



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

---

---

FACULTAD DE QUÍMICA

Manual de quimioinformática para el diseño de  
fármacos: Aplicaciones al estudio de Relaciones  
Estructura-múltiple Actividad (REmA)

T E S I S

QUE PARA OBTENER EL TÍTULO DE  
QUÍMICA FARMACÉUTICO BIÓLOGA

PRESENTA:

FERNANDA ISABEL SALDÍVAR GONZÁLEZ



CIUDAD UNIVERSITARIA, CD. MX.

AÑO 2017



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: Elena Guadalupe Ramírez López

VOCAL: Juan Francisco Cortes Benítez

SECRETARIO: José Luis Medina Franco

1er. SUPLENTE: Carlos Juárez Osornio

2o. SUPLENTE: María Eugenia Mendoza Jasso

## SITIO DONDE SE DESARROLLÓ EL TEMA:

Universidad Nacional Autónoma de México, Ciudad Universitaria, Facultad de Química,  
Departamento de Farmacia, DIFACQUIM, Edificio F, cubículo 108.

## ASESOR DEL TEMA:

Dr. José Luis Medina Franco

---

## SUSTENTANTE:

Fernanda Isabel Saldívar González

---

## PUBLICACIONES

Parte de los conceptos y resultados obtenidos durante el desarrollo de esta tesis están publicados en:

- Saldívar-González Fl., Naveja JJ., Palomino-Hernández O., Medina-Franco JL. Getting SMART in Drug Discovery: Chemoinformatics Approaches for Mining Structure-Multiple Activity Relationships. *RCS Advances*, 2017 7:632-641.
- Saldívar-González F., Prieto-Martínez FD., Medina-Franco JL. Descubrimiento de Fármacos: Un Enfoque Computacional. *Educación Química*, 2017 28:51-58.
- Saldívar-González Fl., Hernández-Luis F., Lira-Rocha A., Medina-Franco JL., Manual de Quimioinformática. *Facultad de Química*, 2017, enviado.

Parte del proyecto se presentó en los siguientes congresos:

- 51° Congreso Mexicano de Química y en el 35° Congreso Nacional de Educación Química llevado a cabo del 28 de septiembre al 1 de octubre del 2016, en la Ciudad del Conocimiento, de la Universidad Autónoma del Estado de Hidalgo, en Pachuca, Hidalgo, México.
- 12<sup>th</sup> German Conference on Chemoinformatics llevado a cabo del 6 al 8 de noviembre del 2016 en Fulda, Alemania.

## **AGRADECIMIENTOS**

A la Universidad Nacional Autónoma de México, especialmente a la Facultad de Química, por haber recibido de ella mi formación profesional, por los excelentes profesores de los que tuve la oportunidad de aprender y por haberme permitido conocer a personas que admiro y quiero tanto.

Mi más sincero agradecimiento al Dr. José Luis Medina Franco, por sus invaluable consejos, por el apoyo y dirección de este trabajo y por la gran calidez humana mostrada. Su gran dedicación y amor por su trabajo es de admirar.

De igual forma, agradezco discusiones y comentarios de mis compañeros de DIFACQUIM, gracias por su gran ayuda y por hacer que mi estancia sea de lo más agradable y provechosa.

A cada uno de los miembros del jurado, por el tiempo dedicado en leer este trabajo y enriquecerlo con sus comentarios.

Finalmente, agradezco a Dirección General de Profesiones (DGAPA) por la beca otorgada por medio del Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIME), proyecto número PE200116 y a la Facultad de Química, UNAM, por la beca otorgada a través del Subprograma 127 "Formación Básica en Investigación".

## ÍNDICE GENERAL

<b>ABREVIATURAS Y ACRÓNIMOS.....</b>	<b>i</b>
<b>ÍNDICE DE FIGURAS.....</b>	<b>ii</b>
<b>INDICE DE TABLAS .....</b>	<b>iii</b>
<b>RESUMEN.....</b>	<b>iv</b>
<b>ABSTRACT.....</b>	<b>v</b>
<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
<b>2. ANTECEDENTES.....</b>	<b>2</b>
<b>3. JUSTIFICACIÓN.....</b>	<b>8</b>
<b>4. OBJETIVOS.....</b>	<b>9</b>
<b>5. HIPÓTESIS.....</b>	<b>9</b>
<b>6. MANUAL DE QUIMIOINFORMÁTICA</b>	
Resumen.....	10
Contenido.....	10
Objetivos.....	11
Metodología.....	11
Capítulo 1. Minería de bases de datos.....	12
Capítulo 2. Espacio químico.....	16
Capítulo 3. Relaciones estructura-actividad.....	19
Capítulo 4. Acoplamiento molecular.....	22
Capítulo 5. Toxicología informática.....	26
Ejercicios Finales.....	29
<b>7. CASO DE ESTUDIO: Estudio quimioinformático del REmA</b>	
Resumen .....	30
Introducción .....	31
Objetivos.....	32

Metodología.....	36
Resultados y discusión.....	42
Conclusiones.....	55
<b>8. CONCLUSIONES.....</b>	<b>56</b>
<b>9. PERSPECTIVAS .....</b>	<b>56</b>
<b>10.REFERENCIAS .....</b>	<b>57</b>
<b>APÉNDICE A. Base de datos de inhibidores de HDACs .....</b>	<b>64</b>
<b>ÁPÉNDICE B. Publicaciones y Congresos .....</b>	<b>69</b>

**ABREVIATURAS Y ACRÓNIMOS**

ADME	Absorción, Distribución, Metabolismo y Excreción
CI50	Concentración inhibitoria 50
CDII	Dominio catalítico II
DAD	Del inglés: <i>Dual Activity Difference</i>
DIFAC	Diseño de fármacos asistido por computadora
DPP IV	Enzima dipeptidil peptidasa IV
ECFP	Del inglés: <i>Extended Conectivity Fingerprints</i> (Huellas digitales moleculares por conectividad extendida)
FDA	Del inglés: <i>Food and Drug Administration</i>
HBA	Del inglés: <i>Hydrogen bond acceptor</i>
HBD	Del inglés: <i>Hydrogen bond donor</i>
HDAC	Histona deacetilasa
MOE	Del inglés: <i>Molecular Operating Environment</i>
PCA	Del inglés: <i>Principal components analysis</i>
PDB	Del inglés: <i>Protein Data Bank</i>
PM	Peso molecular
QSAR	Del inglés: <i>Quantitative structure–activity relationships</i>
SAR	Del inglés: <i>Structure activity relationship</i>
SAS	Del inglés: <i>Structure Activity Similarity</i>
SlogP	Coeficiente de reparto octanol /agua.
SMILES	Del inglés: <i>Simplified Molecular Input Line Entry System</i>
SmAR	Del inglés: <i>Structure multiple Activity Relationship</i>
TAD	Del inglés: <i>Triple Activity Difference</i>
tPSA	Del inglés: <i>topological Polar Surface Area</i>

## ÍNDICE DE FIGURAS

	Pág.
<b>Figura M1</b> Espacio químico obtenido mediante PCA de fármacos aprobados para uso clínico hasta el año 2015	16
<b>Figura M2.</b> Visualización del SAR de inhibidores de la enzima dipeptidilpeptidasa IV mediante el empleo de mapas SAS implementados en DataWarrior.	20
<b>Figura C1</b> Estructura de la enzima HDAC6	32
<b>Figura C2</b> Compuestos inhibidores de HDACs.	33
<b>Figura C3</b> Forma general de un mapa SAS ( <i>Structure Activity Similarity</i> )	39
<b>Figura C4</b> Forma general de un mapa DAD ( <i>Dual Activity Difference</i> )	40
<b>Figura C5</b> Espacio químico de 140 compuestos inhibidores de HDACs obtenido mediante un análisis de componentes principales de la matriz de similitud calculada con ECFP4/Tanimoto.	42
<b>Figura C6</b> Ejemplo de estructuras presentes en cada uno de los grupos identificados por medio del análisis del espacio químico.	43
<b>Figura C7</b> Mapas SAS de las diferentes isoformas de HDAC.	44
<b>Figura C8</b> Pares de compuestos representativos situados en la región II del mapa SAS de HDAC6 y en la región IV de los mapas SAS de HDAC1, 2 y 3.	44
<b>Figura C9</b> Número de <i>activity cliffs</i> asociados con cada compuesto en la base de datos de iHDACs.	45
<b>Figura C10</b> Número de <i>activity cliffs</i> formados por cada compuesto en las diferentes isoformas de HDACs.	48
<b>Figura C11</b> Mapas DAD para las diferentes isoformas de HDAC.	50
<b>Figura C12</b> Modelos de acomplamiento obtenidos con <i>docking</i> con el dominio catalítico 2 de HDC6 de 1 y 14, par de compuestos identificado como <i>activity cliff</i> para HDAC6.	53

## ÍNDICE DE TABLAS

	Pág.
<b>Tabla 1</b> Ejemplos de libros de texto enfocados en Quimioinformática.	6
<b>Tabla M1</b> Páginas disponibles en línea para búsqueda y construcción de bases de datos químicas.	13
<b>Tabla M2</b> Tipos de descriptores usados frecuentemente en Quimioinformática.	17
<b>Tabla M3</b> Métodos desarrollados en el modelado de panoramas de actividad y principales aplicaciones.	20
<b>Tabla M4</b> Tipos de puntuaciones usadas con frecuencia en estudios de acoplamiento molecular.	25
<b>Tabla M5</b> Principales bases de datos utilizadas en Toxicología Informática.	27
<b>Tabla M6</b> Programas de cómputo y sitios web disponibles para realizar predicciones ADME/Tox.	28
<b>Tabla C1</b> Distribución del número de <i>activity cliffs</i> formados por cada compuesto en la base de datos.	46
<b>Tabla C2</b> Compuestos identificados como generadores de <i>activity cliffs</i> .	46
<b>Tabla C3</b> Número de pares de compuestos presentes en cada región de los mapas DAD.	49
<b>Tabla C4</b> Resultados representativos del análisis de los mapas DAD.	51
<b>Tabla C5</b> Resultados de la evaluación de propiedades ADME.	54

## RESUMEN

La Quimioinformática es una disciplina que ha tenido gran aplicación en muchas áreas de la química, en especial en el diseño y desarrollo de fármacos. A pesar de los avances y aplicaciones de esta disciplina, los recursos empleados para su enseñanza a profundidad aún son escasos e incluso inexistentes en algunos centros educativos.

En este trabajo se presenta material didáctico generado para mejorar la enseñanza de quimioinformática en la licenciatura de Química Farmacéutico Biológica. Se incluye un extracto del Manual de Quimioinformática, donde se describen métodos computacionales asociados con frecuencia en el descubrimiento, diseño y desarrollo de compuestos bioactivos. En el Manual se incorporan ejercicios con enfoques actuales para fomentar en los alumnos el interés en el manejo de técnicas computacionales y la comprensión de conceptos básicos de quimioinformática.

Para ejemplificar el uso de las técnicas descritas en el Manual, se planteó y se desarrolló un caso de estudio enfocado al análisis quimioinformático del REmA (Relaciones Estructura-múltiple-Actividad) de un grupo de 140 compuestos derivados de hidroxipirimidina reportados como inhibidores de HDACs. Las técnicas empleadas en este caso de estudio y descritas en el Manual de Quimioinformática son: minería de datos, análisis de espacio químico, modelado de panoramas de actividad, acoplamiento molecular y toxicología informática.

Como parte de la difusión y enseñanza de la quimioinformática en el contexto de la licenciatura de Química Farmacéutico Biológica, parte de los conceptos de esta tesis se publicaron recientemente en un artículo de difusión en la revista de *Educación Química* en el que se discuten métodos y aplicaciones de quimioinformática en el diseño de fármacos, así como algunas de las falsas expectativas del DIFAC. Conceptos del REmA y parte de los resultados del caso de estudio fueron publicados en un artículo de revisión de la revista *RSC Advances*. Ambos artículos y trabajos presentados en congresos relacionados con esta tesis se presentan como Anexos.

## ABSTRACT

Chemoinformatics is a discipline that has had extensive application in many areas of chemistry, especially in the design and development of drugs. Despite the fact there have been several advances and applications of this discipline, the resources used for their in-depth teaching are still scarce and even non-existent in several educational centers. In this work, we introduce didactic material to improve the teaching of chemoinformatics on Chemistry Pharmacy and Biology at the undergraduate level. An extract from theoretical-practical Manual of Chemoinformatics is included, which describes computational methods frequently associated with the discovery, design and development of bioactive compounds. The Manual incorporates exercises with modern approaches to encourage the students to learn the basic concepts of chemoinformatics and the use of computational techniques.

In order to exemplify the use of several techniques described in the Manual, a case study was developed focused on the chemoinformatics analysis of SMART (Structure-Multiple-Activity Relationships) of a data set of 140 pyrimidine hydroxy amine compounds reported as inhibitors of HDACs. The techniques used in this case study and described in the Manual of Chemoinformatics are: data mining, chemical space analysis, activity landscape modeling (ALM), molecular docking, and computational toxicology.

As part of the dissemination and teaching of chemoinformatics in the context of Chemistry Pharmacy and Biology at the undergraduate level, part of the concepts of this thesis were published recently in a scholar article of the journal *Educación Química*, in which the methods and applications of chemoinformatics in drug design are discussed, as well as some of the false expectations of DIFAC. Concepts of the SMART and part of the results of the case study were published in a review article of the journal *RSC Advances*. Both articles and abstracts presented at scientific conferences related to this thesis are included as Attachments.

## 1. INTRODUCCIÓN

El desarrollo de fármacos cada vez más efectivos, seguros y adecuados en el tratamiento de enfermedades no es una tarea sencilla y requiere del esfuerzo conjunto de institutos académicos, industrias farmacéuticas y agencias gubernamentales.

Hoy en día, el proceso de descubrimiento y desarrollo de fármacos ha revolucionado con el advenimiento de la genómica, proteómica, bioinformática y tecnologías eficientes como: química combinatoria, *high-throughput screening* (HTS), diseño *de novo* y el diseño *in silico* o asistido por computadora (1).

En este sentido, el diseño de fármacos asistido por computadora (DIFAC) permite un mejor uso de recursos y del tiempo invertido en el proceso de identificación, desarrollo y optimización de fármacos (2).

Una de las disciplinas ampliamente utilizadas por el DIFAC y que se considera fundamental en la enseñanza de la Química Farmacéutica es la quimioinformática. Esta disciplina facilita el manejo, clasificación e interpretación en forma eficiente de grandes cantidades de datos químicos y la predicción de la actividad biológica (3).

El acoplamiento molecular (*docking*), relaciones cuantitativas estructura-actividad (QSAR), análisis del espacio químico y recientemente la toxicología informática son ejemplos de técnicas comunes desarrolladas dentro de esta área y que se emplean con alta frecuencia en la industria farmacéutica, universidades y centros de investigación públicos y privados.

Con la finalidad de contribuir en la enseñanza de la Química Farmacéutica y subsanar la deficiencia de recursos educativos, se elaboró material didáctico de quimioinformática que incluye un manual teórico-práctico y un artículo de difusión en el que se describen métodos y aplicaciones de quimioinformática en el diseño de fármacos. También se planteó y desarrolló un caso de estudio que enfatiza la aplicación de métodos computacionales en el contexto de conceptos novedosos que han surgido en el área.

## 2. ANTECEDENTES

### Quimioinformática

La quimioinformática es un término que se introdujo recientemente. En 1998 Frank Brown definió por primera vez a la quimioinformática como: “la combinación de aquellos recursos de información para transformar datos en información y la información en conocimiento con el propósito de tomar mejores y más rápidas decisiones en el área de identificación y optimización de compuestos líderes” (4). Esta definición vincula la quimioinformática con la investigación farmacéutica, no obstante, muchas de las técnicas que tienen sus raíces en esta área son de aplicabilidad mucho más amplia. Por este motivo, autores como J. Gasteiger y T. Engel consideran que la quimioinformática puede ser vista como la aplicación de métodos informáticos para resolver problemas químicos (5).

A pesar de que el término de quimioinformática fue introducido recientemente, este campo tiene una larga historia y sus orígenes se remontan a los años sesenta, donde aparecieron diversos métodos como: la representación y la búsqueda de estructuras químicas, las relaciones cuantitativas estructura-actividad, la quimiometría, el modelado molecular, así como la elucidación de estructuras asistida por computadora y el diseño de síntesis (6, 7).

Si bien, los métodos quimioinformáticos han hecho contribuciones importantes desde su aparición en diversas áreas como Química Analítica, Química Orgánica, Química Farmacéutica, Química de Materiales y el área Química de Alimentos (*FoodInformatics*) (8), fue hasta hace pocos años que la quimioinformática ayudó a resolver problemas multidisciplinarios.

El logro más ampliamente aceptado de la quimioinformática es el que proporciona acceso a la información química en bases de datos (9). La enorme cantidad de información de compuestos químicos y sus propiedades físicas, químicas y biológicas, llevó al desarrollo de bases de datos que permiten almacenar y difundir de manera eficiente datos químicos en forma electrónica. En este mismo sentido, con el fin de facilitar la minería de información en bases de datos, se han desarrollado diversos métodos computacionales que permiten realizar búsquedas más eficientes a partir de estructuras completas, subestructuras y por similitud (10).

Para clasificar y analizar la vasta información en química, también se han desarrollado una gran variedad de métodos entre los que se encuentran: métodos estadísticos, métodos de reconocimiento de patrones, redes neuronales artificiales y métodos basados en algoritmos genéticos. Estas estrategias pueden clasificarse en métodos de aprendizaje no supervisado y supervisado y se utilizan para la clasificación o el modelado cuantitativo (11).

Otros logros de la quimioinformática han sido la elucidación y la representación de estructuras químicas. Esto ha permitido modelar las relaciones entre la estructura de un compuesto y su información espectroscópica, así como representar de manera única compuestos químicos y percibir características tales como aromaticidad, estereoquímica y estructuras en 3D.

### **Quimioinformática aplicada en el diseño de fármacos**

La Química Farmacéutica es el área de investigación en la que quizá la quimioinformática ha encontrado la mayor aplicación. El uso de la tecnología y la gestión de la información se han convertido en una parte crítica del proceso de descubrimiento y desarrollo de fármacos.

El gran uso de la quimioinformática en la Química Farmacéutica se debe en gran parte a la aparición de estrategias como química combinatoria y HTS que han conducido a un incremento masivo de datos de estructuras químicas y actividades biológicas (12). A su vez, esto demanda el empleo de métodos computacionales más eficientes que permiten organizar, analizar y entender los datos químicos generados (13).

Por otra parte, el empleo de métodos quimioinformáticos también ha permitido abordar de manera eficiente algunas fallas frecuentes de los enfoques tradicionales del diseño y desarrollo de fármacos tales como farmacocinética deficiente, falta de eficacia y toxicidad de los compuestos.

En resumen, dentro del proceso de descubrimiento y desarrollo de nuevos fármacos, la quimioinformática es utilizada para (10, 13):

- Identificación de nuevos compuestos líderes
- Optimización de compuestos líderes

- Establecimiento de relaciones estructura-actividad
- Comparación de bibliotecas de estructuras químicas
- Análisis cuantitativo y visual de diversidad estructural
- Minería de datos provenientes de HTS
- Acoplamiento molecular ligando-receptor
- Diseño *de novo* de ligandos
- Modelado de propiedades de ADME-Tox
- Análisis de rutas bioquímicas

Asimismo, nuevos enfoques surgidos en el diseño de fármacos como es el reposicionamiento y el concepto de polifarmacología, se están beneficiando de métodos quimioinformáticos, los cuales han demostrado un gran potencial en la predicción de interacciones de un fármaco con una serie de dianas biológicas y en sugerir usos terapéuticos alternos para fármacos ya aprobados para uso clínico (14).

Aún queda pendiente desarrollar nuevos tratamientos farmacéuticos para atender necesidades médicas no cubiertas y abordar retos como son: el desarrollo de descriptores moleculares que describan de una manera más completa a las estructuras químicas, el mejoramiento en la calidad de métodos que asignan valores o puntajes en acoplamiento molecular y la optimización paralela de potencia, selectividad y propiedades ADMET. Se espera que con la mejora de la capacidad de los equipos de cómputo se puedan ofrecer soluciones cada vez más rápidas e innovadoras a estos problemas. También, se espera que el proceso de diseño y desarrollo de nuevos fármacos se acorte y se dirijan de mejor manera los recursos empleados en este proceso.

### **Métodos computacionales utilizados en el diseño de fármacos**

Como se mencionó en el apartado anterior, la implementación de métodos computacionales permite abordar y enfocarse en diferentes aspectos del desarrollo de fármacos. Dependiendo el nivel de información disponible y los objetivos del caso de estudio serán los métodos computacionales que podrán abordarse, los cuales, se pueden clasificar en dos enfoques principales (15):

- a) Basados en la estructura del receptor
- b) Basados en la estructura del ligando

Los métodos computacionales basados en estructura del receptor requieren de información relacionada con el blanco molecular (p. ej. receptor o enzima), por lo que es necesario recurrir a bases de datos como *Protein Data Bank* (RCSB) y *Uniprot*, o bien, recurrir a técnicas de modelado molecular.

Entre los métodos basados en la estructura del receptor se pueden enumerar los siguientes:

- Acoplamiento molecular
- Dinámica molecular
- Modelo del farmacóforo
- Diseño *de novo* de nuevos ligandos

Por otra parte, los métodos basados en la estructura del ligando, son utilizados generalmente en ausencia de información acerca del blanco molecular. Estos métodos se fundamentan en estudios de relaciones estructura-actividad de ligandos o inhibidores enzimáticos. Debido a que aún es insuficiente la cantidad de biomoléculas involucradas en procesos metabólicos asociados a enfermedades, los métodos basados en la estructura del ligando se han posicionado entre los más utilizados en la industria farmacéutica. Ejemplos de estos métodos son:

- Relaciones cuantitativas estructura-actividad (QSAR y métodos comunes)
- Panoramas de actividad (*activity landscape modeling*)
- Huellas digitales moleculares (*molecular fingerprints*) y métricas de similitud.

La finalidad de implementar cualquiera de estos dos enfoques es predecir la actividad biológica u otra propiedad farmacológica de una serie de moléculas.

En quimioinformática, varios de los métodos mencionados anteriormente, así como recursos disponibles en bases de datos y programas web, son utilizados con alta frecuencia en el diseño de fármacos para el manejo, visualización y análisis sistemático de la información química. Una explicación más detallada de los métodos computacionales utilizados en el diseño de fármacos se encuentra descrita en el artículo publicado en *Educación Química* (Anexo B).

## Enseñanza de Quimioinformática

Ante la creciente información acerca de conceptos y técnicas surgidas en el área de quimioinformática, aumenta la necesidad de libros de texto y personas con experiencia en el área que dirijan los esfuerzos de investigación en la dirección correcta y hagan de la quimioinformática un área enriquecedora. Ante esta necesidad, algunas Universidades como Indiana (EUA), Sheffield y Manchester (ambas en Inglaterra) están integrando programas de enseñanza enfocados en quimioinformática. De tal forma que los alumnos se familiaricen con conceptos y técnicas computacionales que actualmente son utilizadas en química (16).

En este contexto, se ha escrito material enfocado a quimioinformática en diferentes libros y revistas que se especializan en el área. La Tabla 1 resume ejemplos de libros de texto enfocados en quimioinformática y con los que actualmente se cuenta para la enseñanza de esta disciplina.

Aunque parece que la literatura relacionada con la quimioinformática es cada vez más abundante, no es fácil para los estudiantes que inician en esta área identificar lecturas generales que les permitan adentrarse en forma sistematizada en este campo. Asimismo, mucho del material generado hasta el momento resalta por ser poco didáctico, lo que dificulta su empleo para la enseñanza a profundidad de temas asociados a quimioinformática y que se consideran importantes en el diseño de fármacos.

**Tabla 1.** Ejemplos de libros de texto enfocados en Quimioinformática.

<b>Libro</b>	<b>Notas</b>
<i>Introduction to Chemoinformatics</i> , Springer; 1 Rev.Ed. 2007, Leach y Gillet, Eds.	Libro de texto básico.
<i>Chemoinformatics: A Text Book</i> , Wiley-VCH; 1 ed.2003, Gasteiger, Engel, Eds.	Libro de texto avanzado.
<i>Chemoinformatics for Drug Discovery</i> , Wiley; 1 ed. 2013, J. Bajorath, Ed.	Aplicaciones de Quimioinformática al diseño de fármacos.
<i>Handbook of Chemoinformatics Algorithms</i> , Chapman and Hall/CRC; 1 ed. 2010, Bender, Faulon Eds.	Fundamentos teóricos de algoritmos diversos.
<i>Statistical Modelling of Molecular Descriptors in QSAR/QSPR</i> , Wiley-Blackwell; 1 ed.2012, Dehmer,Varmuza Eds.	Métodos modernos para hacer análisis QSAR.

Actualmente, se requiere que más centros educativos implementen en sus programas de estudio temas de quimioinformática y que se genere material didáctico para su enseñanza. Este material debe ayudar a reforzar conceptos básicos de química (p.ej. estereoquímica), desarrollar la capacidad para abordar enfoques multidisciplinarios (p. ej. en estudios QSAR y de acoplamiento molecular) y a su vez debe promover competencias transferibles relacionadas con la gestión de datos.

En la Facultad de Química, UNAM, una de las asignaturas en la que se considera indispensable implementar conceptos y técnicas de quimioinformática es *Química Farmacéutica (clave 0110)*. Esta asignatura constituye parte de la última etapa de formación de estudiantes de la licenciatura de Química Farmacéutico Biológica, etapa en la cual se busca que alumnos apliquen y profundicen en temas que les son de interés.

El objetivo principal de la asignatura de Química Farmacéutica es sentar las bases para el diseño racional de nuevos fármacos. Esto a partir del estudio de algunos principios básicos que rigen a los fármacos como son: su obtención, su mecanismo de acción a nivel molecular, las relaciones que existen entre la estructura química y la actividad biológica, sus productos de biotransformación y los mecanismos mediante los cuales estos se producen.

Implementar material para la enseñanza de quimioinformática como el que se propone en este trabajo, contribuirá a que se vea alcanzado el objetivo de la asignatura de Química Farmacéutica. Esto dada la importancia que van adquiriendo los métodos computacionales como herramientas para el diseño racional de nuevos fármacos. De igual forma, se ayudará a muchos estudiantes que se ven atraídos por esta asignatura a que se les facilite la lectura de reportes científicos relacionados con el diseño de fármacos en el contexto actual.

En un corto y mediano plazo se espera que el material didáctico desarrollado sea de utilidad en cursos de Química Farmacéutica y cursos relacionados que se imparten en otras universidades y centros educativos de México y otros países de habla hispana.

### 3. JUSTIFICACIÓN

El número de aplicaciones exitosas que han tenido las técnicas de quimioinformática y otros métodos de Diseño de Fármacos Asistido por Computadora (DIFAC) en el desarrollo de compuestos presentes en el mercado ha hecho que ya se estén incluyendo aspectos teóricos y prácticos de quimioinformática en la asignatura de Química Farmacéutica en la Facultad de Química, UNAM. Sin embargo, aún no es posible profundizar o abarcar otros temas que son relevantes dentro de esta disciplina. Esto se debe, en parte, a la cantidad limitada de libros de texto especializados en esta área y a la disponibilidad aún más restringida de literatura en español sobre quimioinformática.

Implementar nuevos recursos que mejoren la enseñanza de los métodos de quimioinformática a nivel licenciatura permitirá que los métodos computacionales sean percibidos como herramientas útiles en el diseño de fármacos y sean más los alumnos e investigadores los que conozcan y se beneficien con la implementación de estos métodos.

Con el material generado se ayudará fundamentalmente a sentar las bases para priorizar y racionalizar los recursos empleados por instituciones académicas en desarrollo de nuevos fármacos. De igual forma, se espera que el material generado motive a los estudiantes a usar herramientas computacionales innovadoras y de fácil acceso de tal forma que mejoren su aprendizaje y comiencen a familiarizarse con metodologías computacionales de punta que les serán de gran utilidad al término de su carrera.

## 4. OBJETIVOS

### General:

Contribuir a la enseñanza de la Química Farmacéutica abordando temas que son de relevancia y que están poco explorados en los contenidos de cursos actuales. Específicamente, se busca fortalecer en los alumnos la comprensión de conceptos básicos de quimioinformática y fomentar el interés en el manejo e interpretación de técnicas computacionales asociadas a esta disciplina.

### Específicos:

- Desarrollar un manual teórico-práctico con los métodos quimioinformáticos más empleados en el descubrimiento, diseño y desarrollo de fármacos.
- Elaborar material de difusión en idioma español en el que se describan métodos y aplicaciones de quimioinformática en el diseño de fármacos.
- Aplicar e integrar en un caso de investigación actual, diversos conocimientos de este manual. Se busca mostrar ejemplos concretos a los estudiantes y estimularlos en aplicar diversos conocimientos adquiridos para el análisis de un problema de investigación.

## 5. HIPÓTESIS

Realizar ejercicios y resolver problemas de quimioinformática basados en conceptos y problemas de investigación novedosos, mejorará la formación de los estudiantes que cursan la materia de Química Farmacéutica. Asimismo, el acceso a literatura especializada en quimioinformática en el idioma español y la resolución de ejercicios utilizando herramientas computacionales accesibles incentivará el interés en los estudiantes para entender los fundamentos teóricos y manejo de técnicas computacionales aplicadas al diseño de fármacos.

## 6. MANUAL DE QUIMIOINFORMÁTICA

En esta sección se expone de manera breve el contenido del Manual de Quimioinformática. El Manual se elaboró para impulsar la enseñanza de métodos computacionales asociados al diseño de fármacos en la asignatura de Química Farmacéutica y otros cursos afines a nivel Licenciatura y Posgrado. Se pretende dar una visión general de los métodos computacionales aplicados en las primeras etapas del desarrollo de fármacos. Estas etapas son la identificación de “candidatos” (*hits*) y de “líderes o cabezas de serie” (*leads*), ya que son en estas etapas donde se invierte el mayor tiempo y donde el curso de Química Farmacéutica está enfocado.

Para la elaboración de este manual se realizó una investigación con respecto al contenido de publicaciones relacionadas con el área de quimioinformática en el contexto actual. De esta manera, también fue posible plantear diferentes ejercicios enfocados al diseño de fármacos.

El manual completo con ejercicios puede consultarse en la versión editada por la Facultad de Química.

### **Contenido:**

El Manual de Quimioinformática que se presenta en este trabajo está estructurado en cinco capítulos donde se hace hincapié en las técnicas computacionales empleadas en el diseño y en las primeras etapas del desarrollo de fármacos. Estas técnicas son: minería de datos, espacio químico, panoramas de actividad, acoplamiento molecular y toxicología informática. En cada capítulo se mencionan los objetivos y en manera de introducción se señalan los conceptos más importantes de cada tema y la relevancia de dicho estudio. Posteriormente, se describe el procedimiento con ejemplos aplicables en investigación y finalmente se refuerza lo aprendido mediante ejercicios y evaluaciones cortas.

Además de estos capítulos, se incluye un apartado con ejercicios finales en donde se busca que los alumnos integren diversos conceptos revisados en capítulos anteriores.

El hecho de poder aplicar los conocimientos teóricos adquiridos en este Manual y durante la carrera a problemas reales hace aún más interesante la realización de este trabajo.

**OBJETIVOS:**General:

- Diseñar y desarrollar un Manual de Quimioinformática que incluya diversos conceptos y ejercicios aplicables al diseño de fármacos.

Específicos:

- Familiarizarse con el uso de bases de datos y programas (de preferencia de libre acceso) disponibles para la investigación de fármacos.
- Analizar textos científicos donde se implementen uno o más métodos quimioinformáticos. Se busca dar un panorama de las diferentes aplicaciones de cada tema y las diferentes formas de abordar un problema de investigación utilizando métodos computacionales.

**METODOLOGÍA:**

Para la elaboración de este Manual se realizó una investigación en literatura reciente relacionada con el área de quimioinformática y el DIFAC.

Los ejercicios se desarrollaron utilizando diversos recursos disponibles en línea como bases de datos y aplicaciones *web*. Esencialmente se utilizaron bases de datos como: ChEMBL (17), PubChem (18), DrugBank (19) y ZINC (20) así como bases relacionadas a toxicología de compuestos químicos. La aplicación *Activity Landscape Plotter*, elaborada por el grupo DIFACQUIM también se utilizó para plantear un ejercicio de relaciones estructura-actividad (<https://unam-shiny-difacquim.shinyapps.io/ActLSmaps/>).

El manejo de las estructuras químicas y la actividad biológica se realizó con el programa computacional *Molecular Operating Environment* (MOE) (21). Este programa contiene módulos de manejo de bases de datos, cálculos de similitud molecular, análisis de agrupamiento, cálculo de descriptores moleculares, incluyendo propiedades fisicoquímicas ampliamente usadas en desarrollo de fármacos, alineamiento flexible y cribado virtual (usando similitud molecular y *docking*).

Además de MOE, se emplearon programas de distribución libre como *DataWarrior* (<http://www.openmolecules.org/datawarrior>) (22) para visualización y manejo de gráficos y ToxiTree para la predicción de propiedades toxicológicas.

A continuación, se resume el contenido y los objetivos a alcanzar en cada capítulo del Manual, así como los ejercicios que fueron planteados.

## **Capítulo 1. “Bases de datos: Bibliotecas de compuestos químicos”**

La gran cantidad de información en bases de datos la cual es accesible a la comunidad científica ha llevado al reconocimiento de que los experimentos tienen utilidad más allá de su propósito inicial. Esta información no debe ser ignorada ni tratada como resultados aislados e independientes, ya que puede conducir a obtener conclusiones valiosas en muchas áreas de la química, en especial en el diseño de fármacos.

Al abordar este tema, se pretende identificar diferentes bases de datos usadas con frecuencia en la investigación de fármacos y el tipo de información disponible en cada una de ellas, de tal forma, que la búsqueda de información pueda realizarse de una manera sencilla y eficiente. Otro de los objetivos de este tema, es llevar a cabo la construcción de bases de datos con compuestos de interés terapéutico e identificar en estas bases de datos, moléculas que interfieren en los cálculos computacionales, así como las características adecuadas que deben presentar los compuestos para subsiguientes estudios computacionales.

Para esto, el capítulo está organizado en tres secciones que son: 1.1) Minería de datos, 1.2) Construcción de bases de datos y 1.3) Curado de bases de datos.

### 1.1 Minería de datos

La extracción de información de bases de datos, generalmente llamada minería de datos (en inglés, *data mining*) es una técnica que surgió para ayudar a comprender el contenido de un repositorio de datos e identificar relaciones numéricas, patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Hoy en día, se puede acceder a distintas bases de datos que poseen información diversa, ya sean de estructuras moleculares, propiedades fisicoquímicas, propiedades biológicas, entre otras. En la investigación en fármacos, bases de datos usadas con frecuencia son PubChem, ChEMBL, ZINC, Chemspider y DrugBank (23, 24). El tipo de información disponible en cada una de estas páginas se resume en la Tabla M1.

**Tabla M1** Páginas disponibles en línea para búsqueda y construcción de bases de datos químicas

	<b>Tipo de información:</b>	<b>URL</b>
<b>PubChem</b>	Sustancias, estructuras y datos de bioactividad.	<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>
<b>ChemSpider</b>	Estructuras químicas de compuestos de interés y la predicción de muchos parámetros físico-químicos.	<a href="http://www.chemspider.com/">http://www.chemspider.com/</a>
<b>ChEMBL</b>	Dianas moleculares, moléculas bioactivas con propiedades farmacéuticas.	<a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a>
<b>DrugBank</b>	Fármacos aprobados por la FDA (moléculas pequeñas y biotecnológicos), nutraceuticos y fármacos en fase experimental con datos farmacológicos, ADMET.	<a href="http://www.drugbank.ca/">http://www.drugbank.ca/</a>
<b>ZINC</b>	Compuestos químicos disponibles comercialmente.	<a href="http://zinc.docking.org/">http://zinc.docking.org/</a>

A partir de la información disponible en bases de datos de la Tabla M1 se plantearon diversos ejercicios, los cuales incluyen:

- Búsqueda de compuestos para extraer información farmacológica relevante.
- Búsqueda de compuestos con información específica (p. ej. compuestos con propiedades fisicoquímicas para ser administrado oralmente).
- Búsqueda de compuestos por similitud.
- Identificación de un grupo de compuestos con actividad biológica para una diana biológica particular.

## 1.2 Construcción de bases de datos.

Debido al aumento en la cantidad de información química, en donde es común el concepto de *big data* (25), el manejo eficiente de información representa un reto en la actualidad. Ante esta necesidad, la construcción de bases de datos puede ser una forma práctica de ordenar información de acuerdo a los datos disponibles y a los objetivos específicos de estudio.

En quimioinformática, la construcción de bases de datos es una práctica fundamental para realizar diversos estudios computacionales. Entre las aplicaciones que pueden darse a bases de datos se encuentran:

- a) Diseño de bibliotecas químicas.
- b) Caracterización y comparación del espacio químico.
- c) Estudio de relaciones complejidad estructural vs. selectividad.
- d) Selección (filtrado) de compuestos y/o sub-grupo de moléculas (p. ej. por propiedades ADMETox), entre otras.

Ejercicios que se incluyen en el Manual de Quimioinformática:

- Identificación de recursos en línea y revistas útiles para la construcción de bases de datos con relevancia terapéutica.
- Identificación de las aplicaciones a bibliotecas moleculares clasificadas de acuerdo a su origen (bibliotecas combinatoriales, obtenidas por “Síntesis Orientada en la Diversidad” (*Diversity-Oriented-Synthesis, DOS*), Bibliotecas-de-Bibliotecas” (*Libraries-from-Libraries, LOL's*), orientadas en productos naturales).
- Construcción de bases de datos con compuestos provenientes de productos naturales.

### 1.3 Curado de bases de datos.

En modelado molecular y quimioinformática se analizan con frecuencia bases de datos moleculares generadas por otros científicos. Estas bases de datos pueden contener información equivocada o innecesaria. Por lo tanto, el curado es importante para evitar obtener resultados erróneos (26).

El curado de bases de datos consiste en:

1. Eliminar moléculas repetidas.
2. Eliminar moléculas inorgánicas y moléculas organometálicas si estas no son el objeto de estudio central. Cabe mencionar que la mayoría de los programas de quimioinformática actuales están desarrollados para procesar moléculas orgánicas pequeñas.
3. ‘Lavado’ de base de datos: antes de realizar una operación en una base de datos de moléculas, se recomienda aplicar a cada molécula un conjunto de reglas de limpieza tales como la eliminación de sales extrañas o el ajuste de los estados de protonación. Esto con el fin de asegurar que cada estructura estará en forma adecuada para el modelado subsiguiente (por ejemplo, visualización, enumeración de conformación y acoplamiento proteína-ligando).

4. Optimización de la geometría mediante la minimización de energía: consiste en encontrar un conjunto de coordenadas atómicas que corresponden a un mínimo local de la función de energía molecular. Esto se hace mediante la aplicación de técnicas de optimización no lineales a gran escala para calcular una conformación (cercana la geometría de partida) para la cual las fuerzas sobre los átomos son cero. La minimización de la energía es útil para la construcción de la molécula, la determinación de conformaciones de baja energía, la búsqueda conformacional y la preparación para las simulaciones de dinámica molecular y análisis de vibración.

Los ejercicios que se incluyen en esta sección son:

- Curado de una base de datos de productos naturales procedente de ZINC.
- Identificación de moléculas repetidas a partir de información prevista en bases de datos.

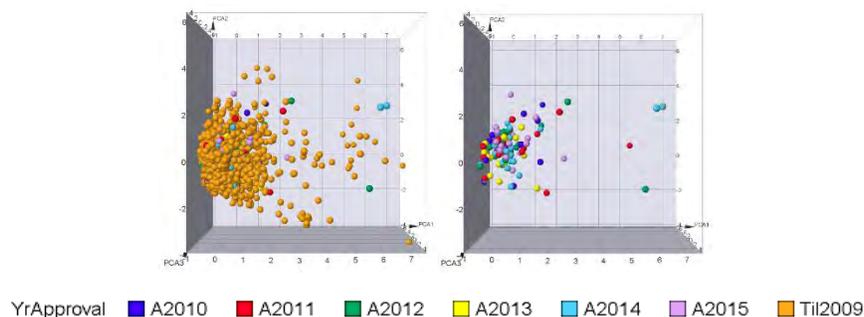
## Capítulo 2. “Espacio químico”

En grupos de investigación dedicados al DIFAC es común que se realicen análisis quimioinformáticos de bases de datos moleculares para obtener su perfil en al menos tres aspectos: propiedades fisicoquímicas, diversidad estructural y distribución en el espacio químico (2).

Para esto, diferentes tipos de descriptores y métodos de visualización pueden ser utilizados. En este capítulo se pretende dar un panorama de los diferentes tipos de descriptores y su utilidad para representar las propiedades de las estructuras químicas, de forma que se pueda realizar una selección de descriptores de acuerdo a los intereses particulares en un determinado estudio. Asimismo, se introduce el método de análisis de componentes principales (PCA por sus siglas en inglés, *Principal Components Analysis*) para llevar a cabo la reducción de dimensiones y facilitar la visualización del espacio químico (27).

Para ejemplificar este concepto, en la Figura M1 se puede observar los resultados de un ejercicio desarrollado en el Manual de Quimioinformática, en donde se visualizó el espacio químico de todos los fármacos aprobados hasta el año 2015. Se utilizó como descriptores 6 propiedades fisicoquímicas (PM, #HBA, #HBD, TPSA, cLogP y número de enlaces rotables) y mediante el método de PCA se obtuvo la visualización en 3D. Debido a que en la visualización del espacio químico por PCA es una reducción de dimensiones, se espera que moléculas (puntos en el gráfico de PCA) con coordenadas semejantes tengan un perfil de propiedades semejantes. Es decir, en cuanto más cerca estén las moléculas, más similares son en cuanto a las propiedades utilizadas para generar el gráfico de PCA (28).

El capítulo 2 del Manual está dividido en dos secciones que son: 2.1) Cálculo de descriptores y 2.2) Visualización del espacio químico.



**Figura M1.** Espacio químico obtenido mediante PCA de fármacos aprobados para uso clínico hasta el año 2015. En color se identifica a los fármacos por año de aprobación.

## 2.1 Calculo de descriptores:

Como primer paso en la visualización del espacio químico es necesario llevar a cabo una selección de descriptores que representen las propiedades de los compuestos que van a ser comparados.

Los descriptores pueden ser determinados a partir de la estructura (constitución, configuración y conformación de la molécula) o por las propiedades (físicas, químicas y biológicas) que presentan las moléculas. Dependiendo el tipo de descriptor utilizado serán las interpretaciones y predicciones que podrán realizarse. La Tabla M2 resume los tipos de descriptores utilizados con frecuencia en la visualización del espacio químico, su función e interpretación, así como algunos ejemplos (29).

**Tabla M2** Tipos de descriptores usados frecuentemente en Quimioinformática.

<b>Descriptor</b>	<b>Función</b>	<b>Interpretación/eficacia</b>	<b>Ejemplos</b>
<b>Índices topológicos</b>	Calcula una variable escalar basado en una gráfica de la molécula sin átomos de hidrógeno.	Solo se tiene la información estructural, aunque esto reduce la interpretación del resultado pueden aplicarse eficientemente.	Índices de Wiener y de Randić.
<b>Propiedades fisicoquímicas</b>	Generalmente estima propiedades físicas de las moléculas con modelos predictivos.	Con valores de estas propiedades se puede hacer un análisis más espontáneo y tomar decisiones informadas.	Constante de disociación del ácido, área polar superficial, solubilidad en agua.
<b>Huellas digitales moleculares</b>	Basados en conocimiento: usan diccionarios. Basados en información: capturan información de la estructura.	Conocimiento: son más fáciles de interpretar, pero tienen poder predictivo bajo. Información: son difíciles de interpretar, pero tienen alto poder predictivo.	Huellas digitales de llaves estructurales ( <i>structure keys</i> ); Huellas digitales de llaves convertidoras ( <i>hash-key</i> )
<b>Modelos del farmacóforo</b>	Generan resultados a partir de posibles puntos de interacción ligando-receptor.	Los resultados son fáciles de interpretar; la eficiencia dependerá de la confiabilidad de los puntos de interacción.	Modelos de 3 y 4 puntos farmacóforicos
<b>Patrones de plantillas moleculares</b>	En base al 'scaffold' (subestructura que soporta a los grupos funcionales).	Su interpretación es intuitiva. En general son buenos predictores.	<i>Scaffolds</i> <i>Scaffold Hopping</i>

Para los ejercicios propuestos en esta sección, se utilizaron principalmente descriptores de propiedades fisicoquímicas ya que son fáciles de interpretar y han resultado de gran utilidad para caracterizar cualidades de agentes terapéuticos como absorción, solubilidad y permeabilidad (30).

---

Los ejercicios incluidos en esta sección están enfocados en:

- Cálculo de propiedades fisicoquímicas de fármacos aprobados hasta el año 2015.
- Cálculo de propiedades fisicoquímicas de bases de datos de productos naturales (base de datos curada en el apartado 1.3).
- Identificar si compuestos cumplen con las reglas de Lipinski, a partir del cálculo de descriptores.

## 2.2 Visualización del espacio químico

Actualmente son numerosos los métodos que existen para dar aproximaciones de la visualización del espacio químico. Estos métodos son generalmente clasificados como *map-based*, que proporcionan la visualización en coordenadas cartesianas, y *network-based*, que proporcionan visualización de los grupos de moléculas como gráficos (27). Todos estos métodos son generalmente aplicables a cualquier base de datos. Sin embargo, el método de elección puede variar dependiendo de las cualidades esperadas de la visualización, o bien, por la habilidad del método de proporcionar gráficos útiles.

Cuando se trabaja con más de 3 descriptores es común realizar una reducción de dimensiones a dos o tres, de forma que la visualización se facilite. Una de las técnicas más reportadas en la literatura que se fundamenta en la reducción de dimensiones es el análisis de componentes principales (PCA, por sus siglas en inglés) (27).

Mediante la localización de moléculas en el espacio se puede caracterizar la diversidad de un grupo de compuestos y de esta forma se pueden diseñar bibliotecas químicas, seleccionar o clasificar compuestos de acuerdo a los descriptores que sean utilizados. De igual forma, la visualización del espacio químico puede ser útil en el diseño de experimentos para proporcionar una cobertura de diversas estructuras, o en la visualización y análisis de relaciones estructura-actividad (*activity landscapes*) (31).

En esta sección se incluyen ejercicios enfocados en:

- Visualización y análisis del espacio químico de fármacos dependiendo su año de aprobación.
- Visualización y análisis del espacio químico de fármacos con diferente sitio de acción.
- Análisis del espacio químico de metabolitos de hongos.

### Capítulo 3. “Relaciones cuantitativas estructura-actividad”

El concepto de relaciones cuantitativas estructura-actividad es un término frecuentemente usado para describir al método QSAR (del inglés, *Quantitative structure–activity relationships*). No obstante, este concepto abarca otros métodos computacionales como son: métodos *rule based* (actividad en escala categórica (activo/inactivo; alto/medio/bajo)), métodos de *machine learning* y el modelado de panoramas de actividad (32).

Todos estos métodos han sido de gran utilidad para describir cambios en la actividad biológica asociados a cambios en las estructuras químicas y en el subsecuente diseño de nuevos compuestos con actividades mejoradas (32).

En general, se recomienda emplear métodos que permitan analizar el SAR de un grupo de compuestos antes de emplear métodos para predecir la actividad biológica. Con respecto a este punto, el modelado de panoramas de actividad es una herramienta muy utilizada.

Es por este motivo, que en este capítulo se busca ejemplificar el concepto de panoramas de actividad y mediante su implementación llevar a cabo la optimización de la actividad biológica de compuestos activos y la identificación sistemática de acantilados de actividad (en inglés, *activity cliffs*).

#### 3.1 Modelado de panoramas de actividad

Modelado de panoramas de actividad (en inglés, *activity landscape modeling*) es un concepto que surgió con el objetivo de describir en forma sistemática cambios en la actividad biológica asociados a cambios en las estructuras químicas. Este método es particularmente útil cuando se analizan grandes colecciones de compuestos (33).

Un panorama de actividad se puede definir como una representación que integra los análisis de similitud estructural y las diferencias en la potencia entre compuestos que comparten la misma actividad biológica (34).

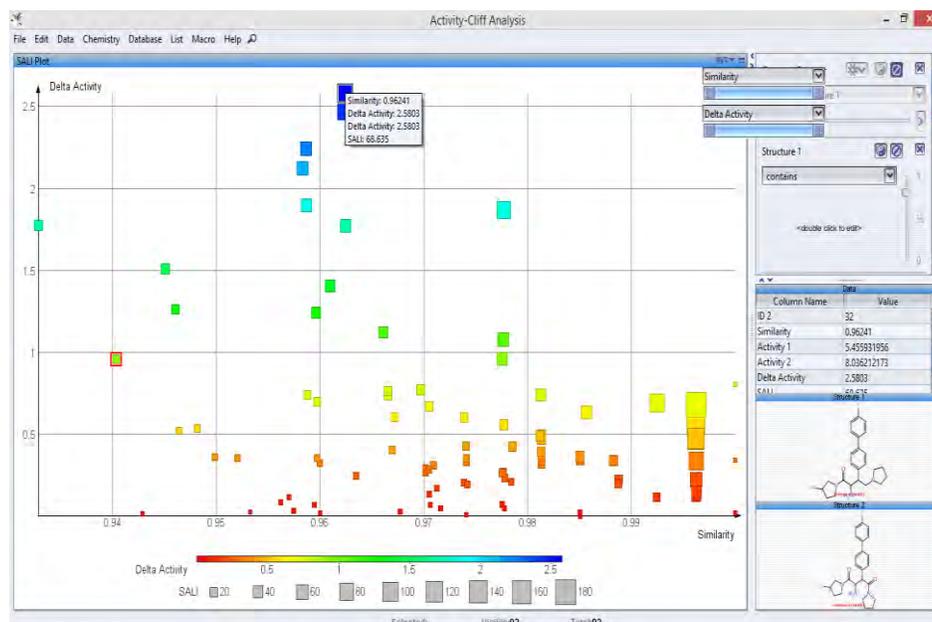
Desde la aparición de este concepto, diversos métodos se han desarrollado con la finalidad de comprender la información prevista en panoramas de actividad y llevar a cabo la

detección de *activity cliffs* (par de compuestos con estructura base muy similar pero una gran diferencia en potencia). Ejemplos de estos métodos se describen en la Tabla M3 (35).

**Tabla M3.** Métodos desarrollados en el modelado de panoramas de actividad y principales aplicaciones.

Método	Aplicaciones
SALI ( <i>Structure-activity landscape index</i> )	Detección de <i>activity cliffs</i>
SARI ( <i>Structure-ctivity relationship index</i> )	Caracterización global y local del SAR y detección de <i>activity cliffs</i>
NSG ( <i>Network-like similarity graphs</i> )	Exploración global y local del SAR. Detección de <i>activity</i> y <i>selectivity cliffs</i>
Mapas SAS ( <i>Structure Activity Similarity</i> )	Detección visual y cuantitativa de <i>activity</i> y <i>selectivity cliffs</i>
Panoramas de actividad 3D	Análisis del SAR considerando la conformación de las moléculas
Árboles de similitud-potencia ( <i>Similarity-potency tres</i> )	Análisis local del SAR

En la Figura M2 se observa un ejemplo de la aplicación del concepto de panoramas de actividad. En esta representación, se visualiza el SAR de inhibidores de la enzima dipeptidilpeptidasa IV integrando los mapas SAS y el índice de SALI.



**Figura M2.** Visualización del SAR de inhibidores de la enzima dipeptidilpeptidasa IV mediante el empleo de mapas SAS implementados en DataWarrior. Se grafica similitud vs diferencia en actividad. El tamaño de cada punto en el gráfico (par de compuestos) se distingue de acuerdo a su valor de SALI.

En esta representación, para cada comparación de pares de compuestos se determina: la similitud estructural, un delta de actividad y un el índice de SALI.

La **similitud estructural** se calcula mediante el coeficiente de Tanimoto. Este coeficiente establece que la similitud estructural  $T_{ij}$  de las moléculas  $i$  y  $j$  se calcula con la expresión:

$$T_{ij} = \frac{C_{ij}}{C_i + C_j - C_{ij}}$$

donde  $T_{ij}$  es la similitud molecular entre la  $i$ -ésima y la  $j$ -ésima molécula;  $C_{ij}$  es el número total de características en común;  $C_i$  es el número total de características en la  $i$ -ésima molécula y  $C_j$  es el número total de características en la  $j$ -ésima molécula.

La **delta de actividad** es la diferencia de la actividad que existe entre cada par de moléculas:

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

donde  $pIC_{50}(P)_i$  y  $pIC_{50}(P)_j$  son las actividades de la  $i$ -ésima y  $j$ -ésima molécula contra un blanco terapéutico. Para la construcción de mapas SAS es frecuente utilizar como índice de similitud el valor absoluto de la diferencia de actividad ( $|\Delta pIC_{50}(P)_{i,j}|$ ).

Finalmente, el **índice se SALI** (36) (*Structure-Activity Landscape Index*) se define como:

$$SALI = \frac{|A_i - A_j|}{1 - sim_{i,j}}$$

donde  $A_i$  y  $A_j$  son las actividades de la  $i$ -ésima y  $j$ -ésima molécula y  $sim_{i,j}$  es el coeficiente de similitud entre las dos moléculas.

Ejercicios planteados y desarrollados en este capítulo están enfocados en:

- Visualización del SAR de inhibidores de la enzima dipeptidilpeptidasa IV mediante mapas SAS e identificación sistemática de *activity cliffs*.
- Verificar que *activity cliffs* identificados en el punto anterior cumplan con criterios para ser considerados como tales (37).
- Estudio de panoramas de actividad utilizando la aplicación *Activity Landscape Plotter*, elaborada por el grupo DIFACQUIM (<https://unam-shiny-difacquim.shinyapps.io/ActLSmaps/>)
- Análisis de los panoramas de actividad de tres diferentes quimiotecas de compuestos evaluadas contra el transportador de noraepinefrina (38).
- Identificación de *activity cliffs* a partir de la estructura química 2D y de datos de diferencia de actividad, similitud molecular e índice de SALI (39).

---

## Capítulo 4. “Acoplamiento molecular”

En el diseño de nuevos compuestos con actividad terapéutica un error que puede ser común es no tomar en cuenta la bioquímica de la enfermedad (mecanismo de acción) y basarse exclusivamente en la estructura química de la molécula líder para diseñar compuestos análogos.

Si se dispone de información relacionada con el blanco terapéutico, el acoplamiento molecular es una técnica que puede ayudar a predecir el modo de unión de una molécula con una diana molecular, a comprender mejor el comportamiento de estos sistemas a nivel molecular y a establecer comparaciones entre teoría y datos experimentales. De igual forma, el acoplamiento molecular es una técnica muy poderosa para diseñar nuevas moléculas con afinidad por un determinado receptor.

En este capítulo, se enfatiza sobre los requisitos en cuanto a la presentación del ligando y de la diana biológica establecidos por diferentes herramientas para realizar acoplamiento molecular. Al realizar este estudio y mediante su interpretación se busca predecir y proponer uno o varios modelos de unión entre ligandos y el blanco molecular, de tal forma que se facilite la selección de compuestos para realizar ensayos biológicos y el diseño de nuevas estructuras químicas con interés terapéutico.

Para esto, el capítulo está dividido en 4 apartados que son: 4.1) Preparación de la diana biológica, 4.2) Preparación del ligando, 4.3) Acoplamiento molecular y 4.4) Análisis de resultados.

### 4.1 Preparación de la diana biológica

Las interacciones proteína-ligando tienen un papel fundamental en varias vías metabólicas. Esto hace que las interacciones ligando-proteína sean muy estudiadas para diseñar compuestos con acción farmacológica.

Para realizar un estudio computacional con proteínas, es indispensable que se corrijan los errores de la estructura y se prepare a la proteína de tal forma que este en las condiciones adecuadas para posteriores análisis. Por ejemplo, un problema con las estructuras cristalográficas de rayos X de proteínas es la falta de información acerca de los hidrógenos, por lo que se debe utilizar una herramienta que asigne estados de protonación e ionización

a la macromolécula de estudio. Asignaciones incorrectas pueden conducir a errores de acoplamiento.

Ejercicios propuestos en el Manual de Quimioinformática:

- Obtener información de proteínas con interés terapéutico mediante la búsqueda del código PDB.
- Identificación de criterios para seleccionar una proteína para estudios de acoplamiento molecular (40).

#### 4.2 Preparación del ligando

Los tipos de ligandos seleccionados para realizar un estudio de acoplamiento molecular deben estar relacionados con el objetivo de estudio. Según sea el caso, pueden aplicarse diferentes filtros para reducir en lo posible el número de compuestos a acoplar y de esta forma evitar un gasto innecesario de tiempo y recursos materiales. Por ejemplo, en el descubrimiento de prototipos se pueden aplicar filtros como carga neta, peso molecular, área polar superficial, solubilidad, disponibilidad comercial y precio por compuesto. En caso de optimización de prototipos pueden añadirse filtros como accesibilidad sintética y propiedades (ADME-Tox) (41).

Como requisitos para la estructura del ligando, la mayoría de las herramientas para realizar acoplamiento molecular destaca la necesidad de utilizar estructuras en 3D. Asimismo, para tener mejores predicciones de unión es importante asegurar que el estado de protonación, tautómeros y formas estereoisoméricas de los ligandos son correctas. Como parte de la preparación, los programas determinan los enlaces rotables del ligando para generar diferentes conformeros durante el acoplamiento molecular.

Para esta sección, los ejercicios planteados en el Manual de Quimioinformática están enfocados en:

- Aplicación de filtros para la identificación de nuevos compuestos líderes como potenciales fármacos antivirales.
- Adquirir librerías basadas en fragmentos y *scaffolds* para acoplamiento molecular.
- Identificación de modificaciones necesarias a implementar en compuestos para que puedan ser utilizados en estudios de acoplamiento molecular.

### 4.3 Acoplamiento molecular

El acoplamiento molecular es uno de los métodos basados en la estructura del receptor más empleados en el diseño de fármacos asistido por computadora (42).

Hasta el momento se han desarrollado programas basados en diferentes algoritmos para llevar a cabo estudios de acoplamiento molecular. El enfoque principal de estos programas es modelar la interacción entre una molécula pequeña y una macromolécula diana, permitiendo a su vez caracterizar el comportamiento de un ligando en el sitio de unión del receptor, así como elucidar procesos bioquímicos fundamentales.

Entre los métodos de acoplamiento molecular existen los que consideran tanto al ligando como al receptor cuerpos rígidos (modelo de la llave-cerradura de Fisher); los que consideran únicamente al ligando como cuerpo flexible y aquellos que consideran tanto al ligando como al receptor como cuerpos flexibles. Por las limitaciones de los recursos informáticos, el modelo más utilizado es el que considera al ligando flexible y al receptor como un cuerpo rígido (43).

Ejercicios planteados en el Manual de Quimioinformática:

- Realizar acoplamiento molecular a compuestos identificados como *activity cliffs* (apartado 3.1).
- Búsqueda del concepto de acoplamiento molecular inverso.
- Identificación de diferentes aplicaciones del acoplamiento molecular en el diseño de fármacos.

### 4.4 Análisis de resultados

Al finalizar el acoplamiento molecular como resultados se obtienen: a) la conformación calculada como la más favorable para el ligando en el sitio de unión y b) un valor de puntuación o *scoring* que es la cuantificación de la energía de interacción entre una conformación determinada del ligando obtenida en el paso anterior con el sitio de unión. El propósito de la función de *scoring* es distinguir los modos de unión correctos de modo de unión incorrectos, o ligandos de compuestos inactivos en un tiempo de cálculo razonable (43). Usualmente se selecciona como mejor resultado aquel que tiene preferencia por el modo de unión con el mejor puntaje y se considera a este como el más representativo.

En la actualidad, hay tres tipos o clases de funciones de puntuación disponibles: basadas en campos de fuerza de mecánica molecular (*molecular mechanics force field-based*), funciones empíricas y funciones de puntuación basadas en el conocimiento. La Tabla M4 resume la función y algunas limitaciones de estos tipos de tipos de puntuación.

**Tabla M4** Tipos de puntuaciones usadas con frecuencia en estudios de acoplamiento molecular.

Función	Función	Limitaciones
<b>Campos de fuerza mecánica molecular</b>	Estiman la energía de enlace por la suma de interacciones electrostáticas y de Van der Waals.	No incluyen términos de solvatación y entrópicos. En ocasiones la puntuación es proporcional al tamaño de ligando.
<b>Funciones empíricas</b>	La puntuación se ajusta a los datos experimentales, p.ej., afinidad de unión, descriptores de enlace de hidrógeno, interacciones estéricas, interacciones lipófilas, solvatación y efectos entrópicos.	Es incierta su capacidad para predecir la afinidad de unión de ligandos que son estructuralmente diferentes de los utilizados en el conjunto de entrenamiento.
<b>Funciones basadas en conocimiento</b>	Están diseñados para reproducir las estructuras experimentales en lugar de energías de enlace.	Su parametrización se basa en la información implícitamente codificada en conjuntos limitados de estructuras complejas de proteína-ligando, que pueden conducir a la diferenciación mínima de tipos de átomos.

Si bien, los métodos computacionales actuales son eficientes para obtener resultados de búsqueda confiable, aún carecen de eficiencia para hacer cálculos precisos del puntaje (*scoring*). Esto se debe al número elevado de aproximaciones que se hacen para hacer los cálculos en forma rápida.

Ejercicios planteados en esta sección:

- *Docking* y análisis de valores de puntuación de compuestos sintetizados como potenciales inhibidores de la transcriptasa inversa del virus del VIH. Se busca reducir el número de evaluaciones biológicas y priorizar compuestos a evaluar.
- Análisis de resultados de *interaction cliffs* (complejos ligando-receptor con alta similitud estructural y tipos de interacciones, pero con una gran diferencia en la potencia de los ligandos) (45).
- Análisis de los resultados de acoplamiento molecular obtenidos para la búsqueda de nuevos núcleos estructurales contra el cáncer colorrectal (46).

---

## Capítulo 5.” Toxicología informática”

Si bien, la afinidad de un ligando con su receptor es una de las características más importantes en el desarrollo de un fármaco, existe una progresiva conciencia acerca de la necesidad de evaluar *a priori* las propiedades farmacocinéticas. Esto con el fin de excluir aquellos compuestos que puedan tener problema de biodisponibilidad, una metabolización inadecuada o bien dar lugar a problemas de toxicidad. En este sentido, disponer de bases de datos y programas que permitan focalizar el esfuerzo exclusivamente en compuestos con buenas cualidades es de vital importancia de cara a reducir el enorme costo asociado al desarrollo de un nuevo medicamento.

Métodos desarrollados dentro de toxicología informática, aunque deben ser tomados con precaución, actualmente proporcionan una indicación rápida de la eficacia y los riesgos potenciales de compuestos químicos.

En este capítulo se abordan diferentes bases de datos con información relacionada a la toxicología de sustancias químicas. Mediante el uso de páginas de libre acceso se busca identificar compuestos con bajo riesgo toxicológico y seleccionar candidatos para la optimización y desarrollo de fármacos potenciales.

### 5.1 Minería de datos

Una prioridad fundamental en los programas de desarrollo de fármacos y en la evaluación de su seguridad llevada a cabo por las autoridades reguladoras sigue siendo la identificación temprana de problemas toxicológicos graves.

Actualmente, métodos *in silico* de toxicología (también conocidos como toxicología informática) están mostrando utilidad para generar información valiosa en las etapas tempranas de diseño de fármacos. Entre la gama de herramientas computacionales que emplea la toxicología informática se encuentran las bases de datos con información toxicológica. Estas bases de datos son útiles para dar seguimiento al uso de fármacos y para elaborar modelos de toxicidad mediante estudios de QSAR, esto por mencionar algunos ejemplos.

La Tabla M5 resume ejemplos de bases de datos de toxicidad disponibles públicamente, útiles en la evaluación y caracterización de riesgos asociados al uso de sustancias químicas.

**Tabla M5** Principales bases de datos utilizadas en Toxicología Informática (consultar una lista más amplia en: *Toxicol. Appl. Pharmacol.*, 2009, 241:356-370).

Base de datos	Descripción	URL
AERS ( <i>Adverse Event Reporting System</i> )	Notificaciones de eventos adversos post-mercado de fármacos aprobados y productos biológicos con actividad terapéutica.	<a href="http://www.fda.gov/Drugs/">http://www.fda.gov/Drugs/</a>
AcTOR ( <i>Aggregated Computational Toxicology Resource</i> )	Datos de toxicidad química que se pueden utilizar para encontrar posibles riesgos químicos para la salud humana y el medio ambiente. Proporcionada por la EPA (Agencia de Protección Ambiental de EE.UU.).	<a href="http://actor.epa.gov/actor/faces/ACToRHome.jsp">http://actor.epa.gov/actor/faces/ACToRHome.jsp</a>
DSSTox ( <i>Distributed Structure-Searchable Toxicity</i> )	Contiene información que ayuda a construir bases de datos públicas para mejorar la estructura-actividad y capacidad de predicción en toxicología. Proyecto del Centro Nacional de Toxicología Computacional de la EPA.	<a href="http://www.epa.gov/ncct/dssto/DataFiles.html">http://www.epa.gov/ncct/dssto/DataFiles.html</a>
ToxNet	Acceso a bases de datos sobre toxicología, productos químicos peligrosos, salud ambiental y emisiones tóxicas, Proporcionada por <i>U.S National Library of Medicine</i> .	<a href="https://toxnet.nlm.nih.gov/">https://toxnet.nlm.nih.gov/</a>
CCRIS	Contiene registros químicos con efectos carcinogénicos, mutagénicos, promoción de tumores y resultados de las pruebas de inhibición de tumores. Fue desarrollada por el Instituto Nacional del Cáncer (NCI)	<a href="http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS">http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS</a>

Ejercicios desarrollados en el Manual de Quimioinformática:

- Buscar información de compuestos aprobados para uso clínico y su implicación en procesos de mutagénesis, promoción de tumores, carcinogénesis y teratogénesis.
- Identificación de causas frecuentes por las que fármacos han sido retirados del mercado.
- Buscar información toxicológica de fármacos a partir de diferentes formas de representación (nombre común, SMILES y estructura 2D).

## 5.2 Introducción a métodos computacionales aplicados en la predicción de propiedades ADME-Tox

Modelos computacionales de farmacocinética, comúnmente referidos como ADME/Tox, pueden ser introducidos en muchas de las etapas del proceso de desarrollo de fármacos. Una de las principales aplicaciones de estos modelos es la predicción de propiedades que permiten priorizar compuestos para síntesis y posteriores evaluaciones biológicas.

Propiedades ADME importantes a considerar para conseguir la exposición sistémica de un fármaco oral son: solubilidad acuosa, permeabilidad intestinal y la unión a proteínas plasmáticas.

El metabolismo y la eliminación del fármaco también determinan el éxito de un medicamento ya que estas propiedades tienen un impacto mayor en la biodisponibilidad y seguridad de un fármaco (48). La Tabla M6 resume ejemplos de programas para predecir propiedades ADME/Tox.

**Tabla M6** Programas de cómputo y sitios web disponibles para realizar predicciones ADME/Tox.

Programa	Descripción	URL
Molinspiration	Calcula propiedades moleculares importantes; predice la puntuación de la bioactividad de los objetivos de medicamentos más importantes (ligandos de GPCRs, inhibidores de quinasa, moduladores de canales de iones y receptores nucleares).	<a href="http://www.molinspiration.com/">www.molinspiration.com/</a>
Admet SAR	Predice propiedades ADMET de candidatos a fármacos y químicos que juega un papel importante en el descubrimiento de fármacos y la evaluación de los peligros al medio ambiente.	<a href="http://lmmmd.ecust.edu.cn:8000/predict/">http://lmmmd.ecust.edu.cn:8000/predict/</a>
ToxCreate	Crea modelos computacionales que predicen la toxicidad de compuestos. Proporcionada por OpenTox.	<a href="http://www.toxcreate.net/create">www.toxcreate.net/create</a>
Toxtree	Estima riesgos de toxicidad mediante métodos de clasificación.	<a href="http://toxtree.sourceforge.net/">http://toxtree.sourceforge.net/</a>
Pre- ADMET	Predice datos ADME, riesgos de toxicidad y la construcción de biblioteca de compuestos químicos.	<a href="https://preadmet.bmdrc.kr">https://preadmet.bmdrc.kr</a>
SwissADME	Permite calcular descriptores fisicoquímicos, así como predecir las propiedades farmacocinéticas y la naturaleza <i>druglike</i> de una o varias moléculas pequeñas. Desarrollado y mantenido por el grupo de Modelado Molecular de SIB ( <i>Swiss Institute of Bioinformatics</i> ).	<a href="http://www.swissadme.ch/">www.swissadme.ch/</a>
MetaPrint2D	Predice el metabolismo de xenobióticos por medio de la extracción de datos y el análisis estadístico de las transformaciones metabólicas conocidas y reportadas en la literatura. Proporcionada por la Universidad de Cambridge.	<a href="http://www-metaprint2d.ch.cam.ac.uk/metaprint2d-react/">http://www-metaprint2d.ch.cam.ac.uk/metaprint2d-react/</a>
SMARTCyp Web Service	Predice los sitios que son más susceptibles de metabolismo mediado por el citocromo P450. Proporcionada por el Departamento de Química Médica de la Universidad de Copenhague.	<a href="http://www.farma.ku.dk/smartcyp/">www.farma.ku.dk/smartcyp/</a>
PPS	Predice las vías posibles para la degradación microbiana de compuestos químicos. Mantenido por la Universidad de Minnesota.	<a href="http://eawag-bbd.ethz.ch/predict/">eawag-bbd.ethz.ch/predict/</a>

Ejercicios desarrollados en el Manual de Quimioinformática:

- Aplicar reglas empíricas para la predicción de la biodisponibilidad oral y toxicidad a un grupo de compuestos diseñados para inhibir p38 $\alpha$  MAPK, diana terapéutica atractiva para el tratamiento de enfermedades como artritis reumatoide y para la enfermedad de Crohn.

- Analizar los tipos de reacciones metabólicas de fase I y II e identificar los sitios más probables para que ocurran estas reacciones para algunos de los compuestos del ejercicio anterior.
- Identificar sitios que son propensos a sufrir reacciones metabólicas.
- Identificar alertas estructurales.
- Identificar compuestos más susceptibles a metabolismo de fase II.

## Ejercicios finales

Contando con información suficiente, se pueden aplicar una o varias de las técnicas mencionadas en los capítulos anteriores aunado a métodos experimentales. Esto es común en investigación en donde diferentes grupos colaboran en la resolución de un determinado problema.

Por este motivo, en este apartado del Manual de Quimioinformática, se busca integrar varios de los conceptos vistos en los capítulos anteriores y analizar artículos en donde se empleen uno o más métodos computacionales.

Para esto, se plantean los siguientes ejercicios:

- Relación de diversos conceptos abordados en el Manual con su definición.
- Aplicación de conceptos de espacio químico, panoramas de actividad y acoplamiento molecular, esto mediante el análisis de resultados del artículo: Naveja JJ, Medina-Franco JL. *RSC Advances*. 2015;5:63882-63895 (49).
- Aplicación de conceptos de descriptores moleculares y toxicología informática, esto mediante el análisis de resultados del artículo: Lan P, et al. *International Journal of Molecular Sciences*. 2010;11:3357 (50).

## 7. CASO DE ESTUDIO: Estudio quimioinformático del REmA

En este capítulo se desarrolla un caso de estudio particular en donde se integran y aplican diversos conceptos del Manual de Quimioinformática presentado en el capítulo anterior de la tesis. El caso de estudio está enfocado en caracterizar Relaciones Estructura-múltiple Actividad (REmA) de una serie de moléculas contenidas en una patente con actividad reportada con una serie de dianas moleculares epigenéticas de relevancia en la actualidad: histonas deacetilasas (HDACs). Hoy en día los estudios REmA tienen un gran impacto en el desarrollo de fármacos ya que forman parte de un cambio hacia un paradigma actual en el diseño de compuestos bioactivos: de una diana contra varias dianas biológicas en forma simultánea. REmA está asociado al término SMART (del inglés *Structure Multiple-Activity Relationships* (51).

### ***Modelado de panoramas de actividad a un grupo de moléculas epigenéticas con actividad reportada contra múltiples HDACs.***

#### **Resumen del caso de estudio**

La cantidad de información REmA ha aumentado notablemente en los últimos años debido a la creciente importancia del concepto de polifarmacología. La gran cantidad de información, que está en línea con el concepto actual de ‘*big data*’, requiere la aplicación de métodos computacionales para el análisis eficiente de datos. En este trabajo se aplicó el concepto de modelado de panoramas de actividad para analizar de manera sistemática el SAR (*Structure Activity- Relationships*) y el REmA de un grupo de compuestos derivados de hidroxipirimidina, reportados como inhibidores de HDACs. Específicamente se aplicaron los mapas SAS (*Structure-Activity Similarity*) y DAD (*Dual Activity Difference*) como parte del modelado de panoramas de actividad. El concepto de ‘limpieza del panorama de actividad’ se incorporó para el análisis de los compuestos generadores de acantilados de actividad (*activity cliffs generators*). Estudios de acopamiento molecular se emplearon con el fin de explicar diferencias en potencias en compuestos identificados en este trabajo como *activity cliffs*.

Finalmente, se modelaron propiedades ADMET a los compuestos incluidos en este estudio para así evaluar propiedades fisicoquímicas y farmacocinéticas de relevancia.

Las técnicas descritas en el Manual de Quimioinformática y que se aplicaron en este caso de estudio fueron las siguientes:

- Minería de datos
- Construcción de base de datos (a partir de una patente)
- Visualización de espacio químico
- Panoramas de actividad
- Acoplamiento molecular (con HDAC6)
- Toxicología informática (predicción de propiedades ADMET de inhibidores de HDACs)

### ***Histonas deacetilasas (HDACs)***

La acetilación de residuos de *N*-acetil lisina es una modificación post-traducciona reversible común en proteínas que está controlada por dos familias de enzimas con acciones opuestas, las histonas acetiltransferasas (HAT) y las histonas deacetilasas (HDAC) (52). Las histonas deacetilasas (HDACs) son enzimas encargadas de remover los grupos acetil de residuos de lisina dentro de una variedad de proteínas incluidos factores de transcripción e histonas (53), por lo que están involucradas en múltiples funciones celulares que incluyen regulación transcripcional, progresión del ciclo celular y diferenciación celular (54).

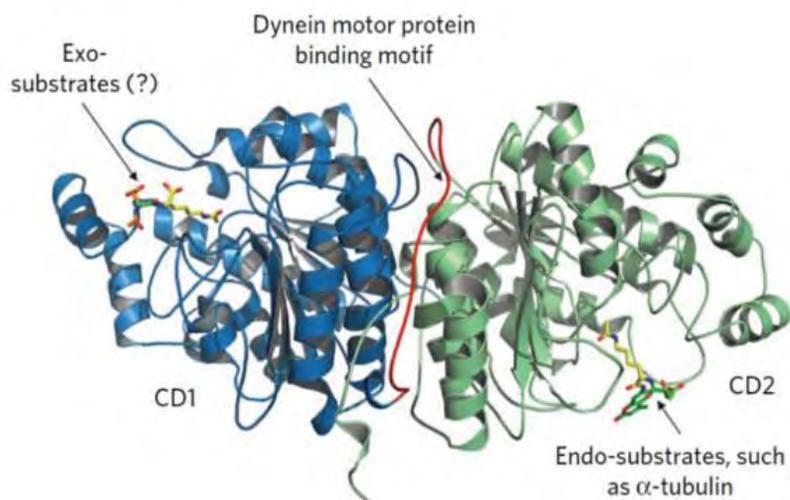
Las histonas deacetilasas dependientes de  $Zn^{2+}$  pueden clasificarse en tres clases dependiendo la homología en su secuencia. La clase I comprende las isoformas 1, 2, 3 y 8. La clase II puede ser subdividida en clase IIa e incluye las isoformas 4, 5, 7 y 9; la clase IIb se compone de las isoformas 6 y 10. Por otra parte, HDAC11 es el único miembro de la clase IV y comparte similitud en la secuencia con la clase I y II de HDACs (52).

La rápida aparición de información relacionada con actividades específicas para cada una de las once isoformas de HDACs y datos clínicos de toxicidad han llevado al diseño de compuestos cada vez más selectivos para cada una de las isoformas de HDAC.

En este sentido, HDAC6, fue la primer isoforma identificada y caracterizada como tubulina deacetilasa (55). Mientras que los sustratos de la mayoría de HDAC son histonas, HDAC6

es la única isoforma que se localiza principalmente en el citoplasma y que desacetila proteínas no histonas, tales como  $\alpha$ -tubulina de microtúbulos. HDAC6 es también inusual porque contiene dos dominios catalíticos desacetilasa y un dominio de unión a ubiquitina que reconoce agregados de proteínas (Figura C1) (56, 57).

Actualmente, existen tres áreas terapéuticas generales que parecen ser más susceptibles a alteraciones en la actividad de HDAC6. Estos son: trastornos autoinmunes, enfermedades neurodegenerativas y ciertos tipos de cáncer (53). La ventaja de usar inhibidores específicos para HDAC6 para el tratamiento de trastornos autoinmunes y condiciones neurodegenerativas radica en que los inhibidores de HDAC6 no parecen ser citotóxicos hacia las células normales lo que es positivo para el tratamiento de enfermedades no oncológicas (57, 58).

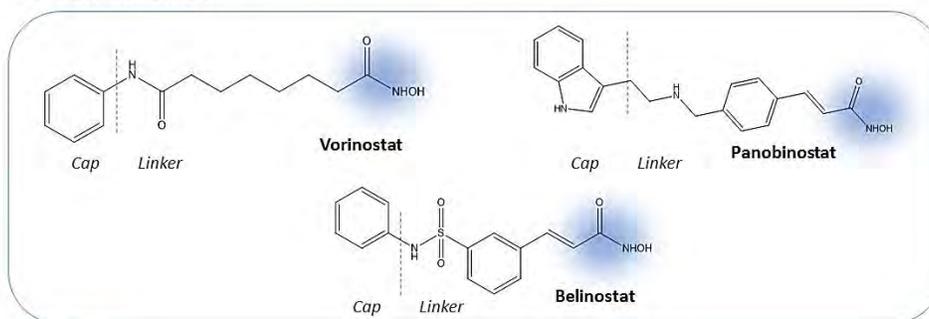


**Figura C1.** Estructura de la enzima HDAC6. Se muestran los dos dominios catalíticos de HDAC6 y el motivo de unión. Fuente: *Nat. Chem. Biol.* 2016, 12:660–661.

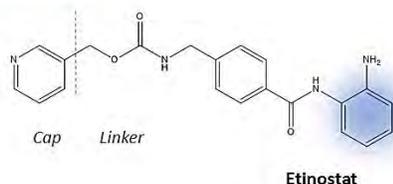
### ***Inhibidores de HDACs***

Con información de la relación entre la expresión elevada de HDACs y la proliferación de diversas células tumorales, varios inhibidores de HDACs como vorinostat [SAHA], romidepsina, belinostat, panobinostat y chidamida se han aprobado para el tratamiento del linfoma cutáneo y periférico de células T, y el mieloma múltiple (59, 60).

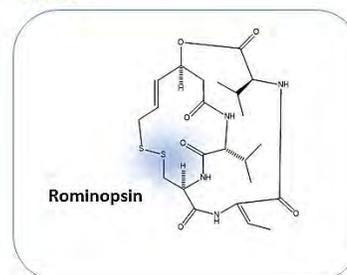
## Ácidos hidroxámicos



## Benzamidas



## Tioles



Región de unión al Zinc

**Figura C2.** Compuestos inhibidores de HDACs. En el recuadro se indican compuestos aprobados por la FDA para uso clínico.

La Figura C2 muestra que, en general, los inhibidores de HDACs tienen tres regiones: una zona *cap* o "dominio de reconocimiento superficial", que incluye la entrada de la bolsa del sitio activo; un "grupo de unión al zinc", que es requerido para la función catalítica; y una región *linker* o "enlazadora", que conecta las dos regiones anteriores. Históricamente, la mayoría de los inhibidores HDACs han sido ácidos hidroxámicos. Sin embargo, ahora se usan otros grupos de unión al zinc para aumentar la especificidad y la selectividad, incluyendo derivados de benzamida, tioles, sulfamidas, cetonas y tritiocarbonatos (61).

Los derivados de ácido hidroxámico (por ejemplo, vorinostat y panobinostat) inhiben generalmente HDACs de varias clases, mientras que los derivados de benzamida (etinostat y mocetinoestat) generalmente se limitan a la inhibición de HDACs de clase I. Por otra parte, la región *cap* parece ser responsable de la selectividad por lo que podría usarse para diseñar inhibidores que se dirijan a HDACs que residen en complejos específicos (60, 62).

## ***Métodos quimioinformáticos aplicados en el estudio del REmA***

Enfoques quimioinformáticos utilizados con frecuencia para conocer asociaciones subyacentes entre estructuras químicas y la actividad biológica son: análisis de espacio químico, modelado de panoramas de actividad y acoplamiento molecular. Estos métodos pueden ser aplicados para el estudio del REmA.

### *Espacio químico*

El concepto de espacio químico se ha utilizado ampliamente en el descubrimiento de fármacos y se utiliza cada vez más en otros campos. Empleando este concepto es posible comparar un conjunto de compuestos para diseñar nuevas colecciones y/o seleccionar bibliotecas de compuestos existentes para el cribado experimental o computacional, clasificar compuestos bioactivos de acuerdo con sus propiedades biológicas y dirigir la síntesis de nuevos compuestos de acuerdo al espacio identificado como activo (63).

Recientemente, el grupo de investigación DIFACQUIM ha introducido el concepto de espacio químico en el estudio de panoramas de actividad para llevar a cabo la limpieza del SAR de un grupo de compuestos inhibidores de DNMT. Esto facilita el análisis visual de los mapas SAS (39). Se espera que la implementación del análisis del espacio químico sea también de gran utilidad en el análisis del REmA.

### *Panoramas de actividad*

Un panorama de actividad es un estudio que integra los análisis de similitud estructural y las diferencias en la potencia entre compuestos que comparten la misma actividad biológica (64). Los mapas SAS y DAD son ejemplos del modelado de panoramas de actividad que permiten caracterizar el SAR y el REmA, respectivamente. Estos mapas ayudan en la detección de *activity cliffs* (compuestos con alta similitud estructural, pero con gran diferencia en actividad).

En este sentido, cabe mencionar que los *activity cliffs* toman cada vez mayor relevancia en la química farmacéutica, ya que su identificación permite construir modelos con mejor capacidad predictiva (QSAR por ejemplo). También permiten identificar elementos estructurales que rigen la actividad biológica y, en algunos casos, selectividad (65, 66).

### *Acoplamiento molecular*

El acoplamiento molecular automatizado es uno de los métodos más empleados para entender el modo de acción de moléculas activas a nivel molecular. Este método tiene como finalidad buscar la conformación y posición óptima de un ligando dentro de un blanco molecular (44). En el diseño de fármacos el acoplamiento molecular suele utilizarse con tres objetivos: 1) encontrar nuevos compuestos líder para su posterior optimización; 2) identificar compuestos selectivos; y 3) optimizar la actividad biológica de compuestos activos (2).

En el estudio del SAR y REmA este método se ha aplicado para ayudar a explicar a nivel molecular diferencias significativas en potencia entre compuestos que son estructuralmente muy similares (49). En este mismo sentido, se han empleado cálculos de acoplamiento molecular para estudiar compuestos generadores de *activity cliffs* y, de esta forma, sugerir puntos clave (*hot spots*) en la proteína objetivo responsable de la formación de *activity cliffs* y que a su vez permitan analizar su implicación en la interacción ligando-enzima (68).

## Objetivos del caso de estudio

### ➤ General

Estudiar relaciones estructura-múltiple actividad (REmA) de grupos de compuestos evaluados contra dos o más blancos terapéuticos, específicamente con una familia de dianas epigenéticas.

### ➤ Particulares

- Identificar *activity cliffs* en compuestos evaluados con las isoformas HDAC1, HDAC2, HDAC3 y HDAC6.
- Adaptar el concepto de 'limpieza del panorama de actividad', que está basado en el espacio químico, para explicar el comportamiento de los compuestos generadores de actividad.
- Explicar a nivel molecular el comportamiento de *activity cliffs* identificados en este trabajo mediante acoplamiento molecular.
- Modelar propiedades ADMET a los compuestos incluidos en este estudio.

## Metodología

En el caso de estudio se caracterizó el SAR y el REmA de 140 inhibidores de HDACs siguiendo tres etapas: a) Caracterización de la diversidad estructural y distribución en el espacio químico de los 140 compuestos; b) Análisis descriptivo del SAR y del REmA utilizando el concepto de modelado de panoramas de actividad, y c) Interpretación a nivel estructural de *activity cliffs* mediante acoplamiento molecular.

Además, se evaluaron propiedades ADMET de los compuestos analizados en este caso de estudio. Esto se hizo porque las moléculas utilizadas para inhibir HDACs en general se metabolizan rápidamente y porque es limitada la información de toxicidad de inhibidores de HDAC de la clase II (57, 69).

### *Base de datos*

Los compuestos utilizados en este caso de estudio fueron identificados en *Binding Database* y forman parte de un programa de optimización para obtener compuestos selectivos para HDAC6 (70).

Para plantear el caso de estudio se buscó que los compuestos tuvieran actividades reportadas para más de una diana biológica o ensayo farmacológico y que la actividad estuviera reportada de preferencia en unidades nano molar (37). Los 140 compuestos derivados de hidroxipirimidinas presentan diferentes estructuras químicas y actividades reportadas contra HDAC1, HDAC2, HDAC3 y HDAC6.

Debido a que las estructuras de las moléculas no se encontraban disponibles en un formato capaz de ser importado a otros sistemas de edición molecular, se usó la interfaz web de OSRA (*Optical Structure Recognition Application*) para convertir las representaciones gráficas de las estructuras publicadas en formato SMILES (<https://cactus.nci.nih.gov/cgi-bin/osra/index.cgi>). Los valores de actividad en pIC<sub>50</sub> y las estructuras químicas de los compuestos se encuentran en el Apéndice A.

### *Similitud estructural*

La similitud estructural de cada compuesto se calculó con *Extended Connectivity Fingerprints* (ECFP) implementada en MayaChemTools (71). La similitud estructural entre cada par de compuestos (7930 pares) fue calculada empleando el coeficiente de Tanimoto (72). Este coeficiente establece que para las moléculas C<sub>i</sub> y C<sub>j</sub>, la similitud estructural por pares (T<sub>ij</sub>) es:

$$T_{ij} = \frac{C_{ij}}{C_i + C_j - C_{ij}}$$

donde T<sub>ij</sub> es la similitud molecular entre la i-ésima y la j-ésima molécula; C<sub>ij</sub> es el número total de características en común; C<sub>i</sub> es el número total de características en la i-ésima molécula y C<sub>j</sub> es el número total de características en la j-ésima molécula.

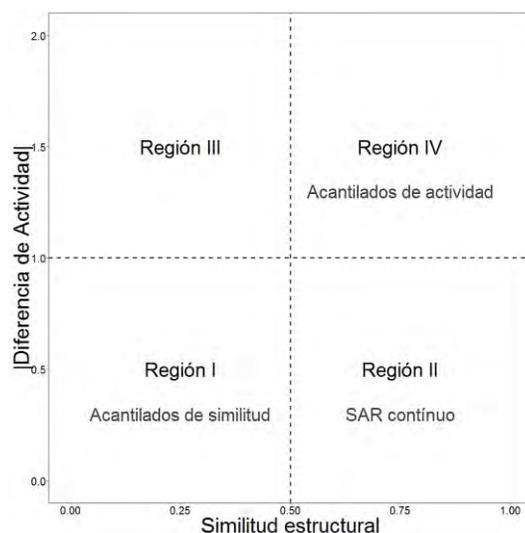
### *Espacio químico*

El espacio químico de los 140 inhibidores de HDACs se visualizó por medio del PCA de la matriz de similitud ECFP4. El análisis PCA se realizó en DataWarrior y el gráfico obtenido se construyó en RStudio. Se indicó con escala de color el valor de  $pIC_{50}$  de los compuestos evaluados con HDAC6.

### *Mapas SAS*

La visualización del SAR de los inhibidores de HDACs se hizo con mapas SAS (*Structure-Activity Similarity*). Estos mapas se construyeron al graficar el valor absoluto de la diferencia de actividad contra la similitud molecular de los 7930 pares de compuestos. La Figura C3 muestra la forma general de un mapa SAS. El gráfico se divide en cuatro regiones utilizando como una unidad logarítmica el límite de la diferencia de actividad y el valor de la mediana de similitud estructural de todos los pares de compuestos como límite de similitud estructural. Los pares de compuestos en la región I (acantilados de similitud), se caracterizan por presentar valores de similitud estructural y diferencia de actividad bajos. En bases de datos con gran diversidad en *scaffolds* esta región toma relevancia, ya que puede sugerir diversas formas de unión entre los ligandos con la diana molecular evaluada. En la región II se encuentran pares de compuestos con un SAR continuo, es decir, compuestos con baja diferencia en actividad y un valor de similitud estructural alto. Esta región puede ser útil en la construcción de estudios QSAR. Por otra parte, la región III se considera poco informativa, ya que es en esta región donde se encuentran los compuestos con alta diferencia en actividad y poca similitud estructural. En la región IV se encuentran los compuestos con un valor de similitud y diferencia de actividad alto, es decir, a los acantilados de actividad (*activity cliffs*).

La identificación de los acantilados de actividad es importante en la construcción de modelos predictivos y para identificar cambios estructurales que rigen la actividad biológica.



**Figura C3.** Forma general de un mapa de similitud estructura-actividad (SAS) que muestra cuatro regiones principales. Las regiones I y II están asociadas con acantilados de similitud y SAR continuo, respectivamente. La región III se considera poco informativa, mientras que la región IV indica un SAR discontinuo asociado a la presencia de acantilados de actividad o *activity cliffs*.

### Generadores de 'activity cliffs'

Se cuantificó el número de *activity cliffs* generados por cada compuesto, es decir, la frecuencia con la que aparecen en la región IV de los mapas SAS. Los compuestos generadores de *activity cliffs* se definieron como aquellos compuestos en la que su frecuencia de formar *activity cliffs* fue superior al valor promedio más dos desviaciones estándar (68). Además, se construyó un gráfico con las frecuencias acumuladas de cada isoforma de HDAC, en donde los compuestos fueron separados en dos grupos de acuerdo a los resultados obtenidos de la visualización del espacio químico.

### Mapas DAD

La caracterización del REmA se llevó a cabo mediante el empleo de los mapas DAD. Estos mapas se basan en las diferencias de la actividad entre pares de compuestos contra dos dianas biológicas y se construyen trazando la diferencia de actividad para una diana biológica (A) vs. la diferencia de actividad con una segunda diana biológica (B).

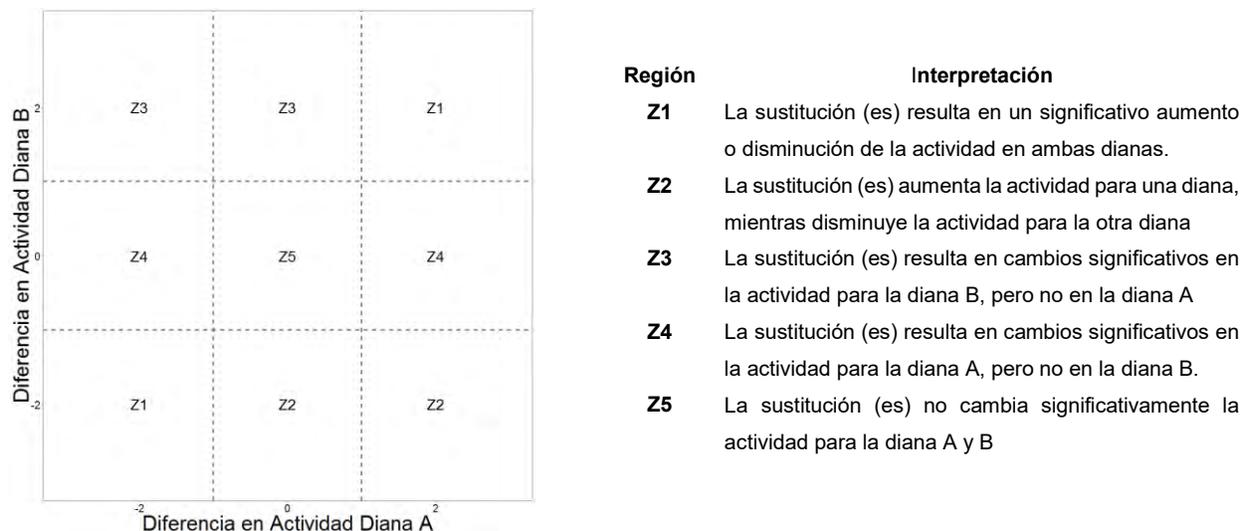
La Figura C4 es un prototipo de mapa DAD. Líneas verticales y horizontales en cada uno de los ejes determinan los límites para una baja o alta deferencia de actividad para las dianas A y B, respectivamente. Se establece como límite 1 unidad logarítmica en

diferencia de actividad, de modo que los pares de compuestos con una diferencia de potencia baja son los que su diferencia de actividad es menor a una unidad logarítmica.

Los límites establecidos dan lugar a cinco zonas generales: Z1-Z5. Si los cambios estructurales entre dos compuestos afectan con una proporción similar la actividad biológica contra las dianas A y B (ya sea un aumento o disminución), este par de compuesto se localizará en la región Z1. Por otra parte, los pares de compuestos situados en Z2 son aquellos en los que el cambio estructural aumenta la actividad hacia una de las dianas, pero disminuye la actividad hacia la otra (*activity switches*); por lo tanto, Z2 está asociado con un "SAR inverso" y es el área más importante para estudios de selectividad. Los pares de compuestos en Z3 o Z4 indican compuestos en el que las diferencias estructurales sólo afectan a la actividad contra una de las dos dianas.

Los puntos de datos en Z5 están asociados con pares de compuestos en los que cambios estructurales tienen poco o ningún impacto en la actividad contra las dos dianas biológicas evaluadas.

Dado que los mapas DAD se basan en comparaciones pareadas, es fácil incorporar la información de la similitud de la estructura (73). Aquí se representa la similitud de la estructura utilizando un color continuo, escala de gris (baja similitud) a morado (alta similitud).



**Figura C4.** Forma general de un mapa DAD (*Dual Activity Difference*) que muestra cinco regiones principales Z1-Z5 al establecer como límite de diferencia de actividad una unidad logarítmica.

### *Acoplamiento molecular*

Para explicar a nivel molecular los *activity cliffs* en HDAC6 se utilizó el dominio catalítico II de la proteína PDB ID: 5EDU, ya que se tiene reportado que es este el dominio funcional en esta proteína y en donde inhibidores tienden a unirse (62).

El dominio catalítico II de HDAC6 se preparó en MOE (21). En resumen, se llevó a cabo la minimización energética y se removieron las moléculas de agua cristalográficas, manteniendo el ión  $Zn^{2+}$  en el sitio catalítico. Se utilizaron los algoritmos implementados en MOE y se realizó la validación correspondiente.

Una vez validado el procedimiento, se procedió a realizar el acoplamiento con los compuestos identificados como *activity cliffs* más representativos en HDAC6 (con valores de similitud superiores al valor de la mediana y con una diferencia de actividad mayor a una unidad logarítmica).

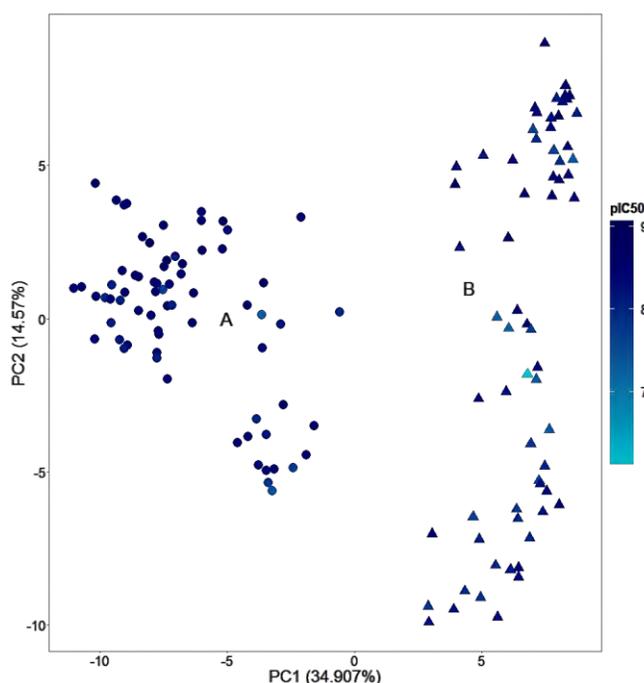
### *Predicción de propiedades ADMET*

Además del estudio del REmA, se realizó la predicción de propiedades ADMET para los compuestos del caso de estudio. Para esto, se utilizó el programa Qikprop (74) y se calcularon descriptores físicoquímicos significativos (log P, peso molecular, HBA, HBD), así como propiedades farmacéuticamente relevantes (absorción oral, número de reacciones metabólicas y unión a proteínas séricas).

## Resultados y discusión del caso de estudio

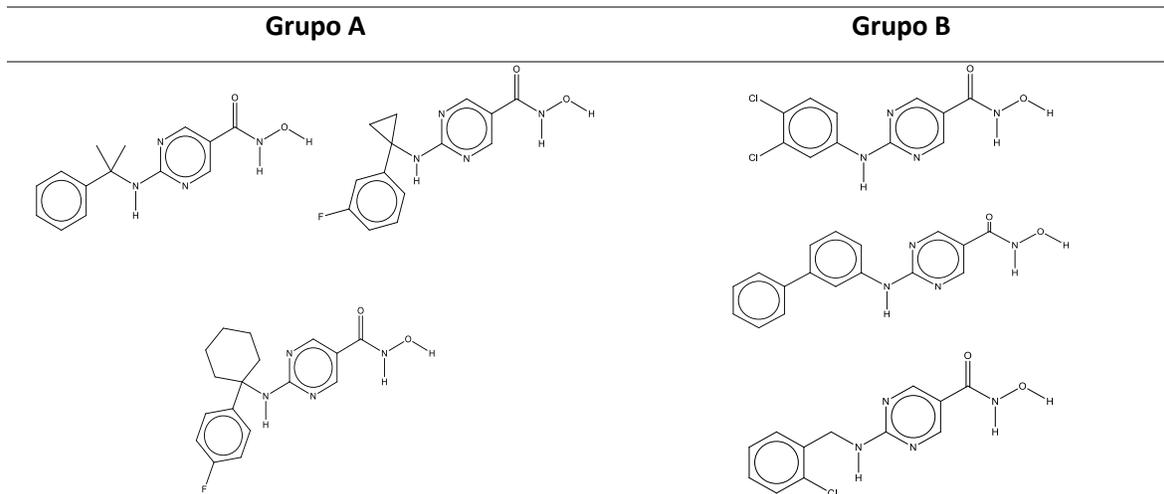
### a) Caracterización de la diversidad estructural y distribución en el espacio químico

Con la finalidad de incorporar el concepto de “limpieza de panoramas de actividad” en el análisis del REmA se exploró la distribución de los 140 compuestos inhibidores de HDACs en el espacio químico. La Figura C5 muestra la caracterización de este espacio químico obtenido a partir del PCA de la matriz de similitud con ECFP. En los ejes se indica las contribuciones porcentuales de los dos primeros componentes, que sumados capturan 49.48% de la covarianza. En esta representación visual del espacio químico se distinguen dos grupos principales, cada uno con 70 compuestos. La visualización sugiere que las moléculas del grupo B son más diversas entre sí que las moléculas del grupo A (28).



**Figura C5.** Espacio químico los 140 compuestos inhibidores de HDAC obtenido mediante un análisis de componentes principales de la matriz de similitud calculada con ECFP4/Tanimoto. La actividad biológica contra HDAC6 se muestra en una escala continua de color. Se distinguen dos grupos de compuestos (A y B).

La Figura 6C muestra ejemplos de estructuras de cada grupo. La diferencia principal entre ambos grupos es la sustitución del carbono unido en la región del 2-amino de la hidroxipirimidina. Para los compuestos del grupo A este carbono es cuaternario, mientras que para los compuestos del grupo B es secundario o aromático.

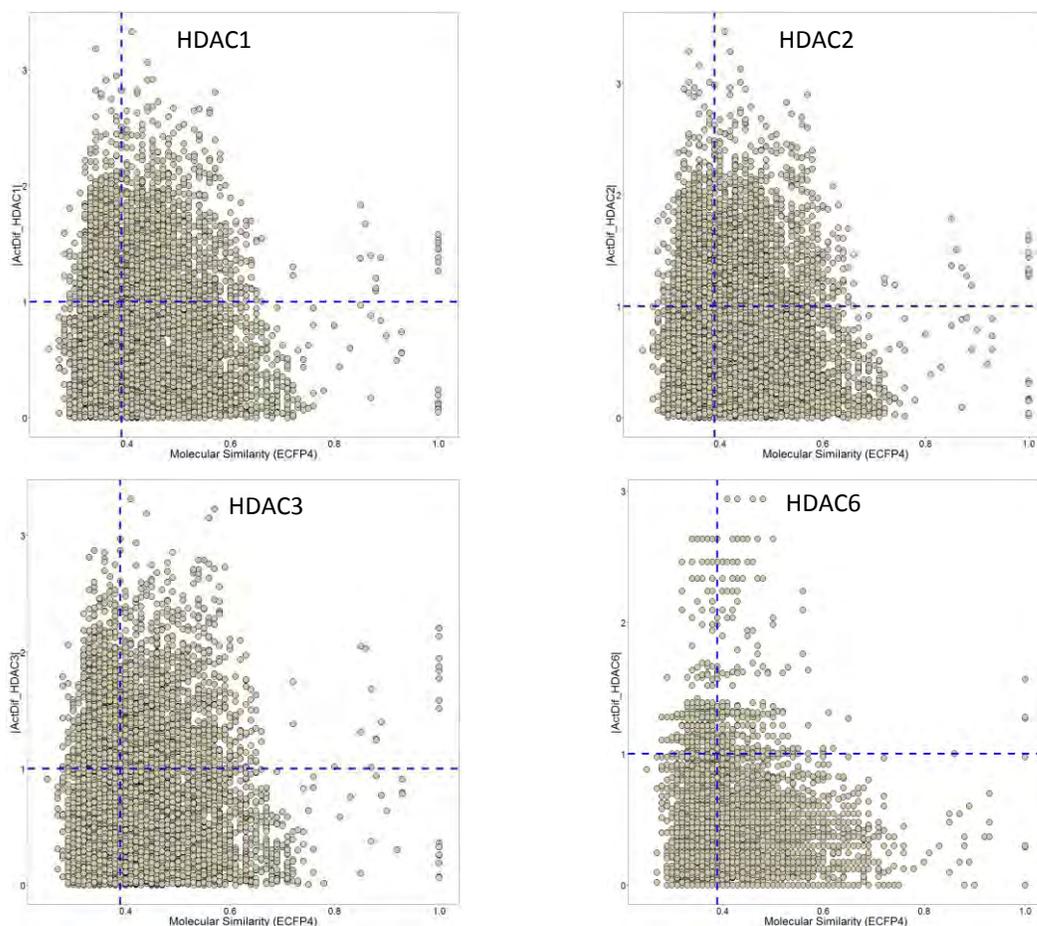


**Figura C6.** Ejemplo de estructuras presentes en cada uno de los grupos identificados por medio del análisis del espacio químico.

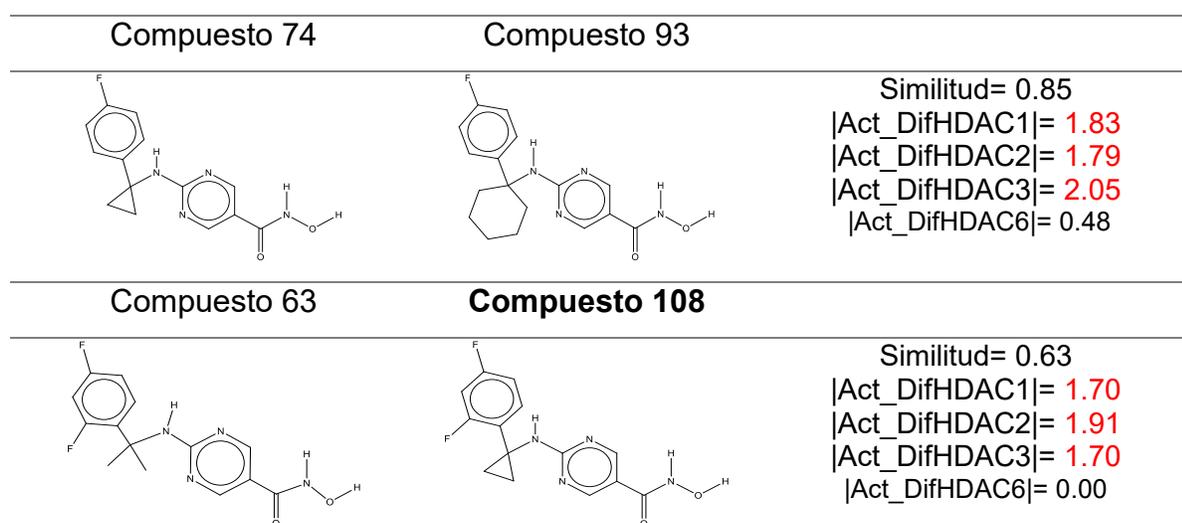
## b) Caracterización del SAR

En esta sección se caracterizó y se analizó el SAR mediante el empleo de los mapas SAS. En la Figura C7 se muestran los mapas SAS para cuatro de las isoformas de la familia de HDACs. En general, se nota una mayor cantidad de pares compuestos en la región IV (*activity cliffs*) para HDAC1, HDAC2 y HDAC3, lo que sugiere una mayor discontinuidad del SAR. Por otra parte, HDAC6 presenta pocos pares de compuestos en esta región. Es interesante observar que el comportamiento en la distribución de pares de compuestos para las isoformas HDAC1, HDAC2 y HDAC3 es similar. Este resultado se puede atribuir a que el dominio catalítico de estas isoformas de la clase I está altamente conservado y tienen alta homología (45-94% de identidad en la secuencia de aminoácidos). Esto no ocurre con HDAC6 que pertenece a otra clase de HDAC y que comparte similitud en la secuencia con la isoforma HDAC10 (55% identidad general) (75).

La observación en la distribución de los puntos en los mapas SAS llevó a analizar los pares de compuestos en la región II del mapa SAS de HDAC6 (SAR continuo) y que a su vez están presentes en la región IV de los mapas SAS de HDAC1, HDAC2 y HDAC3; el análisis de estos pares de compuestos puede sugerir elementos estructurales asociados con la selectividad hacia HDAC6. Ejemplos de pares de compuestos ubicados en estas regiones están en la Figura C8 los cuales pertenecen al grupo A en la caracterización del espacio químico.



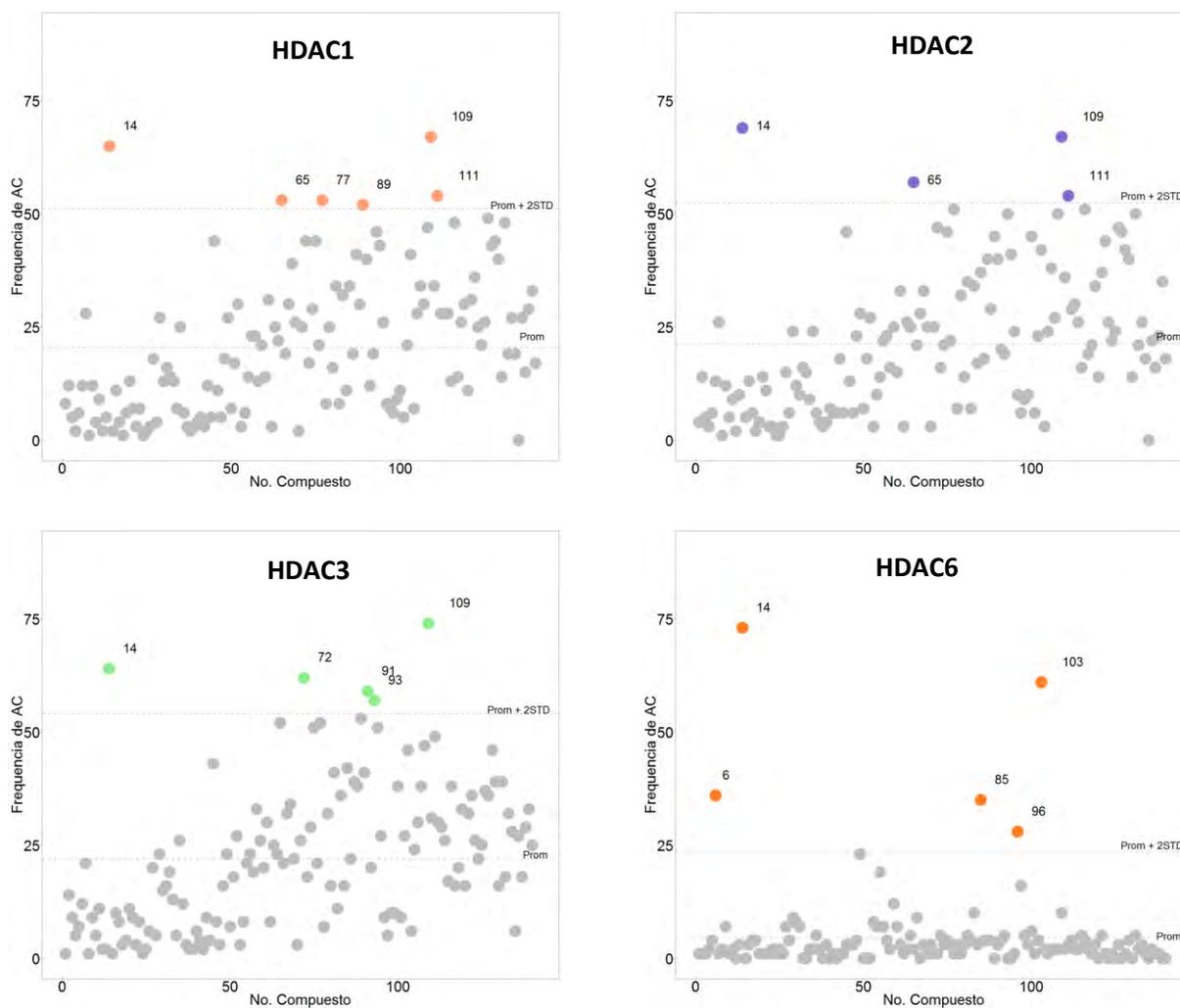
**Figura C7.** Mapas SAS de las diferentes isoformas de HDAC. Cada punto representa un par de compuestos. Se grafica similitud estructural contra el valor absoluto de la diferencia de actividad.



**Figura C8.** Pares de compuestos representativos situados en la región II del mapa SAS de HDAC6 y en la región IV de los mapas SAS de HDAC1, 2 y 3.

### c) Identificación de compuestos generadores de *activity cliffs*

Los gráficos en la Figura C9 muestran los números de *activity cliffs* asociados con cada molécula en la base de datos. La Tabla C1 resume valores estadísticos de esta distribución. En total se identificaron 13 compuestos generadores de *activity cliffs* los cuales están asociados a valores muy altos o bajos de  $CI_{50}$ . La Tabla C2 resume la clave, estructura, la frecuencia y la actividad de estos compuestos.

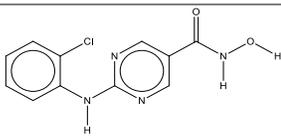
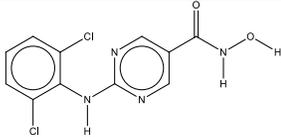
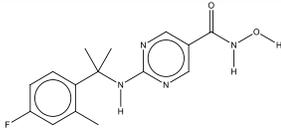
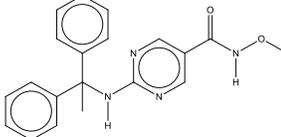
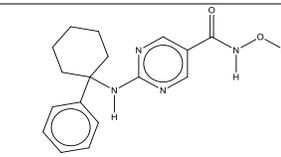


**Figura C9.** Número de *activity cliffs* asociados con cada compuesto en la base de datos. Se resalta con color a los compuestos con un número de *activity cliffs* mayor al promedio más dos desviaciones estándar, considerados como generadores de *activity cliffs*.

**Tabla C1.** Distribución del número de *activity cliffs* formados por cada compuesto en la base de datos.

Estadísticas	HDAC1	HDAC2	HDAC3	HDAC6
<b>Max</b>	67.00	69.00	74.00	73.00
<b>Q3</b>	30.00	29.25	32.00	4.00
<b>Mediana</b>	17.5	21.31	20.00	2.00
<b>Q1</b>	7.00	7.00	8.00	1.00
<b>Min</b>	0.00	0.00	1.00	0.00
<b>Promedio</b>	20.46	29.25	21.99	4.61
<b>STD</b>	15.35	15.53	17.97	9.39

**Tabla C2.** Compuestos identificados como generadores de *activity cliffs*. En rojo se indican frecuencias superiores al valor promedio más dos desviaciones estándar.

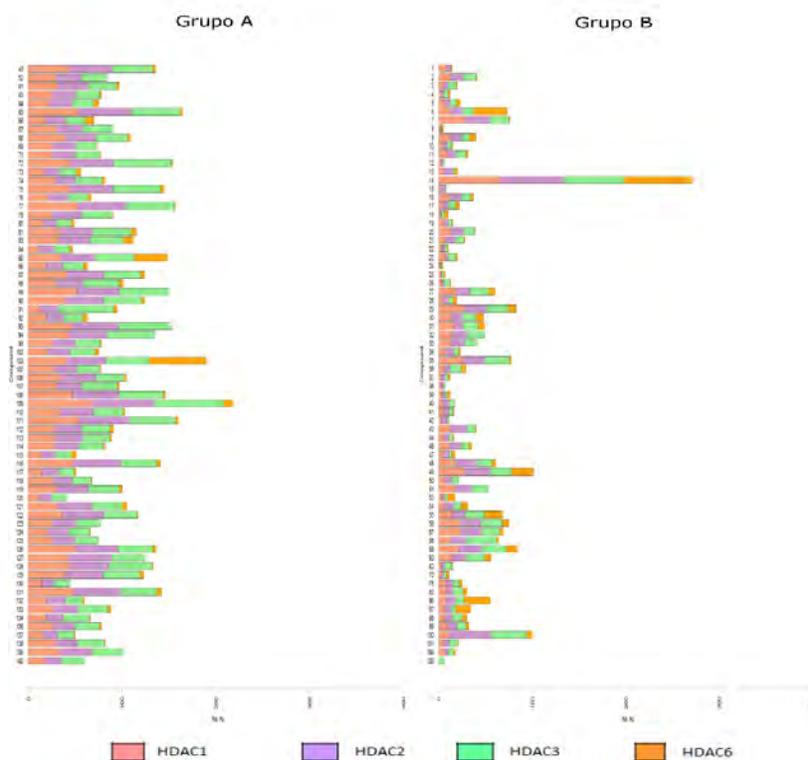
No. Comp (Grupo)	Estructura	Número de <i>activity cliffs</i>				Cl <sub>50</sub> (nM)			
		HDAC1	HDAC2	HDAC3	HDAC6	HDAC1	HDAC2	HDAC3	HDAC6
6 (GB)		12	13	12	36	415	629	606	42
14 (GB)		65	69	64	73	10598	17328	16328	864
65 (GA)		53	57	52	2	5752	8006	12297	19
72 (GA)		44	47	62	1	1268	1830	4676	6
77 (GA)		53	51	52	1	2226	2971	4515	7

85 (GA)		34	37	42	35	1359	2235	4422	33
89 (GA)		52	45	53	0	2341	2465	5709	9
91 (GA)		12	20	59	4	147	239	4252	4
93 (GA)		46	50	57	0	3206	4761	13474	9
96 (GB)		8	10	9	28	578	832	1098	46
103 (GB)		41	42	46	61	1409	1631	3460	41
109 (GA)		67	67	74	10	5	6	8	1
111 (GA)		54	54	49	3	7	9	22	2

Los resultados obtenidos de la frecuencia que tienen cada compuesto de la base de datos de formar *activity cliffs* se compararon con el espacio químico de los 140 inhibidores de HDACs. En la Figura C10 se muestran las frecuencias acumuladas de cada compuesto de formar *activity cliffs* en las cuatro isoformas de HDACs en las que se evaluaron los

compuestos. En esta figura se puede observar que pequeñas modificaciones a la estructura de los compuestos del grupo A tienden a formar más *activity cliffs* en HDAC1, HDAC2 y HDAC3. Por otra parte, pequeñas modificaciones a la estructura de los compuestos de grupo B tienden a formar más *activity cliffs* en HDAC6. Lo anterior coincide con lo observado en la Figura C8, donde las estructuras que sugieren una mayor selectividad a HDAC6 pertenecen a compuestos del grupo A.

Esta observación no puede generalizarse sin antes observar los valores de  $CI_{50}$  de los compuestos generadores de *activity cliffs*. Por ejemplo, los compuestos **109** y **111** tienden a formar un gran número de *activity cliffs* en HDAC1, HDAC2 y HDAC3 y esto se debe a que los valores de  $CI_{50}$  son muy bajos, lo cual no significa que los compuestos sean selectivos hacia HDAC6.



**Figura C10.** Número de *activity cliffs* formados por cada compuesto en las diferentes isoformas de HDACs. Los compuestos están divididos en los dos grupos observados en el espacio químico.

#### d) Caracterización del REMA

El análisis de los pares de compuestos con alta similitud estructural mediante mapas DAD reveló la presencia de *activity cliffs* de un solo, doble, triple y hasta cuádruple objetivo.

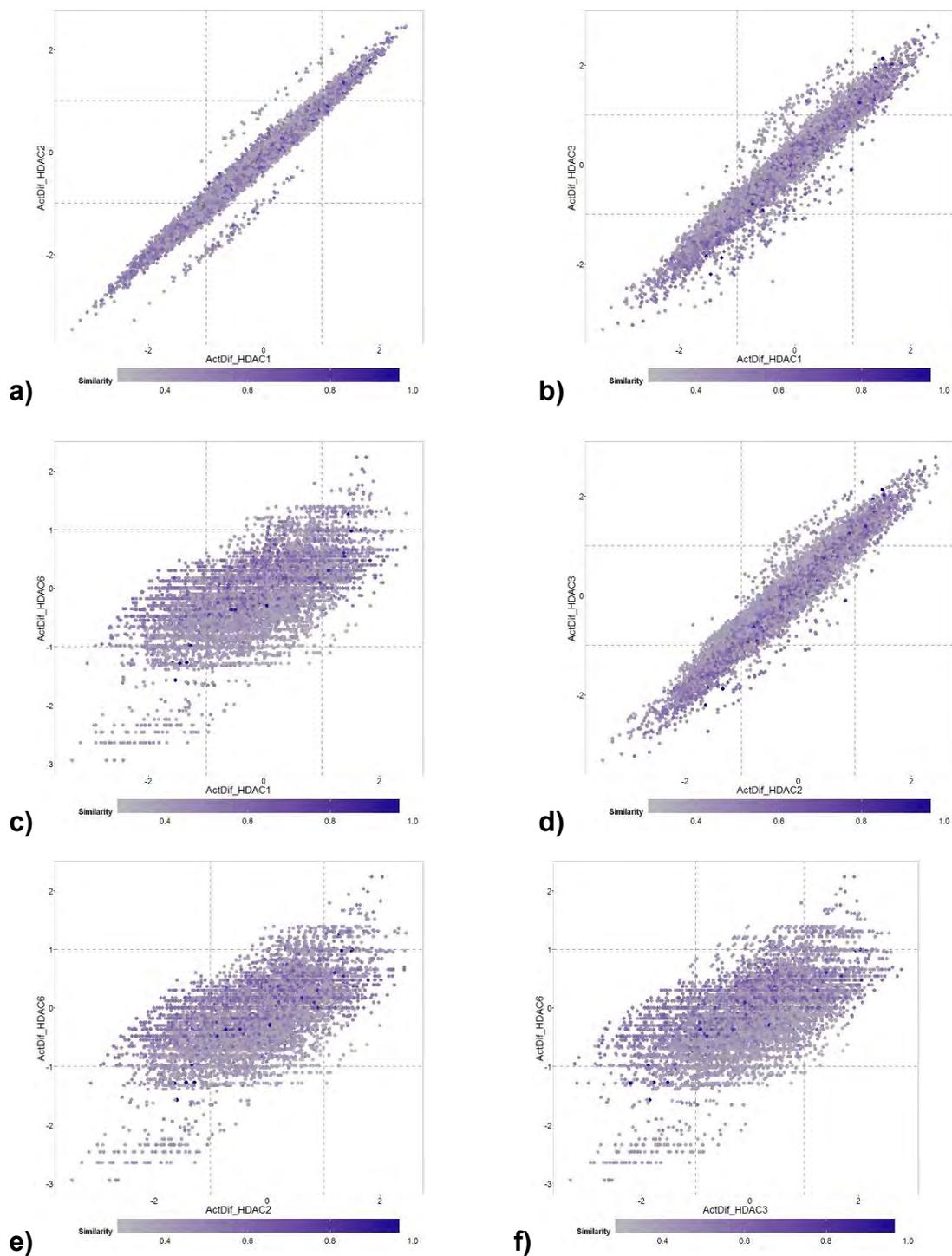
Esto es, pequeños cambios en la estructura con altos cambios de potencia para uno, dos, tres o las cuatro isoformas de HDACs, respectivamente.

Los mapas DAD generados en el caso de estudio se muestra en la Figura C11 y el número de compuestos en cada región se resume en la Tabla C3. En todos los casos se observa una mayor presencia de pares de compuestos en la región Z5 (63.22-67.98%). Esto indica que los compuestos tienen valores de actividad muy similares para ambas isoformas de HDACs con las que se construye el mapa DAD. Los pares identificados en las regiones Z3 y Z4 siguieren que cambios en el núcleo de las hidroxipirimidinas son más susceptibles de presentar cambios en la actividad contra HDAC1, HDAC2 y HDAC3 en comparación con HDAC6. En este mismo sentido, se observa un orden en cómo se afecta a estas isoformas, siendo la isoforma HDAC3 la que presenta un mayor número de *activity cliffs*, seguida de HDAC2 y HDAC1, respectivamente.

Por otra parte, la mayor cantidad de pares de compuestos en la región Z1 comparando con la región Z2 indica que las modificaciones a la estructura de las hidroxipirimidinas afectan la actividad en ambas isoformas en la misma magnitud y dirección. Compuestos presentes en la región Z2 resultaron poco informativos ya que sus valores de similitud están por debajo del valor de la mediana.

**Tabla C3.** Número de pares de compuestos presentes en cada región de los mapas DAD. Total de pares de compuestos 7930.

Región	HDAC1 / HDAC2	HDAC1 / HDAC3	HDAC1 / HDAC6	HDAC2 / HDAC3	HDAC2 / HDAC6	HDAC3 / HDAC6
Z1	2500 (25.69%)	2256 (23.19%)	474 (4.87%)	2211 (22.77%)	489 (6.17%)	454 (5.72%)
Z2	0 (0.00%)	0 (0.00%)	1 (0.01%)	0 (0.00%)	5 (0.05%)	2 (0.02%)
Z3	392 (2.02%)	641 (6.59%)	379 (3.89%)	686 (7.05%)	360 (3.70%)	398 (4.09%)
Z4	237 (2.44%)	481 (4.94%)	2262 (23.25%)	681 (7.00%)	2398 (24.65%)	2441 (25.09%)
Z5	6601 (67.84 %)	6352 (65.28 %)	6614 (67.98%)	6152 (63.22%)	6478 (66.58%)	6435 (66.14%)



**Figura C11.** Mapas DAD para las diferentes isoformas de HDAC. Cada punto es un par de compuestos, se grafica diferencia de actividad para una isoforma de HDAC contra la diferencia de actividad para otra isoforma de HDAC. La similitud estructural se indica con escala de color: gris (baja similitud estructural) y morado (alta similitud estructural). **a)** HDAC1 vs HDAC2, **b)** HDAC1 vs HDAC3, **c)** HDAC1 vs HDAC6, **d)** HDAC2 vs HDAC3, **e)** HDAC2 vs HDAC6, **f)** HDAC3 vs HDAC6.

**Tabla C4.** Resultados representativos del análisis de los mapas DAD. Diferencias de actividad superior a 1 unidad logarítmica están en rojo.

Par	HDAC1	HDAC2	HDAC3	HDAC6	Similitud estructural*
<i>Activity cliffs</i> solo para HDAC1					
105_118	1.19	0.96	0.75	-0.12	0.62
112_118	1.07	0.93	0.79	0.18	0.63
<i>Activity cliffs</i> solo para HDAC2					
30_100	-0.11	-1.19	-0.83	0.0	0.76
77_137	-0.95	-1.13	-0.84	-0.24	0.65
<i>Activity cliffs</i> solo para HDAC3					
66_88	0.79	0.75	1.02	0.18	0.82
<i>Activity cliffs</i> solo para HDAC6					
68_85	0.08	0.36	0.37	1.04	0.62
34_53	0.36	0.32	0.25	1	0.60
Triple <i>activity cliffs</i> para HDAC1,2,3					
74_93	1.83	1.79	2.05	0.48	0.87
80_94	1.40	1.35	1.68	0.54	0.86
77_131	-2.14	-2.13	-1.90	-0.37	0.64
Cuádruple <i>activity cliffs</i> para HDAC1,2,3,4					
1_14	1.60	1.84	1.74	2.24	0.56
90_107	1.67	1.51	2.03	1	0.86
54_55	1.02	1.04	1.14	1.27	0.63
55_56	-1.58	-1.44	-1.77	-1.27	1.00

\* Calculado con Tanimoto/ECFP.

### e) Análisis de *activity cliffs* mediante estudios de acoplamiento molecular

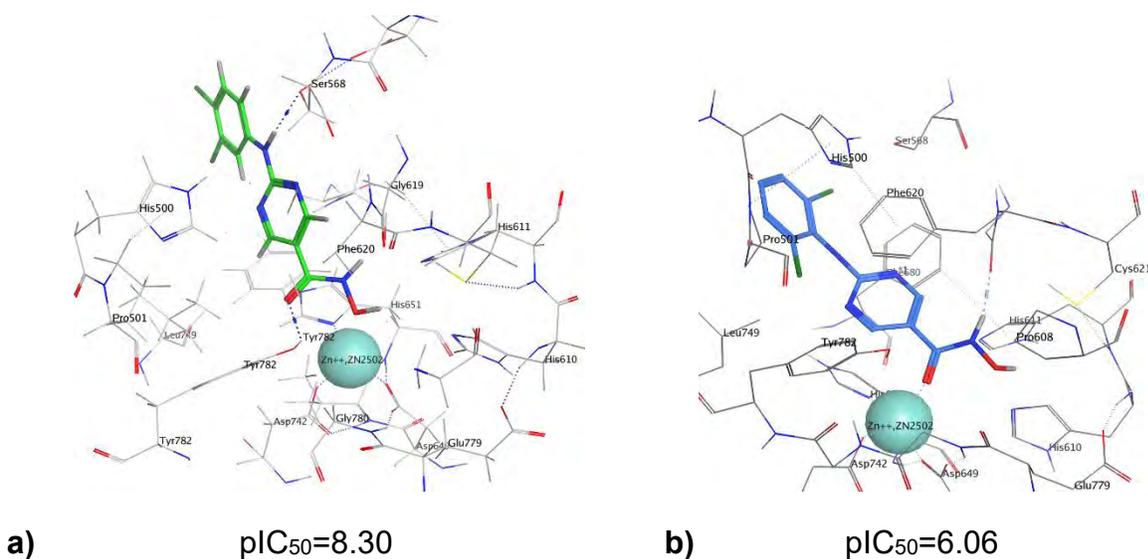
Una vez identificados los *activity cliffs* se realizó acoplamiento molecular con los compuestos **1** y **14**. Este par de compuestos tiene un valor de similitud de 0.56 y una diferencia de actividad para HDAC6 de 2.24 unidades logarítmicas. El compuesto **14**, por su parte, fue identificado como un generador de *activity cliffs* para las cuatro isoformas estudiadas y es el compuesto que contribuye mayoritariamente en la formación de *activity cliffs* en HDAC6.

Los resultados del acoplamiento molecular (Figura C12) muestran que **1** presenta tres interacciones por puente de hidrógeno entre la función del ácido hidroxámico con los residuos His610, Tyr782 y con el  $Zn^{2+}$ . De estas interacciones, los puentes de hidrógeno con His610 y Tyr782 ya se han reportado en la literatura, al igual que la interacción con His611 (62). Estas interacciones parecen ser fundamentales para que los inhibidores lleven a cabo su función.

El acoplamiento molecular entre **14** y el dominio catalítico II de HDAC6 mostró que solo se mantiene la interacción por puente de hidrógeno con His610 y con el  $Zn^{2+}$ . En la representación 2D (Figura C12), se puede observar que **1** se encuentra mejor posicionado en la hendidura de unión (contorno azul) en comparación con **14**. Este resultado puede ayudar a explicar la diferencia significativa en las actividades biológicas de **1** y **14**.

Analizando la estructura de los compuestos de la base de datos que comparten una estructura similar a **1** y **14** (Tabla A1) se observa que grupos funcionales en posición R2 y R6 tienden a presentar valores bajos de  $pIC_{50}$  al igual que **14**.

Por otra parte, los compuestos **96**, **97** y **103**, al carecer de un grupo aromático unido a la porción amino en la estructura de las hidroxipirimidinas y de grupos funcionales que favorezcan interacciones hidrofóbicas entre la región *cap* y el CDII de HDAC6, tienen valores bajos de  $pIC_{50}$  (7.34,7.31,7.39 respectivamente).





En general, los resultados sugieren que los inhibidores de HDACs incluidos en el estudio, presentan parámetros farmacocinéticos favorables para ser administrados oralmente.

A diferencia del fármaco Vorinostat (SAHA), que también contiene una porción de ácido hidroxámico en su estructura, los compuestos de esta base de datos no exhiben valores tan altos de unión a proteínas plasmáticas (Vorinostat 71.3%). Este resultado tiene influencia significativa en el comportamiento farmacocinético y farmacodinámico de los compuestos, ya que sólo la fracción libre puede acceder a los receptores o al sitio de acción y, por lo tanto, ser eficaz.

**Tabla C5.** Resultados de la evaluación de propiedades ADME

Parámetro evaluado	Base de datos HDACs	Valores de referencia
<b>Porcentaje de Absorción Oral (Valor promedio)</b>	77.45%	>80% es alta <25% es baja
<b>Predicción de unión a la albúmina sérica humana.</b>	-0.9-0.047	-1.5 – 1.5
<b>No. De probables reacciones metabólicas.</b>	2-6	1-8
<b>No. De violaciones a la regla de Lipinski (Ro5)</b>	0	Máximo 4
<b>No. De violaciones a la regla de Jorgensen (Ro3)</b>	0	Máximo 3

Ro5: MW < 500, clogP < 5, HBD ≤ 5, HBA ≤ 10.

Ro3: QPlogS > -5.7, QP PCaco > 22 nm/s, No. metabolitos primarios < 7.

### Conclusiones del caso de estudio

- Se caracterizó el REmA de 140 compuestos reportados como inhibidores de HDACs y se identificaron *activity cliffs* para uno, dos, tres y cuatro de las isoformas estudiadas.
- La implementación del concepto de limpieza de panoramas de actividad utilizando espacio químico fue útil para observar el comportamiento del número de *activity cliffs* generados por cada compuesto. Los resultados sugieren que la síntesis de compuestos similares al grupo A serán más selectivos para HDAC6, ya que estos compuestos tienen una alta probabilidad de formar *activity cliffs* en HDAC1, HDAC2 y HDAC3, pero no HDAC6. Esta observación no se puede generalizar para todos los compuestos sin antes considerar los valores de  $CI_{50}$ .
- Estudios de acoplamiento molecular permitieron plantear una hipótesis de la diferencia en potencia entre *activity cliffs* para HDAC6. Para compuestos de la Tabla A1 la sustitución en R2 y R6 disminuye notablemente la capacidad de inhibir a HDAC6. Por otra parte, es necesario la presencia de grupos aromáticos unidos a la región del 2-amino de la hidroxipirimidina y de grupos funcionales que favorezcan interacciones hidrofóbicas entre la región *cap* y el dominio catalítico II de HDAC6.
- La predicción de propiedades ADME sugiere que los compuestos de este caso de estudio presentan parámetros farmacocinéticos favorables para ser administrados oralmente.
- La resolución del problema presentado en el caso de estudio involucra la integración de diversos conceptos y métodos mostrados en el Manual de Quimioinformática.

### Perspectivas del caso de estudio

- Realizar estudios de *docking* adicionales para explicar a nivel molecular los *activity cliffs* identificados en este trabajo.
- Desarrollar modelos predictivos (por ejemplo, QSAR) de las regiones continuas del SAR.

## **CONCLUSIONES**

Se elaboró material didáctico de quimioinformática que, a la fecha de elaboración de esta tesis (febrero 2017), está ayudando a complementar las clases de Química Farmacéutica en la Facultad de Química. Los alumnos del semestre 2017-1 fueron los primeros en profundizar en temas acerca de quimioinformática que son de gran relevancia para esta disciplina y que anteriormente no se abordaban.

Asimismo, la difusión del material didáctico generado en congresos y talleres impartidos por el grupo DIFACQUIM ([www.difacquim.com](http://www.difacquim.com)) ha ayudado a incrementar el interés de alumnos e investigadores provenientes de otros centros educativos por el empleo de métodos computacionales en sus trabajos de investigación. Cabe enfatizar que la apropiación de estas herramientas informáticas por parte de alumnos y profesionistas en el área de Química Farmacéutica ayudará a orientar la escasa inversión en muchas enfermedades, al priorizar y racionalizar los costos del desarrollo de nuevos fármacos.

Se espera que con el material generado nuevas generaciones de la carrera de Química Farmacéutico Biológico egresen con conocimientos sólidos en el uso de metodologías de quimioinformática y puedan integrarse a la industria farmacéutica y proyectos de investigación que se desarrollan en esta área.

## **PERSPECTIVAS**

- Implementación completa del manual en grupos de Química Farmacéutica.
- Registro ISBN del manual por parte de la Facultad de Química (revisión en proceso).
- Difusión del material en congresos, talleres y cursos impartidos por DIFACQUIM.
- Traducción al inglés del manual de quimioinformática. Esto en respuesta a solicitudes explícitas recibidas en congresos por profesores de Rusia y Holanda.
- Incorporación de instrumentos que permitan realizar evaluaciones adecuadas en cada actividad del manual (ej. Apps, Moodle, etc).
- Generar una nueva edición del Manual, actualizándolo con nuevos ejemplos, problemas y conceptos de punta.

---

**REFERENCIAS**

1. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Briefings in Bioinformatics*. 2009; 10(5):579-591.
2. Medina-Franco JL, Fernández-de Gortari E, Naveja JJ. Avances en el diseño de fármacos asistido por computadora. *Educación Química*. 2015;26(3):180-186.
3. Leach AR, Gillet VJ. *An Introduction to Chemoinformatics*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2003.
4. Brown FK. Chemoinformatics: What is it and how does it impact drug discovery. *Annual Reports in Medicinal Chemistry*. 1998, 33: 375-384.
5. Gasteiger J. Introduction. En: *Chemoinformatics*: Wiley-VCH Verlag GmbH & Co. KGaA; 2004. p 1-13.
6. Chen WL. Chemoinformatics: Past, present, and future. *Journal of Chemical Information and Modeling*. 2006; 46(6):2230-2255.
7. Gasteiger J. Chemoinformatics: A new field with a long tradition. *Analytical and Bioanalytical Chemistry*. 2006; 384(1):57-64.
8. Martinez-Mayorga K, Medina-Franco JL, editores. *FoodInformatics: Applications of Chemical Information to Food Chemistry*. New York: Springer; 2014. pp 251.
9. Gasteiger J. Chemoinformatics: Achievements and challenges, a personal view. *Molecules*. 2016; 21(2):151.
10. Engel T. Basic overview of chemoinformatics. *Journal of Chemical Information and Modeling*. 2006; 46(6):2267-2277.
11. Burkard U. Methods for Data Analysis. En: *Chemoinformatics*: Wiley-VCH Verlag GmbH & Co. KGaA; 2004. p. 439-485.
12. Willett P. Chemoinformatics: a history. *WIREs Computational Molecular Science* 2011; 1(1):46-56.
13. Begam BF, Kumar JS. A Study on cheminformatics and its applications on modern drug discovery. *Procedia Engineering*. 2012; 38:1264-1275.
14. Ma D-L, Chan DS-H, Leung C-H. Drug repositioning by structure-based virtual screening. *Chemical Society Reviews*. 2013; 42(5):2130-2141.

15. Medina-Franco JL, Lopez-Vallejo F, Castillo R. Diseño de fármacos asistido por computadora. *Educación Química*. 2006; 17(4):452-457.
16. Wild DJ, Wiggins GD. Challenges for chemoinformatics education in drug discovery. *Drug Discovery Today*. 2006; 11(9–10):436-439.
17. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*. 2012;40(D1):D1100-D1107.
18. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. *Nucleic Acids Research*. 2016;44(Database issue):D1202-D1213.
19. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*. 2008;36:D901-D906.
20. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*. 2005;45(1):177-182.
21. Molecular Operating Environment (MOE), versión 2010. Chemical Computing Group Inc., Montreal, Quebec, Canada. Disponible en <http://www.chemcomp.com>.
22. Sander T, Freyss J, von Korff M, Rufener C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling*. 2015;55(2):460-73.
23. Scior T, Bernard P, Medina-Franco JL, Maggiora GM. Large compound databases for structure-activity relationships studies in drug discovery. *Mini Reviews in Medicinal Chemistry*. 2007;7(8):851-860.
24. Wassermann AM, Lounkine E, Davies JW, Glick M, Camargo LM. The opportunities of mining historical and collective data in drug discovery. *Drug Discovery Today*. 2015;20(4):422-434.
25. Lusher SJ, McGuire R, van Schaik RC, Nicholson CD, de Vlieg J. Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*. 2014;19(7):859-868.

- 
26. Fourches D, Muratov E, Tropsha A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*. 2010;50(7):1189-1204.
  27. Osolodkin DI, Radchenko EV, Orlov AA, Voronkov AE, Palyulin VA, Zefirov NS. Progress in visual representations of chemical space. *Expert Opinion on Drug Discovery*. 2015;10(9):959-973.
  28. Medina-Franco JL, Maggiora GM. Molecular Similarity Analysis. En: *Chemoinformatics for Drug Discovery*: John Wiley & Sons, Inc; 2013. p. 343-399.
  29. Brown N. Chemoinformatics; an introduction for computer scientists. *ACM Computing Surveys*. 2009;41(2):1-38.
  30. Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today*. 2016;21(8):1291-1302.
  31. Medina-Franco JL. Interrogating novel areas of chemical space for drug discovery using cheminformatics. *Drug Development Research*. 2012;73(7):430-438.
  32. Petit J, Meurice N, Medina-Franco JL, Maggiora G. A Rough Set Theory Approach to the analysis of gene expression profiles. En: *Chemoinformatics for Drug Discovery*: John Wiley & Sons, Inc; 2013. p. 51-83.
  33. Bajorath J, Peltason L, Wawer M, Guha R, Lajiness MS, Van Drie JH. Navigating structure-activity landscapes. *Drug Discovery Today*. 2009;14(13-14):698-705.
  34. Wassermann AM, Bajorath J. Chemical substitutions that introduce activity cliffs across different compound classes and biological targets. *Journal of Chemical Information and Modeling*. 2010;50(7):1248-1256.
  35. Naveja JJ, Medina-Franco JL. Activity landscape of DNA methyltransferase inhibitors bridges cheminformatics with epigenetic drug discovery. *Expert Opinion on Drug Discovery*. 2015;10(10):1059-1070.
  36. Guha R, VanDrie JH. Structure-activity landscape index: identifying and quantifying activity cliffs. *Journal of Chemical Information and Modeling*. 2008;48(3):646-658.
  37. Cruz-Monteagudo M, Medina-Franco JL, Perez-Castillo Y, Nicolotti O, Cordeiro MN, Borges F. Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today*. 2014;19(8):1069-1080.

- 
38. Pérez-Villanueva J, Méndez-Lucio O, Soria-Arteche O, Medina-Franco JL. Activity cliffs and activity cliff generators based on chemotype-related activity landscapes. *Molecular diversity*. 2015;19(4):1021-1035.
  39. Aguayo-Ortiz R, Pérez-Villanueva J, Hernandez-Campos A, Castillo R, Meurice N, Medina-Franco JL. Chemoinformatic characterization of activity and selectivity switches of antiprotozoal compounds. *Future Medicinal Chemistry*. 2014;6(3):281-294.
  40. Warren GL, Do TD, Kelley BP, Nicholls A, Warren SD. Essential considerations for using protein-ligand structures in drug discovery. *Drug Discovery Today*. 2012;17(23-24):1270-1281.
  41. Morris GM, Lim-Wilby M. Molecular Docking. En: *Molecular Modeling of Proteins*. Kukol A, editor. Humana Press; 2008. p. 365-82.
  42. Ferreira L, dos Santos R, Oliva G, Andricopulo A. Molecular docking and structure-based drug design strategies. *Molecules*. 2015;20(7):13384.
  43. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*. 2011;7(2):146-157.
  44. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews: Drug Discovery*. 2004;3(11):935-949.
  45. Méndez-Lucio O, Kooistra AJ, Graaf Cd, Bender A, Medina-Franco JL. Analyzing multitarget activity landscapes using protein–ligand interaction fingerprints: interaction cliffs. *Journal of Chemical Information and Modeling*. 2015;55(2):251-262.
  46. Zhang H, Xin M-H, Xie X-X, Mao S, Zuo S-J, Lu S-M, et al. Synthesis and antitumor activity evaluation of PI3K inhibitors containing 3-substituted quinazolin-4(3H)-one moiety. *Bioorganic & Medicinal Chemistry*. 2015;23(24):7765-7776.
  47. Valerio Jr LG. In silico toxicology for the pharmaceutical sciences. *Toxicology and Applied Pharmacology*. 2009;241(3):356-370.
  48. Butina D, Segall MD, Frankcombe K. Predicting ADME properties in silico: methods and models. *Drug Discovery Today*. 2002;7(11):S83-S88.

- 
49. Naveja JJ, Medina-Franco JL. Activity landscape sweeping: insights into the mechanism of inhibition and optimization of DNMT1 inhibitors. *RSC Advances*. 2015;5(78):63882-63895.
  50. Lan P, Huang Z-J, Sun J-R, Chen W-M. 3D-QSAR and molecular docking studies on fused pyrazoles as p38 $\alpha$  mitogen-activated protein kinase inhibitors. *International Journal of Molecular Sciences*. 2010;11(9):3357.
  51. Saldivar-Gonzalez FI, Naveja JJ, Palomino-Hernandez O, Medina-Franco JL. Getting SMART in drug discovery: chemoinformatics approaches for mining structure-multiple activity relationships. *RSC Advances*. 2017; 7(2):632-641.
  52. Roche J, Bertrand P. Inside HDACs with more selective HDAC inhibitors. *European Journal of Medicinal Chemistry*. 2016; 121:451-483.
  53. Kalin JH, Bergman JA. Development and therapeutic implications of selective histone deacetylase 6 inhibitors. *Journal of Medicinal Chemistry*. 2013; 56(16):6297-6313.
  54. Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, et al. HDAC6 is a microtubule-associated deacetylase. *Nature*. 2002; 417(6887):455-458.
  55. Kawaguchi Y, Kovacs JJ, McLaurin A, Vance JM, Ito A, Yao T-P. The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. *Cell*. 2003; 115(6):727-738.
  56. Liu Y, Li L, Min J. Structural biology: HDAC6 finally crystal clear. *Nature Chemical Biology*. 2016; 12(9):660-661.
  57. Best JD, Carey N. Epigenetic therapies for non-oncology indications. *Drug Discovery Today*. 2010; 15(23–24):1008-1014.
  58. Maolanon Alex R, Madsen Andreas S, Olsen Christian A. Innovative strategies for selective inhibition of histone deacetylases. *Cell Chemical Biology*. 2016; 23(7):759-768.
  59. Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nature Reviews Drug Discovery*. 2014; 13(9):673-691.
  60. Mottamal M, Zheng S, Huang LT, Wang G. Histone deacetylase inhibitors in clinical studies as templates for new anticancer agents. *Molecules*. 2015; 20(3): 3898-3941.

- 
61. Praveer G, Robert CR, Abishek I, Matthew JS, David PF. Towards isozyme-selective hdac inhibitors for interrogating disease. *Current Topics in Medicinal Chemistry*. 2012; 12(14):1479-1499.
62. Schäfer S, Saunders L, Eliseeva E, Velená A, Jung M, Schwienhorst A, et al. Phenylalanine-containing hydroxamic acids as selective inhibitors of class IIb histone deacetylases (HDACs). *Bioorganic & Medicinal Chemistry*. 2008; 16(4):2011-2033.
63. Medina-Franco JL. Chemoinformatic Characterization of the Chemical Space and Molecular Diversity of Compound Libraries. En: *Diversity-Oriented Synthesis*: John Wiley & Sons, Inc.; 2013. p. 325-352.
64. Medina-Franco JL, Martínez-Mayorga K, Giulianotti MA, Houghten RA, Pinilla C. Visualization of the chemical space in drug discovery. *Current Computer-Aided Drug Design*. 2008;4(4):322-333.
65. Bajorath J, Peltason L, Wawer M, Guha R, Lajiness MS, Van Drie JH. Navigating structure-activity landscapes. *Drug Discovery Today*. 2009; 14(13-14):698-705.
66. Stumpfe D, Hu Y, Dimova D, Bajorath J. Recent progress in understanding activity cliffs and their utility in medicinal chemistry. *Journal of Medicinal Chemistry*. 2014; 57(1):18-28.
67. Dimova D, Bajorath J. Advances in activity cliff research. *Molecular Informatics*. 2016; 35(5):181-191.
68. Mendez-Lucio O, Perez-Villanueva J, Castillo R, Medina-Franco JL. Identifying activity cliff generators of PPAR ligands using SAS Maps. *Molecular Informatics*. 2012; 31(11-12):837-846.
69. Zhang Y, Kwon S, Yamaguchi T, Cubizolles F, Rousseaux S, Kneissel M, et al. Mice lacking histone deacetylase 6 have hyperacetylated tubulin but are viable and develop normally. *Molecular and Cellular Biology*. 2008; 28(5):1688-1701.
70. J. H. Van duzer, R. Mazitschek, Y. Ding, N. Yu, Y. Cao and Y. Liu, *Pyrimidine Hydroxy Amide Compounds as Protein Deacetylase Inhibitors and Methods of Use Thereof*, Acetylon Pharmaceuticals, Inc., 2014, EP2640709.
71. Sud M. MayaChemTools: An open source package for computational drug discovery. *Journal of Chemical Information and Modeling*. 2016; 56(12):2292-2297.

- 
72. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*. 2015; 7(1):1-13.
73. Perez-Villanueva J, Santos R, Hernandez-Campos A, Giulianotti MA, Castillo R, Medina-Franco JL. Structure-activity relationships of benzimidazole derivatives as antiparasitic agents: Dual activity-difference (DAD) maps. *Medicinal Chemical Communications*. 2011; 2(1): 44-49.
74. QikProp 3.5 User Manual. QikProp, Schrödinger, LLC, New York, NY, 2016.
75. Seto E, Yoshida M. Erasers of histone acetylation: The histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*. 2014;6(4):a018713.

## APÉNDICE A. Estructura y actividades de los 140 compuestos utilizados en el caso de estudio.

Tabla A1. Estructuras y actividades de los compuestos utilizados en el caso de estudio.

Comp.	Sustituciones					Valores de pI <sub>C50</sub>					
	R1	R2	R3	R4	R5	R6	HDAC1	HDAC2	HDAC3	HDAC6	
1	H	H	H	Cl	Cl	H	6.58	6.61	6.52	8.30	
2	H	H	H	F	H	H	6.68	6.46	6.45	8.10	
3	H	H	H	CH <sub>3</sub>	H	H	6.71	6.54	6.54	8.05	
4	H	H	H	OCH <sub>3</sub>	H	H	6.61	6.40	6.38	7.82	
5	H	H	H	Cl	H	H	6.69	6.64	6.58	7.70	
6	H	Cl	H	H	H	H	6.38	6.20	6.22	7.38	
7	H	H	CF <sub>3</sub>	H	CF <sub>3</sub>	H	5.85	5.86	5.73	7.85	
8	CH <sub>3</sub>	H	H	Cl	Cl	H	7.09	7.09	6.66	7.42	
9	H	F	F	H	F	F	6.48	6.42	6.22	7.68	
10	H	H	H	H	H	H	6.74	6.79	6.70	8.30	
11	H	H	F	H	F	H	6.87	6.77	6.75	7.89	
12	H	H	H	NO <sub>2</sub>	CF <sub>3</sub>	H	6.48	6.20	6.37	8.00	
13	H	Cl	H	H	Cl	H	6.67	6.34	6.45	7.82	
14	H	Cl	H	H	H	Cl	4.97	4.76	4.79	6.06	
15	H	H	CF <sub>3</sub>	H	H	H	6.62	6.28	6.32	8.30	
16	H	H	H	H	CF <sub>3</sub>	H	6.65	6.35	6.50	8.52	
17	H	H	CH <sub>3</sub>	H	CH <sub>3</sub>	H	6.60	6.32	6.50	8.40	
18	H	H	OCH <sub>3</sub>	H	OCH <sub>3</sub>	H	6.81	6.46	6.59	8.52	
19	H	H	F	H	CH <sub>3</sub>	H	6.88	6.58	6.71	8.30	
20	H	H	Cl	H	CH <sub>3</sub>	H	6.29	6.03	6.08	8.00	
21	H	H	F	H	CH <sub>3</sub>	H	6.80	6.46	6.57	8.30	
22	H	H	Cl	H	OCH <sub>3</sub>	H	6.38	6.15	6.13	8.15	
23	H	H	H	H	Fenilo	H	6.95	6.73	6.74	8.40	
24	H	H	H	H		H	6.33	6.09	6.28	8.10	
25	H	H	H	H		H	7.06	6.74	6.85	8.52	

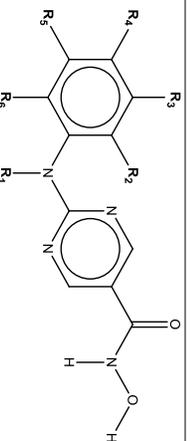
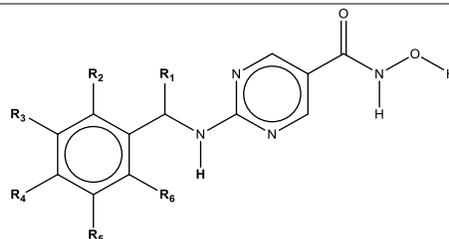


Tabla A2. Estructuras y actividades de los compuestos utilizados en el caso de estudio.

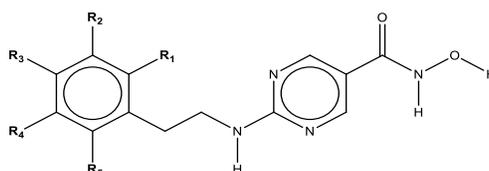


Comp (stereochem)	Sustituciones						Valores de pIC <sub>50</sub>			
	R1	R2	R3	R4	R5	R6	HDAC1	HDAC2	HDAC3	HDAC6
27	CH <sub>3</sub>	H	H	H	H	H	7.33	7.08	7.07	8.70
28	H	H	H	H	H	Cl	6.96	6.65	6.77	8.70
29	H	H	Cl	H	H	H	7.70	7.55	7.47	9.00
30	H	H	H	H	H	H	7.11	6.85	6.89	8.70
31	H	H	H	Cl	H	H	7.40	7.15	7.12	8.70
32	H	H	H	CF <sub>3</sub>	H	H	7.05	6.90	6.83	8.30
33	H	CF <sub>3</sub>	H	H	H	H	6.33	6.14	6.05	8.15
35	H	H	CF <sub>3</sub>	H	H	H	7.42	7.25	7.15	8.70
40	H	H	H	H	H	OCH <sub>3</sub>	6.42	6.28	6.30	8.40
42	H	H	H	H	H	OCF <sub>3</sub>	6.65	6.46	6.52	8.40
43	H	Cl	H	H	H	Cl	6.44	6.22	6.40	8.15
44	H	H	H	F	H	OCH <sub>3</sub>	6.66	6.47	6.52	8.70
45	(CH <sub>3</sub> ) <sub>2</sub>	H	H	H	H	H	5.93	5.82	5.65	8.40
46	H	OCH <sub>3</sub>	H	H	H	OCH <sub>3</sub>	6.33	6.20	6.42	7.96
47	H	H	H	Cl	H	CH <sub>3</sub>	6.73	6.61	6.60	8.52
48	H	H	Cl	F	H	H	7.62	7.52	7.37	8.70
49	H	CH <sub>3</sub>	H	H	H	CH <sub>3</sub>	5.86	5.68	5.80	7.68
50	H	Fenilo	H	H	H	H	6.34	6.23	6.20	8.10
51	H	CF <sub>3</sub>	H	F	H	H	6.28	6.15	6.05	8.22
54	Fenilo	H	H	H	H	H	6.89	6.85	6.63	8.70
(R)-55	CH <sub>3</sub>	H	H	H	H	H	5.87	5.81	5.49	7.43
(S)-56	CH <sub>3</sub>	H	H	H	H	H	7.46	7.25	7.26	8.70
(R)-58	CH <sub>3</sub>	H	H	H	H	H	5.95	5.77	5.13	7.72
(S)-59	CH <sub>3</sub>	H	H	H	H	H	7.41	7.41	7.33	9.00
60	CH <sub>3</sub>	H	H	H	H	H	7.21	7.10	7.00	8.70
61	(CH <sub>3</sub> ) <sub>2</sub>	H	H	F	H	H	6.12	5.97	5.83	8.52
63	(CH <sub>3</sub> ) <sub>2</sub>	H	H	F	H	F	6.19	6.01	5.88	8.52
64	(CH <sub>3</sub> ) <sub>2</sub>	H	H	H	H	F	6.30	6.04	5.90	8.70
65	(CH <sub>3</sub> ) <sub>2</sub>	H	H	F	H	CH <sub>3</sub>	5.24	5.10	4.91	7.72
66	(CH <sub>2</sub> ) <sub>2</sub>	H	H	H	H	H	7.03	6.89	6.66	8.70
67	(CH <sub>3</sub> ) <sub>2</sub>	H	H	CF <sub>3</sub>	H	H	6.05	5.99	5.62	8.22
68	(CH <sub>3</sub> ) <sub>2</sub>	H	F	H	H	H	5.94	6.02	5.73	8.52
69	(CH <sub>3</sub> ) <sub>2</sub>	H	F	F	H	H	6.20	6.22	5.98	8.40
71	(CH <sub>3</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	F	H	H	6.22	6.17	5.88	8.40
73	(CH <sub>3</sub> ) <sub>2</sub>	H	Cl	Cl	H	H	7.07	6.90	6.67	8.70
74	(CH <sub>2</sub> ) <sub>2</sub>	H	H	F	H	H	7.33	7.11	6.92	8.52
75	(CH <sub>2</sub> ) <sub>2</sub>	F	H	H	H	H	7.80	7.74	7.66	8.70
76	(CH <sub>3</sub> ) <sub>2</sub>	H	H	Cl	H	F	6.35	6.34	6.03	8.52
77	(CH <sub>2</sub> ) <sub>5</sub>	H	H	H	H	H	5.65	5.53	5.35	8.15
79	(CH <sub>3</sub> ) <sub>2</sub>	H	H	CF <sub>3</sub>	H	F	6.10	5.89	5.54	7.89
80	(CH <sub>2</sub> ) <sub>2</sub>	H	F	H	H	H	7.07	6.94	6.63	8.70
81	(CH <sub>2</sub> ) <sub>4</sub>	H	H	H	H	H	6.15	6.01	5.65	8.52
83	(CH <sub>2</sub> ) <sub>2</sub>	Cl	H	H	H	H	7.57	7.44	7.39	9.00
84	(CH <sub>2</sub> ) <sub>2</sub>	CF <sub>3</sub>	H	H	H	H	6.80	6.54	6.43	8.70

85	(CH <sub>3</sub> ) <sub>2</sub>	H	CF <sub>3</sub>	H	CF <sub>3</sub>	H	5.87	5.65	5.35	7.48
86	(CH <sub>2</sub> ) <sub>2</sub>	H	H	CH <sub>3</sub>	H	H	7.10	6.97	6.80	8.52
87	(CH <sub>2</sub> ) <sub>2</sub>	H	Cl	H	H	H	7.66	7.49	7.30	8.70
88	THF	H	H	H	H	H	6.23	6.14	5.64	8.52
90	(CH <sub>2</sub> ) <sub>2</sub>	H	H	OCH <sub>3</sub>	H	H	7.72	7.55	7.43	8.70
93	(CH <sub>2</sub> ) <sub>5</sub>	H	H	F	H	H	5.49	5.32	4.87	8.05
94	(CH <sub>2</sub> ) <sub>5</sub>	H	F	H	H	H	5.67	5.59	4.95	8.15
100	H	H	H	Fenilo	H	H	7.22	8.05	7.72	8.70
105	(CH <sub>2</sub> ) <sub>2</sub>	H	OCH <sub>3</sub>	H	H	H	7.42	7.28	6.89	8.40
106	(CH <sub>2</sub> ) <sub>2</sub>	H	CF <sub>3</sub>	H	H	H	7.57	7.49	7.00	8.70
107	(CH <sub>2</sub> ) <sub>5</sub>	H	H	OCH <sub>3</sub>	H	H	6.05	6.05	5.40	7.70
108	(CH <sub>2</sub> ) <sub>2</sub>	F	H	F	H	H	7.89	7.92	7.59	8.52
109	(CH <sub>2</sub> ) <sub>2</sub>	F	H	H	H	F	8.30	8.22	8.10	9.00
110	(CH <sub>2</sub> ) <sub>2</sub>	H	OCF <sub>3</sub>	H	H	H	7.74	7.72	7.24	8.70
111	(CH <sub>2</sub> ) <sub>2</sub>	F	H	H	CF <sub>3</sub>	H	8.15	8.05	7.66	8.70
112	(CH <sub>2</sub> ) <sub>2</sub>	CH <sub>3</sub>	H	H	H	H	7.30	7.24	6.92	8.70
114	(CH <sub>2</sub> ) <sub>2</sub>	H	OCH <sub>3</sub>	H	OCH <sub>3</sub>	H	7.37	7.28	6.91	8.52
115	(CH <sub>2</sub> ) <sub>3</sub>	H	H	H	H	H	6.85	6.80	6.28	8.52
116	(CH <sub>2</sub> ) <sub>2</sub>	H	Cl	F	H	H	7.77	7.72	7.21	8.52
117	(CH <sub>2</sub> ) <sub>2</sub>	H	F	H	F	H	7.02	7.02	6.50	8.40
118	(CH <sub>2</sub> ) <sub>2</sub>	OCH <sub>3</sub>	H	H	H	H	6.23	6.32	6.14	8.52
119	(CH <sub>2</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	H	H	H	7.33	7.40	7.00	8.52
120	(CH <sub>2</sub> ) <sub>2</sub>	H	H	CH <sub>3</sub>	H	H	6.83	6.93	6.46	8.15
121	(CH <sub>2</sub> ) <sub>2</sub>	H	H	Cl	H	H	7.35	7.40	6.97	8.52
122	(CH <sub>2</sub> ) <sub>2</sub>	H	F	CF <sub>3</sub>	H	H	7.57	7.68	7.15	8.40
123	(CH <sub>2</sub> ) <sub>2</sub>	H	OCH <sub>3</sub>	OCH <sub>3</sub>	H	H	7.30	7.31	6.95	8.10
124	(CH <sub>2</sub> ) <sub>2</sub>	OCH <sub>3</sub>	H	H	Cl	H	6.41	6.32	6.11	8.40
125	(CH <sub>2</sub> ) <sub>2</sub>	H	F	F	H	H	7.26	7.21	6.89	8.15
126	(CH <sub>2</sub> ) <sub>2</sub>	H	Cl	Cl	H	H	7.89	7.77	7.27	8.40
127	(CH <sub>2</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	F	H	H	7.77	7.64	7.19	8.22
128	(CH <sub>2</sub> ) <sub>2</sub>	F	F	H	H	H	7.92	7.77	7.54	8.40
129	(CH <sub>2</sub> ) <sub>2</sub>	Cl	H	F	H	H	7.60	7.47	7.24	8.52
130	(CH <sub>2</sub> ) <sub>2</sub>	H	CF <sub>3</sub>	H	CF <sub>3</sub>	H	6.98	6.87	6.42	8.22
131	(CH <sub>2</sub> ) <sub>2</sub>	H	Cl	H	Cl	H	7.80	7.66	7.24	8.52
132	(CH <sub>2</sub> ) <sub>2</sub>	H	CF <sub>3</sub>	H	CF <sub>3</sub>	H	7.19	7.14	6.72	8.52
133	(CH <sub>2</sub> ) <sub>4</sub>	H	H	F	H	H	6.23	6.21	5.67	8.52
134	THF	H	F	H	H	H	6.47	6.48	5.88	8.40
136	(CH <sub>2</sub> ) <sub>4</sub>	H	F	H	H	H	6.23	6.32	5.86	8.52
137	(CH <sub>2</sub> ) <sub>4</sub>	H	H	OCH <sub>3</sub>	H	H	6.61	6.66	6.19	8.40
138	(CH <sub>2</sub> ) <sub>4</sub>	H	Cl	H	H	H	6.28	6.31	5.90	8.10
140	THF	H	F	H	F	H	6.58	6.63	5.91	8.15

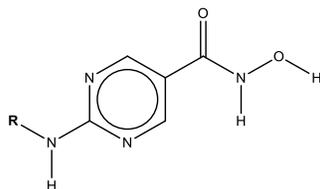
THF: CH<sub>2</sub>CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>

**Tabla A3.** Estructuras y actividades de los compuestos utilizados en el caso de estudio.



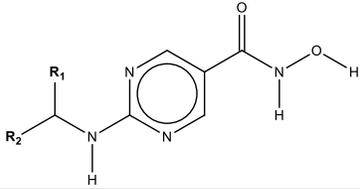
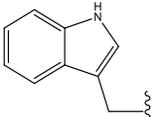
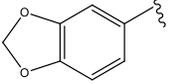
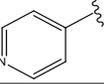
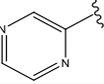
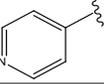
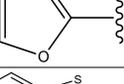
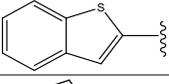
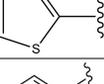
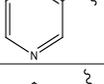
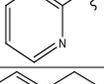
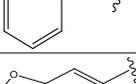
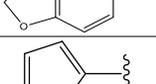
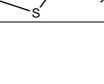
Comp.	Sustituciones					Valores de pIC <sub>50</sub>			
	R1	R2	R3	R4	R5	HDAC1	HDAC2	HDAC3	HDAC6
34	Cl	H	H	H	H	6.58	6.41	6.31	8.40
36	H	H	H	H	H	6.67	6.59	6.49	8.40
37	H	H	H	Cl	H	7.11	7.01	6.89	8.40
39	H	H	Cl	H	H	6.76	6.60	6.51	7.96
53	Fenilo	H	H	H	H	6.22	6.09	6.06	7.40
57	Cl	H	H	H	Cl	5.84	5.75	5.76	7.72

**Tabla A4.** Estructuras y actividades de los compuestos utilizados en el caso de estudio.



Comp.	Sustituciones	Valores de pIC <sub>50</sub>			
	R1	HDAC1	HDAC2	HDAC3	HDAC6
26		6.69	6.44	6.66	8.00
52		6.26	6.15	5.96	8.05
96		6.24	6.08	5.96	7.34
97		6.09	5.93	5.93	7.31
103		5.85	5.79	5.46	7.39
104		6.04	6.12	5.88	7.72
135		6.49	6.48	5.75	8.30

Tabla A5. Estructuras y actividades de los compuestos utilizados en el caso de estudio.

Comp.	Sustituciones		Valores de pIC <sub>50</sub>			
	R1	R2	HDAC1	HDAC2	HDAC3	HDAC6
						
38	H		7.05	6.92	6.97	8.40
41	H		7.40	7.28	7.24	8.70
62	H		6.69	6.62	6.52	8.30
70	H		6.48	6.47	6.36	8.40
78	CH <sub>3</sub>		7.17	7.02	6.87	8.40
91	(CH <sub>2</sub> ) <sub>2</sub>		6.83	6.62	5.37	8.40
92	(CH <sub>2</sub> ) <sub>2</sub>		6.70	6.49	6.46	8.40
95	(CH <sub>2</sub> ) <sub>2</sub>		7.89	7.68	7.48	8.70
98	H		6.67	6.44	6.44	8.40
99	H		7.04	6.88	6.76	8.40
101	H		6.66	6.50	6.45	8.30
102	(CH <sub>2</sub> ) <sub>3</sub>		6.49	6.41	6.07	8.40
113	(CH <sub>2</sub> ) <sub>2</sub>		7.80	7.77	7.31	8.70
139	(CH <sub>2</sub> ) <sub>5</sub>		5.86	5.80	5.47	8.30

**APÉNDICE B**

*Publicaciones y Congresos*

- 51° Congreso Mexicano de Química y el 35° Congreso Nacional de Educación Química (cartel y constancia)



**EDUQ- E2-C**

## Elaboración de un manual teórico-práctico de quimioinformática

**Fernanda I. Saldivar González, Mariana González Medina, José L. Medina Franco**  
 DIFACQUIM, Cubículos 305 y 108, Edificio F, Facultad de Química (UNAM)  
 Teléfono: 56223899, ext. 44458  
 felilang\_12@hotmail.com, medinajl@unam.mx



Introducción
Resultados

La 'quimioinformática' es una herramienta que surge de la combinación de recursos informáticos y los datos químicos y se emplea en el manejo, visualización y análisis sistemático de información química [1]. Esta disciplina puede aplicarse a diferentes áreas de estudio, sin embargo, es en la Química Farmacéutica donde se le asocia con frecuencia, ya que es de gran ayuda en el descubrimiento, diseño y desarrollo de compuestos bioactivos.

Con el objetivo de fortalecer la enseñanza y difusión de los conceptos y métodos fundamentales en quimioinformática, se elaboró un manual teórico-práctico, que incluye ejercicios con enfoques actuales para fomentar el interés de los alumnos de la licenciatura Química Farmacéutico Biológica hacia esta disciplina.

### Justificación

La enseñanza de las bases teóricas de la quimioinformática aplicada a la Química Farmacéutica es escasa y en ocasiones temas de relevancia quedan fuera de los contenidos de cursos actuales de licenciatura. Esto se debe en parte a la cantidad limitada de libros de texto especializados en esta área y a que la literatura en español sobre quimioinformática es aún más restringida.

La implementación de métodos quimioinformáticos en Química Farmacéutica es de suma importancia para sentar las bases del diseño racional de nuevos fármacos.

### Contenido



5 Capítulos

- Capítulo 1. Bases de datos. Bibliotecas de compuestos químicos.
- Capítulo 2. Espacio químico
- Capítulo 3. Relaciones cuantitativas: estructura-actividad
- Capítulo 4. Acoplamiento molecular automatizado
- Toxicología informática
- Ejercicios finales (integrativos)

- Objetivo General
- Objetivos específicos
- Introducción (conceptos más importantes)
- Procedimiento
- Ejercicios prácticos
- Evaluación corta

Mediante artículos científicos donde se utilizan más de un método quimioinformático

### ¿Qué podrán aprender en cada capítulo?



**Cap. 1**

Realizar búsquedas de información en bases de datos usadas con frecuencias en la investigación de fármacos de una forma sencilla y eficiente. Construir bases de datos e identificar moléculas que puedan interferir en cálculos computacionales.



**Cap. 2**

Identificar los diferentes tipos de descriptores moleculares y su utilidad. Obtener perfiles de las bases de datos generadas mediante métodos de visualización.



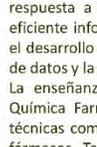
**Cap. 3**

Llevar a cabo la optimización de la actividad biológica de compuestos activos y la identificación sistemática de acantilados de actividad (en inglés, *activity cliffs*)



**Cap. 4**

Identificar los requisitos en cuanto a la presentación del ligando y de la diana biológica establecidos por diferentes herramientas para realizar acoplamiento molecular.



**Cap. 5**

Conocer diferentes bases de datos con información relacionada a la toxicología de sustancias químicas. \*Identificar compuestos con bajo riesgo toxicológico y seleccionar candidatos para la optimización y desarrollo de fármacos potenciales.

Manual disponible en: [www.difacquim.com/educación](http://www.difacquim.com/educación)



### Conclusiones

La quimioinformática es una disciplina científica que ha surgido en respuesta a la necesidad de manejar, clasificar e interpretar en forma eficiente información química, una de sus grandes aplicaciones ha sido en el desarrollo de fármacos, ya que permite el manejo de grandes cantidades de datos y la predicción de la actividad biológica.

La enseñanza de esta disciplina a estudiantes cuya orientación es la Química Farmacéutica puede contribuir a fomentar el interés hacia las técnicas computacionales empleadas en el diseño y desarrollo de nuevos fármacos. También puede ayudar a priorizar y racionalizar los costos del desarrollo de diversos fármacos.

### Referencias

1. Saldivar-González F, Prieto-Martínez FD; Medina-Franco JL. Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química* 2016, <http://dx.doi.org/10.1016/j.eq.2016.06.002>.
2. Medina-Franco JL, Fernández-de-Gortari E, Naveja JJ. Avances en el diseño de fármacos asistido por computadora. *Educación Química* 2015, 26, 180-186.

### Agradecimientos

• Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIIME) PE200116. Se agradecen comentarios y sugerencias del grupo DIFACQUIM en especial al QFI. Fernando Prieto Martínez y Dr. Oscar Méndez Lucio.



# SOCIEDAD QUÍMICA DE MÉXICO, A.C.

*"La química nos une."*

otorga el presente

## *Reconocimiento*

*a*  
*Fernanda Isabel Saldívar González, Mariana González Medina, José Luis Medina Franco*

*por su participación presentando el Trabajo Estudiantil en Modalidad Cartel titulado*

**Elaboración de un manual teórico-práctico de quimioinformática. EDUQ-E2-C**

*en el marco de las actividades del*

**51° Congreso Mexicano de Química**

**35° Congreso Nacional de Educación Química**

Pachuca de Soto, Hidalgo, México.

28 de Septiembre a 1 de Octubre de 2016

Dr. Benjamín Velasco Bejarano  
Presidente Nacional  
Sociedad Química de México, A.C.

Dra. Mayra del Jesús Rosales Hoz  
Presidente de Congresos  
Sociedad Química de México, A.C.

Dr. Antonio Rafael Tapia Benavides  
Presidente Local de Congresos, 2016  
IIAFH

- 12th German Conference on Chemoinformatics llevado a cabo del 6 al 8 de noviembre del 2016 en Fulda, Alemania.

**FINAL PROGRAMME · BOOK OF ABSTRACTS**



GESELLSCHAFT DEUTSCHER CHEMIKER



Fachgruppe  
Chemistry – Information – Computer

**GCC 2016**  
12<sup>th</sup> German  
Conference on Chemoinformatics  
November 6 – 8, 2016  
Fulda · Germany

[www.gdch.de/gcc2016](http://www.gdch.de/gcc2016)



**P12**

**Teaching chemoinformatics at the undergraduate level**  
 Fernanda Saldívar-González, Mariana González-Medina, Alfonso Lira-Rocha,  
 Francisco Hernández-Luís, José L. Medina-Franco  
 Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de  
 México, Mexico City, 04510, Mexico  
 \*medinajl@unam.mx

Despite the fact chemoinformatics is used in many areas of chemistry, the number of text books [1] and courses to teach this discipline at the undergraduate level are limited. Herein, we discuss the development of a manual to teach the basic principles and common applications of chemoinformatics with special emphasis on drug discovery applications [2]. The manual is designed to teach students chemoinformatics at the undergraduate level, in particular, students taking the courses of Medicinal Chemistry. The manual has five major sections: A) Compound databases, B) Chemical Space, C) Structure-Activity Relationships with emphasis on Activity Landscape Modelling, D) Molecular Docking and E) Informatics Toxicology. All sections include basic theory with relevant citations, hands-on exercises, and quizzes.

[1] Johann Gasteiger, Thomas Engel: Chemoinformatics, A Textbook. Wiley-VCH; 2003.  
 [2] Saldívar-González F, Prieto-Martínez FD, Medina-Franco JL. **Drug discovery and development: A computational approach** *Edúc. Quím.* 2016, in press.



## COMUNICACIÓN

# Descubrimiento y desarrollo de fármacos: un enfoque computacional



Fernanda Saldívar-González, Fernando D. Prieto-Martínez y José L. Medina-Franco\*

*Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, Ciudad de México, México*

Recibido el 5 de abril de 2016; aceptado el 20 de junio de 2016

Disponible en Internet el 12 de septiembre de 2016

### PALABRAS CLAVE

Acoplamiento molecular;  
Descubrimiento de fármacos;  
Modelado molecular;  
Quimioinformática;  
Relaciones estructura-actividad

### KEYWORDS

Molecular docking;  
Drug discovery;  
Molecular modeling;  
Chemoinformatics;  
Structure-activity relationships

**Resumen** El descubrimiento de fármacos es un proceso complejo y costoso en el cual convergen diversas áreas del conocimiento. En años recientes métodos computacionales se han integrado a este esfuerzo multidisciplinario y su enseñanza en cursos de Química Farmacéutica es fundamental. En un proyecto determinado, la aplicación de estrategias de cómputo depende de la información disponible del sistema y de los objetivos específicos del estudio. A la fecha, los métodos computacionales han contribuido, entre otras aplicaciones, al análisis eficiente de datos, al filtrado de colecciones de compuestos para seleccionar moléculas para evaluación experimental, a la generación de hipótesis para ayudar a entender el mecanismo de acción de fármacos y al diseño de nuevas estructuras químicas. Además, los métodos de cómputo han tenido aportaciones significativas para desarrollar medicamentos que se encuentran en uso clínico. Sin embargo, quedan muchos retos que afrontar, los mismos que estimulan la innovación y el mejoramiento de métodos que se integren en el esfuerzo multidisciplinario del desarrollo de fármacos.

© 2016 Universidad Nacional Autónoma de México, Facultad de Química. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Drug discovery and development: A computational approach

**Abstract** Drug discovery is a complex and expensive process where different research areas converge. Computational methods have been part of the multidisciplinary efforts and their principles should be included in courses of Medicinal Chemistry. In a given project, the application of computational approaches depends on the information available for the system and the specific goals of the study. Computational approaches have made key contributions to perform

\* Autor para correspondencia.

Correos electrónicos: [medinajl@unam.mx](mailto:medinajl@unam.mx), [jose.medina.franco@gmail.com](mailto:jose.medina.franco@gmail.com) (J.L. Medina-Franco).

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

efficient analyses of data, filtering compounds collections to select molecules for experimental screening, generate hypothesis to understand the mechanism of action of drugs, and the design of new chemical structures. In addition, computational methods have made significant contributions to develop drugs that are in clinical use. However, there are several challenges to face. Addressing these challenges promote innovation and improvement of methodologies that form part of the multidisciplinary effort to develop drugs.

© 2016 Universidad Nacional Autónoma de México, Facultad de Química. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introducción

El desarrollo de fármacos es un proceso complejo el cual inicia con la identificación de compuestos que se unen a un blanco terapéutico o que muestran actividad biológica en un ensayo de tamizaje. Aquellas moléculas que muestran actividad biológica son llamadas *hits*. El siguiente paso es encontrar compuestos que tengan propiedades farmacéuticas atractivas, incluyendo baja toxicidad, solubilidad acuosa adecuada para administrarse vía oral, entre otras propiedades farmacocinéticas. Tales compuestos son llamados «líderes o cabezas de serie». Típicamente, los *hits* son encontrados por tamizaje de un número vasto de moléculas, mientras que los compuestos «cabezas de serie» son desarrollados a partir de los *hits* a través de modificaciones químicas. Considerando que el número de moléculas orgánicas que son sintéticamente factibles se encuentran entre  $10^{20}$  y  $10^{24}$  (Ertl, 2003), es evidente que su análisis sería muy complejo sin el uso de técnicas computacionales.

En este artículo se discuten métodos computacionales comunes que se emplean durante el desarrollo de fármacos. El objetivo es contribuir a la enseñanza de Química Farmacéutica. Trabajos relacionados han sido publicados en la revista *Educación Química*; a diferencia del trabajo de Medina-Franco, Lopez-Vallejo y Castillo (2006) publicado hace 10 años, este artículo enfatiza la aplicación de métodos computacionales en el contexto de conceptos novedosos que han surgido en el diseño de fármacos. Aplicaciones específicas al desarrollo de medicamentos dirigidos a dianas epigenéticas se encuentran publicadas en Medina-Franco, Méndez-Lucio, Yoo y Dueñas, (2015). Se hace notar que el enfoque del artículo publicado en 2015 está centrado en 4 líneas de investigación específicas y casos de estudio particulares.

## Diseño y desarrollo de fármacos

La mayoría de los medicamentos que están en uso clínico son el resultado de un proceso de investigación muy complejo. Por lo mismo, es necesaria la unión de esfuerzos de diferentes disciplinas científicas para descubrir y desarrollar medicamentos con efectos clínicos benéficos y efectos secundarios mínimos. Aunque el descubrimiento y desarrollo de medicamentos se ha hecho durante muchos años usando únicamente métodos experimentales, se espera que el proceso se acelere gracias al uso de métodos

de cómputo (también llamados *in silico*) que permiten codificar con precisión modelos teóricos y son capaces de procesar grandes cantidades de información. Además, en muchos proyectos, la aplicación de modelos generados *in silico* contribuye a entender los mecanismos de acción de los principios activos de los medicamentos o a mejorar las propiedades de los mismos. Un ejemplo es la optimización de los fármacos de origen natural. En la actualidad se hacen esfuerzos para mejorar las propiedades de estos medicamentos (como reducir efectos secundarios) o buscar actividades biológicas potenciales de compuestos químicos aislados de fuentes naturales (Medina-Franco, 2013).

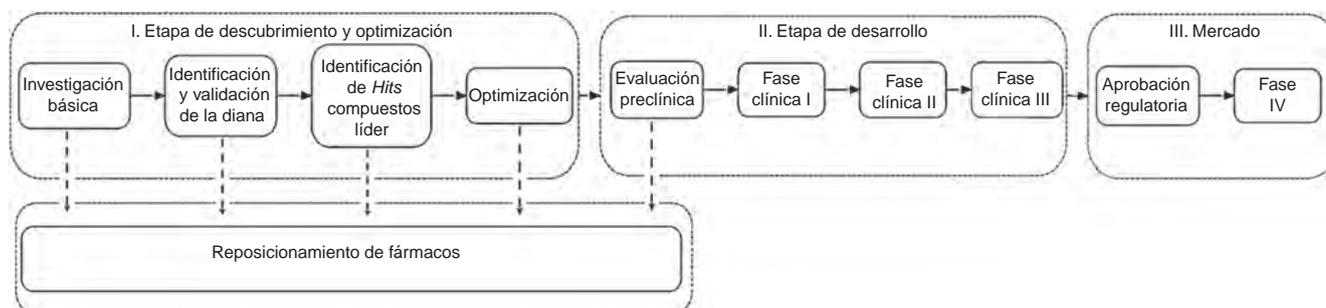
## Etapas del desarrollo de fármacos

La figura 1 muestra las etapas principales de un modelo clásico para el desarrollo de un medicamento. El proceso comienza con la investigación sobre las causas de una enfermedad, que en algunos casos puede llevar a la identificación de una o varias dianas moleculares asociadas con esa enfermedad. Los pasos siguientes involucran la identificación de compuestos activos con la diana molecular y la optimización de su actividad biológica. Estos ensayos se hacen *in vitro* con blancos moleculares aislados de las células. Los compuestos activos se someten a varias evaluaciones experimentales que implican ensayos en líneas celulares, en animales y pruebas clínicas en humanos. Los compuestos que pasan satisfactoriamente por todas las etapas son aprobados para uso clínico por un agente regulatorio, por ejemplo, la Comisión Federal para la Protección contra Riesgos Sanitarios (COFEPRIS) en México o la *Food and Drug Administration* (FDA) en Estados Unidos.

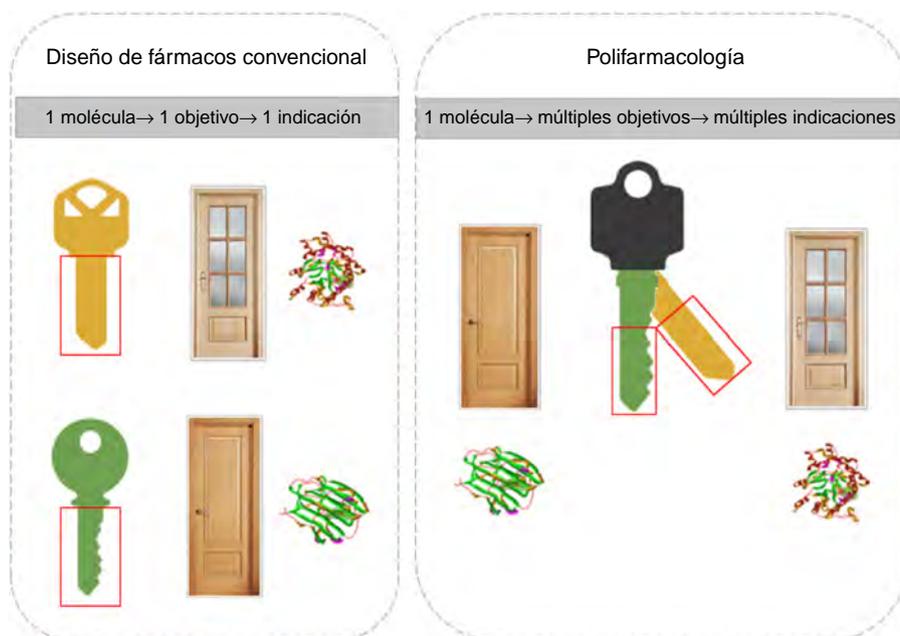
## Probabilidades de éxito

En el desarrollo de fármacos, la mayoría de los compuestos que muestran actividades *in vitro* con las dianas moleculares fallan las pruebas siguientes. Esto se debe frecuentemente a sus pobres propiedades farmacocinéticas y toxicidad. Es decir, además de que un compuesto es activo con los blancos moleculares deseados, también afecta otros procesos fisiológicos y no pueden usarse en forma segura en humanos. Se estima que de cada 9,000 moléculas biológicamente activas solo una tiene uso clínico.

A pesar de que el modelo de desarrollo de fármacos de la figura 1 se sigue aplicando con éxito, no necesariamente es el más eficiente. Cada vez hay mayor evidencia que un



**Figura 1** Etapas principales en el desarrollo clásico de un medicamento.



**Figura 2** Concepto de polifarmacología y diseño de «llaves maestras».

fármaco interacciona con diversos blancos moleculares. En varios casos el efecto clínico se debe a la interacción con múltiples dianas, dando origen al concepto de polifarmacología. De esta manera, la polifarmacología está modificando el paradigma actual de diseño de fármacos, de un diseño dirigido a una sola diana al diseño dirigido simultáneamente a múltiples blancos terapéuticos. Esta estrategia se conoce como diseño multiobjetivo y una de las metas es diseñar una «llave maestra» que se una selectivamente a una serie de blancos moleculares que produzcan una respuesta clínica deseada. La polifarmacología y el diseño de fármacos que actúan como «llaves maestras» se ilustran en la [figura 2](#).

#### Costo y tiempo aproximado

Se estima que el desarrollo de un fármaco tarda aproximadamente entre 10 y 15 años y se invierten en promedio 800 millones de dólares, teniendo el mayor costo y tiempo las pruebas clínicas en humanos. El tiempo y costos tan elevados están asociados en gran medida a la gran cantidad de moléculas que fallan una o varias etapas del desarrollo de fármacos.

#### Estrategias para identificar compuestos líderes

Hay diversas estrategias experimentales y/o computacionales para identificar compuestos activos que se consideran líderes para implementar un programa de optimización:

- Optimización de fármacos ya aprobados para su uso clínico. La optimización puede consistir en mejorar la eficacia del medicamento o reducir efectos secundarios.
- Ensayo biológico sistemático de colecciones de compuestos. Un caso típico son las pruebas biológicas de alto rendimiento (en inglés *high-throughput screening* [HTS]). En HTS se usan robots para evaluar en pocas horas colecciones con miles o millones de moléculas.
- Uso de información biológica disponible. Por ejemplo, si se conoce el mecanismo de acción de un sustrato se diseñan compuestos que se espera sigan el mismo mecanismo de acción.
- Diseño de fármacos asistido por computadora (DIFAC). Esta estrategia involucra una o varias técnicas que se discuten en la siguiente sección.

## Diseño de fármacos asistido por computadora

El DIFAC tiene como principio entender las relaciones estructura-actividad biológica o farmacológica de compuestos (Medina-Franco et al., 2006).

### Objetivos del DIFAC

Los objetivos del DIFAC pueden dividirse en 3:

- *Diseño o identificación de nuevos compuestos.* Esto es, diseñar nuevas estructuras con efecto biológico en una categoría terapéutica deseada. Una alternativa es buscar dentro de colecciones de compuestos existentes moléculas que tengan una actividad biológica específica.
- *Seleccionar candidatos.* Los métodos computacionales ayudan a distinguir las moléculas sobre las cuales se deben enfocar primero las pruebas experimentales, por ejemplo, compra, síntesis, y evaluación biológica. Sin embargo, los métodos de cómputo no reemplazan a los experimentos.
- *Optimizar líderes* para mejorar las propiedades deseadas y disminuir los efectos adversos.

### El DIFAC es una integración de métodos

El DIFAC está integrado por diversas áreas de la investigación que incluyen a la quimioinformática, bioinformática, modelado molecular, química teórica y visualización de datos. Cada una comprende métodos que se emplean dependiendo de las características del sistema, información experimental disponible y de los objetivos específicos del proyecto.

### Falsas expectativas del DIFAC

El concepto de «diseño de fármacos por computadora» está frecuentemente asociado con falsas expectativas. La [tabla 1](#) resume ejemplos los cuales se describen brevemente a continuación.

Una falsa expectativa es que las computadoras pueden diseñar medicamentos por sí solas. Esta percepción conduce con frecuencia a la pregunta: ¿algún medicamento se ha diseñado por métodos computacionales? Los métodos de cómputo no diseñan de forma automatizada los medicamentos. Para que sean efectivos, los modelos teóricos deben integrarse como pruebas biológicas o biofísicas. Los métodos de cómputo forman parte de un esfuerzo multidisciplinario que involucra otras áreas como síntesis orgánica, productos naturales y HTS. Por ejemplo, los cálculos computacionales no demuestran que una molécula puede unirse a una enzima. En la discusión de acoplamiento molecular automatizado se verá que los cálculos «asisten» en la predicción del modo de unión de una molécula con un blanco molecular, pero no demuestran dicha unión. Otras falsas expectativas son que se pueden diseñar fármacos «apretando botones» y que los análisis de cómputo son rápidos. Si es cierto que el procesamiento de datos es rápido (dependiendo del tipo de cálculo, equipo de cómputo y sistema en estudio), el análisis e interpretación de datos son laboriosos. En forma semejante a una prueba experimental, hay que

validar los métodos computacionales y encontrar los parámetros óptimos para que los cálculos estén apegados a la realidad.

### Métodos computacionales comunes

Como se mencionó arriba, el desarrollo de fármacos involucra varias etapas que abarcan desde la identificación de dianas moleculares hasta las fases clínicas ([fig. 1](#)). La mayoría de las técnicas clásicas del DIFAC están centradas en las primeras etapas. En esta sección se discuten 3 métodos que se emplean con frecuencia en la identificación de compuestos líder y su optimización.

### Acoplamiento molecular automatizado

El acoplamiento molecular automatizado (o en inglés *molecular docking*) consiste en buscar la conformación y posición óptima de un ligando (por ejemplo, de una molécula orgánica pequeña) dentro de una diana molecular (por ejemplo, una enzima, un canal iónico o un receptor acoplado a proteína G). La [figura 3](#) ilustra un modelo de acoplamiento de la isoforma BRD4 (*bromodomain* de la familia BET), una diana epigenética. Mediante búsquedas por similitud molecular se encontró que la pentoxifilina es un potencial inhibidor de BRD4; en púrpura se muestra la conformación del ligando propuesto, que es semejante a la orientación del ligando cocrystalizado (A). ¿Cómo podemos confiar en el resultado? Para ello aplicamos el mismo algoritmo sobre el ligando cocrystalizado, para comprobar si es capaz de recuperar el conformero observado experimentalmente (B).

Es posible hacer acoplamiento molecular automatizado entre 2 macromoléculas, como 2 proteínas. Considerando la flexibilidad de las moléculas, el número de posibles conformaciones puede ser muy elevado. De igual manera, si la cavidad del sitio receptor es grande y/o flexible, es más complicado encontrar la posición y orientación que tendrá la molécula pequeña dentro del receptor. Es así que el uso de computadoras ayuda a acelerar el proceso de búsqueda y sugiere modelos de unión.

El acoplamiento molecular automatizado tiene 2 componentes: a) *docking* o proceso de búsqueda de la conformación y orientación de las moléculas, y b) *scoring*, que consiste en asignar un valor o puntaje que mida la interacción entre las 2 estructuras.

De las 2 etapas, la segunda es la más difícil de calcular con precisión de forma rápida. Es por esto que en la actualidad el acoplamiento molecular es especialmente útil para proponer modelos de unión, pero aún no son muy confiables para calcular con precisión la energía de interacción. Esto se debe a la gran cantidad de aproximaciones que se hacen para calcular la energía de interacción de forma rápida. Por ejemplo, no se considera en detalle la flexibilidad y solvatación de la diana molecular. Para solucionar este problema los modelos obtenidos por acoplamiento molecular se someten a otros cálculos más refinados que se calculan con niveles de teoría elevados.

**Tabla 1** Resumen de falsas expectativas típicas del DIFAC y visión realista

Falsa expectativa	Visión realista
Las computadoras pueden diseñar fármacos	Los métodos computacionales no diseñan fármacos por sí solos; forman parte de un esfuerzo multidisciplinario para alcanzar un objetivo extremadamente complejo
Un cálculo teórico (por ejemplo de acoplamiento molecular o <i>docking</i> ) demuestra la unión fármaco-blanco molecular	Cálculos de acoplamiento molecular dan hipótesis y proponen modelos de unión; se utilizan como apoyo a los experimentos que son los que demuestran la unión
Se puede diseñar fármacos apretando botones ( <i>push-a-bottom drug discovery</i> )	El DIFAC es un proceso que requiere gran cantidad de preparación, análisis e interpretación de resultados que se obtienen de la computadora
El DIFAC es rápido porque lo hacen las computadoras	No; la preparación de datos e interpretación de resultados, que forman parte fundamental del DIFAC, son procesos que requieren de un tiempo que puede durar meses

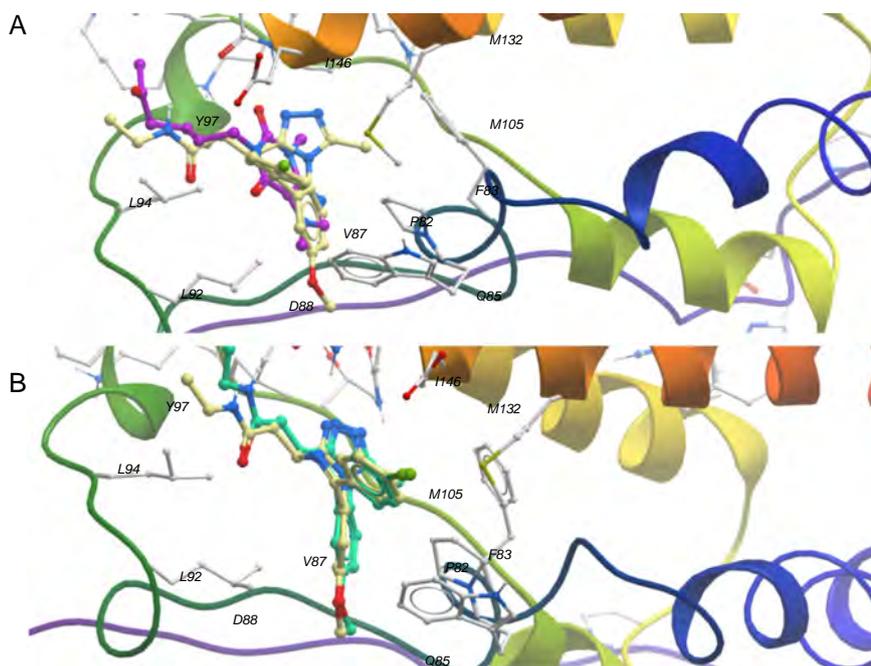
### Modelado del farmacóforo

Otro método computacional muy utilizado es el modelo del «farmacóforo». Este se define como un arreglo tridimensional de las características mínimas necesarias estéricas y electrónicas para asegurar interacciones óptimas con un blanco farmacológico específico, lo cual desencadenará o bloqueará una respuesta biológica. Aunque la representación visual de un farmacóforo puede ser bi o tridimensional, es importante entender que un farmacóforo no representa a una molécula real o una asociación real de grupos funcionales. Un modelo de farmacóforo es un concepto abstracto que indica la capacidad de interacción molecular común de un grupo de compuestos dirigidos a un blanco

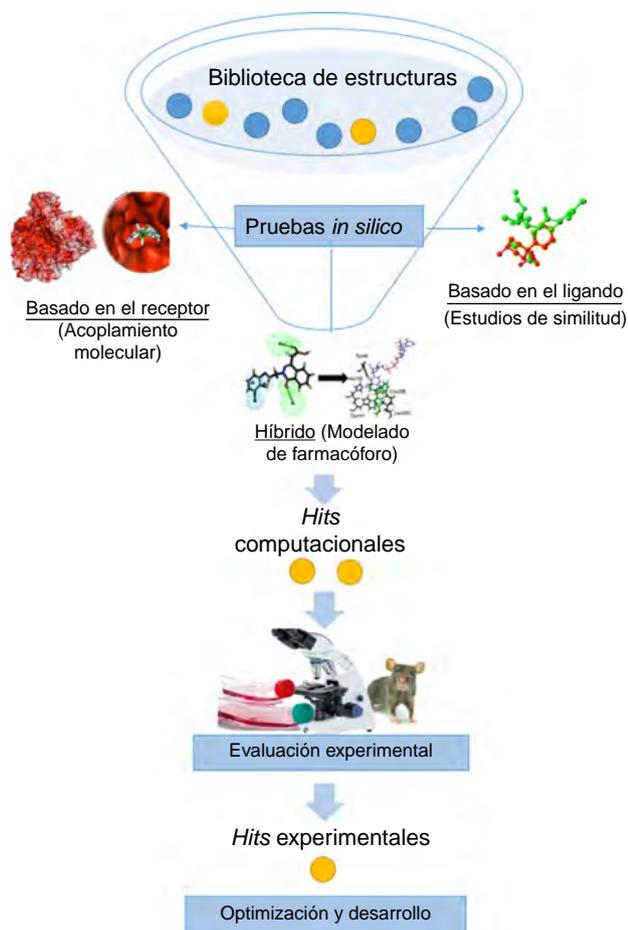
farmacológico específico. En otras palabras, el farmacóforo puede ser considerado como el común denominador de un conjunto de moléculas activas.

### Cribado virtual

El cribado virtual es un filtrado computacional (*in silico*) de moléculas para seleccionar candidatos i.e., *hits* computacionales, para su evaluación experimental (fig. 4) (Scior et al., 2012). De esta manera, el cribado virtual reduce significativamente el número de ensayos biológicos que se harían si no hubiera una selección de compuestos. Sin embargo, es un proceso predictivo que debe integrarse con ensayos



**Figura 3** Modelo de acoplamiento de la diana epigenética BRD4 (PDB 3P50); A) Modelo de acoplamiento de pentoxifilina (átomos de carbono en púrpura) respecto al ligando cocristalizado (amarillo). B) Validación del método de acoplamiento: la conformación en azul es la predicción del algoritmo de *docking*.



**Figura 4** Representación esquemática del proceso de cribado virtual.

experimentales que validen las predicciones de los ensayos *in silico*.

Existen diversos filtros que se utilizan para llevar a cabo el cribado virtual los cuales pueden variar según la complejidad de la base de datos y la información experimental de la que se disponga. Por ejemplo, si se conoce la estructura tridimensional (3D) del receptor se sugiere un cribado basado en la estructura (acoplamiento molecular). Si solo se conocen los compuestos activos, pero no el receptor, entonces la búsqueda se hace basada en el ligando (similitud molecular). Si se conoce la estructura 3D del receptor y de los compuestos activos se pueden combinar filtros para facilitar la búsqueda, tales como descriptores moleculares y propiedades farmacocinéticas, entre otras. Las técnicas para hacer cribado virtual dependen de la información disponible del sistema.

En nuestro grupo de investigación se ha utilizado el cribado virtual usando acoplamiento molecular para la identificación de nuevos inhibidores de las enzimas AKT-2 y ADN metiltransferasa (Kuck, Singh, Lyko y Medina-Franco, 2010; Medina-Franco et al., 2009). Ambas son dianas moleculares para el desarrollo de compuestos contra el cáncer. Los *hits* experimentales han servido como puntos de partida de programas de optimización (Hernández-Campos et al., 2010; Medina-Franco, Fernández-de-Gortari y Naveja, 2015).

### Cribado virtual de fármacos aprobados: reposicionamiento de fármacos

El reposicionamiento de fármacos consiste en encontrar una nueva aplicación terapéutica distinta por la que fue diseñado un fármaco y que no necesariamente tiene que ser la misma dosis. El objetivo de esta estrategia es reducir los gastos y tiempos de investigación. Casos exitosos de reposicionamiento de fármacos se han discutido recientemente (Naveja, Dueñas-González y Medina-Franco, 2016).

A partir de la hipótesis de que «moléculas similares tienen propiedades similares», se han aplicado estrategias computacionales encaminadas a reposicionar fármacos. Un ejemplo es el esfuerzo de reposicionar la olsalazina como anticancerígeno. La olsalazina es un antiinflamatorio aprobado para su uso clínico. La olsalazina fue identificada como un agente hipometilante gracias a la búsqueda de similitud asistida por computadora en DrugBank (Méndez-Lucio, Tran, Medina-Franco, Meurice y Muller, 2014).

### Quimioinformática

La quimioinformática es una herramienta que surge de la combinación de recursos informáticos y datos químicos, y se emplea en el manejo, visualización y análisis sistemático de información química. Esta herramienta permite analizar eficientemente miles de datos. A través de métodos quimioinformáticos es posible manejar y organizar información, visualizar el espacio químico, hacer minería de datos y establecer asociaciones matemáticas entre la estructura y actividad. Un ejemplo clásico son las relaciones cuantitativas estructura-actividad (QSAR, por sus siglas en inglés).

### Bases de datos públicas

La quimioinformática tiene aplicaciones especialmente valiosas en el análisis y manejo de bases de datos moleculares. Ejemplos de colecciones usadas con frecuencia en descubrimiento de fármacos son ZINC, DrugBank, ChEMBL, Binding Database y PubChem. Estas colecciones discutidas en otros artículos (Nicola, Liu y Gilson, 2012; Scior, Bernard, Medina-Franco y Maggiora, 2007) contienen información para realizar cribado virtual, quimiogenómica, análisis estructura-actividad y reposicionamientos de fármacos, entre otros.

### Análisis computacionales en línea

La información contenida en bases de datos públicas se puede analizar con recursos integrados en la misma base de datos. Un ejemplo de este tipo de plataformas es *PubChem BioAssay Database* (<http://pubchem.ncbi.nlm.nih.gov>). PubChem también tiene herramientas que permiten hacer el análisis de todos los datos, tales como análisis de relaciones estructura-actividad y agrupamiento.

### Relaciones estructura-actividad

Cuando no se conoce la estructura del receptor se utilizan métodos basados en el ligando, los cuales dependen de la información disponible para estructuras químicas con

**Tabla 2** Ejemplos recientes de fármacos desarrollados con la ayuda de métodos computacionales

Fármaco	Aplicación	Técnica computacional	Farmacéutica
Crizotinib	Agente antineoplásico (cáncer de pulmón)	Acoplamiento molecular	Pfizer
Rilpivirina	Antiviral (infección por el VIH)	Acoplamiento molecular	Janssen Pharmaceuticals
Zelboraf	Agente antineoplásico (melanoma)	Acoplamiento molecular y análisis de subestructuras	Roche
Boceprevir	Antiviral (infección por el virus de la hepatitis C, VHC)	Acoplamiento molecular	Merck & Co

actividad biológica conocida. Métodos comunes son estudios SAR (cualitativos) y QSAR (cuantitativos).

### QSAR y métodos comunes

De acuerdo con el origen de los descriptores moleculares usados en los cálculos, los métodos QSAR pueden dividirse en 3 grupos. Uno de ellos usa un número relativamente pequeño de descriptores que describen, por ejemplo, efectos hidrófobos, estéricos y electrostáticos. Otro grupo se basa en propiedades calculadas a partir de la conectividad molecular. Debido a que las fórmulas estructurales son bidimensionales, estos métodos se conocen como estudios QSAR en 2 dimensiones o QSAR-2D. Un tercer grupo de métodos se basa en los descriptores obtenidos de la representación tridimensional de las estructuras y se les conoce como QSAR en 3 dimensiones o QSAR-3D. Uno de los ejemplos más conocidos de QSAR-3D es el análisis comparativo de campos moleculares (del inglés, *comparative molecular field analysis* [CoMFA]). Los estudios QSAR también pueden clasificarse por el tipo de métodos de correlación empleados; estos pueden ser métodos lineales, como es el caso de la regresión lineal o regresión lineal múltiple, y regresión no lineal.

### Panoramas de actividad

El modelado de panoramas de actividad (en inglés, *activity landscape modeling*) tiene como objetivo describir sistemáticamente cambios en la actividad biológica asociados con cambios en las estructuras químicas. Una de las aplicaciones es identificar casos en donde pequeños cambios estructurales están asociados a un cambio grande en la actividad biológica. Estos casos llamados «acantilados de actividad» (en inglés, *activity cliffs*) son excepciones al «principio de similitud» y tienen un impacto importante en el diseño de fármacos (Medina-Franco et al., 2013).

### Casos exitosos

Los cálculos computacionales han tenido un papel significativo en la investigación de moléculas que actualmente se encuentran en uso clínico. Por ejemplo, el DIFAC ya ha tenido contribuciones notables en el tratamiento del síndrome de la inmunodeficiencia adquirida, en infecciones por el virus de la influenza y en el tratamiento del glaucoma (Medina-Franco, 2007; Talele, Khedkar y Rigby, 2010). En la tabla 2 se resumen ejemplos de casos de éxito recientes.

En agosto de 2011, crizotinib (Xalkori; Pfizer) fue aprobado por la FDA para el tratamiento de pacientes con cáncer de pulmón de células no pequeñas (NSCLC). Crizotinib fue

descubierto a través de estudios para identificar inhibidores similares a fármacos potentes de la tirosina quinasa del receptor de MET. Posteriormente, también se observó que era un inhibidor potente de ALK. Este descubrimiento se dio a partir de ensayos biológicos y de estudios de acoplamiento molecular. Actualmente se ha observado un desarrollo de resistencia adquirida hacia el crizotinib, y entre los mecanismos identificados se encuentra la adquisición de mutaciones de resistencia secundaria dentro del dominio de tirosina quinasa de ALK, los cuales también pueden ser estudiados por medio de estudios de acoplamiento molecular y dinámica molecular (Cui et al., 2011).

### Conclusiones

En los últimos años el DIFAC ha tenido un desarrollo sin precedentes gracias a los avances de diversas disciplinas como la biología molecular, química teórica, quimiinformática, ciencias de la computación y avances tecnológicos. La incorporación de métodos de cómputo al esfuerzo multidisciplinario que implica el desarrollo de medicamentos ha llevado al desarrollo de fármacos que se encuentran en uso clínico. Además, los métodos de cómputo hacen constantes aportaciones a proyectos de investigación para analizar datos en forma eficiente y plantear hipótesis valiosas que guían el diseño de nuevos fármacos. Se espera que aumente el número y calidad de beneficios que aportan los métodos *in silico* al diseño de medicamentos.

### Perspectivas

El DIFAC tiene aún muchos retos que, sin estar limitado a ellos, incluyen: a) incrementar la eficiencia del cribado virtual; b) aumentar el número y la calidad de los recursos computacionales en línea; c) desarrollar el campo de la quimiogenómica computacional; d) fortalecer el diseño de fármacos dirigidos a múltiples dianas moleculares; e) mejorar la capacidad predictiva de modelos de toxicidad y efectos secundarios, y f) fortalecer la interacción con otras disciplinas para optimizar la búsqueda de moléculas bioactivas para el tratamiento y/o prevención de enfermedades.

### Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Agradecimientos

Se agradecen las discusiones con Vicente F. Kuyoc y el apoyo de la UNAM a través del Programa de Apoyo a Proyectos Para la Innovación y Mejoramiento de la Enseñanza (PAPIME) N.º PE200116. FDP-M agradece al CONACyT la beca N.º 660465/576637. También se agradece al Programa de Apoyo a la Investigación y el Posgrado (PAIP) 5000-9163, Facultad de Química de la UNAM.

## Referencias

- Cui, J. J., Tran-Dubé, M., Shen, H., Nambu, M., Kung, P. P., Pairish, M., et al. (2011). Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition Factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). *Journal of Medicinal Chemistry*, *54*(18), 6342–6363.
- Ertl, P. (2003). Cheminformatics analysis of organic substituents: Identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *Journal of Chemical Information and Computer Sciences*, *43*(2), 374–380.
- Hernández-Campos, A., Velázquez-Martínez, I., Castillo, R., López-Vallejo, F., Jia, P., Yu, Y., et al. (2010). Docking of protein kinase b inhibitors: Implications in the structure-based optimization of a novel scaffold. *Chemical Biology and Drug Design*, *76*(3), 269–276.
- Kuck, D., Singh, N., Lyko, F. y Medina-Franco, J. L. (2010). Novel and selective DNA methyltransferase inhibitors: Docking-based virtual screening and experimental evaluation. *Bioorganic and Medicinal Chemistry*, *18*(2), 822–829.
- Medina-Franco, J. L. (2007). Aplicaciones exitosas del diseño de fármacos utilizando métodos computacionales. *Ciencia (Mexico)*, *59*(1), 1–8.
- Medina-Franco, J. L. (2013). Advances in computational approaches for drug discovery based on natural products. *Revista Latinoamericana de Química*, *41*(2), 95–110.
- Medina-Franco, J. L., Edwards, B. S., Pinilla, C., Appel, J. R., Giulianotti, M. A., Santos, R. G., et al. (2013). Rapid scanning structure-activity relationships in combinatorial data sets: Identification of activity switches. *Journal of Chemical Information and Modeling*, *53*(6), 1475–1485.
- Medina-Franco, J. L., Fernández-de-Gortari, E. y Naveja, J. J. (2015). Avances en el diseño de fármacos asistido por computadora. *Educación Química*, *26*(3), 180–186.
- Medina-Franco, J. L., Giulianotti, M. A., Yu, Y., Shen, L., Yao, L. y Singh, N. (2009). Discovery of a novel protein kinase b inhibitor by structure-based virtual screening. *Bioorganic and Medicinal Chemistry Letters*, *19*(16), 4634–4638.
- Medina-Franco, J. L., Lopez-Vallejo, F. y Castillo, R. (2006). Diseño de fármacos asistido por computadora. *Educación Química*, *17*(4), 452–457.
- Medina-Franco, J. L., Méndez-Lucio, O., Yoo, J. y Dueñas, A. (2015). Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discovery Today*, *20*(6), 569–577.
- Méndez-Lucio, O., Tran, J., Medina-Franco, J. L., Meurice, N. y Muller, M. (2014). Towards drug repurposing in epigenetics: Olsalazine as a novel hypomethylating compound active in a cellular context. *ChemMedChem*, *9*, 560–565.
- Naveja, J. J., Dueñas-González, A. y Medina-Franco, J. L. (2016). Drug repurposing for epigenetic targets guided by computational methods. En J. L. Medina-Franco (Ed.), *Epi-informatics: Discovery and development of small molecule epigenetic drugs and probes* (pp. 327–358). Londres, Reino Unido: Academic Press.
- Nicola, G., Liu, T. y Gilson, M. K. (2012). Public domain databases for medicinal chemistry. *Journal of Medicinal Chemistry*, *55*(16), 6987–7002.
- Scior, T., Bender, A., Tresadern, G., Medina-Franco, J. L., Martínez-Mayorga, K., Langer, T., et al. (2012). Recognizing pitfalls in virtual screening: A critical review. *Journal of Chemical Information and Modeling*, *52*(4), 867–881.
- Scior, T., Bernard, P., Medina-Franco, J. L. y Maggiora, G. M. (2007). Large compound databases for structure-activity relationships studies in drug discovery. *Mini Reviews in Medicinal Chemistry*, *7*(8), 851–860.
- Talele, T. T., Khedkar, S. A. y Rigby, A. C. (2010). Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, *10*(1), 127–141.



CrossMark  
 click for updates

Cite this: *RSC Adv.*, 2017, 7, 632

## Getting SMART in drug discovery: chemoinformatics approaches for mining structure–multiple activity relationships†

Fernanda I. Saldívar-González,<sup>a</sup> J. Jesús Naveja,<sup>abc</sup> Oscar Palomino-Hernández<sup>a</sup> and José L. Medina-Franco<sup>\*a</sup>

In light of the high relevance of polypharmacology, multi-target screening is a major trend in drug discovery. As such, the increasing amount of available structure–activity data requires the application of chemoinformatic approaches to mine structure–multiple activity relationships. To this end, activity landscape methods, initially developed to explore the structure–activity relationships for compounds screened against one target, have been adapted to mine Structure–Multiple Activity Relationships (SMART). Herein, we survey advances in the chemoinformatic approaches to retrieve SMART from screening data sets. Case studies relevant to modern drug discovery are discussed. The methods covered in this survey are general and can be implemented to explore the SMART of other data sets screened across multiple biologically endpoints.

Received 3rd November 2016  
 Accepted 6th December 2016

DOI: 10.1039/c6ra26230a

[www.rsc.org/advances](http://www.rsc.org/advances)

### Introduction

Analysis of structure–activity relationships (SAR) is a common practice in many areas of chemistry. Most medicinal and computational chemists working on drug discovery obtain SAR of compound data sets on a routine basis. This is true not only in academic settings but also in the pharmaceutical industry and research institutes. In several current drug discovery projects, compound data sets are screened across more than one biological endpoint. Depending on the project, it is desirable to identify selective compounds or identify molecules with activity across multiple endpoints. Moreover, in light of the increasing awareness of polypharmacology<sup>1</sup> and multi-target drug discovery,<sup>2</sup> screening small compound data sets or large chemical libraries across more than one biological endpoint is a fundamental task. Therefore, getting Structure–Multiple Activity Relationships (SMART) is a common need in drug discovery.

Methods to get SMART can be broadly classified into qualitative and quantitative. Qualitative approaches can be applied without the need of computational tools and depend on the

experience of the chemist analyzing the data. Thus, qualitative methods are suitable to handle small-to-medium size data sets. In contrast, large screening data sets, in particular those tested across several endpoints, usually require the application of computational procedures in addition to the experience of the chemist.<sup>3</sup> In these cases, *in silico* methods can be performed for either predictive or descriptive purposes. As discussed previously, understanding the SAR of compound data sets should be performed before developing predictive models<sup>4</sup> such as QSAR and QSPR in order to predict novel, potent, and selective compounds.<sup>5,6</sup> In this regard, new computational models that combine multi-target QSAR with machine learning such as artificial neural network algorithms have been developed with the aim of predict the interactions of multiple molecules to targets involved in many diseases and processes of neuroprotection.<sup>5,7</sup>

Activity landscape modeling (ALM) is a chemoinformatic strategy to mine the SAR of compound data sets and it is actively used in academia, industry and other research settings. ALM can be regarded as part of computer-aided drug design and it is an important component in medicinal chemistry.<sup>8</sup> For more than ten years several groups have worked on the development of ALM. These approaches relay on the quantitative comparison of structure similarity with activity similarity (or potency difference) for all pairs of compounds in a screening data set. Over the years a large number of quantitative and visual methods have been developed. Most of these methods started with the main goal of identifying ‘activity cliffs’: pairs of compounds with very similar structure but unexpected high activity difference.<sup>9</sup> Activity cliffs have a ‘dual face’ with a large impact in medicinal and computational chemistry.<sup>10</sup> It has been

<sup>a</sup>Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico. E-mail: medinajl@unam.mx; jose.medina.franco@gmail.com; Tel: +52-55-5622-3899 ext. 44458

<sup>b</sup>Facultad de Medicina, PECEM, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City, 04510, Mexico

<sup>c</sup>Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, 85764, Neuherberg, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ra26230a



largely advised that after identifying activity cliffs, molecular modeling studies should be conducted that help to explain, at the molecular level, the reason associated with the large change in activity due to a small modification the chemical structure. Such studies are highly valuable because add three-dimensional information to the system. To this end, mechanistic studies towards the structural interpretation of activity cliffs in three dimensions have been published.<sup>11,12</sup> Overall, the specific reasons that are associated with the formation of the activity cliffs depend on the system. An alternative approach to add three-dimensional information to the system and consider additional effects of functional groups, conformations and configurations, molecular descriptors that take into account the coordinates space of the compounds or even using several different conformations of the molecules in the data set have been reported.<sup>13,14</sup>

ALM seeks not only to identify activity cliffs but other significant areas of the activity landscape such as 'similarity cliffs' (which are related to scaffold hops)<sup>15</sup> and other continuous regions of the activity landscape. The broad applicability of ALM in medicinal chemistry has been reviewed.<sup>16</sup> Initially developed to describe SARs, ALM has been tested for predictive purposes.<sup>17,18</sup> Similarly, ALM was originally applied to describe the SAR of compound data sets screened for one biological endpoint, for instance, for a single target. However, several methods used in ALM have been adapted to mine SMARTs.

The goal of this work is to survey the progress of ALM to get SMART in drug discovery. We put special emphasis on the development and application of Structure–Activity Similarity (SAS) maps which were one of the first approaches used in ALM.<sup>19</sup> Four years ago the authors reviewed the development of SAS maps to explore SARs.<sup>20</sup> In contrast, this review covers the most recent developments and applications aimed to explore SMARTs. As part of the recent developments the concept of 'pro-activity cliffs' is introduced. The manuscript is organized in five main sections: after this introduction a brief overview of the SAS maps is presented with special emphasis on the development of density SAS maps and activity landscape sweeping strategies. The section after that describes the adaptation of ALM from single to multi-target activity analysis. This section is followed by a discussion of future trends in SAR and SMART analysis using ALM. Concluding remarks are presented at the end.

## Structure–activity similarity (SAS) maps

SAS maps were proposed in 2001.<sup>19</sup> The basic idea of a SAS map is to plot in two-dimensions (2D) the pairwise structure similarity (usually plotted on the X-axis) and activity difference (plotted on the Y-axis) for all pairs of compounds in a data set. A general form of a SAS map is shown in Fig. 1A. To aid in the interpretation, a SAS map can be roughly divided in four major quadrants each one distinguishing pairs of compounds with high/low activity difference and high/low structure similarity. Activity cliffs are located in the quadrant that identifies pairs of molecules with high structure similarity and high activity difference (region IV). Compound pairs with a smooth SAR have high structure similarity and low activity difference (region II).

Scaffold hops (or similarity cliffs) are located in the opposite quadrant of the activity cliffs (region I). Noteworthy, even in the absence of the thresholds with formally defined quadrants, SAS maps are helpful to differentiate major regions in the landscape.

One of the known limitations of the SAS maps is the quantitative criteria to define the thresholds along the X- and Y-axis. A number of approaches to address this issue are discussed elsewhere.<sup>20</sup> Briefly, the thresholds that define high/low activity difference depend on the goal of the project. Usual cutoffs are one, two or more potency units. The thresholds to define high/low structure similarity can be set up based on the distribution of the similarity values of the data set. In some instances, heuristic values of similarity are considered based on author's experience.

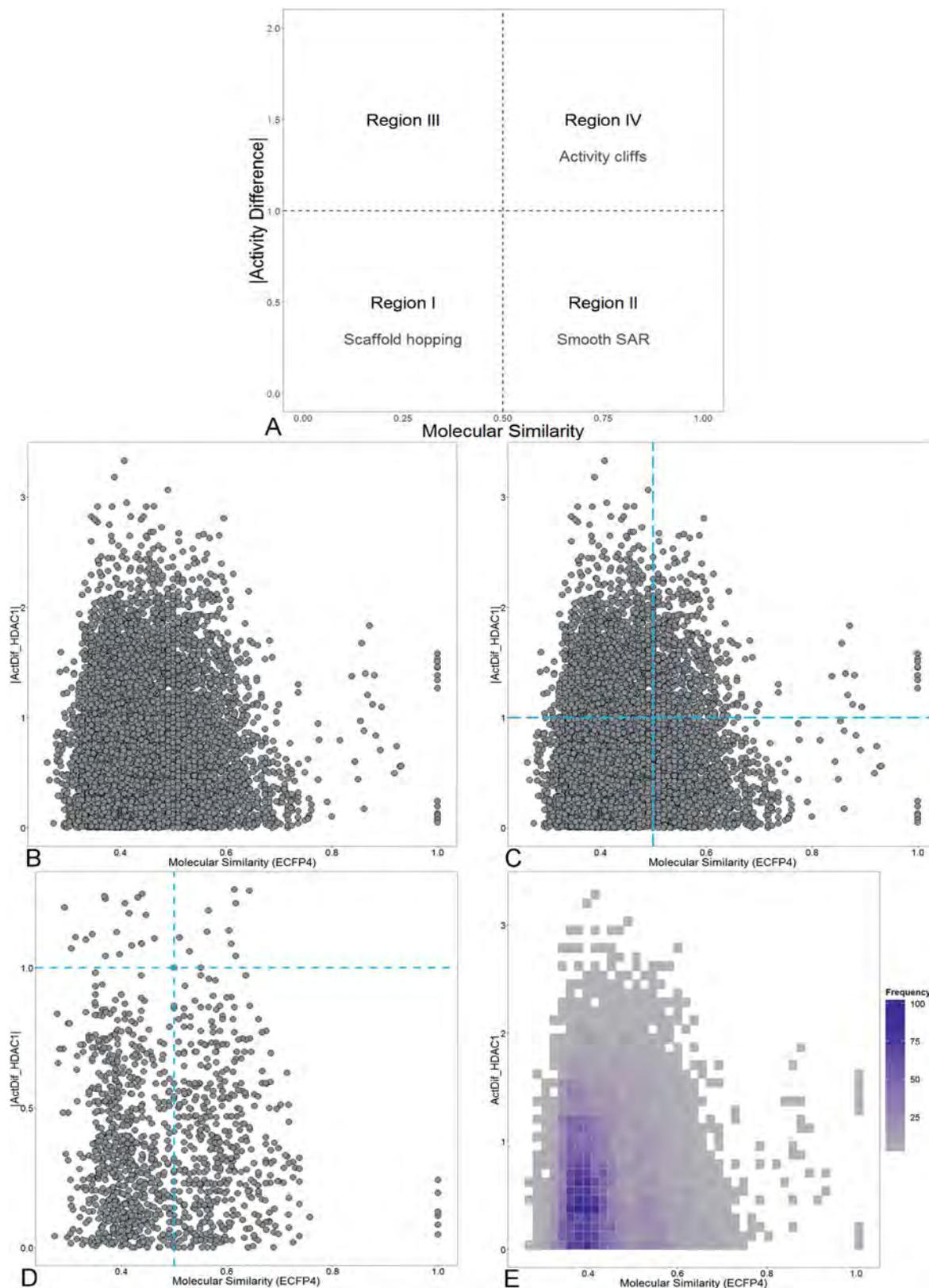
Another limitation of the SAS maps is the large amount of data points that could be generated. Therefore, for large data sets it is challenging the visual interpretation of the SAS maps. To address this issue several strategies have been proposed which are discussed below.

### Density SAS maps

To aid in the visualization of the SAS and related maps three major strategies have been developed: (1) categorical SAS maps;<sup>13</sup> (2) filtered SAS maps showing only the most relevant data points (for instance, the 'active pairs' of compounds defined as pair of molecules containing at least one active compound in the pair) and, more recently (3) density SAS maps that display the amount of data points using a continuous color scale.<sup>21</sup> Fig. 1 shows examples of 'simplified' SAS maps: categorical, filtered and density SAS map for a data set of 140 pyrimidine hydroxyl amide compounds tested with histone deacetylase 1 (HDAC1). These compounds were synthesized and tested as part of a program of optimization to find potent and selective inhibitors of HDAC6, enzyme required for the formation of the aggresome and survival of cancer cells.<sup>22</sup> HDAC is a major epigenetic target and the computational analysis of the SAR can be regarded as part of the emerging research field of Epi-Informatics.<sup>23</sup> SAR analysis of HDAC inhibitors is particularly useful for the treatment of proliferative diseases and disorders by protein deposition, likewise, it is useful for probing biological pathways. A full discussion of the SAR of HDAC inhibitors is out of the scope of this Short Review that is focused on ALM. Fig. S1 in the ESI† shows additional examples of simplified SAS maps for a data set of 91 compounds tested against the parasite *Giardia intestinalis*. Note that density SAS maps provide better information regarding the general distribution of the data points, though sacrificing the chance of including information regarding the individual activity of any of the compounds in the pair.

Several analyses have shown that the similarity cliff region is one of the most populated for several data sets.<sup>13</sup> Results of Maggiora *et al.* further confirmed these observations analyzing many data sets.<sup>15</sup> This is also the case in the activity landscape depicted in Fig. 1 and S1.† Density SAS maps have been employed to analyze the ALM of  $\alpha$ -alpha reductase inhibitors<sup>24</sup>





**Fig. 1** (A) General form of the structure–activity similarity (SAS) maps showing four major regions. Regions I and II are associated with scaffold hopping and smooth SAR, respectively. Region III does not provide relevant information and region IV indicates discontinuous SAR and activity cliffs. Actual (B) and simplified SAS maps for a data set of 140 compounds tested with HDAC1. (C) Categorical map showing the distribution of the data point in each of the four quadrants of the SAS map; (D) filtered map displaying the 'active regions' of the landscape *i.e.*, pairs of compounds that contain at least one active molecular in the pair; and (E) density map that shows the amount of data points in each region using a continuous color scale from purple color (more data points) to grey color (less data points). The simplified SAS maps are designed to aid in the visual representation and interpretation of the SAS maps.



and inhibitors of DNA methyltransferases (DNMTs), other major epigenetic target.<sup>21</sup>

### Activity landscape sweeping

Activity landscape sweeping is a strategy recently developed to 'clean' the SAR/SMART of a data set by filtering first the compounds that are considered to analyze the landscape. An approach is to classify the compounds by the types of molecular scaffold<sup>25</sup> or the relative position in chemical space, to name two criteria. Then, the ALM would be centered on the local SAR of the filtered molecules. In a broad sense, activity landscape sweeping is an approach to analyze local models of SAR/SMARTs. Despite the fact that such models are not general, activity landscape sweeping gives rise to focused analysis of the most interpretable areas of the activity landscape.

In order to illustrate the filtering of compounds before ALM, *i.e.*, activity landscape sweeping, Fig. 2 shows a visual representation of the chemical space of a series of 140 pyrimidine hydroxyl amide compounds synthesized and evaluated as HDAC6 inhibitors. Two main clusters (A: circles, B: triangles) are readily distinguished: compounds of cluster A correspond to formulas IV–VIII described by Van Duzer *et al.* while compounds of cluster B correspond to formulas I–III described in the same work.<sup>22</sup> The main difference between these two groups is the carbon attached to the nitrogen of 2-amino-*N*-hydroxypyrimidine-5-carboxamide. In group A, this carbon is tertiary, while in group B is primary or secondary. Representative chemical structures are shown in Fig. S2 of the ESI.†

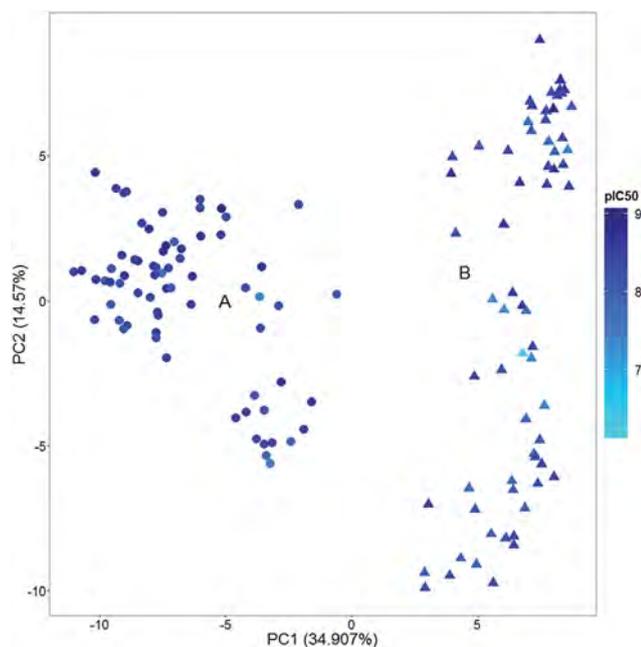


Fig. 2 Example of an activity landscape sweeping. Visual representation of the chemical space of the 140 inhibitors of histone deacetylase 6 (HDAC6). The visualization was obtained by principal component (PC) analysis of the similarity matrix computed with extended connectivity fingerprint 4 (ECFP4). The percentage of variance explained by each PC is indicated in the corresponding axis. Data points are colored by the  $pIC_{50}$  values of HDAC6 in a continuous scale.

Activity landscape sweeping has been recently applied to DNMT inhibitors<sup>21</sup> and 5 $\alpha$ -reductase inhibitors.<sup>24</sup> In both instances activity landscape sweeping was used in conjunction with SAS maps. This approach helped to 'clean' the landscape and facilitated the visual analysis of the SAS maps. Activity landscape sweeping has been used in conjunction with SAS maps but could be implemented in combination with any other ALM strategy such as Structure–Activity Landscape Index (SALI)<sup>26</sup> or other methods.

### SAS maps and PLIFS

Protein–ligand interaction fingerprints (PLIFS) are convenient representations to capture protein–ligands contacts in a systematic manner. PLIFS are at the interface of cheminformatics and molecular modeling<sup>27</sup> and have been designed to 'capture a 1D representation of the interactions between ligand and protein either in complexes of known structure or in docked poses'.<sup>28</sup> Recently SAS maps have been adapted to analyze structure–protein ligand interactions giving rise to the protein–ligand interaction cliffs.<sup>27</sup> These are defined as pairs of compounds with high structure similarity, high protein–ligand contact similarity but very different activity profile. That study was conducted for a series of kinase inhibitors. In that work, Méndez-Lucio *et al.* integrated PLIFS to a multi-target kinase activity landscape analysis. Three data sets, containing the crystallographic structure of the ligand bound to a kinase were used. The authors employed three data sets, containing the crystallographic structure of the ligand bound to a kinase. Pairwise interaction similarity was assessed using PLIFS and the Tanimoto coefficient, whereas twelve 2D and 3D molecular descriptors were used to compute pairwise molecular similarity. Pairwise structure-similarity analysis revealed no correlation with interaction similarity in none of the data sets despite the fact that the kinase ATP binding site is highly conserved. On average, only 33% of the molecular pairs categorized as highly similar showed similar interactions. This approach not only provided structural information of activity cliffs but it also was useful to identify hot spots in the target protein associated with selectivity.<sup>27,29</sup>

### Tuning ALM to get SMART

In addition to SAS maps several other methods have been developed for ALM analysis.<sup>16,20,26,30,31</sup> For instance SALI, the first index developed to rapidly identify activity cliffs, is calculated with the expression:<sup>26</sup>

$$SALI_{i,j} = \frac{|A_i - A_j|}{1 - \text{sim}(i,j)}$$

where  $A_i$  and  $A_j$  are the activities of the  $i$ th and  $j$ th molecules, and  $\text{sim}(i,j)$  is the similarity coefficient between the two molecules. Also, the research group of Bajorath has developed a large number of approaches for ALM.<sup>16</sup>

Several of ALM methods have been adapted to handle SMART. For instance, a straightforward extension of SALI to measure SMART is replacing the numerator of the SALI with the



Table 1 Examples of case studies of SMART studies conducted with SAS-like maps

Study	Major outcome (method)	Major outcome (interpretation)	Ref.
SMART of >50 benzimidazoles tested with <i>T. vaginalis</i> and <i>G. intestinalis</i>	Dual activity difference maps with fingerprint and sub-structure representation	'Activity switches' are introduced: pairs of compounds where one small change in the structure is associated with a different and opposite change in the activity of two biological endpoints	34 and 35
SMART of a series of purine analogs screened against the cysteine protease cathepsins	Triple activity difference maps	The concept of structure–property–activity (SPA) similarity in SAR studies are introduced. SPA maps are analyzed to determine the extent to which property similarities could be applied to characterize SARs	14
SMART of compounds in PubChem	Structure multiple Activity Similarity (SmAS) maps	Bioassay activity landscape is introduced to study the relationship between the structure and bioactivity profiles	37
ADMET analysis of 166 compounds screened for kappa-opioid receptor activity	ADMET property–activity pairwise similarity maps with ADMET descriptors and dimensional 'violation bit vector' representing	Study of the range of ADMET property violations that arise from structural changes, subtle and significant	41
SMART of 15 252 compounds screened across 100 diverse proteins reported by Clemons <i>et al.</i> <sup>38</sup>	SPID measure (Structure–Promiscuity Index Difference)	Structure promiscuity index is introduced to identify the pairs of compounds with high structure–similarity but large activity difference	39

biological profile similarity of the compound pair computed with the Tanimoto coefficient (giving rise to a Structure–Multiple Activity Landscape Index).<sup>32</sup> Representative case studies of the adaptation of ALM to get SMART are summarized in Table 1 and discussed in the next sections.

### SMART with few biological endpoints

One of the first applications of SAS maps applied to analyze data sets across more than one biological endpoint was the SMART exploration of more than 50 benzimidazole analogues tested for their ability to inhibit the growth of the protozoa *Trichomonas vaginalis* and *Giardia intestinalis*.<sup>33</sup> A tool to analyze simultaneously the difference in activity data for both parasites was the Dual Activity Difference (DAD) maps. DAD maps represent in 2D changes in potency difference for two targets.<sup>34</sup> One of the major outcomes of the DAD maps are 'activity switches' defined as pairs of compounds where one small change in the structure is associated with a very different but opposite change in the activity for both biological endpoints. Activity switches have been reviewed in detail.<sup>20</sup> Triple-Activity Difference (TAD) maps were developed later as a natural extension of the DAD maps to analyze SMARTs.<sup>14</sup>

More recently, DAD maps were used to analyze systematically the activity landscape of a series of 91 benzimidazoles tested with the parasites *T. vaginalis* and *G. intestinalis*.<sup>35</sup> In that work the chemical structure of the 91 benzimidazoles was encoded using a fragment-based approach that indicated the presence or absence of six substituents around a common benzimidazole nucleus. Using DAD maps, single and dual substitutions around the benzimidazole scaffold were identified that were

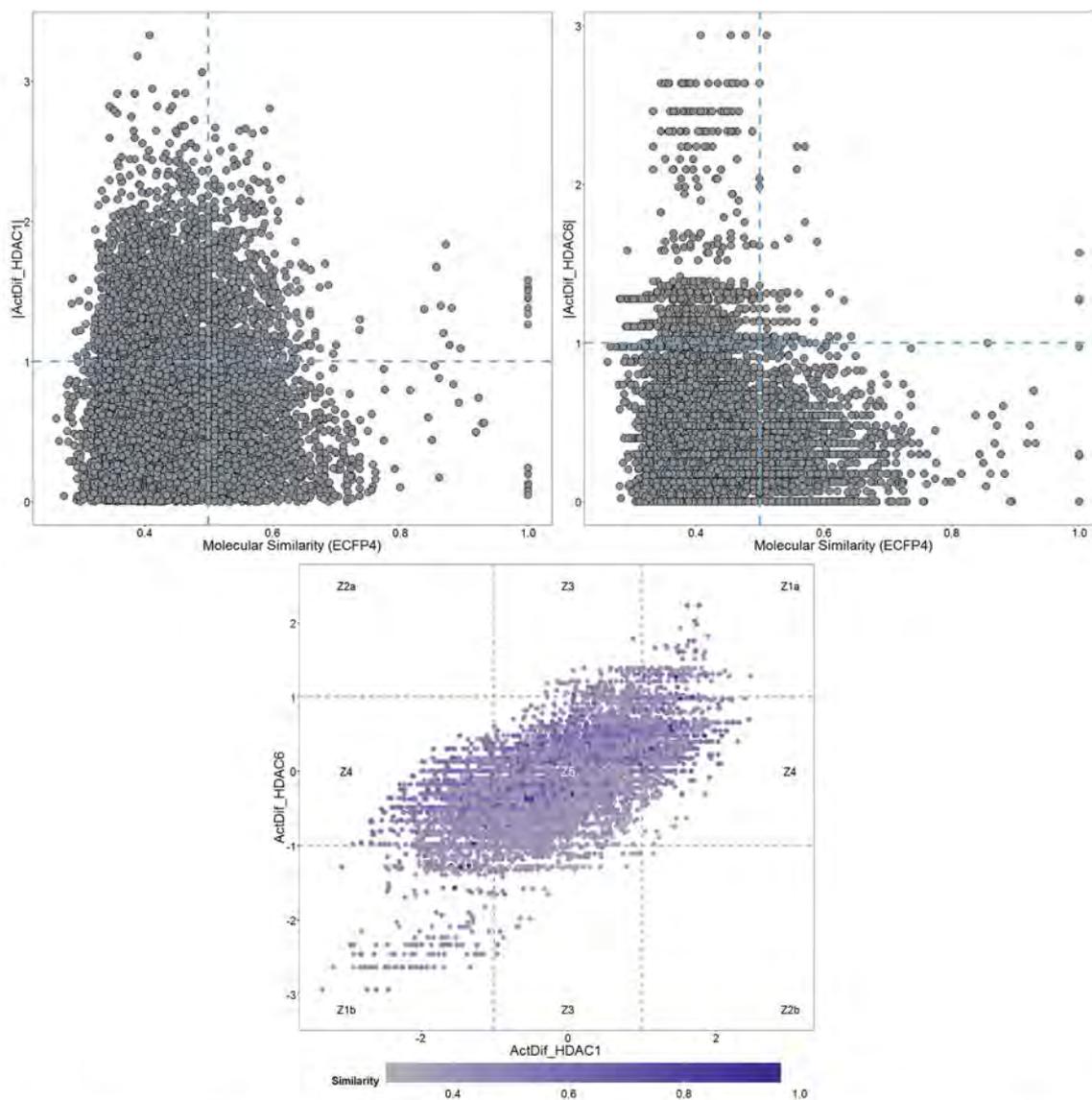
associated with large changes in potency for each of the two parasites. Furthermore, single and dual substitutions associated with large and opposite changes in activity for the two parasites were found.<sup>35</sup>

To illustrate a DAD map, Fig. 3 shows a plot of a data set of 140 molecules tested as HDAC1 and HDAC6 inhibitors.<sup>22</sup> As reference, Fig. 3 also shows the corresponding SAS maps for HDAC1 and HDAC6. In general, the DAD map in Fig. 3 shows that the larger amount of pairs of compounds are located in the region Z5 of the plot (close to 68%), indicating that most of the compounds show activity values very similar for both enzymes. The pairs identified in the Z3 and Z4 regions (simple activity cliffs) suggest that changes in the scaffold are more susceptible to present changes in activity against HDAC1 compared with HDAC6. The increased presence of pairs of compounds in the Z1 region compared to Z2 region indicates that there is a greater likelihood that the modifications affect the activity of both enzymes in the same magnitude and direction.

### SMART with many biological endpoints

ALM have also been applied to analyze the SMART of screening collections tested across a large number of biological endpoints. Different ALM methods have been used including SAS maps. For instance, SAS maps were employed to analyze the SMART obtained from Pubchem.<sup>32</sup> In a proof-of-concept study, Medina-Franco and Wadell analyzed the bioassay activity landscape of 618 molecules tested across 244 confirmatory bioassays. One of the particular challenges in that work was that each bioassay in PubChem has its own specific definition of active, inactive, or inconclusive. A second major challenge was





Region	Interpretation	9730 pairs total
Z1	Substitution (s) result in a significant decrease or increase of activity in both targets	474 (4.87%)
Z2	Substitution (s) increase activity for one target, while decreasing activity for the other target significantly	1 (0.01%)
Z3	Substitution (s) result in significant changes in activity on HDAC6 but not an appreciable change HDAC1	379 (3.89%)
Z4	Substitution (s) result in significant changes in activity on HDAC6 but not an appreciable change HDAC1	2262 (23.25%)
Z5	Substitution does not change significantly the activity for HDAC1 and HDAC6	6614 (67.98%)

Fig. 3 Example of SAS and DAD maps of a data set of 140 compounds tested across two biological endpoints (HDAC1 and HDAC6). Each data point represents a pairwise comparison. The table shows the interpretation and number and percentage of data points in each region of the map for compound pairs.



that not all 618 compounds were tested in all 244 bioassays. A distinctive feature of the SAS-like maps proposed to address those two challenges was the calculation of a pairwise bioassay activity profile similarity (bAPS): for each of the 618 compounds tested in any of the 244 confirmatory assays the bioassay activity profile was represented as a multiset fingerprint encoding of the activity data as follows: 'active' was set to '2'; 'inactive' as '1'; inconclusive or not tested as '0'; the pairwise bAPS was calculated using the Tanimoto coefficient:<sup>36</sup>

$$\text{bAPS}(i,j) = \frac{\sum_{k=1}^n \min[m_k(i), m_k(j)]}{\sum_{k=1}^n \max[m_k(i), m_k(j)]}$$

where  $\text{bAPS}(i,j)$  is the bioassay activity profile similarity of the  $i$ th and  $j$ th molecules,  $m_k(i)$  and  $m_k(j)$  are the activity encodings of the  $i$ th and  $j$ th molecules, respectively, and  $n$  is the total number of assays that the molecules were screened across. This encoding of the activity data enabled the systematic structure- and bioprofile activity similarity and identified bioassay activity profile cliffs *i.e.*, pairs of compounds with high structure similarity but very different bioassay activity profiles.<sup>37</sup>

In a separate work Yongye *et al.* analyzed the ALM of a chemogenomics data set released by Clemons *et al.* The data set contained more than 15 000 compounds from different sources (commercial compounds, natural products and synthetic molecules) that were screened across 100 sequence-unrelated proteins.<sup>38</sup> SMART analysis using SAS maps led to the identification of structural changes that differentiated highly specific from promiscuous compounds. It was also concluded that, in general, similar synthetic structures from academic groups showed greater promiscuity differences than do commercial compounds and natural products.<sup>39</sup> A characteristic metric employed in that work was the Structure-Promiscuity Index Difference (SPID); for each pair of compounds, the relationship between structure similarity and the different number of proteins to which each compound in the pair binds was computed using the expression:

$$\text{SPID}(X_a, X_b) = \frac{|P_{X_a} - P_{X_b}|}{1 - T_n(X_a, X_b)}$$

where  $P_{X_a}$  and  $P_{X_b}$  are the number of proteins to which compounds  $X_a$  and  $X_b$  are bound and  $T_n(X_a, X_b)$  is the pairwise Tanimoto structure similarities of both compounds. The SPID metric is reminiscent of SALI (see above). Noteworthy, SPID focuses on the change in the number of proteins bound associated with a change in the molecular structure but does not account for the specific proteins involved, such as the metric 'binding profile similarity'.<sup>40</sup> In order to address the identity of the proteins Yongye *et al.* also computed the pairwise binding profile similarities employing the binary profile of each compound as a 100-dimensional vector *e.g.*, a pairwise Tanimoto similarity. As such it was also analyzed the multiple-assay profile SAR of the data set using the modified version of SALI: Structure-Multiple Activity Landscape Index (*vide supra*).

Similar to activity landscape analysis with one biological endpoint, the structural interpretation of SMART with many

biological endpoints would require further molecular modeling studies with the three dimensional structures of the targets, if available. An alternative is to incorporate three dimensional molecular descriptors to describe the chemical structures. It is particularly interesting to provide a further rationale of the source of selectivity or promiscuity of the compounds.

## Future directions

In principle, methods employed in ALM can be implemented to explore the SAR or SMART of any screening data evaluated across multiple biological endpoints. Moreover, several methods can be extended to mine biological fingerprints. SMART studies can be further extended to analyze properties such as toxicity. In this regard, Austin *et al.* introduced ADMET property-activity pairwise similarity maps to analyze the relationships between activity, structure and ADMET violations/compliance with particular emphasis on determining structural changes that have a large impact on the ADMET compliance.<sup>41</sup>

In drug discovery, big data is typically obtained from high-throughput screening (HTS). HTS usually is conducted in two general steps: assays at a single dose followed by confirmatory assays at multiple doses. Despite the fact that biological assays at single-dose concentrations have not been considered for activity landscape analysis,<sup>42</sup> such assays do provide valuable information that could be considered in preliminary activity landscape studies. We propose that this is a relevant future direction not only in ALM but in SMART studies in general.

### Pro-activity cliffs

Relevant regions in the activity landscape are analyzed using high quality biological activity data that are obtained after multiple-dose inhibition assays. Due to its rigorous determination, it has been proposed that only those values can be used within the realm of ALM methods in order to minimize errors.<sup>42,43</sup> While the latter remains as the ideal case, there are several cases where only single-dose biological activity data for a given target is available. Herein is proposed that this data can be used for a preliminary activity landscape analysis in order to identify potential areas of interest and guide the next steps towards the acquisition of high quality information. Thus, identification of potential activity cliffs *i.e.*, *pro-activity cliffs* is valuable, as they can be prioritized for additional experimental evaluation.

To temporarily address both the lack of multiple-dose/high quality data and the error involved in the biological activity measurement, the percentage of inhibition frequently obtained at single dose evaluations can be distributed in different categories; for instance, potentially very active, active, inactive and potentially very inactive. Integer indices can be assigned to the different classes: *e.g.*, an integer index of 1 for the least active compounds and 4 for the most active ones. The limits of the inhibitory activity can be fitted to the distribution of the data set; *e.g.*, those compounds with less than 25% of inhibition can be regarded as potentially inactive (*e.g.*, activity index of 1),



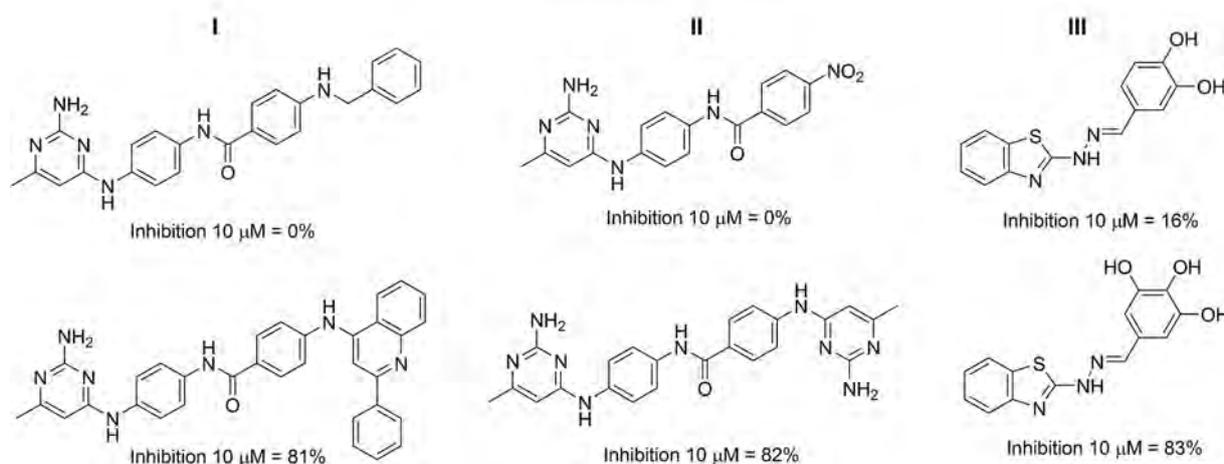
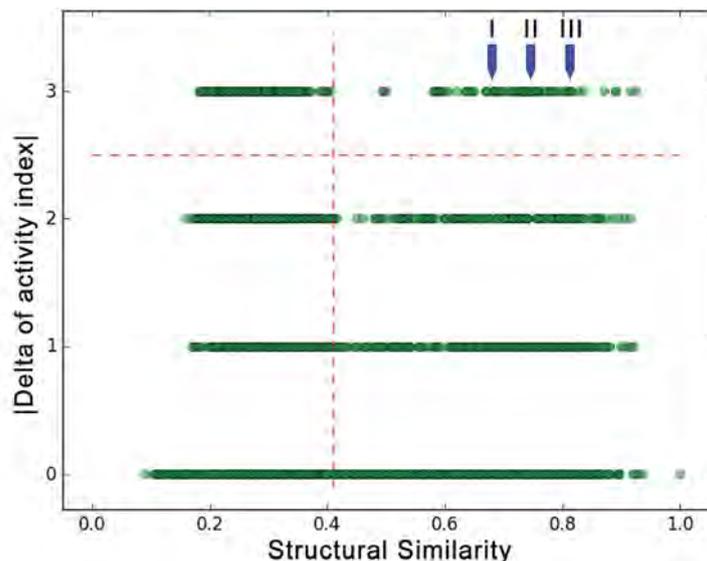


Fig. 4 Examples of pro-activity cliffs for a set of 106 compounds tested as inhibitors of DNMT3A.

while those with more than 75% inhibition can be considered as potentially active compounds (*e.g.*, activity index of 4). After classifying activity data and generating a categorical structure–activity similarity map, four horizontal zones can be defined as the result of comparing the activity index: 0 (as the result of comparing members of the same group), 1 (by comparing members of one unit of difference in the activity index), and so on. Thus, pro-activity cliffs can be defined as pairs of compounds with high structure similarity where one is highly probable to be active and the other is highly probable to be inactive. To illustrate this point, an actual set of single-dose activity data is exemplified for a group of inhibitors of DNA methyltransferase 3A (DNMT3A); for a large number of compounds, only percentages of inhibition obtained at single dose are available (10  $\mu\text{M}$ ). Fig. 4 shows a categorical SAS map for 106 compounds tested as potential modulators of DNMT3A. The SAS map in this figure has 5565 data points; the *x*-axis represents the pair-wise structure similarity computed as the mean of the Tanimoto similarity values computed with

Extended Connectivity Fingerprints (radius 2) and MACCS keys (166 bits). The *y*-axis represents the four regions defined by the difference of the activity indices. The vertical dashed line is marked in the 3<sup>rd</sup> quartile of the pair-wise mean similarity values of the data set (mean similarity of 0.41). In Fig. 4 upper right quadrant identifies the pro-activity cliffs. The same figure illustrates three specific examples of pro-activity cliffs. As shown in Fig. 4, the three pairs of compounds show a remarkable resemblance, and a high difference in their inhibition activities. For instance, the only structural difference in pro-activity cliff “III” is a hydroxyl group. Further multiple-dose testing would confirm or not the status of the potential activity cliffs.

## Concluding remarks

ALM is a quantitative approach to analyze systematically SAR of compound data sets. In many drug discovery programs compound data sets are screened against two, three or many more biological endpoints. To rapidly mine the usually large



data generated, ALM have been adapted to analyze the associated SMART. Among ALM approaches, SAS maps have evolved rapidly to address the increasing need of analyzing SMART. To date, several successful applications have been reported including the analysis of SAR, SMART and protein–ligand interaction cliffs. As part of the development of the SAS maps, a number of metrics and visualization approaches have been developed. Since SMART analysis usually involves analysis of large amount of data, getting smart in drug discovery may require using information available in large screening campaigns that include incomplete chemogenomics data sets or activity data obtained at single concentrations. Bioactivity-profile similarity, activity landscape sweeping and pro-activity cliffs are examples of recently proposed concepts to advance the SMART analysis in drug discovery. One of the major perspectives in the field is to incorporate the principles of quantum mechanics to refine the SMART models and further improve their applicability in drug discovery projects.

## Conflict of interest

The authors declare that they do not have any conflict of interest related to this manuscript.

## Acknowledgements

This work was supported by the Universidad Nacional Autónoma de México (UNAM) grant ‘Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza’ (PAPIME) PE200116 and grant ‘Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica’ (PAPIIT) IA204016. We also thank funding from the ‘Programa de Apoyo a la Investigación y el Posgrado’ (PAIP) 50009163, Facultad de Química, UNAM. OP-H is grateful to CONACyT for the fellowships granted No. 765897/606003. JJ-N is grateful with PECHEM MD PhD program for the organization of the fellowship. Rich discussions with Oscar Méndez-Lucio, Eli Fernández-de-Gortari and Mario Omar García-Sánchez are highly acknowledged.

## References

- O. Méndez-Lucio, J. J. Naveja, H. Vite-Caritino, F. D. Prieto-Martínez and J. L. Medina-Franco, *J. Mex. Chem. Soc.*, 2016, **60**, 168–181.
- J. L. Medina-Franco, M. A. Giulianotti, G. S. Welmaker and R. A. Houghten, *Drug Discovery Today*, 2013, **18**, 495–501.
- X. Hu, Y. Hu, M. Vogt, D. Stumpfe and J. Bajorath, *J. Chem. Inf. Model.*, 2012, **52**, 1138–1145.
- J. L. Medina-Franco, G. Navarrete-Vázquez and O. Méndez-Lucio, *Future Med. Chem.*, 2015, **7**, 1197–1211.
- H. González-Díaz, D. M. Herrera-Ibatá, A. Duardo-Sánchez, C. R. Munteanu, R. A. Orbegozo-Medina and A. Pazos, *J. Chem. Inf. Model.*, 2014, **54**, 744–755.
- G. M. Casañola-Martin, H. Le-Thi-Thu, F. Pérez-Giménez, Y. Marrero-Ponce, M. Merino-Sanjuán, C. Abad and H. González-Díaz, *Mol. Diversity*, 2015, **19**, 347–356.
- F. Durán, N. Alonso, O. Caamaño, X. García-Mera, M. Yañez, F. Prado-Prado and H. González-Díaz, *Int. J. Mol. Sci.*, 2014, **15**, 17035.
- F. Saldívar-González, F. D. Prieto-Martínez and J. L. Medina-Franco, *Educ. Quím.*, 2017, DOI: 10.1016/j.eq.2016.06.002.
- G. M. Maggiora, *J. Chem. Inf. Model.*, 2006, **46**, 1535.
- M. Cruz-Monteaudo, J. L. Medina-Franco, Y. Pérez-Castillo, O. Nicolotti, M. N. D. S. Cordeiro and F. Borges, *Drug Discovery Today*, 2014, **19**, 1069–1080.
- O. Méndez-Lucio, J. Pérez-Villanueva, R. Castillo and J. L. Medina-Franco, *Mol. Inf.*, 2012, **31**, 837–846.
- J. J. Naveja and J. L. Medina-Franco, *RSC Adv.*, 2015, **5**, 63882–63895.
- J. L. Medina-Franco, K. Martínez-Mayorga, A. Bender, R. M. Marin, M. A. Giulianotti, C. Pinilla and R. A. Houghten, *J. Chem. Inf. Model.*, 2009, **49**, 477–491.
- A. Yongye, K. Byler, R. Santos, K. Martínez-Mayorga, G. M. Maggiora and J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2011, **51**, 1259–1270.
- P. Iyer, D. Stumpfe, M. Vogt, J. Bajorath and G. M. Maggiora, *Mol. Inf.*, 2013, **32**, 421–430.
- D. Dimova and J. Bajorath, *Mol. Inf.*, 2016, **35**, 181–191.
- R. Guha, *J. Chem. Inf. Model.*, 2012, **52**, 2181–2191.
- J. Husby, G. Bottegoni, I. Kufareva, R. Abagyan and A. Cavalli, *J. Chem. Inf. Model.*, 2015, **55**, 1062–1076.
- V. Shanmugasundaram and G. M. Maggiora, presented in part at the 222nd ACS National Meeting, Chicago, IL, United States, August 26–30, 2001.
- J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2012, **52**, 2485–2493.
- J. J. Naveja and J. L. Medina-Franco, *Expert Opin. Drug Discovery*, 2015, **10**, 1059–1070.
- J. H. Van duzer, R. Mazitschek, Y. Ding, N. Yu, Y. Cao and Y. Liu, *Pyrimidine Hydroxy Amide Compounds as Protein Deacetylase Inhibitors and Methods of Use Thereof*, Acetylon Pharmaceuticals, Inc., 2014, EP2640709.
- A. Dueñas-González, J. Jesús Naveja and J. L. Medina-Franco, in *Epi-Informatics*, Academic Press, Boston, 2016, pp. 1–20.
- J. J. Naveja, F. Cortés-Benítez, E. Bratoeff and J. L. Medina-Franco, *Mol. Diversity*, 2016, **20**, 771–780.
- J. Pérez-Villanueva, O. Méndez-Lucio, O. Soria-Arteche and J. Medina-Franco, *Mol. Diversity*, 2015, **19**, 1021–1035.
- R. Guha and J. H. VanDrie, *J. Chem. Inf. Model.*, 2008, **48**, 646–658.
- O. Méndez-Lucio, A. J. Kooistra, C. d. Graaf, A. Bender and J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2015, **55**, 251–262.
- S. C. Brewerton, *Curr. Opin. Drug Discovery Dev.*, 2008, **11**, 356–364.
- J. L. Medina-Franco, O. Méndez-Lucio and K. Martínez-Mayorga, *Adv. Protein Chem. Struct. Biol.*, 2014, **96**, 1–37.
- R. Guha, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 829–841.
- V. F. Kuyoc-Carrillo and J. L. Medina-Franco, *Drug Dev. Res.*, 2014, **75**, 313–323.
- J. Waddell and J. L. Medina-Franco, *Bioorg. Med. Chem.*, 2012, **20**, 5443–5452.



- 33 J. Perez-Villanueva, R. Santos, A. Hernandez-Campos, M. A. Giulianotti, R. Castillo and J. L. Medina-Franco, *Bioorg. Med. Chem.*, 2010, **18**, 7380–7391.
- 34 J. Pérez-Villanueva, R. Santos, A. Hernández-Campos, M. A. Giulianotti, R. Castillo and J. L. Medina-Franco, *MedChemComm*, 2011, **2**, 44–49.
- 35 R. Aguayo-Ortiz, J. Perez-Villanueva, A. Hernandez-Campos, R. Castillo, N. Meurice and J. L. Medina-Franco, *Future Med. Chem.*, 2014, **6**, 281–294.
- 36 G. M. Maggiora and V. Shanmugasundaram, in *Cheminformatics and Computational Chemical Biology, Methods in Molecular Biology*, ed. J. Bajorath, Springer, New York, 2011, vol. 672, pp. 39–100.
- 37 J. L. Medina-Franco and J. Waddell, *J. Mex. Chem. Soc.*, 2012, **56**, 163–168.
- 38 P. A. Clemons, N. E. Bodycombe, H. A. Carrinski, J. A. Wilson, A. F. Shamji, B. K. Wagner, A. N. Koehler and S. L. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 18787–18792.
- 39 A. B. Yongye and J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2012, **52**, 2454–2461.
- 40 A. Steffen, T. Kogej, C. Tyrchan and O. Engkvist, *J. Chem. Inf. Model.*, 2009, **49**, 338–347.
- 41 A. B. Yongye and J. L. Medina-Franco, *Drug Discovery Today*, 2013, **18**, 732–739.
- 42 D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2012, **55**, 2932–2942.
- 43 J. L. Medina-Franco, G. M. Maggiora, M. A. Giulianotti, C. Pinilla and R. A. Houghten, *Chem. Biol. Drug Des.*, 2007, **70**, 393–412.

