



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

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

**ESTUDIO COMPUTACIONAL SOBRE LA INHIBICIÓN DE LA CRUZAÍNA
POR PURINO-CARBONITRILOS Y ANÁLISIS QUIMIOINFORMÁTICO DE
BASES DE DATOS**

TESIS
PARA OPTAR POR EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

Q.F.B. OSCAR MÉNDEZ LUCIO



TUTOR: Dr. Rafael Castillo Bocanegra

2012



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO

Presidente	Dr. Juan Raúl Alvarez Idaboy
Vocal	Dr. Rogelio Rodríguez Sotres
Vocal	Dr. Enrique Ramón Ángeles Anguiano
Vocal	Dr. Juan Gabriel Navarrete Vázquez
Secretario	Dr. Fernando Cortés Guzmán

SITIO DONDE SE DESARROLLÓ EL TEMA:

Departamento de Farmacia, Laboratorio 122, Facultad de Química de la Universidad Nacional Autónoma de México, bajo la dirección del Dr. Rafael Castillo Bocanegra.

Los resultados de esta tesis dieron lugar a las siguientes publicaciones:

- O. Méndez-Lucio, J. Perez-Villanueva, A. Romo-Mancillas, R. Castillo. 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). *Med. Chem. Commun.* 2011, 2, 1058-65.
- O. Méndez-Lucio, A. Romo-Mancillas, J.L. Medina-Franco, R. Castillo. Computational study on the inhibition mechanism of cruzain by nitrile-containing molecules. *J. Mol. Graphics Modell.* 2012, 35, 28-35.
- O. Méndez-Lucio, J. Perez-Villanueva, R. Castillo, J.L. Medina-Franco. Activity landscape modeling of PPAR ligands with Dual-Activity Difference maps. *Bioorg. Med. Chem.* 2012, 20, 3523-32.

Y fueron presentados en los siguientes congresos:

- 45º Congreso Mexicano de Química, 18-22 septiembre 2010.
Estudio CoMFA y CoMSIA de purino-carbonitrilos inhibidores de la cruzaína
- 7ª Reunión de la Academia Mexicana de Química Orgánica, Universidad Autónoma del Estado de Morelos, 7 y 8 de abril 2011.
Estudio teórico sobre el mecanismo de inhibición de la cruzaína por purino-carbonitrilos
- Advanced School on Natural Products, Medicinal Chemistry and Organic Synthesis, Universidad de Campinas, Sao Paulo, Brasil, 1-8 de agosto 2011.
Computational study on the inhibition of cruzain by purine-carbonitriles
- XXXVII Congreso de Químicos Teóricos de Expresión Latina (QUITEL 2011), 4-8 de diciembre 2011
Theoretical study on the inhibition mechanism of cysteine proteases by nitrile-containing molecules

AGRADECIMIENTOS

Agradezco a la Universidad Nacional Autónoma de México, especialmente a la Facultad de Química, por haberme brindado la oportunidad y todas las herramientas necesarias para concluir mis estudios de maestría. Al Posgrado de Ciencias Químicas y a todas las personas que trabajan en él por todo su apoyo. Al Consejo Nacional de Ciencia y Tecnología por la beca número 245408 otorgada para la realización de mis estudios de maestría. A todos los sinodales que enriquecieron ampliamente este trabajo con sus observaciones y sugerencias.

Quiero agradecer profundamente al Dr. Rafael Castillo Bocanegra por ser un excelente persona, por contribuir en mi formación académica y personal, y por permitirme formar parte de su grupo de investigación. Al Dr. José Luis Medina Franco por todo el apoyo que me brindó desinteresadamente durante esta etapa, por estimularme a estar en constante aprendizaje y por ser un buen ejemplo a seguir. A la M. en C. Alicia Hernández, Dr. Francisco Hernández Luis, y al Dr. Alfonso Lira Rocha por todas sus enseñanzas dentro y fuera del salón de clases.

A Rodrigo Aguayo, Berenice Ocón, Paulina Flores, Armando Rangel, Pedro García, Silvia Melchor, Irene Escamilla, Laura González, Miguel Flores, Kemish López y a todos los miembros del Laboratorio 122. Especialmente al M. en C. Jaime Pérez, M. en C. Antonio Romo, M. en C. Carlos Méndez "Golden", M. en C. Eduardo Hernández y al QFB. Pedro Trejo quienes me apoyaron directamente en este trabajo de tesis con sus consejos y sugerencias. Gracias a todos por su amistad, apoyo y por hacer que todos los días fueran una experiencia nueva.

A Mario Méndez, Bety Lucio y a Odette Méndez por todo el cariño que me brindan día a día. Por todo su apoyo y por impulsarme para alcanzar mis metas.

A Aracelia Tamariz por haberme acompañado desde el inicio de este trabajo escuchando cada avance y alentándome a seguir adelante.

A todos mis amigos, en especial a Kweilan Yap, Dulce Cruz y Pablo Jardón por los buenos y malos ratos que hemos compartido y que seguiremos compartiendo.

Dedicada con todo cariño y admiración a

Mario, Bety y Odette

ÍNDICE

RESUMEN.....	I
ABSTRACT.....	II
ABREVIATURAS	III
CAPÍTULO 1. INTRODUCCIÓN	1
PARTE I.....	3
CAPÍTULO 2. ANTECEDENTES.....	4
2.1. <i>La enfermedad de Chagas</i>	4
2.2. <i>La cruzaina</i>	6
2.3. <i>Mecanismo catalítico de la cruzaina</i>	7
2.4. <i>Inhibidores de la cruzaina</i>	9
CAPÍTULO 3. METODOLOGÍA	12
3.1. <i>Estudios QSAR-3D</i>	12
3.1.1. Conjunto de datos.....	12
3.1.2. Preparación de ligandos y alineamiento molecular	12
3.1.3. Análisis comparativo de campos moleculares (CoMFA) y análisis comparativo de índices de similitud molecular (CoMSIA)	14
3.1.4. Análisis estadístico y validación	14
3.2. <i>Estudio del mecanismo de reacción</i>	16
3.2.1. Cálculos utilizando la Teoría de Funcionales de la Densidad (DFT).....	16
3.2.2. Cálculos a nivel semi-empírico	18
CAPÍTULO 4. RESULTADOS.....	20
4.1. <i>Estudios QSAR-3D</i>	20
4.1.1. Modelos CoMFA y CoMSIA.....	20
4.1.2. Validación de los modelos CoMFA y CoMSIA.....	26
4.1.3. Mapas de contorno obtenidos utilizando CoMFA	28
4.1.4. Mapas de contorno obtenidos utilizando CoMSIA.....	30
4.2. <i>Estudios químico-cuántico del mecanismo de reacción</i>	35
4.2.1. Superficie de energía potencial	35
4.2.2. Análisis de orbitales moleculares	40
4.2.3. Estudio de la reacción dentro del sitio catalítico.....	42
CAPÍTULO 5. CONCLUSIONES	46

5.1. Conclusiones del estudio QSAR 3D	46
5.2. Conclusiones del estudio químico-cuántico del mecanismo de reacción	47
PARTE II	49
CAPÍTULO 6. ANTECEDENTES	50
6.1. Panorama de actividad	50
6.2. Activity cliffs	52
6.3. Scaffold hops	53
6.4. Receptores activadores de la proliferación de peroxisomas (PPAR)	55
CAPÍTULO 7. METODOLOGÍA	58
7.1. Conjunto de datos	58
7.2. Similitud estructural	58
7.3. Similitud de actividad	59
7.4. Mapas SAS	60
7.5. Mapas DAD y TAD	61
CAPÍTULO 8. RESULTADOS	63
8.1. Distribución de las medidas de similitud	63
8.2. Mapas SAS	65
8.3. Mapas DAD y TAD	69
8.4. Activity cliffs simples	73
8.5. Activity cliffs dobles y triples	75
8.6. Scaffold hops	78
CAPÍTULO 9. CONCLUSIONES	82
9.1. Modelado del panorama de actividad para ligandos de PPAR	82
REFERENCIAS	84
APÉNDICES	91
Apéndice A	92
Apéndice B	93
Apéndice C	99
ARTÍCULOS PUBLICADOS	100

RESUMEN

El trabajo realizado en esta tesis se presenta en dos partes. La primera comprende un estudio teórico sobre algunos inhibidores de la cruzaina y su posible mecanismo de acción; en la segunda, se analizaron, por métodos computacionales, diversos agonistas de los Receptores Activadores de la Proliferación de Peroxisomas (PPARs) con el propósito de identificar las características estructurales relevantes con la actividad y selectividad de estos compuestos.

La cruzaina es la cisteína proteasa más importante de *Trypanosoma cruzi*, el causante etiológico de la enfermedad de Chagas. Recientemente, se reportaron moléculas con un núcleo de purina o triazina como potentes inhibidores no peptídicos de la cruzaina. En esta tesis se realizaron estudios CoMFA y CoMSIA de una serie de purino-carbonitrilos con la finalidad de obtener información acerca de algunos requerimientos estructurales que puedan mejorar la actividad de estas moléculas. Adicionalmente, se estudió la energética asociada con la formación de un enlace covalente entre el grupo nitrilo y la Cys25. Para este propósito se realizaron cálculos químico-cuánticos usando la teoría de funcionales de la densidad y a nivel semi-empírico. Los resultados sugieren que la interacción empieza con un ataque nucleofílico de la Cys25 al inhibidor seguido de una transferencia de protón por parte de la His162.

En la segunda parte de esta tesis se usaron los principios de modelado del *panorama de actividad* para realizar una descripción sistemática del SAR de 168 compuestos probados contra tres subtipos de PPARs. Los PPARs ofrecen una estrategia prometedora para el tratamiento de diabetes mellitus y síndrome metabólico. El análisis de pares de compuestos con alta similitud estructural reveló la presencia de cambios abruptos en la actividad hacia uno, dos o tres receptores (*activity cliffs*). El análisis de las estructuras químicas de algunos puntos seleccionados brinda información acerca de características estructurales que son importantes para el diseño de nuevos agonistas de PPAR.

ABSTRACT

The results presented in this thesis are divided in two parts. The former comprises a theoretical study on some cruzain inhibitors and their possible mechanism of action; in the latter, a set of Peroxisome Proliferator-Activated Receptors (PPARs) agonist were analyzed using computational methods in order to identify important structural characteristics for the biological activity and selectivity of these compounds.

Cruzain has been identified as the major cysteine protease of *Trypanosoma cruzi*, the etiological cause of Chaga's disease. Recently, molecules containing a purine or triazine nucleus have been reported as potent non-peptidic inhibitors of cruzain. In order to gain an insight into the structural requirements that can lead to the improvement of the activity of these molecules, in this thesis we report the CoMFA and CoMSIA studies of a series of purine-carbonitriles as cruzain inhibitors. Additionally, the energetics involved in the formation of a covalent bond between the nitrile group and the Cys25 was studied. For this propose, density functional theory and quantum semi-empirical calculations were conducted. Results suggest an interaction that starts with a nucleophilic attack from the Cys25 to the inhibitor followed by a proton transfer from His162.

In the second part of this thesis, we report a systematic description of the SAR of 168 compounds screened against the three PPAR subtypes using the principles of activity landscape modeling. PPARs offer a promising strategy for the treatment of diabetes mellitus and metabolic diseases. Analysis of pairs of compounds with high structure similarity revealed the presence of single-, dual-, and triple-target activity cliffs. The analysis of the chemical structures of selected data points reported in this paper points to specific structural features that are helpful for the design of new PPAR agonists.

ABREVIATURAS

<i>Activity cliff</i>	Cambios abruptos en la actividad
CoMFA	Análisis comparativo de campos moleculares
CoMSIA	Análisis comparativo de índices de similitud molecular
DAD	Diferencia de actividad dual
DFT	Teoría de funcionales de la densidad
HOMO	Orbital ocupado de mayor energía
IRC	Coordenada intrínseca de reacción
LOO	Validación <i>Leave-One-Out</i>
LUMO	Orbital desocupado de menor energía
PDB	Protein Data Bank
PLS	Mínimos cuadrados parciales
PPAR	Receptores activadores de la proliferación de peroxisomas
QSAR	Relación cuantitativa estructura-actividad
QSAR-3D	Relación cuantitativa estructura-actividad en tres dimensiones
SAR	Relación estructura-actividad
SAS	Similitud estructura-actividad
<i>Scaffold hop</i>	Cambio de estructura base
SEA	Estérico, electrostático y aceptor de puentes de hidrógeno
SED	Estérico, electrostático y donador de puentes de hidrógeno
SEDA	Estérico, electrostático, donador y aceptor de puentes de hidrógeno
SEHDA	Estérico, electrostático, hidrofóbico, donador y aceptor de puentes de hidrógeno
SEP	Superficie de energía potencial
TAD	Diferencia de actividad triple

CAPÍTULO 1. INTRODUCCIÓN

La obtención de un fármaco nuevo es un proceso muy lento y costoso. Las estadísticas indican que de cada 40,000 compuestos que se prueban en animales, 5 llegan a ser evaluados en seres humanos y solamente 1 de ellos es aprobado para ser utilizado como fármaco en algún medicamento. Se estima que se requieren de 7 a 12 años y una inversión de alrededor de 1.2 mil millones de dólares para que un fármaco nuevo llegue al mercado [1]. Por esta razón el uso de métodos computacionales ha sido una buena alternativa para acelerar y reducir los costos de este proceso. Existen diferentes ejemplos de fármacos que se han descubierto o han sido optimizados con el uso de métodos computacionales, algunos de ellos son: Captopril, Dorzolamida, Saquinavir, Zanamivir, Oseltamivir, Aliskiren, Boceprevir, Nilotrexed y Rupintrivir [2].

En los últimos años, se han desarrollado diversas metodologías computacionales que permiten el diseño u optimización de moléculas bioactivas, los cuales se pueden dividir en: (I) diseño basado en la estructura, por ejemplo, el acoplamiento molecular o docking y (II) diseño basado en el ligando en los que destacan los estudios de relación cuantitativa estructura-actividad (QSAR). Dentro de estos últimos, el uso de campos de interacción molecular conjuntamente con el análisis estadístico de mínimos cuadrados parciales dio pie al desarrollo de métodos que relacionan cuantitativamente la conformación tridimensional de las moléculas con su actividad biológica. A estos métodos se les conoce como QSAR-3D. Entre los métodos QSAR-3D más populares se encuentran el análisis comparativo de campos moleculares (CoMFA) y el análisis comparativo de índices de similitud molecular (CoMSIA) [3, 4].

En este trabajo de tesis se estudiaron algunos inhibidores de la cruzaína y diversos agonistas de los receptores activadores de la proliferación de peroxisomas (PPARs) empleando métodos computacionales con la finalidad de obtener más información acerca de las interacciones ligando-receptor. En la primera parte de esta tesis se utilizaron las metodologías CoMFA y CoMSIA para analizar las interacciones estéricas, electrostáticas y la formación de puentes de

hidrógeno entre la cruzaina y sus inhibidores. De manera similar, se utilizaron cálculos químico-cuánticos para estudiar la formación de un enlace covalente entre el sitio catalítico de la enzima y el inhibidor. La segunda parte de la tesis es un claro ejemplo de diseño basado en el ligando, en el cual se empleó el concepto del panorama de actividad para estudiar 168 compuestos con actividad reportada hacia tres subtipos de PPAR. El objetivo principal de este estudio fue analizar los requerimientos esenciales para la actividad biológica, así como algunos determinantes estructurales para la selectividad hacia los diferentes subtipos de PPAR.

Este trabajo forma parte de un esfuerzo multidisciplinario enfocado a la búsqueda de nuevas moléculas bioactivas para el tratamiento de la diabetes mellitus y la enfermedad de Chagas. Los resultados obtenidos en esta tesis brindan información importante que resaltan la importancia del uso de los métodos computacionales para el diseño racional de moléculas bioactivas.

PARTE I

ESTUDIO COMPUTACIONAL SOBRE LA INHIBICIÓN DE LA CRUZAÍNA POR PURINO- CARBONITRILOS

CAPÍTULO 2. ANTECEDENTES

Esta parte del trabajo de tesis consiste de un estudio computacional sobre algunas moléculas con un núcleo de purino-carbonitrilo que tienen actividad como inhibidores covalentes reversibles de la cruzafina, la cisteína proteasa más importante del *Trypanosoma cruzi*. En este estudio se utilizaron metodologías como el análisis comparativo de campos moleculares y el análisis comparativo de índices de similitud molecular para revelar los requerimientos estructurales de estas moléculas para la actividad biológica. Así mismo, se utilizaron cálculos químico-cuánticos a nivel semi-empírico y de teoría de funcionales de la densidad para estudiar el mecanismo de reacción por el cual estos inhibidores se unen covalentemente a la cruzafina.

En este capítulo, se pretende dar una breve introducción acerca de la enfermedad de Chagas, incluyendo su agente causal y sus síntomas. Además, se describe la importancia biológica de la cruzafina, al igual que su uso como blanco terapéutico.

2.1. La enfermedad de Chagas

Esta enfermedad recibe su nombre en honor al médico brasileño Carlos Justiniano Ribeiro Chagas (1879-1934) que describió los síntomas por primera vez en 1909 [5]. La enfermedad de Chagas es un padecimiento causado por el protozoario flagelado *Trypanosoma cruzi*. Este parásito infecta entre 10 y 12 millones de personas y mata alrededor de 15,000 cada año, lo cual convierte a esta enfermedad en un serio problema de salud [6].

Trypanosoma cruzi es transmitido al ser humano mediante las heces de insectos de los géneros *Triatoma*, *Panstrongylus* y *Rhodnius* [5]. Una vez que el humano es infectado se presenta una etapa aguda, la cual suele pasar desapercibida. Después de esta etapa se presenta un periodo asintomático que puede durar de 10 a 20 años. Finalmente, entre un 20-30% de los pacientes presentan un etapa

crónica la cual puede culminar en una muerte súbita por crecimiento de vísceras y/o problemas cardiacos [7].

El ciclo de vida que presenta *Trypanosoma cruzi* consiste en cuatro etapas. La primera de ellas se presenta cuando el parásito es transmitido al humano en forma de tripomastigote metacíclico. Posteriormente, éste se introduce en una célula humana y adquiere su forma reproductiva conocida como amastigote, la cual es intracelular obligada. Dentro de la célula el amastigote se reproduce por fisión binaria y los amastigotes resultantes se convierten nuevamente en tripomastigotes, los cuales salen de la célula y entran en el torrente sanguíneo. Este ciclo infectivo se repite varias veces y puede ser el causante de las manifestaciones clínicas. Si el paciente vuelve a sufrir la picadura de algún insecto, éste puede ingerir los tripomastigotes que se encuentran en el torrente sanguíneo. Una vez en el vector biológico, los tripomastigotes se convierten en epimastigotes, los cuales se multiplican en el intestino. Finalmente, al llegar a la zona rectal del insecto, los epimastigotes se transforman en tripomastigotes metacíclicos para ser transmitidos por las heces y comenzar el ciclo nuevamente [7]. En la Figura 2.1 se muestra esquemáticamente el ciclo de vida de *Trypanosoma cruzi*.

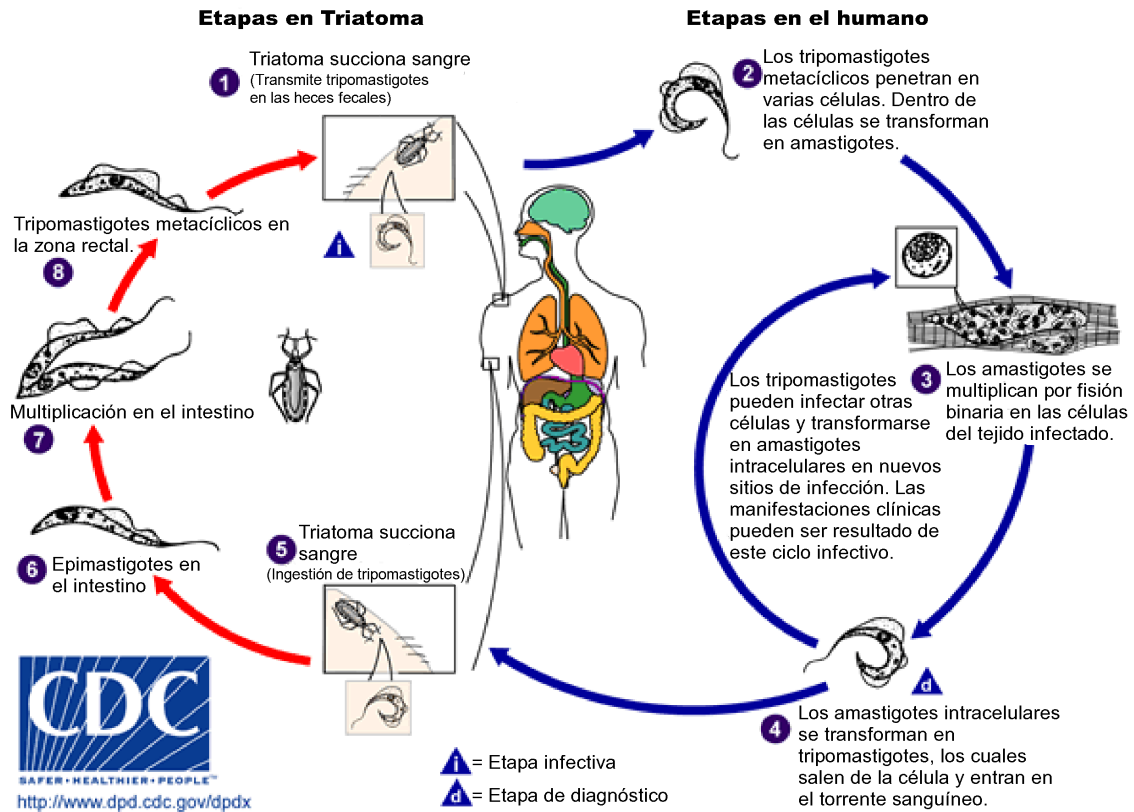


Figura 2.1. Ciclo de vida del *Trypanosoma cruzi*.

2.2. La cruzaina

Las cisteína proteasas toman parte en diferentes procesos biológicos. Por esta razón, una mala regulación de estas enzimas conduce a una gran variedad de enfermedades, representando un buen blanco terapéutico [8, 9]. El gran grupo de las cisteína proteasas se divide en 20 familias y se agrupa en tres clanes. El clan A comprende a las familias C1, C2, y C10 las cuales se describen como “proteasas tipo-papaína” [10]. Las cisteína proteasas que forman parte de este clan son consideradas blancos terapéuticos prometedores en contra de infecciones parasitarias, así como contra enfermedades sistémicas en humanos [8, 9].

La cruzaina pertenece al clan A de las cisteína proteasas y es un blanco terapéutico contra *Trypanosoma cruzi*, causante etiológico de la enfermedad de

Chagas. Esta enzima está codificada por diversos genes polimórficos organizados en tándem resultando en una gran variedad isoformas [9]. En general, la cruzaina es una glicoproteína de aproximadamente 60 kDa que se compone de una cadena polipeptídica de 215 aminoácidos, los cuales se pliegan en dos dominios. Al igual que en la papaína, un dominio consiste principalmente de α -hélices, denominado dominio L, y el otro de hojas β -plegadas con interacciones antiparalelas, llamado dominio R [11]. A pesar de las similitudes entre la cruzaina y la papaína, se pueden encontrar diferencias estructurales importantes entre las asas y los giros de las dos proteínas [7].

La cruzaina es considerada la cisteína proteasa de mayor importancia en el parásito y se localiza principalmente en un organelo pre-lisosomal llamado "reservosoma", el cual sólo es específico de los epimastigotes; sin embargo, estudios más recientes indican que también se puede localizar en la membrana plasmática [7]. Entre otras funciones, se ha reportado que esta enzima es esencial para la penetración del parásito en las células del hospedero, en la defensa contra el sistema inmune del organismo infectado y tiene un papel importante en la nutrición, además de que es importante para la diferenciación del microorganismo [7]. Debido al papel vital que juega la cruzaina para el parásito y a su participación en múltiples procesos, diferentes autores la han identificado como un nuevo blanco terapéutico [9].

2.3. Mecanismo catalítico de la cruzaina

El sitio activo de la cruzaina, al igual que el de otras cisteína proteasas del clan A, está compuesto por la triada catalítica Cys25, His162 y Asn182 [11]. El mecanismo propuesto para esta triada es el siguiente: la Asn182 actúa como un aceptor de puente de hidrógeno hacia la His162, esto permite la formación permanente del par iónico imidazolio-tiolato entre la Cys25 y la His162 [12]. Se ha propuesto que otro aminoácido importante en el sitio catalítico es la Gln19, la cual actúa como parte de un agujero de oxoanión que estabiliza el complejo acilo-

enzima mediante la formación de un puente de hidrógeno entre el grupo amino de la Gln19 y el sustrato [13].

Se han propuesto dos mecanismos catalíticos principales para explicar la actividad enzimática de las cisteína proteasas. El primero consiste en un mecanismo basado en las serina proteasas propuesto por Drenth *et al* [14] el cual se lleva a cabo en tres pasos (Figura 2.2). El primer paso es un ataque nucleofílico por parte del tiolato hacia el sustrato. El estado de transición de esta reacción se estabiliza mediante un puente de hidrógeno formado entre el sustrato y el NH de la cadena principal de la Cys25. Como productos se obtiene un intermediario tetraédrico, el cual es estabilizado por un puente de hidrógeno formado la Gln19 en el agujero de oxoanión y el ligando. Posteriormente, se efectúa la transferencia de un protón desde el ion imidazolico hacia el nitrógeno de la amida en el sustrato para formar el complejo acilo-enzima. Finalmente, existe un paso de desacilación durante el cual una molécula de agua ataca el átomo de carbono en el complejo acilo-enzima para producir los productos finales [14].

En un mecanismo alternativo se propone que la protonación del oxígeno o nitrógeno en el sustrato, causada por la transferencia de un protón desde la His162, ocurre antes o de manera concertada al ataque nucleofílico del tiolato en la Cys25 y sin la existencia de ningún intermediario [15-17]. Esto puede ser posible debido a que las serina proteasas deben superar la energía de activación mediante la estabilización del intermediario tetraédrico, mientras que las cisteína proteasas comienzan la reacción desde un estado basal de mayor energía lo que disminuye la energía de activación [8].

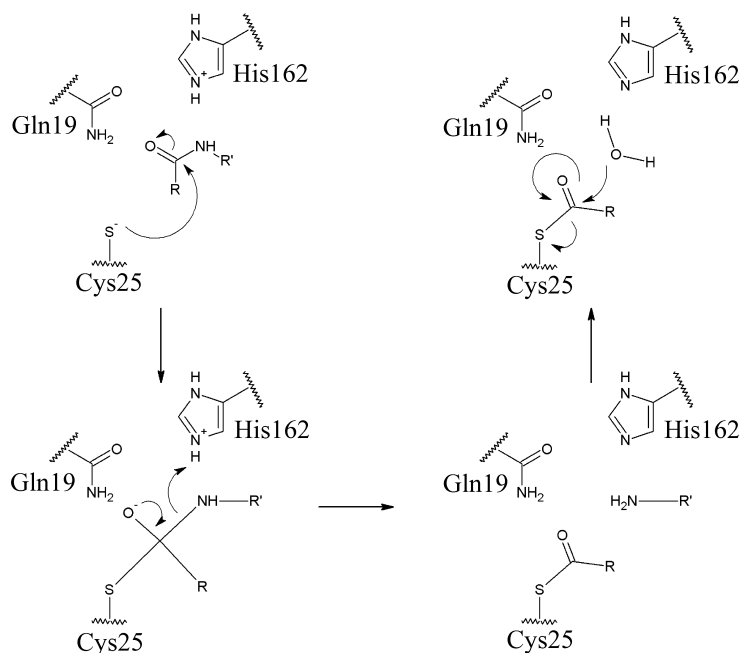


Figura 2.2. Mecanismo catalítico propuesto por Drenth *et al.*, para las cisteína proteasas [14].

2.4. Inhibidores de la cruzaina

Debido a la importancia de la cruzaina y de las cisteína proteasas, se han diseñado varias moléculas con la capacidad de inhibir su actividad catalítica. Entre ellas se encuentra un amplio rango de moléculas peptídicas y no peptídicas con un grupo susceptible al ataque nucleofílico por parte de la enzima. Por ejemplo, las moléculas con un grupo fluorofenoximetil cetona, vinilsulfonamida, epóxido, hidrazona, tiosemicarbazona, tiosemicarbazida, tiazolidona, o nitrilo han presentado buena actividad inhibiendo a la cruzaina [18-25]. A pesar de la gran diversidad de inhibidores de la cruzaina, el mecanismo de inhibición de algunos pocos ha sido estudiado experimentalmente o con cálculos teóricos [26-31]. En la Figura 2.3, se muestran las estructuras químicas de diversos inhibidores de la cruzaina. Los avances recientes en los métodos computacionales han facilitado los estudios acerca del mecanismo de inhibición de algunas moléculas, lo cual puede resultar en nuevas clases de inhibidores [32].

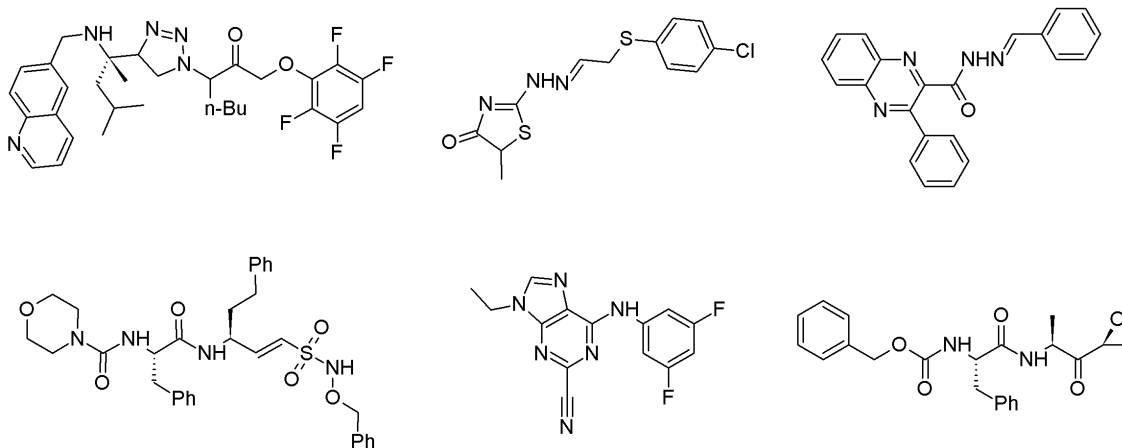


Figura 2.3. Estructuras químicas de algunos inhibidores de la cruzaña.

Se han realizado estudios sobre el mecanismo tripanocida de estas moléculas donde se reporta que la inhibición de la cruzaña aminora la invasión de las células del hospedero por parte del parásito; adicionalmente, bloquea la replicación del tripomastigote y la diferenciación de tripomastigote a amastigote. Otros resultados indican que los inhibidores irreversibles pertenecientes a la clase de las vinil sulfonas inducen la acumulación de una proforma de la cruzaña en el aparato de Golgi, resultando en la muerte del parásito [9].

Recientemente, Mott *et al.* sintetizaron una serie de triazino- y purino-carbonitrilos que se unen covalentemente a la cruzaña y la inhiben con Cl_{50} que van desde 2.23 μ M hasta 10 nM [25]. A pesar de que las actividades inhibitorias obtenidas son aceptables, aun se puede realizar una optimización de estos compuestos. Una técnica ampliamente utilizada para este fin es el desarrollo de modelos que relacionen cuantitativamente la estructura tridimensional de los compuestos con su actividad biológica (QSAR-3D) como un análisis comparativo de campos moleculares (CoMFA) o un análisis comparativo de índices de similitud molecular (CoMSIA).

El objetivo de este trabajo es realizar un estudio CoMFA y CoMSIA de los compuestos reportados por Mott *et al.* y, de esta manera, poder obtener más

información de las interacciones ligando-receptor que puedan llevar a la optimización de estas moléculas. Esta es la primera vez que se realiza un estudio QSAR-3D para la optimización de purino-carbonitrilos como inhibidores de la cruzafina [33]. En adición a esto, se utilizaron cálculos químico-cuánticos, a nivel semi-empírico y de teoría de funcionales de la densidad, para estudiar el perfil energético de la reacción por la cual los purino-carbonitrilos inhiben a la cruzafina [31]. Los resultados obtenidos de este estudio son de gran utilidad para entender más a fondo la interacción entre la cruzafina y los purino-carbonitrilos, además de que brindarán información valiosa para el diseño de nuevos inhibidores.

CAPÍTULO 3. METODOLOGÍA

3.1. Estudios QSAR-3D

3.1.1. Conjunto de datos

El conjunto de datos utilizado durante este estudio se conformó por 32 triazino- y purino-carbonitrilos cuyas estructuras y su actividad biológica fueron reportadas por Mott *et al.* anteriormente y se enlistan en el Apéndice A [25]. Este grupo de compuestos se dividió en un grupo de entrenamiento, a partir del cual se construyeron los modelos QSAR-3D, y un grupo de prueba usado para la validación estadística de los modelos. El grupo de prueba se eligió manualmente intentado abarcar el mayor rango de actividades y la mayor diversidad estructural posible. Para facilitar la generación de los modelos QSAR-3D todas las actividades biológicas se utilizaron como pCl_{50} ($-\log Cl_{50}$).

3.1.2. Preparación de ligandos y alineamiento molecular

Todas las moléculas fueron dibujadas utilizando como plantilla la estructura del inhibidor 6-[(3,5-difluorofenil)amino]-9-etil-9*H*-purino-2-carbonitrilo, el cual se encuentra co-cristalizado con la cruzaina y cuyas coordenadas fueron extraídas del Protein Data Bank (PDB) con número de acceso 3I06 [25]. Posteriormente, se preparó un cálculo de optimización de geometría utilizando el programa TRITON [34], el cual nos permitió minimizar la energía de los ligandos dentro de una cavidad formada por 20 aminoácidos del sitio activo de la cruzaina (Glu19, Cys22, Gly23, Ser24, Cys25, Trp26, Cys63, Ser64, Gly65, Gly66, Leu67, Met68, Val137, Ala138, Val139, Leu160, Asp161, His162, Gly163, Glu208). Finalmente, se realizó el cálculo de optimización de geometría con el programa MOPAC2009 utilizando el hamiltoniano PM6, el método de orbitales moleculares localizados MOZYME [35-39], y el modelo de solvatación continua COSMO con una constante dieléctrica de 78.4 D que corresponde a la del agua [40]. Durante el cálculo se fijaron las posiciones de todos los átomos de la cadena principal de la proteína; por el

contrario todos los átomos de las cadenas laterales, así como todos los del ligando fueron sometidos a una optimización de geometría utilizando el método BFGS [41-44].

Una vez terminados los cálculos, las moléculas se extrajeron de la cavidad utilizadas para la minimización de geometría para proceder al alineamiento molecular. En este trabajo de tesis se utilizaron dos hipótesis de alineamiento molecular debido a que esta etapa juega un papel fundamental en la obtención de modelos QSAR-3D que sean estadísticamente aceptables y con alto poder predictivo. La primera hipótesis de alineamiento utilizada en este trabajo consistió en traslapar el grupo imino de todas las moléculas, el cual corresponde a una subestructura que comparten todos los compuestos (Figura 3.1 A). La segunda hipótesis consistió en rotar 180° el enlace que une el anillo de purina y el grupo nitrilo en 5 compuestos de la base de datos con la finalidad de que el grupo fenilo de todos los compuestos coincidiera con el del inhibidor 6-[(3,5-difluorofenil)amino]-9-etil-9*H*-purino-2-carbonitrilo dentro una cavidad cercana a la Cys25 en el sitio activo de la cruzaina. Después de esta modificación, las moléculas se sometieron a una segunda optimización de geometría siguiendo el procedimiento mencionado anteriormente y, por último, se alinearon utilizando el grupo imino nuevamente (Figura 3.1 B).

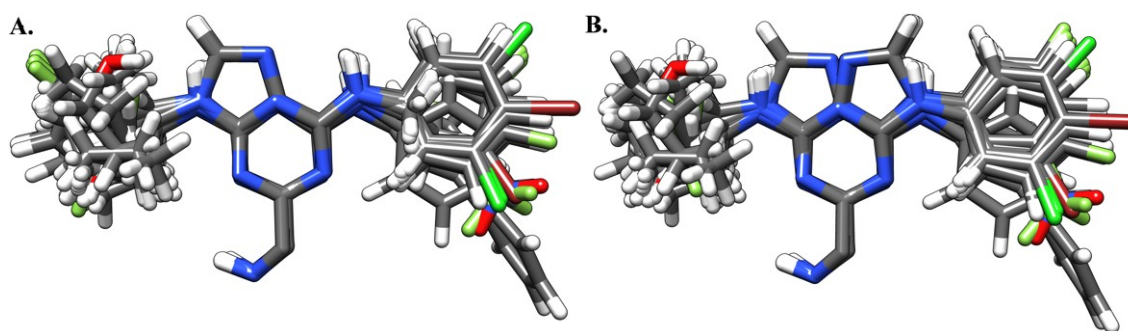


Figura 3.1. Alineación de las moléculas usando diferentes hipótesis.

3.1.3. Análisis comparativo de campos moleculares (CoMFA) y análisis comparativo de índices de similitud molecular (CoMSIA)

Para el cálculo de los descriptores CoMFA y CoMSIA se utilizó el programa SYBYL8.0 [45] para generar una malla tridimensional con separaciones de 2 Å, la cual se extendió 4 Å más allá de las moléculas. También se generaron modelos utilizando la misma malla con separaciones de 1 Å. Los descriptores CoMFA consisten en calcular las interacciones estéricas y electrostáticas. Las primeras se calculan entre la molécula y un átomo de carbono sp³ con un radio de van der Waals de 1.52 Å colocado en cada una de las intersecciones de la malla utilizando el potencial de Lennard-Jones. De igual forma, las interacciones electrostáticas son calculadas con un potencial coulombico entre la molécula y una carga +1 localizada en las intersecciones de la malla. Para estos cálculos se utilizó un valor de corte igual a 30 kcal/mol. Los descriptores CoMSIA fueron calculados utilizando la misma hipótesis de alineamiento con la cual se obtuvo el mejor modelo CoMFA. Así mismo, se utilizó una malla con espaciamiento de 2 Å para calcular cinco índices de similitud molecular: estérico (S), electrostático (E), hidrofóbico (H), donador de puente de hidrógeno (D) y aceptor de puente de hidrógeno (A). En este caso se utilizó un factor de atenuación igual a 0.3 para la función gaussiana de los índices de similitud.

3.1.4. Análisis estadístico y validación

Se utilizó el método de mínimos cuadrados parciales (PLS) para poder elaborar los modelos CoMFA y CoMSIA. En ambos casos se utilizaron los descriptores como variable independiente y el valor de pCl₅₀ como variable dependiente. Una vez que se obtuvieron los modelos, se les realizó una validación cruzada con el método *Leave-One-Out (LOO)* para determinar el número óptimo de componentes con el fin de evitar un sobreajuste de los resultados. Se eligieron los mejores modelos tomando como parámetro los coeficientes de correlación (r^2) y de correlación en la validación cruzada (q^2) más altos.

El coeficiente de correlación en la validación cruzada se calculó de la siguiente manera:

$$q^2 = 1 - \frac{\sum(Y_{pred} - Y_{exp})^2}{\sum(Y_{pred} - Y_{promedio})^2}$$

donde Y_{pred} , Y_{exp} y $Y_{promedio}$ son los valores de la actividad predicha, experimental y promedio, respectivamente.

Se realizó una validación adicional al utilizar todos los modelos CoMFA y CoMSIA para predecir la actividad biológica de un grupo de prueba externo que consistió de 6 moléculas. Las predicciones se examinaron con un análisis de regresión aplicando los criterios propuestos por Golbraikh y Tropsha [46]. En términos generales, los criterios propuestos por Golbraikh y Tropsha evalúan qué tan grande es la diferencia entre un modelo “ideal” y el modelo que se generó. Para que un modelo sea aceptable debe satisfacer los criterios que se enlistan a continuación:

1. $q^2 > 0.5$
2. $R^2 > 0.6$
3. $[(R^2 - R_o^2)/R^2] < 0.1$ ó $[(R'^2 - R'o^2)/R'^2] < 0.1$
4. $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$

donde q^2 es el coeficiente de correlación en la validación cruzada, R^2 es el coeficiente de correlación de actividad experimental (y) vs actividad predicha (\tilde{y}) de las moléculas en el grupo de prueba; R_o^2 y $R'o^2$ son los coeficientes de correlación para la regresión a través del origen para y vs \tilde{y} y \tilde{y} vs y , respectivamente; k y k' son las pendientes de la regresión a través del origen $y^o = k\tilde{y}$ and $\tilde{y}^o = k'y$.

3.2. Estudio del mecanismo de reacción

3.2.1. Cálculos utilizando la Teoría de Funcionales de la Densidad (DFT)

Se realizaron cálculos utilizando la Teoría de Funcionales de la Densidad con la finalidad de estudiar el mecanismo de reacción por el cual se unen los inhibidores covalentemente a la cruzafina. Para estos cálculos se empleó un modelo aislado que consistió de las cadenas laterales de dos aminoácidos importantes para el mecanismo (Cys25 e His162) y un purino-carbonitrilo que se utilizó como modelo del inhibidor (Figura 3.2). Durante estos cálculos se fijaron los grupos metilo que corresponden a los carbonos α de los aminoácidos.

Se realizó una búsqueda a través de la superficie de energía potencial correspondiente a la inhibición de la cruzafina por un purino-carbonitrilo utilizando el funcional híbrido B3LYP [47, 48] con el conjunto de base 6-31+G(d). El estado estacionario correspondiente al estado de transición fue caracterizado por una frecuencia imaginaria resultante del cálculo de frecuencias vibracionales armónicas. De igual manera los estados estacionarios correspondientes a los reactivos y productos fueron confirmados por la ausencia de frecuencias imaginarias. Con la finalidad de confirmar la conexión del estado de transición con los reactivos y productos, se efectuaron cálculos de coordenada intrínseca de reacción (IRC) utilizando el algoritmo HPC y un paso de 0.13 Bohr. Posteriormente, se calculó la energía de cada punto resultante del cálculo IRC con un punto simple a un nivel de teoría M05-2X/6-31+G(d) tal y como fue propuesto por Goodman *et al.* [49]. Es importante mencionar que durante todos los cálculos se empleó un modelo continuo polarizable (PCM) aplicando la variante del formalismo de la ecuación integral (IEFPCM) y utilizando agua como solvente. Esto debido a que el sitio catalítico de la cruzafina se encuentra en una región externa de la proteína y, por consiguiente, es altamente accesible para el solvente. Todos los cálculos se realizaron con el programa Gaussian 09 para el cálculo de propiedades moleculares [50].

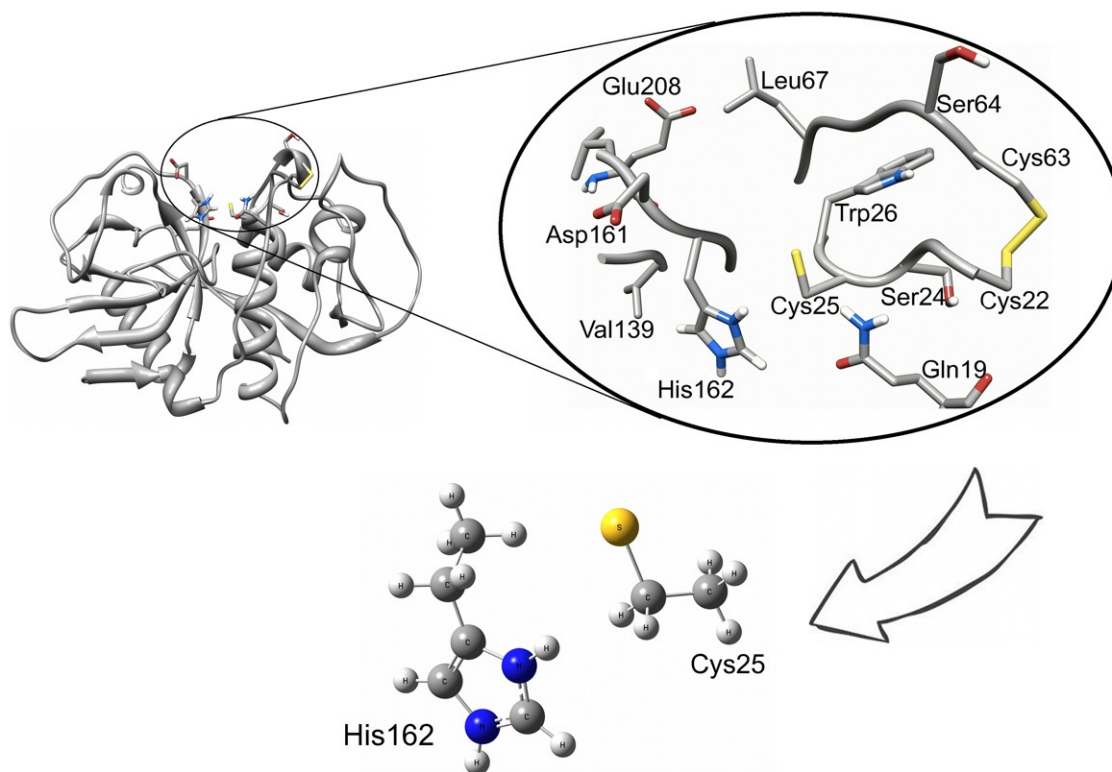


Figura 3.2. Modelos utilizados en los cálculos.

Los índices de Fukui [51] se calcularon con la finalidad de evaluar sitios reactivos en la molécula. Estos índices están definidos de la siguiente manera:

$$f^{-}(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_v^{-} = \rho_{N_0}(\mathbf{r}) - \rho_{N_0-1}(\mathbf{r}),$$

$$f^{+}(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_v^{+} = \rho_{N_0+1}(\mathbf{r}) - \rho_{N_0}(\mathbf{r}),$$

y

$$f^0(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_v^0 = \frac{1}{2} (f^{-}(\mathbf{r}) + f^{+}(\mathbf{r})) = \frac{1}{2} (\rho_{N_0+1}(\mathbf{r}) - \rho_{N_0-1}(\mathbf{r})),$$

donde ρ_{N_0-1} , ρ_{N_0} y ρ_{N_0+1} son las densidades electrónicas de los sistemas con $N_0 - 1$, N_0 y $N_0 + 1$ electrones, respectivamente. La función $f^-(\mathbf{r})$ define la región donde la remoción de una fracción de electrón genera un cambio energéticamente favorable; de forma contraria, $f^+(\mathbf{r})$ define la región donde la adición es más favorable. En términos de reactividad, un buen nucleófilo presentaría valores grandes de $f^-(\mathbf{r})$, mientras las regiones susceptibles a un ataque nucleofílico estarían caracterizadas por valores grandes de $f^+(\mathbf{r})$ [52].

3.2.2. Cálculos a nivel semi-empírico

Se realizaron cálculos a nivel semi-empírico con la finalidad de estudiar las interacciones entre el inhibidor y el sitio activo de la cruzaína. Para estos cálculos se utilizó la estructura cristalográfica de la cruzaína unida al inhibidor 6-[(3,5-difluorofenil)amino]-9-etil-9*H*-purino-2-carbonitrilo (PDB ID: 3I06). Se mapeó la superficie de energía potencial variando dos coordenadas de reacción simultáneamente: las distancias C2-S y N1-H. Estos átomos se aproximaron gradualmente y se realizó una minimización de energía cada 0.05 Å. Esta metodología ha sido utilizada anteriormente para estudiar diversos mecanismos de reacción enzimáticos [53-56].

Los cálculos se realizaron utilizando explícitamente 32 átomos del inhibidor y 20 aminoácidos del sitio activo de la cruzaína (Glu19, Cys22, Gly23, Ser24, Cys25, Trp26, Cys63, Ser64, Gly65, Gly66, Leu67, Met68, Val137, Ala138, Val139, Leu160, Asp161, His162, Gly163, Glu208), los cuales suman 275 átomos y se muestran en la Figura 3.2. Las coordenadas de los átomos correspondientes a la cadena principal de los aminoácidos se fijaron durante todos los cálculos; de manera contraria, aquellas correspondientes a las cadenas laterales y al inhibidor se optimizaron mediante el método BFGS [41-44].

Se utilizó el programa TRITON [34] como interfaz gráfica para preparar los cálculos, los cuales se efectuaron con el programa MOPAC2009 empleando el hamiltoniano semi-empírico PM6 para todos los átomos y el método de orbitales

moleculares localizados MOZYME [35-39]. Adicionalmente, se utilizó el modelo de solvatación continua COSMO con una constante dieléctrica de 78.4 D que corresponde a la del agua durante todos los cálculos realizados a nivel semi-empírico.

CAPÍTULO 4. RESULTADOS

4.1. Estudios QSAR-3D

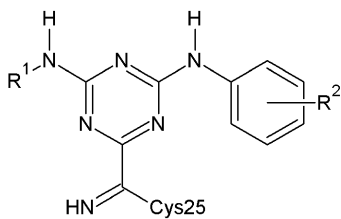
4.1.1. Modelos CoMFA y CoMSIA

Los modelos QSAR-3D reportados en este trabajo de tesis se realizaron con una base de datos compuesta por 32 triazino- y purino-carbonitrilos cuyas estructuras, su Cl_{50} (μM) y su pCl_{50} ($-\log Cl_{50}$) se enlistan en la Tabla 4.1 [25]. Se construyeron diversos modelos CoMFA con la finalidad de obtener uno en el cual la conformación de los ligandos sea la más cercana a la esperada en el sitio de unión. Para esto, se utilizaron las diferentes alineaciones mencionadas en la sección de Metodología 3.2.1.

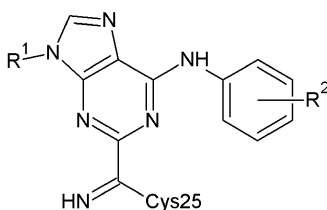
El modelo 1 se elaboró con la alineación mostrada en la Figura 3.1A en la cual todos las moléculas fueron orientadas tomando como base el anillo de purina del compuesto **23** el cual fue co-cristalizado unido a la cruzaina. En términos generales, los resultados obtenidos en el modelo 1 son aceptables para un estudio QSAR-3D con una q^2 de 0.623 y una r^2 de 0.983. Los resultados estadísticos de todos los modelos se muestran en la Tabla 4.2.

El modelo 2 se construyó utilizando un alineamiento alternativo en el cual los grupos fenilo de todas las moléculas fueron orientados hacia el mismo lado del anillo de purina (Figura 3.1B). Esta alineación también está basada en la estructura del compuesto **23** co-cristalizado con la cruzaina. En esta estructura se puede observar que el grupo fenilo del compuesto **23** se encuentra dentro de una cavidad la cual es análoga a la cavidad S_2 de la papaína, otra cisteína proteasa con un alto nivel de homología a la cruzaina. Evidencia experimental sugiere que la cavidad S_2 tiene una alta afinidad por residuos hidrofóbicos, especialmente por la fenilalanina [57], razón por la cual también podría tener afinidad por el grupo fenilo de las moléculas en la base de datos. En términos generales el modelo 2 mostró resultados estadísticos muy similares al modelo 1, esto se puede observar al comprar el valor de r^2 (0.980). Sin embargo, este modelo presentó mejores

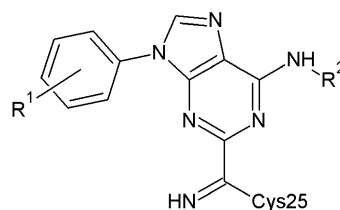
Tabla 4.1. Estructuras químicas y valores de pCl₅₀ de las moléculas usadas en este estudio



Compuestos 1-22



Compuestos 23-27



Compuestos 28-32

Compuesto	R ¹	R ²	Cl ₅₀ (μM)	pCl ₅₀
1	etilo	-	0.710	6.1487
2	ciclopentilo	-	0.710	6.1487
3	propilo	-	0.790	6.1024
4	isopropilo	-	0.890	6.0506
5*	metilciclopropilo	-	0.890	6.0506
6	isobutilo	-	0.890	6.0506
7	butilo	-	0.890	6.0506
8	4-hidroxibutilo	-	0.890	6.0506
9	metilo	-	1.260	5.8996
10*	ciclohexilo	-	1.410	5.8508
11	2-hidroxietilo	-	1.580	5.8013
12	ciclopentilo	3-nitro	0.063	7.2006
13*	ciclopentilo	3-fluoro	0.079	7.1024
14	ciclopentilo	3-cloro	0.079	7.1024
15	ciclopentilo	3-bromo	0.100	7.0000
16	ciclopentilo	4-fluoro	0.126	6.8996
17	ciclopentilo	3-metilo	0.126	6.8996
18	ciclopentilo	4-bromo	1.122	5.9500
19	ciclopentilo	3-fenilo	2.239	5.6499
20	ciclopentilo	3,5-difluoro	0.063	7.2007
21*	ciclopentilo	3,5-dicloro	0.398	6.4001
22	2,2-difluoroetilo	3,5-difluoro	0.025	7.6021
23	etilo	3,5-difluoro	0.010	8.0000
24*	2,2-difluoroetilo	3,5-difluoro	0.013	7.8860
25	ciclopentilo	3,5-difluoro	0.018	7.7447
26	etilo	3-cloro	0.018	7.7447
27	ciclopentilo	3-cloro	0.040	7.3979
28	3,5-difluoro	2,2-difluoroetilo	0.050	7.3010
29	3-cloro	etilo	0.063	7.2007
30	3-cloro	2,2-difluoroetilo	0.071	7.1487
31*	3,5-difluoro	etilo	0.251	6.6003
32	3,5-difluoro	ciclopentilo	0.316	6.5003

* Compuestos utilizados en el grupo de prueba.

Tabla 4.2. Resumen de los modelos CoMFA 1-3 y de los modelos CoMSIA utilizando los campos: SED, SEA, SEDA y SEHDA

Estadísticos	CoMFA			CoMSIA			
	Modelo 1	Modelo 2	Modelo 3	SED	SEA	SEDA	SEHDA
Alineamiento	A	B	B	B	B	B	B
Espaciamiento (Å)	2.0	2.0	1.0	2.0	2.0	2.0	2.0
q^{2a}	0.623	0.736	0.692	0.627	0.618	0.565	0.460
PRESS ^b	0.488	0.408	0.441	0.473	0.479	0.523	0.533
r^{2c}	0.983	0.980	0.977	0.888	0.874	0.901	0.712
S^d	0.104	0.112	0.122	0.259	0.275	0.249	0.389
F^e	227.096	196.155	166.508	41.591	36.544	36.586	59.474
R^{2f}	0.856	0.763	0.753	0.806	0.833	0.815	0.758
NOC ^g	5	5	5	4	4	5	1
Contribución:							
Estérica	0.815	0.816	0.804	0.367	0.461	0.336	0.159
Electrostática	0.185	0.184	0.196	0.259	0.268	0.235	0.080
Hidrofóbica							0.307
Donador de puente de H				0.374		0.347	0.304
Aceptor de puente de H					0.271	0.082	0.150

^a Coeficiente de correlación de la validación cruzada LOO.

^b Error estándar de las predicciones obtenidas con el método LOO.

^c Coeficiente de correlación r^2 .

^d Error estándar estimado.

^e Valor de la prueba de Fisher.

^f Coeficiente de correlación obtenido de la predicción del grupo de prueba.

^g Número óptimo de componentes principales.

resultados en la validación cruzada donde se obtuvo un valor de q^2 igual a 0.736. Estos resultados sugieren que ambas hipótesis de alineamiento son lo suficientemente aceptables para generar un modelo CoMFA altamente confiable. Sin embargo, el valor de q^2 del modelo 2 indica un mayor poder predictivo aunado al hecho de que su alineamiento está basado en evidencia experimental. Por estas razones se eligió al modelo 2 como base para generar los modelos QSAR-3D subsecuentes.

Debido a que el cálculo de descriptores CoMFA y CoMSIA se realiza utilizando una malla tridimensional, se evaluó el espaciado de la malla como otra posible

variable durante la generación de los modelos. Recientemente, Mittal *et al.* realizó un análisis de diferentes modelos QSAR-3D con la finalidad de estudiar la influencia de los parámetros utilizados al calcular los descriptores en el poder predictivo de un modelo CoMFA [58]. Sus resultados indican que un valor alto de q^2 generalmente se encuentra asociado con un espaciamiento de la malla pequeño. Basado en este estudio se generó el modelo 3, en el cual se consideró la misma hipótesis de alineamiento que en el modelo 2, pero con un espaciamiento de la malla de 1 Å. Los parámetros estadísticos resultantes del modelo 3 ($q^2 = 0.692$ and $r^2 = 0.977$) fueron menores a los obtenidos con el modelo 2. Por esta razón, el modelo 2 fue elegido como el mejor modelo CoMFA. La predicción de la actividad biológica de las moléculas en el grupo de entrenamiento y en el grupo de prueba utilizando el modelo 2 se encuentran enlistadas y graficadas en la Tabla 4.3 y en la Figura 4.1, respectivamente.

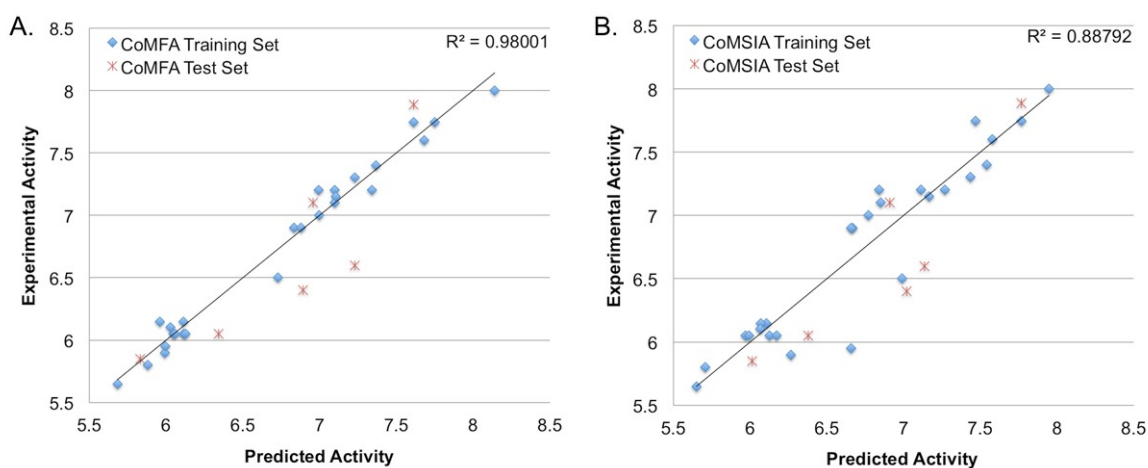


Figura 4.1. Graficas que muestran la correlación entre los valores experimentales vs. predichos del (A) modelo 2 CoMFA y del (B) modelo CoMSIA incluyendo los campos SED.

Tabla 4.3. Predicción de la actividad biológica utilizando el modelo 2 CoMFA y el modelo CoMSIA incluyendo los campos SED

Compuesto	pCl ₅₀	CoMFA		CoMSIA	
		Predicción	Residual	Predicción	Residual
1	6.1487	6.1119	0.0368	6.107	0.0417
2	6.1487	5.9593	0.1894	6.0697	0.0791
3	6.1024	6.0296	0.0728	6.0662	0.0362
4	6.0506	6.0463	0.0043	5.9688	0.0818
5*	6.0506	6.3420	-0.2914	6.3777	-0.3271
6	6.0506	6.0562	-0.0056	6.1253	-0.0747
7	6.0506	6.1141	-0.0635	6.1739	-0.1233
8	6.0506	6.125	-0.0744	5.9938	0.0568
9	5.8996	5.9918	-0.0922	6.2667	-0.3671
10*	5.8508	5.8323	0.01844	6.0114	-0.1606
11	5.8013	5.8808	-0.0795	5.7081	0.0933
12	7.2006	7.3393	-0.1387	7.1142	0.0864
13*	7.1024	6.9560	0.14637	6.9091	0.1932
14	7.1024	7.0989	0.0035	6.8515	0.2509
15	7.0000	6.9979	0.0021	6.7704	0.2296
16	6.8996	6.8324	0.0672	6.6669	0.2327
17	6.8996	6.8793	0.0204	6.6588	0.2408
18	5.9500	5.9941	-0.0441	6.6577	-0.7077
19	5.6499	5.6839	-0.0339	5.6500	-0.0001
20	7.2007	7.0986	0.1021	7.2680	-0.0674
21*	6.4001	6.8935	-0.4934	7.0180	-0.6179
22	7.6021	7.6803	-0.0782	7.5783	0.0238
23	8.0000	8.1405	-0.1405	7.9463	0.0537
24*	7.8860	7.6108	0.2752	7.7680	0.1180
25	7.7447	7.6108	0.1339	7.768	-0.0233
26	7.7447	7.7499	-0.0051	7.4673	0.2775
27	7.3979	7.3678	0.0302	7.5408	-0.1429
28	7.3010	7.2295	0.0716	7.4335	-0.1325
29	7.2007	6.9939	0.2067	6.8395	0.3611
30	7.1487	7.1043	0.0444	7.1660	-0.0173
31*	6.6003	7.2303	-0.6300	7.1365	-0.5361
32	6.5003	6.73	-0.2297	6.9896	-0.4893

*Compuestos utilizados en el grupo de prueba.

Los modelos CoMSIA fueron construidos con base en el modelo 2; es decir, se utilizó el alineamiento mostrado en la Figura 3.2B y se empleó un espaciado de la malla de 2 Å. Se calcularon cinco campos: estérico (S), electrostático (E), hidrofóbico (H), donador de puentes de hidrógeno (D) y aceptor de puentes de

hidrógeno (A), y se evaluaron las 31 posibles combinaciones entre estos (Figura 4.2). Las combinaciones SED, SEA y SEDA fueron aquellas que presentaron los valores de q^2 más altos (0.627, 0.618 y 0.565, respectivamente), mientras que la combinación SEHDA presentó un valor de $q^2 < 0.5$. Este es un resultado interesante debido a que muestra cómo el campo H no está correlacionado con la actividad biológica. Por esta razón, la relación estructura-actividad de esta base de datos no puede ser explicada con base en la hidrofobicidad de los compuestos. Sin embargo, debido a la similitud que existe entre la cavidad S_2 de la papaína y la cruzaina, sería interesante enriquecer la base de datos con inhibidores que tengan una mayor diversidad estructural (principalmente modificaciones que afecten la hidrofobicidad) en la posición del fenilo. Un detalle interesante en la Figura 4.2 es que la mayoría de las combinaciones con $q^2 > 0.5$ contienen a los campos estérico y/o electrostático, descriptores similares a los utilizados para generar los modelos CoMFA, lo cual podría explicar parcialmente la similitud en los resultados obtenidos con estas dos metodologías.

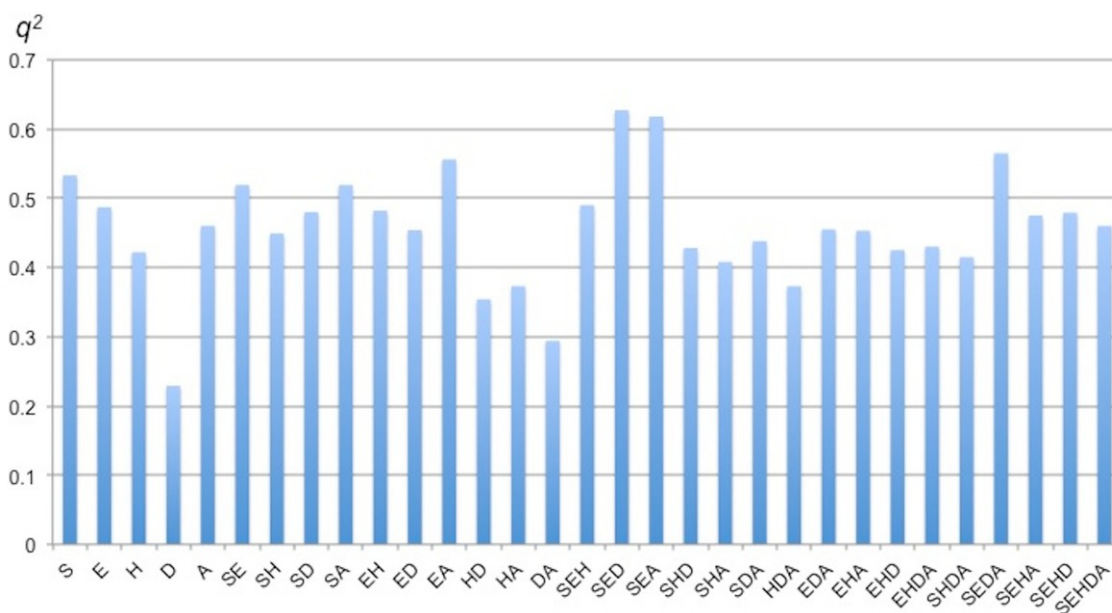


Figura 4.2. Histograma que presenta los valores de q^2 para las 31 posibles combinaciones de los campos utilizados en los modelos CoMSIA.

4.1.2. Validación de los modelos CoMFA y CoMSIA

A pesar de que el valor de q^2 se utiliza frecuentemente como parámetro para evaluar el capacidad predictiva de un modelo QSAR, se ha observado que un valor alto ($q^2 > 0.5$) es una condición necesaria, pero no suficiente para asegurar el poder predictivo y la robustez de un modelo. Por esta razón, Golbraikh y Tropsha desarrollaron una metodología para evaluar modelos QSAR [46]. Esta consiste en medir la diferencia entre las actividades biológicas predichas por el modelo QSAR que se evalúa y aquellas hipotéticamente determinadas por un modelo QSAR ideal. Esta metodología de validación se utilizó para evaluar a todos los modelos realizados en este trabajo de tesis con base en las predicciones de la actividad biológica de los compuestos en el grupo de prueba. Los resultados de la validación se encuentran enlistados y graficados en la Tabla 4.4 y en la Figura 4.3, respectivamente.

Tabla 4.4. Resumen de la validación para los modelos 1-3 CoMFA y los modelos CoMSIA utilizando los campos: SED, SEA, SEDA y SEHDA

Datos estadísticos	CoMFA			CoMSIA			
	Modelo 1	Modelo 2	Modelo 3	SED	SEA	SEDA	SEHDA
R^{2a}	0.856	0.763	0.753	0.806	0.833	0.815	0.758
R_0^{2b}	0.705	0.654	0.540	0.676	0.695	0.725	0.553
k^c	1.028	1.021	1.030	1.030	1.026	1.024	1.026
$(R^2 - R_0^2) / R^2$	0.176	0.142	0.282	0.161	0.166	0.110	0.269
$R_0'^{2d}$	0.829	0.761	0.739	0.795	0.815	0.810	0.744
k^e	0.971	0.976	0.968	0.968	0.972	0.975	0.972
$(R^2 - R_0'^2) / R^{2g}$	0.032	0.002	0.018	0.013	0.022	0.006	0.017

^a Coeficiente de correlación derivado de las predicciones de las moléculas en el grupo de prueba.

^{b,d} Coeficiente de correlación derivado de la regresión a través del origen para las actividad experimental vs. predicha y predicha vs. experimental, respectivamente.

^{c,e} Pendientes derivadas de la regresión a través del origen para las actividad experimental vs. predicha y predicha vs. experimental, respectivamente.

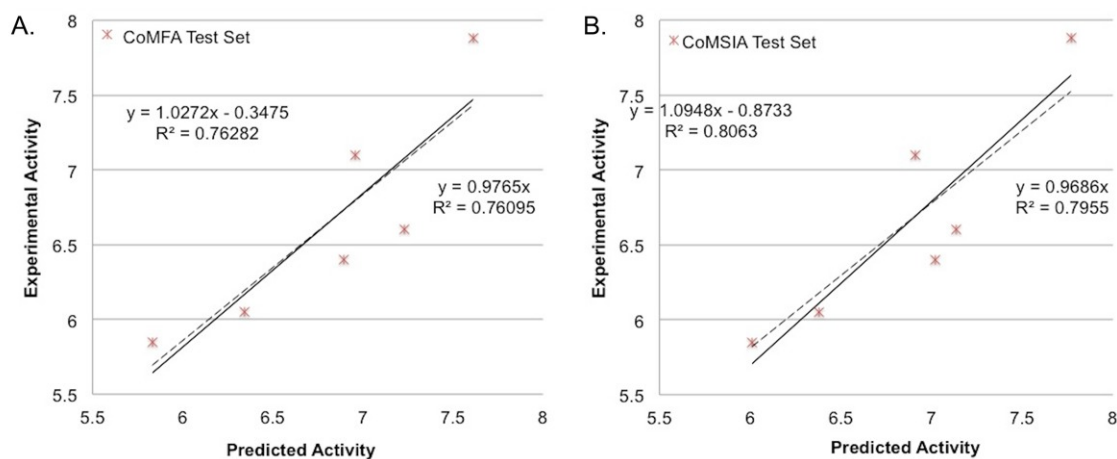


Figura 4.3. Resultados de la validación utilizando el conjunto de prueba para (A) el modelo 2 CoMFA y (B) modelo CoMSIA incluyendo los campos SED.

En términos generales, todos los modelos CoMFA y CoMSIA predijeron de manera aceptable las actividades de las moléculas en el grupo de prueba, lo cual se refleja en valores de $R^2 > 0.75$. Analizando detenidamente los resultados de la validación para los modelos CoMFA se puede observar que el valor de R^2 para el modelo 1 es mayor que el presentado por el modelo 2 (0.856 y 0.763, respectivamente); esto indica que el modelo 1 predice de mejor manera las actividades de las moléculas en el grupo de prueba. No obstante, todos los modelos cumplen con los criterios de $[(R^2 - R_0^2)/R^2]$ y k (descritos en la sección 3.1.4) por lo que se puede decir que tiene un poder predictivo aceptable. Cabe destacar que el modelo 2 presentó los mejores resultados de todos los modelos CoMFA; es decir, los valores de $[(R^2 - R_0^2)/R^2]$ y k más cercanos a 1. Por esta razón, aunado al hecho de que la hipótesis de alineamiento está basada en evidencia experimental, el modelo 2 se eligió como el mejor modelo CoMFA.

Los resultados de la validación para los modelos CoMSIA muestran una situación similar. El modelo que contiene los campos SEDA cumple de mejor manera los criterios $[(R^2 - R_0^2)/R^2]$ y k ; sin embargo, no presenta los valores más altos de R^2 . Este hecho puede ser explicado en términos de la baja diversidad molecular que existe en grupo de prueba. Es probable que al utilizar un grupo de prueba más numeroso y con mayor diversidad molecular, el valor de R^2 se correlacione con el

valor de q^2 . Un resultado que llama la atención es que los valores de R^2 y de q^2 que presentan todos los modelos CoMSIA son superiores a los presentados por el modelo 2 realizado con la metodología CoMFA. Esta diferencia sugiere que los modelos CoMSIA construidos en este estudio tienen un mayor poder predictivo que los modelos CoMFA.

4.1.3. Mapas de contorno obtenidos utilizando CoMFA

Los mapas de contorno, junto con las predicciones de actividad biológica, son parte de los resultados más interesantes que se obtienen con las metodologías QSAR 3D. Los mapas de contorno sirven de guía en el proceso de optimización de moléculas líder debido a que indican las regiones de una molécula donde ciertas modificaciones estructurales aumentan o disminuyen la actividad biológica. En este estudio se analizaron los mapas de contorno obtenidos en el modelo 2 con el propósito de obtener información relevante sobre las características estructurales de los purino-carbonitrilos asociadas con la actividad biológica. En esta sección se aplicó un filtrado de columna de 2 kcal/mol sobre el modelo 2 con la finalidad de obtener mapas mejor definidos y para facilitar el análisis. Es importante mencionar que el poder predictivo del modelo 2 no se afectó al realizar esta modificación.

Normalmente, la metodología CoMFA genera dos tipos de mapas de contorno: los estéricos y los electrostáticos. En este trabajo los mapas de contorno estéricos se denotan por poliedros verdes, en regiones donde un grupo voluminoso favorece la actividad biológica, y amarillos, si un grupo voluminoso disminuye la actividad. De igual manera, los mapas de contorno electrostáticos están representados por poliedros rojos o azules en regiones donde un potencial negativo aumenta o disminuye la actividad biológica, respectivamente. La Figura 4.4 muestra los mapas de contorno estéricos y electrostáticos obtenidos con el modelo 2. En esta figura, el compuesto más activo (**23**) está representado por bolas y tubos, mientras que el resto de los compuestos están representados por líneas.

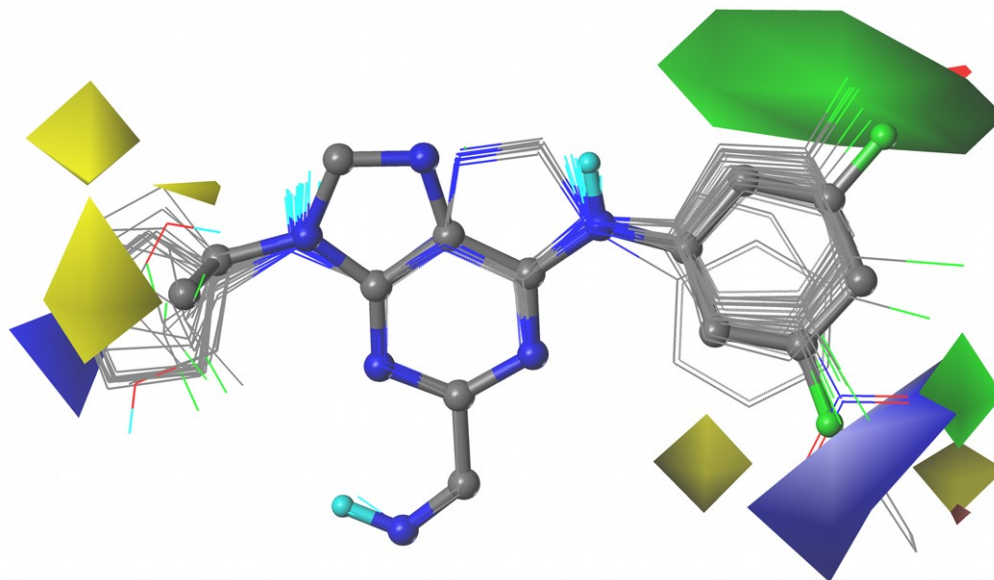


Figura 4.4. Mapas de contorno CoMFA generados con el modelo 2. Los contornos verdes y amarillos denotan regiones favorecidas o desfavorecidas, respectivamente, por grupos voluminosos. Los contornos rojos y azules están asociados con regiones donde un potencial negativo aumenta o disminuye la actividad biológica, respectivamente.

Al analizar la Figura 4.4 se pueden observar dos contornos verdes cerca de las posiciones 3 y 5 del fenilo. Este resultado puede indicar que los halógenos en estas posiciones incrementan la actividad biológica. De hecho, las moléculas **12-32**, de las cuales la mayoría tiene grupos voluminosos en estas posiciones, presentan una actividad biológica aceptable con una $pCl_{50} > 6.4$. En contraste, las moléculas **1-11**, las cuales no tienen sustituciones en el grupo fenilo, presentan una actividad biológica considerablemente baja ($pCl_{50} < 6.2$). Una excepción a lo mencionado anteriormente es el compuesto **19**, el cual tiene como sustituyente un 3-fenilo que se encuentra cerca de un contorno amarillo. Este contorno amarillo indica que un grupo voluminoso en esta región, por ejemplo el 3-fenilo, está asociado con una reducción en la actividad biológica y concuerda con la baja actividad presentada por el compuesto **19** (pCl_{50} 5.6499). Adicionalmente, este contorno nos da información acerca de la cavidad S_2 donde se va a contener al grupo fenilo. Con base en el conjunto de compuestos utilizados en este estudio se puede concluir que la cavidad S_2 es lo suficientemente grande para contener

grupos como el 3,5-difluorofenilo o el 3,5-diclorofenilo (compuestos **20** y **21**, respectivamente), pero no para albergar un grupo tan voluminoso como el 3-bifenilo del compuesto **19**.

En la Figura 4.4 también se pueden observar otros contornos amarillos cerca del ciclopentilo que tienen las moléculas **12-21**, **25**, **27** y **32** en la posición 9 del anillo de purina. En esta área, los grupos voluminosos están asociados con una disminución en la actividad. Un ejemplo de esta observación se obtiene al comparar los compuestos **25**, **27** y **32** con las moléculas **23**, **26** y **31**, respectivamente. En todos los casos las moléculas que tienen un etilo en lugar de un ciclopentilo en esta posición son más activas. Es posible que este efecto esté relacionado con algún impedimento estérico al momento en que los inhibidores se aproximan a su sitio de unión en la cruzaina.

Los mapas de contorno electrostáticos generados por la metodología CoMFA para el modelo 2 se pueden observar de color azul y rojo en la Figura 4.4. En esta figura, resalta un contorno azul de gran tamaño cerca de la posición 3 del fenilo y está relacionado, principalmente con el grupo 3-fenilo del compuesto **19**, el compuesto con la actividad más baja de toda la base de datos. Este contorno azul atraviesa por el centro al grupo fenilo, exactamente donde se concentra una alta densidad electrónica debido a los electrones π del fenilo. Lo que indica, una vez más, que este sustituyente puede ser el responsable de la baja actividad biológica presentada por el compuesto **19**.

4.1.4. Mapas de contorno obtenidos utilizando CoMSIA

La metodología CoMSIA se caracteriza por generar mapas de contorno mejor definidos que CoMFA. Esta circunstancia surge del hecho de que los índices de similitud evita el uso de valores de corte definidos arbitrariamente. Por consiguiente los descriptores pueden ser calculados en todos los puntos de la malla incluyendo aquellos que se encuentran muy cerca de las moléculas. En la

Figura 4.5 se muestran los mapas de contorno generados por CoMSIA para el modelo que incluye los campos SEDA; una vez más el compuesto más activo de la base de datos (**23**) se encuentra representado por bolas y tubos, mientras que el resto de los compuestos están representados por líneas. En la Figura 4.5A se muestran los mapas de contorno estéricos los cuales no presentan diferencias significativas respecto a los generados con CoMFA. En otras palabras, se observa un contorno verde (regiones donde grupos voluminosos favorecen la actividad) sobre la posición 3 del grupo fenilo y un contorno amarillo (regiones donde grupos voluminosos disminuyen la actividad) por encima del grupo ciclopentilo. No obstante, se pueden observar algunas diferencias en cuanto a los mapas de contorno electrostáticos generados por CoMISA. En estos mapas, aparecen contornos rojos (regiones donde un potencial negativo favorece la actividad) cerca de las posiciones 3 y 5 del grupo fenilo, muy cerca de los contornos verdes discutidos anteriormente. Estos contornos rojos están relacionados con los halógenos de los compuestos **20-25**, **28**, **31**, **31** y el grupo 3-nitro del compuesto **12**. Es importante resaltar que entre estos compuestos se encuentran los más activos de toda la base de datos. Cabe señalar que los contornos azules, relacionado con las regiones donde un potencial negativo está asociado con una disminución en la actividad, casi desaparecieron. Este es un claro ejemplo de cómo las metodologías CoMFA y CoMSIA se complementan entre sí. En este caso, CoMFA fue capaz de detectar que los electrones π y el tamaño del grupo 3-fenilo disminuyen la actividad del compuesto **19**, mientras que CoMSIA detectó la importancia de los halógenos en las posiciones 3 y 5 del fenilo para la actividad biológica.

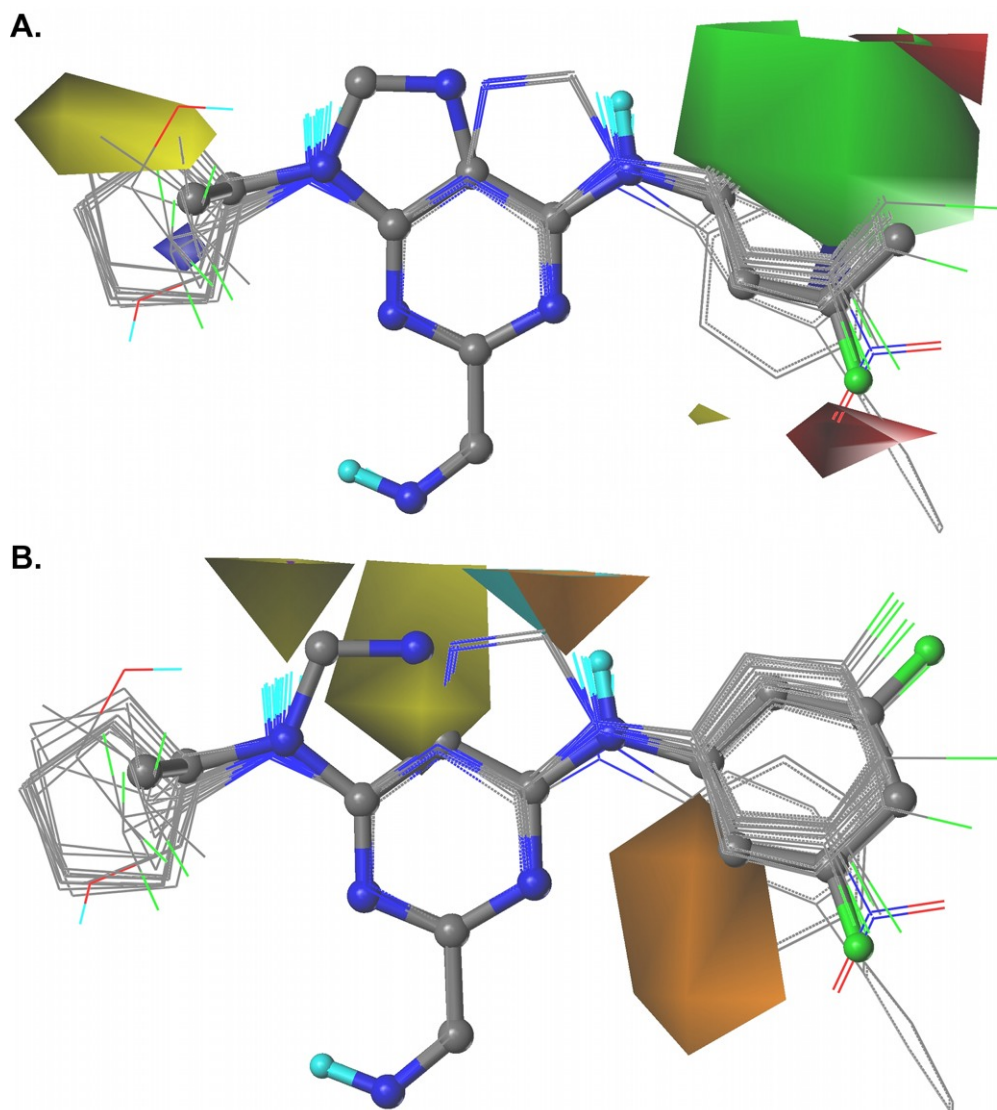


Figura 4.5. Mapas de contorno CoMSIA que incluyen los campos SEDA. El panel (A) muestra los contornos estéricos y electrostáticos los cuales siguen el mismo código de colores que CoMFA. El panel (B) muestra contornos anaranjados en regiones donde aceptores de puentes de hidrógeno son favorables, mientras que los amarillos muestran regiones donde no son favorables. De la misma manera, los contornos cian resaltan regiones donde los donadores de puentes de hidrógeno incrementan la actividad, mientras que los morados indican regiones donde la disminuyen.

La Figura 4.5B muestra los mapas de contorno relacionados con los grupos aceptores y donadores de puentes de hidrógeno. Los grupos aceptores de puentes de hidrógeno son favorables para la actividad biológica si se encuentran

localizados en regiones cercanas a los contornos anaranjados y son desfavorables si están próximos a los contornos amarillos. En el caso de los grupos donadores de puentes de hidrógeno, estos favorecen la actividad si se encuentran contiguos a los contornos cian y la desfavorecen cuando están cerca de contornos morados. Como se puede observar en la Figura 4.5B, existe un contorno amarillo que rodea el nitrógeno de la triazina el cual se encuentra traslapado con el carbono cabeza de puente de la purina. Este contorno puede estar relacionado con la preferencia de un carbono sobre un nitrógeno en esta posición en las moléculas con un anillo de triazina. En términos de los compuestos utilizados en este estudio, este resultado refleja que es preferible utilizar un anillo de purina en lugar de uno de triazina como estructura base. Al comparar el compuesto **14** con el **27** y el **22** con el **24** se puede observar claramente que los compuestos con un anillo de purina son mejores inhibidores de la cruzaina que aquellos que contienen un anillo de triazina. Un punto que llama la atención es que al analizar la estructura cristalográfica de la cruzaina unida al compuesto **23** no se logra observar ninguna posible interacción entre el anillo de purina y la enzima. Esto hace pensar en la posibilidad de que el anillo de purina incremente la electrofilicidad en el grupo nitrilo haciéndolo más susceptible a un ataque nucleofílico por parte del tiolato de la Cys25 como lo propone Oballa *et al* [59]. Basado en lo anterior, la diferencia en la actividad biológica de los compuestos con un anillo de purina y aquellos con un anillo de triazina se podría explicar con base en su diferente reactividad hacia la enzima.

Un segundo contorno amarillo aparece sobre la amina secundaria que une al anillo de purina con los sustituyentes en posición R² de los compuesto **29-32** (ver Tabla 4.1). De la misma manera, se observa un contorno naranja y uno cian sobre la amina que une al anillo de purina con el fenilo en los compuestos **22-28**. Es interesante mencionar que en esta ocasión los contornos no muestran una posible interacción de la amina como aceptor o donador de puente hidrógeno, no obstante nos brindan información sobre la orientación del anillo de purina con base en la posición de la amina secundaria. En este grupo de moléculas, los contornos se pueden interpretar como una preferencia de que el grupo fenilamino se encuentre

en la posición 6 del anillo de purina como en los compuestos **22-28**. Al comparar directamente los compuestos **23-26** con las moléculas **31, 28, 32 y 29**, respectivamente, se puede observar que el grupo amino en la posición 6 (y no en la posición 9) del anillo de purina confiere una mayor actividad biológica a estos compuestos.

4.2. Estudios químico-cuántico del mecanismo de reacción

4.2.1. Superficie de energía potencial

Se utilizó la teoría de funcionales de la densidad (DFT) con la finalidad de estudiar la superficie de energía potencial (SEP) correspondiente a la inhibición covalente de la cruzaina por un purino-carbonitrilo. Para esta reacción se encontró una trayectoria de mínima energía cuando el inhibidor se aproxima al tiolato de la Cys25 para empezar un ataque nucleofílico. Este no es un resultado inesperado ya que el carbono del nitrilo tiene un déficit de carga que facilita el ataque nucleofílico del tiolato. En la Figura 4.6 se muestra el perfil de reacción calculado a dos niveles de teoría en el cual se observa que el estado de transición que separa a los reactivos de los productos se encuentra a 1.39 kcal/mol por arriba de la energía de los reactivos. Este estado de transición se alcanza cuando la distancia C2-S es de 2.39 Å. En esta estructura, el tiolato interacciona directamente con el carbono del nitrilo produciendo una migración electrónica desde el carbono hacia un orbital *p* del tiolato lo que debilita el enlace N1-C2 en el nitrilo. Este hecho se refleja en una elongación del enlace N1-C2 hasta alcanzar una distancia de 1.19 Å al mismo tiempo que se rompe el triple enlace del nitrilo para formar un doble enlace. Un aspecto interesante del estado de transición es que el N1 y H se encuentran a 1.78 Å de distancia lo cual hace pensar que su interacción es relativamente débil.

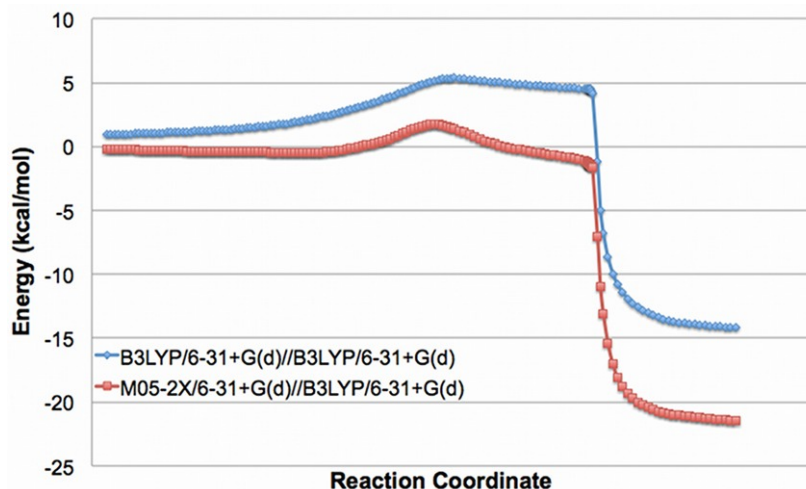
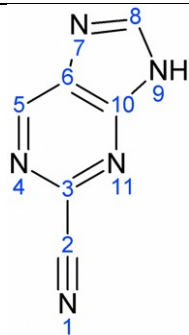
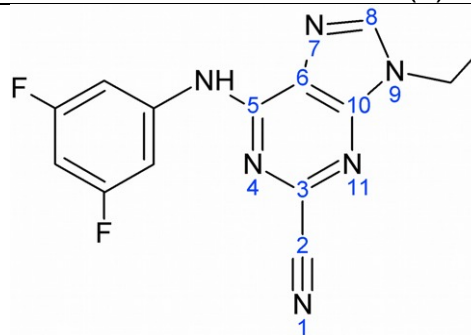


Figura 4.6. Curvas de energía potencial obtenida para la inhibición de la cruzaina por un purino-carbonitrilo.

Tabla 4.5. Valores calculados de la función de Fukui utilizando B3LYP/6-31+G(d)



Purino-carbonitrilo



6-[(3,5-difluorofenil)amino]-9etil-9H-purino-2-carbonitrilo

Átomo	NBO			Mulliken			NBO			Mulliken		
	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0
N1	-0.180	-0.203	-0.191	-0.174	-0.181	-0.177	-0.154	-0.089	-0.121	-0.149	-0.085	-0.117
C2	0.018	0.01	0.014	-0.007	-0.015	-0.011	0.008	0.032	0.020	0.007	0.001	0.004
C3	-0.127	-0.113	-0.120	-0.072	-0.05	-0.061	-0.118	-0.031	-0.074	-0.055	0.029	-0.013
N4	-0.018	-0.042	-0.030	-0.068	-0.096	-0.082	0.000	-0.034	-0.017	-0.049	-0.04	-0.044
C5	-0.144	-0.079	-0.111	-0.146	-0.033	-0.089	-0.094	0.029	-0.032	-0.025	0.041	0.008
C6	-0.056	-0.09	-0.073	-0.078	-0.087	-0.082	-0.046	-0.062	-0.054	-0.039	-0.036	-0.037
N7	-0.05	-0.095	-0.072	-0.064	-0.125	-0.094	-0.031	-0.016	-0.023	-0.055	-0.042	-0.048
C8	-0.143	-0.136	-0.139	-0.106	-0.066	-0.086	-0.118	-0.069	-0.093	-0.129	-0.038	-0.083
N9	-0.024	-0.013	-0.018	-0.011	-0.041	-0.026	-0.029	-0.015	-0.022	0.027	-0.004	0.011
C10	-0.003	-0.063	-0.033	0.053	-0.009	0.022	-0.003	-0.03	-0.016	0.001	-0.031	-0.015
N11	-0.155	-0.056	-0.105	-0.157	-0.11	-0.133	-0.114	-0.081	-0.097	-0.116	-0.092	-0.104

Después que la reacción pasa el estado de transición llega a una zona en la SEP que asemeja una plataforma la cual está relacionada con la transferencia de un hidrógeno desde la His162 hacia el grupo nitrilo del inhibidor. Los valores resultantes de calcular las funciones de Fukui indican que la carga negativa que se añade al nitrilo durante el ataque nucleofílico se va a localizar en el nitrógeno del nitrilo, esto se ve reflejado en un valor $f_A^+ = 0.180$ para N1 (Tabla 4.5). Al mismo tiempo, un exceso de carga en el nitrilo convierte a N1 en un buen nucleófilo (con un valor de f_A^- de 0.203) lo que facilita la transferencia del hidrógeno de la histidina hacia el inhibidor. Otro cambio estructural que se lleva a cabo durante la región plana de la SEP es la variación del ángulo diedro S-C2-C3-N11. Este ángulo se

modifica con la finalidad de conjugar los electrones π del anillo de purina con los del doble enlace N1-C2. Un aspecto a resaltar es que la distancia C2-S no cambia significativamente durante la región plana de la SEP y no es lo suficientemente corta para completar la formación de un enlace covalente. Las estructuras correspondientes a los estados estacionarios de la reacción se muestran en la Figura 4.7 y la información sobre los parámetros geométricos y energéticos se resume en la Tabla 4.6.

Es importante mencionar que la forma tan particular que presenta esta SEP, en la que aparece una región muy plana, se ha observado en ocasiones anteriores y se atribuye a la posible existencia de interacciones de vibracionales cerca del estado de transición [60]. Las interacciones vibracionales son favorecidas por la existencia de estados excitados que se encuentran en niveles bajos de energía y toman mayor importancia cerca del estado de transición donde puede llevarse a cabo el cruce de diferentes estados electrónicos. Estas interacciones vibracionales tienden a aumentar la diferencia de energía entre dos SEP y previenen el cruce de dos estados electrónicos diferentes; a consecuencia de esto, la SEP de menor energía tiende a aplanarse [60].

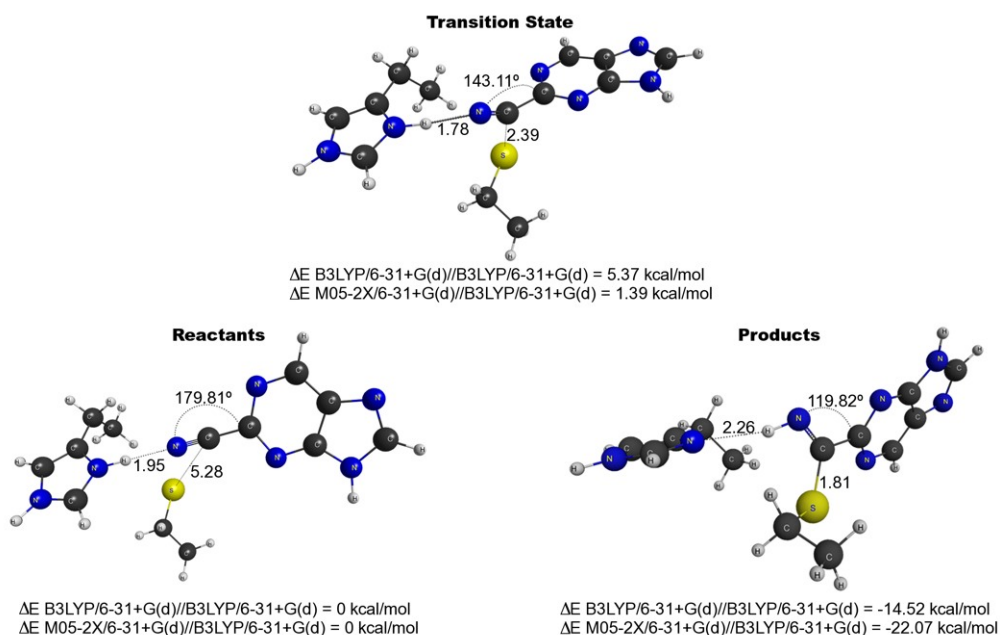


Figura 4.7. Estructuras de mínima energía calculadas a nivel B3LYP/6-31+G(d) correspondientes a los estados estacionarios de la reacción. Las distancias se encuentran indicadas en Å.

Tabla 4.6. Parámetros geométricos y energéticos de los estados estacionarios en la inhibición de la cruzaina

Parametros*	Purino-carbonitrilo			6-[(3,5-difluorofenil)amino]-9-etil-9H-purino-2-carbonitrilo		
	Reactivos	Estado de Transición	Productos	Reactivos	Estado de Transición	Productos
r (N1-C2)	1.1601	1.1920	1.2723	1.1602	1.1926	1.2727
r (N1-H)	1.9514	1.7763	1.0279	1.9441	1.7552	1.0266
r (N1-N _{His})	2.9749	2.8221	3.1898	2.9705	2.8063	3.2256
r (C2-C3)	1.4492	1.4742	1.5048	1.4509	1.4774	1.5075
r (C2-S)	5.2787	2.3900	1.8143	5.1646	2.3829	1.8128
r (N _{His} -H)	1.0276	1.0488	2.2601	1.0284	1.0531	2.3212
θ (N1-C2-C3)	179.811	143.108	119.817	179.624	142.394	119.885
θ (N1-C2-S)	52.949	118.651	127.363	56.049	118.496	127.361
θ (H-N1-C2)	178.611	127.319	113.715	178.490	130.261	113.607
D (S-C2-C3-N11)	-62.592	-108.634	-153.342	-73.239	-109.024	-154.944
D (S-C2-N1-H)	162.889	-36.166	-5.877	-13.455	-25.275	-5.609
E (B3LYP/6-31+G(d)//B3LYP/6-31+G(d))	0.00	5.37	-14.52	0.00	6.27	-13.94
E (M05-2X/6-31+G(d)//B3LYP/6-31+G(d))	0.00	1.39	-22.07	0.00	1.50	-22.42
G (B3LYP/6-31+G(d)//B3LYP/6-31+G(d))	0.00	8.92	-9.68	0.00	8.32	-13.00

* r , distancia; θ , ángulo de enlace; D , ángulo diedro; E, energía relativa; G, energía libre relativa.

A medida en que la coordenada de reacción incrementa, la SEP llega al borde de la región plana y de este punto en adelante sigue una trayectoria cuesta abajo hasta el producto, un tioimidato, que se encuentra 22.07 kcal/mol por debajo de la energía de los reactivos. Esta diferencia de energía se debe a dos principales razones: la formación del enlace covalente entre el azufre del tioimidato y el carbono del nitrilo, y la estabilización del enlace entre el protón y el nitrógeno del ligando. Las distancias de enlace C2-S, N1-C2 y N1-H son 1.87, 1.27 y 1.02 Å, respectivamente; presenta un ángulo N1-C2-C3 de 119.82° y un ángulo diedro S-C2-C3-N11 de 153.34°.

Es bien sabido que las moléculas que contienen un nitrilo forman un éster de tioimidato, el cual es reversible, entre el inhibidor y el sitio activo de la enzima, en este caso la Cys25 [61]. Basándonos en la SEP calculada en este trabajo se puede decir que la reacción es reversible durante la región plana posterior al estado de transición y antes de que se alcance el borde (Figura 4.6); en otras palabras, antes de que se complete la transferencia del protón hacía el inhibidor. Este resultado indica que se podría modular la reversibilidad de la reacción, lo cual sería muy interesante para el diseño de nuevos inhibidores.

Con la finalidad de analizar el efecto de los sustituyentes en el anillo de purina sobre la reactividad hacia la cruzaina, se calcularon los estados estacionarios para el inhibidor 6-[(3,5-difluorofenil)amino]-9-etil-9*H*-purino-2-carbonitrilo el cual tiene una CI_{50} de 0.01 μ M hacia la cruzaina [25]. Los resultados obtenidos resultan de gran interés porque no se observan diferencias significativas en los parámetros geométricos ni energéticos respecto a los obtenidos con el purino-carbonitrilo (Tabla 4.6). Estos resultados sugieren que los sustituyentes de este inhibidor no influyen directamente en la reactividad del grupo nitrilo. No obstante, el estudio de la reactividad y actividad biológica presentada por inhibidores con grupos electroattractores o electrodonadores unidos directamente al anillo de purina podría aportar más información al respecto.

Estos resultados se pueden utilizar como guía en el diseño de nuevos inhibidores de la cruzaina. Por ejemplo, se puede sugerir que la reversibilidad de lo inhibidores con un grupo carbonitrilo puede ser modulada reduciendo o aumentando la carga negativa en el nitrógeno del nitrilo. Una menor carga negativa en este nitrógeno podría retardar la transferencia del protón haciendo a la reacción más reversible. En contraste, una mayor carga negativa en el nitrógeno aceleraría la transferencia de protón al mismo tiempo que facilitaría la formación del tioimidato. Por otro lado, el tamaño de la barrera energética se podría reducir al aumentar la electrofilia del carbono en el nitrilo con la finalidad de facilitar el ataque nucleofílico proveniente del tiolato en la Cys25.

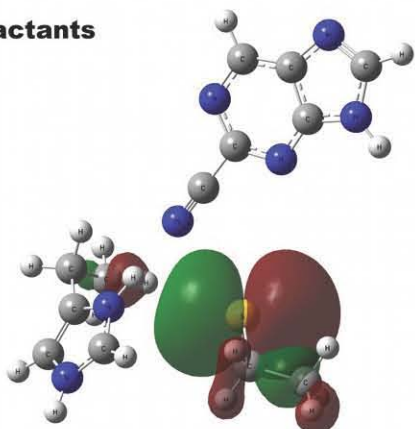
4.2.2. Análisis de orbitales moleculares

El modelo de orbitales moleculares, en especial la teoría de orbitales frontera, es un herramienta muy útil para evaluar y analizar la reactividad de las diferentes especies en una reacción. Esta teoría se basa en asumir que el enlace químico se forma por un flujo de electrones desde el orbital ocupado de mayor energía (HOMO) hacia el orbital desocupado de menor energía (LUMO).

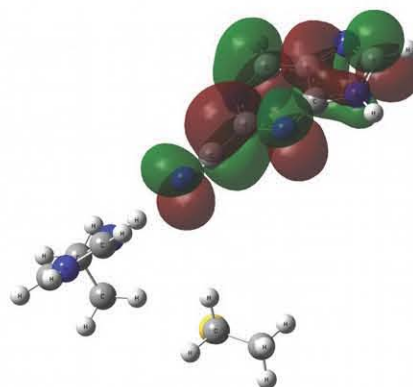
La Figura 4.8 muestra el HOMO y LUMO de los estados estacionarios durante la inhibición de la cruzafina por un purino-carbonitrilo. La diferencia de energía entre el HOMO y el LUMO durante los reactivos es de apenas 50.88 kcal/mol, lo cual indica que la reacción se puede llevar a cabo con una barrera energética relativamente baja. En los reactivos, se puede observar que el HOMO se encuentra localizado en el cisteína, principalmente el orbital p del tiolato, que actúa como donador de electrones. En contraste, el LUMO se forma por un sistema de orbitales p en todo el purino-carbonitrilo. Para que la interacción entre los orbitales p de ambos reactivos se lleve a cabo es necesario que el ataque nucleofílico proveniente del tiolato proceda de forma perpendicular al anillo de purina. Es importante resaltar que el nucleófilo puede atacar por cualquier lado del anillo de purina por lo que surge una característica estereoquímica que determina el modo de unión del ligando. Sin embargo, la interacción más favorable es cuando el ataque procede mientras que el C5 de la purina apunta hacia la His162, ya que esta interacción se encuentra permitida por la simetría de los orbitales.

Al analizar el HOMO del siguiente paso de la reacción, el estado de transición, se puede observar que los orbitales p localizados en el carbonitrilo del inhibidor y en el tiolato de la cisteína se traslapan. Esto es debido a la transferencia de electrones desde el S hacia el C2 lo que comienza a formar un enlace covalente. Un detalle interesante de los orbitales frontera en el estado de transición es que el orbital p del nitrógeno en el nitrilo, el cual se encuentra apuntando hacia el hidrógeno de la His162, se traslapa con el orbital s de un hidrógeno localizado en la cadena lateral de la cisteína. Este hecho se puede asociar con una posible estabilización del estado de transición mediante una interacción con la cadena

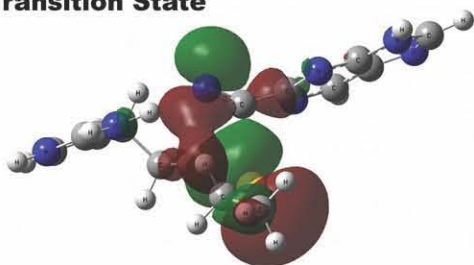
lateral de la Cys25. Además, se observa que el grupo carbonitrilo se encuentra unido al anillo de purina solamente por la interacción de orbitales s.

Reactants

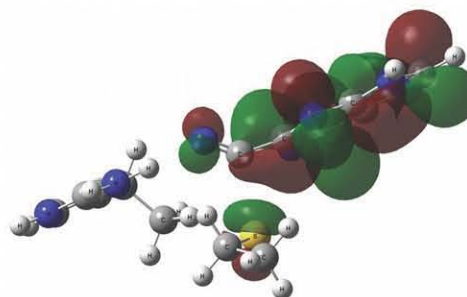
HOMO
-0.17504



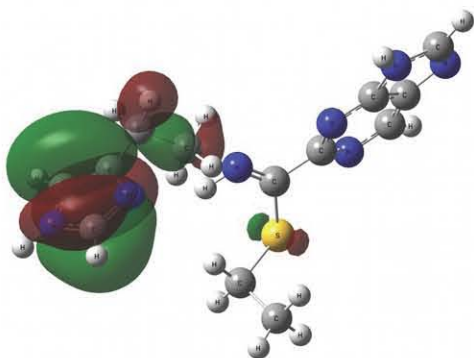
LUMO
-0.09395

Transition State

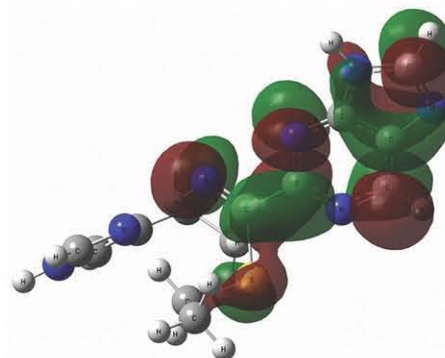
HOMO
-0.19311



LUMO
-0.06668

Products

HOMO
-0.23235



LUMO
-0.07661

Figura 4.8. Orbitales frontera y energías orbital, en unidades atómicas, correspondientes a los estados estacionarios de la reacción.

Finalmente, los productos de la reacción presentan una diferencia de energía entre el HOMO y el LUMO de 97.72 kcal/mol, la cual es muy alta y se relaciona con la estabilidad de los productos. En la Figura 4.8 se puede observar que el HOMO de los productos se encuentra localizado en el sistema de orbitales p del imidazol en la His162, mientras que el LUMO se compone de orbitales p de antienlace en el anillo de purina. A pesar de que los orbitales frontera de los productos no son tan interesantes como los presentados por los reactivos y el estado de transición, se puede observar que el HOMO y el LUMO son incapaces de reaccionar entre sí, lo cual explica la gran diferencia de energía entre el HOMO y el LUMO.

4.2.3. Estudio de la reacción dentro del sitio catalítico

Es posible que otros aminoácidos en el sitio activo, diferentes de aquellos que actúan directamente en el mecanismo de reacción, intervengan estabilizando el estado de transición o los productos; es por eso que en esta sección se estudia el papel que toman otros aminoácidos durante la reacción de inhibición. Se realizaron cálculos utilizando el hamiltoniano semi-empírico PM6 con el propósito de analizar las posibles interacciones entre el inhibidor y 20 aminoácidos en el sitio activo de la cruzaina. En particular se estudió la SEP correspondiente a la inhibición de la cruzaina por el compuesto 6-[(3,5-difluorofenil)amino]-9-etil-9H-purino-2-carbonitrilo. En este estudio fue posible generar una SEP de 2 dimensiones, la cual se muestra en la Figura 4.9, al seguir dos coordenadas de reacción simultáneamente: la distancia C2-S entre el tiolato y el nitrilo; y la distancia N1-H entre el nitrilo y la histidina. Se puede observar que la trayectoria de menor energía corresponde a aquella en la que el S y el C2 se acercan de tal manera que parece un ataque nucleofílico. Una vez que el S y el C2 se encuentran lo suficientemente cerca, se observa una migración del protón desde la histidina hacia el inhibidor para que la reacción llegue a los productos finales. Por el contrario, no se observa ninguna trayectoria favorable para el proceso en el que la protonación del nitrilo sucede antes del ataque nucleofílico por parte del

tiolato; de hecho, se puede observar que este proceso no se encuentra favorecido energéticamente.

Cabe señalar que los resultados obtenidos utilizando cálculos químico-cuánticos a nivel semi-empírico están en buen acuerdo con aquellos realizados empleando la teoría de funcionales de la densidad. Este hecho se refleja en la similitud del mecanismo de reacción obtenido para este proceso y en la energía relativa de los productos, los cuales yacen 17.07 kcal/mol por debajo de la energía de los reactivos. Adicionalmente, los parámetros geométricos de las estructuras optimizadas utilizando el hamiltoniano PM6 son muy similares a los presentados en las estructuras optimizadas mediante cálculos DFT. Por ejemplo, en los productos las distancias de enlace C2-S, N1-C2 y N1-H son 1.82, 1.26 y 1.00 Å, respectivamente; se presenta un ángulo N1-C2-C3 de 125.60° y un ángulo diedro S-C2-C3-N1 de 134.50°. Las estructuras de los reactivos y productos optimizadas con el hamiltoniano PM6 se muestran en la Figura 4.10.

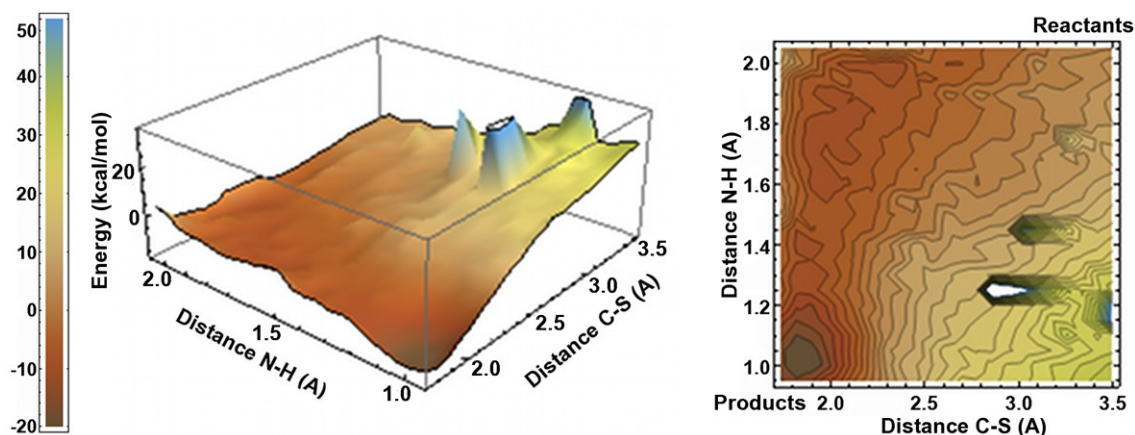


Figura 4.9. Superficie de energía potencial correspondiente a la inhibición de la cruzaina. Los cálculos fueron realizados utilizando el hamiltoniano semi-empírico PM6.

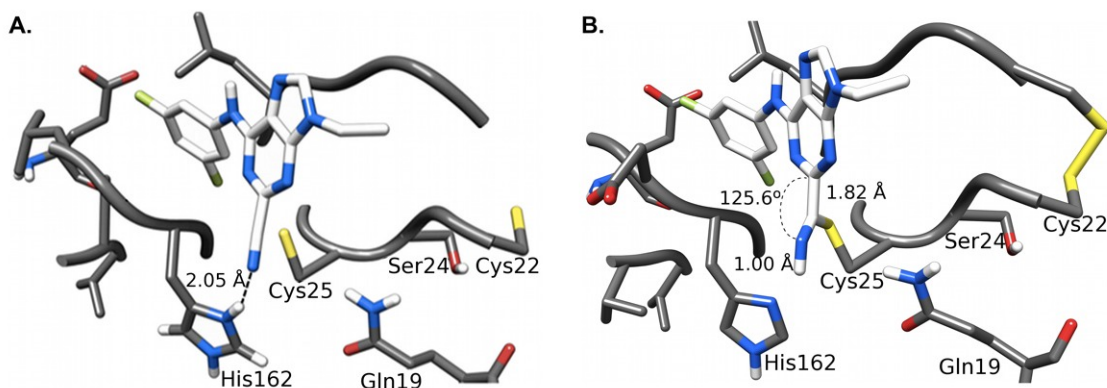


Figura 4.10. Estructuras de los (A) reactivos y (B) productos optimizadas utilizando el hamiltoniano semi-empírico PM6.

La variación de la carga total de Mulliken en cada aminoácido se utilizó como un parámetro para evaluar las posibles interacciones durante la reacción. La carga de cada aminoácido se calculó como la suma de las cargas parciales de cada átomo en el residuo. En la Figura 4.11 se presenta un histograma que compara la carga de cada aminoácido en los reactivos y en los productos. Es el histograma se puede observar que la mayoría de los residuos no presentan un variación de carga significativa. No obstante, los dos aminoácidos que intervienen en la reacción (Cys25 e His162) presentan una variación de carga debido a la transferencia del protón. Un aspecto a resaltar es que la carga de la Gln19, la cual actúa como un oxoanión en mecanismo catalítico, no varía drásticamente durante la reacción. Este hecho sugiere que la Gln19 no juega un papel importante en la estabilización del inhibidor.

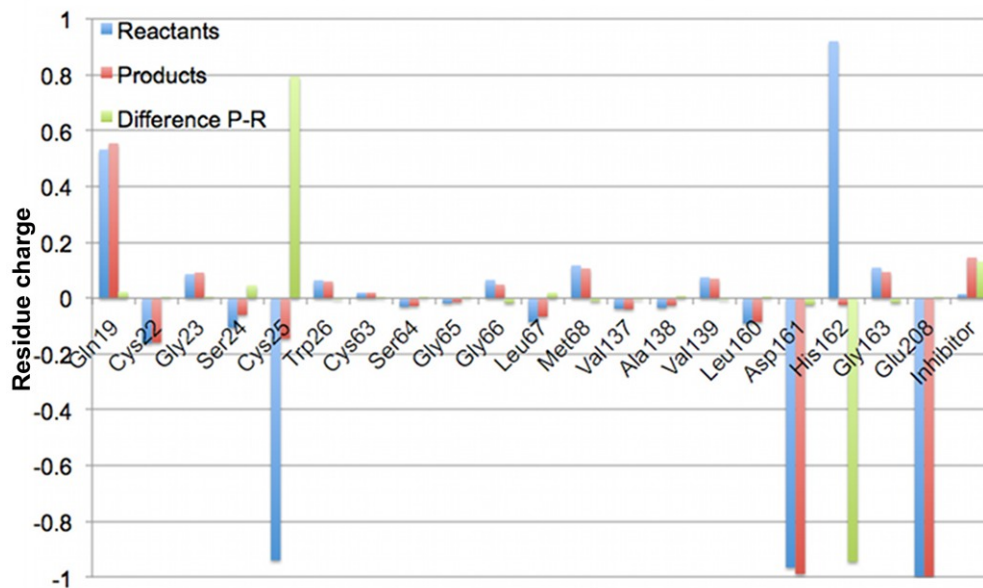


Figura 4.11. Histograma que muestra la variación de carga de los 20 aminoácidos en el sitio activo durante la reacción.

A pesar de que los cálculos semi-empíricos ofrecen resultados poco cuantitativos, nos brindan muy buena información de cómo interactúa el ligando con todo el sitio activo de la enzima; además, representan un punto de partida para estudiar el mecanismo de reacción de manera más detallada. En futuros estudios sería interesante evaluar la contribución de toda la proteína en la formación del complejo enzima-ligando o estudiar la capacidad del ligando para producir cambios conformacionales en la enzima.

CAPÍTULO 5. CONCLUSIONES

5.1. Conclusiones del estudio QSAR 3D

Se realizó un estudio QSAR 3D con un conjunto de 32 purino- y triazino-carbonitrilos, los cuales han sido identificados como inhibidores de la cruzaina, la cisteína proteasa más importante del *Trypanosoma cruzi*. La geometría de las moléculas se optimizó utilizando métodos semi-empíricos, los cuales resultaron ser una buena aproximación a la conformación real de los inhibidores dentro del sitio activo. La alineación de los ligandos, la cual es un paso crucial en el desarrollo de modelos QSAR 3D, se realizó con base en información experimental (la estructura cristalográfica del compuesto **23** [25] e información acerca del sitio catalítico de la cruzaina [57]) la cual permitió generar dos hipótesis de alineamiento con las que se generaron los modelos CoMFA 1 y 2. El modelo CoMFA 2 presentó los parámetros estadísticos más aceptables, con una q^2 de 0.736 y una r^2 de 0.980.

También se estudiaron los modelos resultantes de las 31 combinaciones posibles entre los 5 campos diferentes disponibles en CoMSIA. El modelo CoMSIA que incluye los campos SEDA presentó parámetros estadísticos aceptables con una q^2 de 0.565, una r^2 de 0.901 y una R^2 de 0.815; adicionalmente, es el modelo que nos brinda más información acerca de los requerimientos estructurales para la actividad biológica. Cabe señalar que todos los modelos CoMFA y CoMSIA cumplen los criterios de la validación propuesta por Golbraikh y Tropsha [46].

Los resultados de este estudio revelaron cuatro requerimientos estructurales esenciales para la actividad biológica:

- Se prefiere un anillo de purina sobre uno de triazina como estructura base de los inhibidores.
- Los grupos voluminosos con un potencial electrostático negativo, como los halógenos, en las posiciones 3 y 5 del grupo fenilo están asociados con un aumento en la actividad.

- Se prefieren grupos pequeños en la posición 9 del anillo de purina.
- El grupo fenilo debe estar unido al anillo de purina en la posición 6 mediante una amina secundaria.

En conclusión, los modelos reportados en este trabajo de tesis mostraron un buen poder predictivo. Aunado a esto, la información obtenida acerca de los requerimientos estructurales necesarios para la actividad de estas moléculas puede ser útil en la optimización o diseño de nuevos purino- y triazino-carbonitrilos como inhibidores de la cruzafina.

5.2. Conclusiones del estudio químico-cuántico del mecanismo de reacción

Se realizó un estudio teórico utilizando cálculos semi-empíricos y DFT con la finalidad de revelar el mecanismo por el cual las moléculas con un grupo nitrilo inhiben a las cisteína proteasas, especialmente a la cruzafina. Los resultados de este trabajo de tesis sugieren que la inhibición ocurre vía un mecanismo que inicia con un ataque nucleofílico por parte de la Cys25 hacia el inhibidor seguido la transferencia de un protón de la His162. La reacción presenta una SEP muy peculiar, la cual consiste de un estado de transición seguido de una región plana. Estos resultados están de acuerdo con el mecanismo catalítico propuesto para la papaína, otra cisteína proteasa, en el que la hidrólisis peptídica se lleva a cabo por un mecanismo concertado [15-17]. Por otro lado, no se encontró ningún intermediario en la reacción como se ha hecho en modelos para la inhibición de cisteína proteasas por dicetonas, halometil cetonas, diazometil cetonas y aziridinas [26-30].

La evidencia presentada en este trabajo de tesis sugiere que la reversibilidad de la interacción puede ser manipulada modulando la transferencia del protón desde la His162 hacia el inhibidor. Adicionalmente, se sugiere que al aumentar la electrofilicidad del carbono en el nitrilo se podría facilitar el ataque nucleofílico por parte de la Cys25. De esta manera, se podría reducir la barrera energética que se

para a los productos de lo reactivos, al mismo tiempo de que se incrementaría la actividad biológica. Sería interesante investigar de qué manera se podría controlar la reactividad y la reversibilidad de éstas moléculas con la finalidad de diseñar fármacos más efectivos y seguros contra la enfermedad de Chagas.

En términos generales, este estudio presenta información valiosa para entender de mejor manera el mecanismo por el cual las moléculas con un grupo nitrilo son capaces de inhibir a la cruzaina. La información obtenida puede ser de gran ayuda en el diseño de nuevos y mejores inhibidores de las cisteína proteasas, los cuales pueden conducir a nuevos agentes tripanocidas.

Los resultados presentados en esta parte dieron lugar a las siguientes publicaciones:

[31] O. Méndez-Lucio, A. Romo-Mancillas, J.L. Medina-Franco, R. Castillo. Computational study on the inhibition mechanism of cruzain by nitrile-containing molecules. *J. Mol. Graphics Modell.* 2012, 35, 28-35.

[33] O. Mendez-Lucio, J. Perez-Villanueva, A. Romo-Mancillas, R. Castillo. 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). *Med. Chem. Commun.* 2011, 2, 1058-65.

PARTE II

ANÁLISIS QUIMIOINFORMÁTICO DE BASES DE DATOS

CAPÍTULO 6. ANTECEDENTES

6.1. *Panorama de actividad*

El estudio de la relación estructura-actividad (SAR) se ha convertido en un tema de gran importancia al momento de diseñar moléculas bioactivas [62]. Los resultados obtenidos de estos estudios nos brindan información importante durante la búsqueda de una molécula líder o en la optimización de ésta. En los últimos años se ha observado un aumento en la aplicación de estos métodos, principalmente aquellos que permiten relacionar, de manera cuantitativa, la estructura tridimensional de una molécula con su actividad; por ejemplo, el análisis de campos moleculares (CoMFA) y el análisis comparativo de índices de similitud molecular (CoMSIA) [3, 4]. Estas metodologías se usan principalmente con bases de datos pequeñas o medianas, generalmente no más de 80 compuestos. Sin embargo, los recientes avances en química combinatoria y en la búsqueda de alto rendimiento (*high-throughput screening*) hacen que surja la necesidad de utilizar nuevas metodologías para analizar el SAR de bases de datos compuestas por cientos de moléculas. Una metodología utilizada ampliamente para este fin es el modelado del *panorama de actividad* [62, 63].

El *panorama de actividad* se puede conceptualizar como una superficie en la que se observa la respuesta biológica respecto al espacio químico, la cual se obtiene al agregar una dimensión de actividad a este espacio [63]. Estas representaciones son capaces de integrar y facilitar el análisis de similitud estructural y diferencia de potencia entre compuestos que comparten la misma actividad biológica. Por ejemplo, si los compuestos son muy parecidos estructuralmente, entonces estarán localizados en regiones cercanas del espacio químico; por el contrario, compuestos diferentes se encontrarán en regiones alejadas [62]. Como se muestra en la Figura 6.1, los *panoramas de actividad* se pueden considerar como mapas topográficos, los cuales son muy parecidos a los mapas geográficos, donde es posible distinguir diferentes tipos de SAR. Por ejemplo, un SAR continuo está asociado con un *panorama de actividad* que presenta pequeñas colinas, las cuales indican que pequeños cambios en la estructura molecular producen

pequeños cambios en la actividad biológica [64]. En contraste, un SAR discontinuo presenta un *panorama de actividad* más accidentado (asemejando riscos y montañas) plagado de moléculas con pequeñas diferencias estructurales que generan un gran cambio en la actividad biológica [64]. La tarea de analizar una base de datos es un poco más compleja ya que el *panorama de actividad* suele ser una combinación de regiones continuas y discontinuas, a lo que se le denomina un SAR heterogéneo.

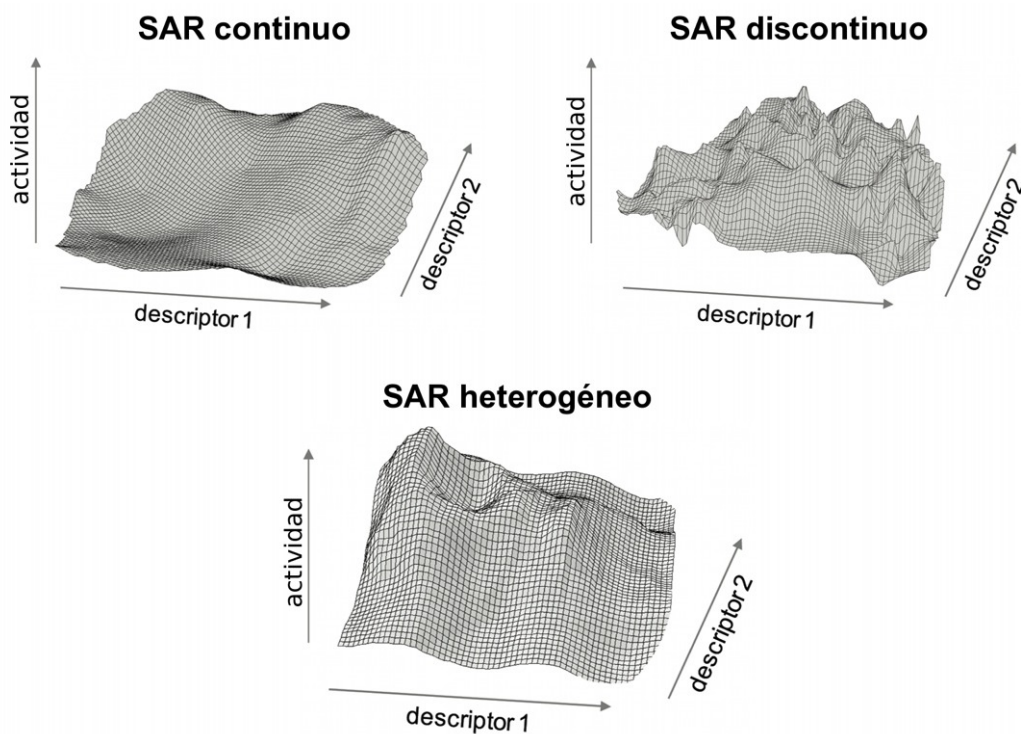


Figura 6.1 Diferentes tipos de SARs visualizados empleando el concepto de *panorama de actividad* (figura tomada de referencia [62]).

En la actualidad, el modelado del panorama de actividad se establece como una técnica descriptiva que ayuda a analizar y caracterizar sistemáticamente el SAR de bases de datos de gran tamaño. Este método se ha beneficiado del desarrollo de representaciones en 2D o 3D que facilitan la caracterización del *panorama de actividad*. Entre las representaciones más utilizadas se encuentran los mapas de similitud estructura-actividad (SAS), los mapas de diferencia de actividad dual

(DAD) y los mapas de diferencia de actividad triple (TAD), en las que se pueden detectar y analizar SARs continuos y discontinuos fácilmente [65-67]. A pesar de que el modelado del *panorama de actividad* se encuentra en pleno desarrollo, es una técnica que ya se está aplicando para estudiar diversas bases de datos que presentan un SAR complejo [66-70]. No obstante, aun se encuentran en estudio diversos alcances de este método; por ejemplo, el desarrollo de modelos predictivos o la comparación cuantitativa de *panoramas de actividad*.

6.2. Activity cliffs

Un *activity cliff* se define generalmente como un par de moléculas que presentan una alta similitud estructural, pero difieren en la potencia de la actividad biológica, por lo que se encuentran asociados con un SAR discontinuo [64, 71, 72]. Generalmente, un SAR discontinuo está determinado por interacciones ligando-receptor específicas. Por esta razón, la formación de *activity cliffs* se puede atribuir a la presencia o ausencia de una o más interacciones ligando-receptor que son críticas para la actividad; por ejemplo, la existencia de un donador o aceptor de puente de hidrógeno como se refleja en la Figura 6.2 [72].

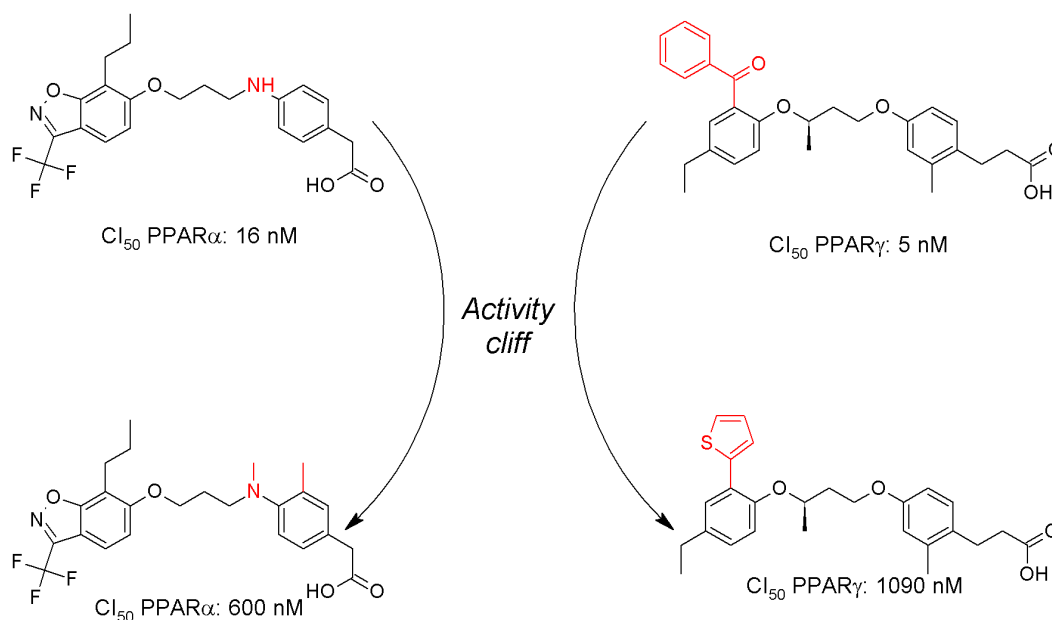


Figura 6.2. Ejemplos de *activity cliffs* para PPAR α y γ . Las principales diferencias estructurales están resaltadas en color rojo

En diferentes estudios se han realizado algunas extensiones del concepto de *activity cliff* las cuales son más apropiadas para ciertos casos específicos [72]; por ejemplo:

- *Consensus activity cliffs*: son *activity cliffs* que pueden ser identificados con diferentes descriptores moleculares. En otras palabras, no son afectados por la dependencia que existe entre el espacio químico y la representación molecular utilizada para describirlo.
- *R-cliffs*: son *activity cliffs* que se dan dentro de una serie de análogos, generalmente con la misma estructura base y donde la diferencia entre las moléculas se encuentra sólo en los sustituyentes (grupos R).
- *Selectivity cliffs*: son un par de compuestos altamente parecidos que presentan una diferencia de actividad muy grande hacia uno o más receptores. En este caso, las diferencias entre las moléculas están asociadas con características estructurales que determinan selectividad hacia cierto receptor.
- *Multitarget activity cliffs*: están formados por compuestos similares que presentan diferente potencia contra una serie de receptores. Estos son interesantes al estudiar la afinidad hacia diferentes miembros de una familia de proteínas.
- *Mechanism cliffs*: están compuestos por pares de moléculas en las cuales los pequeños cambios estructurales inducen la transición de un mecanismo molecular a otro, por ejemplo, de agonista a antagonista de algún receptor.

6.3. Scaffold hops

La palabra *scaffold* se ha convertido en un término común entre los químicos a través de los últimos años. Esta palabra hace referencia al núcleo o estructura base de una molécula; sin embargo, llama la atención que existen diferentes definiciones de *scaffold*, pero ninguna es apropiada para todas las situaciones [73]. Bemis y Murcko propusieron una serie de reglas que pueden ser aplicadas

para determinar *scaffolds* a diferentes niveles de resolución usando análisis computacionales [74]. El primer paso consiste generar un *scaffold* molecular al retirar todas las cadenas laterales de la molécula, sin realizar más cambios en la estructura resultante. La resolución de esta estructura se puede reducir al convertir todos los átomos y enlaces a átomos de carbono y enlaces sencillos, respectivamente. Finalmente, la representación de menor resolución, llamada *scaffold* reducido, se obtiene al convertir los sistemas de anillos en nodos. La Figura 6.3A muestra un ejemplo de los diferentes niveles de abstracción que puede tener un *scaffold* al utilizar estas reglas.

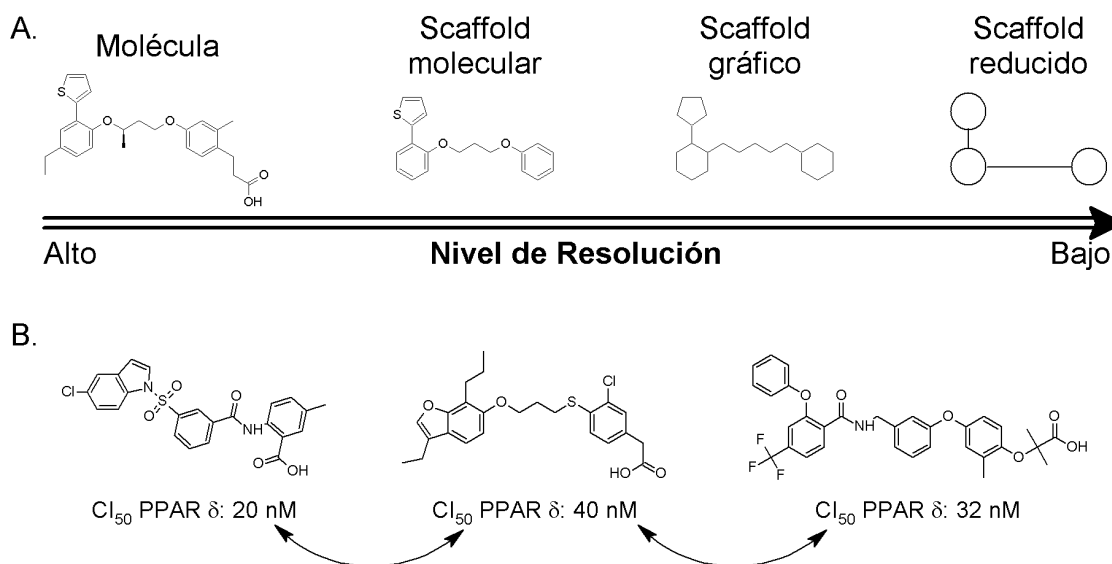


Figura 6.3. (A) Muestra un ejemplo de los diferentes niveles de resolución utilizados para determinar un *scaffold* según Bemis y Murcko [74]. El panel (B) muestra ejemplos de *scaffold hops* para PPAR δ .

Un *scaffold hop* se define como un par de moléculas con diferente *scaffold* que presentan la misma actividad biológica; es decir, un cambio en la estructura base [73, 75]. El *scaffold hopping* se basa en que compuestos similares generalmente presentan actividades biológicas similares como lo dicta el principio de similitud [76]. Esta técnica suele emplearse para explorar diferentes regiones del espacio químico con posible actividad biológica. Adicionalmente, se usa para mejorar propiedades ADME, disminuir toxicidad, facilitar la síntesis de los compuestos o

por simples razones de propiedad intelectual [67, 73, 75]. Por estas razones es importante conocer la frecuencia con la que se forman los *scaffold hops* dentro de una base de datos y cómo se distribuyen al utilizar diferentes receptores.

6.4. Receptores activadores de la proliferación de peroxisomas (PPAR)

La Diabetes Mellitus (DM) se ha convertido en un problema global de salud, en el 2010 la prevalencia mundial de esta enfermedad fue alrededor de 6.4% afectando a 285 millones de personas y algunos estudios indican que aumentará a 7.7% y 439 millones de adultos para el 2030 [77]. Entre los factores de riesgo podemos encontrar que el sobrepeso juega un papel importante en el desarrollo de esta enfermedad [78, 79]. En vista del número creciente de pacientes, se han propuesto diferentes blancos terapéuticos con la finalidad de modular algunos aspectos importantes de la DM. Especialmente, la estimulación de los receptores activadores de la proliferación de peroxisomas (PPARs) ofrecen una estrategia prometedora para el tratamiento de la DM y algunos de sus factores de riesgo [80].

Los receptores de la activación de peroxisomas son factores de transcripción activados por lípidos los cuales han sido identificados como reguladores importantes del metabolismo de lípidos y de la glucosa. Por esta razón, los PPARs contribuyen significativamente a algunas enfermedades como la diabetes, obesidad y complicaciones cardiovasculares [81]. Los PPARs son miembros de la subfamilia 1 (receptores relacionados a hormonas tiroideas) de la superfamilia de los receptores nucleares de hormonas [82]. Como todos los miembros de esta superfamilia, los PPARs se localizan en el núcleo de la célula y son regulados por fosforilación [83, 84]. En los últimos años se describió la expresión de tres subtipos diferentes de PPAR (PPAR α , PPAR δ y PPAR γ), los cuales están codificados en diferentes genes [81, 83]. Cada subtipo presenta un patrón de distribución diferente a lo largo del cuerpo humano, así como una respuesta farmacológica distinta [81].

El PPAR α se expresa en hígado, riñón y corazón, y juega un papel importante en el metabolismo de lípidos regulando positivamente la producción de ApoA-I y regulando el metabolismo de ácidos grasos libres [85, 86]. Los fibratos fueron identificados como agonistas de PPAR α a pesar de que presentan baja afinidad y poca selectividad hacia este subtipo de PPAR [81]. Así mismo, PPAR δ se expresa en todos los tejidos del cuerpo humano y su función está relacionada con la regulación de la respuesta inmune. Sin embargo, algunos agonistas de PPAR δ han sido empleados para tratar diferentes aspectos de DM 2 y del síndrome metabólico [81, 87]. Finalmente, PPAR γ se distribuye en el músculo esquelético, hígado y tejido adiposo, y se asocia principalmente con la regulación de la adipogénesis y la señalización intracelular de la insulina [81, 88, 89]. Este subtipo de PPAR es el blanco de las tiazolidindionas (TDZs), compuestos que se emplean actualmente en el tratamiento de DM 2 como la Rosiglitazona and Pioglitazona [81, 89].

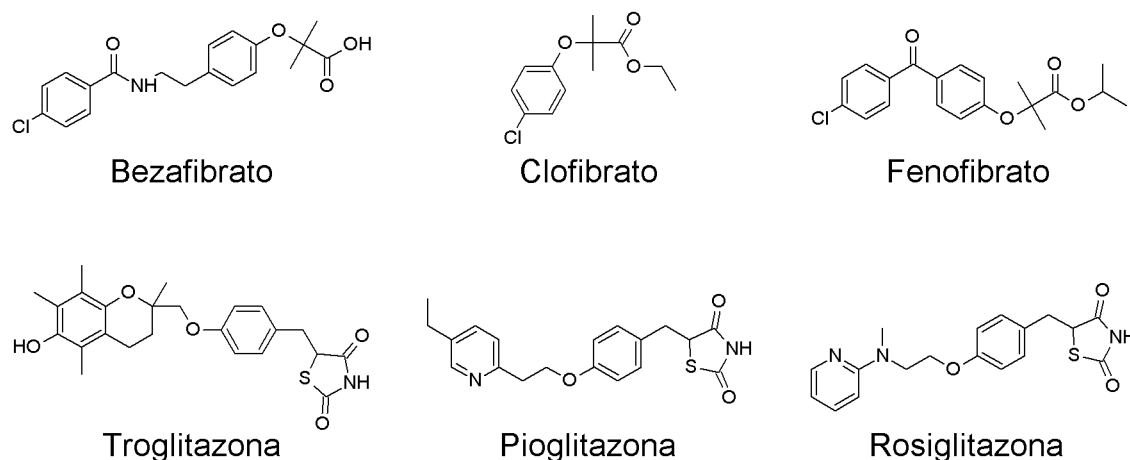


Figura 6.4. Estructuras químicas de algunos fibratos y tiazolidindionas, agonistas de PPAR α y γ , respectivamente.

Recientemente, se han diseñado agonistas duales de los PPAR como posible terapia contra la obesidad, la diabetes y el síndrome metabólico [81, 90, 91]. En los últimos años, los agonistas duales de PPAR α/γ se han utilizado como alternativa para disminuir algunos efectos secundarios de los agonistas de PPAR γ ;

por ejemplo, la resistencia a la insulina y el desbalance de lípidos [90-92]. Desafortunadamente, algunos agonistas duales de PPAR α/γ han presentado problemas de seguridad durante las fases clínicas I y II, los cuales incluyen carcinogénesis en roedores, incremento del riesgo cardiovascular, incremento en la creatinina plasmática, entre otros [93]. En vista de estos resultados, el estudio de requerimientos estructurales que determinen la selectividad hacia un subtipo de PPAR toma una gran relevancia.

CAPÍTULO 7. METODOLOGÍA

7.1. Conjunto de datos

En este estudio se utilizó un conjunto de 168 agonistas de los PPAR que fueron obtenidos de la base de datos pública Binding Database [94, 95]. La determinación de la actividad agonista contra los tres subtipos de PPAR fue utilizada como parámetro de exclusión para seleccionar los 168 compuestos. Los valores de pCl_{50} ($-\log Cl_{50}$) y la estructura química de los compuestos en representación SMILES se enlista en el Apéndice B. Los valores de pCl_{50} para PPAR α se encuentran en el intervalo de 4.36 a 8.32, para PPAR δ van de 3.40 a 8.76 y para PPAR γ de 4.39 a 8.30. Es importante mencionar que el intervalo de actividad biológica hacia los tres subtipos de PPAR es muy parecido, lo que facilita la comparación de los resultados obtenidos para los diferentes blancos biológicos.

7.2. Similitud estructural

Se utilizaron once diferentes descriptores moleculares 2D para describir el *panorama de actividad* de estas moléculas. Esto se realizó con la finalidad de disminuir la dependencia que existe entre el *panorama de actividad* y la representación molecular utilizada. Los descriptores se calcularon con el programa de código libre MayaChemTools [96], el cual es un conjunto de scripts, módulos y clases en Perl que permite generar descriptores correspondientes a: atom neighborhoods [97], atom types, electrotopological state indices (EStateIndices) [98], extended connectivity [99], MACCS (322 bits) [100], path length, topological atom pairs (TopAtomPairs) [101], topological atom torsions (TopAtomTorsions) [102], topological atom triplets (TopAtomTriplets), topological pharmacophore atom pairs (TopPh4Pairs) [103] y topological pharmacophore atom triplets (TopPh4Triplets) [104]. Todos ellos son descriptores ampliamente utilizados y validados para realizar estudios de similitud molecular (Apéndice C). Adicionalmente, se calcularon descriptores moleculares 3D donde se emplearon

ComboScore y ShapeTanimoto, implementados en el programa ROCS (Openeye), utilizando sólo una conformación de mínima energía calculada con el campo de fuerza MMFF94x. Estudios recientes han demostrado que a pesar de que el uso de un conformero para representar la estructura tridimensional de la molécula sólo es una aproximación a la conformación asociada con la actividad biológica, los descriptores moleculares 3D son de gran importancia al caracterizar el *panorama de actividades* [68, 71]. No obstante, estos estudios se pueden realizar utilizando las moléculas en múltiples conformaciones, lo cual sería importante al estudiar compuestos muy flexibles [70]. Durante este estudio se utilizó el coeficiente de Tanimoto para calcular la similitud entre los compuestos en la base de datos [105]. Para cada representación molecular se obtuvo un total de 14,028 datos de similitudes pareadas a partir de los 168 compuestos. Los valores obtenidos con los descriptores EStateIndices, TopAtomPairs, TopPh4Triplets y ComboScore fueron combinadas para generar una nueva medida de similitud llamada similitud promedio.

7.3. Similitud de actividad

La similitud de actividad fue calculada como la diferencia de actividad que existe entre cada agonista de los PPAR en la base de datos. Para esto se utilizó la siguiente expresión:

$$\Delta pCI_{50}(R)_{i,j} = pCI_{50}(R)_i - pCI_{50}(R)_j$$

donde $pCI_{50}(R)_i$ y $pCI_{50}(R)_j$ son las actividades de la i -ésima y j -ésima molécula ($j > i$) hacia el subtipo de receptor (R). En este trabajo R = PPAR α , δ o γ . Es importante mencionar que el valor absoluto de la diferencia de actividades fue utilizado como índice de similitud en los mapas SAS. En contraste, los valores positivos y negativos de esta expresión nos brindan información acerca de la dirección del SAR al generar los mapas DAD.

7.4. Mapas SAS

En la Figura 7.1 se muestra un prototipo de los mapas SAS. Estos mapas se basan en comparaciones pareadas y fueron construidos al graficar el valor absoluto de la diferencia de actividad contra la similitud molecular. Para facilitar su interpretación, cada mapa fue dividido en cuatro regiones utilizando 1 unidad logarítmica como límite de la similitud de actividad. El límite de la similitud molecular fue definido como la mediana de la similitud estructural entre los compuestos activos en la base de datos ($pCl_{50} > 7.3$). Los pares de moléculas localizados en la región I se caracterizan por tener alta similitud de actividad, pero baja similitud estructural, por lo que se asocian con *scaffold hopping*. Las moléculas con alta similitud estructural y alta similitud de actividad se encuentran en la región II, por lo que los compuestos en esta región presentan un SAR continuo. Los pares menos interesantes se encuentran en la región III donde moléculas diferentes tienen un potencia diferente. Finalmente, en la región IV se encuentran aquellos compuestos con alta similitud molecular pero baja similitud de actividad por lo que corresponden a un SAR discontinuo o *activity cliffs*. Se identificaron algunos pares consenso entre las cuatro representaciones moleculares utilizadas (EStateIndices, TopAtomPairs, TopPh4Triplets y ComboScore). Un par consenso está definido como un par que se encuentra en la misma región del mapa SAS utilizando diferentes descriptores moleculares.

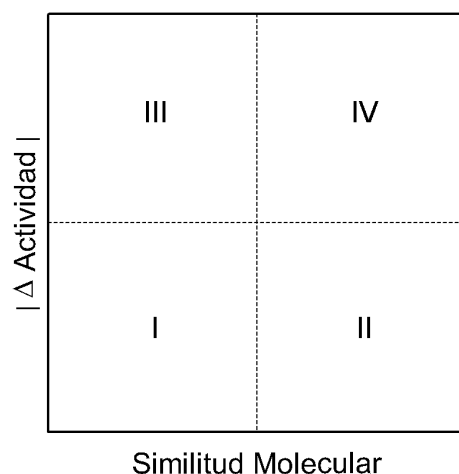


Figura 7.1. Prototipo de un mapa SAS. Los puntos en la región I se asocian con *scaffold hopping*, aquellos en la región II con un SAR continuo y si se encuentran en la región IV son *activity cliffs* o moléculas con SAR discontinuo.

7.5. Mapas DAD y TAD

Los mapas DAD se han utilizado para caracterizar conjuntos de moléculas con actividad biológica hacia dos blancos terapéuticos [66, 67]. Estos mapas están basados en comparaciones pareadas de actividad contra dos receptores y se construyen al graficar la diferencia de actividad hacia un receptor contra la diferencia de actividad hacia un segundo receptor como se muestra en la Figura 7.2. Las líneas verticales y horizontales en $\Delta pCl_{50} \pm t$ determinan las fronteras entre una diferencia de actividad alta y baja para el receptor **A** o **B**, respectivamente. En esta ocasión $t = 1$ unidad logarítmica por lo que un par de moléculas se considera con baja diferencia de actividad si $-1 \leq \Delta pCl_{50} \leq 1$ para cada receptor. Estas fronteras generan cinco zonas en los mapas DAD, las regiones Z1-Z5 [67]. Si los cambios estructurales entre los dos compuestos del par afectan la actividad biológica hacia ambos receptores (**A** o **B** y sin importar si aumenta o disminuye) en una proporción similar, entonces este par estará localizado en la región Z1. Los pares de compuestos localizados en Z2 serán aquellos en los cuales los cambios estructurales aumentan la actividad hacia un receptor, pero la disminuyen hacia el otro; por esta razón, la región Z2 está asociada con un “SAR inverso” y es la zona más importante para estudios de selectividad. Los pares de compuestos localizados en Z3 y Z4 están asociados con compuestos en los que las diferencias estructurales sólo afectan la actividad contra un receptor, sin cambiar la actividad hacia el otro. Los pares localizados en el centro de los mapas DAD, región Z5, son aquellos en los que las diferencias estructurales tienen un pequeño impacto en la actividad hacia los dos receptores. En términos generales, los puntos que nos brindan más información relacionada con la selectividad hacia algún receptor son los que se encuentran en las regiones Z2, Z3 y Z4. Es importante mencionar que la clasificación de los puntos en los mapas DAD es independiente de la similitud estructural de los compuestos; sin embargo, ésta se puede incorporar empleando un código de color. En este trabajo se utilizó una escala continua de color que va de verde a rojo para indicar baja o alta similitud, respectivamente. El incluir la información de similitud molecular en los mapas DAD y TAD nos permite identificar fácilmente la presencia *activity cliffs*

simples, dobles y triples, así como regiones con SAR continuo para uno o varios receptores [67]. Un *activity cliff* simple es aquel en el que las diferencias estructurales entre las moléculas del par afectan drásticamente (más de 1 unidad logarítmica) la actividad hacia un receptor. Del mismo modo, un *activity cliff* doble es aquel en el que las diferencias estructurales influyen en la actividad hacia dos receptores. Siguiendo el mismo criterio, las diferencias estructurales de un *activity cliff* triple afectarían la actividad hacia tres receptores. Los mapas DAD pueden extenderse a masa de diferencia a actividad triple al añadir una dimensión que representa las diferencias de actividades hacia un tercer receptor. Los mapas TAD se pueden interpretar como el agrupamiento de tres mapas DAD donde cada uno representa dos de los tres receptores.

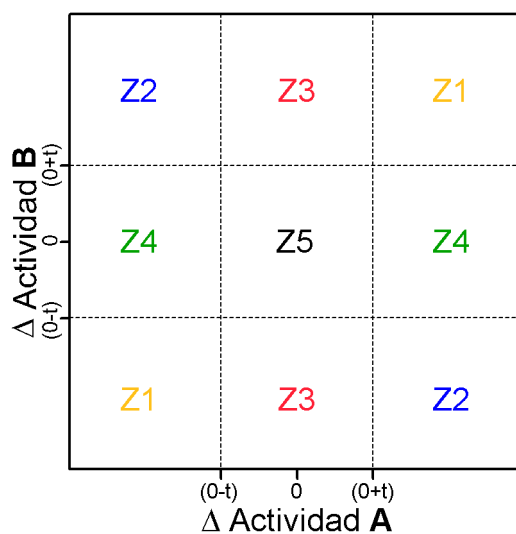


Figura 7.2. Prototipo de un mapa DAD. La región Z1 contiene pares en los cuales los cambios estructurales afectan en la misma proporción a los dos receptores. La región Z2 contiene pares de moléculas con un SAR inverso. La región Z3 y Z4 contiene pares donde los cambios estructurales sólo afectan a uno de los dos receptores. En la región Z5 se encuentran los pares de moléculas en las cuales las diferencias estructurales no afectan la actividad de ninguno de los receptores.

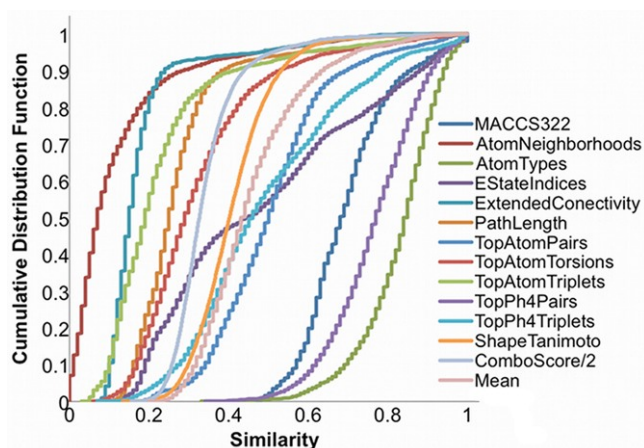
CAPÍTULO 8. RESULTADOS

8.1. Distribución de las medidas de similitud

En la Figura 8.1 se resume en forma de funciones de distribución acumulada la información de las 14,028 similitudes pareadas calculadas con 13 representaciones moleculares 2D y 3D. La tabla en esta figura resume los estadísticos de cada curva indicando el máximo, mínimo, primer y tercer cuartil, mediana, promedio y desviación estándar. En términos generales, la distribución de los descriptores 2D y 3D muestran una variación muy amplia. Se ha observado que otras bases de datos también presentan distribuciones similares a la obtenida en este trabajo [68, 70, 71]. Se observa que los descriptores atom types, TopPh4Pairs y MACCS (322 bits) muestran los valores de similitud más altos con valores en sus de medianas de 0.84, 0.77 y 0.68, respectivamente. Sin embargo, también se encontraron diversos descriptores que presentan una distribución normal, lo cual se puede observar en una curva sigmoidea y en valores de mediana y promedio cercanos a 0.5.

Tabla 8.1 Matriz de correlación entre los descriptores moleculares

	Atom Neighborhoods	AtomTypes	EstateIndices	Extended Conectivity	MACCS322	PathLength	TopAtomPairs	TopAtomTorsions	TopAtomTriplets	TopPh4Pairs	TopPh4Triplets	ShapeTanimoto	ComboScore/2
AtomNeighborhoods	1.000												
AtomTypes	0.491	1.000											
EStateIndices	0.556	0.447	1.000										
ExtendedConectivity	0.896	0.413	0.472	1.000									
MACCS322	0.623	0.365	0.479	0.642	1.000								
PathLength	0.800	0.350	0.456	0.901	0.704	1.000							
TopAtomPairs	0.751	0.804	0.503	0.709	0.612	0.659	1.000						
TopAtomTorsions	0.859	0.601	0.523	0.791	0.583	0.726	0.824	1.000					
TopAtomTriplets	0.865	0.587	0.497	0.845	0.639	0.791	0.920	0.865	1.000				
TopPh4Pairs	0.511	0.415	0.347	0.471	0.527	0.442	0.611	0.474	0.572	1.000			
TopPh4Triplets	0.527	0.383	0.363	0.591	0.435	0.617	0.545	0.496	0.553	0.579	1.000		
ShapeTanimoto	0.310	0.126	0.202	0.334	0.218	0.304	0.281	0.313	0.369	0.159	0.142	1.000	
ComboScore/2	0.480	0.252	0.290	0.497	0.359	0.456	0.455	0.466	0.548	0.285	0.270	0.820	1.000



	Max	Q3	Mediana	Q1	Min	Mean	STD
AtomNeighborhoods	1.00	0.15	0.07	0.03	0.00	0.12	0.15
AtomTypes	1.00	0.90	0.84	0.77	0.33	0.83	0.10
EStateIndices	1.00	0.68	0.45	0.26	0.06	0.49	0.25
ExtendedConectivity	1.00	0.19	0.15	0.12	0.07	0.18	0.12
MACCS322	1.00	0.76	0.68	0.62	0.45	0.70	0.11
PathLength	1.00	0.31	0.25	0.20	0.10	0.28	0.14
TopologicalAtomPairs	1.00	0.58	0.50	0.41	0.13	0.51	0.15
TopologicalAtomTorsions	1.00	0.39	0.29	0.22	0.05	0.33	0.16
TopologicalAtomTriplets	1.00	0.26	0.19	0.13	0.03	0.23	0.16
TopologicalPharmacophoreAtomPairs	1.00	0.85	0.77	0.69	0.34	0.77	0.11
TopologicalPharmacophoreAtomTriplets	1.00	0.63	0.46	0.35	0.07	0.50	0.19
ShapeTanimoto	1.00	0.46	0.40	0.35	0.13	0.41	0.10
ComboScore/2	1.00	0.37	0.33	0.29	0.16	0.34	0.09
Similitud Promedio	1.00	0.52	0.43	0.37	0.20	0.46	0.13

Figura 8.1. Funciones de distribución acumulada para las 14,028 similitudes pareadas calculadas con 13 representaciones moleculares.

La Tabla 8.1 muestra los coeficientes de correlación de Pearson, en forma de matriz, calculados para cada par de los 13 descriptores 2D y 3D. Se observa una alta correlación entre los siguientes descriptores: topological atom pairs y topological atom triplets; atom neighborhoods y extended connectivity; path length y extended connectivity; topological atom triplets y topological atom torsions; topological atom triplets y atom neighborhoods; topological atom torsions y atom neighborhoods. Todos ellos con un coeficiente de correlación mayor a 0.85. El descriptor EStateIndices presentó un coeficiente de correlación menor a 0.55 con el resto de las representaciones moleculares. Este se puede explicar con base en

el diseño de este descriptor, el cual mide el estado electrónico de un átomo tomando en cuenta la perturbación electrónica debido a la influencia de los demás átomos en la molécula [98]. Por el contrario, los descriptores que siguen un diseño similar están altamente correlacionados, por ejemplo, topological atom pairs y topological atom triplets con un coeficiente de correlación de 0.92.

Una estrategia para reducir la dependencia que existe entre las representaciones estructurales y el espacio químico es combinar las diferentes medidas de similitud obtenidas por diferentes métodos. En este trabajo se utilizaron los descriptores EStateIndices, topological atom pairs, topological pharmacophore triplets, y Combo Score para calcular la similitud promedio. Esta metodología ha sido utilizada anteriormente para generar modelos consenso del *panorama de actividad* empleando otras bases de datos [67, 68, 70]. Estos cuatro descriptores fueron escogidos debido a que capturan aspectos estructurales diferentes al resto de los 13 descriptores como se ha sugerido recientemente [106]. Adicionalmente, estos descriptores muestran una baja correlación entre sí para las 14,028 similitudes calculadas, con un coeficiente de correlación menor a 0.55 (Tabla 8.1).

8.2. Mapas SAS

En la Figura 8.2 se muestran los mapas SAS para PPAR α , δ y γ calculados con EStateIndices, topological atom pairs, topological pharmacophore triplets, y Combo Score. Estos mapas muestran la relación entre la similitud estructural y la similitud en la actividad para un grupo de 168 moléculas agonistas de los PPAR. Cada punto en los mapas SAS representa una de las 14,028 comparaciones pareadas que se calcularon. El número total de compuestos en cada región de los mapas SAS, así como el número de pares consenso para los tres subtipos de PPAR se enlista en la Tabla 8.2. Como era de esperarse, tomando en cuenta las funciones distribución acumulada, los valores de similitud obtenidos con estos descriptores moleculares se encuentran distribuidos a través de todo el eje X. Es interesante notar que topological atom pairs fue la representación molecular 2D

Tabla 8.2. Distribución de los puntos en las diferentes regiones de los mapas SAS

PPAR	Descriptor	Similitud mediana de los activos	I	II	III	IVA(1-2)	IVB>2
Alfa	EStateIndices	0.55	4998	3383	3498	1503	646
	TopAtomPairs	0.45	2790	5591	2427	2462	758
	TopPh4Trip	0.47	4376	4005	3181	1825	641
	ComboScore/2	0.34	4908	3473	3689	1451	507
	Consenso	0.45	1155	1219	989	313	116
Delta	EStateIndices	0.45	3049	3347	4035	1354	2243
	TopAtomPairs	0.56	4059	2337	5785	710	1137
	TopPh4Trip	0.55	3946	2450	5202	811	1619
	ComboScore/2	0.34	3607	2789	4990	1040	1602
	Consenso	0.47	1274	744	1884	105	224
Gamma	EStateIndices	0.47	3347	3212	3933	1950	1586
	TopAtomPairs	0.58	4496	2063	6089	914	466
	TopPh4Trip	0.73	5293	1266	6819	459	191
	ComboScore/2	0.33	3434	3125	4376	1731	1362
	Consenso	0.53	1642	545	2304	181	27

Como se mencionó en el capítulo 7, los descriptores mostrados en la Figura 8.2 se combinaron con la finalidad de generar una medida de similitud denominada similitud promedio. Los mapas SAS construidos con la similitud promedio se pueden observar en la Figura 8.3. En esta misma figura se encuentran marcados ejemplos representativos de pares consenso y sus estructuras químicas se muestran en la Figura 8.4. Como ejemplo de la región I tenemos a los pares **11_83**, **13_114** y **99_155** todos ellos con baja similitud estructural, pero con actividades biológicas muy parecidas; por lo tanto, estos compuestos están asociados con *scaffold hopping*. Es interesante notar que el par **11_83** no permanece en la misma región relativa al utilizar diferentes descriptores moleculares, sino que también entre los diferentes subtipos de PPAR. Los pares **1_10**, **11_16** y **59_63** se localizan en la región II, donde los pares tienen alta similitud molecular y pCl₅₀ muy parecidas por lo que estos pares están asociados con una región de SAR continuo. Finalmente, los pares **73_100** y **58_103** son buenos ejemplos de la región IV o *activity cliffs* debido a que presentan alta similitud molecular, pero baja similitud de actividad, por lo que están localizados en una región de SAR discontinuo del *panorama de actividad*.

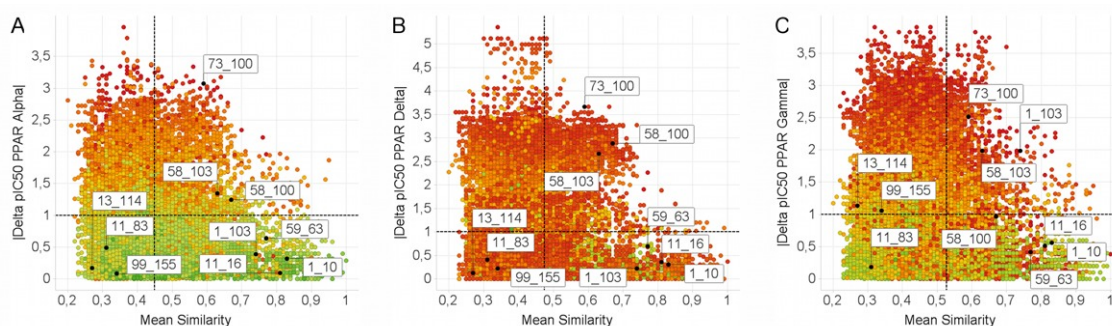


Figura 8.3. Mapas SAS para PPAR α , δ y γ calculados con la similitud promedio. Cada mapa muestra 14,028 puntos resultantes de las comparaciones pareadas de 168 compuestos. El color de cada punto corresponde a la actividad del compuesto más activo en el par y va de rojo (más activos) a verde (menos activos).

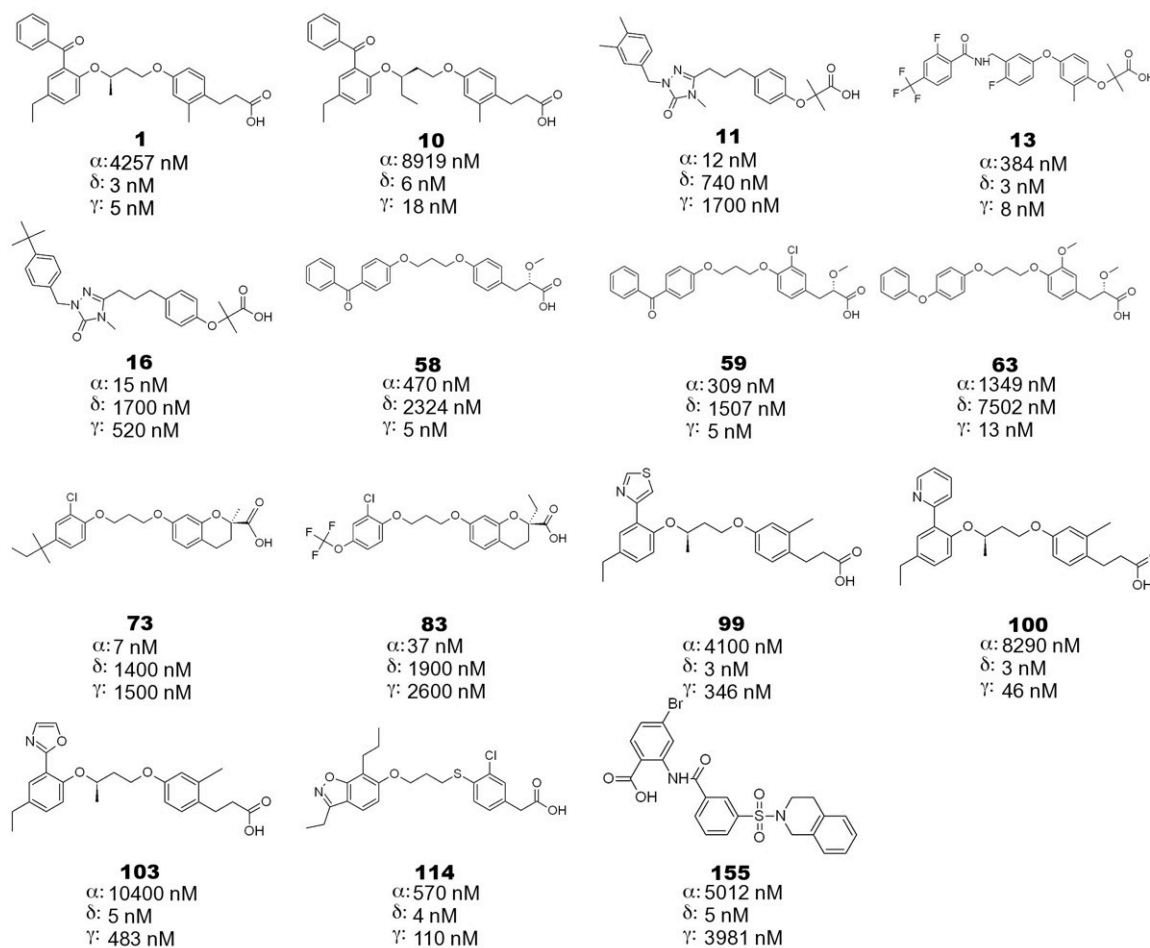


Figura 8.4. Estructuras químicas y valores de pCl₅₀ correspondientes a los pares representativos de los mapas SAS. La posición de cada par en los mapas SAS se muestra en la Figura 8.3.

Vale la pena mencionar que algunos pares consenso que aparecen en una cierta región del mapa SAS para algún receptor pueden aparecer en regiones diferentes en los mapas para los demás receptores. Un ejemplo de este comportamiento es el par **1_103**, el cual es un par consenso que aparece en la región II para los PPAR α y δ , pero es un *activity cliff* (región IV) para PPAR γ . De igual manera el par **58_100** el cual se localiza en la región IV para PPAR α y δ , aparece en la región II para PPAR γ . Este comportamiento se asocia con compuestos selectivos hacia algún subtipo de PPAR, donde las pequeñas modificaciones estructurales disminuyen o incrementan la actividad hacia un receptor en específico. Por esta razón, el estudio de los pares que presentan este comportamiento puede dar lugar a un mejor entendimiento de los requerimientos estructurales necesarios para obtener compuestos selectivos.

8.3. Mapas DAD y TAD

Los mapas DAD y TAD mostrados en la Figura 8.5 fueron construidos con la finalidad de obtener una descripción más detallada del *panorama de actividad*, además de que nos brindan caracterización sistemática del perfil de selectividad. Cada mapa muestra las 14,028 diferencias de actividades entre los 168 compuestos de la base de datos. A diferencia de los mapas SAS, en esta ocasión la distribución de los puntos en las gráficas es independiente de la similitud estructural entre los compuestos. Los puntos fueron coloreados con base en la similitud promedio del par utilizando una escala continua que va de los más similares (rojo) a los menos similares (verde). Al analizar los compuestos de la base de datos utilizando los mapas DAD podemos identificar algunos compuestos con alta similitud estructural que están asociados con diferencias de actividades muy grandes lo que facilita la identificación de *activity cliffs* simples, dobles o triples y *scaffold hops*. Como se ha mencionado anteriormente, un *activity cliff* simple es aquel en el que las diferencias estructurales entre las moléculas del par afectan drásticamente (más de 1 unidad logarítmica) la actividad hacia un

receptor. Del mismo modo, un *activity cliff* doble es aquel en el que las diferencias estructurales influyen en la actividad hacia dos receptores. Siguiendo el mismo criterio, las diferencias estructurales de un *activity cliff* triple afectarían la actividad hacia los tres receptores.

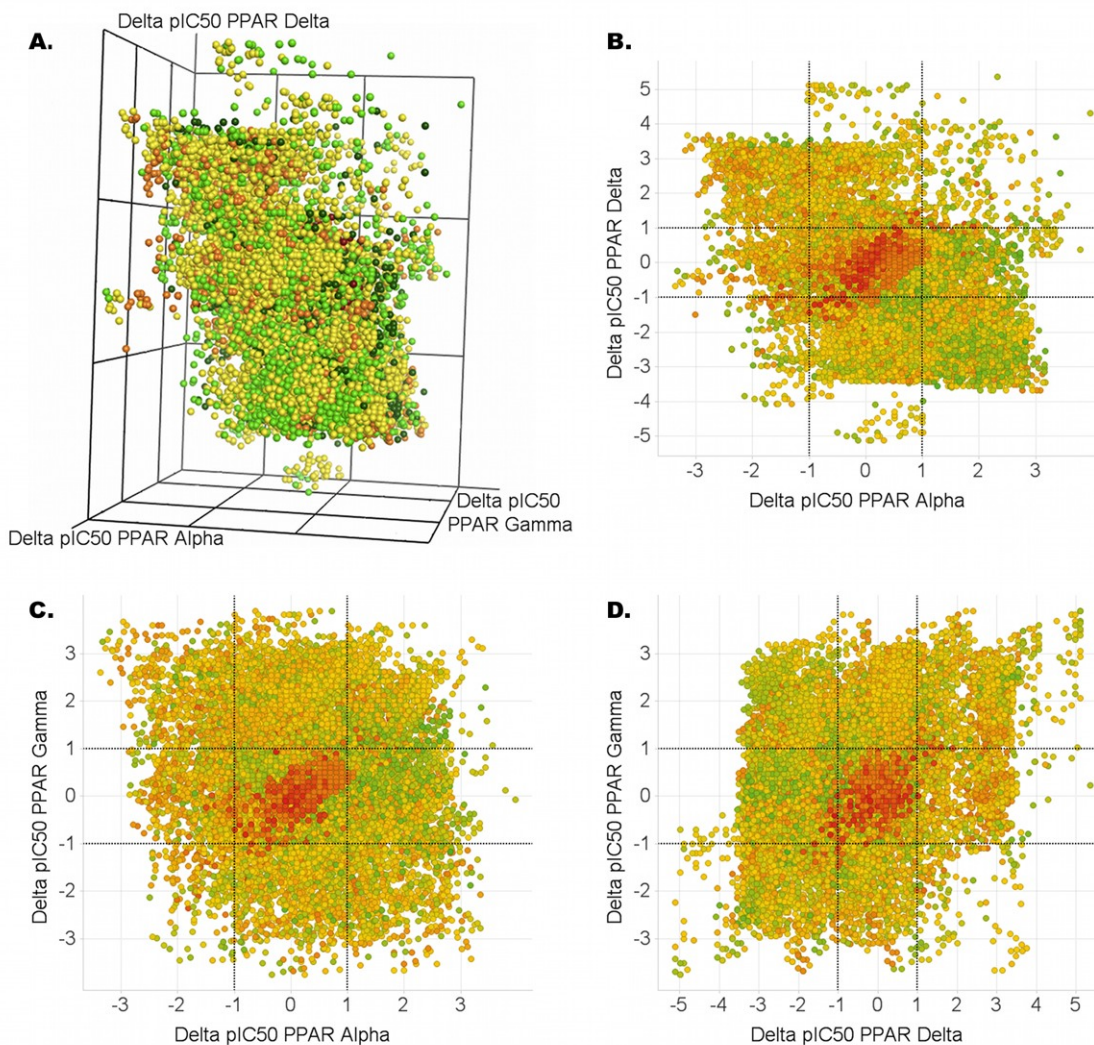


Figura 8.5. (A) Mapas de diferencia de actividad tripe para los tres subtipos de PPAR. Mapas de diferencia de actividad doble para PPAR: (B) α - δ , (C) α - γ y (D) δ - γ . Cada mapa contiene 14,028 puntos resultantes de las comparaciones pareadas de 168 compuestos. Los puntos están coloreados con base en la similitud promedio siguiendo una escala continua de verde (poco similares) a rojo (muy similares).

Como era de esperarse, la mayoría de los compuestos con alta similitud estructural se encuentran localizados en la región Z5 de los mapas DAD y la similitud disminuye a medida que los pares se alejan del centro de la gráfica. No obstante, es interesante que algunos compuestos muy similares no presenten este comportamiento ya que se encuentran en las regiones Z1-Z4 de la gráfica. Como se explica en el capítulo 7, las regiones Z1-Z4 de los mapas DAD indican una gran diferencia en la actividad biológica que al ser combinada con una alta similitud estructural puede dar indicios de características estructurales asociadas con la selectividad hacia uno o dos receptores.

En vista del gran número de puntos en cada mapa DAD y lo difícil que sería analizarlos todos simultáneamente, se optó por generar mapas que sólo muestran compuestos con alta o baja similitud (similitud promedio > 0.75 y < 0.25 , respectivamente). La Figura 8.6 muestra los mapas DAD y TAD que incluyen sólo los 534 pares de moléculas con la similitud estructural más alta. Un total de 393 pares tienen una diferencia de actividad menor a 1 unidad logarítmica hacia los tres subtipos de PPAR, por lo que están asociados con un SAR continuo; 78 pares fueron identificados como *activity cliffs* simples, 48 pares como *activity cliffs* dobles con SAR directo, 4 pares como *activity cliffs* dobles con SAR inverso, y 11 pares como *activity cliffs* triples. La Tabla 8.3 resume el número de *activity cliffs* simples, dobles y triples identificados en los mapas DAD y TAD de cada subtipo de PPAR. Para los 168 compuestos estudiados en este trabajo, el mayor número de *activity cliffs* para PPAR γ , comparado con el número de *activity cliffs* presentados en PPAR α y δ , sugiere que el SAR para PPAR γ es más discontinuo que el de los otros dos receptores. Estos resultados indican que sería más difícil desarrollar un modelo QSAR que cuente con buenas propiedades predictivas para PPAR γ . Cabe señalar que 21 de los 41 *activity cliffs* para PPAR γ tienen una diferencia de actividad mayor de 1.5 unidades logarítmicas, de los cuales 6 presentan más de 2 unidades logarítmicas de diferencia. Por otro lado, para PPAR α y δ sólo se detectó un pequeño número de *activity cliffs* con una gran diferencia de actividad.

Es interesante que el mapa DAD α - δ fue el que presentó menor número de *activity cliffs* dobles, comparado con los otros dos. Este resultado sugiere que el SAR de los 168 compuestos probados contra PPAR α y δ es muy parecido. Por el contrario, el mapa DAD α - γ fue el que presentó mayor número de *activity cliffs* dobles sugiriendo que el SAR de los compuestos probados contra estos subtipos de PPAR es muy diferente.

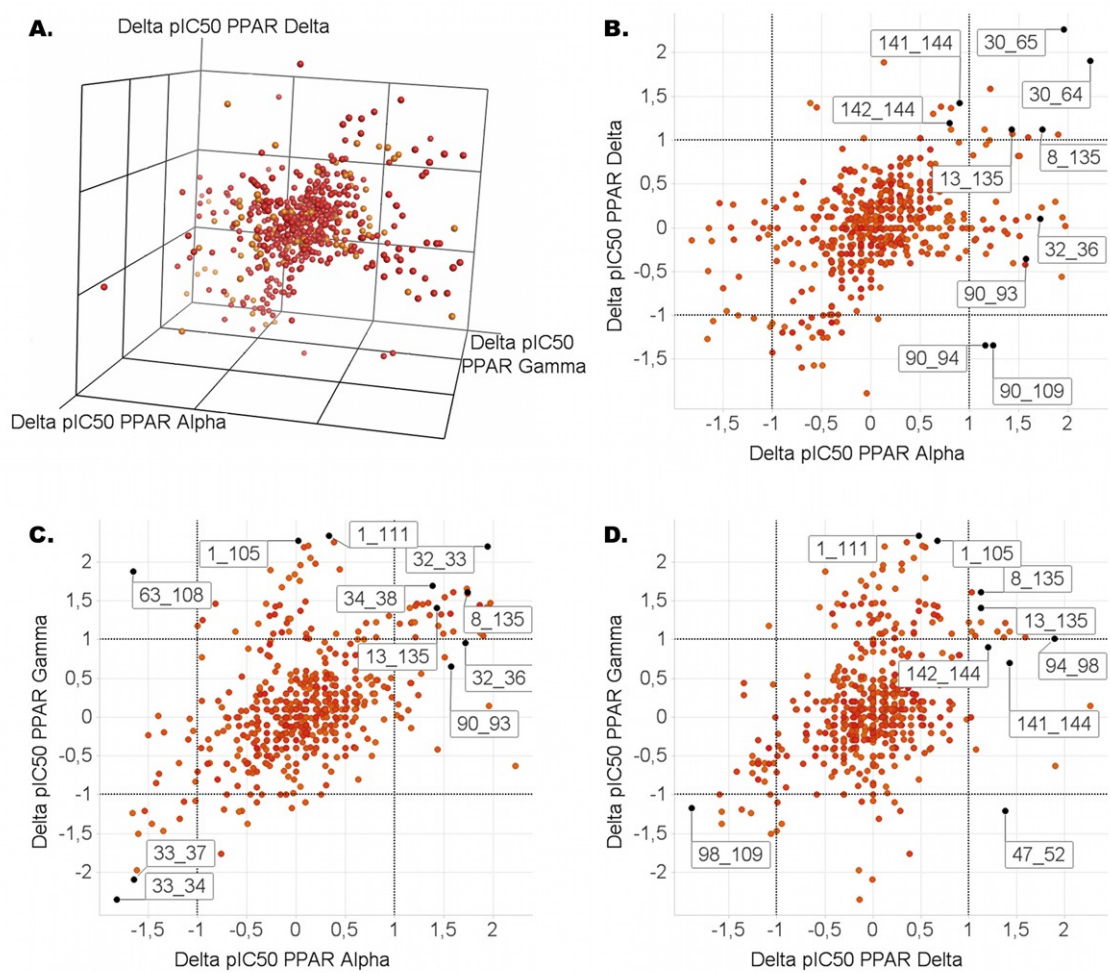


Figura 8.6. (A) Mapas de diferencia de actividad tripe para los tres subtipos de PPAR. Mapas de diferencia de actividad doble para PPAR: (B) α - δ , (C) α - γ y (D) δ - γ . Cada mapa muestra 534 puntos, todos ellos con una similitud promedio > 0.75. Los puntos seleccionados son ejemplos de *activity cliffs*.

Tabla 8.3. Conteo de *activity cliffs* y de *scaffold hops*

Tipo	Subtipo de PPAR	<i>Scaffold Hops</i>		<i>Activity Cliffs</i> ^a	
		Total	Total	Directo	Inverso
Simple	α	14	18 (3)		
	δ	20	19 (0)		
	γ	30	41 (15) ^b		
Doble	α - δ	4	8 (2)	6 (2)	2 (0)
	α - γ	23	31 (16)	30 (15)	1 (1)
	δ - γ	7	13 (5)	12 (5)	1 (0)
Triple	α - δ - γ	11	11 (7)	10 (7)	1 (0)

^aEl número en el paréntesis indica *activity cliffs* con diferencia de actividad mayor a 1.5 unidades logarítmicas

^bSeis *activity cliffs* tienen una diferencia de actividad mayor a 2 unidades logarítmicas

8.4. *Activity cliffs* simples

La Figura 8.7 muestra algunos ejemplos de *activity cliffs* simples para los tres subtipos de PPAR. Los pares **32_36** y **90_93** (Figura 8.7A) son ejemplos representativos de *activity cliffs* simples para PPAR α . En ambos casos, las moléculas que componen el par tiene una alta similitud estructural (similitud promedio > 0.75). Las únicas diferencias que existen es la adición de un metilo en el primer par y el cambio de una amina secundaria por una terciara en el segundo. Estos pequeños cambios estructurales están asociados con una diferencia de actividad mayor de 1.5 unidades logarítmicas hacia PPAR α . De manera similar, los pares **141_144** y **142_144** (Figura 8.7B) son ejemplos de *activity cliffs* simples para PPAR δ . En estos casos sustituir el grupo metoxi por un metilo o un bromo es suficiente para producir una diferencia significativa en la actividad hacia PPAR δ . Por último, los pares **1_105** y **1_111** (Figura 8.7C), donde el cambio del fenilcarbonil en **1** por un furano en **105** o un tiofeno en **111** reduce la actividad en más de dos unidades logarítmicas, son buenos ejemplos de *activity cliffs* simples para PPAR γ . Es importante mencionar que en todos estos ejemplos las diferencias estructurales sólo afectan la actividad hacia un subtipo de PPAR con un pequeño impacto en la actividad hacia los otros. La posición relativa de cada ejemplo en los diferentes mapas DAD se encuentra indicada en la Figura 8.6. Así mismo, la

información correspondiente a la diferencia de actividad y la similitud estructural de todos los ejemplos, utilizando diferentes descriptores, se resume en la Tabla 8.4.

Tabla 8.4. Diferencia de actividad y similitud molecular para los *activity cliffs*

Tipo de <i>activity cliff</i>	Par	ΔpCl_{50} α	ΔpCl_{50} δ	ΔpCl_{50} γ	TopAto mPairs	TopPh4 Triplets	EState Indices	Combo Score/2	Mean
Simple α	32_36	1.722	0.106	0.961	0.92	0.98	1.00	0.56	0.86
Simple α	90_93	1.574	-0.353	0.654	0.85	0.92	0.98	0.43	0.79
Simple δ	141_144	0.900	1.421	0.700	0.95	0.91	0.99	0.65	0.87
Simple δ	142_144	0.800	1.199	0.900	0.94	0.95	0.99	0.62	0.87
Simple γ	1_111	0.330	0.477	2.338	0.85	0.84	0.84	0.58	0.78
Simple γ	1_105	0.024	0.669	2.274	0.85	0.84	0.81	0.59	0.77
Doble α - δ directo	30_64	2.226	1.903	-0.628	0.86	0.99	0.99	0.22	0.76
Doble α - δ directo	30_65	1.959	2.265	0.150	0.85	0.97	0.99	0.29	0.77
Doble α - δ inverso	90_94	1.158	-1.342	0.439	0.89	0.93	0.96	0.70	0.87
Doble α - δ inverso	90_109	1.243	-1.342	0.281	0.89	0.94	0.97	0.57	0.84
Doble α - γ directo	32_33	1.944	0.117	2.204	0.86	0.92	1.00	0.35	0.78
Doble α - γ directo	34_38	1.385	-0.072	1.691	0.94	0.98	0.98	0.43	0.83
Doble α - γ directo	33_37	-1.643	-0.007	-2.095	0.86	0.92	1.00	0.43	0.80
Doble α - γ directo	33_34	-1.819	-0.139	-2.350	0.79	0.88	0.98	0.51	0.79
Doble α - γ inverso	63_108	-1.653	-0.497	1.878	0.80	0.93	0.82	0.57	0.78
Doble δ - γ directo	94_98	0.130	1.889	1.015	0.87	0.98	0.96	0.45	0.81
Doble δ - γ directo	98_109	-0.044	-1.889	-1.172	0.86	0.94	0.97	0.45	0.81
Doble δ - γ inverso	47_52	-0.554	1.376	-1.204	0.88	0.96	0.93	0.50	0.82
Triple	8_135	1.739	1.125	1.609	0.81	0.99	0.87	0.58	0.81
Triple	13_135	1.429	1.125	1.404	0.81	0.94	0.97	0.53	0.81

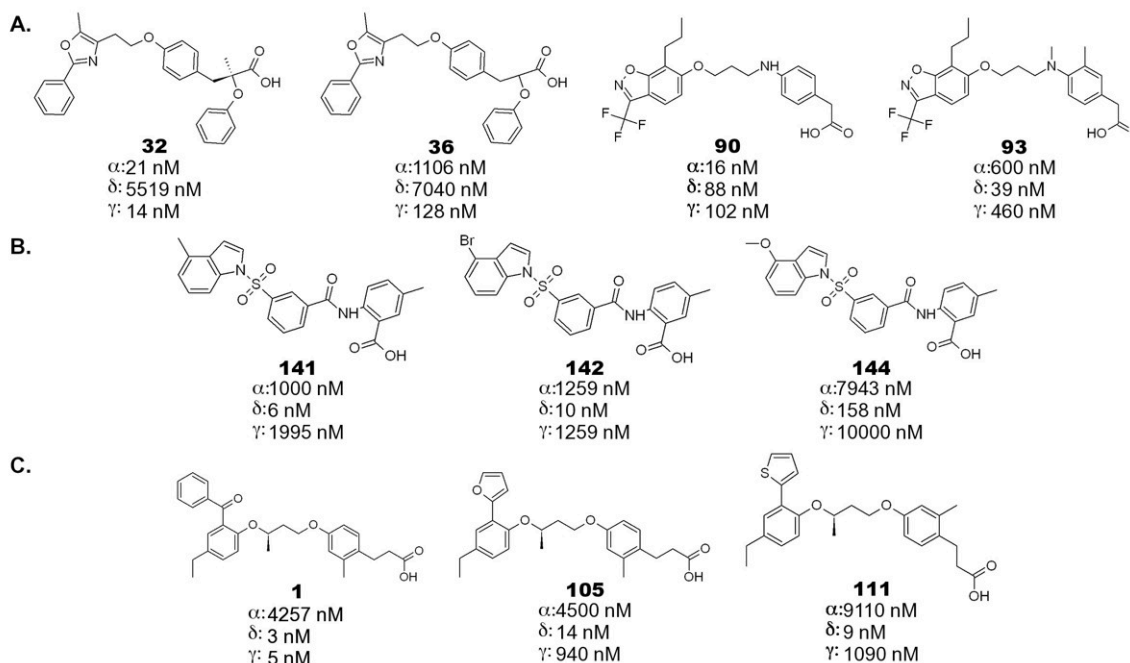


Figura 8.7. Estructuras químicas y valores de pCl_{50} para los *activity cliffs* simples de PPAR (A) α , (B) δ y (C) γ . La posición relativa de cada par en los mapas DAD se indica en la Figura 8.6.

8.5. *Activity cliffs* dobles y triples

La Figura 8.8 muestra las estructuras de algunos ejemplos de *activity cliffs* dobles y triples. Los pares de compuestos **30_64**, **30_65**, **90_94** y **90_109** (Figura 8.8A) son ejemplos de *activity cliffs* dobles para PPAR α y δ . Los pares **30_64** y **30_65** exhiben un SAR continuo ya que están localizados en la región Z1 de los mapas DAD. En otras palabras, los cambios estructurales en estos compuestos afectan de la misma manera la actividad hacia los dos subtipos de PPAR. En estos ejemplos, la apertura del anillo de benzopirano en **64** y **65** para obtener un metoxibenceno, como en el compuesto **30**, resulta en un incremento de la actividad hacia ambos receptores. Por el contrario, los cambios estructurales entre los pares **90_94** y **90_109**, localizados en la región Z2, representan un efecto inverso en la actividad. En estos ejemplos la adición de un cloro y el cambio de una amina secundaria en **90** por un átomo de oxígeno o azufre en **94** y **109**, respectivamente, disminuye la actividad biológica hacia PPAR α , mientras que los

mismos cambios aumentan la actividad hacia PPAR δ . Basados en esta observación podemos sugerir que la actividad biológica hacia PPAR α depende del heteroátomo de la siguiente manera: NHR2 > O > S.

Se detectaron diversos *activity cliffs* dobles para PPAR α y γ . Algunos ejemplos de estos son los pares **32_33**, **33_37**, **33_34**, **34_38**, y **63_108** (Figura 8.8B), de los cuales los primeros cuatro exhiben un SAR directo mientras que el último presenta un SAR inverso. La diferencia estructural principal en los pares **32_33** y **33_37**, localizados en la región Z1 del mapa DAD α - γ , es la longitud de la cadena alifática. Una cadena más larga estaría asociada con una disminución en la actividad hacia PPAR α y γ , pero no tendría efecto en la actividad hacia δ . De la misma manera, el par **33_34** presenta dos diferencias estructurales: la longitud de la cadena alifática y el cambio de un fenilo por un tiofeno. El par **32_34** muestra que el cambio de un fenilo por un tiofeno no tiene impacto sobre la actividad biológica, por lo que la cadena alifática es la principal responsable de este *activity cliff*. De forma similar, en el par **34_38**, que también presenta un SAR directo, la diferencia entre las estructuras de estos compuestos es el feniléter, el cual incrementa la actividad hacia PPAR α y γ . A diferencia de todos los ejemplos anteriores, el par **63_108**, asociado con un SAR inverso, se encuentra localizado en la región Z2 del mapa DAD α - γ . La presencia de halógenos y la falta del grupo metoxi en el compuesto **108** están asociadas con un incremento en la actividad hacia PPAR α y una disminución hacia γ . Por el contrario, la ausencia de halógenos y la presencia del grupo metoxi en el compuesto **63** mejora la actividad hacia PPAR γ y disminuye la actividad hacia PPAR α . Es importante mencionar que estos pares no presentaron una diferencia de actividad significativa hacia PPAR δ ya que están localizados en las regiones Z4 y Z3 de los mapas DAD α - δ y δ - γ , respectivamente.

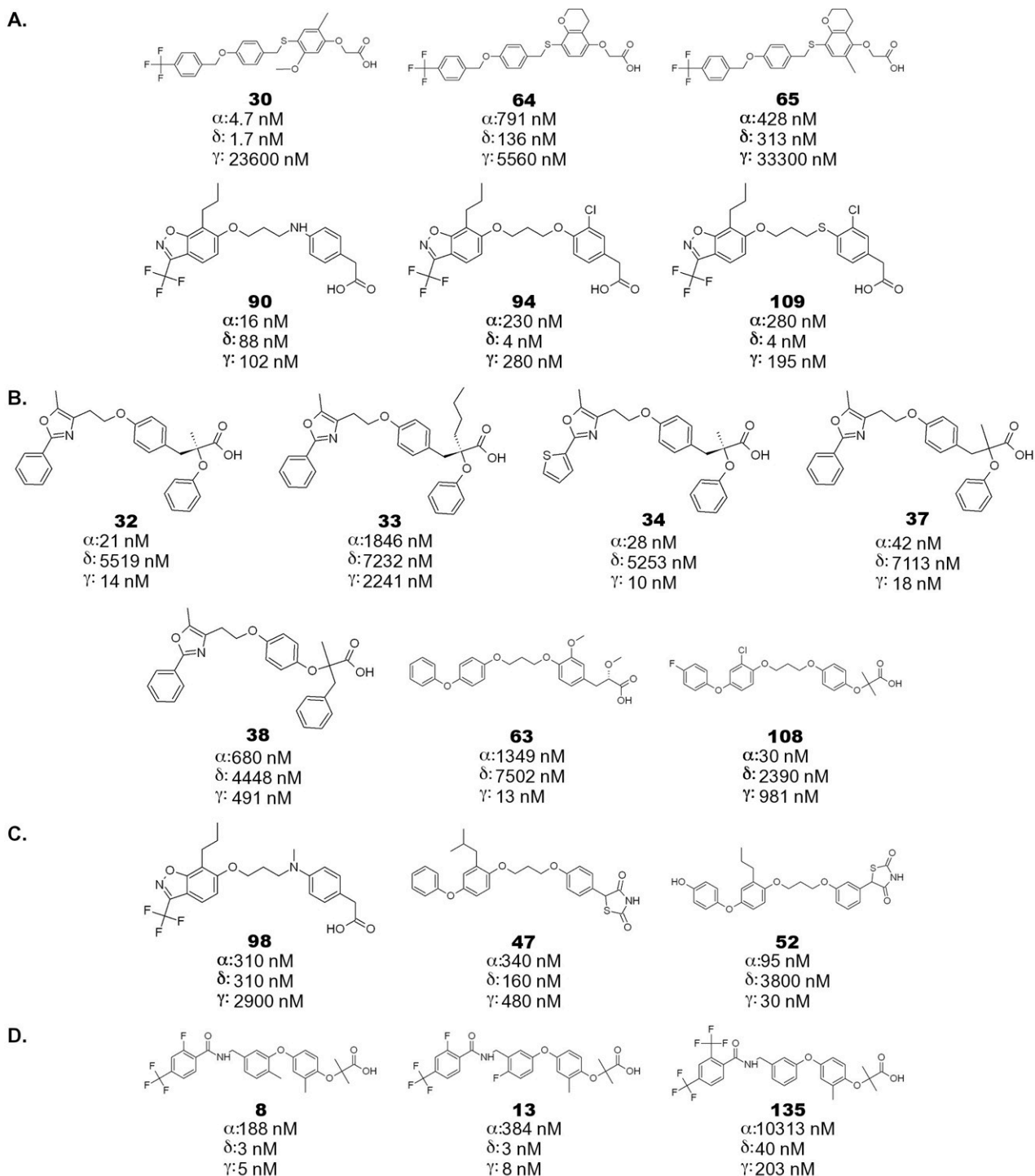


Figura 8.8. Estructuras químicas y valores de pCl_{50} para los *activity cliffs* dobles de PPAR (A) α - δ , (B) α - γ , y (C) δ - γ . El panel D presenta las estructuras de los *activity cliffs* triples. La posición relativa de cada par en los mapas DAD se indica en la Figura 8.6.

Por último, los pares **8_135** y **13_135** (Figura 8.8D) son ejemplos de *activity cliffs* triples, los cuales presentan una alta similitud molecular y se localizan en la región Z1 de los tres mapas DAD (Figura 8.6). En estos pares, el cambio del trifluorometilo en **135** por un simple fluoruro, como en **8** y **13**, incrementa significativamente la actividad hacia los tres subtipos de PPAR.

8.6. Scaffold hops

La Figura 8.9 muestra los mapas DAD y TAD que incluyen sólo los 126 pares de moléculas con baja similitud estructural (similitud promedio > 0.25). De estos, un total de 109 pares tienen una diferencia de actividad menor a 1 unidad logarítmica, por lo que están asociados con *scaffold hopping*. La Tabla 8.3 resume el número de *scaffold hops* simples, dobles y triples. La Figura 8.9 muestra la posición de algunos pares seleccionados en los mapas DAD. Los valores correspondientes a la diferencia de actividad para cada uno de los tres subtipos, así como la similitud estructural empleando diferentes descriptores se encuentra resumida en la Tabla 8.5. La Figura 8.10 muestra las estructuras químicas y los valores de pCl₅₀ para cada molécula en los ejemplos seleccionados en los mapas DAD.

Tabla 8.5. Diferencia de actividad y similitud molecular para los ejemplos de *scaffold hops*

Tipo de <i>scaffold hop</i>	Par	Δ pCl ₅₀ α	Δ pCl ₅₀ δ	Δ pCl ₅₀ γ	TopAto mPairs	TopPh4 Triplets	EStateI ndices	Combo Score/2	Mean
Simple α	7_62	0.082	2.835	2.232	0.33	0.17	0.17	0.35	0.25
Simple α	97_126	0.067	-1.366	-2.632	0.35	0.24	0.19	0.24	0.25
Simple δ	95_151	2.421	0.000	1.169	0.33	0.21	0.16	0.22	0.23
Simple δ	95_165	2.721	0.097	1.369	0.35	0.21	0.18	0.26	0.25
Simple γ	57_168	1.092	-2.635	0.087	0.31	0.13	0.19	0.33	0.24
Doble α - δ directo	31_135	0.319	-0.477	-1.832	0.38	0.29	0.07	0.24	0.25
Doble α - γ directo	57_141	0.092	-3.061	-0.313	0.19	0.20	0.17	0.40	0.24
Doble δ - γ directo	33_83	-1.698	-0.581	0.065	0.44	0.13	0.25	0.19	0.25
Triple	92_143	0.698	-0.455	-0.079	0.28	0.27	0.18	0.29	0.25
Triple	110_131	-0.067	-0.097	-0.691	0.27	0.22	0.20	0.25	0.23

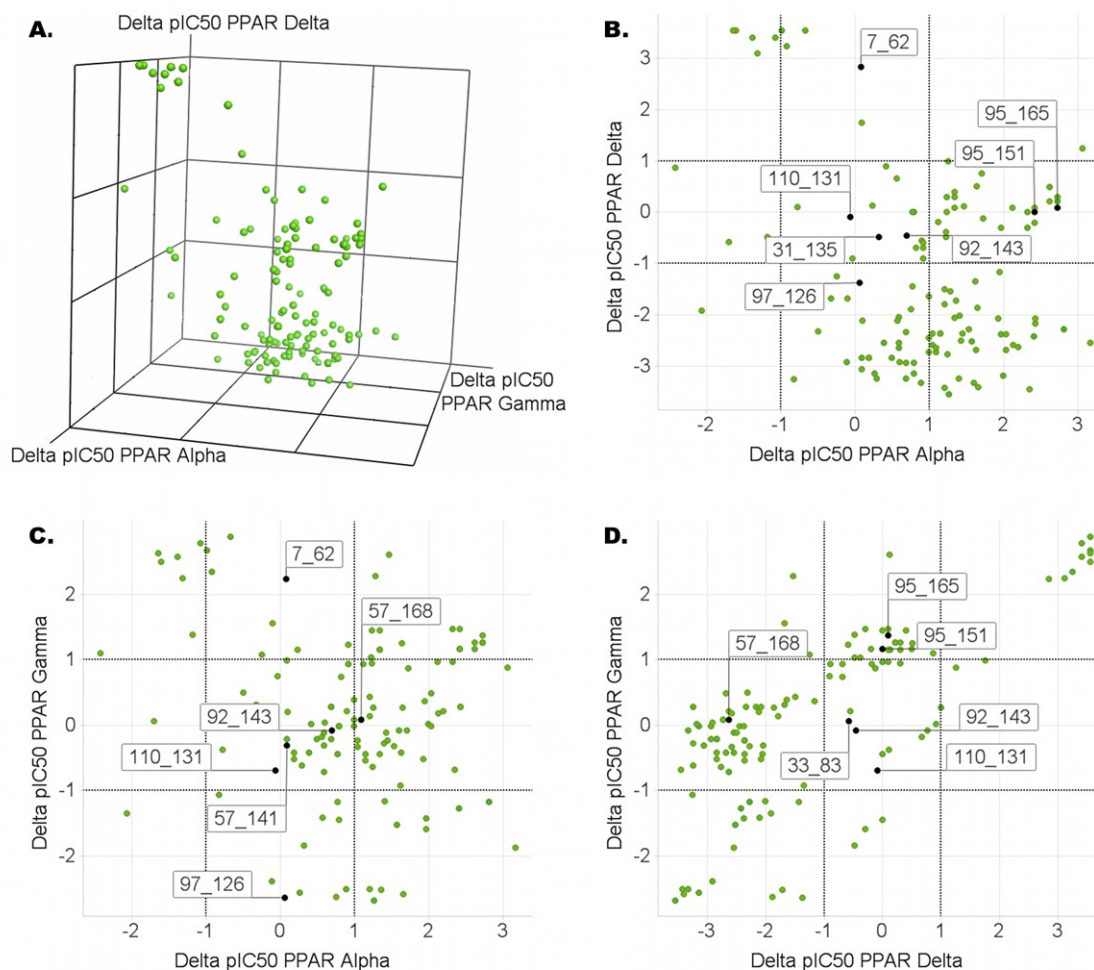


Figura 8.9. (A) Mapas de diferencia de actividad tripe para los tres subtipos de PPAR. Mapas de diferencia de actividad doble para PPAR: (B) α - δ , (C) α - γ y (D) δ - γ . Cada mapa muestra 126 puntos, todos ellos con una similitud promedio < 0.25 . Los puntos seleccionados son ejemplos de *scaffold hops*.

Algunos ejemplos de pares de compuestos con baja similitud molecular y una diferencia de actividad muy pequeña hacia un sólo receptor son los pares **7_62** y **97_126** para PPAR α , los pares **95_151** y **96_165** para PPAR δ , y el par **57_168** para PPAR γ ; todos ellos están localizados en la región Z3 o Z4 de los mapas DAD (Figura 8.9). Adicionalmente, los pares **31_135**, **57_141** y **33_83**, localizados en la región Z5 sólo en uno de los mapas DAD, son ejemplos de *scaffold hops* dobles para α - δ , α - γ y δ - γ , respectivamente. Los pares de compuestos **92_143** y **110_131**, que se encuentran en la región Z5 en los tres mapas DAD, son ejemplos de

scaffold hops triples (Figura 8.10). Vale notar que los compuestos **92**, **143**, **100** y **131**, que conforman los *scaffold hops* triples, presentan selectividad hacia PPAR δ . Este resultado sugiere que la estructura base de estos cuatro compuestos puede ser atractiva para desarrollar compuestos selectivos hacia esta isoforma.

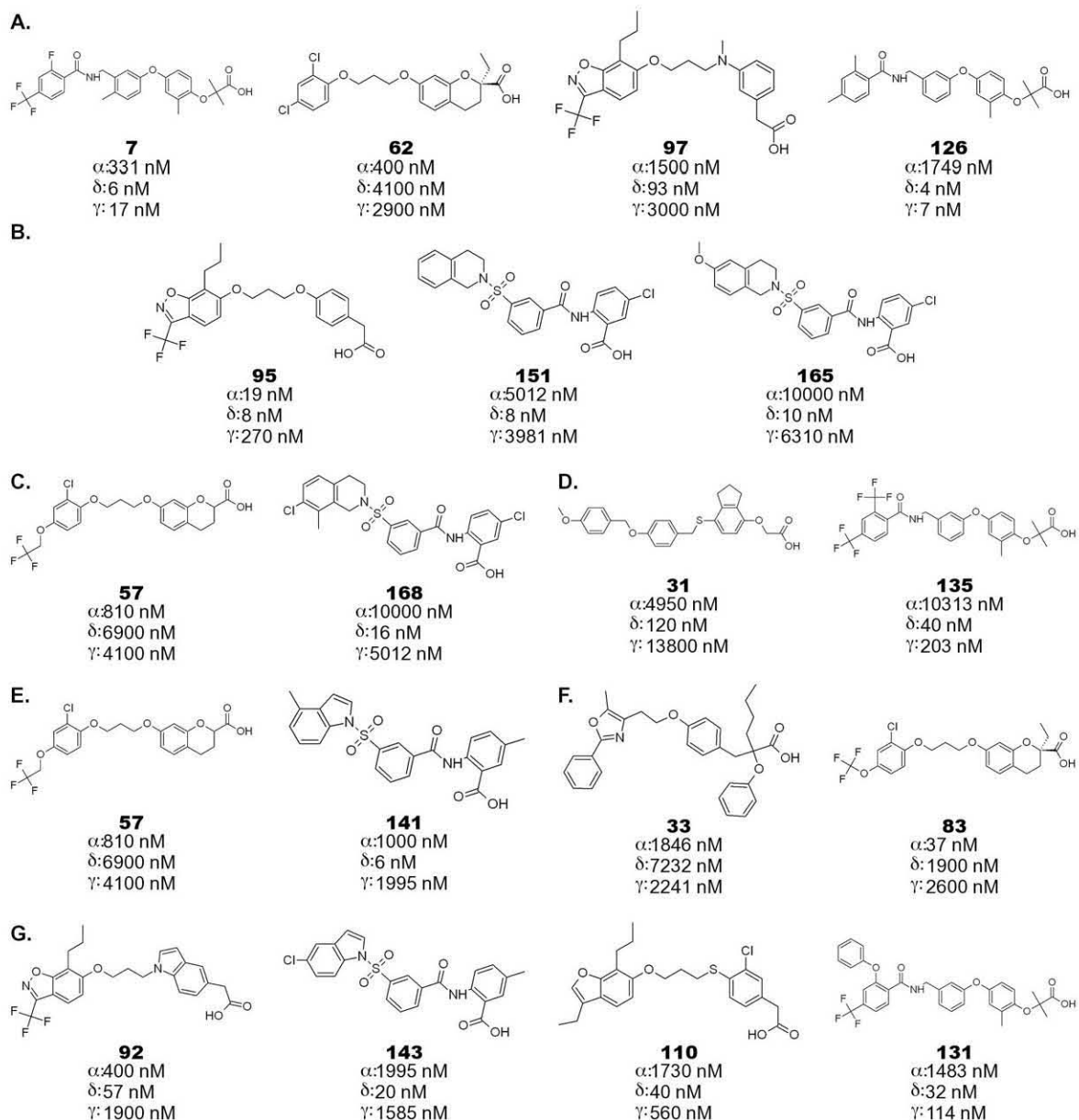


Figura 8.10. Estructuras químicas y valores de pCl₅₀ para *scaffold hops*.

El *scaffold hopping* es una herramienta útil con diversas aplicaciones; por ejemplo, sirve para mejorar propiedades farmacocinéticas, se aplicado para evitar restricciones de propiedad intelectual o simplemente para analizar compuestos que pertenecen a diferentes regiones del espacio químico. Sin embargo, hace falta desarrollar métodos que permitan un análisis más completo y más profundo de los *scaffold hops*.

CAPÍTULO 9. CONCLUSIONES

9.1. Modelado del panorama de actividad para ligandos de PPAR

En este trabajo se empleó el concepto de *panorama de actividad* para describir el SAR de 168 compuestos con actividad reportada hacia tres subtipos de PPAR. El análisis se basó en comparaciones pareadas entre diferencias de actividad y similitud estructural. Con la finalidad de reducir la dependencia que existe entre las representaciones estructurales y el espacio químico, se utilizó una combinación de cuatro descriptores 2D y 3D. El SAR de los compuestos fue analizado utilizando mapas SAS y DAD, los cuales son herramientas útiles para estudiar el SAR de una base de datos. Los mapas SAS se construyeron al combinar los descriptores moleculares con la similitud estructural. Se logró identificar una gran cantidad de pares consenso en los mapas SAS para los tres subtipos de PPAR, la mayoría de ellos se localizan en la región II asociada con un SAR continuo. Sólo una pequeña cantidad de los pares consenso se caracterizó como *activity cliffs* simples, los cuales están asociados con un SAR discontinuo.

En el conjunto de datos se pudieron identificar diversos ejemplos de *activity cliffs* simples, dobles y triples utilizando los mapas DAD y TAD. A partir de los *activity cliffs* se pudieron reconocer algunos determinantes estructurales para la selectividad. Por ejemplo, si tomamos en cuenta que la mayoría de los agonistas de los PPAR se componen por una cola aromática, un espaciador y una cabeza ácida [90, 91], podemos sugerir que los heteroátomos en el espaciador incrementan la actividad hacia PPAR α en el siguiente orden: $\text{HNR}_2 > \text{O} > \text{S} > \text{NR}_3$ (como se observa en los compuestos **90**, **94**, **109** y **98**, respectivamente). Es interesante notar que la actividad hacia los tres subtipos de PPAR se incrementa cuando el ligando contiene NHR_2 en lugar de NR_3 . Esto puede ser consecuencia del comportamiento como aceptor o donador de puente de hidrógeno que presenta la amina secundaria, mientras que la amina terciaria sólo se comporta como aceptor de puente de hidrógeno. Adicionalmente, los resultados sugieren que una cadena alifática larga cercana a la cabeza ácida (por ejemplo el butilo en

el compuesto **33** comparado con el compuesto **32** y **37**) tiende a reducir la actividad biológica hacia PPAR α y γ , convirtiéndola en una característica importante para el diseño de nuevos agonistas selectivos de PPAR δ . Por el contrario, parece ser que una cadena alifática cercana a la cola aromática aumenta la actividad hacia los tres receptores ya que es una característica compartida por muchos compuestos en la base de datos. Debido a su importancia, esta cadena a sido propuesta anteriormente como una característica farmacofórica para los agonistas de los PPAR [107].

Los resultados en este trabajo son un ejemplo de cómo los mapas SAS y DAD/TAD pueden ser utilizados para analizar un conjunto de compuestos con actividad hacia más de un receptor. Cabe señalar que este es un primer intento de analizar sistemáticamente el SAR de algunos agonistas de los PPAR empleando el modelado del *panorama de actividad*.

REFERENCIAS

- [1] I.M. Kapetanovic. Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. *Chemico-Biological Interactions*. 2008, 171, 165-76.
- [2] T.T. Talele, S. Khedkar, A.C. Rigby. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* 2010, 10, 127-41.
- [3] R. Cramer, D. Patterson, J. Bunce. Comparative Molecular-Field Analysis (COMFA) .1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988, 110, 5959-67.
- [4] G. Klebe, U. Abraham, T. Mietzner. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 1994, 37, 4130-46.
- [5] J. Rodrigues-Coura, P. Albajar-Viñas. Chagas disease: a new worldwide challenge. *Nature Outlook*. June 24th, 2010, 56-7.
- [6] J. Clayton. Chagas disease 101. *Nature Outlook*. June 24th, 2010, 54-5.
- [7] J. Cazzulo, V. Stoka, V. Turk. The major cysteine proteinase of *Trypanosoma cruzi*: a valid target for chemotherapy of Chagas disease. *Curr. Pharm. Des.* 2001, 7, 1143-56.
- [8] H.H. Otto, T. Schirmeister. Cysteine proteases and their inhibitors. *Chem. Rev.* 1997, 97, 133-72.
- [9] F. Lecaille, J. Kaleta, D. Brömme. Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* 2002, 102, 4459-88.
- [10] N.D. Rawlings, A.J. Barrett. Families of cysteine peptidases. *Methods Enzymol.* 1994, 244, 461-86.
- [11] M. McGrath, A. Eakin, J. Engel, J. McKerrow, C. Craik, R. Fletterick. The crystal structure of cruzain: a therapeutic target for Chagas' disease. *J. Mol. Biol.* 1995, 247, 251-9.
- [12] J. Nishihira, H. Tachikawa. Theoretical evaluation of a model of the catalytic triads of serine and cysteine proteases by ab initio molecular orbital calculation. *J. Theor. Biol.* 1999, 196, 513-9.
- [13] E. Dufour, A. Storer, R. Menard. Peptide aldehydes and nitriles as transition state analog inhibitors of cysteine proteases. *Biochemistry.* 1995, 34, 9136-43.
- [14] J. Drenth, K.H. Kalk, H.M. Swen. Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry.* 1976, 1-8.
- [15] D. Arad, R. Langridge, P. Kollman. A simulation of the sulfur attack in catalytic pathway of papain using molecular mechanics and semiempirical quantum mechanics. *J. Am. Chem. Soc.* 1990, 112, 491-502.
- [16] W. Welsh, Y. Lin. Discussion of the catalytic pathway of cysteine proteases based on AM1 calculations. *J. Mol. Struct.: THEOCHEM.* 1997, 401, 315-26.
- [17] M. Harrison, N. Burton, I. Hillier. Catalytic mechanism of the enzyme papain: predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.* 1997, 119, 12285-91.

- [18] K. Brak, I.D. Kerr, K.T. Barrett, N. Fuchi, M. Debnath, K. Ang, et al. Nonpeptidic Tetrafluorophenoxymethyl Ketone Cruzain Inhibitors as Promising New Leads for Chagas Disease Chemotherapy. *J. Med. Chem.* 2010, 53, 1763-73.
- [19] W.R. Roush, J. Cheng, B. Knapp-Reed, A. Alvarez-Hernandez, J.H. McKerrow, E. Hansell, et al. Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease cruzain. *Bioorg. Med. Chem. Lett.* 2001, 11, 2759-62.
- [20] W.R. Roush, F.V. González, J.H. McKerrow, E. Hansell. Design and synthesis of dipeptidyl α',β' -epoxy ketones, potent irreversible inhibitors of the cysteine protease cruzain. *Bioorg. Med. Chem. Lett.* 1998, 8, 2809-12.
- [21] F.V. González, J. Izquierdo, S. Rodríguez, J.H. McKerrow, E. Hansell. Dipeptidyl- α,β -epoxyesters as potent irreversible inhibitors of the cysteine proteases cruzain and rhodesain. *Bioorg. Med. Chem. Lett.* 2007, 17, 6697-700.
- [22] W.R. Roush, A.A. Hernandez, J.H. McKerrow, P.M. Selzer, E. Hansell, J.C. Engel. Design, Synthesis and Evaluation of d-Homophenylalanyl Epoxysuccinate Inhibitors of the Trypanosomal Cysteine Protease Cruzain. *Tetrahedron.* 2000, 56, 9747-62.
- [23] N.C. Romeiro, G. Aguirre, P. Hernández, M. González, H. Cerecetto, I. Aldana, et al. Synthesis, trypanocidal activity and docking studies of novel quinoxaline-N-acylhydrazones, designed as cruzain inhibitors candidates. *Bioorg. Med. Chem.* 2009, 17, 641-52.
- [24] A.C.L. Leite, R.S. de Lima, D.R.d.M. Moreira, M.V.d.O. Cardoso, A.C. Gouveia de Brito, L.M. Farias dos Santos, et al. Synthesis, docking, and in vitro activity of thiosemicarbazones, aminoacyl-thiosemicarbazides and acyl-thiazolidones against *Trypanosoma cruzi*. *Bioorg. Med. Chem.* 2006, 14, 3749-57.
- [25] B.T. Mott, R.S. Ferreira, A. Simeonov, A. Jadhav, K.K.-H. Ang, W. Leister, et al. Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB. *J. Med. Chem.* 2010, 53, 52-60.
- [26] M. Tarnowska, S. Oldziej, A. Liwo, P. Kania, F. Kasprzykowski, Z. Grzonka. MNDO study of the mechanism of the inhibition of cysteine proteinases by diazomethyl ketones. *Eur. Biophys. J.* 1992, 21, 217-22.
- [27] R. Shankar, P. Kolandaivel, K. Senthilkumar. Reaction mechanism of cysteine proteases model compound HSH with diketone inhibitor PhCOCOCH₃-nX_n, (X= F, Cl, n= 0, 1, 2). *Int. J. Quantum Chem.* 2010, 110, 1660-74.
- [28] S. Vijayakumar, P. Kolandaivel. Reaction mechanism of HSH and CH₃SH with NH₂CH₂COCH₂X (X= F and Cl) molecules. *Int. J. Quantum Chem.* 2008, 108, 927-36.
- [29] G. Barreiro, R. De Alencastro, J. Neto. A semiempirical study on leupeptin: An inhibitor of cysteine proteases. *Int. J. Quantum Chem.* 1997, 65, 1125-34.
- [30] R. Vicik, H. Helten, T. Schirmeister, B. Engels. Rational design of aziridine-containing cysteine protease inhibitors with improved potency: studies on inhibition mechanism. *ChemMedChem.* 2006, 1, 1021-8.
- [31] O. Méndez-Lucio, A. Romo-Mancillas, J.L. Medina-Franco, R. Castillo. Computational study on the inhibition mechanism of cruzain by nitrile-containing molecules. *J. Mol. Graphics Modell.* 2012, 35, 28-35.
- [32] M.K. Lindvall. Molecular modeling in cysteine protease inhibitor design. *Curr. Pharm. Des.* 2002, 8, 1673-81.

- [33] O. Mendez-Lucio, J. Perez-Villanueva, A. Romo-Mancillas, R. Castillo. 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). *Med. Chem. Commun.* 2011, 2, 1058-65.
- [34] M. Prokop, J. Adam, Z. Kriz, M. Wimmerova, J. Koca. TRITON: a graphical tool for ligand-binding protein engineering. *Bioinformatics.* 2008, 24, 1955-6.
- [35] MOPAC2009, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net), 2008.
- [36] J.J.P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* 2007, 13, 1173-213.
- [37] J.J.P. Stewart. Application of the PM6 method to modeling proteins. *J. Mol. Model.* 2009, 15, 765-805.
- [38] J.J.P. Stewart. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int. J. Quantum Chem.* 1996, 58, 133-46.
- [39] J.J.P. Stewart. Calculation of the geometry of a small protein using semiempirical methods. *J. Mol. Struct.: THEOCHEM.* 1997, 401, 195-205.
- [40] A. Klamt, G. Schuurmann. COSMO - A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2.* 1993, 799-805.
- [41] C.G. Broyden. The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Its Appl.* 1970, 6, 222-31.
- [42] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal.* 1970, 13, 317-22.
- [43] D. Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.* 1970, 24, 23-6.
- [44] D.F. Shanno. Conditioning of Quasi-Newton methods for function minimization. *Math. Comput.* 1970, 24, 647-56.
- [45] SYBYL 8.0, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- [46] A. Golbraikh, A. Tropsha. Beware of q²! *J. Mol. Graphics Modell.* 2002, 20, 269-76.
- [47] A. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 1993, 98, 5648.
- [48] C. Lee, W. Yang, R.G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.* 1988, 37, 785.
- [49] L. Simon, J.M. Goodman. How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals. *Org. Biomol. Chem.* 2011, 9, 689-700.
- [50] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, et al. Gaussian 09, Revision A.1. Gaussian, Inc. Eds., Wallingford CT, 2009.
- [51] R.G. Parr, W. Yang. Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.* 1984, 106, 4049-50.

- [52] J.L. Gázquez. Perspectives on the density functional theory of chemical reactivity. *J. Mex. Chem. Soc.* 2008, 52, 3-10.
- [53] M. Boháč, Y. Nagata, Z. Prokop, M. Prokop, M. Monincová, M. Tsuda, et al. Halide-stabilizing residues of haloalkane dehalogenases studied by quantum mechanic calculations and site-directed mutagenesis. *Biochemistry.* 2002, 41, 14272-80.
- [54] E. Dyguda, B. Szefczyk, W. Sokalski. The mechanism of phosphoryl transfer reaction and the role of active site residues on the basis of ribokinase-like kinases. *Int. J. Mol. Sci.* 2004, 5, 141-53.
- [55] P. Banáš, M. Otyepka, P. Jeřábek, M. Petřek, J. Damborský. Mechanism of enhanced conversion of 1, 2, 3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. *J. Comput.-Aided Mol. Des.* 2006, 20, 375-83.
- [56] H. Yu, J. Liu, Z. Shen. Modeling catalytic mechanism of nitrile hydratase by semi-empirical quantum mechanical calculation. *J. Mol. Graphics Modell.* 2008, 27, 522-8.
- [57] I. Schechter, A. Berger. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochemical and Biophysical Research Communications.* 1968, 32, 898-902.
- [58] R.R. Mittal, R.A. McKinnon, M.J. Sorich. Effect of steric molecular field settings on CoMFA predictivity. *J. Mol. Model.* 2008, 14, 59-67.
- [59] R.M. Oballa, J.-F. Truchon, C.I. Bayly, N. Chauret, S. Day, S. Crane, et al. A generally applicable method for assessing the electrophilicity and reactivity of diverse nitrile-containing compounds. *Bioorg. Med. Chem. Lett.* 2007, 17, 998-1002.
- [60] G.N. Sastry, T. Bally, V. Hrouda, P. Cársky. The C₄H₆^{•+} potential energy surface. 1. The ring-opening reaction of cyclobutene radical cation and related rearrangements. *J. Am. Chem. Soc.* 1998, 120, 9323-34.
- [61] J. Moon, R. Coleman, R. Hanzlik. Reversible covalent inhibition of papain by a peptide nitrile. Carbon-13 NMR evidence for a thioimidate ester adduct. *J. Am. Chem. Soc.* 1986, 108, 1350-1.
- [62] A.M. Wassermann, M. Wawer, J.r. Bajorath. Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* 2010, 53, 8209-23.
- [63] J. Bajorath, L. Peltason, M. Wawer, R. Guha, M.S. Lajiness, J.H. Van Drie. Navigating structure-activity landscapes. *Drug Discovery Today.* 2009, 14, 698-705.
- [64] G.M. Maggiora. On outliers and activity cliffs – Why QSAR often disappoints. *J. Chem. Inf. Model.* 2006, 46, 1535.
- [65] V. Shanmugasundaram, G.M. Maggiora. Characterizing property and activity landscapes using an information-theoretic approach. In: 222nd American Chemical Society National Meeting, 2001.
- [66] J. Pérez-Villanueva, R. Santos, A. Hernandez-Campos, M.A. Giulianotti, R. Castillo, J.L. Medina-Franco. Structure-activity relationships of benzimidazole derivatives as antiparasitic agents: Dual activity-difference (DAD) maps. *Med. Chem. Commun.* 2010, 2, 44-9.

- [67] J.L. Medina-Franco, A.B. Yongye, J. Pérez-Villanueva, R.A. Houghten, K. Martínez-Mayorga. Multitarget structure–activity relationships characterized by activity-difference maps and consensus similarity measure. *J. Chem. Inf. Model.* 2011, 51, 2427-39.
- [68] J. Pérez-Villanueva, R. Santos, A. Hernández-Campos, M.A. Giulianotti, R. Castillo, J.L. Medina-Franco. Towards a systematic characterization of the antiprotozoal activity landscape of benzimidazole derivatives. *Bioorg. Med. Chem.* 2010, 18, 7380-91.
- [69] J. Waddell, J.L. Medina-Franco. Bioactivity landscape modeling: Chemoinformatic characterization of structure–activity relationships of compounds tested across multiple targets. *Bioorg. Med. Chem.* 2011, doi:10.1016/j.bmc.2011.11.051.
- [70] A.B. Yongye, K. Byler, R. Santos, K. Martínez-Mayorga, G.M. Maggiora, J.L. Medina-Franco. Consensus models of activity landscapes with multiple chemical, conformer, and property representations. *J. Chem. Inf. Model.* 2011, 51, 1259-70.
- [71] J. Medina-Franco, K. Martínez-Mayorga, A. Bender, R. Marín, M. Giulianotti, C. Pinilla, et al. Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. *J. Chem. Inf. Model.* 2009, 49, 477-91.
- [72] D. Stumpfe, J. Bajorath. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* 2012, 55, 2932-42.
- [73] S.R. Langdon, P. Ertl, N. Brown. Bioisosteric Replacement and Scaffold Hopping in Lead Generation and Optimization. *Molecular Informatics.* 2010, 29, 366-85.
- [74] G.W. Bemis, M.A. Murcko. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* 1996, 39, 2887-93.
- [75] H.-J. Böhm, A. Flohr, M. Stahl. Scaffold hopping. *Drug Discovery Today: Technologies.* 2004, 1, 217-24.
- [76] Y.C. Martin, J.L. Kofron, L.M. Traphagen. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* 2002, 45, 4350-8.
- [77] J.E. Shaw, R.A. Sicree, P.Z. Zimmet. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res. Clin. Pract.* 2010, 87, 4-14.
- [78] G.A. Colditz, W.C. Willett, A. Rotnitzky, J.E. Manson. Weight gain as a risk factor for clinical diabetes mellitus in women. *Ann. Intern. Med.* 1995, 122, 481-6.
- [79] J.M. Chan, E.B. Rimm, G.A. Colditz, M.J. Stampfer, W.C. Willett. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care.* 1994, 17, 961-9.
- [80] D.E. Moller. New drug targets for type 2 diabetes and the metabolic syndrome. *Nature.* 2001, 414, 821-7.
- [81] T.M. Willson, P.J. Brown, D.D. Sternbach, B.R. Henke. The PPARs: From orphan receptors to drug discovery. *J. Med. Chem.* 2000, 43, 527-50.
- [82] I.J. McEwan. Nuclear receptors: One big family. *Methods Mol. Biol.* 2009, 505, 3-18.
- [83] P. Arck, B. Toth, A. Pestka, U. Jeschke. Nuclear receptors of the peroxisome proliferator-activated receptor (PPAR) family in gestational diabetes: from animal models to clinical trials. *Biol. Reprod.* 2010, 83, 168-76.

- [84] O. Hyyti, M. Portman. Molecular mechanisms of cross-talk between thyroid hormone and peroxisome proliferator activated receptors: Focus on the heart. *Cardiovasc. Drugs Ther.* 2006, 20, 463-9.
- [85] D.H. van Raalte, M. Li, P.H. Pritchard, K.M. Wasan. Peroxisome proliferator-activated receptor (PPAR)- α : A pharmacological target with a promising future. *Pharm. Res.* 2004, 21, 1531-8.
- [86] A. Shah, D.J. Rader, J.S. Millar. The effect of PPAR- α agonism on apolipoprotein metabolism in humans. *Atherosclerosis.* 2010, 210, 35-40.
- [87] A.N. Billin. PPAR- β/δ agonists for Type 2 diabetes and dyslipidemia: an adopted orphan still looking for a home. *Expert Opin. Invest. Drugs.* 2008, 17, 1465-71.
- [88] P. Tontonoz, E. Hu, B.M. Spiegelman. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell.* 1994, 79, 1147-56.
- [89] B.M. Spiegelman. PPAR-gamma: adipogenic regulator and thiazolidinedione receptor. *Diabetes.* 1998, 47, 507-14.
- [90] H. Pingali, M. Jain, S. Shah, P. Patil, P. Makadia, P. Zaware, et al. Modulation of PPAR receptor subtype selectivity of the ligands: Aliphatic chain vs aromatic ring as a spacer between pharmacophore and the lipophilic moiety. *Bioorg. Med. Chem. Lett.* 2008, 18, 6471-5.
- [91] P. Zaware, S.R. Shah, H. Pingali, P. Makadia, B. Thube, S. Pola, et al. Modulation of PPAR subtype selectivity. Part 2: Transforming PPAR α/γ dual agonist into α selective PPAR agonist through bioisosteric modification. *Bioorg. Med. Chem. Lett.* 2011, 21, 628-32.
- [92] B.R. Henke. Peroxisome proliferator-activated receptor α/γ dual agonists for the treatment of type 2 diabetes. *J. Med. Chem.* 2004, 47, 4118-27.
- [93] A. Rubenstrunk, R. Hanf, D.W. Hum, J.-C. Fruchart, B. Staels. Safety issues and prospects for future generations of PPAR modulators. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids.* 2007, 1771, 1065-81.
- [94] X. Chen, Y. Lin, M. Liu, M.K. Gilson. The Binding Database: data management and interface design. *Bioinformatics.* 2002, 18, 130-9.
- [95] T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007, 35, D198-D201.
- [96] M. Sud. MayaChemTools. URL: <http://www.MayaChemTools.org>.
- [97] D. Filimonov, V. Poroikov, Y. Borodina, T. Glorizova. Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. *J. Chem. Inf. Comput. Sci.* 1999, 39, 666-70.
- [98] L.H. Hall, L.B. Kier. Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* 1995, 35, 1039-45.
- [99] D. Rogers, M. Hahn. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 2010, 50, 742-54.
- [100] J.L. Durant, B.A. Leland, D.R. Henry, J.G. Nourse. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* 2002, 42, 1273-80.

- [101] R.E. Carhart, D.H. Smith, R. Venkataraghavan. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inform. Comput. Sci.* 1985, 25, 64-73.
- [102] R. Nilakantan, N. Bauman, J.S. Dixon, R. Venkataraghavan. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* 1987, 27, 82-5.
- [103] S. Renner, U. Fechner, G. Schneider. Alignment-free pharmacophore patterns – A correlation-vector approach. In: *Pharmacophores and Pharmacophore Searches*, Wiley-VCH Verlag GmbH & Co. KGaA, 2006, pp. 49-79.
- [104] F. Bonachéra, B. Parent, F. Barbosa, N. Froloff, D. Horvath. Fuzzy tricentric pharmacophore fingerprints. 1. Topological fuzzy pharmacophore triplets and adapted molecular similarity scoring schemes. *J. Chem. Inf. Model.* 2006, 46, 2457-77.
- [105] P. Jaccard. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaudoise Sci. Nat.* 1901, 37, 241-72.
- [106] A. Bender, J.L. Jenkins, J. Scheiber, S.C.K. Sukuru, M. Glick, J.W. Davies. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J. Chem. Inf. Model.* 2009, 49, 108-19.
- [107] S. Sundriyal, P.V. Bharatam. Important pharmacophoric features of pan PPAR agonists: Common chemical feature analysis and virtual screening. *Eur. J. Med. Chem.* 2009, 44, 3488-95.

APÉNDICES

Apéndice A

Representación en SMILES de las estructuras y valores de CI_{50} y pCI_{50} de 32 purino-carbonitrilos inhibidores de la cruzafina.

Compuesto	Molécula	CI_{50} (μ M)	pCI_{50}
1	<chem>n1c(nc(nc1Nc1cccc1)NCC)C=N</chem>	0.710	6.1487
2	<chem>n1c(nc(nc1NC1CCCC1)Nc1cccc1)C=N</chem>	0.710	6.1487
3	<chem>n1c(nc(nc1Nc1cccc1)NCCC)C=N</chem>	0.790	6.1024
4	<chem>n1c(nc(nc1Nc1cccc1)NC(C)C)C=N</chem>	0.890	6.0506
5	<chem>n1c(nc(nc1Nc1cccc1)NCC1CC1)C=N</chem>	0.890	6.0506
6	<chem>n1c(nc(nc1Nc1cccc1)NCC(C)C)C=N</chem>	0.890	6.0506
7	<chem>n1c(nc(nc1Nc1cccc1)NCCCC)C=N</chem>	0.890	6.0506
8	<chem>OCCCNc1nc(nc(n1)Nc1cccc1)C=N</chem>	0.890	6.0506
9	<chem>n1c(nc(nc1Nc1cccc1)NC)C=N</chem>	1.260	5.8996
10	<chem>n1c(nc(nc1NC1CCCC1)Nc1cccc1)C=N</chem>	1.410	5.8508
11	<chem>OCCNc1nc(nc(n1)Nc1cccc1)C=N</chem>	1.580	5.8013
12	<chem>O=[N+](([O-])c1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)ccc1</chem>	0.063	7.2006
13	<chem>Fc1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)ccc1</chem>	0.079	7.1024
14	<chem>Clc1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)ccc1</chem>	0.079	7.1024
15	<chem>Brc1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)ccc1</chem>	0.100	7.0000
16	<chem>Fc1ccc(Nc2nc(nc(n2)NC2CCCC2)C=N)cc1</chem>	0.126	6.8996
17	<chem>n1c(nc(nc1NC1CCCC1)Nc1cc(ccc1)C)C=N</chem>	0.126	6.8996
18	<chem>Brc1ccc(Nc2nc(nc(n2)NC2CCCC2)C=N)cc1</chem>	1.122	5.9500
19	<chem>n1c(nc(nc1NC1CCCC1)Nc1cc(ccc1)-c1cccc1)C=N</chem>	2.239	5.6499
20	<chem>Fc1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)cc(F)c1</chem>	0.063	7.2007
21	<chem>Clc1cc(Nc2nc(nc(n2)NC2CCCC2)C=N)cc(Cl)c1</chem>	0.398	6.4001
22	<chem>Fc1cc(Nc2nc(nc(n2)NCC(F)F)C=N)cc(F)c1</chem>	0.025	7.6021
23	<chem>Fc1cc(Nc2nc(nc3n(cnc23)CC)C=N)cc(F)c1</chem>	0.010	8.0000
24	<chem>Fc1cc(Nc2nc(nc3n(cnc23)C2CCCC2)C=N)cc(F)c1</chem>	0.013	7.8860
25	<chem>Fc1cc(Nc2nc(nc3n(cnc23)C2CCCC2)C=N)cc(F)c1</chem>	0.018	7.7447
26	<chem>Fc1cc(Nc2nc(nc3n(cnc23)CC)C=N)ccc1</chem>	0.018	7.7447
27	<chem>Fc1cc(Nc2nc(nc3n(cnc23)C2CCCC2)C=N)ccc1</chem>	0.040	7.3979
28	<chem>Fc1cc(-n2c3nc(nc(NCC(F)F)c3nc2)C=N)cc(F)c1</chem>	0.050	7.3010
29	<chem>Fc1cc(-n2c3nc(nc(NCC)c3nc2)C=N)ccc1</chem>	0.063	7.2007
30	<chem>Fc1cc(-n2c3nc(nc(NCC(F)F)c3nc2)C=N)ccc1</chem>	0.071	7.1487
31	<chem>Fc1cc(-n2c3nc(nc(NCC)c3nc2)C=N)cc(F)c1</chem>	0.251	6.6003
32	<chem>Fc1cc(-n2c3nc(nc(NC4CCCC4)c3nc2)C=N)cc(F)c1</chem>	0.316	6.5003

Apéndice B

Representación en SMILES de las estructuras de 168 agonistas de los PPARs y sus valores de pCl₅₀ hacia cada subtipo de PPAR (α , δ y γ).

Comp	Número en BD	Molécula	pCl ₅₀ α	pCl ₅₀ δ	pCl ₅₀ γ
1	50194627	<chem>O(CC[C@H](Oc1ccc(cc1C(=O)c1ccccc1)CC)C)c1cc(C)c(cc1)CCC(O)=O</chem>	5.371	8.523	8.301
2	50194628	<chem>FC(F)(F)c1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(C)c(cc2)CCC(O)=O)C)cc1</chem>	5.984	8.301	8.222
3	50194625	<chem>FC(F)(F)Oc1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(C)c(cc2)CCC(O)=O)C)cc1</chem>	6.047	8.301	8.222
4	50194626	<chem>O(CCCOc1ccc(OC(C(O)=O)(C)C)cc1)c1ccc(cc1C(=O)c1ccccc1)CC</chem>	5.854	8.222	7.420
5	50194629	<chem>O(C(C(O)=O)(C)C)c1ccc(OCC[C@H](Oc2ccc(cc2C(=O)c2ccccc2)C)C)cc1</chem>	5.878	8.301	8.222
6	50194624	<chem>O(CC[C@H](Oc1ccc(cc1Cc1ccccc1)CC)C)c1cc(C)c(cc1)CCC(O)=O</chem>	4.963	8.097	7.229
7	50227737	<chem>Fc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1C)C(F)(F)F</chem>	6.480	8.222	7.770
8	50227736	<chem>Fc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)c(cc1)C)C(F)(F)F</chem>	6.726	8.523	8.301
9	50194637	<chem>Clc1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(F)c(cc2)CCC(O)=O)C)cc1</chem>	5.572	8.398	7.201
10	50194638	<chem>O(CC[C@H](Oc1ccc(cc1C(=O)c1ccccc1)CC)CC)c1cc(C)c(cc1)CCC(O)=O</chem>	5.050	8.222	7.745
11	50135773	<chem>O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(Cc2cc(C)c(cc2)C)C(=O)N1C</chem>	7.921	6.131	5.770
12	50135774	<chem>O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(CC2CCCC2)C(=O)N1CCC</chem>	6.602	6.056	5.770
13	50227738	<chem>Fc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1F)C(F)(F)F</chem>	6.416	8.523	8.097
14	50227739	<chem>FC(F)(F)Oc1cc(C)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1</chem>	5.314	7.268	7.328
15	50135775	<chem>O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(Cc2ccc(cc2)C)C(=O)N1CCC</chem>	7.796	5.921	5.310
16	50135777	<chem>O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(Cc2ccc(cc2)C(C)(C)C)C(=O)N1C</chem>	7.824	5.770	6.284
17	50194630	<chem>Clc1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(CC)c(cc2)CCC(O)=O)C)cc1</chem>	5.159	8.398	7.721
18	50194631	<chem>O(c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC)c1ccccc1</chem>	5.425	8.398	8.222
19	50194633	<chem>Clc1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(C(F)(F)F)c(cc2)CCC(O)=O)C)cc1</chem>	5.358	8.301	7.215
20	50194634	<chem>O(CC[C@H](Oc1ccc(cc1C(=O)c1ccccc1)CC)C)c1cc(C)c(cc1)CCC(O)=O</chem>	5.083	8.155	7.921
21	50194635	<chem>Clc1cc(Oc2ccccc2)c(O[C@@H](CCOc2cc(C)c(cc2)CCC(O)=O)C)cc1</chem>	5.275	8.398	7.824
22	50227740	<chem>O(c1cc(C)c(OC(C(O)=O)(C)C)cc1)c1cc(ccc1)CNC(=O)c1ccc(OC)cc1C</chem>	5.377	7.921	8.097
23	50227746	<chem>Clc1cc2cc([nH]c2cc1)C(=O)NCc1cc(Oc2cc(C)c(cc2)CCC(O)=O)ccc1</chem>	4.999	8.523	7.886
24	50227745	<chem>Brc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1)C(F)(F)F</chem>	5.798	8.523	7.921
25	50227748	<chem>Brc1cc2cc([nH]c2cc1)C(=O)NCc1cc(Oc2cc(C)c(cc2)CCC(O)=O)ccc1</chem>	5.127	8.398	8.097
26	50227747	<chem>FC(F)(F)c1ccc(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1</chem>	5.458	8.222	7.745

APÉNDICE B

27	50227741	FC(F)(F)c1cc(CC)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1	5.948	8.398	8.000
28	50227743	FC(F)(F)c1cc(C)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1	5.754	8.523	8.046
29	50212695	S(Cc1ccc(OCc2ccc(cc2)C(F)(F)F)cc1)c1cc(C)c(OCC(O)=O)c2c1CC C2	6.507	6.232	4.666
30	50212693	S(Cc1ccc(OCc2ccc(cc2)C(F)(F)F)cc1)c1cc(C)c(OCC(O)=O)cc1OC	8.328	8.770	4.627
31	50212697	S(Cc1ccc(OCc2ccc(OC)cc2)cc1)c1c2c(CCC2)c(OCC(O)=O)cc1	5.305	6.921	4.860
32	50145722	o1c(nc(CCOc2ccc(cc2)C[C@](Oc2ccccc2)(C(O)=O)C)c1C)-c1ccccc1	7.678	5.258	7.854
33	50145720	o1c(nc(CCOc2ccc(cc2)C[C@@](Oc2ccccc2)(CCCC)C(O)=O)c1C)-c1ccccc1	5.734	5.141	5.650
34	50145721	s1cccc1-c1oc(C)c(n1)CCOc1ccc(cc1)C[C@](Oc1ccccc1)(C(O)=O)C	7.553	5.280	8.000
35	50145719	o1c(nc(CCOc2ccc(cc2)C[C@](Oc2ccccc2)(C(O)=O)C)c1C)C1CCC CC1	7.745	5.014	7.824
36	50145713	o1c(nc(CCOc2ccc(cc2)C[C@H](Oc2ccccc2)C(O)=O)c1C)-c1ccccc1	5.956	5.152	6.893
37	50145712	o1c(nc(CCOc2ccc(cc2)C[C@](Oc2ccccc2)(C(O)=O)C)c1C)-c1ccccc1	7.377	5.148	7.745
38	50145711	o1c(nc(CCOc2ccc(O[C@](Cc3ccccc3)(C(O)=O)C)cc2)c1C)-c1ccccc1	6.167	5.352	6.309
39	50145717	s1cccc1-c1oc(C)c(n1)CCOc1ccc(cc1)C[C@@](Oc1ccccc1)(C(O)=O)C	6.209	5.171	6.527
40	50145716	o1c(nc(CCOc2ccc(cc2)C[C@@](Oc2ccccc2)(CC)C(O)=O)c1C)-c1ccccc1	6.325	5.295	6.538
41	50145715	o1c(nc(CCOc2ccc(cc2)C[C@@](Oc2ccccc2)(C(O)=O)C)c1C)-c1ccccc1	5.811	5.571	6.411
42	50135780	O(C(C(O)=O)(C)C)c1ccc(cc1)CCCc1nc(O)n(n1)Cc1ccc(cc1)C	7.620	5.420	5.187
43	50145710	s1cccc1-c1oc(C)c(n1)CCOc1ccc(cc1)C[C@](Oc1ccccc1)(C(O)=O)C	7.347	5.289	7.620
44	50135782	O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(Cc2cc(OC)ccc2)C(=O)N1CCC	7.721	5.699	5.208
45	50135781	O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN([C@H](C)c2ccccc2)C(=O)N1CCC	6.658	5.824	5.215
46	50135779	O(C(C(O)=O)(C)C)c1ccc(cc1)CCCC1=NN(Cc2ccc(cc2)C)C(=O)N1CCO	7.569	5.387	5.215
47	50127928	S1[C@@H](c2ccc(OCCCOc3ccc(Oc4ccccc4)cc3CC(C)C)cc2)C(=O)NC1=O	6.469	6.796	6.319
48	50181912	O(CCCOc1cc2n(ccc2cc1)CC(O)=O)c1cc2c(cc(cc2)C(=O)c2ccccc2)cc1	5.348	5.590	5.801
49	50181911	O(CCCOc1cc2c(n(cc2)CC(O)=O)cc1)c1ccc2c(ccc(c2)C(=O)c2ccccc2)c1CCC	6.947	6.652	7.301
50	50181910	Fc1ccc(cc1)C(=O)c1cc(CCC)c(OCCCOc2cc3n(ccc3cc2)CC(O)=O)c c1	5.980	5.118	5.890
51	50168545	Clc1cc(Oc2ccc(F)cc2)ccc1OCCCOc1cc2O[C@](CCc2cc1)(C(O)=O)C	6.854	5.301	6.046
52	50131514	S1[C@@H](c2ccc(OCCCOc3ccc(Oc4ccc(O)cc4)cc3CCC)ccc2)C(=O)NC1=O	7.022	5.420	7.523
53	50157060	O(c1ccc(OCCC#Cc2ccc(cc2)C[C@H](OC)C(O)=O)cc1)c1ccccc1	5.938	5.405	7.886
54	50157061	O(CCCOc1ccc(cc1)C[C@H](OC)C(O)=O)c1ccc(cc1)-c1ccccc1	6.246	5.857	7.678
55	50168560	Clc1cc(ccc1OCCCOc1cc2O[C@](CCc2cc1)(C(O)=O)C)C1CCCC1	7.398	4.979	5.420
56	50168565	Clc1cc(ccc1OCCCOc1cc2O[C@@](CCc2cc1)(CC)C(O)=O)C(F)(F)F	7.046	5.638	5.367
57	50168564	Clc1cc(OCC(F)(F)F)ccc1OCCCOc1cc2O[C@@H](CCc2cc1)C(O)=O	6.092	5.161	5.387
58	50157059	O(CCCOc1ccc(cc1)C[C@H](OC)C(O)=O)c1ccc(cc1)C(=O)c1ccccc1	6.328	5.634	8.301
59	50157055	Clc1cc(ccc1OCCCOc1ccc(cc1)C(=O)c1ccccc1)C[C@H](OC)C(O)=O	6.510	5.822	8.301

APÉNDICE B

60	50157056	O(c1ccc(OCCCOc2ccc(cc2)C[C@H](OC)C(O)=O)cc1)c1ccccc1	5.567	5.718	7.469
61	50157057	O(c1ccc(OCCCC#Cc2ccc(cc2)C[C@H](OC)C(O)=O)cc1)c1ccccc1	6.114	5.385	8.155
62	50168559	Clc1cc(Cl)ccc1OCCCOc1cc2O[C@@](CCc2cc1)(CC)C(O)=O	6.398	5.387	5.538
63	50157058	O(c1ccc(OCCCOc2ccc(cc2OC)C[C@H](OC)C(O)=O)cc1)c1ccccc1	5.870	5.125	7.886
64	50212706	S(Cc1ccc(OCc2ccc(cc2)C(F)(F)F)cc1)c1c2OCCc2c(OCC(O)=O)c1	6.102	6.866	5.255
65	50212705	S(Cc1ccc(OCc2ccc(cc2)C(F)(F)F)cc1)c1cc(C)c(OCC(O)=O)c2c1OCC2	6.369	6.504	4.478
66	50188032	FC(F)(F)c1ccc(cc1)-c1oc(C)c(n1)Cn1c2c(cc(cc2)C[C@@H](OCCC)C(O)=O)cc1	7.398	6.587	6.399
67	50188033	FC(F)(F)c1ccc(cc1)-c1oc(C)c(n1)Cn1c2c(cc(cc2)C[C@@H](OCC)C(O)=O)cc1C	6.453	6.429	5.097
68	50157054	O(c1ccc(OCCCOc2ccc(cc2)C[C@H](OCC)C(O)=O)cc1)c1ccccc1	5.773	5.804	7.357
69	50212701	Clc1cc(OCC(O)=O)c(cc1SCc1ccc(OCc2ccc(cc2)C(F)(F)F)cc1)C	6.893	7.699	5.040
70	50188039	FC(F)(F)c1ccc(cc1)-c1oc(C)c(n1)Cn1c2c(cc(cc2)C[C@@H](OC(C)C)C(O)=O)cc1	7.215	6.050	6.857
71	50181903	Fc1ccc(cc1)C(=O)c1cc(CCC)c(OCCCOc2cc3c(n(cc3)CC(O)=O)cc2)cc1	6.328	6.188	6.889
72	50181902	Fc1ccc(cc1)C(=O)c1ccc(OCCCOc2cc3n(ccc3cc2)CC(O)=O)cc1	6.936	5.501	5.684
73	50168554	Clc1cc(ccc1OCCCOc1cc2O[C@](CCc2cc1)(C(O)=O)C)C(C)C	8.155	4.854	4.824
74	50181905	O(CCCOc1c2c(n(cc2)CC(O)=O)ccc1)c1ccc2c(ccc(c2)C(=O)c2ccccc2)c1CCC	6.238	5.269	6.084
75	50168557	Clc1cc(ccc1OCCCOc1cc2O[C@](CCc2cc1)(C(O)=O)C)CC(C)C	7.796	5.119	5.523
76	50181904	O(CCCOc1cc2c(n(cc2)CC(O)=O)cc1)c1cc2c(cc(cc2)C(=O)c2ccccc2)cc1	6.315	5.354	6.201
77	50188040	o1c(nc(Cn2c3c(cc2)c(ccc3)C[C@H](OCC)C(O)=O)c1C)-c1ccccc1	6.903	5.456	5.824
78	50181907	Fc1ccc(cc1)C(=O)c1cc(CCC)c(OCCCOc2c3c(n(cc3)CC(O)=O)ccc2)cc1	6.383	5.487	5.618
79	50131502	S1[C@@H](c2cc(OCCCOc3ccc(Oc4ccccc4)cc3)ccc2)C(=O)NC1=O	7.328	6.721	7.119
80	50181909	O(CCCOc1cc2n(ccc2cc1)CC(O)=O)c1ccc2c(ccc(c2)C(=O)c2ccccc2)c1CCC	5.348	5.620	5.694
81	50213329	s1c(nc(C)c1COc1cc2sc(c2cc1C)CC(O)=O)-c1ccc(cc1)C(F)(F)F	6.166	6.625	5.203
82	50213330	s1cc(c2c1cc(OCc1ccc(cc1)COc1ccc(cc1)C(F)(F)F)cc2)CC(O)=O	6.640	8.444	5.428
83	50168572	Clc1cc(OC(F)(F)F)ccc1OCCCOc1cc2O[C@@](CCc2cc1)(CC)C(O)=O	7.432	5.721	5.585
84	50213309	Clc1ccc(cc1)-c1onc(c1)COc1cc(c2c(scc2CC(O)=O)c1)C	6.507	7.638	5.062
85	50213310	s1cc(c2cc(C)c(OCc3ccccc3COc3ccc(cc3)C(F)(F)F)cc12)CC(O)=O	5.330	7.367	5.120
86	50213311	s1cc(c2c1cc(OCc1cc(ccc1)COc1ccc(cc1)C(F)(F)F)cc2)CC(O)=O	6.016	7.824	5.237
87	50213317	FC(F)(F)c1ccc(OCc2cc(ccc2)COc2cc3occ(c3cc2)CC(O)=O)cc1	5.602	7.618	5.268
88	28661	s1c(nc(C)c1CSc1cc(C)c(OCC(O)=O)cc1)-c1ccc(cc1)C(F)(F)F	5.700	8.301	5.200
89	50085040	Clc1cc(ccc1SCCCOc1ccc2c(occ2-c2ccccc2)c1CCC)CC(O)=O	7.268	8.222	7.027
90	50184274	FC(F)(F)c1noc2c1ccc(OCCCNc1ccc(cc1)CC(O)=O)c2CCC	7.796	7.056	6.991
91	50184256	FC(F)(F)c1noc2c1ccc(OCCCN(C)c1cc(C)c(cc1)CC(O)=O)c2CCC	6.060	6.824	5.337
92	50184255	FC(F)(F)c1noc2c1ccc(OCCCN1c3c(cc(cc3)CC(O)=O)cc1)c2CCC	6.398	7.244	5.721
93	50184258	FC(F)(F)c1noc2c1ccc(OCCCN(C)c1ccc(cc1C)CC(O)=O)c2CCC	6.222	7.409	6.337
94	50184263	Clc1cc(ccc1OCCCOc1ccc2c(onc2C(F)(F)F)c1CCC)CC(O)=O	6.638	8.398	6.553
95	50184261	FC(F)(F)c1noc2c1ccc(OCCCOc1ccc(cc1)CC(O)=O)c2CCC	7.721	8.097	6.569
96	50184249	FC(F)(F)c1noc2c1ccc(OCCCN1c3c(cc(cc3)CCC(O)=O)cc1)c2CCC	6.347	7.155	5.921

APÉNDICE B

97	50184253	FC(F)(F)c1noc2c1ccc(OCCCN(C)c1cc(ccc1)CC(O)=O)c2CCC	5.824	7.032	5.523
98	50184254	FC(F)(F)c1noc2c1ccc(OCCCN(C)c1ccc(cc1)CC(O)=O)c2CCC	6.509	6.509	5.538
99	50199778	s1cc(nc1)-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	5.387	8.523	6.461
100	50199779	O(CC[C@H](Oc1ccc(cc1-c1ncccc1)CC)C)c1cc(C)c(cc1)CCC(O)=O	5.081	8.523	7.337
101	50199776	O(CC[C@H](Oc1ccc(cc1-c1ncccc1)CC)C)c1cc(CC)c(cc1)CCC(O)=O	5.134	8.523	7.456
102	50199777	FC(F)(F)c1cc(-c2ncccc2)c(O[C@@H](CCOc2cc(C)c(cc2)CCC(O)=O)C)cc1	5.231	8.398	7.699
103	50199782	o1ccnc1-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	4.983	8.301	6.316
104	50199781	FC(F)(F)c1cc(-c2ncccc2)c(O[C@@H](CCOc2cc(CC)c(cc2)CCC(O)=O)C)cc1	5.215	8.523	7.796
105	50199780	o1cccc1-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	5.347	7.854	6.027
106	50199784	o1cc(nc1)-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	5.302	7.886	6.009
107	50133652	O1[C@@H](c2cc(OCCCOc3ccc(cc3CCC)-c3cccc3)ccc2)C(=O)NC1=O	5.578	5.821	6.347
108	50179835	Clc1cc(Oc2ccc(F)cc2)ccc1OCCCOc1ccc(OC(C(O)=O)(C)C)cc1	7.523	5.622	6.008
109	50126018	Clc1cc(ccc1SCCCOc1ccc2c(onc2C(F)(F)F)c1CCC)CC(O)=O	6.553	8.398	6.710
110	50126017	Clc1cc(ccc1SCCCOc1ccc2c(occ2CC)c1CCC)CC(O)=O	5.762	7.398	6.252
111	50199775	s1cccc1-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	5.040	8.046	5.963
112	50199772	s1cc(cc1)-c1cc(ccc1O[C@@H](CCOc1cc(C)c(cc1)CCC(O)=O)C)CC	5.225	8.097	6.246
113	50126019	Clc1cc(ccc1SCCCOc1ccc(/C(=N/O)/CC)c(O)c1CCC)CC(O)=O	7.000	8.000	7.319
114	50126016	Clc1cc(ccc1SCCCOc1ccc2c(onc2CC)c1CCC)CC(O)=O	6.244	8.398	6.959
115	50179236	FC(F)(F)c1noc2c1ccc(OCCCOc1cc3c(n(cc3)CC(O)=O)cc1)c2CCC	6.305	7.018	6.818
116	50133667	S1[C@@H](c2cc(OCCCOc3ccc(OC4CC4)cc3CCC)ccc2)C(=O)NC1=O	7.377	6.036	7.201
117	50133661	S1[C@@H](c2cc(OCCCOc3ccc(cc3CCC)-c3noc4c3cccc4)ccc2)C(=O)NC1=O	7.432	6.854	7.523
118	50133663	S1[C@@H](c2cc(OCCCOc3ccc(cc3CCC)C(=O)c3cccc3)ccc2)C(=O)NC1=O	7.046	6.770	7.721
119	50133662	O1[C@H](c2cc(OCCCOc3ccc(cc3CCC)C3CCCC3)ccc2)C(=O)NC1=O	6.201	5.415	5.975
120	50179821	Clc1cccc1OCCCOc1ccc(cc1)C[C@](C)(C(O)=O)C	7.377	5.334	6.070
121	50126021	Clc1cc(ccc1SCCCOc1ccc(C(=O)C)c(O)c1CCC)CC(O)=O	5.131	7.699	6.237
122	50126020	Clc1cc(ccc1SCCCOc1ccc(C(=O)CC)c(O)c1CCC)CC(O)=O	6.222	7.699	6.523
123	50264403	S(CC(=O)Nc1ccc(cc1)C(O)=O)c1nnc(n1CC)-c1oc2c(c1)cccc2OC	6.000	3.409	4.398
124	50264404	O1CCN(CC1)c1ccc(\N=C\c2c3c(ccc2OCc2ccc(cc2)C(O)=O)cccc3)c1	4.367	4.469	4.699
125	50227765	Clc1cc2c([nH]c(C(=O)NCc3cc(Oc4cc(C)c(cc4)CCC(O)=O)ccc3)c2C)cc1	5.677	8.523	7.569
126	50227766	O(c1cc(C)c(OC(C(O)=O)(C)C)cc1)c1cc(ccc1)CNC(=O)c1ccc(cc1)C	5.757	8.398	8.155
127	50227763	Clc1cc2c(n(C)c(C(=O)N[C@H](C)c3cc(Oc4cc(CC)c(cc4)CCC(O)=O)cc(F)c3)c2C)cc1	4.989	8.301	7.409
128	50227764	Fc1cc(ccc1C(=O)NCc1cc(cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)c1)C)C(F)(F)F	6.504	8.301	7.921
129	50227761	Fc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)cc(F)c1)C(F)(F)F	6.138	8.523	8.097
130	50227762	Clc1cc(C)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1	5.145	8.222	8.046
131	50227760	FC(F)(F)c1cc(Oc2cccc2)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)	5.829	7.495	6.943

APÉNDICE B

		=O)(C)C)cc2)ccc1			
132	50227757	Fc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1)C(F)(F)F	6.194	8.398	7.921
133	50227759	Clc1cc2[nH]c(cc2cc1)C(=O)NCc1cc(Oc2cc(C)c(cc2)CCC(O)=O)ccc1	5.946	8.523	6.633
134	50227752	Clc1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1)C(F)(F)F	5.910	8.523	7.824
135	50227753	FC(F)(F)c1cc(ccc1C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1)C(F)(F)F	4.987	7.398	6.693
136	50227754	Clc1cc2c(sc(C(=O)N[C@H](C)c3cc(Oc4cc(CC)c(cc4)CCC(O)=O)cc(F)c3)c2C)cc1	5.030	8.301	7.301
137	50227750	Brc1c2[nH]c(cc2ccc1)C(=O)NCc1cc(Oc2cc(C)c(cc2)CCC(O)=O)ccc1	5.313	8.046	7.432
138	50227751	Clc1cc2cc([nH]c2cc1Cl)C(=O)NCc1cc(Oc2cc(C)c(cc2)CCC(O)=O)cc1	5.361	8.301	7.022
139	50227749	FC(F)(F)c1cc(OC)c(cc1)C(=O)NCc1cc(Oc2cc(C)c(OC(C(O)=O)(C)C)cc2)ccc1	6.334	8.398	8.155
140	28665	S(=O)(=O)(N1CCc2c(C1)cccc2)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.000	7.886	5.300
141	28682	S(=O)(=O)(n1c2c(cc1)c(ccc2)C)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	6.000	8.222	5.700
142	28683	Brc1c2c(n(S(=O)(=O)c3cc(ccc3)C(=O)Nc3ccc(cc3C(O)=O)C)cc2)cc1	5.900	8.000	5.900
143	28686	Clc1cc2c(n(S(=O)(=O)c3cc(ccc3)C(=O)Nc3ccc(cc3C(O)=O)C)cc2)c1	5.700	7.699	5.800
144	28687	S(=O)(=O)(n1c2c(cc(OC)cc2)cc1)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.100	6.801	5.000
145	28684	S(=O)(=O)(n1c2c(cc1)c(OC)ccc2)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.900	8.000	5.800
146	28685	S(=O)(=O)(n1c2c(cc(cc2)C)cc1)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	6.000	8.000	5.600
147	28689	Clc1cc2n(S(=O)(=O)c3cc(ccc3)C(=O)Nc3ccc(cc3C(O)=O)C)ccc2cc1	5.500	7.201	5.700
148	28688	S(=O)(=O)(n1c2cc(ccc2cc1)C)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.600	8.000	5.600
149	28668	S(=O)(=O)(N1CCc2c(C1)cccc2)c1cc(ccc1)C(=O)Nc1ccc(F)cc1C(O)=O	5.200	7.602	5.200
150	28669	S(=O)(=O)(N1CCc2c(C1)cccc2)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C(F)(F)F	5.800	8.301	5.500
151	28666	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4)ccc2)cc1	5.300	8.097	5.400
152	28667	Brc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4)ccc2)cc1	5.400	8.097	5.600
153	28690	S(=O)(=O)(n1c2cc(OC)ccc2cc1)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.400	7.495	5.100
154	28695	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)n3c4c(cc3)c(OC)ccc4)ccc2)cc1	6.100	8.222	5.700
155	28670	Brc1cc(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4)ccc2)c(cc1)C(O)=O	5.300	8.301	5.400
156	28696	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)n3c4c(cc(OC)cc4)cc3)ccc2)cc1	5.500	7.886	5.400
157	28697	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)n3c4cc(OC)ccc4cc3)ccc2)cc1	5.800	7.886	5.600
158	28672	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4OC)cc2)cc1	5.100	7.886	5.300
159	28698	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)n3c4c(cc3)c(C)c(OC)cc4)ccc2)cc1	5.500	7.796	5.500
160	28691	S(=O)(=O)(n1c2c(cccc2C)cc1)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.800	8.398	6.000
161	28674	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4C(O)=	5.400	8.398	5.100

APÉNDICE B

		O)ccc2)cc1			
162	28692	Clc1c2n(S(=O)(=O)c3cc(ccc3)C(=O)Nc3ccc(cc3C(O)=O)C)ccc2ccc1	5.500	7.796	6.100
163	28675	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(C3)cccc4OCC(O)=O)ccc2)cc1	5.300	8.000	5.100
164	28694	S(=O)(=O)(n1c2c(cc1)c(C)c(OC)cc2)c1cc(ccc1)C(=O)Nc1ccc(cc1C(O)=O)C	5.500	7.699	5.500
165	28676	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4cc(OC)ccc4C3)cc2)cc1	5.000	8.000	5.200
166	28678	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(cc(Cl)cc4)C3)cc2)cc1	5.100	7.602	5.400
167	28677	Clc1cc(C(O)=O)c(NC(=O)c2cc(S(=O)(=O)N3CCc4c(cc(F)cc4)C3)cc2)cc1	5.200	7.796	5.500
168	28679	Clc1ccc2c(CN(S(=O)(=O)c3cc(ccc3)C(=O)Nc3ccc(Cl)cc3C(O)=O)C2)c1C	5.000	7.796	5.300

Descriptor	Tipo de Valores	Parámetros/Descripción
Atom Neighborhoods	Vector	Valores: Vector alfanumérico; Radio mínimo: 0; Radio máximo: 2; Tipo de átomo: Invariantes atómicas (AS, X, BO, H,F C)
Atom Types	Bit-vector o vector	Valores: Vector numérico; Tipo de átomo: Invariantes atómicas (AS, X, BO, H, FC)
E-state Indices	Vector	Valores: Vector numérico; Tamaño del conjunto de tipos de átomos: Arbitrario
Extended Connectivity	Bit-vector o vector	Valores: Vector alfanumérico; Radio: 2; Tipo de átomo: Invariantes atómicas (AS, X, BO, H, FC, MN)
MACCS	Bit-vector o vector	Valores: Bit-vector; Tamaño: 166; Tamaños disponibles: 166 and 322; Conteo de características disponibles
Path Lengths	Bit-vector o vector	Valores: Bit-vector; Tamaño: 1024; Tipo de átomo: Invariantes atómicas (AS); Longitud mínima: 1; Longitud máxima: 8; Trayectorias disponibles
TopAtomPairs	Vector	Valores: Vector numérico; Tipo de átomo: Invariantes atómicas (AS, X, BO, H, FC); Distancia mínima: 1; Distancia máxima: 10
TopAtomTriplets	Vector	Valores: Vector numérico; Tipo de átomo: Invariantes atómicas (AS,X,BO,H,FC); Distancia mínima: 1; Distancia máxima: 10; Desigualdad del triángulo: No
TopAtomTorsions	Vector	Valores: Vector numérico; Tipo de Átomo: Invariantes Atómicas (AS, X, BO, H, FC)
TopPh4Pairs	Vector	Valores: Vector numérico; Tipo de átomos: HBD, HBA, PI, NI, H; Distancia mínima: 1; Distancia máxima: 10; Peso del tipo de átomo: Ninguno; Normalización: Ninguna; Contar pares de átomos difusos: No
TopPh4Triplets	Vector	Valores: Vector numérico; Tipo de átomos: HBD, HBA, PI, NI, H, Ar; Distancia mínima: 1; Distancia máxima: 10; Tamaño de distancia binaria: 2; Desigualdad del triángulo: Si

Tipo de átomo: Invariantes atómicas, grupos funcionales, DREIDING, EState, MMFF94, SLogP, SYBYL, TPSA y UFF

Invariantes atómicas: AS (símbolo atómico), X (número de átomos vecinos), BO (suma de orden de enlace de los átomos), LBO (orden de enlace más grande), SB (número de enlaces sencillos), DB (número de dobles enlaces), TB (número de enlaces triples), H (número de hidrógenos implícitos y explícitos), Ar (aromático), RA (átomo en un anillo), FC (carga formal), MN (masa atómica), SM (multiplicidad de spin)

Grupos funcionales: HBD (donador de puente de hidrógeno), HBA (aceptor de puente de hidrógeno), IP (ionizable positivamente), NP (ionizable negativamente), Ar (aromático), Hal (halógeno), H (hidrofóbico), RA (átomo en un anillo), CA (átomo en una cadena)

ARTÍCULOS PUBLICADOS

Cite this: *Med. Chem. Commun.*, 2011, **2**, 1058

www.rsc.org/medchemcomm

CONCISE ARTICLE

3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA)Oscar Méndez-Lucio,^a Jaime Pérez-Villanueva,^b Antonio Romo-Mancillas^a and Rafael Castillo^{*a}

Received 23rd June 2011, Accepted 9th August 2011

DOI: 10.1039/c1md00163a

Cruzain has been identified as the major cysteine protease of *Trypanosoma cruzi*, the etiological cause of Chaga's disease. For this reason, many efforts have been undertaken to design new inhibitors against this enzyme. Recently, molecules having a purine or triazine nucleus have been reported as potent non-peptidic inhibitors of cruzain. In order to gain an insight into the structural requirements that can lead to the improvement of the activity of these molecules, in this paper we report the CoMFA and CoMSIA studies of a series of purine-carbonitriles as cruzain inhibitors. Quantum semi-empirical calculations of the inhibitors inside the active site of cruzain were used as an approach to obtain reliable conformations for molecular alignment. Two different molecular alignments were used, resulting in 3 CoMFA models and 31 CoMSIA models. These models correspond to all of the possible combinations among five fields: steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor. Highly predictive models were obtained. Based on the q^2 values, the best CoMFA model had an $r^2 = 0.98$ and a $q^2 = 0.73$, whereas the best CoMSIA model retrieved an $r^2 = 0.88$ and a $q^2 = 0.62$. All models were validated with a rigorous procedure using an external test set. Contour maps obtained from these models show a preference toward the purine ring and indicate that bulky groups with a negative potential at the 3- and 5-positions of the phenyl ring are important structural requirements for inhibitory activity against cruzain.

Introduction

American trypanosomiasis is an infection caused by *Trypanosoma cruzi*, a flagellated protozoan parasite which infects around 10–12 million people, kills more than 15 000 each year and has become a world health problem.¹ This parasite is transmitted to humans mainly by insects from the genera *Triatoma*, *Panstrongylus* and *Rhodnius*.² Once a human has been infected, there appears an acute phase characterized by mild symptoms. After this phase, an asymptomatic period starts and can last between 10 and 20 years. Finally, the patient presents a chronic phase which could lead to sudden death caused by the enlarging of hollow viscera and/or heart problems.³ Benznidazole and nifurtimox have been used to treat patients; nevertheless, serious side effects and the inefficacy of these drugs in chronic phase patients have stimulated the interest of several research groups to search for new therapeutic agents.^{3,4}

The life cycle of this parasite consists of four stages: the epimastigote and metacyclic trypomastigote, which take place in the

insect vector, and the amastigote and bloodstream trypomastigote, which appear only in the infected organism.³ An essential enzyme for the differentiation between these life cycle stages is cruzain, the major cysteine protease of *Trypanosoma cruzi*. Among other functions, cruzain plays an essential role in the penetration into the host cell, the defense against the immune system of the infected organism and in the parasite's nutrition.³ Based on the vital functions of cruzain, it has been proposed as a promising target in the design of new drugs against this parasite. Some examples of reversible and irreversible inhibitors are the previously reported molecules with a sulfone, a fluoro methyl ketone or a nitrile as a susceptible group to nucleophilic attack by the enzyme.^{5–8}

Recently, a set of 33 purine nitriles as reversible covalent inhibitors of cruzain have been synthesized and tested.⁸ Although some of these compounds show good activity, an optimization of these molecules can be carried out. For this purpose, computational methodologies are an important tool in the study of the three-dimensional quantitative structure–activity relationships (3D-QSAR). Particularly, Comparative Molecular Field Analysis (CoMFA)⁹ and Comparative Molecular Similarity Indices Analysis (CoMSIA)¹⁰ are two powerful methodologies widely used. These two methodologies allow us to study a set of molecules, their putative active conformation, and binding mode. A

^aFacultad de Química, Departamento de Farmacia, UNAM, México, DF 04510, Mexico. E-mail: rafaalc@servidor.unam.mx; Fax: +525 56 22 53 29; Tel: +525 56 22 52 87

^bDepartamento de Sistemas Biológicos, División de Ciencias Biológicas y de la Salud, UAM-X, México, DF 04960, Mexico

great advantage of CoMFA and CoMSIA is the contour maps that highlight the most important structural features that could be useful in the optimization or design of new active compounds. When doing this kind of study, it is important to consider that molecules in the dataset have the same binding mode, act *via* the same mechanism and have a common pharmacophore (not necessarily the same molecular skeleton).¹¹ In particular, molecules with two different scaffolds (purine and triazine) were analyzed during this study; nevertheless it is important to notice that the triazine is a substructure of the purine nucleus.

In order to obtain more information concerning the ligand–protein interactions that can lead to the improvement of purine nitrile cruzain inhibitors, CoMFA and CoMSIA studies were carried out, obtaining high quality models. As far as we know, this is the first time that the inhibition of cruzain by purine nitrile molecules has been studied by 3D-QSAR.

Results and discussion

CoMFA and CoMSIA models

In general, all molecules used in this study have a similar structure as seen in Table 1. A rational procedure to obtain a molecular alignment consists of using a molecular conformation similar to that of the crystallographic protein–ligand complex. Therefore, the first stage of this study was to build CoMFA models derived from two different molecular alignments (Fig. 1) in order to obtain the closer conformation to the one expected in the binding site.

Model 1 was developed using the alignment shown in Fig. 1A, in which the purine scaffold of all molecules is oriented in the same way as the crystallographic structure of compound **23** bound to cruzain. For this alignment, model 1 had a q^2 of 0.623 and an r^2 of 0.983, which are quite acceptable results for a 3D-QSAR model. The statistical results for model 1 are shown in Table 2.

An alternative alignment was chosen for model 2 (Fig. 1B). In this approach, the phenyl groups of all the molecules were oriented on the same side of the purine scaffolds. This new alignment is based on the fact that the phenyl group of compound **23** is inside a pocket of cruzain which is very similar to the S_2 pocket of papain, another cysteine protease. It is important to mention that previous studies on papain have found that the S_2 pocket has a predilection for hydrophobic residues, especially for phenylalanine.¹² For this CoMFA model, a similar value of r^2 (0.980) was obtained, but it had a slightly better q^2 of 0.736. These results indicate that both alignments are good enough to derive 3D-QSAR models. Nevertheless, model 2 has a better predictive power and the alignment used to build this model is based on experimental data. For these reasons, the alignment shown in Fig. 1B was used to generate the subsequent models.

Since the calculation of the CoMFA and CoMSIA descriptors takes place in a three-dimensional lattice, another variable that was evaluated in this study is the lattice grid spacing. An exhaustive analysis of several 3D-QSAR models indicates that the smallest grid spacing results in a higher q^2 .¹³ To prove this hypothesis in our database, model 3 was built considering the same alignment and parameters as in model 2, but with a grid

spacing of 1 Å, whereas the grid spacing of model 2 was 2 Å. Lower values for the statistical parameters ($q^2 = 0.692$ and $r^2 = 0.977$) were obtained from model 3. This unexpected result has been observed before as an indication of the poor effect of more grid points, as compared with a careful choice of parameters at 2 Å.¹⁴ Thus, model 2 was selected as the best CoMFA model and the predicted values of activity are plotted and listed in Fig. 2A and Table 3, respectively.

The CoMSIA studies were carried out using the 1B alignment and a grid spacing of 2 Å. Five fields were used: steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D) and hydrogen bond acceptor (A), and the 31 possible combinations among these were evaluated (Fig. 3). Interestingly, the higher values of q^2 were obtained for the combinations SED, SEA, and SEDA (0.627, 0.618 and 0.565, respectively) and not for the one that includes the five fields. A detail that stands out is that almost all the combinations with a $q^2 > 0.5$ include the steric and/or the electrostatic fields, the same kind of descriptors used in the CoMFA models. These results suggest that the structure–activity relationship for this database cannot be explained based on hydrophobicity; however, the similarities of the cruzain pocket with the S_2 pocket of papain suggest that it would be interesting to generate inhibitors with a higher molecular diversity (including hydrophobic modifications) at this position.

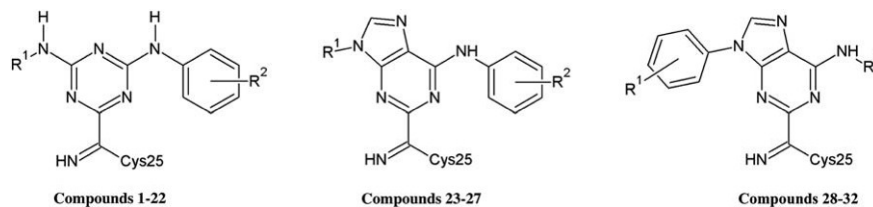
Validation of CoMFA and CoMSIA models

Although the q^2 value is used in several papers to define the predictive ability of a QSAR model, it has been observed that the high value ($q^2 > 0.5$) is a necessary condition, but insufficient to assure the predictive power and model robustness.¹⁵ For this reason, Golbraikh and Tropsha¹⁵ developed a methodology to validate QSAR models. This methodology consists of evaluating the difference between test set predictions obtained by the derived QSAR model and those determined by a hypothetical ideal model.

Hence, this validation methodology was applied to the activity predictions of compounds in the test set of all models reported in this paper. The results of the validation procedure are shown in Table 4 and plotted in Fig. 4.

It can be seen that all CoMFA and CoMSIA models predicted, in a good manner, the activities of the compounds in the test set (this is represented by values of $R^2 > 0.75$). Examining the results for the CoMFA models, it can also be seen that the R^2 value of model 1 is higher than the one presented by model 2 (0.856 and 0.763, respectively); this is indicative that model 1 has better predictive power than model 2. Nevertheless, all the models fulfil the criteria for $[(R^2 - R_0^2)/R^2]$ and k (see Methods); in other words, all the models have an acceptable predictive power. By analyzing the CoMFA validation results in a detailed manner, it can be seen that model 2 has the better results; that is to say, a value of $[(R^2 - R_0^2)/R^2]$ near zero and the k value closest to 1. Since the only difference between these two models is the molecular alignment, model 2 is still selected as the best CoMFA model because the alignment used for this model is supported by experimental evidence.

Now focusing on the CoMSIA validation results, a similar situation was found. The model based on SEDA fields fulfils the validation criteria better; even so it does not have the highest

Table 1 Chemical structures and pIC₅₀ values of molecules used in this study

Compound	R ¹	R ²	IC ₅₀ (μM)	pIC ₅₀
1	Ethyl		0.71	6.1487
2	Cyclopentyl		0.71	6.1487
3	Propyl		0.79	6.1024
4	Isopropyl		0.89	6.0506
5 ^a	Methylcyclopropyl		0.89	6.0506
6	Isobutyl		0.89	6.0506
7	Butyl		0.89	6.0506
8	4-Hydroxybutyl		0.89	6.0506
9	Methyl		1.26	5.8996
10 ^a	Cyclohexyl		1.41	5.8508
11	2-Hydroxyethyl		1.58	5.8013
12	Cyclopentyl	3-Nitro	0.063	7.2006
13 ^a	Cyclopentyl	3-Fluoro	0.079	7.1024
14	Cyclopentyl	3-Chloro	0.079	7.1024
15	Cyclopentyl	3-Bromo	0.1	7.0000
16	Cyclopentyl	4-Fluoro	0.126	6.8996
17	Cyclopentyl	3-Methyl	0.126	6.8996
18	Cyclopentyl	4-Bromo	1.122	5.9500
19	Cyclopentyl	3-Phenyl	2.239	5.6499
20	Cyclopentyl	3,5-Difluoro	0.063	7.2007
21 ^a	Cyclopentyl	3,5-Dichloro	0.398	6.4001
22	2,2-Difluoroethyl	3,5-Difluoro	0.025	7.6021
23	Ethyl	3,5-Difluoro	0.01	8.0000
24 ^a	2,2-Difluoroethyl	3,5-Difluoro	0.013	7.8860
25	Cyclopentyl	3,5-Difluoro	0.018	7.7447
26	Ethyl	3-Chloro	0.018	7.7447
27	Cyclopentyl	3-Chloro	0.04	7.3979
28	3,5-Difluoro	2,2-Difluoroethyl	0.05	7.3010
29	3-Chloro	Ethyl	0.063	7.2007
30	3-Chloro	2,2-Difluoroethyl	0.071	7.1487
31 ^a	3,5-Difluoro	Ethyl	0.251	6.6003
32	3,5-Difluoro	Cyclopentyl	0.316	6.5003

^a Compounds in the test set.

value of R^2 , but this can be explained in terms of the molecular diversity in the test set. It would be more useful for the future to use numerous test groups with greater molecular diversity to observe if the value of R^2 correlates with the one of q^2 . An interesting aspect of these results is that in all the three CoMSIA models with the highest value of q^2 , the R^2 value is higher than the one presented in model 2 of the CoMFA study. This difference suggests that the CoMSIA models, in this study, have a better predictive ability than the CoMFA models.

CoMFA contour maps

The most important information that a CoMFA model can give, in addition to the prediction of pIC₅₀, is the contour maps, which can guide to a more rationalized lead optimization. The 3D contour maps obtained from model 2 give information about the structural features to assess active compounds. In order to obtain better-defined maps, a column filtering of 2 kcal was included in

model 2; it is important to mention that the predictive ability of model 2 was not affected by this modification. The steric (green and yellow) and electrostatic (red and blue) contour maps are shown in Fig. 5. The most potent inhibitor (**23**) is shown in balls and sticks, while the rest of the compounds are represented by lines.

By analyzing contour maps in Fig. 5, it can be seen that a large green contour (sterically favored) and a smaller one are found near positions 3 and 5 of the phenyl ring. This could be indicative that halogens at these positions on the phenyl could increase the biological activity. Indeed, the inhibition presented by molecules **12–32**, where most of them are substituted in those positions by bulky groups, mainly halogens, has a pIC₅₀ > 6.4. On the other hand, compounds **1–11**, which have no substitution on the phenyl ring, presented considerably low biological activity (<6.2). One exception to the above mentioned is compound **19**, which has a 3-phenyl substituent close to a yellow contour (sterically unfavored). This contour is in good agreement with

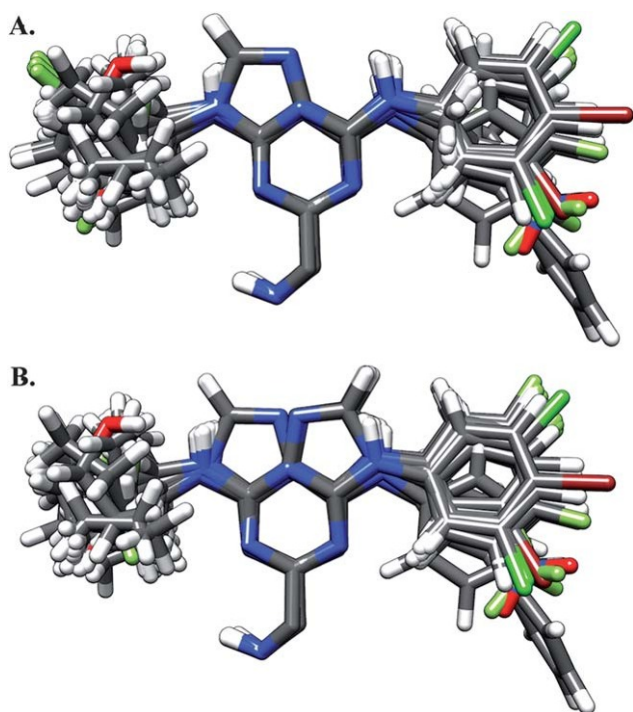


Fig. 1 Different alignments used in this study: (A) is based on crystallographic data of compound **23** bound to cruzain; (B) is based on the fact that the S_2 pocket has a predilection for hydrophobic residues.

the extremely low pIC_{50} presented by this compound (5.6499), and gives information about the size of the S_2 pocket where this group is placed. Based on the results obtained with the current database we conclude that the S_2 pocket is large enough to hold bulky groups such as 3,5-difluorophenyl or 3,5-dichlorophenyl, but it is not big enough to contain groups such as a 3-biphenyl.

Another yellow contour appears near the cyclopentyl substituent at position 9 of the purine ring in molecules **12–21**, **25**, **27** and **32**. In this area, bulky groups would decrease activity. In

fact, this observation is exemplified by comparing compounds **25**, **27** and **32** with molecules **23**, **26** and **31**, respectively. In all cases, molecules with an ethyl group instead of the cyclopentyl are more active. This effect could be related to the fact that these groups are exposed to the solvent; the less lipophilic these groups are, the more interactive they are with water molecules in the solvent, nevertheless more specialized calculations are needed to support this idea.

The electrostatic contours are in red when a negative potential is favorable for the enzyme inhibition, whereas blue contours are close to areas where a negative potential is unfavorable. A blue contour appears near position 3 of the phenyl group and it is related to the 3-phenyl substituent of compound **19**, the compound with lowest biological activity of the entire database. As it can be seen in Fig. 5, the blue contour crosses through the center of a phenyl group, exactly where a high electron density is concentrated due to the π electrons.

CoMSIA contour maps

One of the advantages of the CoMSIA over CoMFA methodology is that the former produces better-defined contour maps. This circumstance arises from the fact that the use of molecular similarity indices avoids the arbitrary definitions of cutoff values, and the descriptors can be calculated in all grid points.¹⁰ In this section, contour maps obtained from the CoMSIA model in which SEDA fields were included (Fig. 6) will be discussed. In Fig. 6A the contour maps corresponding to steric and electrostatic fields are displayed; compound **23** is shown in a ball and stick representation while the rest of the compounds are just in lines. Interestingly, the contour maps for steric interactions are very similar to those of CoMFA; that is, a green contour (sterically favored) upon position 3 of the phenyl group, while a yellow one (sterically disfavored) is above the cyclopentyl substituent. Nevertheless, some differences are found in electrostatic contours as compared with CoMFA. In CoMSIA, red contours (areas where a negative potential is favorable) are near positions

Table 2 Summary of CoMFA models 1–3 and CoMSIA models using SED, SEA, SEDA and SEHDA fields

Statistics	CoMFA			CoMSIA			
	Model 1	Model 2	Model 3	SED	SEA	SEDA	SEHDA
Alignment	A	B	B	B	B	B	B
Grid spacing/ \AA	2.0	2.0	1.0	2.0	2.0	2.0	2.0
q^{2a}	0.623	0.736	0.692	0.627	0.618	0.565	0.460
Press ^b	0.488	0.408	0.441	0.473	0.479	0.523	0.533
r^{2c}	0.983	0.980	0.977	0.888	0.874	0.901	0.712
S^d	0.104	0.112	0.122	0.259	0.275	0.249	0.389
F^e	227.096	196.155	166.508	41.591	36.544	36.586	59.474
R^{2f}	0.856	0.763	0.753	0.806	0.833	0.815	0.758
ONC ^g	5	5	5	4	4	5	1
Contribution							
Steric	0.815	0.816	0.804	0.367	0.461	0.336	0.159
Electrostatic	0.185	0.184	0.196	0.259	0.268	0.235	0.080
Hydrophobic							0.307
Hydrogen bond donor				0.374		0.347	0.304
Hydrogen bond acceptor					0.271	0.082	0.150

^a Cross-validated correlation coefficient from LOO. ^b Standard error of predictions derived from the LOO method. ^c Non-crossvalidated r^2 . ^d Standard error estimate. ^e F -Test value. ^f Correlation coefficient derived from predictions of test set molecules. ^g Optimum number of principal components.

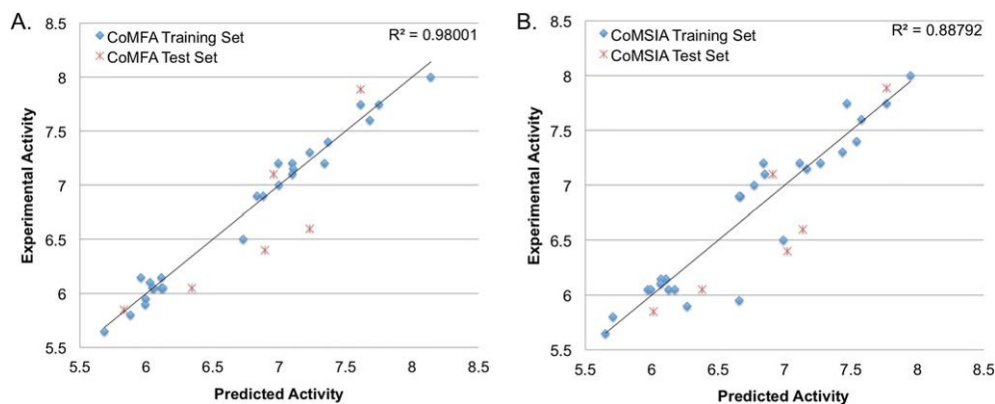


Fig. 2 Plots of experimental vs. predicted values of (A) CoMFA model 2 and (B) CoMSIA including SED fields.

Table 3 Training and test sets predictions using CoMFA model 2 and CoMSIA including SED

Compound	pIC ₅₀	CoMFA		CoMSIA	
		Prediction	Residual	Prediction	Residual
1	6.1487	6.1119	0.0368	6.107	0.0417
2	6.1487	5.9593	0.1894	6.0697	0.0791
3	6.1024	6.0296	0.0728	6.0662	0.0362
4	6.0506	6.0463	0.0043	5.9688	0.0818
5 ^a	6.0506	6.3420	-0.2914	6.3777	-0.3271
6	6.0506	6.0562	-0.0056	6.1253	-0.0747
7	6.0506	6.1141	-0.0635	6.1739	-0.1233
8	6.0506	6.125	-0.0744	5.9938	0.0568
9	5.8996	5.9918	-0.0922	6.2667	-0.3671
10 ^a	5.8508	5.8323	0.0184	6.0114	-0.1606
11	5.8013	5.8808	-0.0795	5.7081	0.0933
12	7.2006	7.3393	-0.1387	7.1142	0.0864
13 ^a	7.1024	6.9560	0.1464	6.9091	0.1932
14	7.1024	7.0989	0.0035	6.8515	0.2509
15	7.0000	6.9979	0.0021	6.7704	0.2296
16	6.8996	6.8324	0.0672	6.6669	0.2327
17	6.8996	6.8793	0.0204	6.6588	0.2408
18	5.9500	5.9941	-0.0441	6.6577	-0.7077
19	5.6499	5.6839	-0.0339	5.6500	-0.0001
20	7.2007	7.0986	0.1021	7.2680	-0.0674
21 ^a	6.4001	6.8935	-0.4934	7.0180	-0.6179
22	7.6021	7.6803	-0.0782	7.5783	0.0238
23	8.0000	8.1405	-0.1405	7.9463	0.0537
24 ^a	7.8860	7.6108	0.2752	7.7680	0.1180
25	7.7447	7.6108	0.1339	7.768	-0.0233
26	7.7447	7.7499	-0.0051	7.4673	0.2775
27	7.3979	7.3678	0.0302	7.5408	-0.1429
28	7.3010	7.2295	0.0716	7.4335	-0.1325
29	7.2007	6.9939	0.2067	6.8395	0.3611
30	7.1487	7.1043	0.0444	7.1660	-0.0173
31 ^a	6.6003	7.2303	-0.6300	7.1365	-0.5361
32	6.5003	6.73	-0.2297	6.9896	-0.4893

^a Compounds in the test set.

3 and 5 of the phenyl group, very close to the green contours discussed before. These red contours are related to the halogens in compounds **20–25**, **28**, **31**, **32**, and to the 3-nitro substituent of compound **12**. It is important to remark that within these compounds they are the most active molecules in this database. Blue contours related to areas where negative potentials are unfavorable almost disappeared. This is a good example of how

CoMSIA and CoMFA complement each other. In this case CoMFA was able to detect that the π electrons decrease the activity of compound **19**, whereas CoMSIA detected the importance of halogens for biological activity.

Fig. 6B shows the contour maps related to the hydrogen bond donor and acceptor moieties. Hydrogen bond acceptor moieties are favorable for biological activity in regions near orange contours and unfavorable near yellow contours. In the case of hydrogen bond donors, these favor the activity if they are close to cyan contours and disfavor when near purple contours. Fig. 6B shows that a yellow contour surrounds the triazine nitrogen that is overlapped with the bridgehead carbon of purine. This contour might be explained as the preference of a carbon instead of nitrogen at this position in molecules with a triazine ring; in terms of the compounds used in this study, this result reflects a preference for the purine over the triazine ring. Comparing compounds **14** with **27** and **22** with **24**, it is clearly seen that compounds with a purine scaffold are better cruzain inhibitors than those with a triazine ring. Interestingly, no interaction was found between cruzain and the purine scaffold in the crystal structure of compound **23**. There exists the possibility that the purine scaffold could increase the electrophilicity of the nitrile group making it more susceptible to a nucleophilic attack from the thiolate of Cys25 as studied by Oballa *et al.*¹⁶ For this reason, a possible explanation for the differences in biological activity among the triazine and the purine inhibitors could be the different reactivity toward the enzyme, but it would be necessary to have a deeper study to confirm this hypothesis.

Another yellow contour appears pointing to the secondary amine that links the purine ring with the R² substituents in compounds **29–32**; in the same way, an orange and a cyan contour point to the amine linking the purine and the phenyl ring in compounds **22–28**. It is worth mentioning that these contours are giving information about the purine orientation based on the secondary amine position, instead of showing the interaction of the amine as a hydrogen bond donor or hydrogen bond acceptor. In this set of compounds, these contours can be interpreted as a preference for the phenylamino group at position 6 of the purine ring as it is in compounds **22–28**. A direct comparison between compounds **23–26** and **31**, **28**, **32**, **29**, respectively, shows that the phenylamino group at position 6 (and not at position 9) of purine confers a major activity on these compounds.

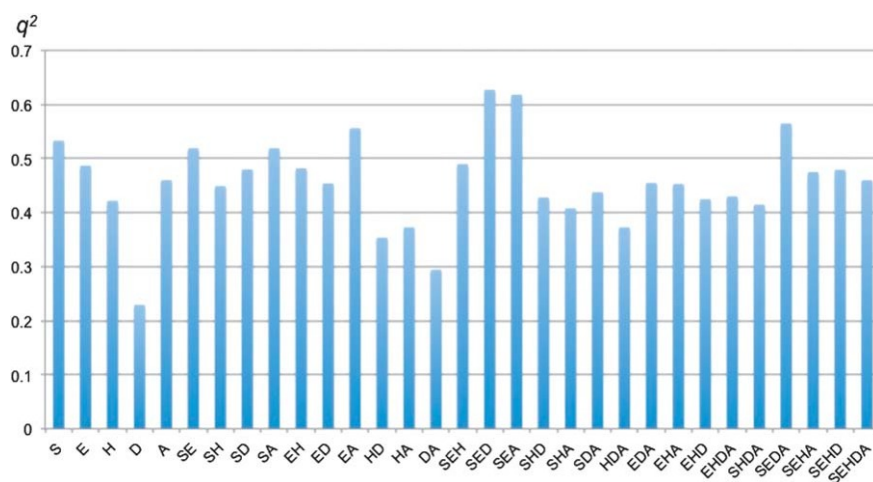


Fig. 3 Histogram presenting the q^2 value of the 31 possible CoMSIA field combinations.

Methods

Dataset

The dataset used in this study contained 32 triazine and purine nitrile molecules which have been previously published.⁸ These compounds were divided into a training set used to derive the 3D-QSAR models and a test set used for the validation procedure. The test set was selected manually in order to screen the whole range of activities and diversity of structures. All the activities were converted to pIC_{50} ($-\log IC_{50}$) values shown in Table 1.

Preparation of ligands and molecular alignment

All molecules were constructed using as a template the structure of the inhibitor 6-[(3,5-difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile (**23**) co-crystallized with cruzain obtained from the protein data bank (access number: 3I06).⁸ Then, the calculations of the ligands' energy minimization were aided by using TRITON,¹⁷ a graphic user interface for MOPAC that allowed us to minimize the ligands inside a cavity formed by 20 amino acids in the cruzain's active site (Glu19, Cys22, Gly23,

Ser24, Cys25, Trp26, Cys63, Ser64, Gly65, Gly66, Leu67, Met68, Val137, Ala138, Val139, Leu160, Asp161, His162, Gly163, Glu208). Finally, the calculations were carried out with MOPAC2009 employing the PM6 hamiltonian, the localized molecular orbital method MOZYME^{18–22} and a continuum COSMO model with the water dielectric constant of 78.4.²³ The backbone atoms of the amino acids were fixed and all the side chain and ligand atoms were optimized using the BFGS method.^{24–27}

After the calculations, the molecules were extracted from the cavity used for the minimization to start the alignment procedure. It is well known that the alignment of molecules plays a fundamental role in 3D-QSAR to obtain high quality models; for this reason, two different alignments were used in this research. The first one consisted of overlapping the imine group, which is a common substructure in all molecules (Fig. 1A). In a second approach, the bond between the purine ring and the nitrile moiety was rotated 180° in compounds **28–32**, thus making the phenyl group match with that of compound **23** inside a pocket next to Cys25. After these modifications, the molecules were submitted to a second geometry optimization inside the cavity and finally aligned using the common substructure previously defined (Fig. 1B).

Table 4 Summary of validation data for CoMFA models 1–3 and CoMSIA models using SED, SEA, SEDA and SEHDA fields

Statistics	CoMFA			CoMSIA			
	Model 1	Model 2	Model 3	SED	SEA	SEDA	SEHDA
R^{2a}	0.856	0.763	0.753	0.806	0.833	0.815	0.758
R_0^{2b}	0.705	0.654	0.540	0.676	0.695	0.725	0.553
k^c	1.028	1.021	1.030	1.030	1.026	1.024	1.026
$(R^2 - R_0^2)/R^2$	0.176	0.142	0.282	0.161	0.166	0.110	0.269
R_0^{12b}	0.829	0.761	0.739	0.795	0.815	0.810	0.744
$k^{c'}$	0.971	0.976	0.968	0.968	0.972	0.975	0.972
$(R'^2 - R_0'^2)/R'^2$	0.032	0.002	0.018	0.013	0.022	0.006	0.017

^a Correlation coefficient derived from predictions of test set molecules. ^b Correlation coefficients for the regression through origin for experimental vs. predicted and predicted vs. experimental activity, respectively. ^c Slopes for regression through origin for experimental vs. predicted and predicted vs. experimental, respectively.

CoMFA and CoMSIA

For the CoMFA calculations, a three-dimensional lattice with a grid spacing of 2.0 Å was created and extended 4 Å away from the molecule for each model using SYBYL8.0.²⁸ As an alternative approach, the same lattice was used with a grid space of 1.0 Å. In each lattice intersection a sp^3 atom with a van der Waals radius of 1.52 Å was used as a steric probe and a +1 charge as an electrostatic probe. The energy cut-off default value of 30 kcal mol⁻¹ was used. The CoMSIA models were derived using the same training set and alignment as those of the best CoMFA model. A grid spacing of 2.0 Å was used to calculate five different similarity indices: steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor. An attenuation factor of 0.3 was used for the Gaussian function.

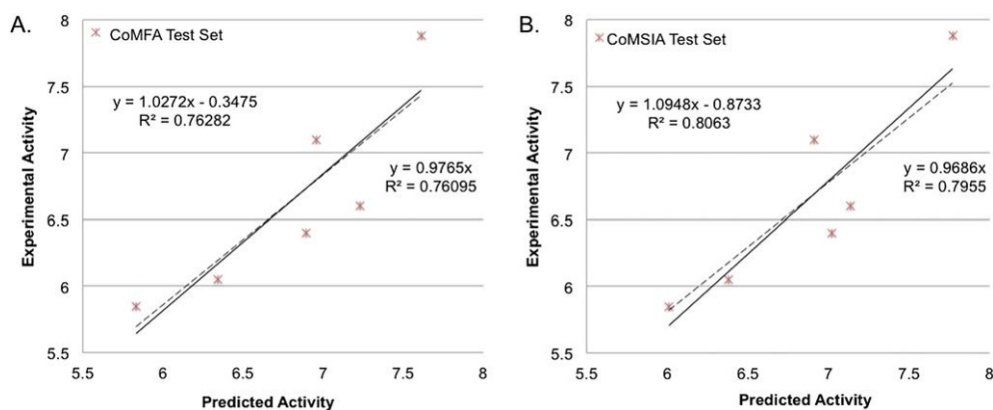


Fig. 4 Validation using the test set for (A) CoMFA model 2 and (B) CoMSIA including SED fields.

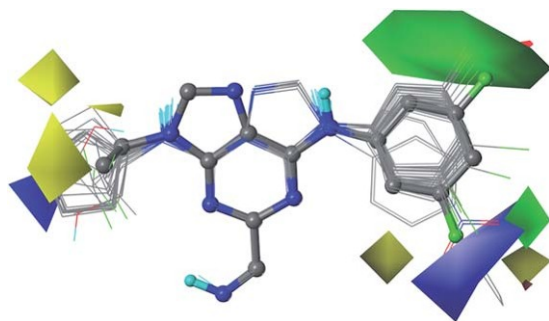


Fig. 5 CoMFA contour maps obtained from model 2 (with a column filtering of 2 kcal). Green contours denote sterically favored regions and yellow contours sterically unfavored regions. Blue contours indicate areas where a negative potential increases activity, whereas red contour regions indicate areas where a negative potential decreases it.

Statistical analysis and validation

In order to derive CoMFA and CoMSIA models, partial least squares analysis was carried out using the descriptors as independent variables and the pIC_{50} as a dependent variable. Cross-validation analysis was performed using the Leave-One-Out (LOO) method to determine the optimal number of components and to avoid data over fitting. The best model was chosen using as parameters the cross-validation (q^2) and the correlation (r^2) coefficients.

Additional validation was performed predicting the activity of an external test set with 6 compounds using all the CoMFA and

CoMSIA models. The predictions were analyzed by regression analysis, applying the criteria proposed by Golbraikh and Tropsha,¹⁵ as follows:

- (1) $q^2 > 0.5$.
- (2) $R^2 > 0.6$.
- (3) $[(R^2 - R_0^2)/R^2] < 0.1$ or $[(R^2 - R_0'^2)/R^2] < 0.1$.
- (4) $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$.

where q^2 is the cross-validated correlation coefficient from LOO; R^2 is the correlation coefficient for experimental (y) vs. predicted (\bar{y}) activities for test set molecules; R_0^2 and $R_0'^2$ are the correlation coefficients for the regression through origin for y vs. \bar{y} and \bar{y} vs. y respectively; k and k' are the slopes for regression through origin $y^0 = k\bar{y}$ and $\bar{y}^0 = k'y$.

Conclusions

A 3D-QSAR (CoMFA and CoMSIA) study was carried out on a set of 32 purine and triazine carbonitriles, which have been identified as inhibitors of cruzain, the most important cysteine protease of *Trypanosoma cruzi*. Geometry optimization of the ligands inside the active site of the enzyme using semi-empirical calculations yielded a good approach to obtain reliable conformations. The problem of molecular alignment was faced by using valuable experimental data (crystallographic structure of compound **23** (ref. 8) and information of the catalytic site of papain¹²) that led us to two different alignments, from which CoMFA models 1 and 2 were derived. Model 2 was selected as the best CoMFA model, which yielded quite satisfactory values of q^2 (0.736) and r^2 (0.980).

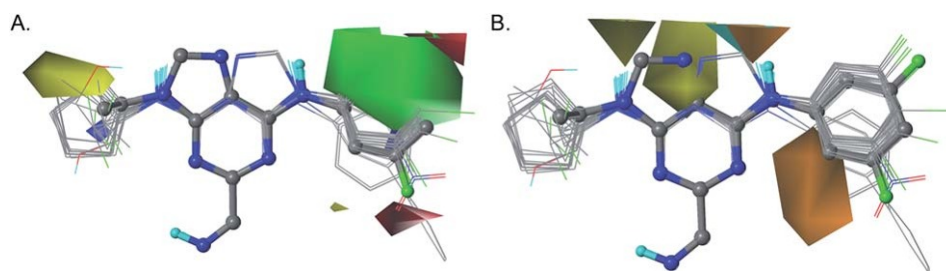


Fig. 6 CoMSIA contour maps including SEDA fields. (A) Depicts steric and electrostatic contour maps, the ones that have the same color code as CoMFA. (B) Orange contours depict regions in which hydrogen bond acceptors are favorable, whereas those near yellow contours are unfavorable. In the same way cyan contours indicate regions where hydrogen bond donors increase activity, whereas those near purple contours decrease activity.

The 31 possible combinations among the 5 different fields of CoMSIA were studied. The CoMSIA model that included the SEDA fields presented good statistical parameters with a q^2 of 0.565, an r^2 of 0.901 and a R^2 of 0.815 and it was the one that has given us more information about the structural requirements for biological activity. All the CoMFA and CoMSIA models fulfil the validation procedure proposed by Golbraikh and Tropsha.¹⁵

Four main structural requirements for biological activity were revealed by this study: a purine scaffold is preferred over the triazine ring; bulky groups with a negative electrostatic potential, such as halogens, seem to be important at positions 3 and 5 of the phenyl group; small groups are preferred at position 9 of the purine scaffold; the phenyl group must be linked to purine at position 6 by a secondary amine.

In conclusion, the models reported in this paper show good predictive power, and the information concerning the structural requirements could be useful for optimization of these molecules or for the design of new compounds.

Acknowledgements

This research was supported by CONACyT project no. 80093. O. M.-L and A. R.-M are very grateful to CONACyT for the fellowships granted (no. 245408 and 173861, respectively).

References

- 1 J. Clayton, *Nature Outlook*, June 24th, 2010, 54–55.
- 2 J. Rodrigues-Coura and P. Albajar-Viñas, *Nature Outlook*, June 24th, 2010, 56–57.
- 3 J. Cazzulo, V. Stoka and V. Turk, *Curr. Pharm. Des.*, 2001, **7**, 1143–1156.
- 4 M. McGrath, A. Eakin, J. Engel, J. McKerrow, C. Craik and R. Fletterick, *J. Mol. Biol.*, 1995, **247**, 251–259.
- 5 E. Dufour, A. Storer and R. Menard, *Biochemistry*, 1995, **34**, 9136–9143.
- 6 J. Palmer, D. Rasnick, J. Klaus and D. Bromme, *J. Med. Chem.*, 1995, **38**, 3193–3196.
- 7 J. Drenth, K. H. Kalk and H. M. Swen, *Biochemistry*, 1976, 1–8.
- 8 B. T. Mott, R. S. Ferreira, A. Simeonov, A. Jadhav, K. K.-H. Ang, W. Leister, M. Shen, J. T. Silveira, P. S. Doyle, M. R. Arkin, J. H. Mckerrow, J. Inglese, C. P. Austin, C. J. Thomas, B. K. Shoichet and D. J. Maloney, *J. Med. Chem.*, 2010, **53**, 52–60.
- 9 R. Cramer, D. Patterson and J. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959–5967.
- 10 G. Klebe, U. Abraham and T. Mietzner, *J. Med. Chem.*, 1994, **37**, 4130–4146.
- 11 H. Kubinyi, *Drug Discovery Today*, 1997, **2**, 457–467.
- 12 I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, 1968, **32**, 898–902.
- 13 R. R. Mittal, R. A. McKinnon and M. J. Sorich, *J. Mol. Model.*, 2008, **14**, 59–67.
- 14 J. L. Melville and J. D. Hirst, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1294–1300.
- 15 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.
- 16 R. M. Oballa, J.-F. Truchon, C. I. Bayly, N. Chauret, S. Day, S. Crane and C. Berthelette, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 998–1002.
- 17 M. Prokop, J. Adam, Z. Kriz, M. Wimmerova and J. Koca, *Bioinformatics*, 2008, **24**, 1955–1956.
- 18 J. J. P. Stewart, *Mopac 2009 Version 10.105L*, Fujitsu Limited, Tokyo, Japan, 2006.
- 19 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.
- 20 J. J. P. Stewart, *J. Mol. Model.*, 2009, **15**, 765–805.
- 21 J. J. P. Stewart, *Int. J. Quantum Chem.*, 1996, **58**, 133–146.
- 22 J. J. P. Stewart, *J. Mol. Struct.*, 1997, **401**, 195–205.
- 23 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- 24 C. G. Broyden, *J. Inst. Math. Appl.*, 1970, **6**, 222–231.
- 25 R. Fletcher, *Comput. J.*, 1970, **13**, 317–322.
- 26 D. Goldfarb, *Math. Comput.*, 1970, **24**, 23–26.
- 27 D. F. Shanno, *Math. Comput.*, 1970, **24**, 647–656.
- 28 SYBYL 8.0, Tripos International, 1699 South Hanley Rd., St Louis, Missouri, 63144, USA.



Computational study on the inhibition mechanism of cruzain by nitrile-containing molecules

Oscar Méndez-Lucio^a, Antonio Romo-Mancillas^a, José L. Medina-Franco^b, Rafael Castillo^{a,*}

^a Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, México 04510, D.F., Mexico

^b Torrey Pines Institute of Molecular Studies, 11350 SW Village Parkway, Port St. Lucie, FL 34987, USA

ARTICLE INFO

Article history:

Received 25 October 2011

Received in revised form 6 January 2012

Accepted 10 January 2012

Available online 20 January 2012

Keywords:

Cruzain

Cysteine proteases

Density functional theory

Semi-empirical quantum mechanics

Trypanosoma cruzi

ABSTRACT

Cysteine proteases from parasites as well as from mammals are promising drug targets for parasitic infections and systemic human diseases, respectively. Many reversible and irreversible inhibitors of this very large class of proteins have been designed. Among others, molecules with a nitrile moiety, which is a group that is susceptible to a nucleophilic attack by the enzyme, have been identified as good inhibitors. Although it is known that the nitrile group binds covalently to Cys25, there are no reports about the energetics involved in the mechanism of this process. Herein, density functional theory and quantum semi-empirical calculations were conducted in order to study the molecular recognition of cysteine proteases by nitrile-containing molecules. Results reported in this paper suggest an interaction that starts with a nucleophilic attack from the Cys25 to the inhibitor followed by a proton transfer from His162. Only one transition state was detected; however, we found the existence of an energy plateau in the potential energy surface. Based on the proposed mechanism, some structural features that could improve the biological activity of nitrile-containing molecules toward cysteine proteases are discussed.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Cysteine proteases are involved in many biological processes. The misregulation of these enzymes conduces to a variety of diseases; therefore, they represent promising drug targets [1,2]. The huge class of cysteine proteases are divided into 20 families and grouped into three clans. Clan A comprises the C1, C2 and C10 families, which can be described as “papain-like proteases” [3]. Cysteine proteases from parasites as well as from mammals that belong to this clan are promising drug targets for parasitic infections and systemic human diseases [2]. For example, cathepsins B, F, K and L are involved in rheumatoid arthritis, cancer, osteoporosis and atherosclerosis, while falcipain, cruzain, and rhodesain are drug targets against malaria (*Plasmodium falciparum*), Chagas' disease (*Trypanosoma cruzi*) and sleeping sickness (*Trypanosoma brucei*), respectively [1,2].

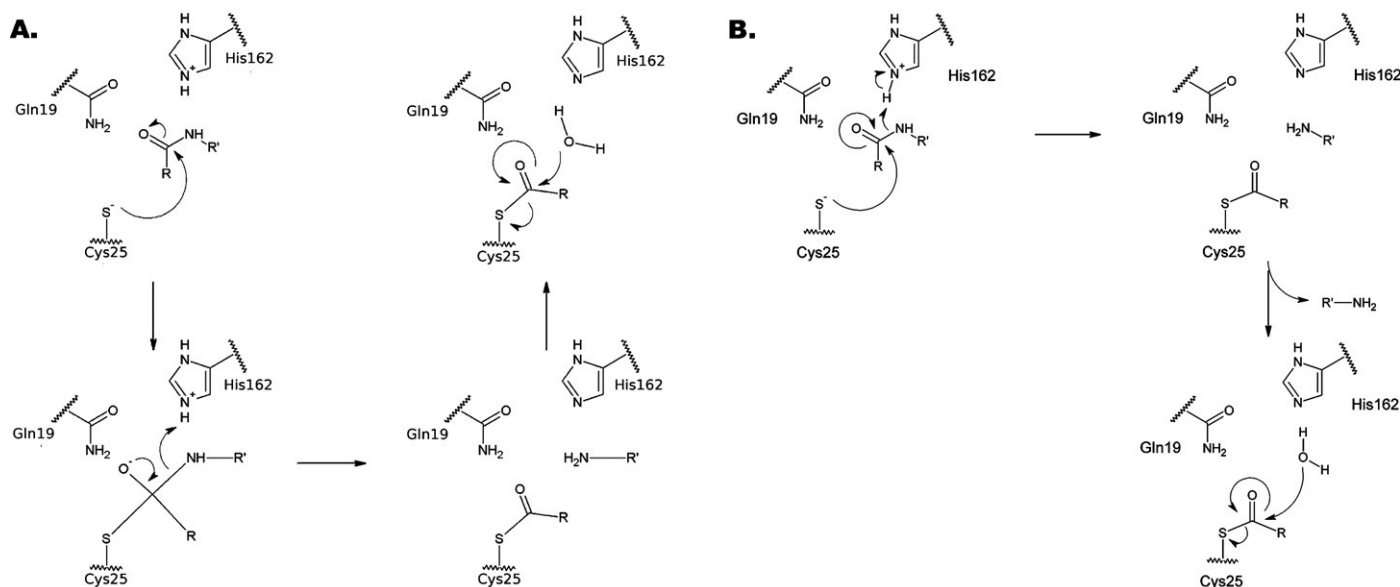
The active site of the cysteine proteases is composed of the catalytic triad Cys25, His162, and Asn182 (cruzain numbering) [4]. The proposed mechanism for this triad operates as follows: Asn182 acts as a hydrogen bond acceptor toward the His162, enabling the formation of a permanent imidazolium–thiolate ion pair between Cys25 and His162 [5]. It has been proposed that another

important amino acid in the catalytic site is Gln19, which acts as part of an oxyanion hole that stabilizes an acyl-enzyme intermediate by forming a hydrogen bond between the amino group in the Gln19 and the substrate [6].

Two main catalytic mechanisms of enzymatic activity of cysteine proteases have been proposed. The former consists of a serine protease-like mechanism proposed by Drenth et al. [7] that takes place in three steps (Scheme 1A). The first step is the nucleophilic attack on the substrate by the thiolate. A stabilization of the transition state in this reaction occurs by the hydrogen bond between the substrate and the backbone NH of Cys25. The resulting tetrahedral intermediate is stabilized by a hydrogen bond formed between Gln19 in the oxyanion hole and the ligand. The latter is a proton transfer from the imidazolium ion to the nitrogen of the amide in the substrate to form an acyl-enzyme complex. Finally, the deacylation occurs via a water molecule which attacks the acyl-enzyme complex at the carbonyl carbon to produce the final products [7]. An alternative mechanism has been proposed in which the protonation of the oxygen or nitrogen in the substrate caused by a proton transfer from His162 must occur prior, or in a concerted manner, to the nucleophilic attack [8–10].

A wide range of peptidic or non-peptidic molecules with a group susceptible to a nucleophilic attack can inhibit cysteine proteases. For example, vinyl sulfones, fluoro methyl ketones, aziridines or nitriles have shown good inhibition activity [1,2]. Nevertheless, the inhibition mechanism of only a few of them has been studied, both

* Corresponding author. Tel.: +52 5 56 22 52 87; fax: +52 5 56 22 53 29.
E-mail addresses: rafaelc@unam.mx, rafaelc@servidor.unam.mx (R. Castillo).



Scheme 1. (A) Serine protease-like catalytic mechanism proposed for cysteine proteases. (B) Alternative catalytic mechanism in which the proton transfer occurs prior, or in a concerted manner, to the nucleophilic attack.

experimentally and theoretically [11–15]. Further advances in theoretical methods should facilitate mechanistic studies that could result in new inhibitor classes [16].

Cruzain is the major cysteine protease of *T. cruzi*, the causal agent of Chagas' disease, and has been identified as an important enzyme for the parasite survival on the host [17,18]. Because of the biological importance of cruzain, many attempts to inhibit this enzyme have been undertaken. Recently, the inhibitor 6-[(3,5-difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile (Table 1) was identified as an inhibitor of cruzain with an IC_{50} of 0.01 μ M [19]. As stated above, molecules with a nitrile moiety have been used as a reversible covalent inhibitor of cysteine proteases [1,2], although their inhibition mechanism has not yet been studied from a computational perspective. For this reason, the principal aim of this paper is to gain insights into the mechanism whereby the nitrile inhibitors bind to cysteine proteases, especially to cruzain. Density functional theory and quantum semi-empirical calculations were employed to analyze the energy profile of the putative mechanism of inhibition of cruzain. This paper is part of our on going efforts to rationalize, at the molecular level, the structure–activity relationships and mechanism of inhibition of purine-carbonitriles that serve as guide to design new inhibitors of cruzain [20].

2. Computational methods

2.1. Density functional theory calculations

For these calculations an isolated model including the side chains of two residues (Cys25 and His162) and a purine-carbonitrile as a model of the inhibitors were used (shown in Table 1). The methyl groups corresponding to the protein backbone were the only atoms fixed in these calculations. A three-dimensional representation of the protein model used for these calculations is depicted in Fig. 1.

A search along the potential energy surface of the binding recognition of cruzain by a purine-carbonitrile was performed using the B3LYP hybrid functional [21,22] with the 6-31+G(d) basis set. The stationary point corresponding to a transition state was characterized by an imaginary frequency resulting from the

harmonic vibrational frequency calculations. On the other hand, reactants and products were confirmed by the absence of imaginary frequencies. Intrinsic reaction coordinate calculations (using a step size of 0.13 Bohr) were performed to confirm the connectivity between the transition state and the desired reactants and products. Single-point calculations were computed at M05-2X/6-31+G(d)//B3LYP/6-31+G(d) level of theory for all the IRC and stationary points, as proposed by Goodman and co-worker after a benchmark study [23]. Since the catalytic site of the cruzain is located in an external region of the protein, which is exposed to solvent, a polarizable continuum model (PCM) was used applying the integral equation formalism variant (IEFPCM) with water as solvent during all the calculations. All calculations were carried out using Gaussian09 software [24].

Fukui indices [25], which were computed in order to assess reactive sites within a molecule, can be defined as:

$$f^-(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_v^- = \rho_{N_0}(r) - \rho_{N_0-1}(r)$$

$$f^+(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_v^+ = \rho_{N_0+1}(r) - \rho_{N_0}(r)$$

and

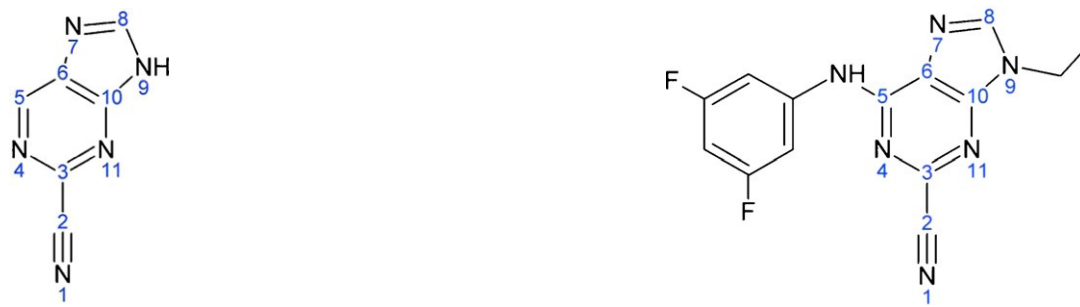
$$f^0(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_v^0 = \frac{1}{2}(f^-(r) + f^+(r)) = \frac{1}{2}(\rho_{N_0+1}(r) - \rho_{N_0-1}(r))$$

where ρ_{N_0-1} , ρ_{N_0} , and ρ_{N_0+1} are the electron densities of the $N_0 - 1$, N_0 , and $N_0 + 1$ electron systems. The function $f^-(r)$ defines the place where the removal of a fraction of electron is more energetically favorable, while $f^+(r)$ defines the place where it is more favorable to add it. In terms of reactivity, large values of $f^-(r)$ would indicate a good nucleophile; on the other hand, large values of $f^+(r)$ highlight susceptible areas for a nucleophilic attack [26].

2.2. Quantum semi-empirical calculations

In order to investigate possible interactions between inhibitors with a nitrile group and the active site of cruzain, the crystallographic structure of cruzain co-crystallized with the

Table 1
Calculated Fukui function values computed at B3LYP/6-31+G(d).



Atom	NBO			Mulliken			NBO			Mulliken		
	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0	f_A^+	f_A^-	f_A^0
N1	-0.18	-0.203	-0.191	-0.174	-0.181	-0.177	-0.154	-0.089	-0.121	-0.149	-0.085	-0.117
C2	0.018	0.01	0.014	-0.007	-0.015	-0.011	0.008	0.032	0.020	0.007	0.001	0.004
C3	-0.127	-0.113	-0.120	-0.072	-0.05	-0.061	-0.118	-0.031	-0.074	-0.055	0.029	-0.013
N4	-0.018	-0.042	-0.030	-0.068	-0.096	-0.082	0.000	-0.034	-0.017	-0.049	-0.04	-0.044
C5	-0.144	-0.079	-0.111	-0.146	-0.033	-0.089	-0.094	0.029	-0.032	-0.025	0.041	0.008
C6	-0.056	-0.09	-0.073	-0.078	-0.087	-0.082	-0.046	-0.062	-0.054	-0.039	-0.036	-0.037
N7	-0.05	-0.095	-0.072	-0.064	-0.125	-0.094	-0.031	-0.016	-0.023	-0.055	-0.042	-0.048
C8	-0.143	-0.136	-0.139	-0.106	-0.066	-0.086	-0.118	-0.069	-0.093	-0.129	-0.038	-0.083
N9	-0.024	-0.013	-0.018	-0.011	-0.041	-0.026	-0.029	-0.015	-0.022	0.027	-0.004	0.011
C10	-0.003	-0.063	-0.033	0.053	-0.009	0.022	-0.003	-0.03	-0.016	0.001	-0.031	-0.015
N11	-0.155	-0.056	-0.105	-0.157	-0.11	-0.133	-0.114	-0.081	-0.097	-0.116	-0.092	-0.104

inhibitor 6-[(3,5-difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile (Table 1) was obtained from the Protein Data Bank (PDB ID: 3I06) [19].

The potential energy surface for this reaction was mapped by scanning two different reaction coordinates at the same time: the distances C2–S and N1–H. These atoms were approached and geometry optimizations were carried out every 0.05 Å using the PM6 quantum semi-empirical method. A similar methodology has been used before studying different enzyme reaction mechanisms [27–30].

Since the aim of this paper is to study a reaction that takes place in the catalytic site of the enzyme, 275 atoms in the active site and 32 atoms in the ligand were explicitly treated by the PM6 quantum semi-empirical method. The calculations were prepared using TRITON, a free access software that acts as a graphic user interface for MOPAC and other programs [31]. The calculations were carried out using only a cavity formed by 20 amino acids in the active site (Gln19, Cys22, Gly23, Ser24, Cys25, Trp26, Cys63, Ser64, Gly65, Gly66, Leu67, Met68, Val137, Ala138, Val139, Leu160, Asp161, His162, Gly163, Glu208), as shown in Fig. 1. For these amino acids, the backbone atoms were fixed through all the calculations and all the side chains and ligand atoms were optimized using the BFGS method [32–35].

All the calculations were conducted with MOPAC2009 software for semi-empirical calculations employing the PM6 Hamiltonian for all atoms and the localized molecular orbital method MOZYME [36–40]. A continuum COSMO model with the water dielectric constant of 78.4 was used in all the semi-empirical calculations [41].

3. Results and discussion

3.1. Potential energy surface

The potential energy surface (PES) of the binding recognition of cruzain by a purine-carbonitrile was studied using DFT calculations. A low-energy reaction pathway was found when the inhibitor approaches the thiolate of Cys25 starting a nucleophilic attack. This result is not surprising since the carbon of the nitrile has a charge deficit that enhances the nucleophilic attack from the thiolate. The

PES of this inhibition is shown in Fig. 2, and it can be seen that the reactants must surpass an energetic barrier no higher than 1.39 kcal/mol. The transition state is reached at a C2–S distance of 2.39 Å. In this structure the thiolate interacts with the carbon of the nitrile; this interaction produces an electron migration from the carbon to the *p* orbital of the thiolate making the N1–C2 bond of the nitrile weaker. This fact is reflected in the elongation of the N1–C2 distance up to 1.19 Å at the same time as the triple bond of the nitrile breaks to form a double bond. Interestingly, the interaction between the N1 and H atoms seem to be relatively weak since they are separated by 1.78 Å.

After the transition state, an unusual energy plateau related to the hydrogen transfer from the His162 to the nitrile moiety can be observed. The calculated Fukui function value $f_A^+ = 0.180$ (Table 1) indicates that a negative charge added to the purine-carbonitrile will be located on the nitrogen of the nitrile. An increased charge in the nitrile makes the nitrogen a good nucleophile (f_A^- value of 0.203) thus facilitating the proton transfer from the histidine to the inhibitor. Moreover, in the energy plateau the dihedral angle S–C2–C3–N11 starts to change in order to conjugate the π electrons of the N1–C2 double bond with those of the purine ring. This kind of PES that presents an energy plateau near the transition state has been observed before and it may be consequence of the vibronic interactions that prevail in this region [42]. Vibronic interactions are favored by low-lying excited states that become very important near transition states, where different electronic states may cross. Although the crossing of the two surfaces is prevented by the vibronic interactions which tend to maximize the gap between them, they frequently cause the flattening of the lower one [42]. It is important to mention that the C2–S distance does not change significantly at the plateau and is not short enough to complete the formation of a covalent bond. The structures of the stationary points are depicted in Fig. 3. Some geometrical parameters and the energetics of the reaction are listed in Table 2.

As the reaction coordinate increases, the PES reaches the edge of the energy plateau and from this point on, the PES follows a downhill path up to the final product, a thioimidate, which lies 22.07 kcal/mol below the energy of the reactants. The difference in energy between the reactants and the product is due to two

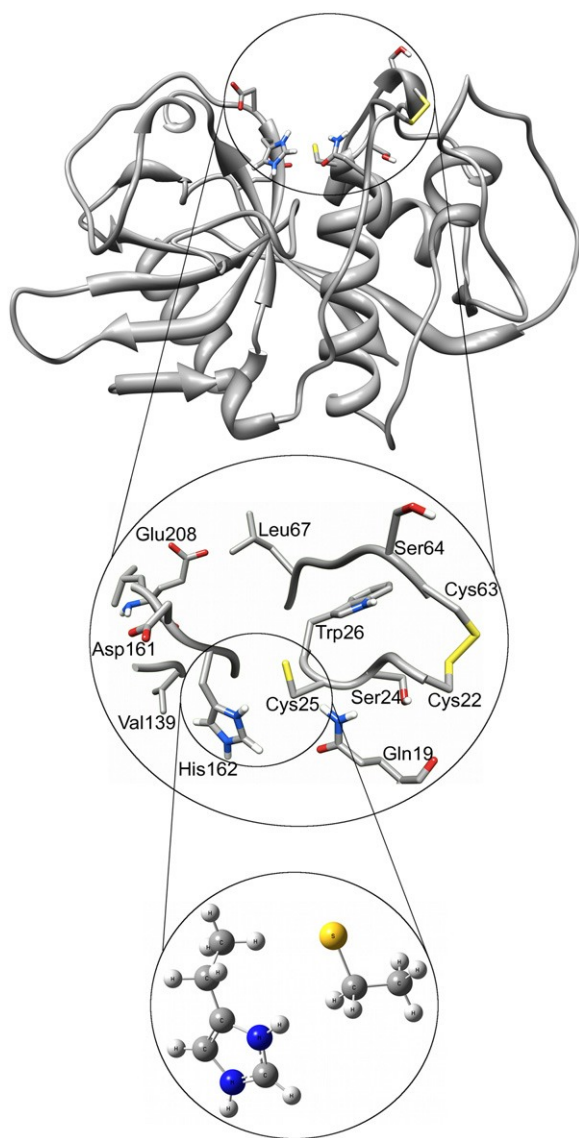


Fig. 1. Crystallographic structure of cruzain (PDB ID: 3I06) used in this study (top). The model used during the semi-empirical calculations is shown in the middle and the one used in DFT calculations is shown at the bottom.

factors: the formation of a chemical bond between the sulfur atom of Cys25 and the carbon on the nitrile, and the stabilization of the bond between the proton and the nitrogen of the ligand. In the resulting thioimidate, the bond lengths of C2–S, N1–C2 and N1–H are 1.81, 1.27, 1.02 Å, respectively; it has a N1–C2–C3 angle of 119.82° and a dihedral S–C2–C3–N11 angle of 153.34°.

It is known that the nitrile-containing molecules form a reversible thioimidate ester adduct between the inhibitor and the active site of the enzyme, in this case, the Cys25 [43]. Based on the PES, we can say that this interaction is reversible at the energy plateau and before the edge is reached; in other words, before the proton transfer from His162 is completed. This is an important observation as it would be interesting to modulate the reversibility of this interaction.

In order to analyze how the substituents on the purine scaffold could affect the reactivity toward the cysteine protease, we computed the stationary points of the co-crystal inhibitor 6-[(3,5-difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile which has an IC_{50} of 0.01 μM against cruzain [19]. Interestingly, neither the relative energy nor the structural parameter differs from those

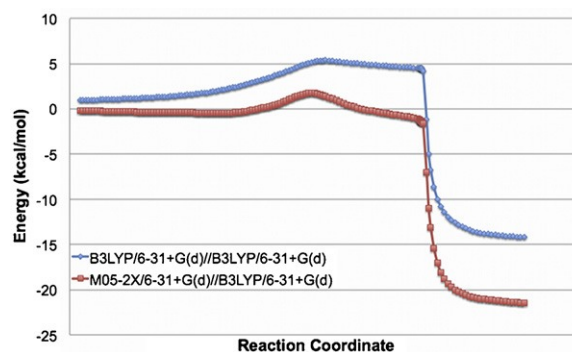


Fig. 2. Potential energy surfaces obtained for the inhibition of cruzain by a purine-carbonitrile.

of the purine-carbonitrile (Table 2). This result suggests that at least these substituents do not have an important effect on the reactivity of the nitrile group; nonetheless, it would be interesting to study inhibitors with electron withdrawing and electron donor groups directly bound to the purine ring and their influence on the reactivity and biological activity.

Using these results as a guideline for drug design, it can be suggested that the reversibility of carbonitrile inhibitors could be regulated by reducing or increasing the negative charge in the nitrile nitrogen. A reduced negative charge on this nitrogen would delay the proton transfer causing a more reversible interaction. On the other hand, an increased negative charge on the nitrogen would facilitate the formation of the thioimidate. Besides, the energetic barrier could be reduced by increasing the electrophilicity of the carbon on the nitrile in order to facilitate the nucleophilic attack from the Cys25 thiolate.

3.2. Molecular orbitals analysis

A useful model to assess and analyze reactivity is the molecular orbital model, especially the frontier orbital approach. This approach is based on the assumption that a bond formation is caused by the flow of electrons from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO).

HOMO and LUMO of the stationary points are depicted in Fig. 4. The HOMO–LUMO energy gap of the reactants is 50.88 kcal/mol, indicating that the reaction may take place with a low energetic barrier. The HOMO of the reactants is localized on the cysteine, mainly at the p orbital of the thiolate, which acts as an electron donor. On the other hand, the LUMO is formed by the p orbital system in the purine. Since these two p orbitals must interact, it is necessary that the nucleophilic attack from the thiolate proceed perpendicularly to the plane that contains the purine ring. It is important to highlight that a stereochemical feature arises because the nucleophile is free to attack from either side of the plane. Nevertheless, the more favorable interaction is at the moment the attack proceeds when the C5 in the purine is pointing to the His162, since this interaction is allowed by orbital symmetry.

Analyzing the HOMO of the next step of the reaction, the transition state, it can be noticed that the p orbitals localized on the purine and the cysteine overlap because of the electron transfer from the S to the C2, and they start the formation of a chemical bond. A detail that stands out is the p orbital of the nitrile nitrogen, which points to the histidine hydrogen but is overlapped with the s orbital of the cysteine side chain hydrogen. An issue that must be addressed is that the nitrile is bonded to the purine ring by the interaction of an s orbital.

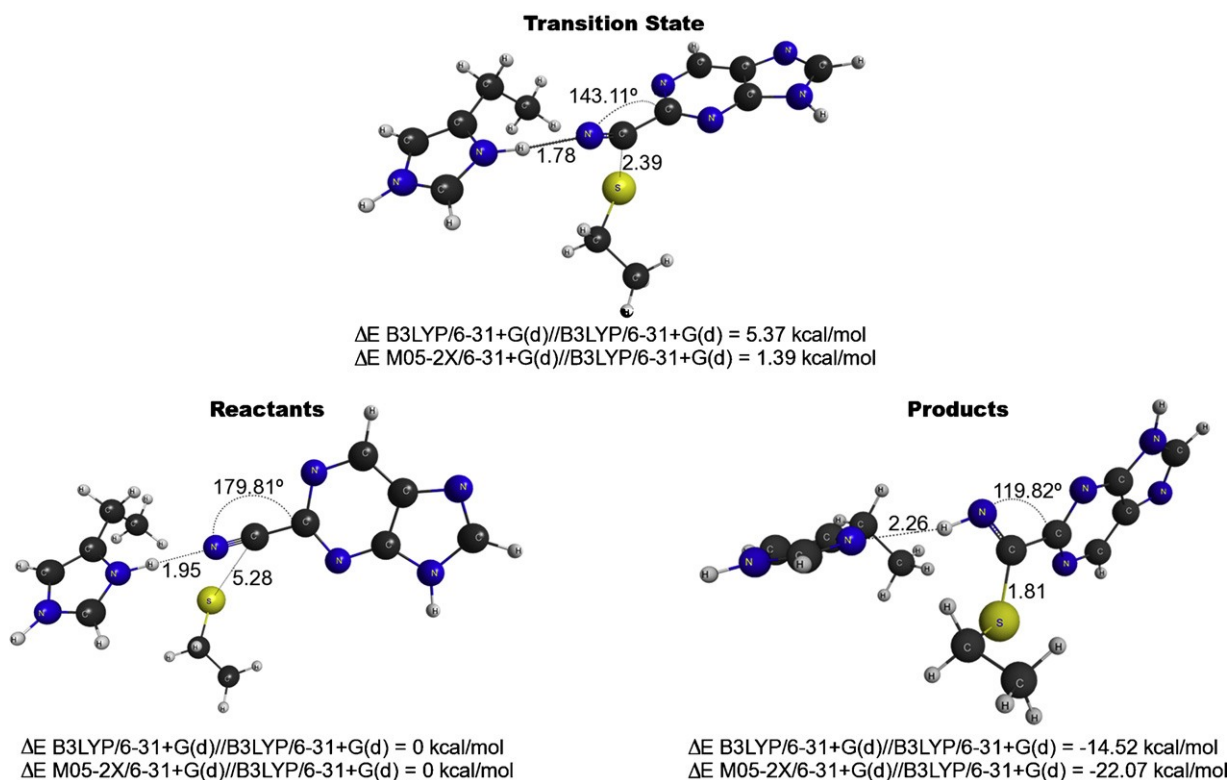


Fig. 3. Optimized structures of the stationary points at B3LYP/6-31+G(d) level of theory. The distances and bond lengths are indicated in Å.

Finally, the products of the reaction are characterized by a large HOMO–LUMO energy gap of 97.72 kcal/mol, which is related to the high stability of the products. Although the frontier orbitals of the products are not as interesting as those of the reactants or transition state, it can be seen in Fig. 4 that these HOMO and LUMO cannot react with each other, favoring the stability of the products. The HOMO of the products is located mainly in the *p* orbital system of the imidazole in the histidine, while the LUMO is composed by antibonding *p* orbitals in the purine.

3.3. Reaction inside the active site

Since it is possible that other amino acids, different from the catalytic ones, could take part in the inhibition reaction by stabilizing the transition state or products, in this section the role of other amino acids in the catalytic site during the molecular interaction was studied. For this purpose, quantum

semi-empirical calculations were carried out in order to analyze possible interactions between the inhibitor and the active site of the enzyme during the reaction; specifically, the PES of the cruzain inhibition by 6-[(3,5-difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile was mapped. Two reaction coordinates were followed inside a cavity of 20 amino acids of the cruzain catalytic active site: the C2–S distance between thiolate and the nitrile; and the N1–H distance between the nitrile and histidine. The 2D PES obtained from these calculations is depicted in Fig. 5. It can be seen that the lowest energy pathway is the one in which the S and the C2 approach in such a way that seems to be a nucleophilic attack. Once the S and the C2 are close enough, a proton migration from the histidine to the nitrogen nitrile occurs so that the final products may be reached. On the other hand, no energetically favored pathway was found for the process in which the protonation of the nitrile takes place before the nucleophilic attack from the thiolate; in fact, this is a very disfavored process, as shown in Fig. 5.

Table 2

Selected geometrical parameters and energetics for the stationary points in the inhibition of cruzain.

Parameters	Purine-carbonitrile			6-[(3,5-Difluorophenyl)amino]-9-ethyl-9H-purine-2-carbonitrile		
	Reactants	Transition state	Products	Reactants	Transition state	Products
$r(\text{N1}-\text{C2})$	1.1601	1.1920	1.2723	1.1602	1.1926	1.2727
$r(\text{N1}-\text{H})$	1.9514	1.7763	1.0279	1.9441	1.7552	1.0266
$r(\text{N1}-\text{N}_{\text{His}})$	2.9749	2.8221	3.1898	2.9705	2.8063	3.2256
$r(\text{C2}-\text{C3})$	1.4492	1.4742	1.5048	1.4509	1.4774	1.5075
$r(\text{C2}-\text{S})$	5.2787	2.3900	1.8143	5.1646	2.3829	1.8128
$r(\text{N}_{\text{His}}-\text{H})$	1.0276	1.0488	2.2601	1.0284	1.0531	2.3212
$\theta(\text{N1}-\text{C2}-\text{C3})$	179.811	143.108	119.817	179.624	142.394	119.885
$\theta(\text{N1}-\text{C2}-\text{S})$	52.949	118.651	127.363	56.049	118.496	127.361
$\theta(\text{H}-\text{N1}-\text{C2})$	178.611	127.319	113.715	178.490	130.261	113.607
$D(\text{S}-\text{C2}-\text{C3}-\text{N11})$	-62.592	-108.634	-153.342	-73.239	-109.024	-154.944
$D(\text{S}-\text{C2}-\text{N1}-\text{H})$	162.889	-36.166	-5.877	-13.455	-25.275	-5.609
$E(\text{B3LYP/6-31+G(d)//B3LYP/6-31+G(d)})$	0.00	5.37	-14.52	0.00	6.27	-13.94
$E(\text{M05-2X/6-31+G(d)//B3LYP/6-31+G(d)})$	0.00	1.39	-22.07	0.00	1.50	-22.42

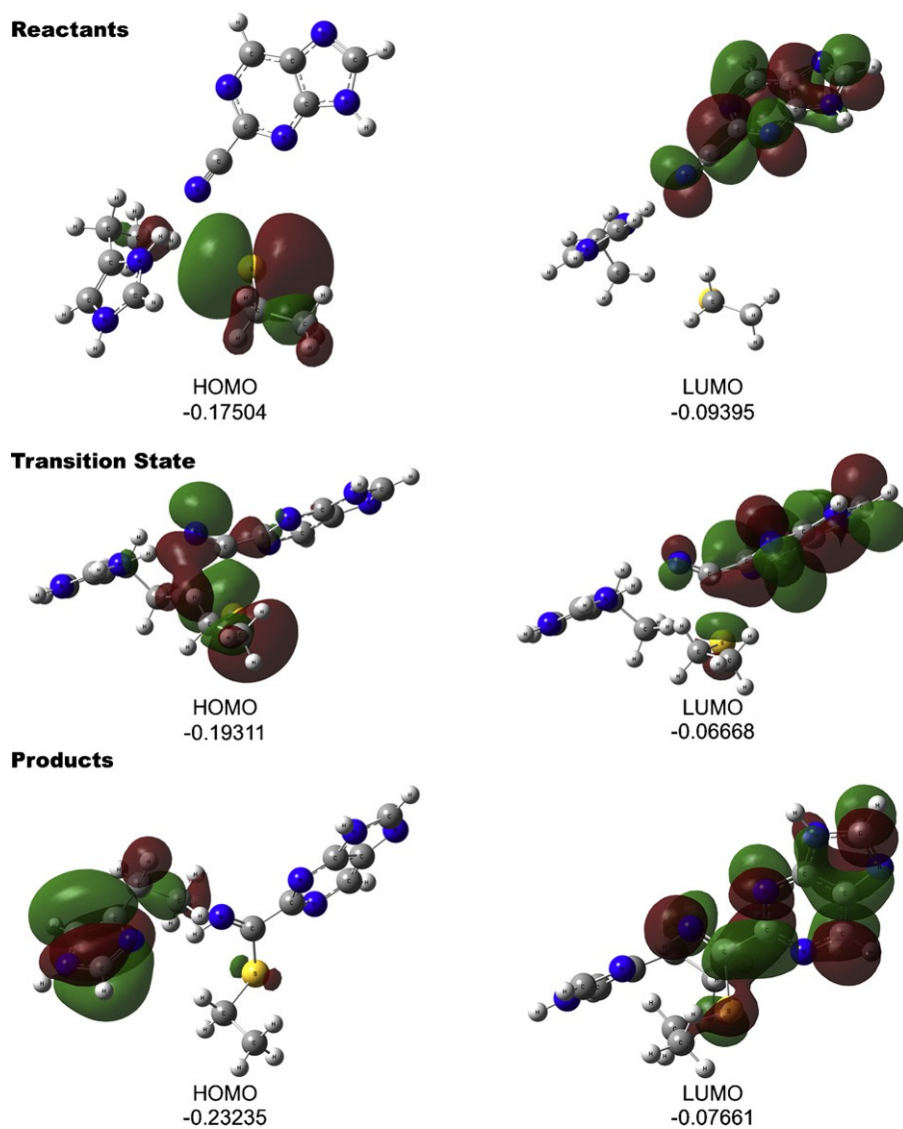


Fig. 4. B3LYP frontier molecular orbitals and orbital energies (in atomic units) of the stationary points.

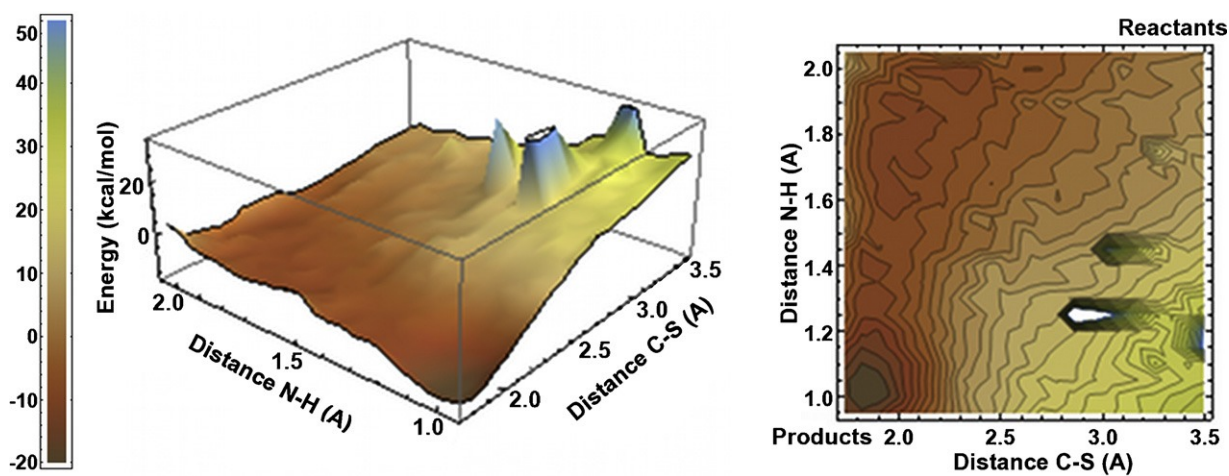


Fig. 5. 2D potential energy surface for the inhibition of cruzain computed with the PM6 Hamiltonian.

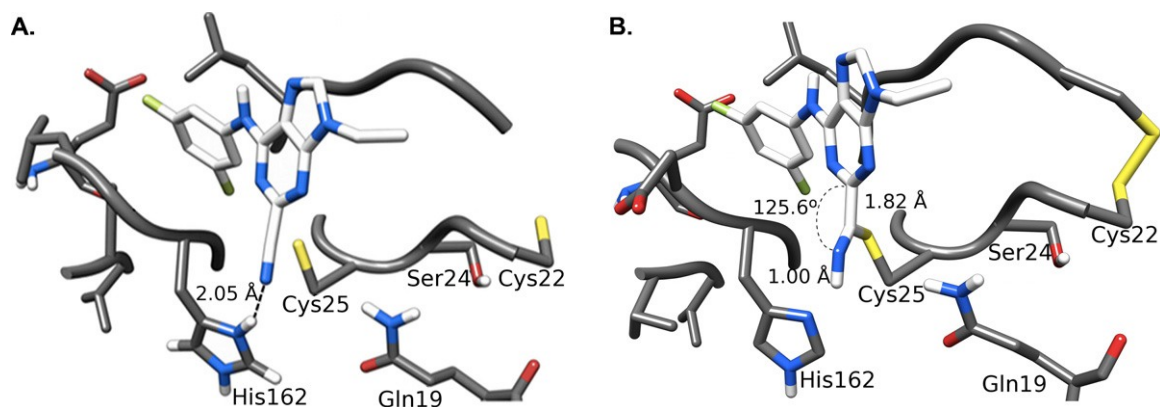


Fig. 6. Optimized structures of reactants (A) and products (B) employing the PM6 Hamiltonian.

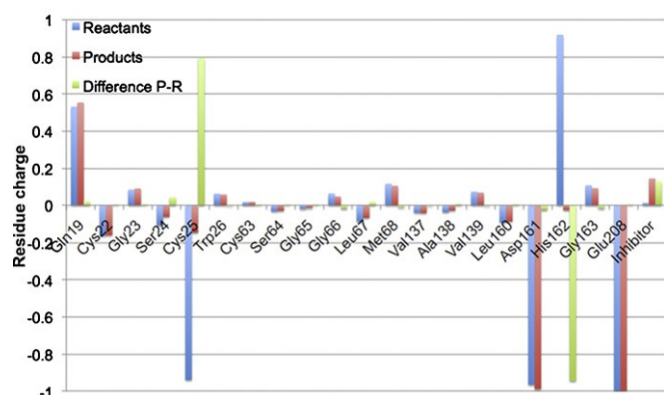


Fig. 7. Histogram showing the charge variation of 20 amino acids in the active site during the reaction.

The results obtained using quantum semi-empirical methods are in good agreement with the DFT calculations. This fact is reflected by the resulting reaction pathway for this process and by the relative energy of the products, which is 17.07 kcal/mol below the reactants. Also, the geometrical parameters are very similar to those of the structures optimized using DFT; for example, the C2–S, N1–C2 and N1–H bond lengths are 1.82, 1.26, 1.00 Å, respectively; the N1–C2–C3 angle is 125.60° and the dihedral S–C2–C3–N1 angle is 134.50° (Fig. 6).

The variation of the total Mulliken charge on the amino acid residue was used to assess possible interactions during the reaction. The residue charge was calculated as the sum of partial charges from all its atoms. The charge variation between reactants and products is shown in Fig. 7. It can be seen that most of the residues do not present any significant charge variation. Nonetheless, as was expected, the two amino acids that take part in the reaction (Cys25 and His162) presented a charge variation due to the proton transfer. A fact that must be highlighted is that Gln19, which acts as part of an oxyanion hole, has a low variation of the positive charge during the reaction. This could be indicative of the fact that the Gln19 does not play a major role in the stabilization of the inhibitor.

4. Conclusions and perspectives

A theoretical study, using DFT and quantum semi-empirical calculations, was carried out to reveal the mechanism of the inhibition of cysteine proteases by nitriles. This paper shows evidence suggesting that the inhibition occurs via a mechanism which starts with a nucleophilic attack from the Cys25 to the inhibitor followed by a proton transfer from the His162. The reaction presents a very

particular PES that consists of one transition state and an energy plateau. These results are in good agreement with the proposed peptide hydrolysis mechanism of papain, another cysteine protease, in which the peptide hydrolysis occurs in a concerted manner [8–10]. On the other hand, no intermediates were found as had been in the models for the inhibition of cysteine proteases by diketones, halomethyl ketones, diazomethyl ketones and aziridines [11–15].

Results presented in this paper suggest that the reversibility of this interaction could be managed by modulating the proton transfer from the His162 to the inhibitor. Additionally, an increased electrophilicity in the carbon on the nitrile might facilitate the nucleophilic attack from Cys25; in this way, the activation energy would be reduced at the same time as the inhibition activity would increase. This study presents information that is helpful for a better understanding of the mechanism of inhibition of cruzain by nitriles. The data obtained is useful for the design of better cysteine protease inhibitors that could lead to new trypanocidal agents. Future work is focused on studying the effect of different substituents in the purine scaffold on the energy profile of the reaction. A second perspective is to investigate how the reactivity and reversibility of the inhibitors could be modulated in order to design safe and effective drugs against Chagas' disease.

Acknowledgments

Authors are grateful to Karina Martinez-Mayorga for helpful discussions. This research was supported by CONACyT project No. 80093. O. M-L and A. R-M are very grateful to CONACyT for the fellowships granted (Nos. 245408 and 173861, respectively). J. L. M-F thanks the State of Florida for funding.

References

- [1] H.H. Otto, T. Schirmeister, Cysteine proteases and their inhibitors, *Chem. Rev.* 97 (1997) 133–172.
- [2] F. Lecaille, J. Kaleta, D. Brömme, Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design, *Chem. Rev.* 102 (2002) 4459–4488.
- [3] N.D. Rawlings, A.J. Barrett, Families of cysteine peptidases, *Methods Enzymol.* 244 (1994) 461–486.
- [4] M. McGrath, A. Eakin, J. Engel, J. McKerrow, C. Craik, R. Fletterick, The crystal structure of cruzain: a therapeutic target for Chagas' disease, *J. Mol. Biol.* 247 (1995) 251–259.
- [5] J. Nishihira, H. Tachikawa, Theoretical evaluation of a model of the catalytic triads of serine and cysteine proteases by ab initio molecular orbital calculation, *J. Theor. Biol.* 196 (1999) 513–519.
- [6] E. Dufour, A. Storer, R. Menard, Peptide aldehydes and nitriles as transition state analog inhibitors of cysteine proteases, *Biochemistry* 34 (1995) 9136–9143.
- [7] J. Drenth, K.H. Kalk, H.M. Swen, Binding of chloromethyl ketone substrate analogues to crystalline papain, *Biochemistry* (1976) 1–8.
- [8] D. Arad, R. Langridge, P. Kollman, A simulation of the sulfur attack in catalytic pathway of papain using molecular mechanics and semiempirical quantum mechanics, *J. Am. Chem. Soc.* 112 (1990) 491–502.

- [9] W. Welsh, Y. Lin, Discussion of the catalytic pathway of cysteine proteases based on AM1 calculations, *J. Mol. Struct.: Theochem.* 401 (1997) 315–326.
- [10] M. Harrison, N. Burton, I. Hillier, Catalytic mechanism of the enzyme papain: predictions with a hybrid quantum mechanical/molecular mechanical potential, *J. Am. Chem. Soc.* 119 (1997) 12285–12291.
- [11] M. Tarnowska, S. Oldziej, A. Liwo, P. Kania, F. Kasprzykowski, Z. Grzonka, MNDO study of the mechanism of the inhibition of cysteine proteinases by diazomethyl ketones, *Eur. Biophys. J.* 21 (1992) 217–222.
- [12] R. Shankar, P. Kolandaivel, K. Senthilkumar, Reaction mechanism of cysteine proteases model compound HSH with diketone inhibitor PhCOCOCH₃-*n*Xn (X = F, Cl, *n* = 0, 1, 2), *Int. J. Quantum Chem.* 110 (2010) 1660–1674.
- [13] S. Vijayakumar, P. Kolandaivel, Reaction mechanism of HSH and CH₃SH with NH₂CH₂COCH₂X (X = F and Cl) molecules, *Int. J. Quantum Chem.* 108 (2008) 927–936.
- [14] G. Barreiro, R. De Alencastro, J. Neto, A semiempirical study on leupeptin: an inhibitor of cysteine proteases, *Int. J. Quantum Chem.* 65 (1997) 1125–1134.
- [15] R. Vicik, H. Helten, T. Schirmeister, B. Engels, Rational design of aziridine-containing cysteine protease inhibitors with improved potency: studies on inhibition mechanism, *ChemMedChem* 1 (2006) 1021–1028.
- [16] M.K. Lindvall, Molecular modeling in cysteine protease inhibitor design, *Curr. Pharm. Des.* 8 (2002) 1673–1681.
- [17] J. Cazzulo, V. Stoka, V. Turk, The major cysteine proteinase of *Trypanosoma cruzi*: a valid target for chemotherapy of Chagas disease, *Curr. Pharm. Des.* 7 (2001) 1143–1156.
- [18] J.C. Engel, P.S. Doyle, I. Hsieh, J.H. McKerrow, Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection, *J. Exp. Med.* 188 (1998) 725–734.
- [19] B.T. Mott, R.S. Ferreira, A. Simeonov, A. Jadhav, K.K.-H. Ang, W. Leister, et al., Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB, *J. Med. Chem.* 53 (2010) 52–60.
- [20] O. Méndez-Lucio, J. Perez-Villanueva, A. Romo-Mancillas, R. Castillo, 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), *Med. Chem. Commun.* 2 (2011) 1058–1065.
- [21] A. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* 98 (1993) 5648.
- [22] C. Lee, W. Yang, R.G. Parr, Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* 37 (1988) 785.
- [23] L. Simon, J.M. Goodman, How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals, *Org. Biomol. Chem.* 9 (2011) 689–700.
- [24] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, et al., Gaussian 09, Revision A.1, Gaussian, Inc., Wallingford, CT, 2009.
- [25] R.G. Parr, W. Yang, Density functional approach to the frontier-electron theory of chemical reactivity, *J. Am. Chem. Soc.* 106 (1984) 4049–4050.
- [26] J.L. Gázquez, Perspectives on the density functional theory of chemical reactivity, *J. Mex. Chem. Soc.* 52 (2008) 3–10.
- [27] M. Boháč, Y. Nagata, Z. Prokop, M. Prokop, M. Monincová, M. Tsuda, et al., Halide-stabilizing residues of haloalkane dehalogenases studied by quantum mechanic calculations and site-directed mutagenesis, *Biochemistry* 41 (2002) 14272–14280.
- [28] E. Dyguda, B. Szefczyk, W. Sokalski, The mechanism of phosphoryl transfer reaction and the role of active site residues on the basis of ribokinase-like kinases, *Int. J. Mol. Sci.* 5 (2004) 141–153.
- [29] P. Banáš, M. Otyepka, P. Jeřábek, M. Petřek, J. Damborský, Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling, *J. Comput.-Aided Mol. Des.* 20 (2006) 375–383.
- [30] H. Yu, J. Liu, Z. Shen, Modeling catalytic mechanism of nitrile hydratase by semiempirical quantum mechanical calculation, *J. Mol. Graphics Modell.* 27 (2008) 522–528.
- [31] M. Prokop, J. Adam, Z. Kriz, M. Wimmerova, J. Koca, TRITON: a graphical tool for ligand-binding protein engineering, *Bioinformatics* 24 (2008) 1955–1956.
- [32] C.G. Broyden, The convergence of a class of double-rank minimization algorithms, *J. Inst. Math. Appl.* 6 (1970) 222–231.
- [33] R. Fletcher, A new approach to variable metric algorithms, *Comput. J.* 13 (1970) 317–322.
- [34] D. Goldfarb, A family of variable-metric methods derived by variational means, *Math. Comput.* 24 (1970) 23–26.
- [35] D.F. Shanno, Conditioning of Quasi-Newton methods for function minimization, *Math. Comput.* 24 (1970) 647–656.
- [36] MOPAC2009, James J.P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, <http://OpenMOPAC.net>, 2008.
- [37] J.J.P. Stewart, Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* 13 (2007) 1173–1213.
- [38] J.J.P. Stewart, Application of the PM6 method to modeling proteins, *J. Mol. Model.* 15 (2009) 765–805.
- [39] J.J.P. Stewart, Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations, *Int. J. Quantum Chem.* 58 (1996) 133–146.
- [40] J.J.P. Stewart, Calculation of the geometry of a small protein using semiempirical methods, *J. Mol. Struct.: Theochem.* 401 (1997) 195–205.
- [41] A. Klamt, G. Schuurmann, COSMO – a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin Trans. 2* (1993) 799–805.
- [42] G.N. Sastry, T. Bally, V. Hroudá, P. Cársky, The C₄H₆^{••} potential energy surface. 1. The ring-opening reaction of cyclobutene radical cation and related rearrangements, *J. Am. Chem. Soc.* 120 (1998) 9323–9334.
- [43] J. Moon, R. Coleman, R. Hanzlik, Reversible covalent inhibition of papain by a peptide nitrile Carbon-13 NMR evidence for a thioimidate ester adduct, *J. Am. Chem. Soc.* 108 (1986) 1350–1351.



Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Activity landscape modeling of PPAR ligands with dual-activity difference maps

Oscar Méndez-Lucio^a, Jaime Pérez-Villanueva^b, Rafael Castillo^{a,*}, José L. Medina-Franco^{c,*}

^a Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, México DF 04510, Mexico

^b Departamento de Sistemas Biológicos, División de Ciencias Biológicas y de la Salud, UAM-X, México DF 04960, Mexico

^c Torrey Pines Institute of Molecular Studies, 11350 SW Village Parkway, Port St. Lucie, FL 34987, USA

ARTICLE INFO

Article history:

Received 15 February 2012

Revised 27 March 2012

Accepted 4 April 2012

Available online 19 April 2012

Keywords:

Activity cliffs

Diabetes mellitus

Dual-activity difference maps

PPAR agonist

Structure–activity relationships

ABSTRACT

Activation of peroxisome proliferator-activated receptor (PPAR) subtypes offers a promising strategy for the treatment of diabetes mellitus and metabolic diseases. Selective and dual PPAR agonists have been developed and the systematic characterization of their structure–activity relationships (SAR) is of major significance. Herein, we report a systematic description of the SAR of 168 compounds screened against the three PPAR subtypes using the principles of activity landscape modeling. As part of our effort to develop and apply chemoinformatic tools to navigate through activity landscapes, we employed consensus dual-activity difference maps recently reported. The analysis is based on pairwise relationships of potency difference and structure-similarity which were calculated from the combination of four different 2D and 3D structure representations. Dual-activity difference maps uncovered regions in the landscape with similar SAR for two or three receptor subtypes as well as regions with inverse SAR, that is, changes in structure that increase activity for one subtype but decrease activity for the other subtype. Analysis of pairs of compounds with high structure similarity revealed the presence of single-, dual-, and 'pan-receptor' activity cliffs, that is, small changes in structure with high changes in potency for one, two, or three receptor subtypes, respectively. Single-, dual-, and pan-receptor scaffold hops are also discussed. The analysis of the chemical structures of selected data points reported in this paper points to specific structural features that are helpful for the design of new PPAR agonists. The approach presented in this work is general and can be extended to analyze larger data sets.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Diabetes mellitus (DM) is a global health problem. Estimates indicate that in 2010 there were 285 million people worldwide with diabetes and that in 2030 there will be 439 million adults with diabetes.¹ Different molecular targets have been proposed in order to modulate important aspects of DM pathogenesis. Specially, the activation of peroxisome proliferator-activated receptors (PPARs) offers a promising strategy for the treatment of DM and metabolic diseases.²

PPARs are nuclear lipid-activated transcription factors that have been identified as major regulators in glucose and lipid metabolism; thereby they contribute significantly to some disorders such as diabetes, obesity and cardiovascular complications.³ The expression of three different PPAR subtypes, PPAR α , PPAR δ and PPAR γ has

been described.^{3,4} Each subtype shows a different tissue distribution pattern, and pharmacological response.³ For example, PPAR α is amply expressed in the liver, kidney, heart, and skeletal muscle. Experimental data indicates that this subtype has a critical role in the uptake and oxidation of fatty acids and also in lipoprotein metabolism. Activation of PPAR α decreases serum triglycerides and increases serum high-density lipoprotein cholesterol, which together restores lipid and glucose homeostasis. PPAR γ is predominantly expressed in adipose tissue and is a key modulator of adipocyte differentiation. Activation of PPAR γ improves glycemic control by increasing insulin sensitivity. PPAR δ also seems to play an important role in the regulation of lipid metabolism and cholesterol efflux.^{5–7} Selective, dual and pan PPAR agonists have been studied as potential therapy against diabetes and the metabolic syndrome.^{3,5,6} Progress in the development of some dual PPAR α / γ agonists has been hampered because of safety issues during phase II and phase III clinical trials.⁸ Therefore, the systematic study of the structural requirements to determine selective compounds toward one PPAR subtype or for the design of new dual PPAR agonists is of great significance.^{5–7} Computational studies, such as pharmacophore modeling^{9,10} have been reported directed towards the rationalization of the SAR of PPAR ligands.

Abbreviations: DAD, dual activity difference; DM, diabetes mellitus; EStateIndices, electrotopological state indices; PPAR, peroxisome-proliferator-activated receptor; ROCS, Rapid Overlay of Chemical Structures; SAR, structure–activity relationships; TAD, triple activity difference.

* Corresponding authors. Tel.: +1 772 345 4685; fax: +1 772 345 3649 (J.L.M.-F.).

E-mail address: jmedina@tpims.org (J.L. Medina-Franco).

Systematic description of the structure–activity relationships (SAR) of data sets can be conveniently achieved using the principles and methods of activity landscape modeling.^{11–13} The activity landscape has been defined as ‘any representation that integrates the analysis of the structural similarity and potency differences between compounds sharing the same biological activity’.¹² The activity landscape has been compared to rolling hills or continuous SAR where small changes in molecular structure are associated with small changes in activity.¹⁴ A discontinuous SAR or rugged activity landscape, however, is populated with molecules with small changes in structure but large changes in activity, that is, ‘activity cliffs’ (chemical compounds with highly similar structures but significantly different biological activities).¹⁴ Understanding the activity landscape and early detection of activity cliffs can be crucial to the success of predictive computational models such as QSAR.^{15–17} However, understanding the activity landscape of data sets tested with one or multiple biological endpoints is not straightforward because the landscape can be highly complex with a combination of smooth and rugged regions.¹⁸ Another major challenge is the dependence of chemical space on molecular representation.^{19,20}

Herein, we report a systematic characterization of the SAR of 168 compounds screened against the three PPAR subtypes. The analysis of the SAR was based on pairwise comparisons of the activity similarity and structure similarity. The activity landscape across multiple PPAR subtypes, that is, ‘multi-target activity landscape modeling’,^{21,22} was characterized using the dual and triple activity difference (DAD/TAD) maps we recently developed.^{23,24} As pointed out by Peltason and co-workers, multi-target activity landscape modeling is attractive to characterize the SAR of compound data sets associated with selectivity²² or promiscuity, for example, ‘scaffold-based promiscuity’ of the targets. Single-, dual-, and triple-receptor subtype activity cliffs as well as scaffold hops (chemical compounds with highly dissimilar structures but significantly similar biological activities)²⁵ were identified.

2. Methods

2.1. Data set

For this study, a set of 168 unique PPAR agonists was obtained from the Binding Database,^{26,27} which is a public repository of chemical compounds annotated with biological activity well suited to perform SAR studies.²⁸ All the 168 compounds have IC₅₀ values reported against the three receptor subtypes. The SMILES representation of the structures and the pIC₅₀ (–logIC₅₀) values are listed in the [Supplementary data \(Table S1\)](#). For PPAR α , the pIC₅₀ values range from 4.36 to 8.32, for PPAR δ from 3.40 to 8.76 and for PPAR γ from 4.39 to 8.30. It is noteworthy that the activity range for the three subtypes is very similar, thus facilitating the cross-comparisons of the activity landscapes.²⁹

2.2. Structure similarity

Eleven different 2D fingerprint representations were calculated for each molecule using the open-source software MayaChemTools.³⁰ This software is a set of Perl scripts, modules and classes for the generation of fingerprints corresponding to atom neighborhoods,³¹ atom types, electrotopological state indices (EStateIndices),³² extended connectivity,³³ MACCS (322 bits),³⁴ path length, topological atom pairs (TopAtomPairs),³⁵ topological atom torsions (TopAtomTorsions),³⁶ topological atom triplets (TopAtomTriplets), topological pharmacophore atom pairs (TopPh4Pairs),³⁷ and topological pharmacophore atom triplets (TopPh4Triplets).³⁸ All of them are widely used and validated molecular descriptors.

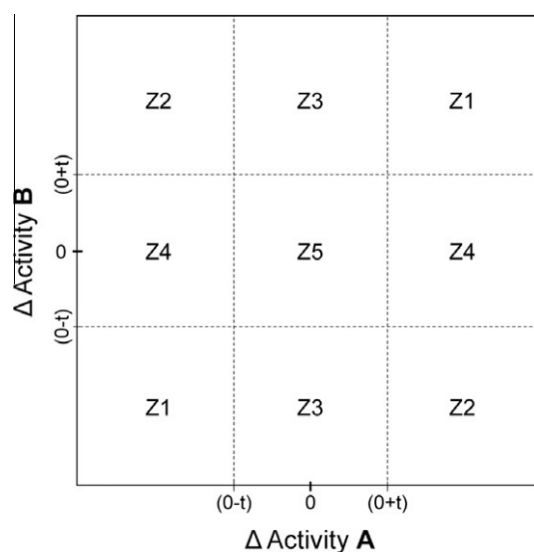


Figure 1. Prototype dual activity-difference (DAD) map divided in five major regions labeled as Z1–Z5. Region Z1 contains pairs in which structural changes affect in similar proportion the activity against both receptors. Pairs in Z2 are those in which the structural change increases the activity for one target, but decreases the activity toward the other (inverse SAR). Pairs in which a structural difference does not affect the activity against one target, but it does against the other, are located in region Z3 or Z4. Region Z5 is associated with pairs with similar activities against both targets.

Additionally, ComboScore and ShapeTanimoto, as implemented in Rapid Overlay of Chemical Structures (ROCS) (OpenEye Scientific Software) were computed as 3D fingerprints using a single low-energy conformation calculated with the MMFF94x force field. The Combo Score was scaled by dividing the 3D similarity values by two, in order to obtain a number within the same scale as the other computed similarity measures. Recent works have shown that the use of 3D fingerprints are important in characterizing activity landscapes, even if the use of a single conformer to represent 3D structures is an approximation of the real active conformation.^{39–41} However, activity landscape modeling can also be carried out using multiple 3D conformers as we have previously shown.²⁹ Pairwise similarity was calculated using the Tanimoto coefficient^{42,43} but other similarity measures can be used. For each fingerprint representation, 14,028 pairwise similarities were calculated for the 168 compounds. Similarity values calculated with EStateIndices, TopAtomPairs, TopPh4Triplets, and ComboScore were combined through mean fusion, that is, computing the mean similarity of the four selected fingerprints (see below). Other fusion rules can be applied.

2.3. Pairwise activity relationships

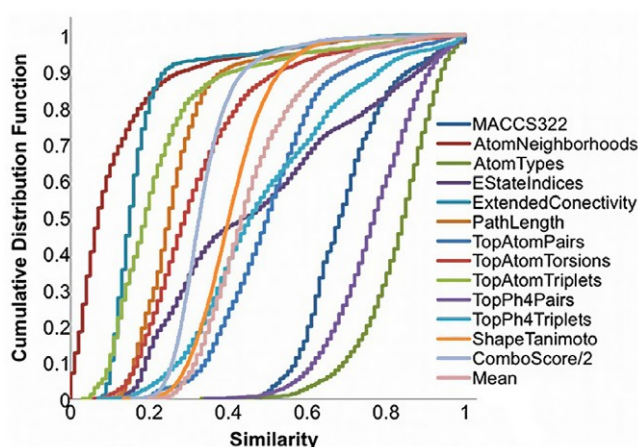
For each pair of PPAR agonists, potency differences were calculated using the expression:

$$\Delta \text{pIC}_{50}(\text{R})_{ij} = \text{pIC}_{50}(\text{R})_i - \text{pIC}_{50}(\text{R})_j$$

where pIC₅₀(R)_i and pIC₅₀(R)_j are the activities of the *i*th and *j*th molecules (*j* > *i*) against receptor subtype R. In this paper R = PPAR α , δ , or γ . The positive or negative values of the expression above provided information concerning the direction of SAR.²⁴

2.4. Dual- and triple-receptor activity landscapes

We recently developed DAD maps to characterize activity landscapes of data sets with biological activity for two biological endpoints.^{23,44} These maps are based on pairwise activity differences



	Max	Q3	Median	Q1	Min	Mean	STD
MACCS322	1.00	0.76	0.68	0.62	0.45	0.70	0.11
Atom Neighborhoods	1.00	0.15	0.07	0.03	0.00	0.12	0.15
Atom Types	1.00	0.90	0.84	0.77	0.33	0.83	0.10
EStateIndices	1.00	0.68	0.45	0.26	0.06	0.49	0.25
ExtendedConnectivity	1.00	0.19	0.15	0.12	0.07	0.18	0.12
PathLength	1.00	0.31	0.25	0.20	0.10	0.28	0.14
TopAtomPairs	1.00	0.58	0.50	0.41	0.13	0.51	0.15
TopAtomTorsions	1.00	0.39	0.29	0.22	0.05	0.33	0.16
TopAtomTriplets	1.00	0.26	0.19	0.13	0.03	0.23	0.16
TopPh4Pairs	1.00	0.85	0.77	0.69	0.34	0.77	0.11
TopPh4Triplets	1.00	0.63	0.46	0.35	0.07	0.50	0.19
ShapeTanimoto	1.00	0.46	0.40	0.35	0.13	0.41	0.10
ComboScore/2	1.00	0.37	0.33	0.29	0.16	0.34	0.09
Mean	1.00	0.52	0.43	0.37	0.20	0.46	0.13

Figure 2. Cumulative distribution functions (CDF) of 14,028 pairwise structural similarities using different 2D and 3D structure representations. The table summarizes the information of the distributions. Q3 and Q1 indicate the third and first quartile, respectively.

against two receptors and are built plotting the activity difference for one receptor against the activity difference for the second receptor as depicted in the prototype DAD map in Figure 1. Vertical and horizontal lines at $\Delta IC_{50} \pm t$ determine boundaries for low/high potency difference for receptors **A** and **B**, respectively. Herein, we set $t = 1$ (1 log unit) so that data points were considered with low potency difference if $-1 \leq \Delta pIC_{50} \leq 1$ for each receptor. The boundaries give rise to five general zones, Z1–Z5.⁴⁴ If the structural changes between two compounds in the pair affect in similar proportion the biological activity against both **A** and **B** receptors (either an increase or decrease), this pair will be located in Z1. Pairs of compounds located in Z2 are those in which the structural change increases the activity toward one receptor, but decreases the activity toward the other; therefore Z2 is associated with an ‘inverse SAR’ and is the most important zone for selectivity studies. Pairs of compounds located in Z3 or Z4 denote compounds in which structural differences do not affect the activity against one receptor, but do affect the activity against the other. Data points in Z5 are associated with pairs of compounds for which structural changes have little or no impact on the activity against the two receptors. In general, the most informative data points associated with receptor selectivity are those located at the outer regions of zones Z2, Z3, and Z4. The classification of data points in Figure 1 is independent of the structure similarity. However, since DAD maps are based on pairwise comparisons, it is straightforward to incorporate structure similarity information.

Here we represent structure similarity using a continuous color scale from green (low similarity) to red (high similarity). Mapping structure similarity information into the DAD maps easily reveals the presence of single-, dual-, or triple-receptor activity cliffs as well as regions with continuous SAR for one or several receptors.⁴⁴ DAD maps can be extended to triple activity-difference (TAD) maps by adding a third dimension that represents the activity difference for a third receptor. TAD maps can be interpreted as the grouping of three DAD maps for each combination of two receptors.⁴⁴

3. Results and discussion

3.1. Distribution of similarity measures

Figure 2 summarizes the 14,028 pairwise similarities calculated with 13 2D and 3D structure representations as cumulative distribution functions (CDF). The table at the bottom of the figure summarizes the statistics of each curve indicating the maximum, third and first quartile, median, mean, and standard deviation. 2D and 3D representations showed a wide variation of distributions. Similar results have been obtained for other data sets.^{29,39,40} Atom Types, TopPh4Pairs and MACCS (322 bits) showed the highest similarity values with median values of 0.84, 0.77, and 0.68, respectively. Fingerprints with the lowest similarity values were atom neighborhoods, extended connectivity, and topological atom

Table 1
Correlation matrix between molecular representations

	AtomNeighbors	AtomTypes	EStateIndices	ExtendedConnectivity	MACCS322	PathLength	TopAtomPairs	TopAtomTorsions	TopAtomTriplets	TopPh4Pairs	TopPh4Triplets	ShapeTanimoto	ComboScore/2
AtomNeighbors	1.000												
AtomTypes	0.491	1.000											
EStateIndices	0.556	0.447	1.000										
ExtendedConnectivity	0.896	0.413	0.472	1.000									
MACCS322	0.623	0.365	0.479	0.642	1.000								
PathLength	0.800	0.350	0.456	0.901	0.704	1.000							
TopAtomPairs	0.751	0.804	0.503	0.709	0.612	0.659	1.000						
TopAtomTorsions	0.859	0.601	0.523	0.791	0.583	0.726	0.824	1.000					
TopAtomTriplets	0.865	0.587	0.497	0.845	0.639	0.791	0.920	0.865	1.000				
TopPh4Pairs	0.511	0.415	0.347	0.471	0.527	0.442	0.611	0.474	0.572	1.000			
TopPh4Triplets	0.527	0.383	0.363	0.591	0.435	0.617	0.545	0.496	0.553	0.579	1.000		
ShapeTanimoto	0.310	0.126	0.202	0.334	0.218	0.304	0.281	0.313	0.369	0.159	0.142	1.000	
ComboScore/2	0.480	0.252	0.290	0.497	0.359	0.456	0.455	0.466	0.548	0.285	0.270	0.820	1.000

triplets with median values of 0.07, 0.15, and 0.19, respectively. Several fingerprints had a normal distribution as revealed by the sigmoidal shape of the corresponding curves and the similar median and mean values.

Table 1 shows the correlation matrix of the Pearson's correlation coefficient between the 14,028 pairwise similarities for each pair of the 13 2D and 3D representations. High correlations between representations occur for topological atom pairs and topological atom triplets; atom neighborhoods and extended connectivity; path length and extended connectivity; topological atom triplets and topological atom torsions; topological atom triplets and atom neighborhoods; topological atom torsions and atom neighborhoods (correlation >0.85). EStateIndices showed low correlation (<0.55) with all other representations. This result is explained because EStateIndices is the only fingerprint used in this study that encodes the intrinsic electronic state of the atom as perturbed by the electronic influence of all other atoms in the molecule.³² In contrast, other fingerprints with similar design showed high correlations, for example, topological atom pairs and topological atom triplets (correlation of 0.92).

3.2. Selection of structure representations for consensus models

In order to reduce the dependence of structure representation with chemical space, we combined similarity measures obtained by different methods using the principles of data fusion. In this work we computed the mean similarity, that is, mean fusion, of four selected representations. Our group has applied this approach to derive consensus models of the activity landscape of other data sets.^{29,40,44} We computed the mean similarity of EStateIndices, topological atom pairs, topological pharmacophore triplets, and Combo Score (scaled). Following a similar approach recently reported,^{29,45} from the initial pool of 13 2D and 3D similarities (cf. Section 2) we selected fingerprints that capture different aspects of the chemical structures⁴⁶ and showed relatively low linear correlations for the 14,028 pairwise structure similarities. The maximum correlation between any of the four selected representations was <0.55 (Table 2).

3.3. Consensus models of activity landscapes

Figure 3 shows DAD and TAD maps with the 14,028 pairwise potency differences of the 168 compounds of the data set. The distribution of the data points in the plots is independent of the structure similarity (see above). Data points are colored by mean structure similarity using a continuous scale from less similar (green) to more similar (red) structures. Mapping similarity values into the DAD maps clearly revealed activity differences associated with large and small differences in chemical structures making straightforward the identification of single-, dual-, and triple-receptor activity cliffs and scaffold hops (see below).

Figure 4 shows DAD and TAD maps displaying only 534 molecule pairs with the highest mean structure similarity (>0.75). A total of 393 pairs had potency difference lower than 1 log unit for all three receptors subtypes and are associated with a continuous SAR; 78 pairs were identified as single-receptor activity cliffs, 48 pairs as dual-receptor activity cliffs with 'direct SAR', 4 pairs as dual-receptor activity cliffs with 'inverse SAR', and 11 pairs as triple-receptor activity cliffs. Table 2 summarizes the number of single- and dual-receptor activity cliffs for each subtype identified in the DAD and TAD maps. For the set of 168 compounds studied in this work, the larger number of activity cliffs for PPAR γ as compared to the number of cliffs for PPAR α and δ suggests that the SAR for PPAR γ is more discontinuous than the SAR of the other two subtypes. This is also an indicative that it would be more challenging to develop quantitative models such

Table 2
Counts of activity cliffs and scaffold hops

Type	PPAR subtype	Scaffold Hops Total	Activity cliffs ^a		
			Total	Direct	Inverse
Single-target	α	14	18 (3)		
	δ	20	19 (0)		
	γ	30	41 (15) ^b		
Dual-target	α - δ	4	8 (2)	6 (2)	2 (0)
	α - γ	23	31 (16)	30 (15)	1 (1)
	δ - γ	7	13 (5)	12 (5)	1 (0)
Triple-target (pan cliff)	α - δ - γ	11	11 (7)	10 (7)	1 (0)

^a Number in parenthesis indicate activity cliffs with potency difference of more than 1.5 log units.

^b Six activity cliffs have potency difference of more than 2 log units.

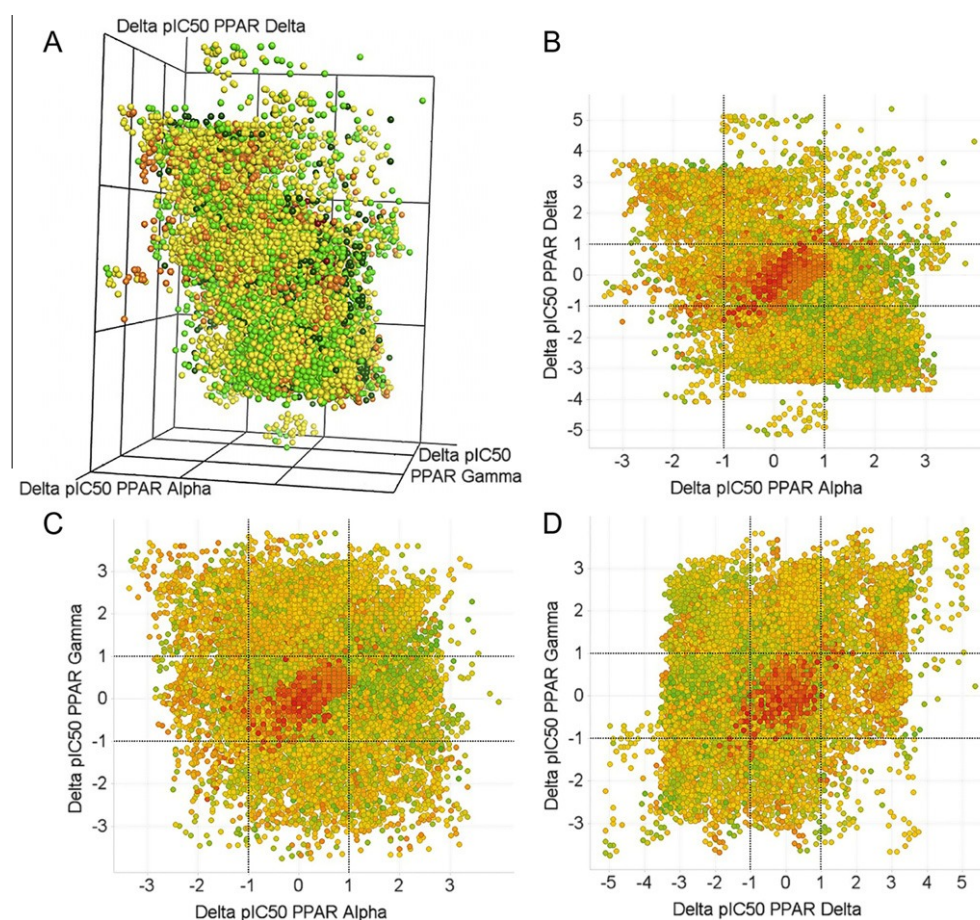


Figure 3. (A) Triple activity-difference maps for the three PPAR subtypes. Dual activity-difference maps for PPAR: (B) α - δ ; (C) α - γ ; and (D) δ - γ . Each map contains 14,028 data points resulting from the pairwise comparisons of 168 compounds. Data points are color-coded by the mean structure similarity using a continuous scale from more similar (red) to less similar (green).

as QSAR for PPAR γ .¹⁶ Of note, 15 out of the 41 activity cliffs for PPAR γ have a potency difference of more than 1.5 log units and six cliffs of more than 2 log units (Table 2). Activity cliffs with a very large potency difference can be further identified as 'deep activity cliffs'.⁴⁰ The number of deep cliffs for PPAR α and PPAR δ is rather low (Table 2).

The α - δ DAD maps had a lower number of dual-receptor activity cliffs than the other two DAD maps (Table 2). This result suggests more of an agreement between the SAR of the 168 compounds tested against PPAR α and δ . In contrast, the α - γ DAD map showed a larger number of dual-receptor cliffs indicating more disagreement between the SAR of the compounds tested against these two subtypes.

Selected pairs of compounds which are examples of single-, dual- and triple-receptor activity cliffs are labeled in Figure 4. Figures 5 and 6 show the chemical structures of the selected pairs. The corresponding values of potency difference for the three subtypes, and the structure similarity from different representations are summarized in Table 3.

3.3.1. Single-receptor activity cliffs

Figure 5 shows examples of single-target activity cliffs for the three PPAR subtypes. Pairs **32_36** and **90_93** (Fig. 5A) are representative examples of single-target activity cliffs for PPAR α . In both cases compounds in the pair have a high structural similarity (mean similarity >0.75). The only structural difference is the

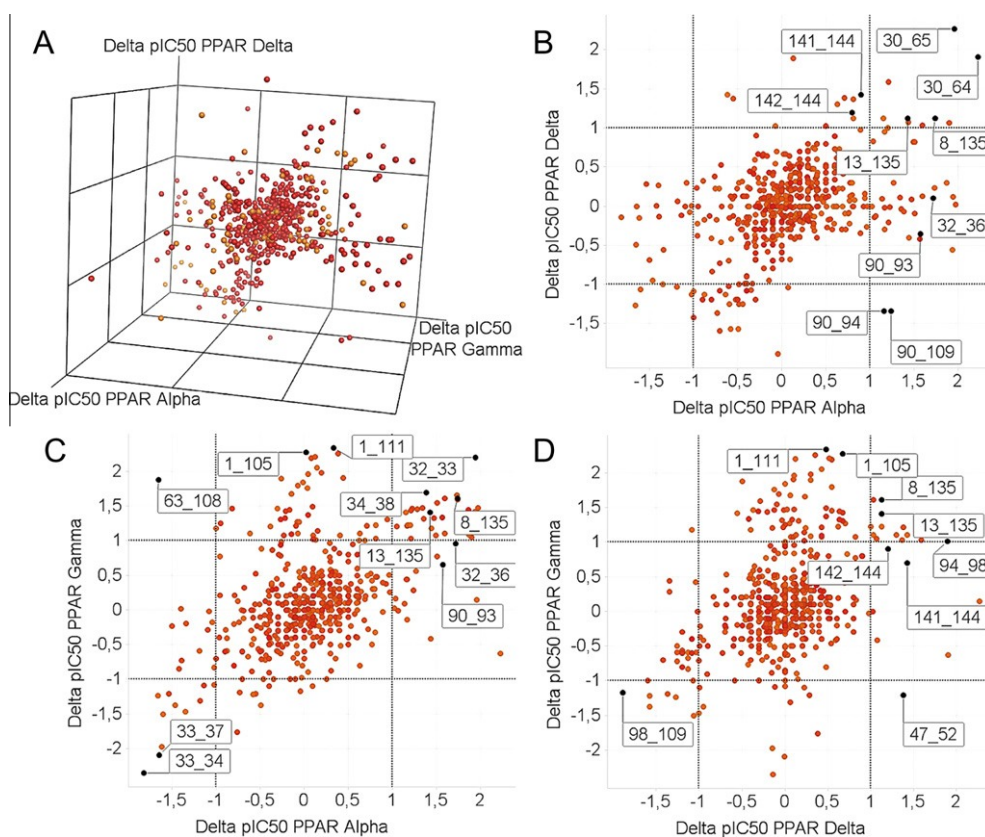


Figure 4. (A) Triple activity-difference map for the three PPAR subtypes. Dual activity-difference maps for PPAR: (B) α - δ ; (C) α - γ ; and (D) δ - γ . Each graph shows 534 data points, all with a mean structure similarity >0.75. Selected points are examples of activity cliffs.

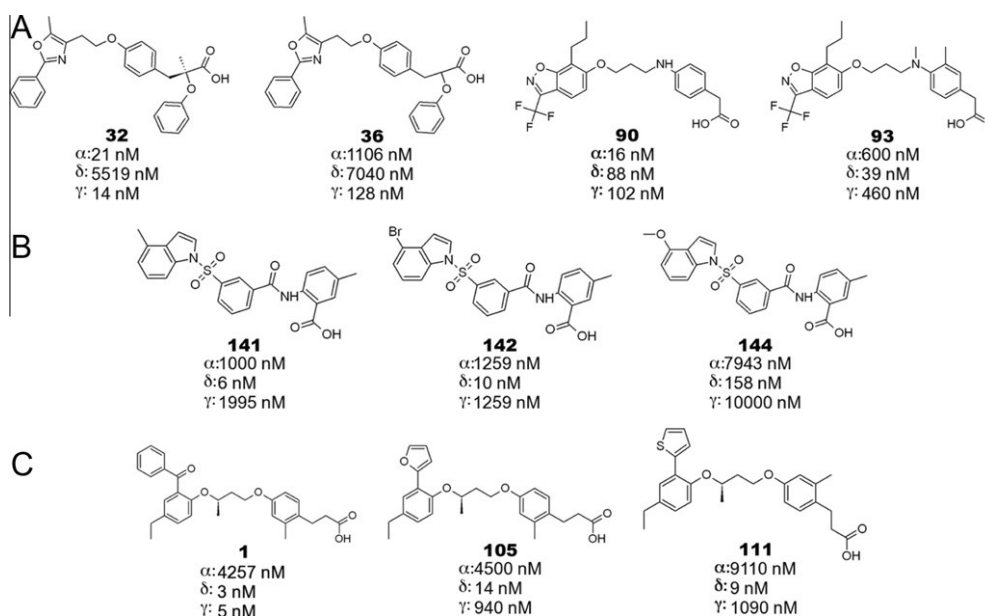


Figure 5. Chemical structures and IC_{50} values for single-target activity cliffs of PPAR (A) α (compound pairs **32_36** and **90_93**); (B) δ (pairs **141_144** and **142_144**); and (C) γ (pairs **1_105** and **1_111**). The position of each pair in the DAD maps is depicted in Figure 4.

addition of a methyl in the first pair and the change of a secondary amine for a tertiary amine in the second one. These small structural changes are associated with a potency difference of more than 1.5 log units against PPAR α . Similarly, pairs **141_144** and **142_144** (Fig. 5B) are examples of single-target activity cliffs for PPAR δ . In these cases, the substitution of the methoxy by a methyl

or a bromide is enough to produce a significant change in the activity against PPAR δ . Pairs of compounds **1_105** and **1_111** (Fig. 5C), in which the change of the phenylcarbonyl in **1** by a furan in **105** or a thiophene in **111** produces a decrease of more than 2 log units in the activity, exemplify single-target activity cliffs for PPAR γ . It is important to mention that in all these examples the structural

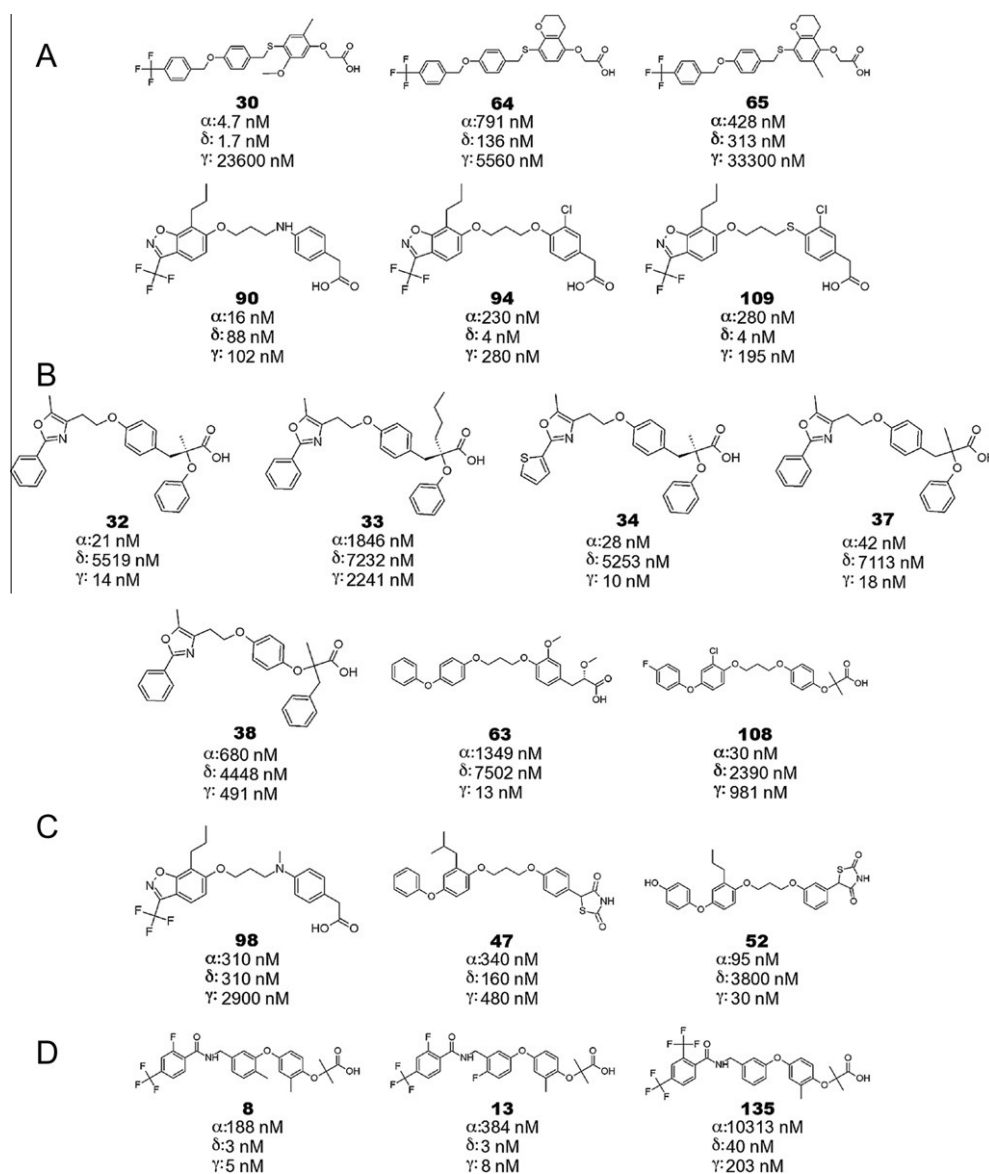


Figure 6. Chemical structures and IC_{50} values for compound pairs that form dual-target activity cliffs of PPAR (A) α - δ ; (B) α - γ ; and (C) δ - γ . Panel D represent the chemical structures and IC_{50} for compound pairs that form triple-target activity cliffs. The position of each pair in the DAD maps is depicted in Figure 4. See text for details.

differences only affect the activity against one PPAR subtype with a small impact in the other subtypes.

3.3.2. Dual- and triple-receptor activity cliffs

The pairs of compounds **30_64**, **30_65**, **90_94** and **90_109** (Fig. 6A) are examples of dual-receptor activity cliffs for PPAR α and δ . Pairs **30_64** and **30_65** exhibit a continuous SAR as they are located in region Z1 of the α - δ DAD map. In other words, structural changes in these compounds affect in the same way the activity against the two subtypes. In these examples, the ring opening of the benzopyrrolone in **64** and **65** into a methoxybenzene, as in **30**, results in an increase of the activity against both subtypes. On the other hand, structural changes between pairs **90_94** and **90_109**, located near region Z2, represent an inverse effect in the activity; the addition of a chloride and the change of the secondary amine in **90** by an oxygen or sulfur atom in **94** and **109**, respectively, decrease the biological activity towards PPAR α whereas the same changes increase the activity toward PPAR δ . Based on this observation we can suggest that the activity for PPAR α depends on the heteroatom in the following order: $NHR_2 > O > S$.

Several dual-receptor activity cliffs were identified for PPAR α and γ . Representative examples are the compound pairs **32_33**, **33_37**, **33_34**, **34_38** and **63_108** (Fig. 6B) of which the first four exhibit a direct SAR, whereas the last one exhibits an inverse SAR. The main structural difference in pairs **32_33**, **33_37**, located in region Z1 of the α - γ DAD map, is the length of the aliphatic chain. A longer chain would be associated with a decrease of the activity against PPAR α and γ , but no effect on the activity against δ . In the same way, pair **33_34** presents two structural differences: the length of the aliphatic chain and the substitution of a phenyl by a thiophene ring. The pair **32_34** shows that replacing a phenyl by a thiophene ring does not impact activity, indicating that the length of the aliphatic chain is the main cause of the activity cliff. Similarly, in pair **34_38**, also presenting a direct SAR, the important difference between the structures of these compounds is the phenyl ether, which increases the activity toward PPAR α and γ . In contrast, pair **63_108**, associated with an inverse effect in activity, is located in region Z2 of the α - γ DAD map. The presence of halogens and the lack of methoxyl groups in **108** are associated with an increase of the activity against PPAR α and with a decrease against γ . On the

Table 3
Potency difference and structure similarity of selected single-, dual- and triple-receptor activity cliffs and scaffold hops

	Pair	$\Delta\text{pIC}_{50\alpha}$	$\Delta\text{pIC}_{50\delta}$	$\Delta\text{pIC}_{50\gamma}$	Mean Similarity	TopAtomPairs	TopPh4Triplets	Estate Indices	Combo Score/2
<i>Activity cliffs</i>									
Single α	32_36	1.722	0.106	0.961	0.86	0.92	0.98	1.00	0.56
Single α	90_93	1.574	-0.353	0.654	0.79	0.85	0.92	0.98	0.43
Single δ	141_144	0.900	1.421	0.700	0.87	0.95	0.91	0.99	0.65
Single δ	142_144	0.800	1.199	0.900	0.87	0.94	0.95	0.99	0.62
Single γ	1_111	0.330	0.477	2.338	0.78	0.85	0.84	0.84	0.58
Single γ	1_105	0.024	0.669	2.274	0.77	0.85	0.84	0.81	0.59
Dual α - δ direct	30_64	2.226	1.903	-0.628	0.76	0.86	0.99	0.99	0.22
Dual α - δ direct	30_65	1.959	2.265	0.150	0.77	0.85	0.97	0.99	0.29
Dual α - δ inverse	90_94	1.158	-1.342	0.439	0.87	0.89	0.93	0.96	0.70
Dual α - δ inverse	90_109	1.243	-1.342	0.281	0.84	0.89	0.94	0.97	0.57
Dual α - γ direct	32_33	1.944	0.117	2.204	0.78	0.86	0.92	1.00	0.35
Dual α - γ direct	34_38	1.385	-0.072	1.691	0.83	0.94	0.98	0.98	0.43
Dual α - γ direct	33_37	-1.643	-0.007	-2.095	0.80	0.86	0.92	1.00	0.43
Dual α - γ direct	33_34	-1.819	-0.139	-2.350	0.79	0.79	0.88	0.98	0.51
Dual α - γ inverse	63_108	-1.653	-0.497	1.878	0.78	0.80	0.93	0.82	0.57
Dual δ - γ direct	94_98	0.130	1.889	1.015	0.81	0.87	0.98	0.96	0.45
Dual δ - γ direct	98_109	-0.044	-1.889	-1.172	0.81	0.86	0.94	0.97	0.45
Dual δ - γ inverse	47_52	-0.554	1.376	-1.204	0.82	0.88	0.96	0.93	0.50
Triple	8_135	1.739	1.125	1.609	0.81	0.81	0.99	0.87	0.58
Triple	13_135	1.429	1.125	1.404	0.81	0.81	0.94	0.97	0.53
<i>Scaffold hops</i>									
Single α	7_62	0.082	2.835	2.232	0.25	0.33	0.17	0.17	0.35
Single α	97_126	0.067	-1.366	-2.632	0.25	0.35	0.24	0.19	0.24
Single δ	95_151	2.421	0.000	1.169	0.23	0.33	0.21	0.16	0.22
Single δ	95_165	2.721	0.097	1.369	0.25	0.35	0.21	0.18	0.26
Single γ	57_168	1.092	-2.635	0.087	0.24	0.31	0.13	0.19	0.33
Dual α - δ direct	31_135	0.319	-0.477	-1.832	0.25	0.38	0.29	0.07	0.24
Dual α - γ direct	57_141	0.092	-3.061	-0.313	0.24	0.19	0.20	0.17	0.40
Dual δ - γ direct	33_83	-1.698	-0.581	0.065	0.25	0.44	0.13	0.25	0.19
Triple	92_143	0.698	-0.455	-0.079	0.25	0.28	0.27	0.18	0.29
Triple	110_131	-0.067	-0.097	-0.691	0.23	0.27	0.22	0.20	0.25

other hand, the absence of halogens and the presence of methoxy groups in **63** improve the activity towards PPAR γ and decreases the activity toward PPAR α . It is important to say that there is no significant difference in the activity of these pairs toward PPAR δ as they are located in regions Z4 and Z3 of the α - δ and δ - γ DAD maps, respectively.

Compound pairs **94_98** and **98_109** exemplify dual-target activity cliffs for PPAR δ and γ with direct SAR. These pairs present a high mean molecular similarity (0.81). As it can be seen in Figure 6C, the only difference is the substitution of the tertiary amine in **98** by an ether in **94** or thioether in **109** and the addition of the chloride. These changes confer a major activity against PPAR δ and γ with a minor effect in the activity against PPAR α . On the other hand, pair **47_52** represents a dual-target activity cliff for PPAR δ and γ with and inverse SAR. In this example a different substitution pattern and a hydroxyl group decrease the activity against PPAR δ while the activity against PPAR γ is increased in more than 1 log unit.

Examples of triple-target activity cliffs are pairs **8_135** and **13_135** (Fig. 6D), which have high structure similarity and are located in region Z1 of all three DAD maps (Fig. 4). In these pairs, the substitution of the trifluoromethyl in **135** by a single fluoride, as in **8** and **13**, significantly increases the activity against the three PPAR subtypes.

3.3.3. Scaffold hops

Figure 7 shows DAD and TAD maps displaying 126 compound pairs with low mean structure similarity (<0.25). A total of 109 pairs had a potency difference lower than 1 log unit and are associated with scaffold hopping. Table 2 summarizes the number of single-, dual- and triple-receptor scaffold hops. Figure 7 shows the position of selected compound pairs in the DAD maps. The corresponding values of potency difference for the three subtypes, and the structure similarity from different representations are summarized in

Table 3. Figure 7E and Figure S1 in the Supplementary data shows the chemical structures of the selected pairs.

Examples of compounds with low structure similarity but low potency difference for single receptors subtypes are pairs **7_62** and **97_126** for PPAR α , pairs **95_151** and **95_165** for PPAR δ , and pair **57_168** for PPAR γ ; all of them located in region Z3 or Z4 of the DAD maps (Fig. 7). Additionally, pairs **31_135**, **57_141** and **33_83**, located in region Z5 in only one DAD map, are examples of dual-receptor scaffold hops for α - δ , α - γ and δ - γ , respectively. Compound pairs **92_143** and **110_131**, which are located in region Z5 in all three DAD maps are examples of triple-target scaffold hops (Fig. 7E). It is worth noting that all four compounds in Figure 7E have some selectivity towards PPAR δ . The scaffolds present in these four compounds are attractive scaffolds to develop selective compounds towards this isoform.

4. Conclusions and perspectives

Several structural features associated with ligand selectivity were identified from the analysis of single- and dual-activity cliffs identified in the data set of 168 compounds with reported activity against three PPAR subtypes. For example, if we consider that most of the PPAR agonists are composed by an aromatic tail, a linker or spacer, and an acidic head,^{5,6} it is possible to suggest that heteroatoms in the spacer increase the activity for PPAR α in the following order: NHR₂ > O > S > NR₃, for example, compounds **90**, **94**, **109**, **98**, respectively. Interestingly, the activity towards the three PPARs is increased when NHR₂ is used instead of NR₃, this may be a consequence of the hydrogen-bond donor and acceptor behavior of the secondary amine, while the tertiary amine only behaves as a hydrogen-bond acceptor. Additionally, results suggest that a large aliphatic chain near the acidic head, for example, the butyl of compound **33** compared with compound **32** and **37**, tends to reduce the

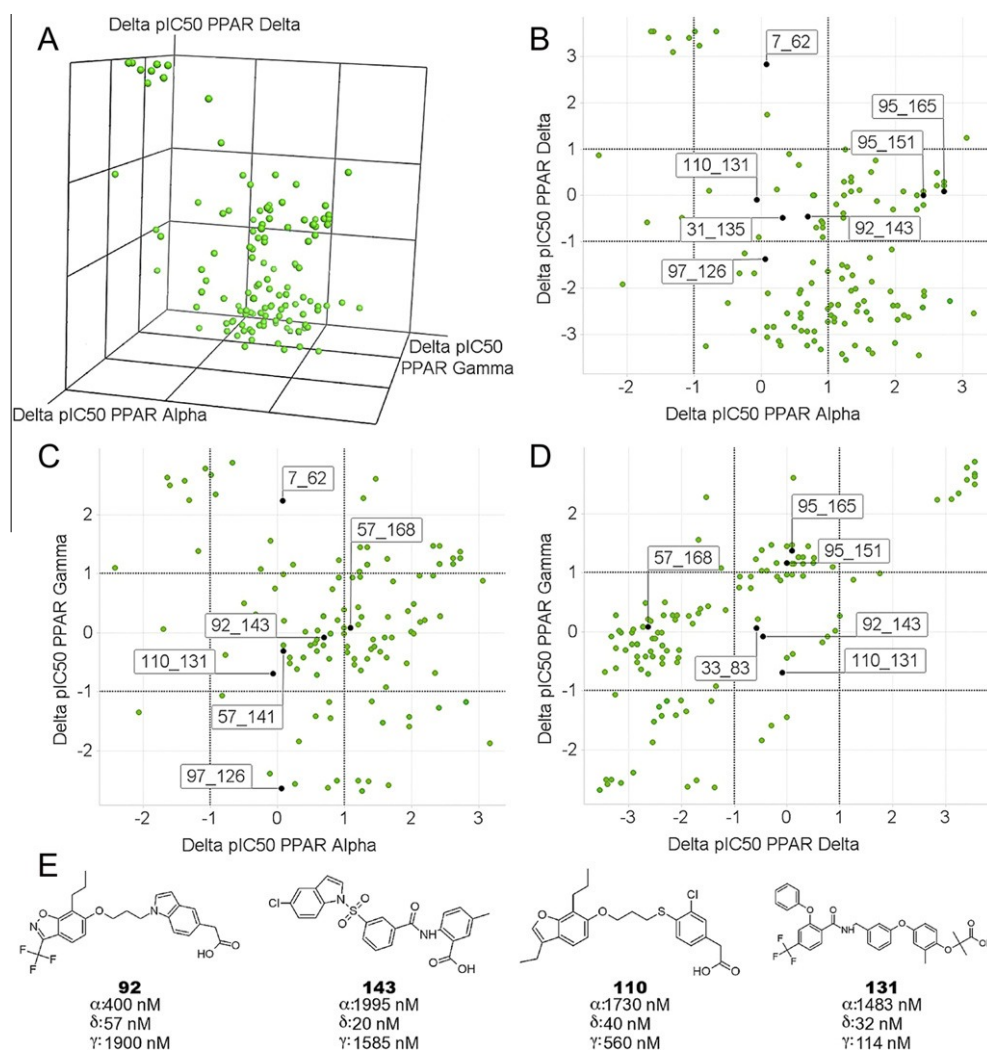


Figure 7. (A) Triple activity-difference map for the three PPAR subtypes. Dual activity-difference maps for PPAR: (B) α - δ ; (C) α - γ ; and (D) δ - γ . Each graph shows 126 data points, all with a mean similarity <0.25. Selected points are examples of scaffold hops. (E) Chemical structures and IC_{50} values of triple-target scaffold hops (compound pairs **92_143** and **110_131**).

biological activity toward PPAR α and γ , making it an important characteristic that could be used for the design of new PPAR δ selective agonists. On the other hand, an aliphatic chain near the aromatic tail seems to be important for activity and many compounds share this structural feature. Because of its importance, this aliphatic chain has been proposed before as an important pharmacophoric feature of pan PPAR agonists.⁹ This work represents a first attempt to systematically analyze the SAR of PPAR ligands using the emerging approach of activity landscape modeling. Perspectives of this work include extending this approach to the SAR analysis of more compounds tested across different PPAR subtypes and the use of the models developed in this work to predict the activity landscape of other data sets.⁴⁷

Acknowledgments

We are grateful to Jacob Waddell for proofreading the manuscript and to the anonymous reviewers for their comments. Authors thank OpenEye Scientific Software, Inc. for providing ROCS. This work was supported by CONACyT project No. 80093. O.M.-L is very grateful to CONACyT for the fellowship granted (No. 245408). J.L.M.-F thanks the State of Florida, Executive Office of the Governor's Office of Tourism, Trade, and Economic Development for funding.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bmc.2012.04.005>.

References and notes

- Shaw, J. E.; Sicree, R. A.; Zimmet, P. Z. *Diabetes Res. Clin. Pract.* **2010**, *87*, 4.
- Moller, D. E. *Nature* **2001**, *414*, 821.
- Willson, T. M.; Brown, P. J.; Sternbach, D. D.; Henke, B. R. *J. Med. Chem.* **2000**, *43*, 527.
- Arck, P.; Toth, B.; Pestka, A.; Jeschke, U. *Biol. Reprod.* **2010**, *83*, 168.
- Zaware, P.; Shah, S. R.; Pingali, H.; Makadia, P.; Thube, B.; Pola, S.; Patel, D.; Priyadarshini, P.; Suthar, D.; Shah, M.; Jamili, J.; Sairam, K. V. V. M.; Giri, S.; Patel, L.; Patel, H.; Sudani, H.; Patel, H.; Jain, M.; Patel, P.; Bahekar, R. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 628.
- Pingali, H.; Jain, M.; Shah, S.; Patel, P.; Makadia, P.; Zaware, P.; Sairam, K. V. V. M.; Jamili, J.; Goel, A.; Patel, M.; Patel, P. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6471.
- Henke, B. R. *J. Med. Chem.* **2004**, *47*, 4118.
- Rubenstrunk, A.; Hanf, R.; Hum, D. W.; Fruchart, J.-C.; Staels, B. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2007**, *1771*, 1065.
- Sundriyal, S.; Bharatam, P. V. *Eur. J. Med. Chem.* **2009**, *44*, 3488.
- Al-Najjar, B. O.; Wahab, H. A.; Tengku Muhammad, T. S.; Shu-Chien, A. C.; Ahmad Noruddin, N. A.; Taha, M. O. *Eur. J. Med. Chem.* **2011**, *46*, 2513.
- Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. *Drug Discovery Today* **2009**, *14*, 698.
- Wassermann, A. M.; Wawer, M.; Bajorath, J. *J. Med. Chem.* **2010**, *53*, 8209.

13. Medina-Franco, J. L.; Yongye, A. B.; López-Vallejo, F. In *Statistical Modeling of Molecular Descriptors in QSAR/QSPR*; Matthias, D., Kurt, V., Danail, B., Eds.; Wiley-VCH, 2012; p 307.
14. Maggiora, G. M. *J. Chem. Inf. Model.* **2006**, *46*, 1535.
15. Guha, R.; VanDrie, J. H. *J. Chem. Inf. Model.* **2008**, *48*, 646.
16. Guha, R.; Van Drie, J. H. *J. Chem. Inf. Model.* **2008**, *48*, 1716.
17. Scior, T.; Medina-Franco, J. L.; Do, Q. T.; Martínez-Mayorga, K.; Yunes Rojas, J. A.; Bernard, P. *Curr. Med. Chem.* **2009**, *16*, 4297.
18. Peltason, L.; Bajorath, J. *Chem. Biol.* **2007**, *14*, 489.
19. Eckert, H.; Bajorath, J. *Drug Discovery Today* **2007**, *12*, 225.
20. Medina-Franco, J. L.; Martínez-Mayorga, K.; Giulianotti, M. A.; Houghten, R. A.; Pinilla, C. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 322.
21. Peltason, L.; Hu, Y.; Bajorath, J. *ChemMedChem* **2009**, *1864*, 4.
22. Wassermann, A. M.; Peltason, L.; Bajorath, J. *ChemMedChem* **2010**, *5*, 847.
23. Pérez-Villanueva, J.; Santos, R.; Hernández-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L. *Med. Chem. Comm.* **2011**, *2*, 44.
24. Medina-Franco, J. L.; Yongye, A. B.; Pérez-Villanueva, J.; Houghten, R. A.; Martínez-Mayorga, K. *J. Chem. Inf. Model.* **2011**, *51*, 2427.
25. Brown, N.; Jacoby, E. *Mini-Rev. Med. Chem.* **2006**, *6*, 1217.
26. Chen, X.; Lin, Y.; Liu, M.; Gilson, M. K. *Bioinformatics* **2002**, *18*, 130.
27. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007**, *35*, D198.
28. Scior, T.; Bernard, P.; Medina-Franco, J. L.; Maggiora, G. M. *Mini-Rev. Med. Chem.* **2007**, *7*, 851.
29. Yongye, A.; Byler, K.; Santos, R.; Martínez-Mayorga, K.; Maggiora, G. M.; Medina-Franco, J. L. *J. Chem. Inf. Model.* **2011**, *51*, 1259.
30. Sud, M. URL: www.MayaChemTools.org.
31. Filimonov, D.; Poroikov, V.; Borodina, Y.; Glorizova, T. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 666.
32. Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1995**, *1039*, 35.
33. Rogers, D.; Hahn, M. *J. Chem. Inf. Model.* **2010**, *50*, 742.
34. Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273.
35. Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. *J. Chem. Inform. Comput. Sci.* **1985**, *25*, 64.
36. Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82.
37. Renner, S.; Fechner, U.; Schneider, G. In *Pharmacophores and Pharmacophore Searches*; Wiley-VCH GmbH KGaA, 2006; p 49.
38. Bonachéra, F.; Parent, B.; Barbosa, F.; Froloff, N.; Horvath, D. *J. Chem. Inf. Model.* **2006**, *46*, 2457.
39. Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. *J. Chem. Inf. Model.* **2009**, *49*, 477.
40. Pérez-Villanueva, J.; Santos, R.; Hernández-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L. *Bioorg. Med. Chem.* **2010**, *18*, 7380.
41. Yongye, A. B.; Bender, A.; Martínez-Mayorga, K. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 675.
42. Jaccard, P. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 241.
43. Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983.
44. Medina-Franco, J. L.; Yongye, A. B.; Pérez-Villanueva, J.; Houghten, R. A.; Martínez-Mayorga, K. *J. Chem. Inf. Model.* **2011**, *51*, 2427.
45. Waddell, J.; Medina-Franco, J. L. *Bioorg. Med. Chem.* in press. doi:<http://dx.doi.org/10.1016/j.bmc.2011.11.051>.
46. Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Glick, M.; Davies, J. W. *J. Chem. Inf. Model.* **2009**, *49*, 108.
47. Kuwabara, N.; Oyama, T.; Tomioka, D.; Ohashi, M.; Yanagisawa, J.; Shimizu, T.; Miyachi, H. *J. Med. Chem.* **2012**, *55*, 893.