + aceptor de puentes de hidrógeno (CoMSIA-SHDA) presentó un valor mayor de q^2 que el obtenido en CoMFA-SE, siendo este de 0.598. El otro parámetro estadístico importante es el r^2_{test} , que evalúa la capacidad predictiva externa del modelo. Para CoMFA-SE, el valor del r^2_{test} fue de 0.823, mientras que el modelo CoMSIA-SHDA obtuvo un valor de 0.925. A pesar de la existencia de combinaciones con un valor de q^2 mayor a esta, esas combinaciones no presentaron un resultado de r^2_{test} como el obtenido en el modelo CoMSIA-SHDA. Otro parámetro que se consideró fue el número óptimo de componentes (N). Idealmente, un buen modelo debe tener la menor cantidad posible de componentes, el valor de N debe ser menor a un tercio del número total de compuestos estudiados, lo que asegura que las predicciones se basarán en información significativa de las contribuciones de campo, en lugar de generar un modelo “sobre-entrenado”. El valor de N para los modelos CoMFA y CoMSIA fue de 7. El término SEP corresponde al error estándar de predicción, en

el cual las dos mejores combinaciones generaron errores estándar bajos. El término r_{ncv}^2 es el cuadrado del coeficiente de validación no cruzada, en el cual se obtuvieron valores altos. También están los términos F y SEE que corresponden al valor de F-test de Fisher y al error estándar de estimación del análisis de validación no cruzada. Por otra parte, el modelo CoMFA, presentó un porcentaje de contribución de un 49,7% para la contribución estérica y un 50,3% para la electrostática. Por otro lado, el modelo CoMSIA presentó un 19,5% para la contribución estérica, 20.8% para la hidrofóbica, 27.6% para la capacidad de donar puentes de hidrógeno, y un 32.1% para la capacidad aceptora de puentes de hidrógeno. Estos porcentajes respaldan la fiabilidad de las conclusiones obtenidas de cada mapa de contorno.

Tabla 2: Parámetros estadísticos y combinaciones de campos para CoMFA y CoMSIA.

Modelo	q ²	N	SEP	SEE	r ² _{ncv}	F	Contribuciones					
							S	E	H	D	A	
CoMFA-S	0.439	9	0.516	0.141	0.958	193.662	1					
CoMFA-E	0.406	5	0.518	0.315	0.780	56.579		1				
CoMFA-SE	0.563	7	0.451	0.165	0.941	176.85	0.497	0.503				
CoMSIA-S	0.185	5	0.607	0.448	0.555	19.991	1					
CoMSIA-E	0.454	20	0.551	0.192	0.934	45.880		1				
CoMSIA-H	0.434	4	0.502	0.362	0.707	48.771			1			
CoMSIA-D	0.376	5	0.531	0.438	0.575	21.607				1		
CoMSIA-A	0.429	5	0.508	0.412	0.624	26.603						1
CoMSIA-SE	0.399	3	0.514	0.395	0.646	49.950	0.294	0.706				
CoMSIA-SEH	0.439	4	0.500	0.363	0.704	48.100	0.226	0.553	0.221			
CoMSIA-SEHD	0.488	3	0.475	0.375	0.680	58.15	0.151	0.351	0.172	0.327		
CoMSIA-SEHA	0.477	4	0.483	0.348	0.728	54.3	0.146	0.381	0.157			0.317
CoMSIA-SED	0.467	3	0.485	0.384	0.664	54.124	0.184	0.447		0.369		
CoMSIA-SEA	0.448	12	0.523	0.175	0.938	92.047	0.217	0.436				0.347
CoMSIA-SEDA	0.487	3	0.475	0.372	0.686	59.762	0.138	0.332		0.307	0.223	
CoMSIA-SH	0.339	4	0.543	0.382	0.674	41.788	0.414		0.586			
CoMSIA-SD	0.391	2	0.515	0.468	0.496	40.866	0.287			0.713		
CoMSIA-SA	0.476	7	0.493	0.311	0.792	42.319	0.336					0.664
CoMSIA-SHD	0.463	5	0.492	0.359	0.714	39.918	0.3		0.303	0.397		
CoMSIA-SHA	0.539	8	0.465	0.221	0.896	82.682	0.264		0.298			0.438

CoMSIA-SDA	0.43	10	0.524	0.259	0.860	46.222	0.247		0.403	0.35
CoMSIA-SHDA	0.595	11	0.445	0.175	0.937	100.486	0.181	0.227	0.306	0.285
CoMSIA-EH	0.44	4	0.5	0.376	0.682	43.522		0.699	0.301	
CoMSIA-ED	0.453	3	0.491	0.381	0.671	55.652		0.587	0.413	
CoMSIA-EA	0.521	13	0.49	0.195	0.924	67.41		0.582		0.418
CoMSIA-EHD	0.523	12	0.486	0.17	0.942	97.966		0.413	0.236	0.351
CoMSIA-EHA	0.536	11	0.476	0.162	0.946	118.195		0.4	0.239	0.361
CoMSIA-EDA	0.48	3	0.479	0.388	0.659	52.821		0.374	0.349	0.377
CoMSIA-EHDA	0.586	11	0.45	0.167	0.943	111.506		0.279	0.184	0.292
CoMSIA-HD	0.521	6	0.468	0.35	0.732	35.924		0.441	0.559	
CoMSIA-HA	0.557	10	0.462	0.201	0.916	82.218		0.459		0.541
CoMSIA-HDA	0.626	11	0.427	0.193	0.924	81.745		0.284	0.377	0.339
CoMSIA-DA	0.428	8	0.52	0.338	0.756	29.857			0.54	0.46
CoMSIA-ALL	0.572	11	0.482	0.158	0.949	125.294	0.126	0.24	0.152	0.255
CoMSIA-HDA	0.6	7	0.43	0.252	0.863	70.274		0.281	0.349	0.37
CoMSIA-SHDA	0.598	7	0.433	0.217	0.899	98.221	0.195		0.208	0.276

La **Tabla 3** presenta un resumen de la validación externa de los modelos CoMFA-SE y CoMSIA-SHDA. Ambos modelos presentan un valor alto para el r^2_{test} (0.8226 y 0.9245 respectivamente), lo que es una indicación de una capacidad predictiva externa adecuada. Sin embargo, de acuerdo con Golbraikh y Tropsha, los valores altos de q^2 y r^2_{test} (condiciones 1 y 2) son necesarias, pero no son condiciones suficientes para la validación de un modelo. Para que un modelo QSAR tenga una capacidad predictiva confiable, la línea para la actividad experimental frente a la predicha debe estar lo más cerca posible de la línea $y = x$.

Esto se observa en el cumplimiento de las condiciones [3a o 3b], [4a o 4b], [5a o 5b] y 6 enumeradas en la **Tabla 3**. Finalmente, la condición 7, conocida como r^2_m métrico, es una medida cuantitativa para determinar la proximidad entre la actividad observada y la predicha para el conjunto de pruebas. Los modelos CoMFA-SE y CoMSIA-SHDA informados aquí cumplieron las condiciones para la validación interna y externa y, en general, el modelo CoMSIA-SHDA muestra mejores parámetros estadísticos que el modelo

CoMFA-SE. Los valores de actividad experimental, actividad predicha y valores residuales para los mejores modelos CoMFA y CoMSIA se muestran en la **Tabla 4**.

Tabla 3: Resumen de los parámetros de validación externa para el CoMFA y el CoMSIA.

Condition	Parameters	Threshold value	CoMFA	CoMSIA
1	q^2	>0.5	0.563	0.598
2	r_{test}^2	>0.6	0.8226	0.925
3a	r_0^2	Close to value of r_{test}^2	0.696	0.830
3b	$r_0'^2$	Close to value of r_{test}^2	0.809	0.891
4a	k	$0.85 < k < 1.15$	0.985	0.996
4b	k'	$0.85 < k' < 1.15$	0.102	0.100
5a	$(r_{test}^2 - r_0^2)/r_{test}^2$	<0.1	0.155	0.102
5b	$(r_{test}^2 - r_0'^2)/r_{test}^2$	<0.1	0.016	0.036
6	$ r_0^2 - r_0'^2 $	<0.3	0.114	0.061
7	r_m^2	>0.5	0.529	0.641

Tabla 4: Valores de pIC₅₀ experimentales, predictivos y valores residuales para los modelos CoMFA y CoMSIA.

Moléculas	pIC ₅₀ experimental	CoMFA		CoMSIA	
		pIC ₅₀ predichas	Residuales	pIC ₅₀ predichas	Residuales
1	6.921	6.889	0.032	6.985	-0.064
2	6.854	6.793	0.061	6.971	-0.117
3	6.824	6.980	-0.156	6.913	-0.089
4	7.398	7.347	0.051	7.379	0.019
5	7.284	7.249	0.035	7.354	-0.070
6	7.602	7.268	0.334	7.321	0.281
7	6.721	6.767	-0.046	6.820	-0.099
8	7.292	7.419	-0.127	7.384	-0.092
9	7.301	6.639	0.662	7.140	0.161
10	7.456	7.622	-0.166	7.623	-0.167
11	7.538	7.567	-0.029	7.392	0.146
12	7.481	7.443	0.038	7.371	0.110

13	7.215	7.248	-0.033	7.155	0.060
14	6.959	7.035	-0.076	7.258	-0.299
15	7.125	7.049	0.076	7.151	-0.026
16	7.328	7.061	0.267	7.196	0.132
17	7.538	7.351	0.187	7.106	0.432
18	6.815	6.666	0.149	6.644	0.171
19	7.149	7.155	-0.006	7.159	-0.010
20	6.854	6.691	0.163	6.654	0.200
21	6.793	6.687	0.106	6.822	-0.029
22	6.380	6.547	-0.167	6.147	0.233
23	7.481	7.525	-0.044	7.939	-0.458
24	7.721	7.874	-0.153	7.741	-0.020
25	7.959	7.644	0.315	7.613	0.346
26	7.721	7.899	-0.178	7.811	-0.090
27	6.706	6.705	0.001	7.023	-0.317
28	6.699	7.060	-0.361	7.208	-0.509
29	7.276	7.252	0.024	7.321	-0.045
30	7.469	7.321	0.148	7.237	0.232
31	7.398	7.280	0.118	7.257	0.141
32	7.167	7.356	-0.189	7.127	0.040
33	8.046	7.229	0.817	6.919	1.127
34	6.907	7.692	-0.785	6.979	-0.072
35	6.110	6.359	-0.249	6.199	-0.089
36	7.538	7.649	-0.111	7.611	-0.073
37	6.780	6.571	0.209	6.318	0.462
40	7.620	7.607	0.013	7.637	-0.017
41	6.054	7.649	-1.595	7.008	-0.954
42	6.460	6.640	-0.180	6.709	-0.249
44	6.190	7.191	-1.001	7.154	-0.964
45	7.509	7.365	0.144	7.424	0.085
46	8.222	7.729	0.493	7.731	0.491
47	7.420	7.605	-0.185	7.341	0.079
48	7.569	7.620	-0.051	6.647	0.922
50	7.658	7.529	0.129	7.857	-0.199
51	6.757	6.872	-0.115	6.734	0.023
52	7.456	7.325	0.131	7.388	0.068
53	7.585	7.409	0.176	7.574	0.011
54	7.569	7.390	0.179	7.629	-0.060
55	7.824	7.807	0.017	8.017	-0.193
56	7.495	7.717	-0.222	7.610	-0.115
57	6.947	6.989	-0.042	6.971	-0.024
58	6.917	6.908	0.009	7.089	-0.172
59	9.000	8.891	0.109	8.984	0.016

60	7.620	7.415	0.205	6.465	1.155
61	7.602	7.578	0.024	7.797	-0.195
62	7.252	7.161	0.091	7.157	0.095
63	7.092	7.182	-0.090	7.290	-0.198
64	6.967	6.972	-0.005	6.854	0.113
65	6.699	7.207	-0.508	6.949	-0.250
66	6.697	6.808	-0.111	6.912	-0.215
67	6.652	6.890	-0.238	7.012	-0.360
68	6.627	6.930	-0.303	6.812	-0.185
69	6.419	6.239	0.180	6.399	0.020
70	6.278	6.206	0.072	6.335	-0.057
71	6.240	7.664	-1.424	7.220	-0.980
77	8.000	8.003	-0.003	7.757	0.243
78	8.046	8.008	0.038	8.053	-0.007
79	8.222	8.374	-0.152	8.020	0.202
80	7.602	7.520	0.082	7.783	-0.181
81	7.620	7.688	-0.068	7.918	-0.298
82	7.854	7.667	0.187	7.813	0.041
83	7.569	7.669	-0.100	7.605	-0.036
84	7.796	7.828	-0.032	8.053	-0.257
85	8.155	8.201	-0.046	8.084	0.071
86	7.770	7.696	0.074	7.733	0.037
87	7.824	7.865	-0.041	7.707	0.117
88	7.509	7.500	0.009	7.820	-0.311
89	8.222	7.980	0.242	7.979	0.243
90	7.824	7.936	-0.112	7.756	0.068
91	8.398	8.378	0.020	8.482	-0.084
92	8.155	8.064	0.091	8.000	0.155
93	7.745	7.815	-0.070	7.696	0.049
94	8.222	8.288	-0.066	8.093	0.129
95	8.155	7.715	0.440	7.752	0.403
96	8.398	8.238	0.160	8.305	0.093
97	8.398	7.972	0.426	8.154	0.244
98	9.000	8.955	0.045	9.291	-0.291
99	8.398	8.517	-0.119	8.129	0.269
100	7.456	7.600	-0.144	7.253	0.203
102	7.509	7.915	-0.406	7.554	-0.045
103	8.155	7.908	0.247	7.856	0.299
104	7.292	7.674	-0.382	7.421	-0.129
105	7.921	8.060	-0.139	7.899	0.022
106	8.155	7.881	0.274	8.087	0.068
107	8.046	7.659	0.387	7.669	0.377
108	7.328	7.325	0.003	7.520	-0.192

109	7.420	7.540	-0.120	7.643	-0.223
110	7.495	7.600	-0.105	7.492	0.003
111	8.155	8.154	0.001	7.970	0.185
112	8.398	8.266	0.132	8.220	0.178
113	7.886	7.730	0.156	7.863	0.023
114	8.301	7.798	0.503	7.712	0.589
115	8.301	7.978	0.323	7.779	0.522
116	7.721	7.582	0.139	7.659	0.062
117	7.886	7.783	0.103	7.635	0.251
118	8.398	8.346	0.052	8.103	0.295
119	8.222	8.410	-0.188	8.203	0.019
120	8.398	8.174	0.224	8.087	0.311
121	7.721	8.000	-0.279	7.847	-0.126
122	6.604	8.111	-1.507	7.561	-0.957
123	8.097	8.148	-0.051	8.049	0.048
124	8.699	8.032	0.667	7.986	0.713
125	8.301	8.403	-0.102	8.097	0.204
126	9.000	9.082	-0.082	8.719	0.281
127	7.469	8.174	-0.705	7.988	-0.519
128	7.678	7.618	0.060	7.721	-0.043
129	7.569	7.387	0.182	7.888	-0.319
130	9.000	9.009	-0.009	9.099	-0.099
131	8.523	7.877	0.646	8.248	0.275
132	7.658	7.488	0.170	7.845	-0.187

En la **Figura 20** se muestra un gráfico de regresión donde se utilizaron los valores logarítmicos de actividad biológica experimentales versus los valores logarítmicos de la actividad biológica predichas para el modelo CoMFA. Se puede apreciar que los datos presentan una distribución lineal en el eje trazado, es decir que están cerca de la línea $Y = X$. Esto quiere decir que el modelo presenta un buen equilibrio en términos de capacidad predictiva. La **Figura 21** muestra el gráfico de regresión del modelo CoMSIA. En este gráfico también se puede apreciar una distribución lineal en la zona de la línea $Y = X$, por lo que este modelo también presenta un buen equilibrio en términos de capacidad predictiva.

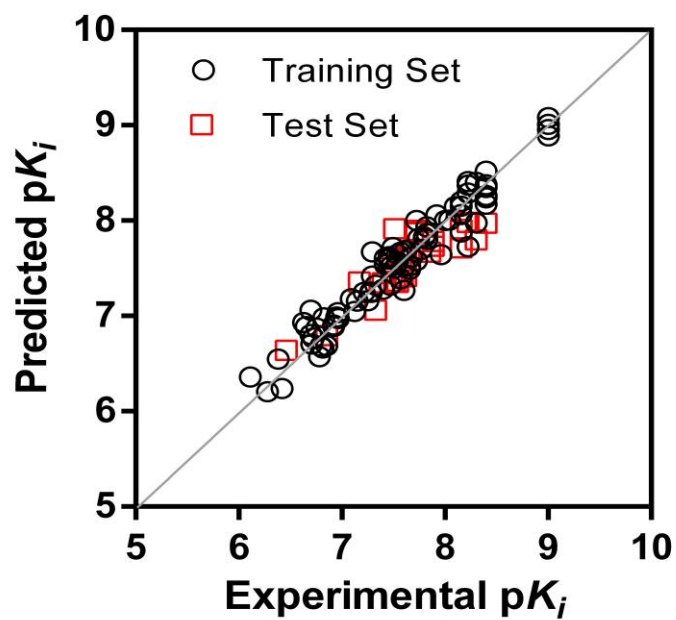


Figura 20: Gráfica de regresión del modelo CoMFA.

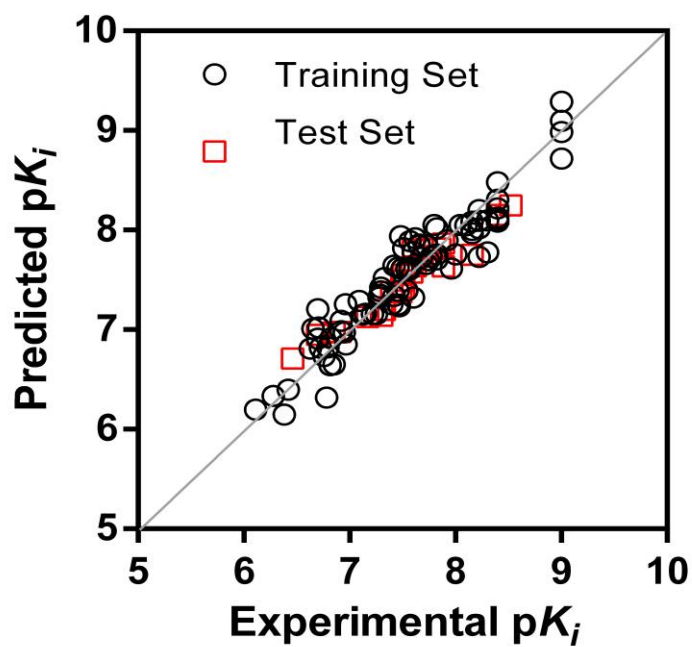


Figura 21: Gráfica de regresión del modelo CoMSIA.

En la **Figura 22** y **Figura 23** se aprecian los gráficos de residuales para los modelos CoMFA y CoMSIA respectivamente. En la **Figura 22**, se pueden ver 59 compuestos con residuales positivos, mientras que 50 compuestos presentan valores residuales negativos. El rango de residuales fue de -0.41 a 0.50 para el modelo CoMFA. Por otro lado, en la **Figura 23**, se pueden apreciar 55 compuestos con residuales positivos, mientras que también presenta 54 compuestos con valores residuales negativos. El rango de residuales fue de -0.51 a 0.52 para el modelo CoMSIA.

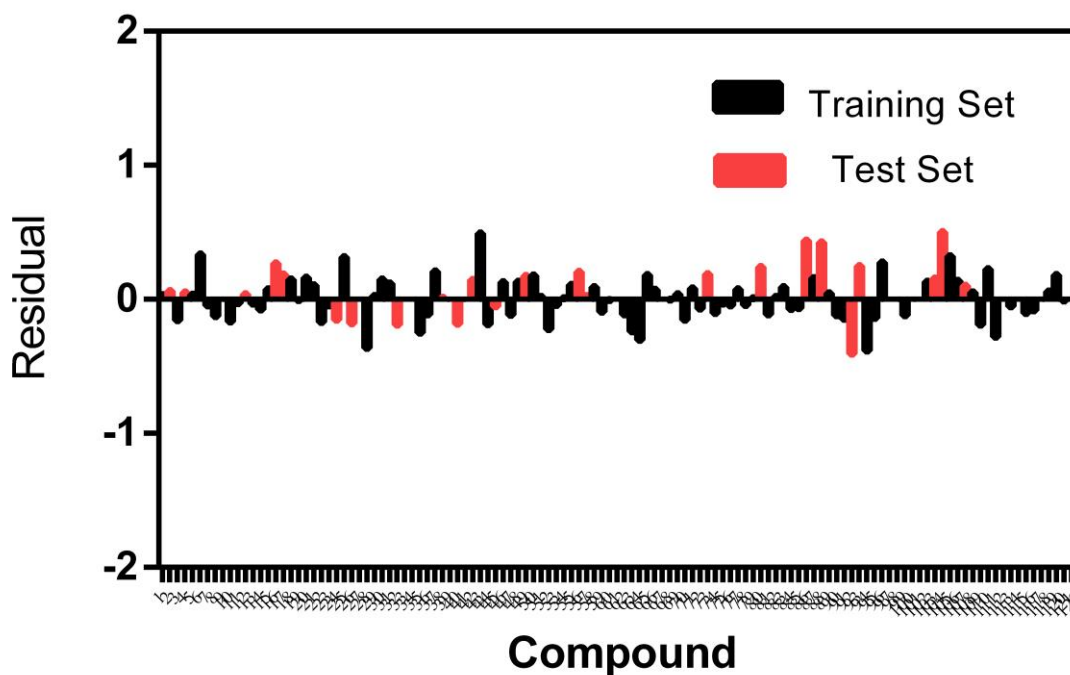


Figura 22: Gráfico de residuales del modelo CoMFA.

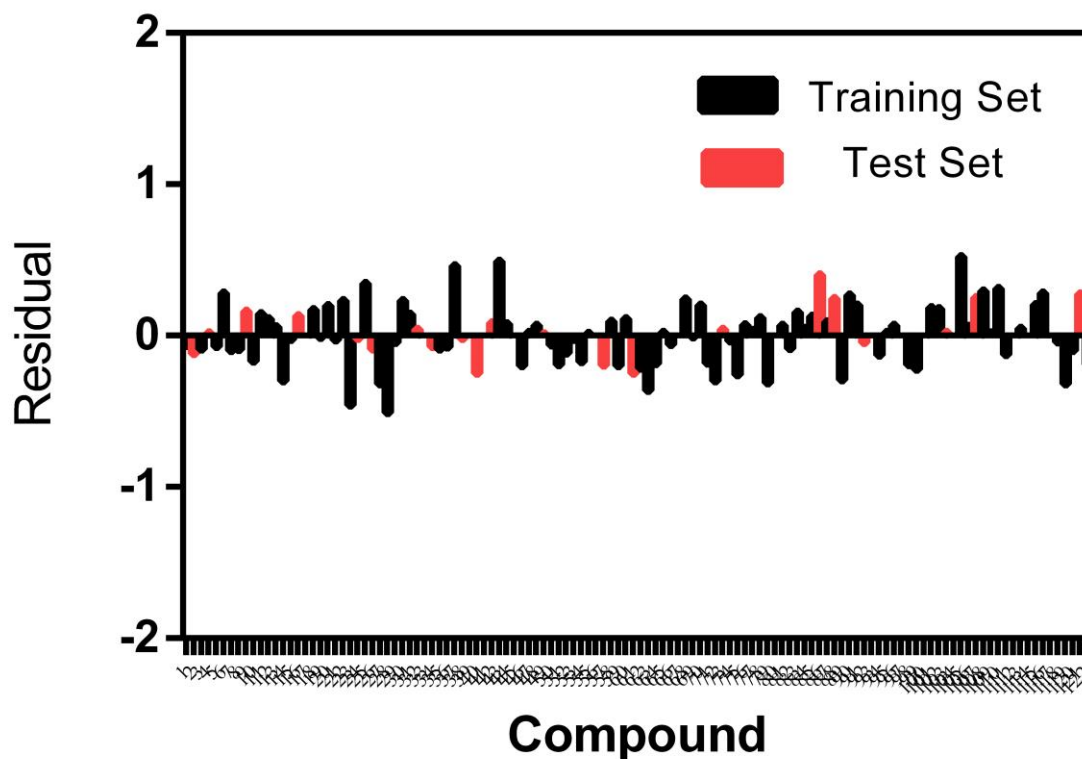


Figura 23: Gráfico de residuales del modelo CoMSIA.

En las **Figuras 24** y **25**, los modelos CoMFA-SE y CoMSIA-SHDA muestran una capacidad predictiva que considera la relación de valores de actividad biológica experimentales contra los valores de actividad predicha, la cual es satisfactoria a lo largo de todo el conjunto de datos para las moléculas menos activas (**35**, **42** y **69**) y las más activas (**59**, **98**, **126** y **130**).

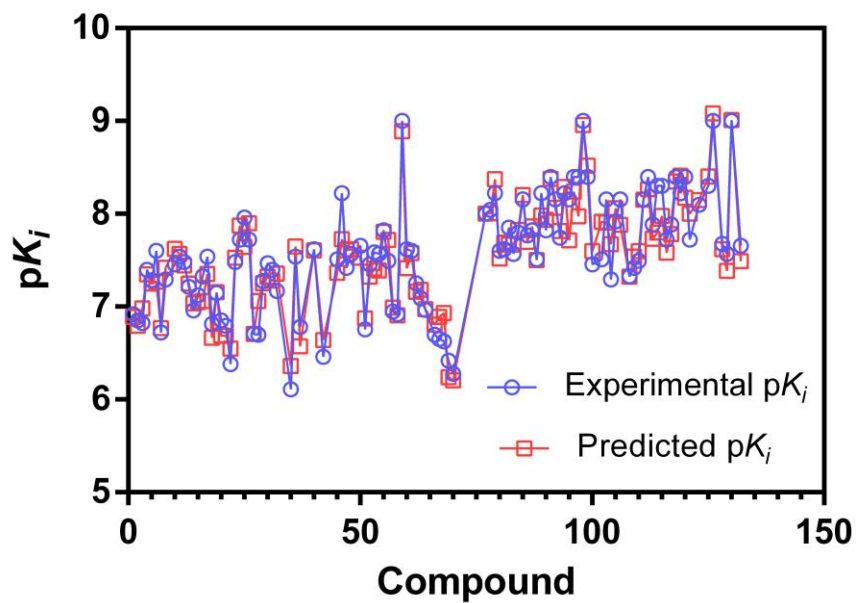


Figura 24: Gráfico relación actividad experimental versus actividad predicha del modelo CoMFA.

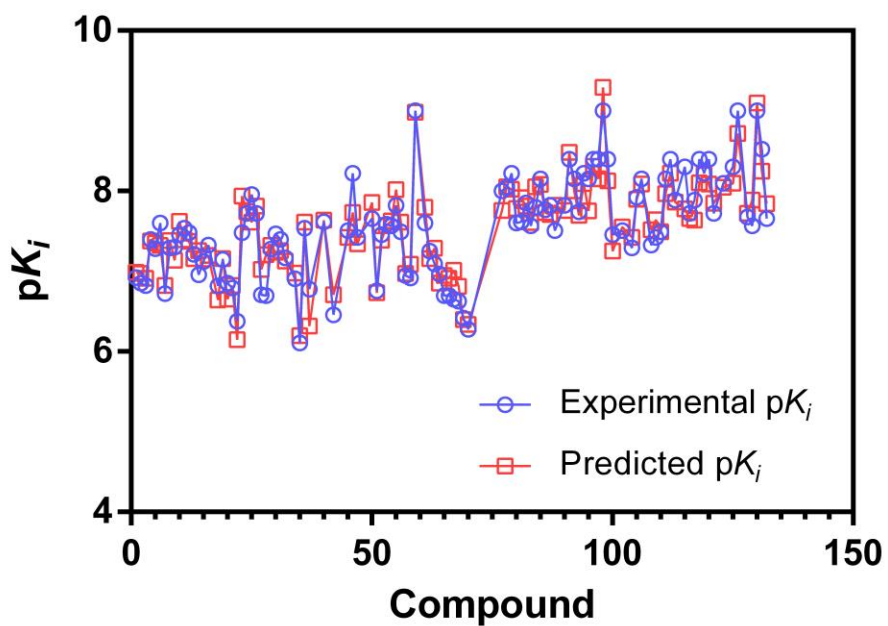


Figura 25: Gráfico relación actividad experimental versus actividad predicha del modelo CoMSIA.

Además, para evaluar la solidez del modelo, se aplicó la prueba de aleatorización de Y^{107} . La variable dependiente (actividad biológica) se barajó aleatoriamente y se desarrolló un nuevo modelo QSAR utilizando la matriz original de variables independientes. Si después de múltiples aleatorizaciones los nuevos valores de q^2 y r^2_{ncv} son negativos o están por debajo del límite de aceptabilidad ($q^2 < 0.5$ y $r^2_{ncv} < 0.6$), entonces se corrobora que los resultados obtenidos en la formulación de los modelos finales no son fruto del azar. En este caso, los nuevos modelos QSAR (después de muchas repeticiones) tienen valores q^2 y r^2_{ncv} bajos (**Tabla 5**).

Tabla 5: Valores de q^2 y r^2_{ncv} después de varias pruebas de la aleatorización-Y.

	CoMFA		CoMSIA	
	q^2	r^2_{ncv}	q^2	r^2_{ncv}
random 1	-0.300	0.198	-0.026	0.110
random 2	-0.308	0.233	-0.043	0.129
random 3	-0.323	0.148	-0.128	0.081
random 4	-0.258	0.143	-0.296	0.095
random 5	-0.110	0.144	-0.012	0.103
random 6	-0.059	0.160	-0.115	0.096
random 7	-0.546	0.229	-0.206	0.136
random 8	-0.058	0.199	-0.007	0.154
random 9	-0.029	0.185	-0.189	0.159
random 10	-0.029	0.185	-0.189	0.159

En síntesis, se seleccionaron los mejores modelos de CoMFA y CoMSIA sobre la base de su solidez estadística y su buena predictibilidad externa validada. En el caso del CoMFA, ambos potenciales contribuyen igualmente a la actividad biológica (49.7% para el campo estérico y 50.3% para el campo electrostático). En el caso CoMSIA hay una distribución que va del 19.5% (campo estérico) al 32.1% (campo aceptor de puentes de hidrógeno).

5.2. Análisis de mapas de contornos

5.2.1. CoMFA

Para la discusión de los mapas de contornos se utilizó el esquema de la **Figura 26**.

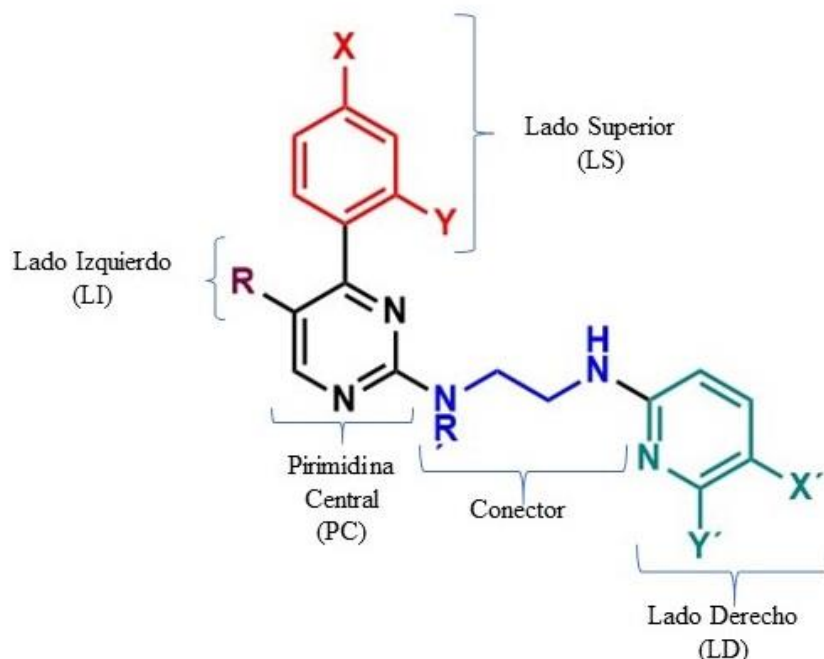


Figura 26: Esquema general de las moléculas utilizadas para el estudio QSAR y en la discusión de los mapas de contornos.

En el mapa de contornos estérico (**Figura 27A y B**) se aprecian dos poliedros verdes en la posición *orto*- y *para*- del benceno unido a la pirimidina (LS); un poliedro amarillo y otro verde cerca del sustituyente de la posición 3- de la pirimidina (LI), y un gran poliedro amarillo en la zona del grupo nitro del anillo piridínico (LD). Esto quiere decir que, en las zonas de los poliedros verdes, es favorable el adicionar sustituyentes voluminosos para incrementar a la actividad biológica. Mientras que, en las zonas de los poliedros amarillos, para mejorar la actividad antidiabética, es favorable la utilización de sustituyentes poco

voluminosos. En la **Figura 26A** se presenta la molécula **130** ($pIC_{50} = 9.00$) versus la molécula **41** en la **Figura 26B** ($pIC_{50} = 6.05$). La molécula **130** presenta dos átomos de cloro en la zona LS del compuesto, y un anillo de benceno unido a un imidazol en la zona LI, los cuales están en contacto con los poliedros verdes del mapa de contornos estérico; por lo que esto favorece a la actividad biológica de dicho compuesto. Por otro lado, la molécula **41** no presenta sustituyentes en las zonas de los poliedros verdes, excepto por un OCH_3 - en la posición *orto*- del benceno en LS, por lo que la ausencia de sustituyentes en las zonas de los poliedros verdes afectaría al valor de la actividad biológica. A modo de ejemplo, la molécula **46** ($pIC_{50} = 8.22$) presenta un cloro en la posición *orto*- del benceno unido a la pirimidina, y otro cloro en la posición *para*- del mismo sustituyente. Una propuesta para mejorar su actividad biológica es reemplazar estos sustituyentes por Br- o I- ya que, aun siendo halógenos, presentan un radio atómico mayor que el cloro y, por tanto, son más voluminosos.

Por otra parte, esta misma molécula, presenta un sustituyente imidazol en la zona LI. Cerca de este sustituyente hay un poliedro verde, por lo que otra propuesta aceptable para mejorar su actividad antidiabética es la adición de otro sustituyente como un *ter*-butil, isopropil, o un grupo fenil.

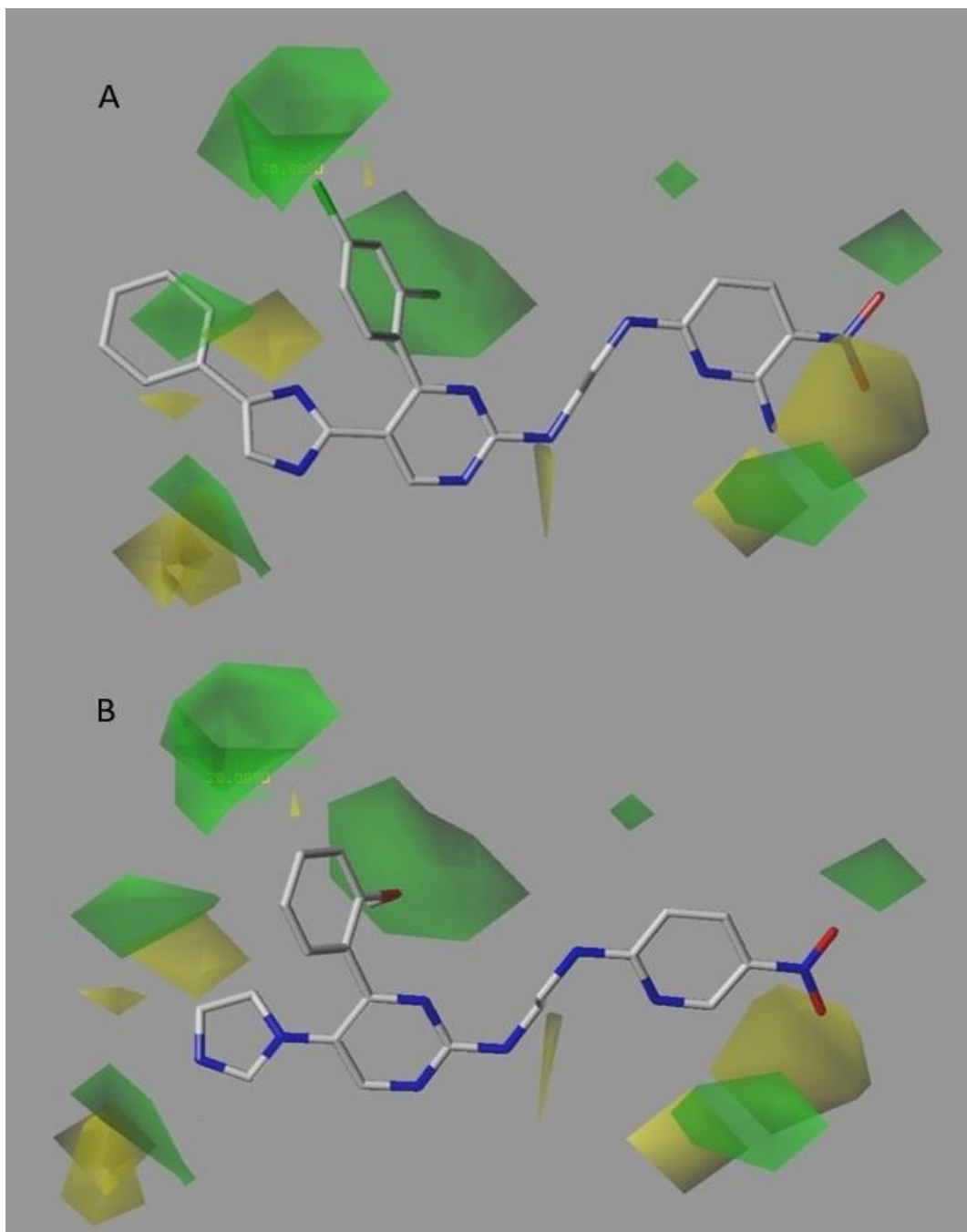


Figura 27: Mapa de contornos estéricos del modelo CoMFA. Los poliedros verdes indican que es favorable para la actividad biológica la presencia de compuestos voluminosos, mientras que los poliedros amarillos indican que no es favorable el uso de sustituyentes voluminosos. En A) está la molécula más activa (**130**) y en B) la menos activa (**41**).

En la **Figura 28A** y **B** se presenta el mapa de contornos electrostático del modelo CoMFA-SE, en el cual se aprecia un gran poliedro azul en la zona de la posición 3- del anillo pirimidínico (LI) y otro en la posición 2- en el anillo piridínico (LD). Esto quiere decir que en estas zonas es favorable el uso de sustituyentes electropositivos para potenciar su actividad antidiabética. A su vez, el mapa de contornos presenta una cierta cantidad de poliedros rojos siendo el más notorio el ubicado en la proximidad de los sustituyentes de la posición 3- del anillo pirimidínico (LI). Estos poliedros indican que es favorable para la actividad antidiabética la presencia de átomos electronegativos. En la **Figura 28A** se presenta la molécula **130** y la **Figura 28B** presenta la molécula **41**. En este caso, la molécula **130** presenta un grupo NH₂- en la posición 2- del anillo piridínico (LD). Este grupo amino se encuentra junto a un poliedro rojo, lo cual significa que es favorable para la actividad antidiabética la presencia de este sustituyente electronegativo. Sin embargo, la molécula **41** no presenta sustituyentes en esa zona, lo que se traduce en un bajo valor de pIC₅₀. Por ejemplo, la molécula **125** (pIC₅₀ = 8.30) muestra un grupo metil unido a un imidazol en la zona LI, y un grupo CN- en el anillo piridínico (LD). El grupo metil toca al poliedro azul de la posición 3- (LI), y el átomo de nitrógeno presente en el grupo imidazol se encuentra junto al poliedro rojo descrito anteriormente. Ambos casos favorecen a la actividad biológica, por lo que una propuesta para potenciarla aún más es el cambio del metil por algo más electropositivo como un CH₃CH₂-.

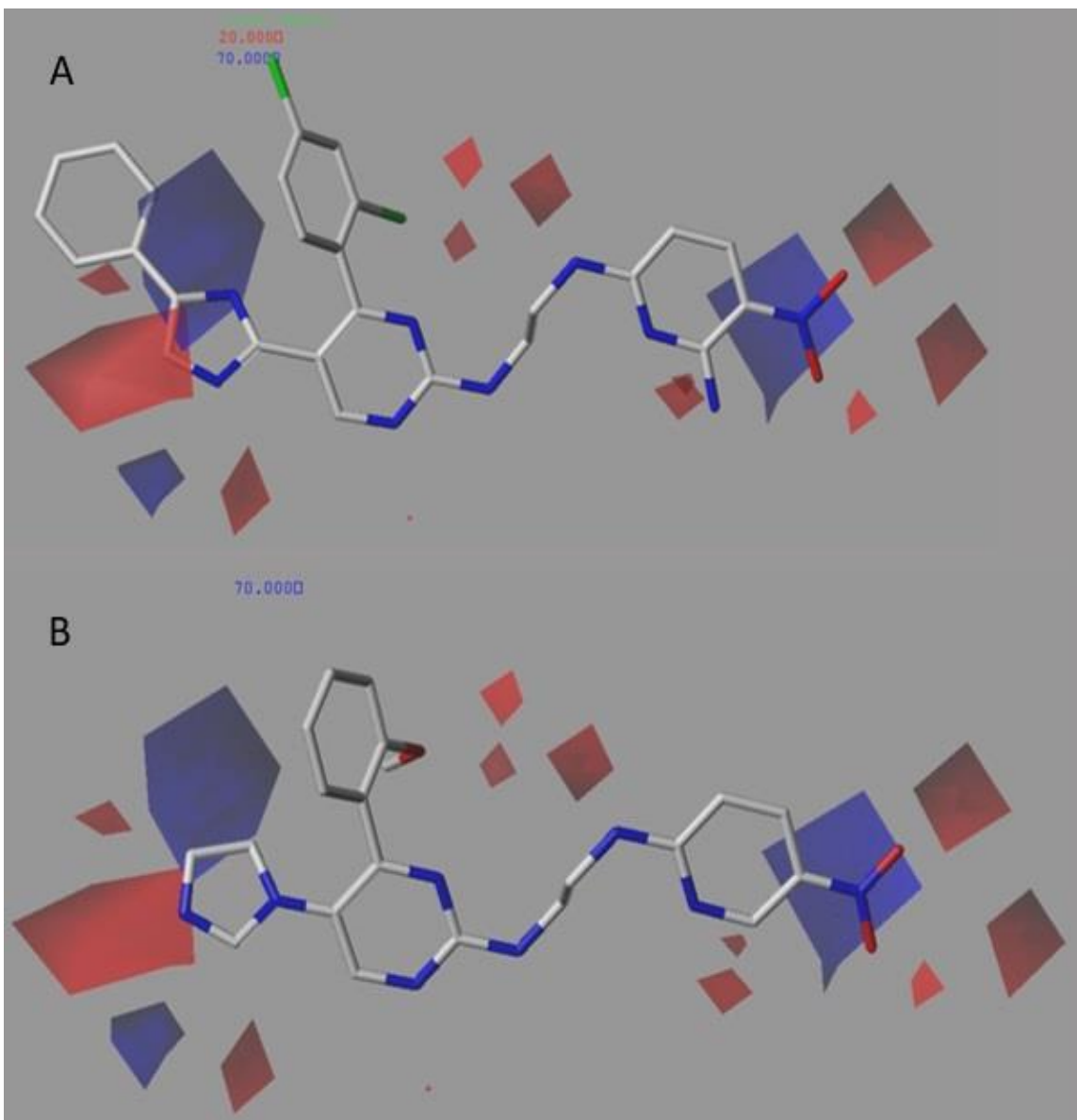


Figura 28: Mapa de contornos electrostáticos del modelo CoMFA. Los poliedros azules indican que es favorable para la actividad biológica la presencia de sustituyentes electropositivos, mientras que los poliedros rojos indican que es favorable la presencia de sustituyentes electronegativos. En A) está la molécula más activa (**130**), y en B) la menos activa (**41**).

Esta propuesta puede relacionarse con lo obtenido en el mapa de contorno estérico generando propuestas más interesantes por evaluar. También, el grupo CN-, que se ubica en la posición 2- del anillo piridínico (LD), puede ser reemplazado por un sustituyente electropositivo como un CH₂- unido a CN-, o bien cadenas propil, butil, isobutil, entre otros, que tengan deficiencia electrónica.

5.2.2. CoMSIA

En el mapa de contorno estérico (**Figura 29A y B**) se pueden apreciar dos poliedros verdes de gran tamaño ubicados en la posición *para*- del anillo benzoico de la zona LS, y el otro poliedro verde se encuentra en la posición 2- de la pirimidina (LI). Como también existe un poliedro verde en la posición *orto*- del mismo anillo benzoico. Esto quiere decir que agregar sustituyentes voluminosos en esas zonas favorecerán al aumento de la actividad antidiabética de las moléculas. También, está la presencia de un poliedro amarillo en la posición 2- del anillo piridínico (LD), lo que indica que no es favorable para la actividad biológica el adicionar sustituyentes voluminosos en esa zona. En la **Figura 29A**, la molécula **130** presenta un grupo fenil unido a un imidazol en la zona LI del compuesto, como también posee un átomo de cloro en la posición *orto*- y *para*- del anillo benzoico en LS. La presencia de estos sustituyentes en contacto con los poliedros verdes se traduce en un alto valor de actividad biológica, ya que el mapa de contornos indica que es favorable para dicho valor la presencia de sustituyentes voluminosos. Por otro lado, en la **Figura 29B**, la molécula **41** sólo presenta el grupo imidazol en la zona LI de la molécula, y un OCH₃- en la posición *orto*- del benceno en LS. Como la molécula **130** presenta una mayor cantidad de sustituyentes voluminosos que la molécula **41**, presenta un valor de actividad biológica mayor.

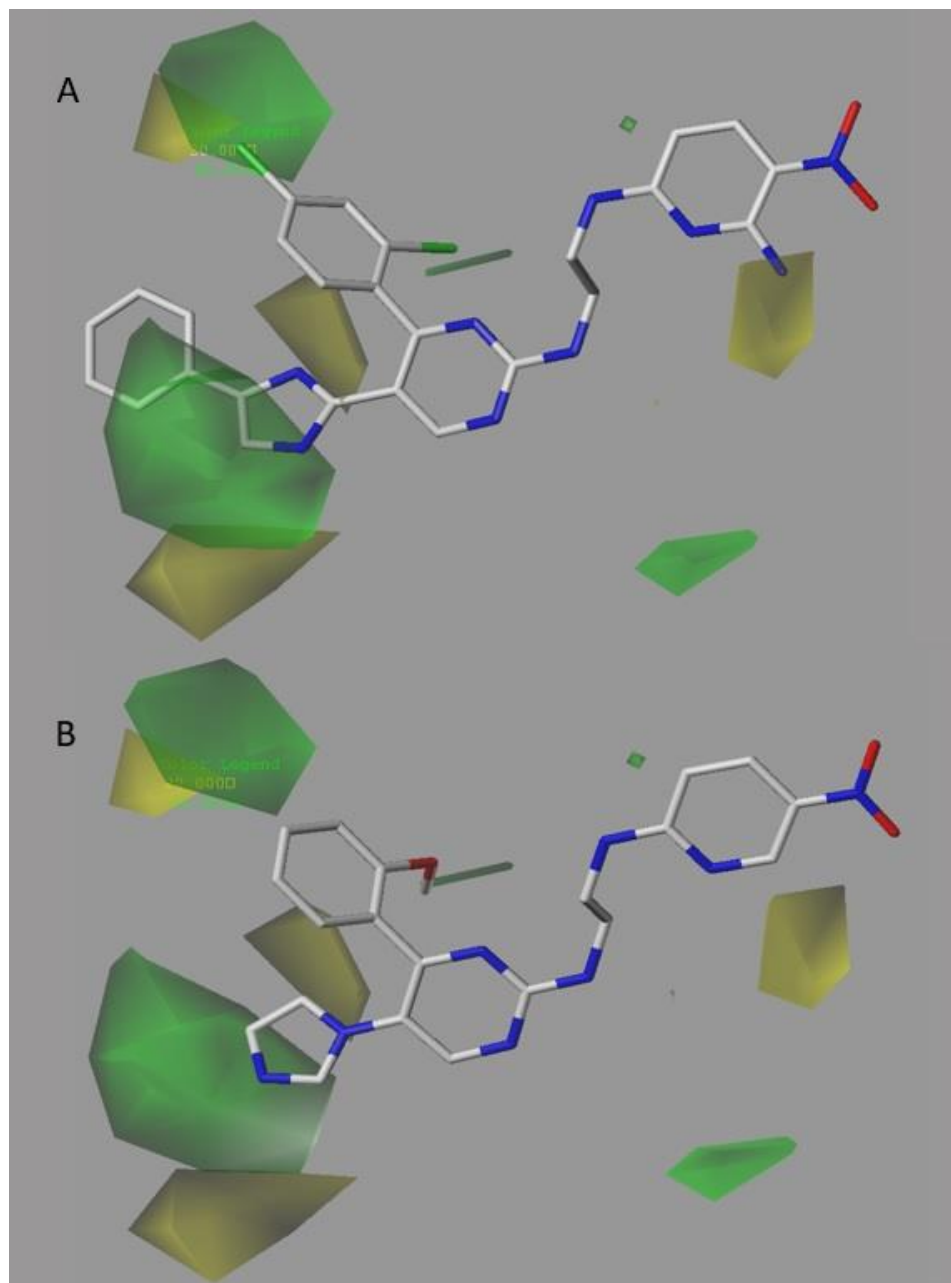


Figura 29: Mapa de contornos estéricos del modelo CoMSIA. Los poliedros verdes indican que es favorable para la actividad biológica la presencia de sustituyentes voluminosos, mientras que los poliedros amarillos indican que no es favorable el uso de sustituyentes voluminosos. En A) está la molécula más activa (**130**) y en B) la menos activa (**41**).

Por ejemplo, la molécula **77** ($pIC_{50} = 8.00$) presenta 2 átomos de cloro en las posiciones *orto*- y *para*- del benceno (LS). Para mejorar la actividad biológica una propuesta aceptable es el utilizar otros halógenos voluminosos como Br- o I-. Esto se complementa con el mapa de contorno estérico obtenido en el modelo CoMFA. A su vez, otra propuesta posible, es el adicionar sustituyentes más voluminosos en la posición 2- del anillo pirimidínico. Por ejemplo, un grupo fenil, bifenil, fenil-propano, o fenil-butano. Finalmente, como se aprecia un poliedro amarillo en la posición 2- de la piridina (LD), una posible propuesta para mejorar la actividad biológica del compuesto **77** es reemplazar el grupo NH₂- del anillo piridínico por un átomo de hidrógeno.

En el mapa de contorno hidrofóbico (**Figura 30A y B**) se pueden apreciar dos poliedros grises: uno en la zona LD, y otro entre el benceno (LS) y la posición 2- del anillo pirimidínico (LI). A su vez, existen dos poliedros amarillos: uno ubicado en la posición 2- del anillo piridínico (LD), y el otro en la posición *orto*- del anillo benzoico de la zona LS. Estos poliedros quieren decir que, en las zonas grises, es favorable para la actividad antidiabética el uso de sustituyentes hidrofílicos, mientras que en las zonas de los poliedros amarillos es beneficioso para la actividad biológica utilizar sustituyentes hidrofóbicos. En la **Figura 30A** se puede apreciar que el grupo imidazol de la molécula **130** en LI presenta un nitrógeno en la zona del poliedro gris, mientras que, en la **Figura 30B**, la molécula **41** tiene dos átomos de carbono en esa misma posición correspondientes a su grupo imidazol. Esta diferencia es clave para la explicación de sus valores de actividad biológica, ya que el átomo de nitrógeno tiene un comportamiento hidrofílico, y los átomos de carbono presentan un comportamiento lipofílico.

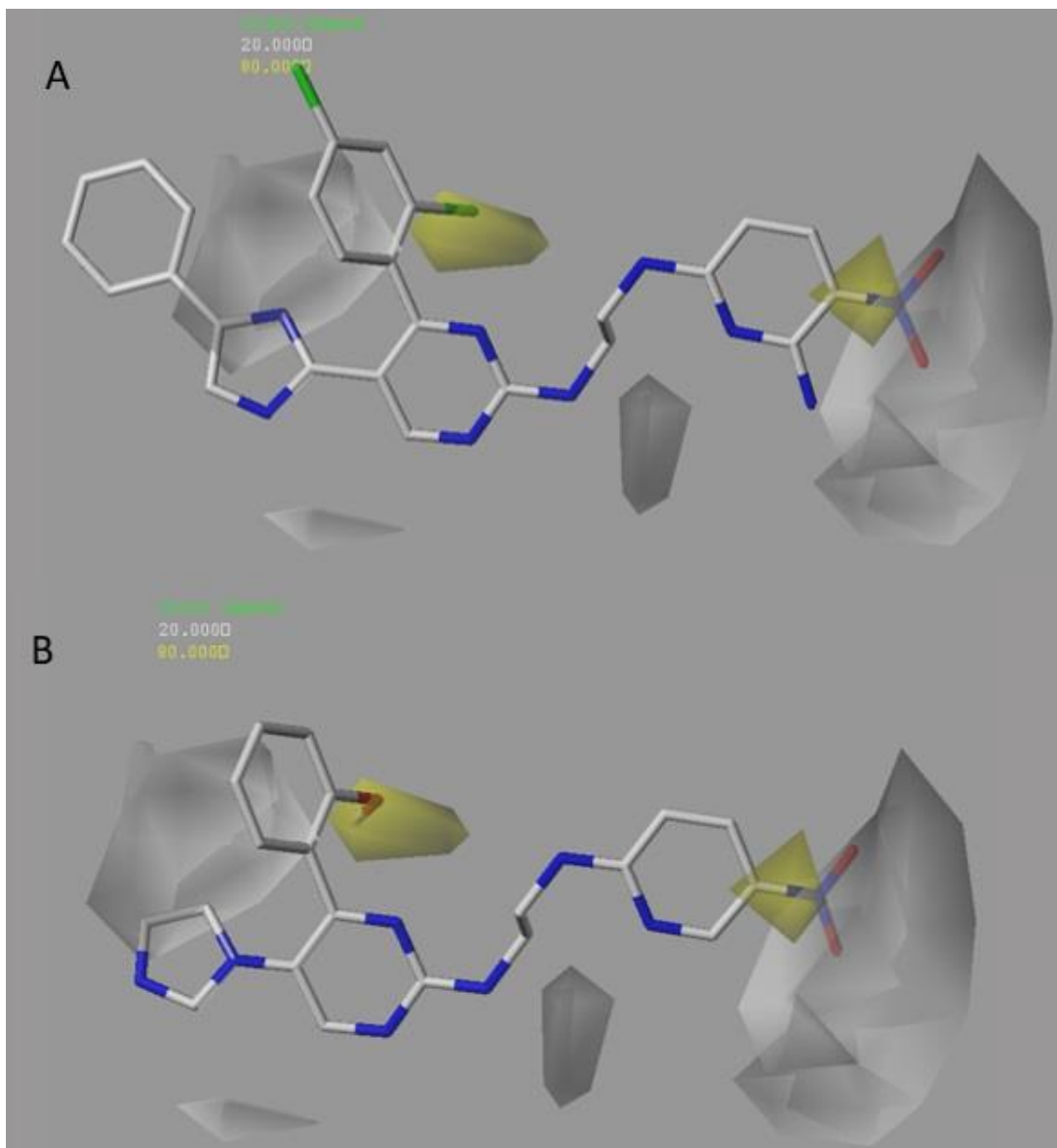


Figura 30: Mapas de contornos hidrofóbico. El poliedro amarillo indica que es favorable para la actividad biológica el uso de sustituyentes hidrofóbicos, mientras que el poliedro gris indica que es favorable el uso de sustituyentes hidrofílicos. En A) está la molécula más activa (**130**) y en B) la menos activa (**41**).

A modo de ejemplo, la molécula **99** ($pIC_{50} = 8.40$) tiene un grupo imidazol en la posición 2- del anillo pirimidínico (LI), por lo que una posible propuesta para mejorar su actividad biológica es la adición de sustituyentes hidrocarbonados como un CH_3CH_2- , *sec*-butil o *ter*-butil. También, posee un grupo nitrilo en la posición 2- del anillo piridínico, por lo que, considerando que existe un poliedro amarillo antes del poliedro gris, una propuesta es la utilización de grupos amino como $NHCH_2CH_2CH_3$, o un N-butilamino, o un N-*sec*-butil-amino, como algún carbonilo como $COCH_2CH_3$, entre otros. Y, por último, esta molécula posee un cloro en la posición orto- y para- del benceno unido a la pirimidina. Por lo tanto, no sería favorable que uno de estos cloros que se encuentra en la posición del poliedro amarillo sea reemplazado por un grupo OH- o un $-CH_2OH$, sino que deben insertarse átomos o grupos que incrementen la lipofilia en esa posición. Por ejemplo: Br-, I-, NO_2- , y cadenas o ciclos alquílicos o aromáticos.

En el mapa de contornos dador de puentes de hidrógeno (**Figura 31A y B**) se puede apreciar un gran poliedro cian y tres poliedros purpura entre la zona del conector y el lado derecho (LD) de los compuestos. Esto quiere decir que en la zona del poliedro cian es favorable la presencia de compuestos capaces de donar puentes de hidrógeno, mientras que los poliedros purpura indican que no es favorable para la actividad biológica el utilizar sustituyentes con capacidad de donar puentes de hidrógeno. En la **Figura 31A** se presenta la molécula **130** como la más activa versus la molécula **41** (**Figura 31B**) como el compuesto menos activo. En el caso de la molécula **130**, la presencia de su grupo NH_2- podría ser el factor que marque la diferencia en estas moléculas para el caso de este mapa de contornos.

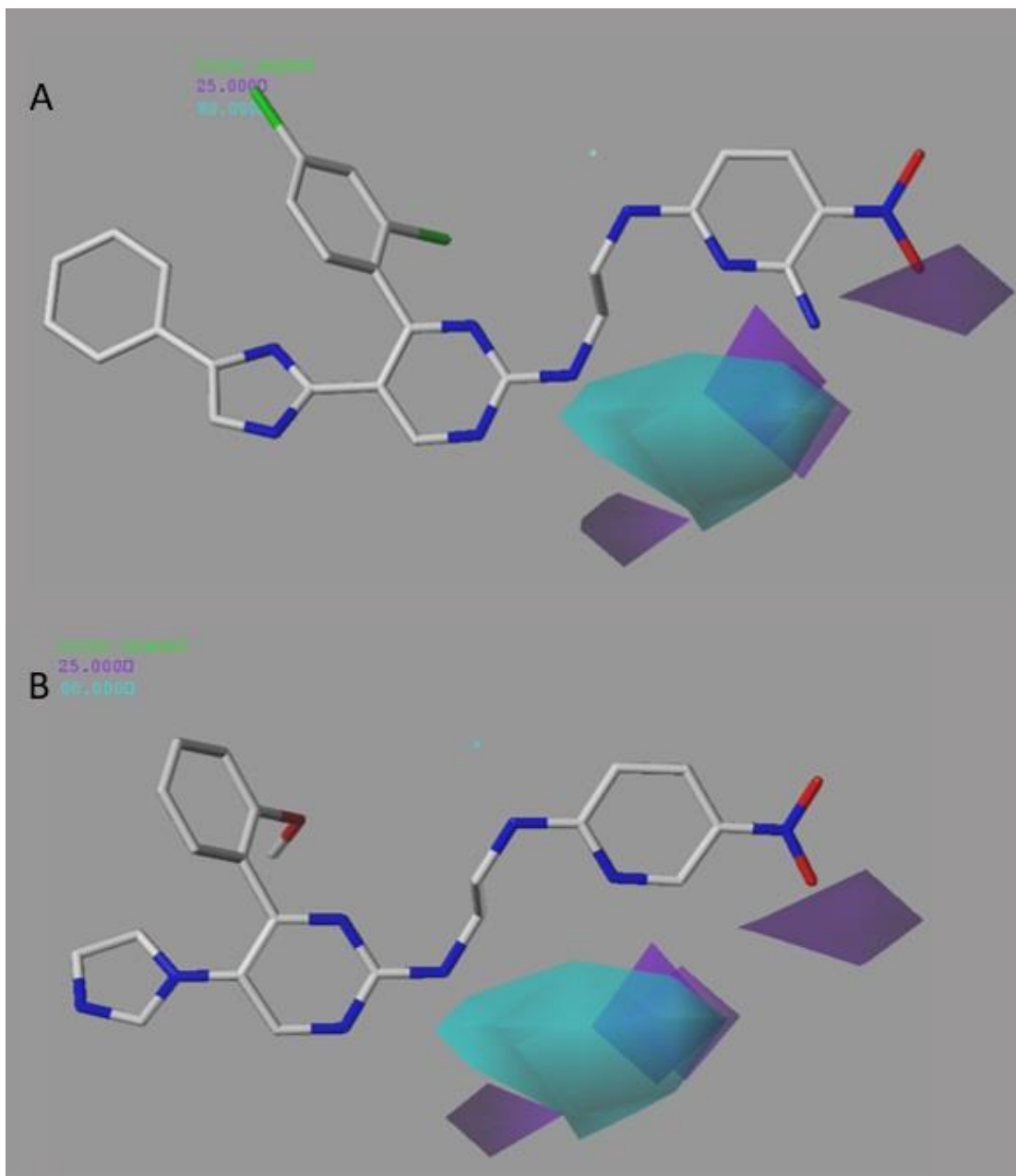


Figura 31: Mapas de contornos dador de puentes de hidrógeno. El poliedro cian indica que es favorable para la actividad biológica el uso de sustituyentes dadores de puentes de hidrógeno, mientras que el poliedro púrpura indica que no es favorable el uso de sustituyentes dadores de puentes de hidrógeno. En A) está la molécula más activa (**130**) y en B) la menos activa (**41**).

Esto se debe a que el átomo de hidrógeno del grupo amino entra en contacto con el poliedro cian. La interacción intermolecular denominada “puente de hidrógeno” se forma cuando un átomo de hidrógeno está unido a un átomo muy electronegativo como un átomo de oxígeno o como un átomo de nitrógeno, y este interactúa con otro de estos átomos electronegativos generando la interacción de Van der Waals más resistente¹⁰⁸. Por lo tanto, el átomo de hidrógeno del grupo amino de la molécula **130** es capaz de generar puentes de hidrógeno.

Por ejemplo, la molécula **52** ($pIC_{50} = 7.46$) presenta una cadena alquílica en el grupo amino más próximo al anillo pirimidínico. Este pasa por el poliedro cian y culmina en un poliedro púrpura. En el poliedro púrpura tiene un $CH_3NCH_3^-$, el que se protona a pH fisiológico, por lo tanto, sería conveniente la eliminación del átomo de N en esta cadena. Además, considerando el poliedro cian, una propuesta interesante para potenciar su actividad biológica en función de lo obtenido en este mapa de contorno es la adición de un sustituyente que tenga la capacidad dadora como algún alcohol como un OH^- o bien CH_2OH^- en la cercanía al N de la cadena unida a la pirimidina (conector).

Finalmente, en el mapa de contorno aceptor de puentes de hidrógeno (**Figura 32A y B**), se observa un gran poliedro rojo en la posición 2- del anillo pirimidínico (LI), como otro poliedro rojo en la cercanía del anillo piridínico (LD), y la presencia de poliedros magenta en las posiciones *para*- del anillo benzoico en la zona LS, mientras que el otro poliedro se encuentra en la posición 2- del anillo piridínico (LD). Esto quiere decir que, para mejorar su actividad antidiabética, en las zonas de los poliedros magenta es favorable la presencia de compuestos capaces de aceptar puentes de hidrógeno, mientras que en las ubicaciones de los

poliedros rojos no es favorable la presencia de grupos aceptores de puentes de hidrógeno. En la **Figura 32A** se presenta a la molécula **130** como la más activa y la molécula **41** en la **Figura 32B** como la menos activa. Para este caso, el grupo amino de la molécula **130**, no es favorable que esté en esa posición ya que se encuentra cerca de un poliedro rojo, el cual indica que no es favorable la presencia de compuestos capaces de aceptar puentes de hidrógeno. Sin embargo, el grupo fenil del imidazol presente en la zona LI entra en contacto con el poliedro rojo, por lo que le daría un punto a favor con respecto a la molécula **41**. La molécula **41** no presenta un grupo amino en su anillo piridínico, pero sí presenta un grupo éter en la posición *orto*-del benceno en la zona LS. Este grupo, a pesar de ser un aceptor de puentes de hidrógeno, se encuentra más próximo al poliedro rojo que al poliedro magenta por lo que un éter en esa posición no es favorable para este compuesto.

Por ejemplo, en la molécula **59** ($pIC_{50} = 9.00$) se encuentra un grupo NH_2 - cerca del poliedro rojo, por lo que una propuesta para mejorar la actividad de esta molécula sea utilizar sustituyentes que no tengan la capacidad de aceptar puentes de hidrógeno como un CH_3 , CH_2CH_3 , o bien un *ter*-butil. Por otro lado, esta misma molécula presenta un grupo imidazol en la posición 2- del anillo pirimidínico, por lo que otra propuesta aceptable para mejorar la actividad biológica es reemplazar este sustituyente con grupos fenil. Aunque, otra propuesta que sería bueno explorar es que, en vez de reemplazarlo, adicionarle al imidazol sustituyentes como metil o etil. Finalmente, esta molécula presenta un Cl- en la posición *para*- del anillo benzoico y, considerando que el mapa propone utilizar sustituyentes que puedan aceptar puentes de hidrógeno, puede ser un átomo de flúor.

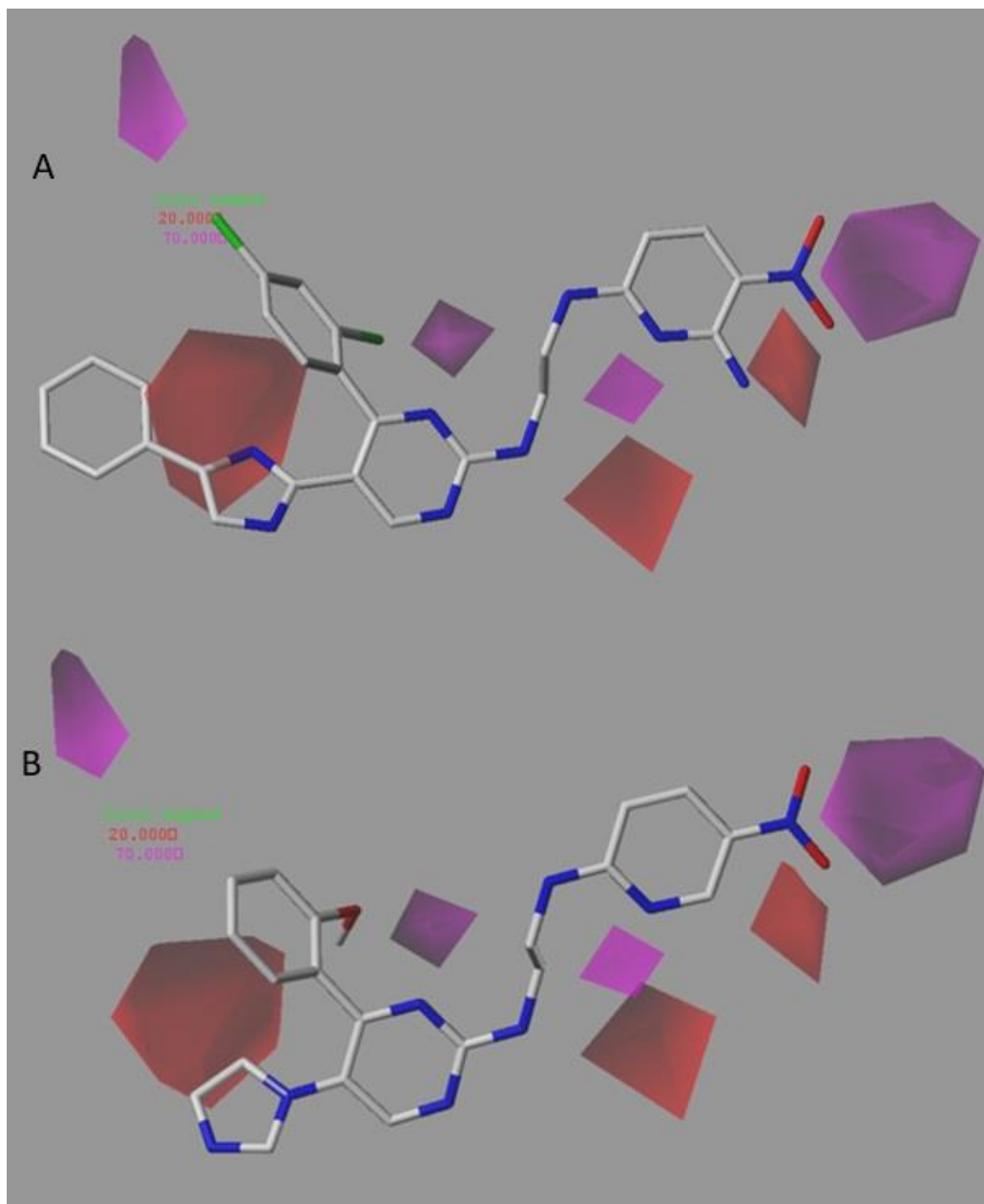


Figura 32: Mapas de contornos aceptor de puentes de hidrógeno. El poliedro magenta indica que es favorable para la actividad biológica el uso de sustituyentes aceptores de puentes de hidrógeno, mientras que el poliedro rojo indica que no es favorable la presencia de sustituyentes aceptores de puentes de hidrógeno. En A) está la molécula más activa (**130**) y en B) la menos activa (**41**).

5.3. Valores atípicos (“Outliers”)

En el modelo CoMFA-SE, las moléculas **9, 33, 34, 41, 44, 48, 65, 71, 107, 122, 124, 127, 131** fueron valores “Outliers”. La molécula **44** posee un átomo de cloro en la posición meta- del anillo benzoico (LS), en el cual se encuentra un poliedro amarillo y, el átomo de cloro, en comparación a un átomo de hidrógeno, es mucho más voluminoso, por lo que presentó valores residuales altos y arruinaba la linealidad predictiva del modelo. La molécula **48** presenta un grupo metil en la posición de un poliedro rojo en el mapa de contorno electrostático, por lo que, aunque no presente un bajo valor de pIC_{50} , genera valores residuales que también afectaban al análisis. Las moléculas **122, 124, y 127** presentaron valores residuales muy altos que, al momento de predecir, afectaron la linealidad de los resultados. Estas moléculas son muy semejantes entre sí, con pequeñas diferencias en sus sustituyentes. La molécula **127**, por ejemplo, presenta un CF_3 - en la cercanía de un poliedro azul del mapa de contorno electrostático. El átomo de flúor es altamente electronegativo, por lo que induce a que la nube electrónica del carbono sea desplazada hacia los F- a los que está enlazado. Por lo tanto, existe una gran densidad electrónica en la ubicación del poliedro azul, cuando en esa zona se recomienda la presencia de átomos electropositivos. Sin embargo, en el mapa de contorno estérico, tanto las moléculas **122, 124, y 127** presentan buenos resultados, ya que contienen sustituyentes voluminosos en las zonas de los poliedros verdes, pero al momento de analizar la contribución electrostática los resultados cambian y generan residuales altos, como el caso de la molécula **127**.

En términos generales estos compuestos presentan sustituyentes que aportan bastante lipofilia a la estructura, por lo que la predictibilidad dispar para las moléculas podría deberse a razones farmacocinéticas.

En el modelo CoMSIA-SHDA, las moléculas **17, 33, 41, 44, 48, 60, 71, 103, 107, 114, 122, 124, y 127** fueron “Outliers”. Como observación, la molécula **33** presenta un grupo etil en su anillo benzoico en la zona LS. Este grupo etilo está próximo a un pequeño poliedro amarillo que se observa en el mapa de contornos estéricos CoMSIA (Figura 28 A y B). Generando valores predictivos distintos a los experimentales y, por lo tanto, arruinando la linealidad de los valores experimentales versus valores predichos. La molécula **107**, según el mapa de contornos aceptor de puentes de hidrógeno (Figura 31 A y B), presenta dos sustituyentes capaces de aceptar puentes de hidrógenos en zonas donde se encuentran poliedros rojos, una en la zona LI donde presenta un grupo imidazol con un nitrógeno en el poliedro rojo, y un grupo amino en la zona LD en la posición 2- del anillo piridínico con su átomo de nitrógeno en la posición del otro poliedro rojo. Esto nos indica que no es favorable para la actividad biológica la presencia de dichos compuestos aceptores de puentes de hidrógeno. Generando valores predictivos distintos a los experimentales y, como consecuencia, valores residuales altos. A grandes rasgos, estas moléculas aumentan la lipofilia a los compuestos, por lo que la razón de obtener distintos valores predictivos se deba a parámetros farmacocinéticos.

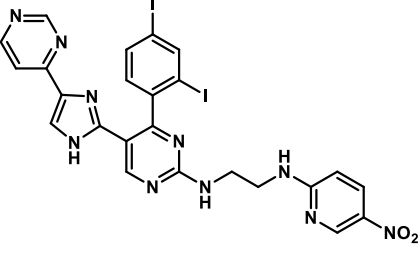
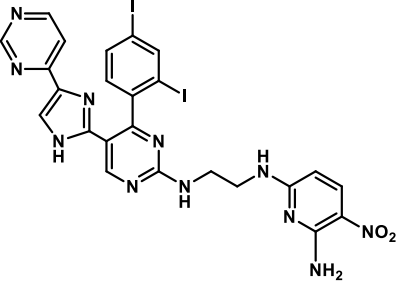
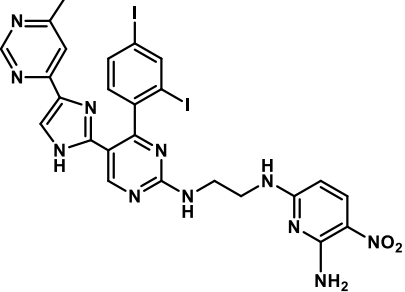
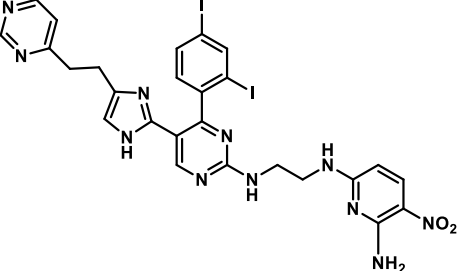
5.4. Dominio de Aplicabilidad

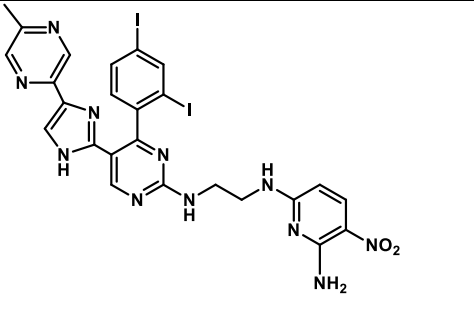
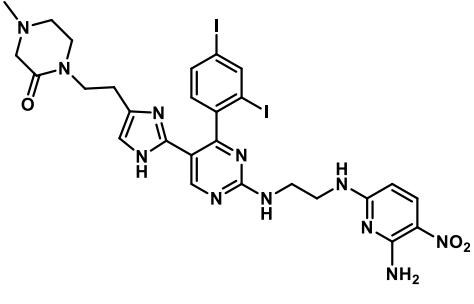
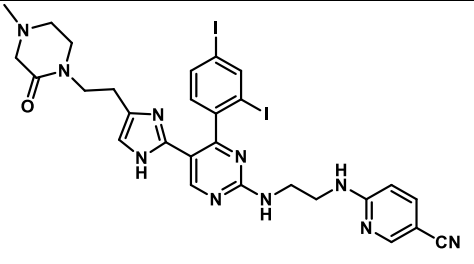
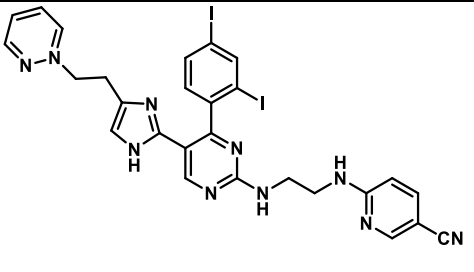
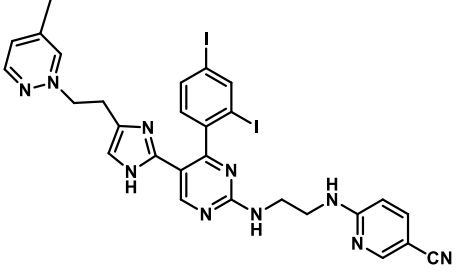
El dominio de aplicabilidad (AD) es una región teórica en el espacio químico que abarca tanto los descriptores del modelo como la respuesta modelada, lo que permite estimar la incertidumbre en la predicción de un compuesto basándose en cuán similar es a los compuestos de entrenamiento (Training Set) empleados en el desarrollo del modelo. En este trabajo, se utilizó el método desarrollado por Roy et al.¹⁰⁶ para la determinación del AD. Este método se basa en la teoría básica del enfoque de estandarización. El cálculo se realizó utilizando la aplicación gratuita disponible en la página del autor, tras lo cual se obtuvo que todos los compuestos estaban dentro del dominio de aplicación.

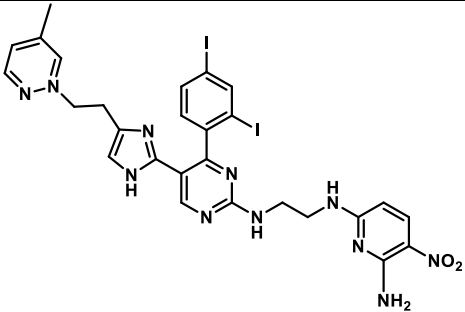
5.5. Predicciones

A partir de la información obtenida en los mapas de contornos del análisis CoMFA y CoMSIA, se logró la generación de nuevas moléculas que presentaron valores de actividad biológica mayores a 9.00. Considerando las propuestas individuales de cada mapa, no fue posible encontrar compuestos con actividad biológica alta, sin embargo, probando sustituyentes que cumplieran con más de una condición de las discutidas en los mapas de contornos, fue posible el hallazgo de nuevos compuestos. Si sólo se realizan combinaciones dándole importancia a las contribuciones estéricas, los compuestos tendrían sustituyentes que aumentarían la lipofilia, generando que sus valores de actividad biológica disminuyan en lugar de aumentar. Se ensayaron 118 combinaciones, de las cuales 10 obtuvieron valores de $pIC_{50} > 9.00$ para CoMFA, CoMSIA, o ambas (**Tabla 6**). Las combinaciones n° **72x**, **89x**, **91x**, **102x**, **110x**, y **113x** obtuvieron valores de $pIC_{50} > 9.00$ para CoMSIA, la **114x** y **115x** para CoMFA, y la **95x** y **116x** en CoMFA y CoMSIA.

Tabla 6: Estructuras de los compuestos predichos a partir de los resultados obtenidos en los mapas de contornos CoMFA y CoMSIA.

N° de Combinación	Moléculas	IC ₅₀ pred CoMFA	IC ₅₀ pred CoMSIA
72x		8.292	9.012
89x		8.474	9.076
91x		8.701	9.243
95x		9.271	9.310

102x		8.747	9.367
110x		8.562	9.025
113x		8.701	9.075
114x		9.192	8.979
115x		9.364	8.997

116x	 <p>The chemical structure of compound 116x is a complex molecule consisting of several fused and linked rings. It features a central pyrimidine ring system. Attached to this system are: a 2-methylimidazole ring via a methylene bridge; a 2,4-diiodophenyl ring; and a 2-amino-5-nitrophenyl ring via a propyl chain. The 2-amino-5-nitrophenyl ring has an amino group (NH₂) at the 2-position and a nitro group (NO₂) at the 5-position.</p>	9.011	9.193
-------------	---	-------	-------

6. CONCLUSIONES

Con el fin de generar nuevos derivados aminopirimidínicos con potencial inhibitorio de GSK-3 como alternativa para el tratamiento de la diabetes tipo II, se realizó un extenso estudio 3D-QSAR basado en CoMFA y CoMSIA, donde se buscó la combinación de Test set que arrojará resultados estadísticos óptimos para la predicción. Para CoMFA se obtuvo una secuencia con contribuciones estéricas y electrostáticas (CoMFA-SE) dando como resultado un q^2 de 0.563, un r^2_{ncv} de 0.941, y un r^2_{test} de 0.823. Mientras que para CoMSIA se obtuvo una secuencia con contribuciones estéricas, hidrofóbicas, dadora y aceptora de puentes de hidrógeno (CoMSIA-SHDA) dando como resultado un q^2 de 0.598, un r^2_{ncv} de 0.899, y un r^2_{test} de 0.925. Finalmente, en la **Figura 33** se presenta el SAR (relación estructura actividad) que resume lo obtenido en los mapas de contornos CoMFA y CoMSIA.

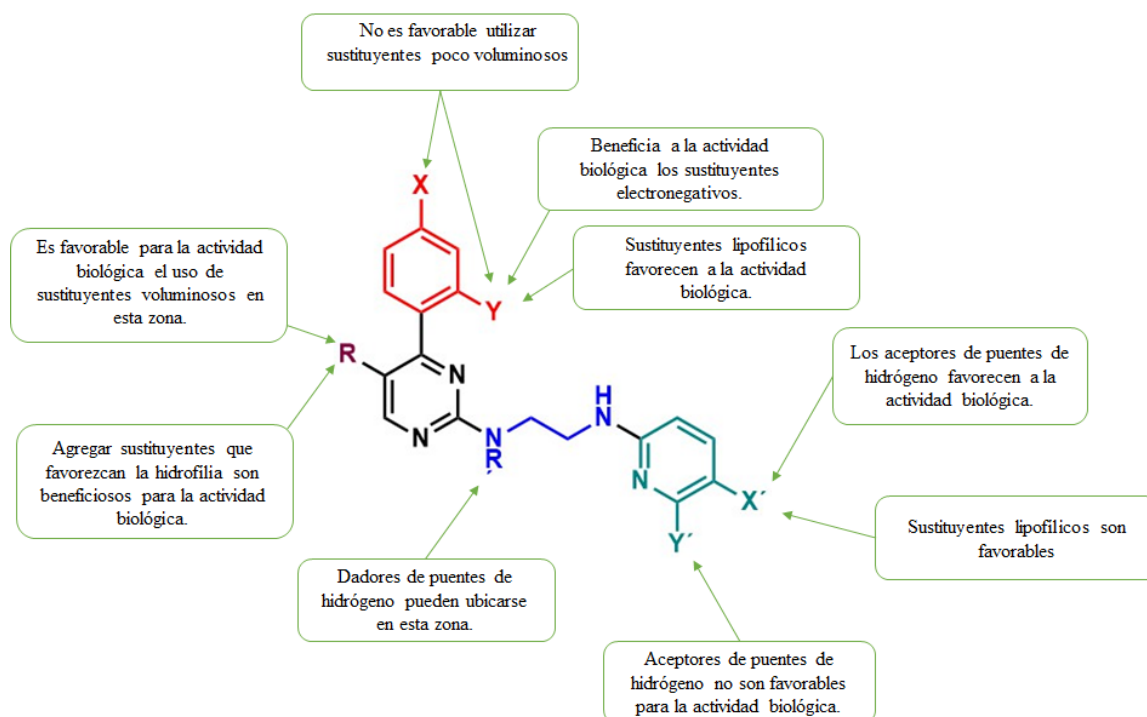


Figura 33: Principales relaciones estructura-actividad encontradas en este estudio.

7. PROYECCIONES

Realizar estudios de acoplamiento molecular inducido (Docking) con el fin de estudiar las interacciones intermoleculares que pueden presentar los compuestos al acoplarse a la glucógeno sintasa quinasa 3 y validar las propuestas del estudio QSAR.

Validar las características farmacocinéticas de los derivados propuestos en ensayos in silicio y luego in vitro.

Realizar la síntesis de los compuestos propuestos mediante los estudios CoMFA y CoMSIA.

Evaluar la actividad biológica de los compuestos sintetizados en modelos in vitro y en modelos animales que presenten síntomas de diabetes tipo II y en modelos con dieta controlada para comprobar la real utilidad de estos compuestos como posible tratamiento para esta enfermedad.

8. REFERENCIAS

1. Doble, B. W. GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116**, 1175–1186 (2003).
2. Kaidanovich-Beilin, O. & Woodgett, J. R. GSK-3: Functional Insights from Cell Biology and Animal Models. *Front. Mol. Neurosci.* **4**, 1–25 (2011).
3. Linding, R. *et al.* Systematic Discovery of In Vivo Phosphorylation Networks. *Cell* **129**, 1415–1426 (2007).
4. Sutherland, C. What are the bona fide GSK3 substrates? *Int. J. Alzheimers. Dis.* **2011**, (2011).
5. Rylatt, D. B.; Aitken, A.; Bilham, T.; Condon, G. D.; Embi, N. . C. P. Glycogen synthase from rabbit skeletal muscle. *Eur. J. Biochem.* **107**, 529–537 (1980).
6. Lehninger, N. Bioquímica. *Lehninger - Principios de Bioquímica 5ed* 184, 201, 202, 203 (2009).
7. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, (2011).
8. Woodgett, J. R. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J.* **9**, 2431–8 (1990).
9. Hoeflich, K. P. *et al.* Requirement for glycogen synthase kinase-3 β in cell survival and NF- κ B activation. *Nature* **406**, 86–90 (2000).
10. Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S. & Baltimore, D. Embryonic lethality and liver degeneration in mice lacking the κ B component of NF- κ B. *Nature* **376**, 167–170 (1995).

11. Woodgett, J. R. Physiological roles of glycogen synthase kinase-3: potential as a therapeutic target for diabetes and other disorders. *Curr Drug Targets Immune Endocr Metab. Disord* **3**, 281–290 (2003).
12. Álamo González, C. & López Muñoz, F. *Historia de la psicofarmacología*. (Médica Panamericana, 2007).
13. Beurel, E., Grieco, S. F. & Jope, R. S. Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. *Pharmacol. Ther.* **148**, 114–131 (2015).
14. King, T. D., Bijur, G. N. & Jope, R. S. Caspase-3 activation induced by inhibition of mitochondrial complex I is facilitated by glycogen synthase kinase-3 β and attenuated by lithium. *Brain Res.* **919**, 106–114 (2001).
15. Meares, G. P. & Jope, R. S. Resolution of the nuclear localization mechanism of glycogen synthase kinase-3: Functional effects in apoptosis. *J. Biol. Chem.* **282**, 16989–17001 (2007).
16. Bijur, G. N. & Jope, R. S. Proapoptotic Stimuli Induce Nuclear Accumulation of Glycogen Synthase Kinase-3*. (2001). doi:10.1074/jbc.M105725200
17. Yaffe, Y. *et al.* Distinct Molecular Regulation of Glycogen Synthase Kinase-3 Isozyme Controlled by Its N-terminal Region FUNCTIONAL ROLE IN CALCIUM/CALPAIN SIGNALING * □ S Inbar Azoulay-Alfaguter ‡1. (2011). doi:10.1074/jbc.M110.127969
18. Gupta, C., Kaur, J. & Tikoo, K. Regulation of MDA-MB-231 cell proliferation by GSK-3 β involves epigenetic modifications under high glucose conditions. *Exp. Cell Res.* **324**, 75–83 (2014).
19. Eickholt, B. J., Walsh, F. S. & Doherty, P. An inactive pool of GSK-3 at the leading

- edge of growth cones is implicated in Semaphorin 3A signaling. *J. Cell Biol.* **157**, 211–7 (2002).
20. Zhou, F.-Q., Zhou, J., Dedhar, S., Wu, Y.-H. & Snider, W. D. NGF-Induced Axon Growth Is Mediated by Localized Inactivation of GSK-3 β and Functions of the Microtubule Plus End Binding Protein APC. *Neuron* **42**, 897–912 (2004).
 21. Wood-Kaczmar, A., Kraus, M., Ishiguro, K., Philpott, K. L. & Gordon-Weeks, P. R. An alternatively spliced form of glycogen synthase kinase-3 β is targeted to growing neurites and growth cones. *Mol. Cell. Neurosci.* **42**, 184–194 (2009).
 22. Hernández, F. *et al.* The role of GSK3 in Alzheimer disease. *Brain Res. Bull.* **80**, 248–250 (2009).
 23. Magdesian, M. H. *et al.* Amyloid-beta binds to the extracellular cysteine-rich domain of Frizzled and inhibits Wnt/beta-catenin signaling. *J. Biol. Chem.* **283**, 9359–68 (2008).
 24. Takashima, A. *et al.* Presenilin 1 associates with glycogen synthase kinase-3 β and its substrate tau. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9637–41 (1998).
 25. Townsend, M., Mehta, T. & Selkoe, D. J. SOLUBLE A β INHIBITS SPECIFIC SIGNAL TRANSDUCTION CASCADES COMMON TO THE INSULIN RECEPTOR PATHWAY. (2007). doi:10.1074/jbc.M610390200
 26. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. beta-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.* **16**, 3797–804 (1997).
 27. Cross, D. A. E., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).

28. Marambaud, P. *et al.* A presenilin-1/gamma-secretase cleavage releases the E-cadherin intracellular domain and regulates disassembly of adherens junctions. *EMBO J.* **21**, 1948–1956 (2002).
29. Bryja, V. *et al.* The Extracellular Domain of Lrp5/6 Inhibits Noncanonical Wnt Signaling In Vivo. *Mol. Biol. Cell* **20**, 924–936 (2009).
30. Cselenyi, C. S. *et al.* LRP6 transduces a canonical Wnt signal independently of Axin degradation by inhibiting GSK3's phosphorylation of beta-catenin. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8032–7 (2008).
31. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
32. Schaffer, B. A. J. *et al.* Association of GSK3B With Alzheimer Disease and Frontotemporal Dementia. *Arch. Neurol.* **65**, 1368–1374 (2008).
33. Stambolic, V., Ruel, L. & Woodgett, J. R. Lithium inhibits glycogen synthase kinase-3 activity and mimics Wingless signalling in intact cells. *Curr. Biol.* **6**, 1664–1669 (1996).
34. Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8455–9 (1996).
35. O'Brien, W. T. & Klein, P. S. Validating GSK3 as an in vivo target of lithium action. *Biochem. Soc. Trans.* **37**, 1133–8 (2009).
36. Jope, R. S. Glycogen Synthase Kinase-3 in the Etiology and Treatment of Mood Disorders. *Front. Mol. Neurosci.* **4**, 16 (2011).
37. Manoukian, A. S. & Woodgett, J. R. Role of glycogen synthase kinase-3 in cancer: Regulation by Wnts and other signaling pathways. *Adv. Cancer Res.* **84**, 203–229

- (2002).
38. Ougolkov, A. V & Billadeau, D. D. Targeting GSK-3: a promising approach for cancer therapy? *Futur. Oncol.* **2**, 91–100 (2006).
 39. Mills, C. N., Nowsheen, S., Bonner, J. A. & Yang, E. S. Emerging Roles of Glycogen Synthase Kinase 3 in the Treatment of Brain Tumors. *Front. Mol. Neurosci.* **4**, 47 (2011).
 40. McCubrey, J. A. *et al.* Multifaceted roles of GSK-3 and Wnt/ β -catenin in hematopoiesis and leukemogenesis: opportunities for therapeutic intervention. *Leukemia* **28**, 15–33 (2014).
 41. Polakis, P. The many ways of Wnt in cancer. *Curr. Opin. Genet. Dev.* **17**, 45–51 (2007).
 42. Ma, C. *et al.* The role of glycogen synthase kinase 3beta in the transformation of epidermal cells. *Cancer Res.* **67**, 7756–64 (2007).
 43. Leis, H., Segrelles, C., Ruiz, S., Santos, M. & Paramio, J. M. Expression, localization, and activity of glycogen synthase kinase 3? during mouse skin tumorigenesis. *Mol. Carcinog.* **35**, 180–185 (2002).
 44. Farago, M. *et al.* Kinase-inactive glycogen synthase kinase 3beta promotes Wnt signaling and mammary tumorigenesis. *Cancer Res.* **65**, 5792–801 (2005).
 45. Armanious, H. *et al.* Clinical and biological significance of GSK-3 β inactivation in breast cancer—an immunohistochemical study. *Hum. Pathol.* **41**, 1657–1663 (2010).
 46. Dembowy, J., Adissu, H. A., Liu, J. C., Zacksenhaus, E. & Woodgett, J. R. Effect of glycogen synthase kinase-3 inactivation on mouse mammary gland development and oncogenesis. *Oncogene* **34**, 3514–3526 (2015).

47. Vincent, T., Kukalev, A., Andäng, M., Pettersson, R. & Percipalle, P. The glycogen synthase kinase (GSK) 3 β represses RNA polymerase I transcription. *Oncogene* **27**, 5254–5259 (2008).
48. Shakoori, A. *et al.* Inhibition of GSK-3 β activity attenuates proliferation of human colon cancer cells in rodents. *Cancer Sci.* **98**, 1388–1393 (2007).
49. Ougolkov, A. V, Fernandez-zapico, M. E., Savoy, D. N., Urrutia, R. A. & Billadeau, D. D. Glycogen Synthase Kinase-3 B Participates in Nuclear Factor K B – Mediated Gene Transcription and Cell Survival in Pancreatic Cancer Cells. 2076–2082 (2005).
50. Garcea, G. *et al.* Glycogen Synthase Kinase-3 Beta; A New Target in Pancreatic Cancer? *Curr. Cancer Drug Targets* **7**, 209–215 (2007).
51. Wang, Z. *et al.* Glycogen synthase kinase 3 in MLL leukaemia maintenance and targeted therapy. *Nature* **455**, 1205–1209 (2008).
52. Mulholland, D. J., Dedhar, S., Wu, H. & Nelson, C. C. PTEN and GSK3 β : key regulators of progression to androgen-independent prostate cancer. *Oncogene* **25**, 329–337 (2006).
53. Darrington, R. S. *et al.* Distinct expression and activity of GSK-3 α and GSK-3 β in prostate cancer. *Int. J. Cancer* **131**, E872–E883 (2012).
54. Flórez, J., Armijo, J. A. & Mediavilla, A. *Farmacologia humana*.
55. Sutherland, C., Leighton, I. A. & Cohen, P. Inactivation of glycogen synthase kinase-3 beta by phosphorylation: new kinase connections in insulin and growth-factor signalling. *Biochem. J.* **296** (Pt 1), 15–9 (1993).
56. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes

- (UKPDS 33). *Lancet* **352**, 837–853 (1998).
57. Eldar-Finkelman, H., Schreyer, S., Diabetes, M. S.- & 1999, undefined. Increased glycogen synthase kinase-3 activity in diabetes-and obesity-prone C57BL/6J mice. *Am Diabetes Assoc Diabetes Assoc*
 58. Eldar-Finkelman, H. & Kaidanovich, O. The role of glycogen synthase kinase-3 in insulin resistance and Type 2 diabetes. *Expert Opin. Ther. Targets* **6**, 555–561 (2002).
 59. Nikoulina, S. E. *et al.* Potential role of glycogen synthase kinase-3 in skeletal muscle insulin resistance of type 2 diabetes. *Diabetes* **49**, 263–71 (2000).
 60. Ciaraldi, T. P., Nikoulina, S. E. & Henry, R. R. Role of glycogen synthase kinase-3 in skeletal muscle insulin resistance in Type 2 diabetes. *J. Diabetes Complications* **16**, 69–71 (2002).
 61. Jope, R. S. Lithium and GSK-3: one inhibitor, two inhibitory actions, multiple outcomes. *Trends Pharmacol. Sci.* **24**, 441–443 (2003).
 62. Martinez, A., Castro, A., Dorronsoro, I. & Alonso, M. Glycogen synthase kinase 3 (GSK-3) inhibitors as new promising drugs for diabetes, neurodegeneration, cancer, and inflammation. *Med. Res. Rev.* **22**, 373–384 (2002).
 63. Eldar-Finkelman, H. & Ilouz, R. Challenges and opportunities with glycogen synthase kinase-3 inhibitors for insulin resistance and Type 2 diabetes treatment. *Expert Opin. Investig. Drugs* **12**, 1511–1519 (2003).
 64. Van Wauwe, J. & Haefner, B. Glycogen synthase kinase-3 as drug target: from wallflower to center of attention. *Drug News Perspect.* **16**, 557–65 (2003).
 65. Wagman, A., Johnson, K. & Bussiere, D. Discovery and Development of GSK3 Inhibitors for the Treatment of Type 2 Diabetes. *Curr. Pharm. Des.* **10**, 1105–1137

- (2004).
66. Davies, S. P., Reddy, H., Caivano, M. & Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.* **351**, 95–105 (2000).
 67. Bain, J., McLauchlan, H., Elliott, M. & Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem. J.* **371**, 199–204 (2003).
 68. Meijer, L. *et al.* GSK-3-Selective Inhibitors Derived from Tyrian Purple Indirubins. *Chem. Biol.* **10**, 1255–1266 (2003).
 69. Knockaert, M. *et al.* *Intracellular targets of paullones: identification following affinity purification on immobilized inhibitor Downloaded from. J. Biol. Chem* **5**, (JBC Papers in Press. Published on, 2002).
 70. Meijer, L., Flajolet, M. & Greengard, P. Pharmacological inhibitors of glycogen synthase kinase 3. *Trends Pharmacol. Sci.* **25**, 471–480 (2004).
 71. Cohen, P. & Goedert, M. GSK3 inhibitors: development and therapeutic potential. *Nat. Rev. Drug Discov.* **3**, 479–487 (2004).
 72. Coghlan, M. P. *et al.* Selective small molecule inhibitors of glycogen synthase kinase-3 modulate glycogen metabolism and gene transcription. *Chem. Biol.* **7**, 793–803 (2000).
 73. Henriksen, E. J. *et al.* Modulation of muscle insulin resistance by selective inhibition of GSK-3 in Zucker diabetic fatty rats. *Am. J. Physiol. Metab.* **284**, E892–E900 (2003).
 74. Cline, G. W. *et al.* Effects of a novel glycogen synthase kinase-3 inhibitor on insulin-stimulated glucose metabolism in Zucker diabetic fatty (fa/fa) rats. *Diabetes* **51**, 2903–10 (2002).

75. Wermuth, C. G. *The Practice of Medicinal Chemistry*. **3^o edition**, 942 (2008).
76. Nantasenamat, C., Isarankura-Na-Ayudhya, C. & Prachayasittikul, V. Advances in computational methods to predict the biological activity of compounds. *Expert Opin. Drug Discov.* **5**, 633–654 (2010).
77. Ghafourian, T. & Fooladi, S. The effect of structural QSAR parameters on skin penetration. *Int. J. Pharm.* **217**, 1–11 (2001).
78. Betageri, G. V. & Rogers, J. A. Correlation of Partitioning of Nitroimidazoles in the n-Octanol/Saline and Liposome Systems with Pharmacokinetic Parameters and Quantitative Structure–Activity Relationships (QSAR). *Pharm. Res.* **06**, 399–403 (1989).
79. Chuman, H. Toward basic understanding of the partition coefficient log *P* and its application in QSAR 1. *SAR QSAR Environ. Res.* **19**, 71–79 (2008).
80. Jonathan J. Sullivan, *, †, A. Daniel Jones, § and & Tanji†, K. K. QSAR Treatment of Electronic Substituent Effects Using Frontier Orbital Theory and Topological Parameters. (2000). doi:10.1021/CI990222K
81. Fernández, I. & Frenking, G. Correlation between Hammett substituent constants and directly calculated pi-conjugation strength. *J. Org. Chem.* **71**, 2251–6 (2006).
82. Longo, F. R. *et al.* *The Biological Basis of Radiation Ther-apy*. Hayatsu and S. Iida *Tetrahedron Lett* **3**, (Soc. C, 1970).
83. Daunes, S., D'silva, C., Kendrick, H., Yardley, V. & Croft, S. L. QSAR Study on the Contribution of Log P and E s to the in Vitro Antiprotozoal Activity of Glutathione Derivatives. (2001). doi:10.1021/jm000502n
84. Xi, Z., Yu, Z., Niu, C., Ban, S. & Yang, G. Development of a general quantum-

- chemical descriptor for steric effects: Density functional theory based QSAR study of herbicidal sulfonylurea analogues. *J. Comput. Chem.* **27**, 1571–1576 (2006).
85. Toshinobu Korenaga, *, Koichi Kadowaki, Tadashi Ema, and & Sakai*, T. Reestimation of the Taft's Substituent Constant of the Pentafluorophenyl Group. (2004). doi:10.1021/JO048939G
 86. Sigman, M. S. & Miller, J. J. Examination of the Role of Taft-Type Steric Parameters in Asymmetric Catalysis. *J. Org. Chem.* **74**, 7633–7643 (2009).
 87. Kaliszan, R., Osmialowski, K. & Kobylczyk, K. In vitro activity of sulphonamides as a function of their molar refractivity. *Chemotherapy* **25**, 5–8 (1979).
 88. Heinzen, V. E. ., Cechinel Filho, V. & Yunes, R. A. Correlation of activity of 2-(X-benzyloxy)-4,6-dimethoxyacetophenones with topological indices and with the Hansch equation. *Farm.* **54**, 125–129 (1999).
 89. Wei, D. B., Zhang, A. Q., Han, S. K. & Wang, L. S. Joint QSAR Analysis Using the Free-Wilson Approach and Quantum Chemical Parameters. *SAR QSAR Environ. Res.* **12**, 471–479 (2001).
 90. Craig, P. N. Proceedings: Comparison of Hansch and free-Wilson methods for structure-activity correlation. *Cancer Chemother. Rep.* **24**, 39 (1974).
 91. Chen, H. *et al.* Beyond the Scope of Free-Wilson Analysis: Building Interpretable QSAR Models with Machine Learning Algorithms. *J. Chem. Inf. Model.* **53**, 1324–1336 (2013).
 92. Cramer, R. D., Patterson, D. E. & Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **291**, 161–5 (1989).
 93. Byberg, J. R., Jørgensen, F. S. & Klemmensen, P. D. Towards an identification of the

- pyrethroid pharmacophore. A molecular modelling study of some pyrethroid esters. *J. Comput. Aided. Mol. Des.* **1**, 181–195 (1987).
94. Cramer, R. D., Clark, R. D., Patterson, D. E. & Ferguson, A. M. *Bioisosterism as a Molecular Diversity Descriptor: Steric Fields of Single "Topomeric" Conformers.* (1996).
 95. Konovalov, D. A., Llewellyn, L. E., Vander Heyden, Y. & Coomans, D. Robust Cross-Validation of Linear Regression QSAR Models. *J. Chem. Inf. Model.* **48**, 2081–2094 (2008).
 96. Klebe, G. & Abraham, U. On the prediction of binding properties of drug molecules by comparative molecular field analysis. *J. Med. Chem.* **36**, 70–80 (1993).
 97. Klebe, G., Abraham, U. & Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **37**, 4130–4146 (1994).
 98. Stähle, L. & Wold, S. 6 Multivariate Data Analysis and Experimental Design in Biomedical Research. *Prog. Med. Chem.* **25**, 291–338 (1988).
 99. Wagman, A. S. *et al.* Synthesis, Binding Mode, and Antihyperglycemic Activity of Potent and Selective (5-Imidazol-2-yl-4-phenylpyrimidin-2-yl)[2-(2-pyridylamino)ethyl]amine Inhibitors of Glycogen Synthase Kinase 3. *J. Med. Chem.* **60**, 8482–8514 (2017).
 100. Powell, M. J. D. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **7**, 155–162 (1964).
 101. SYBYL-X 1.2 Tripos International, South Hanley Rd., St. Louis, Missouri, 63144, U. SYBYL-X 1.2.

102. Clark, M., Cramer, R. D. & Van Opdenbosch, N. Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* (1989). doi:10.1002/jcc.540100804
103. Golbraikh, A. & Tropsha, A. Beware of q²! *J. Mol. Graph. Model.* **20**, 269–276 (2002).
104. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* **29**, 476–488 (2010).
105. Roy, K. & Kar, S. The rm2 metrics and regression through origin approach: Reliable and useful validation tools for predictive QSAR models (Commentary on ‘Is regression through origin useful in external validation of QSAR models?’). *Eur. J. Pharm. Sci.* **62**, 111–114 (2014).
106. Roy, K., Kar, S. & Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* **145**, 22–29 (2015).
107. Rücker, C., Gerta Rücker, & Meringer, M. y-Randomization and Its Variants in QSPR/QSAR. (2007). doi:10.1021/CI700157B
108. Brown, T. L., LeMay, H. E. (Harold E., Bursten, B. E. & Escalona y García, H. J. *Química : la ciencia central*. (Pearson Education, 2004).

9. ANEXO

Tabla S1: Fármacos inhibidores de GSK-3 reportados por Meijer et al⁷⁰.

Inhibitor ^b	Class	IC ₅₀ (μM)	
		GSK-α and GSK-3β ^c	CDK1–cyclin B complex
Hymenialdisine	Pyrroloazepine	0.010 (β)	0.022
Flavopiridol	Flavone	0.450	0.400
Kenpaullone	Benzazepinone	0.023 (β)	0.400
Alsterpaullone	Benzazepinone	0.004 (α); 0.004 (β)	0.035
Azakenpaullone	Benzazepinone	0.018 (β)	2.000
Indirubin-3'-oxime	Bis-Indole	0.022 (β)	0.018
6-Bromoindirubin-3'-oxime (BIO)	Bis-Indole	0.005	0.320
6-Bromoindirubin-3'-acetoxime	Bis-Indole	0.010	63.000
Aloisine A	Pyrrolopyrazine	0.650	0.150
Aloisine B	Pyrrolopyrazine	0.750	0.850
TDZD8	Thiadiazolidinone	2.000 (β); 7.000 (α/β) ^d	> 100; > 10 ^d
Compound 12	Pyridyloxadiazole	0.390 (β); 8.000 (α/β) ^d	> 10 ^d
Pyrazolopyridine 18	Pyrazolopyridine	0.018 (α)	Inhibits CDK2–cyclin A (95% at 10 μM)
Pyrazolopyridine 9	Pyrazolopyridazine	0.022 (α)	Inhibits CDK2–cyclin A (90% at 10 μM)
Pyrazolopyridine 34	Pyrazolopyridine	0.007 (α)	> 10 (CDK2–cyclin A)
CHIR98014	Aminopyrimidine	0.00065 (α); 0.00058 (β)	3.700
CHIR99021 (CT99021)	Aminopyrimidine	0.010 (α); 0.007 (β)	8.800
CT20026	Aminopyridine	0.004 (α/β)	?
Compound 1	Pyrazoloquinoxaline	1.000	0.600
SU9516	Oxindole (indolinone)	0.330; 0.35 (α/β) ^d	0.040; 0.022 ^d
ARA014418	Thiazole	0.104 (β)	> 100 (CDK2 and CDK5)
Staurosporine	Bisindolylmaleimide	0.015; 0.089	0.006; 0.008
Compound 5a	Bisindolylmaleimide	0.018 (β)	0.24
Compound 29	Azaindolylmaleimide	0.034 (β)	> 10
Compound 46	Azaindolylmaleimide	0.036 (β)	> 10
GF109203x (bisindolylmaleimide I)	Bisindolylmaleimide	0.190 (β)	2.300 ^d
Ro318220 (bisindolylmaleimide IX)	Bisindolylmaleimide	0.003–0.038 (β)	?
SB216763	Arylindolemaleimide	0.034 (α); 0.075 (α/β) ^d	0.550 ^d
SB415286	Anilinomaleimide	0.078 (α); 0.13 (α/β) ^d	0.900 ^d
I5	Anilinoarylmaleimide	0.076 (α); 0.160 (β)	> 10 (CDK2–cyclin A)
CGP60474	Phenylaminopyrimidine	0.010 ^d	0.017; 0.0006 ^d
Compound 8b	Triazole	0.280 (β)	> 250 (CDK2–cyclin A)
TWS119	Pyrrolopyrimidine	0.030 (β)	?
Compound 1A	Pyrazolopyrimidine	0.016 (β)	?
Compound 17	Chloromethyl thienyl ketone	1.00 (β)	?
Lithium	Atom (competition with Mg ²⁺)	2000.0	No effect
Beryllium	Atom (competition with Mg ²⁺ and ATP)	6.00	Inhibits CDK1
Zinc	Atom (uncompetitive)	15.00	No effect

Tabla S2: Valores de π para diversos sustituyentes en el anillo aromático.

Sustituyente	π aromático
H	0,00
Br	0,86
Cl	0,71
F	0,14
NO ₂	-0,28
NH ₂	-1,23
OH	-0,67
OCH ₃	-0,02
CN	-0,57
CF ₃	0,88
CH ₃	0,56
CH ₂ CH ₃	1,02
CH ₂ CH ₂ CH ₃	1,55

Tabla S3: Valor de σ para diversos sustituyentes en el anillo aromático. Se muestran valores para el sustituyente en la posición meta- (σ_m) y en la posición para- (σ_p).

Sustituyente	σ_m	σ_p
H	0,00	0,00
Br	0,39	0,23
Cl	0,37	0,23
F	0,34	0,06
NO ₂	0,71	0,78
NH ₂	-0,16	-0,66
OH	0,12	-0,37
OCH ₃	0,12	-0,27
CN	0,56	0,66
CF ₃	0,43	0,54
CH ₃	-0,07	-0,17
CH ₂ CH ₃	-0,07	-0,15
CH ₂ CH ₂ CH ₃	-0,07	-0,13

Tabla S4: Valores de E_s para distintos sustituyentes.

Sustituyente	E_s
H	0,00
Br	-1,16
Cl	-0,97
F	-0,46
NO ₂	-2,52
NH ₂	-0,61
OH	-0,55
OCH ₃	-0,55
CN	-0,51
CF ₃	-2,40
CH ₃	-1,24
CH ₂ CH ₃	-1,31
CH ₂ CH ₂ CH ₃	-1,60