



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**FACULTAD DE QUÍMICA**

**TESIS**

**DISEÑO MOLECULAR DE NOVO DE INHIBIDORES DE DNA  
METILTRANSFERASAS ASISTIDO POR COMPUTADORA**

**QUE PARA OBTENER EL TÍTULO DE**

**QUÍMICO FARMACÉUTICO BIÓLOGO**

**PRESENTA**

**DAVID ERICK CHÁVEZ PONCE DE LEÓN**



**Ciudad Universitaria, Cd. MX., 2019**



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: Profesor: PAULINA DEL VALLE PEREZ**

**VOCAL: Profesor: JOSÉ LUIS MEDINA FRANCO**

**SECRETARIO: Profesor: MARIA EUGENIA MENDOZA JASSO**

**1er. SUPLENTE: Profesor: AUDIFÁS SALVADOR MATUS MEZA**

**2° SUPLENTE: Profesor: CHRISTY HERNANDEZ SALAZAR**

**SITIO DONDE SE DESARROLLÓ EL TEMA: FACULTAD DE QUIMICA, EDIFICIO F, OFICINA 177, GRUPO DE INVESTIGACIÓN DIFACQUIM (DISEÑO DE FÁRMACOS ASISTIDO POR COMPUTADORA).**

**ASESOR DEL TEMA:**

**DR. JOSÉ LUIS MEDINA FRANCO**

**SUSTENTANTE:**

**DAVID ERICK CHÁVEZ PONCE DE LEÓN**

# Índice

1.	Siglas.....	8
2.	Prefacio.....	9
3.	Introducción.....	10
	3.1 Diseño de fármacos asistido por computadora (DIFAC) como una herramienta importante para el desarrollo actual de fármacos.....	10
	3.2 DNA metiltransferasas (DNMT) como reguladores de la expresión epigenética: asociación con neoplasias hematopoyéticas.....	14
	3.3 Estado del arte de los Inhibidores de DNMT (I-DNMT) disponibles actualmente .....	15
	3.4 Enfoques actuales para el diseño molecular <i>de novo</i> (DMDN) .....	19
	3.4.1 Aspectos generales del DMDN.....	19
	3.4.2 DMDN basado en la estructura de la diana biológica (DMDN-D)..	21
	3.4.2.1 Moléculas pequeñas.....	21
	3.4.3 DMDN basado en el ligando (DMDN-L) .....	24
	3.4.3.1 Moléculas pequeñas.....	24
	3.4.3.2 Péptidos.....	24
	3.4.3.3 Productos naturales.....	25
4.	Objetivos.....	26
5.	Material y recursos.....	27
6.	Metodología.....	28
	6.1 Diseño molecular <i>de novo</i> basado en la diana biológica (DMDN-D) de I-DNMT.....	28
	6.2 Diseño molecular <i>de novo</i> basado en el ligando (DMDN-L) de I-DNMT.....	28

6.2.1	Uso de bases de datos de acceso libre para obtener plantillas moleculares.....	28
6.2.1.1	DMDN-L por cambio de núcleo base ( <i>scaffold replacement</i> ).....	28
6.2.1.2	DMDN-L por fragmentos moleculares generados a través de reglas basadas en análisis retrosintéticos estadísticos ( <i>RECAP analysis</i> y <i>RECAP synthesis</i> ).....	29
6.2.2	Uso de la información estructural de la coenzima S-adenosilhomocisteína (SAH) co-cristalizada como plantilla molecular.....	29
7.	Resultados.....	30
7.1	DMDN-D de I-DNMT.....	30
7.1.1	Estrategia de exploración ( <i>Exploring strategy</i> ).....	30
7.1.1.1	Detección de la cavidad y modelo del farmacóforo de I-DNMT3A ( <i>Cavity function</i> ).....	30
7.1.1.2	Nuevas entidades químicas ( <i>NME</i> ) de I-DNMT3A generadas por algoritmos genéticos y reglas químicas ( <i>Build function</i> ).....	31
7.1.1.3	Análisis de accesibilidad y rutas sintéticas de las <i>NME</i> De I-DNMT3A.....	31
7.2	DMDN-L de I-DNMT.....	36
7.2.1	Uso de bases de datos de acceso libre para obtener plantillas moleculares.....	36
7.2.1.1	DMDN-L por cambio de núcleo base ( <i>scaffold replacement</i> ).....	36
7.2.1.2	DMDN-L por fragmentos moleculares generados a través de reglas basadas en análisis retrosintéticos estadísticos ( <i>RECAP analysis</i> y <i>RECAP synthesis</i> ).....	44
7.2.2	Uso de la información estructural de la coenzima S-adenosilhomocisteína (SAH) co-cristalizada como plantilla molecular.....	45
7.2.2.1	Análogos de S-adenosilhomocisteína (SAH).....	45
7.2.2.2	Propiedades fisicoquímicas, toxicológicas y químico	

farmacéuticas de los análogos de SAH.....	47
<b>7.2.3 Material educativo y de difusión para la aplicación del DMDN y dinámica molecular.....</b>	<b>49</b>
<b>7.2.3.1 Diseño <i>de novo</i> basado en fragmentos y dinámica molecular (segunda edición manual de Quimioinformática)....</b>	<b>49</b>
<b>7.2.3.2 DMDN para el desarrollo de I-DNMT enfocado a productos naturales.....</b>	<b>50</b>
<b>8. Conclusiones.....</b>	<b>51</b>
<b>9. Perspectivas.....</b>	<b>51</b>
<b>10. Referencias.....</b>	<b>52</b>
<b>Anexos</b>	
<b>A.1 Glosario.....</b>	<b>58</b>
<b>A.2 Segunda edición del manual de Quimioinformática.....</b>	<b>64</b>
<b>A.3 Artículo.....</b>	<b>81</b>

## **Agradecimientos**

Se agradece el otorgamiento del apoyo económico correspondiente al Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIME), Proyecto PE200118.

También se agradece el apoyo del proyecto Ciencia Básica de CONACyT, número 282785.

Parte de los resultados de esta tesis fueron publicados en:

Saldívar-González FI, Gómez-García A, Chávez-Ponce de León DE, Sánchez-Cruz N, Ruiz-Rios J, Pilon-Jiménez BA, Medina-Franco JL. Inhibitors of DNA methyltransferases from natural sources: A computational perspective. *Frontiers in Pharmacology* **2018**, 9, 1144.

Y los resultados completos de esta tesis fueron presentados en:

Chávez-Ponce de León DE, Medina-Franco JL. *De novo* design of DNA methyltransferases inhibitors 257<sup>th</sup> American Chemical Society National Meeting, Orlando, Florida, USA, March 31 - April 4, **2019**.



# 1. Siglas

- + **ADME/T:** Propiedades farmacocinéticas concernientes a la absorción, distribución, metabolismo y excreción. Además de la toxicidad ligada al fármaco.
- + **BD:** Base de datos.
- + **DiFAC:** Diseño de fármacos asistido por computadora.
- + **DMDN:** Diseño molecular *de novo*.
- + **DMDN-D:** Diseño molecular *de novo* basado en la estructura de la diana biológica.
- + **DMDN-L:** Diseño molecular *de novo* basado en el ligando.
- + **DNMT:** DNA metiltransferasa.
- + **FBDD:** Siglas en inglés de descubrimiento de fármacos basado en fragmentos.
- + **FDA:** Siglas en inglés de agencia de alimentos y medicamentos (estadounidense).
- + **HTS:** siglas en inglés de ensayos de alto rendimiento.
- + **I-DNMT:** Inhibidores de DNA metiltransferasa.
- + **I-DNMT1-NN:** Inhibidores de DNA metiltransferasa de la isoenzima 1 no nucleósidos.
- + **I-DNMT3A-NN:** Inhibidores de DNA metiltransferasa de la isoenzima 3A no nucleósidos.
- + **I-DNMT-NN:** Inhibidores de DNA metiltransferasa no nucleósidos
- + **NME:** Siglas en inglés de Nuevas entidades moleculares.
- + **PAINS:** Siglas en inglés que hace referencia a compuestos falsos positivos comúnmente identificados en ensayos de HTS.
- + **PN:** Productos naturales.
- + **RPD:** Restricciones primarias de la diana biológica.
- + **RSD:** Restricciones secundarias de la diana biológica.
- + **SAH:** S-adenosilhomocisteína (coenzima SAM, pero sin grupo metilo en el tioéter).
- + **SAM:** S-adenosil metionina (coenzima de las DNMT).
- + **SAR:** Relaciones estructura actividad.
- + **TPSA:** área superficial polar topológica.

## 2. Prefacio

En el presente trabajo se buscó desarrollar, promover y participar en tres aspectos fundamentales y centrales en el quehacer universitario: investigación; educación y divulgación científica. Específicamente:

1. Se realizó investigación en Química Farmacéutica, bajo un enfoque *in silico*, para la generación de nuevas moléculas epi-farmacológicas potencialmente bioactivas, dirigidas al tratamiento contra el cáncer (con énfasis en neoplasias hematopoyéticas). Se pretende que, como perspectivas de este trabajo, las moléculas puedan ser sintetizadas, evaluadas y optimizadas.
2. Con las metodologías computacionales empleadas en la investigación básica (punto 1), se contribuyó con la elaboración de material educativo y didáctico, que puede ser utilizado por alumnos de licenciatura, posgrado y usuarios externos. Se pretende que el material desarrollado sirva como una guía introductoria para realizar un diseño racional de fármacos, mediante la implementación de los métodos de diseño molecular *de novo* (DMDN) y dinámica molecular.
3. Uniendo las metodologías computacionales empleadas en el proyecto de investigación con su aplicación en el desarrollo de compuestos bioactivos, se desarrolló una sección de un artículo de revisión ([\*Inhibitors of DNA methyltransferases from natural sources: A computational perspective\*](#). *Front. in Pharmacol.*, **2018**, 9, 1144). En esta sección se discute cómo se puede aplicar el DMDN dentro de un contexto general para el desarrollo de inhibidores de DNA metiltransferasas (I-DNMT) enfocado a productos naturales.

## 3. Introducción

### 3.1 El diseño de fármacos asistido por computadora (DIFAC) como una herramienta importante para el desarrollo actual de fármacos

El descubrimiento de fármacos comprende una tarea multidisciplinaria que además de ser compleja, es costosa y tardada. Se estima que, para desarrollar un nuevo fármaco, el gasto promedio de la inversión es alrededor de 1000 millones de dólares (Mohs y Greig, 2017). El tiempo promedio de desarrollo dura 11 años (DiMasi et al., 2016). Además, durante el periodo comprendido de 2005 - 2010, disminuyó la productividad de nuevas entidades químicas (*NME*, por sus siglas en inglés) aprobadas por la agencia de regulación de medicamentos y alimentos estadounidense (*FDA*, por sus siglas en inglés) en nueve empresas farmacéuticas representativas (AstraZeneca, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Merck, Novartis, Pfizer, Roche y Sanofi-Aventis). Lo anterior, a pesar de haberse incrementado la inversión en investigación y desarrollo (*R&D*) (Bunnage, 2011). Los principales factores que contribuyeron al fracaso de los nuevos fármacos durante la etapa preclínica y las diferentes fases clínicas fueron la eficacia y la farmacocinética (Khanna, 2012). Como media, de 1993 a 2017 se aprobaron 26 *NME* por año, y se estima que la productividad no cambie mucho (Mullard, 2018).

Existen múltiples estrategias que se pueden seguir para guiar un diseño racional de fármacos. En general, lo más conveniente y productivo es combinar métodos de una forma racional que ayude a agilizar los procesos de descubrimiento, optimización y desarrollo. Tradicionalmente, la síntesis orgánica y los productos naturales (plantas, microorganismos y animales) fueron y continúan siendo la primordial fuente de obtención de moléculas bioactivas (Sneader, 2005; Bernardini et al., 2018). Con el paso de los años y hasta la fecha, dichas moléculas resultan ser cada vez más complicadas de obtener, ya que, para el caso de la síntesis orgánica, cada molécula convencionalmente se diseñaba, sintetizaba y se evaluaba una a la vez. Esto hacía que el proceso se prolongará por muchos años. La situación es semejante para los productos naturales: la extracción, caracterización y síntesis a gran escala retardan el proceso (Li y Vederas, 2009).

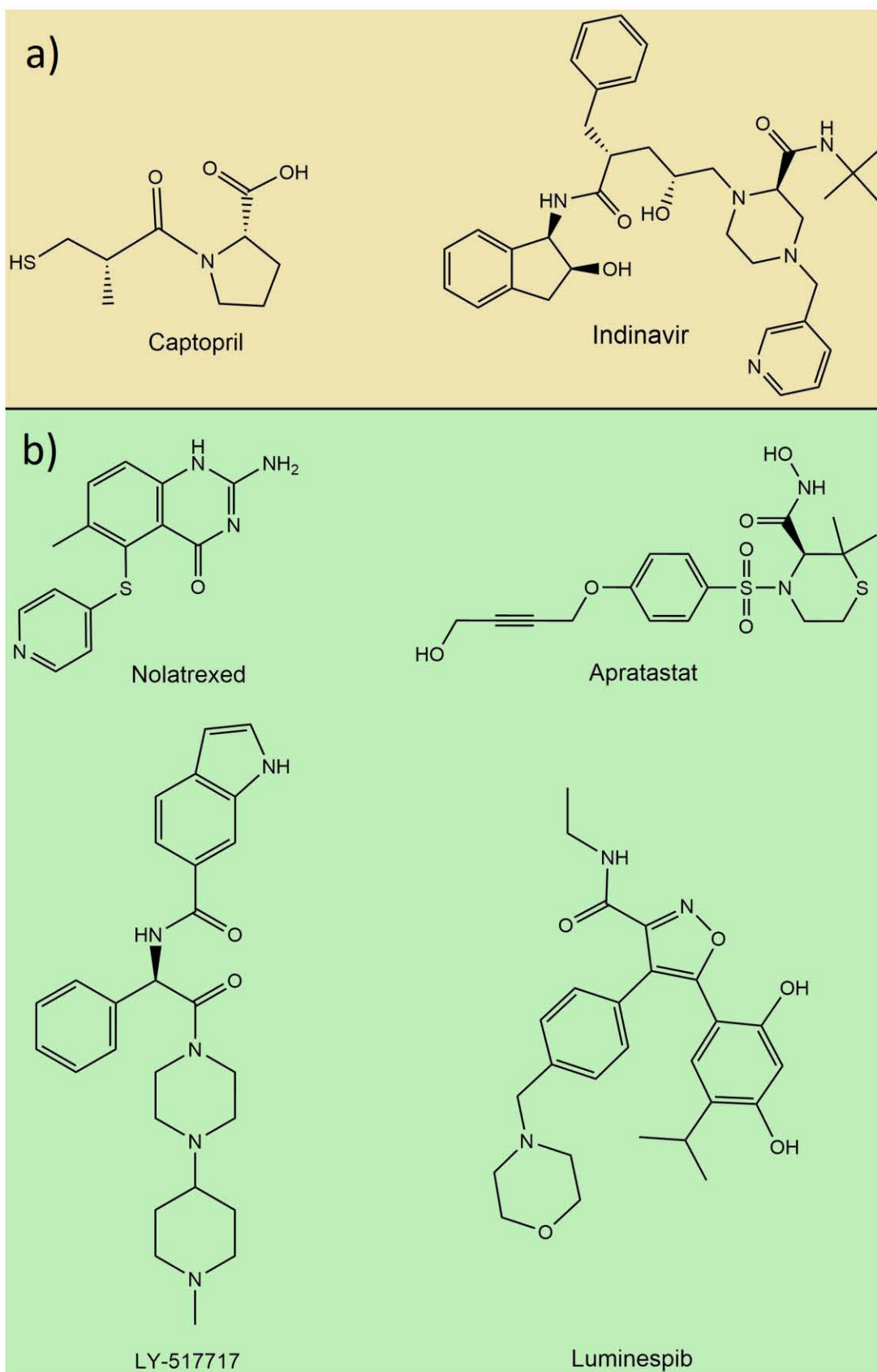
En respuesta a estos problemas, tanto en la iniciativa privada como en la academia, se buscan constantemente nuevas soluciones que conlleven a agilizar, tanto la síntesis como la evaluación biológica de las posibles *NME*. Para ello, durante el siglo pasado, se implementó la química combinatoria para obtener una colección vasta de compuestos estructuralmente diversos en un tiempo más corto que la síntesis convencional (Liu et al., 2017). A su vez, los ensayos de alto rendimiento (*high-throughput screening*, *HTS*) se utilizaron para identificar compuestos prototipo (*hits*) a través de la evaluación biológica automática de millones de compuestos en cuestión de sólo semanas (Wermuth, 2015).

El diseño de fármacos asistido por computadora (DiFAC) es un área multidisciplinaria que comprende el uso de la quimioinformática, bioinformática, modelado molecular, química computacional, y química teórica, entre otras disciplinas (Saldívar-González et al., 2017). Una de las aplicaciones inmediatas del DiFAC recae en la Química Farmacéutica para la identificación, selección y optimización de nuevos compuestos bioactivos.

El DiFAC emplea principalmente dos enfoques de diseño: 1) Basado en el receptor aplicando métodos como el modelado por homología, acoplamiento y dinámica molecular y 2) Basado en el ligando utilizando métodos como cribado virtual, generación de modelos *QSAR*, búsquedas de similitud, y modelados del farmacóforo (Sliwosk et al., 2014). La elección del enfoque dependerá del conocimiento previo que se tenga de la enfermedad que se quiera combatir, además de las necesidades de cada caso particular. El DiFAC junto con el uso conjunto de *HTS*, química combinatoria y la implementación de las nuevas tendencias como inteligencia artificial (*artificial intelligence*, *AI*) y *machine learning* aplicadas a la síntesis robotizada, para el descubrimiento, planeación y optimización de reacciones químicas orgánicas, así como para la predicción de propiedades *ADME/T* (Granda et al., 2018; Luechtefeld et al., 2018; Sellwood et al., 2018) pueden ser utilizados como una herramienta que ayude a disminuir los costos y riesgos asociados al actual descubrimiento de fármacos (Sinha y Vohora, 2018). A la fecha, se han aprobado por la FDA al menos 14 fármacos descubiertos mediante el uso de métodos de DiFAC. Entre estos destaca el captopril (antihipertensivo) y el indinavir (antirretroviral contra el VIH). Otros candidatos a

fármacos como nolatrexed (tratamiento para el cáncer de hígado), apratastat (artritis reumatoide), LY-517717 (inhibidor del factor Xa activado) y luminespib (inhibidor de HSP90) se encuentran en pruebas clínicas (Bisht y Singh, 2018). Las estructuras químicas de estos fármacos están en la **Figura 1**.

El uso de la computación como herramienta de investigación enfocada al descubrimiento de fármacos, se ha vuelto indispensable debido al aumento en la cantidad de información química actualmente disponible (Brown et al., 2018). La necesidad y demanda por contar con personal con aptitudes en quimioinformática, es cada vez mayor en la industria farmacéutica y se prevé que ésta aumente en un futuro (Chen et al., 2018). La búsqueda bibliográfica, construcción y minería de bases de datos, análisis de similitud, realización de estudios de cribado virtual, diseño de bibliotecas virtuales, visualización del espacio químico, análisis de resultados de HTS, diseño molecular *de novo*, generación de modelos QSAR y uso de métodos de *machine learning* aplicados a la predicción de propiedades ADME/T y sintéticas, son los principales usos de la quimioinformática en los proyectos actuales de desarrollo de fármacos (Xu y Hagler, 2002).



**Fig. 1. Fármacos y candidatos desarrollados con la asistencia directa de métodos DiFAC.**  
 a) Fármacos aprobados por la FDA para uso clínico. b) Candidatos a fármacos en pruebas clínicas.

### **3.2 DNA metiltransferasas (DNMT) como reguladores de la expresión epigenética: asociación con neoplasias hematopoyéticas**

La epigenética estudia cambios heredables en la estructura de la cromatina en donde no existe una modificación directa a la secuencia de nucleótidos en el DNA (Bonasio et al., 2010). La metilación de DNA representa un mecanismo de regulación de la actividad de genes funcionales a través del silenciamiento de la expresión génica. Este silenciamiento involucra modificaciones covalentes por medio de la transferencia de grupos metilo por la coenzima S-adenosilmetionina (SAM) a la posición 5-C de las citocinas precedentes a guanina en las islas CpG, comúnmente ubicadas en las regiones promotoras de dichos genes (Auclair y Weber, 2012).

La reacción de transferencia es catalizada por la familia de isoenzimas DNA metiltransferasas (DNMT). El genoma humano codifica para cinco de éstas: DNMT1, DNMT2, DNMT3A, DNMT3B y DNMT3L. (Lyko, 2018). DNMT2 está involucrada en la metilación de tRNA (Goll et al. 2006), mientras que DNMT3L, a pesar de presentar homología estructural con DNMT3, no presenta actividad catalítica, no obstante, se sabe que se puede asociar a otras DNMT para regular su actividad (Jia et al., 2007). DNMT1 participa en el mantenimiento de los patrones de metilación en los procesos de replicación de DNA durante la división celular. Por tanto, DNMT1 sólo puede modificar DNA hemimetilado (Jurkowska, 2010). Por otro lado, DNMT3A y DNMT3B están implicadas en la metilación *de novo* de DNA estableciendo los patrones de metilación permanentes durante el desarrollo embrionario temprano (Okano et al., 1999).

Está bien identificada la relación entre la aparición de neoplasias hematopoyéticas y el síndrome mielodisplásico asociada a una mala función en la regulación epigenética (Bennett y Licht, 2018). En células neoplásicas se han observado patrones de metilación aberrantes. Dichos patrones están caracterizados por hipermetilaciones anormales en las regiones promotoras de genes supresores de tumores (Ning et al., 2016). Los mecanismos moleculares por los cuales las islas CpG son hipermetiladas en los genes supresores de tumores implicados en el fenómeno de carcinogénesis, en neoplasias hematopoyéticas, han sido elucidados recientemente. *XAF1* es un gen

supresor de tumores que favorece la apoptosis mediante la inhibición de la AIP (proteínas de inhibición de la apoptosis), inhibe el crecimiento celular y regula el ciclo celular (Zhu et al., 2014). Por otra parte, CTCF es una proteína multifuncional que regula directamente la actividad de *XAF1* a través de su unión a regiones sensibles a la metilación a -386 pb previas al sitio de inicio transcripcional. La ausencia de interacción de CTCF con *XAF1* promueve la hipermetilación de dinucleótidos CpG en las regiones promotoras de los genes mencionados anteriormente (Victoria-Acosta et al., 2015).

Los patrones de metilación en el DNA son reversibles. Esto abre la posibilidad de eliminarlos mediante la inhibición de las DNMT. De este modo, *XAF1* o cualquier otro gen supresor de tumores implicado en un silenciamiento mediado por una hipermetilación aberrante se puede reactivar (Herman y Stephen, 2003) y se puede restaurar la homeostasis en el ambiente celular. Además, se tienen avances en pruebas de fase preclínica para tratar tumores sólidos (Ramachandran et al., 2015). También se ha descubierto un nuevo dominio de aplicación para combatir patologías no oncológicas. Ejemplos de esto son enfermedades metabólicas, autoinmunes e inflamatorias: diabetes, obesidad, lupus eritematoso sistémico y artritis reumatoide, respectivamente (Best y Carey, 2010). Lo anterior hace que DNMT1, DNMT3A y DNMT3B sean blancos moleculares relevantes para el desarrollo de fármacos.

### **3.3 Estado del arte de los Inhibidores de DNMT (I-DNMT)**

Actualmente, azacitidina (5-azacitidina) y decitabina (2'-desoxi-5-azacitidina) (**Figura 2**) son los únicos fármacos aprobados por la FDA para su uso clínico en el tratamiento del síndrome mielodisplásico, leucemia mieloide aguda y leucemia mielomonocítica crónica (Keating, 2009; Saba y Wijermans, 2005). Ambos fármacos son nucleósidos análogos de la citidina. La actividad antineoplásica de estos compuestos se da a través de dos mecanismos: 1) Inhibición de la DNMT1 a dosis bajas, causando hipometilación y 2) Citotoxicidad mediada por su incorporación en el RNA y DNA a altas dosis (Martens, 2010).



Ambos fármacos son incorporados al interior de la célula por el transportador hENT1 (Transportador 1 de nucleósidos equilibrativo humano). Posteriormente, son convertidos a su forma activa trifosforilada por la uridina citidina cinasa, cinasa CMP y difosfocinasa. Después de estas modificaciones, los fármacos se incorporan al núcleo celular para actuar, de acuerdo con la dosis administrada, por los dos mecanismos de acción mencionados anteriormente (Agrawal et al., 2018).

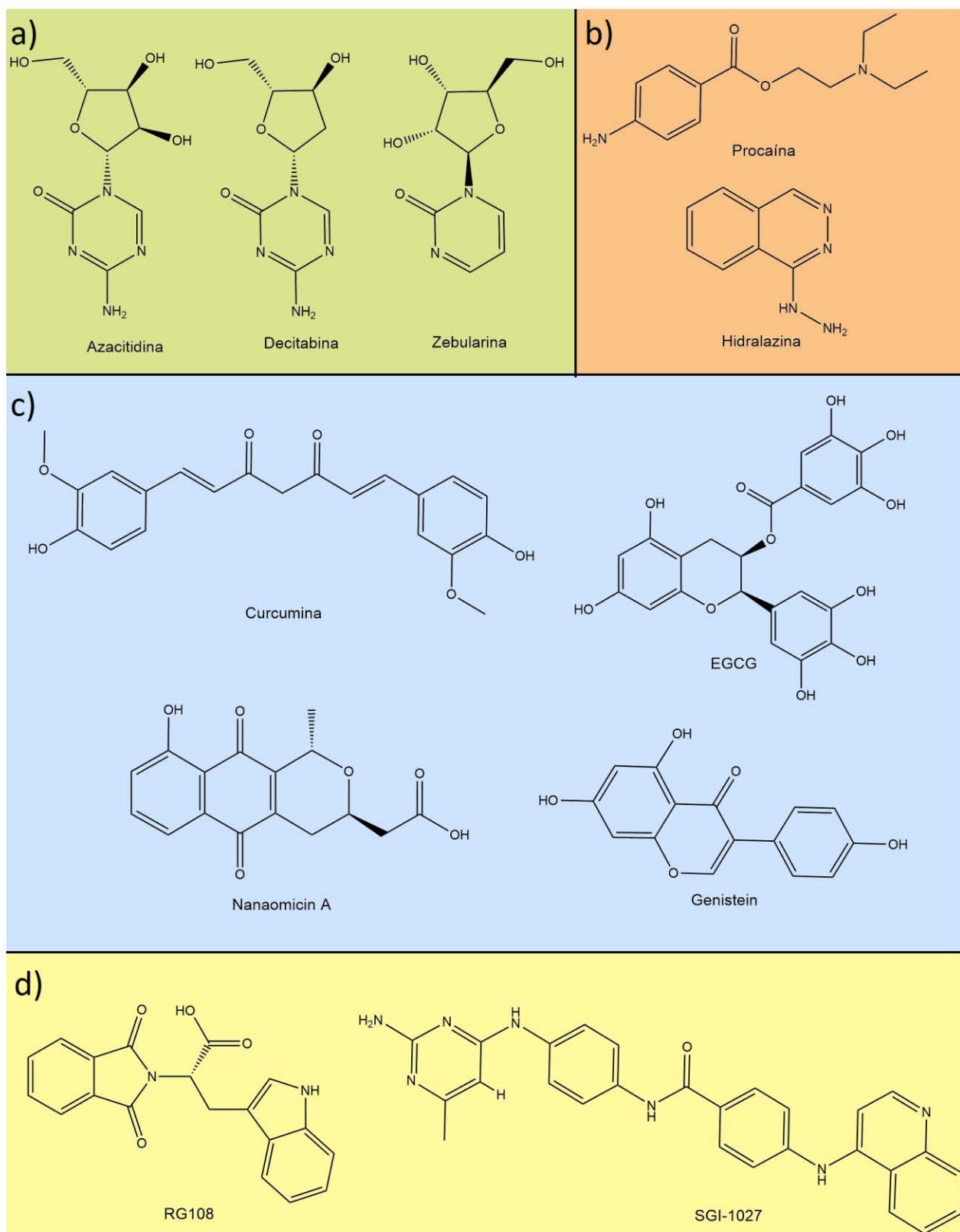
El mecanismo de acción por inhibición de las DNMT de los dos azanucleósidos se basa en la formación de un enlace covalente en la posición 6 del anillo de la citosina, a través de un ataque nucleofílico proveniente del residuo catalítico de la cisteína de la DNMT1. De este modo, se inactiva la enzima, ya que en la posición 5 del mismo anillo hay un nitrógeno que impide su liberación por medio de una reacción tipo  $\beta$ -eliminación (Stresemann y Lyko, 2008). La formación del aducto en la enzima conlleva a su degradación vía proteosomal (Ghoshal et al., 2005).

Como la azacitidina y decitabina son ribonucleósidos, se incorporan en el RNA en mayor proporción (80-90%) que en el DNA (10-20%). La incorporación al RNA impide la síntesis de proteínas induciendo un efecto citotóxico. La incorporación al DNA restringe la síntesis de éste durante los procesos de replicación provocando la inhibición de la proliferación celular (Agrawal et al., 2018).

El segundo mecanismo de acción de los azanucleósidos es el que ocasiona los efectos colaterales de citotoxicidad, ya que, aunque se controle las dosis, se sabe que debido a la variabilidad biológica el metabolismo de fármacos es diferente en cada paciente (Renaissance, 2016). Este último hecho hace que los metabolizadores lentos eleven sus concentraciones plasmáticas aún a bajas dosis. La citotoxicidad no es el problema *per se* (ya que al tratarse de una terapia oncológica esta propiedad podría suponer una ventaja), sino la baja selectividad relacionada a este mecanismo de acción. Otro problema importante es que la eficacia terapéutica depende, en gran medida, de la activación metabólica de ambos fármacos. Las células cancerígenas utilizan este último hecho como mecanismo de resistencia al tratamiento ante estos agentes terapéuticos (Stresemann y Lyko, 2008). El bajo perfil farmacocinético y su poca estabilidad en solución acuosa son otras de las problemáticas de dichos fármacos (Brueckner y Lyko,

2004). De aquí surge la necesidad de diseñar nuevos inhibidores de DNA metiltransferasas no nucleósidos (I-DNMT-NN) que sean selectivos, con un perfil farmacocinético adecuado y que presenten mínima toxicidad (o idealmente, nula) a dosis terapéuticas.

A la fecha, se han hecho esfuerzos para descubrir nuevos I-DNMT-NN (Xu et al., 2016; Erdmann et al., 2016; Saldívar-González, 2018). Ejemplos representativos son los compuestos EGCG, hidralazina, SGI-1027 y RG108 (**Figura 2**). Estos se han identificado, obtenido o desarrollado de productos naturales, procesos de reposicionamiento de fármacos, HTS y cribado virtual. Sin embargo, estos compuestos aún no cuentan con la actividad o selectividad adecuada. Dentro de los métodos de diseño y descubrimiento de fármacos, no se ha reportado el uso de diseño *de novo*.



**Fig. 2. Inhibidores de DNMT.** a) inhibidores nucleósidos. b) inhibidores obtenidos mediante reposicionamiento de fármacos. c) inhibidores provenientes de productos naturales. d) inhibidores obtenidos a través de HTS (SGI-1027) y cribado virtual (RG108).

### 3.4 Enfoques actuales para el diseño molecular *de novo* (DMDN)

#### 3.4.1 Aspectos generales del DMDN

El descubrimiento de fármacos basado en fragmentos (FBDD, *Fragment-Based Drug Discovery*, por sus siglas en inglés), es un método experimental de diseño basado en la diana biológica que consiste en evaluar la afinidad de una biblioteca de fragmentos. Entiéndase por “fragmentos” a moléculas con peso molecular menor a 150 Da hasta llegar a considerar átomos por sí solos. Una de las premisas principales de FBDD es identificar aquellos fragmentos que sean activos dentro de una determinada diana biológica (Erlanson, 2004). Este proceso es seguido por la unión química de los fragmentos y la optimización del ligando generado. En general, esta metodología está guiada por técnicas biofísicas (experimentales) como resonancia magnética nuclear (NMR), cristalografía de rayos X, resonancia de plasmones superficiales (SPR) o titulación calorimétrica isotérmica (ITC). Los resultados son confirmados por ensayos bioquímicos como espectroscopia de correlación de fluorescencia (FCS) (Erlanson, 2012). Eso hace que FBDD sea un proceso costoso y tardado. No obstante, la principal ventaja es que la molécula obtenida por FBDD tiene, en general, una gran afinidad por la diana biológica. Ejemplos de fármacos diseñados por esta técnica son AT9283 (inhibidor de aurora cinasa), AT7519 (antagonista de cinasas dependientes de ciclinas, CDK), PPM204 (inhibidor de receptores activados por proliferadores de peroxisomas, PPAR) y ABT-737 (inhibidor de BCL-2 y BCL-XL) (Hajduk y Greer, 2007).

Por su parte, el DMDN se puede considerar como una aproximación computacional del FBDD. Ambos comparten el objetivo de generar nuevos compuestos bioactivos a partir de fragmentos moleculares. Sin embargo, en el DMDN se busca que las *NME* posean un perfil de propiedades fisicoquímicas y farmacológicas deseables, las cuales son tomadas como criterios de filtro al momento de realizar el diseño. Las *NME* se obtienen mediante la construcción creciente de un ligando dentro un modelo de diana biológica generado a partir de la información estructural proveniente de cristalografía de rayos X o NMR (Schneider y Lloyd, 2013). Las bases del DMDN se asentaron desde hace más de 20 años. Para realizar un DMDN se tenía que implementar diversos programas computacionales por separado para realizar las diferentes etapas del diseño (ver

**página 24)** (Schneider y Fechner, 2005). Hoy en día, gran parte de los programas computacionales para DMDN consideran múltiples características de interés farmacéutico en forma simultánea como: 1) aspectos farmacocinéticos ligados a las propiedades ADME (estimadas a partir de reglas empíricas e.g. reglas de Lipinski); 2) probables efectos tóxicos (anticipados con alertas estructurales) y 3) accesibilidad sintética de las *NME* propuestas (estimada a través de filtros retrosintéticos obtenidos mediante bibliotecas de reacciones orgánicas plenamente identificadas) (Hartenfeller y Schneider, 2010). Lo anterior abre la posibilidad de generar nuevas moléculas sintéticamente accesibles, con diferente núcleo base (derechos de propiedad intelectual), con probable bioactividad y propiedades biofarmacéuticas deseables.

A partir de lo anterior, se puede decir que el DMDN es menos costoso y más rápido de implementar que FBDD. Además, al tomar en cuenta en forma paralela las características de interés farmacéutico, la búsqueda en el espacio químico se ve mejor direccionada sin dejar de ser amplia y diversa, ya que se estima que el número de posibles moléculas tipo fármaco (*drug-like*) entre un peso molecular de 300 a 500 Da tiene un rango de  $10^{20}$  a  $10^{200}$  moléculas (Ruddigkeit et al., 2012).

La principal limitante del DMDN basado en la diana biológica es que para la evaluación y selección de las moléculas generadas se emplea un puntaje (*scoring function*) el cual es una estimación muy aproximada de la afinidad de la unión del ligando con la diana biológica. El puntaje se expresa en términos energéticos ( $\Delta G$  de unión) donde no siempre las contribuciones energéticas individuales de los fragmentos son aditivas (Schneider, 2014). Dicho puntaje o cálculo de la energía de unión está aproximado porque, al igual que en el acoplamiento molecular, no se consideran adecuadamente los factores entrópicos y cinéticos propios del sistema (Prieto-Martínez et al., 2018). A pesar de estas limitaciones, el DMDN se ha aplicado exitosamente en el descubrimiento de nuevos inhibidores de trombina, timidilato sintasa, anhidrasa II carbónica, enzima convertidora de angiotensina, aldosa reductasa, antagonistas de receptores de estrógenos, agentes antitripanosómicos, antifúngicos y algunos antivirales (Suryanarayanan et al., 2018; Krueger et al., 2009). Recientemente, se desarrollaron moduladores de receptores activados por el proliferador de peroxisomas (PPAR) y

receptores de retinoide X (RXR) con actividad nano y micromolar, utilizando conjuntamente el DMDN e inteligencia artificial (Merk et al., 2018).

### 3.4.2 DMDN basado en la estructura de la diana biológica (DMDN-D)

#### 3.4.2.1 Moléculas pequeñas

El DMDN-D tiene cinco etapas: 1) elección del tipo de elemento de construcción; 2) derivación de las restricciones primarias de la diana biológica (RPD) o (*primary target constraints*); 3) selección de los fragmentos moleculares; 4) construcción del ligando y 5) evaluación y selección de las mejores moléculas generadas.

- 1) El DMDN-D puede llevarse a cabo usando tanto fragmentos de bajo peso molecular como sólo átomos. Se ha demostrado que el DMDN basado en fragmentos es mejor que el DMDN basado en átomos. Esto debido a que el DMDN basado en fragmentos genera candidatos más adecuados (ligandos “químicamente racionales”), a pesar de que DMDN basado en átomos es superior en cuanto a diversidad estructural (Schneider, 2014).
- 2) Las RPD se definen como toda la información relacionada con las interacciones del complejo ligando-diana biológica asociadas a la afinidad del mismo complejo. En DMDN-D se aplican tres métodos para calcular las RPD:
  - a) Métodos basados en reglas (*rule-based methods*): Se centran en determinar si se cumplen las condiciones necesarias para formar cierto tipo de interacciones tales como formación de puentes de hidrógeno o interacciones hidrofóbicas (Danziger y Dean, 1989; Gillet et al., 1995). En el proceso se genera un mapa de regiones favorables hacia dichas interacciones. La información cristalográfica se considera como criterio empírico para definir los valores de la geometría y los rangos de tolerancia para obtener el mapa.
  - b) Métodos basados en celdas (*grid-based methods*): Se basan en establecer a lo largo del sitio de unión regiones espaciales de muestreo delimitadas por formas cúbicas (celdas), en las que, dentro de éstas se estiman las mejores zonas para

desempeñarse como sitios de anclaje. Esto se hace a través del cálculo de la energía de unión de diferentes fragmentos “sonda”. De este modo, por medio de un análisis estadístico de todas las celdas estudiadas (utilizado en la mayoría de los programas para realizar DMDN-D) se precisan la RPD (Schneider y Fechner, 2005).

- c) Búsqueda por multicopias simultáneas (*multiple copy simultaneous search, MCSS*): Consiste en posicionar aleatoria y simultáneamente múltiples copias de una biblioteca de grupos funcionales dentro del sitio de unión, a modo de evaluar sus posiciones y orientaciones más favorables por medio de la implementación de campos de fuerza (*force fields*, conjunto de funciones matemáticas que asocia los parámetros necesarios de energía, mecánica y dinámica molecular, para realizar los cálculos de conformación, flexibilidad e interacciones moleculares de un determinado sistema). Los grupos funcionales que posean una energía de unión inferior a un determinado punto de corte son descartados automáticamente (Miranker y Karplus, 1991).

3) Para la selección de los fragmentos moleculares se pueden usar diversas estrategias:

- a) Algoritmos de búsqueda combinatorios e heurísticos (reglas probabilísticas basadas en conocimientos previos que en general tienen buenos resultados, pero que no hay prueba de que esto siempre ocurra). Ejemplos son las estrategias *breadth-first* y *depth-first search*.
- b) Algoritmos aleatorios, como el método Metropolis-Monte Carlo.
- c) Algoritmos de optimización basados en poblaciones, como los algoritmos evolutivos.

Independientemente del algoritmo(s) empleado(s), es importante considerar a las restricciones secundarias de la diana biológica (RSD) o “*second target constraints*”. Entiéndase como RSD a las propiedades esenciales de la molécula que las hace adecuadamente farmacológicas, diferentes de la afinidad de unión con la diana biológica (propiedades ADME/T, accesibilidad sintética y selectividad) (Schneider, 2014).

4) Para la construcción del ligando los métodos más comúnmente aplicados son:

- a) Unión (*linking*): Consiste en posicionar los mejores fragmentos en sus respectivos puntos de anclaje previamente identificados. Una vez hecho esto, se unen los fragmentos mediante conectores (*linkers*) para completar la estructura del ligando.
- b) Crecimiento (*growing*): Consiste en unir a los fragmentos secuencialmente, a modo de que una vez que se posiciona el primer fragmento en su sitio de anclaje, el segundo fragmento es colocado por medio de una ampliación del primer fragmento, considerando el repertorio de fragmentos previamente seleccionados. Se continúa este proceso hasta unir a todos los sitios de anclaje y completar el ligando (Schneider, 2014).

5) La selección de los candidatos más prometedores se hace con base a un puntaje el cual puede ser calculado a partir de las siguientes estrategias:

- a) Basados en campos de fuerza: Se obtienen a través de términos de interacciones no enlazantes, tales como interacciones de Van der Waals y Coulomb.
- b) Empíricos: Se basan en la suma ponderada de las interacciones individuales del complejo ligando-diana biológica.
- c) Basados en conocimiento: Puntaje generado a partir de un análisis estadístico de las interacciones de las moléculas generadas dentro del sitio unión, en donde aquellas que resultan menos frecuentes son consideradas como repulsivas, mientras que las que aparecen más frecuentemente son catalogadas como atractivas (Hartenfeller y Schneider, 2011).



### 3.4.3 DMDN basado en el ligando (DMDN-L)

#### 3.4.3.1 Moléculas pequeñas

En ocasiones se desconoce la diana biológica o la estructura tridimensional aún no está disponible. Sin embargo, sí se conocen ligandos activos contra la diana biológica en cuestión. En este caso, dichos ligandos pueden ser utilizados como plantillas moleculares para generar nuevas moléculas bioactivas. En este enfoque las RPD se obtienen a través de la generación de un pseudoreceptor, el cual es propuesto con la ayuda de un modelo del farmacóforo de ligando tridimensional. Posteriormente, este modelo puede ser usado en los programas computacionales de DMDN-D, siguiendo las etapas propias de este enfoque (ver **página 24**).

A diferencia del DMDN-D, en el DMDN-L el proceso de evaluación y selección de las moléculas candidatas no depende únicamente de un puntaje basado en la estimación de la energía de unión del complejo ligando-diana biológica entre las moléculas candidatas y la molécula de referencia. En DMDN-L el puntaje puede calcularse a partir de:

- a) Un modelo de relaciones estructura-actividad cuantitativo (QSAR, por sus siglas en inglés) previamente establecido.
- b) Medidas de similitud molecular entre los candidatos y la molécula de referencia (Schneider, 2014).

Las RSD pueden ser obtenidas mediante programas computacionales o plataformas externas.

#### 3.4.3.2 Péptidos

Para el DMDN-L enfocado en péptidos se requiere de una plantilla molecular de algún ligando conocido del cual se pueda extraer la información de elementos estructurales para seleccionar la secuencia de aminoácidos del nuevo péptido. Para hacer esto hay dos métodos generales:

- a) Métodos basados en la estructura (e.g.  $\alpha$  hélices, interfaces proteína-proteína o proteínas superficiales).
- b) Métodos basados en secuencias, donde se considera a una secuencia de aminoácidos "padre" o "semilla" como plantilla de diseño sin considerar explícitamente los aspectos conformacionales.

Posterior a esto, el diseño complementario del péptido puede ser guiado a través de la aplicación del algoritmo de reempaque de Monte Carlo (Schneider, 2014).

En ocasiones se carece de información necesaria para realizar DMDN, ya que no se dispone de la información estructural de la diana biológica, relaciones estructura-actividad (SAR, por sus siglas en inglés) o ligandos de referencia. En este caso, se debe realizar una búsqueda aleatoria del espacio químico de péptidos que cubra una amplia diversidad en secuencias (las cuales son medidas a través del cálculo de la entropía de Shannon), a modo de encontrar regiones específicas de interés, reducir progresivamente la diversidad en dichas regiones y construir bibliotecas de péptidos candidatos. Para este fin los algoritmos inspirados en la naturaleza (*nature-inspired algorithms*) han mostrado buenos resultados. Ejemplos son:

- a) Algoritmos evolutivos (*evolutionary algorithms, EA*).
- b) Optimización de partículas de enjambre (*particle swarm optimization, PSO*).
- c) Optimización de colonia de hormigas (*ant colony optimization, ACO*) (Schneider, 2014).

### 3.4.3.3 Productos naturales

Por razones conceptuales el DMDN enfocado a productos naturales (PN) siempre estará dirigido al DMDN-L. Existen dos estrategias que pueden ser utilizadas para la generación de nuevas moléculas bioactivas obtenidas a partir de plantillas moleculares provenientes de PN. La primera estrategia proporciona una clasificación de la diversidad estructural de los núcleos base (*scaffolds*) de los PN, a través de la implementación del algoritmo mapa topográfico generativo, a menudo llamado quimiografías. Las quimiografías permiten visualizar la distribución del panorama

(*landscape*) del espacio químico de los PN y sus compuestos sinteticomiméticos (Miyao et al., 2015). Dado que las quimiografías pueden generarse por medio de características farmacofóricas y descriptores moleculares, es accesible obtener compuestos *scaffold hop* de los PN estudiados (Rodrigues et al., 2016). El segundo enfoque está basado en la simplificación de los núcleos base de los PN que se puede adaptar para generar compuestos semisintéticos. Esta estrategia reduce el marco molecular (*molecular framework*) de los PN mediante la implementación de un algoritmo *scaffold tree* fundamentado en la descomposición basada en reglas de los sistemas de anillo (Bajorath, 2018).

## 4. Objetivos

### General

Proponer nuevas entidades químicas (*NME*) de inhibidores de DNA metiltransferasas no nucleósidos (I-DNMT-NN) mediante diseño molecular *de novo* (DMDN) basado en el modo de unión de la coenzima S-adenosil metionina (SAM) de las DNMT.

### Específicos

1. Utilizar recursos quimioinformáticos adicionales al diseño molecular *de novo* (DMDN) que ayuden a complementar y enriquecer los resultados obtenidos a partir de la metodología implementada.
2. Generar material educativo y tutorial que ayude en el aprendizaje y desarrollo del diseño molecular *de novo* (DMDN), así como de otros recursos quimioinformáticos.

## 5. Material y recursos

A continuación, se listan los materiales, recursos y su propósito:

- *ChemBioDraw Ultra* 14.0: elaboración de estructuras químicas.
- *DataWarrior*: análisis de datos moleculares, cálculo de descriptores y visualización del espacio químico por *PCA*.
- *KNIME Analytics Platform*: curado y filtrado de bases de datos.
- *LigBuilder* v2.0: diseño molecular *de novo* basado en la estructura de la diana biológica (DMDN-D).
- *Molecular Operating Environment (MOE)* 2018.1: acoplamiento y diseño molecular *de novo* basado en el ligando (DMDN-L), usando la función de *Scaffold Replacement*, *RECAP analysis* y *RECAP synthesis*.
- *Openbabel* 2.3.2: Programa soporte de *Ligbuilder* v2.0 convertidor de formatos moleculares.
- *PyMOL* 1.7.2.1: Visualización de estructuras químicas 3D del DMDN-D.
- *RCSB Protein Data Bank (PDB)*: fuente de información cristalográfica de DNMT1 y DNMT3A.
- Servidores *AdmeSAR*, *SwissADME* y aplicación *Toxtree*: evaluación de las propiedades ADME/T de las moléculas generadas a partir del diseño *de novo* basado en el ligando para análogos de SAH

## 6. Metodología

### 6.1 DMDN-D de inhibidores de DNMT

1. A partir de la información estructural de la isoenzima DNMT3A (PDB ID: 2QRV) se generaron *NME* por DMDN-D empleando el programa computacional Ligbuilder v2.0, implementado la estrategia exploración (*Exploring strategy*).
2. Los mejores candidatos obtenidos en el punto 1 fueron sometidos a estudios de acoplamiento molecular con la estructura 3D de DNMT3A con el mismo programa Ligbuilder v2.0.

### 6.2 DMDN-L de I-DNMT

#### 6.2.1 Uso de bases de datos de acceso libre para obtener plantillas moleculares

##### 6.2.1.1 DMDN-L por cambio de núcleo base (*scaffold replacement*)

1. Se tomó como punto de partida una base de datos (BD) de 16,102 inhibidores dirigidos contra diversas dianas epigenéticas provenientes de las BD de acceso libre ChEMBL y BindingDB (Naveja y Medina-Franco, 2018). Esta última BD se curó y filtró para generar una nueva BD de 26 I-DNMT3A-NN reportados como activos.
2. La BD obtenida en el punto 1 fue sometida a estudios de acoplamiento molecular. Los primeros estudios se enfocaron en el sitio catalítico y los segundos fueron acoplamientos moleculares ciegos.
3. Se realizó un segundo filtrado para considerar únicamente a los I-DNMT3A-NN que tuviesen una energía de unión ( $\Delta G$  aproximado) más favorable que la coenzima SAH (*S*-adenosilhomocisteína, -7.563 Kcal/mol).
4. La BD resultante tuvo diez I-DNMT3A-NN con cuatro núcleos base diferentes. De estos, aplicando la función *scaffold replacement* en MOE.2018.1, se obtuvieron 13,468 moléculas candidatas.
5. Las moléculas candidatas fueron sometidas a estudios de acoplamiento molecular.

6. Las BD de los acoplamientos de las moléculas candidatas fueron filtradas para seleccionar a los nuevos I-DNMT3A-NN que tuviesen una energía de unión ( $\Delta G$  aproximado) más favorable que SAH (-7.563 Kcal/mol) o PS-3114 (compuesto prototipo (*hit*) previamente identificado por el Dr. Alexandre Gagnon con una energía de unión calculada de -5.649 Kcal/mol).

#### **6.2.1.2 DMDN-L por fragmentos moleculares generados a través de reglas basadas en análisis retrosintéticos estadísticos (*RECAP analysis* y *RECAP synthesis*)**

1. Para DNMT1 se generaron *NME* aplicando la función *RECAP analysis* y *RECAP synthesis* en MOE.2018.1.
2. Las *NME* propuestas en el paso anterior se filtraron con cribado virtual usando acoplamiento molecular ciego para seleccionar aquellas *NME* que sólo tuviesen potencial de unirse al sitio catalítico.

#### **6.2.2 Uso de la información estructural de la coenzima S-adenosilhomocisteína (SAH) co-cristalizada como plantilla molecular**

1. Se determinaron las interacciones clave de SAH co-cristalizada en la enzima DNMT3A (ID PDB: 2QRV) a través de acoplamiento molecular (como parte del mismo protocolo del acoplamiento molecular).
2. Por medio de la función *scaffold replacement* en MOE.2018.1 se generó una BD de 426 moléculas análogas a SAH.
3. Se realizó acoplamiento molecular a la BD anterior.
4. De la BD anterior se obtuvieron cinco moléculas con una energía de unión más favorable a PS-3114.
5. Las RSD de los mejores candidatos fueron estimadas mediante los servidores y programa externos: *AdmeSAR*, *SwissADME* y *Toxtree*, respectivamente.

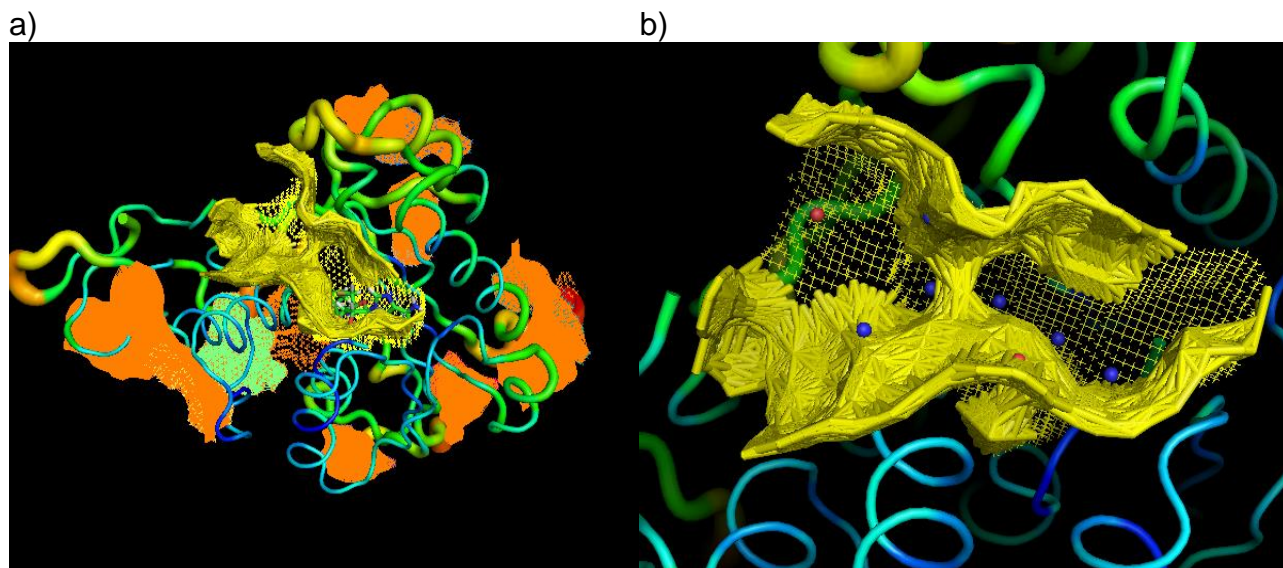
## 7. Resultados

### 7.1 DMDN-D de I-DNMT

#### 7.1.1.1 Detección de la cavidad y modelo del farmacóforo de I-DNMT3A (*cavity function*)

El DMDN-D se realizó con *Ligbuilder v2.0*. La primera etapa del DMDN-D con dicho programa (utilizando la función *cavity*) consiste en determinar y detectar las posiciones, formas y volúmenes de los posibles sitios de unión en la superficie de la diana biológica, así como estimar cuantitativamente su capacidad de afinidad en dichos sitios de unión (*druggability*) (Yuan, Pei, y Lai, 2011). Esta etapa corresponde a la derivación de las RPD.

Se detectaron nueve posibles sitios de unión en la superficie de la DNMT3A (**Fig. 3a**). Para el DMDN-D, se eligió al que está asociado con el sitio catalítico de la enzima. La **Fig. 3b** ilustra al sitio de unión asociado al sitio catalítico y el modelo del farmacóforo propuesto por el programa computacional.



**Fig. 3. Sitios de unión y modelo del farmacóforo en la DNMT3A propuestos por *Ligbuilder v2.0*.** a) naranja: sitios de unión alostéricos, amarillo: sitio de unión asociado al sitio catalítico, verde: *S*-adenosilhomocisteína; b) puntos azules: áreas donadoras de puente de hidrógeno, puntos rojos: áreasceptoras de puente de hidrógeno. La función *cavity* de *Ligbuilder v2.0* fue entrenada por un conjunto de prueba de 1300 complejos ligando-proteína presentando una precisión del 86% (Yuan, Pei, y Lai, 2011).

### **7.1.1.2 Nuevas entidades químicas (*NME*) de I-DNMT3A generadas por algoritmos genéticos y reglas químicas (*Build function*)**

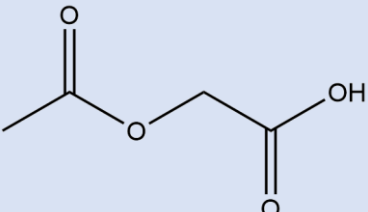
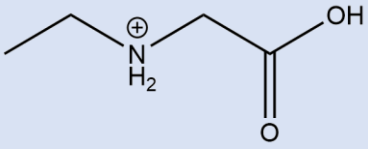
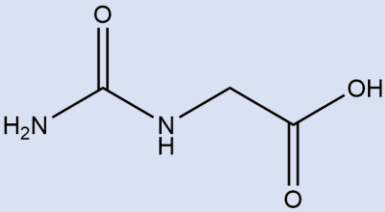
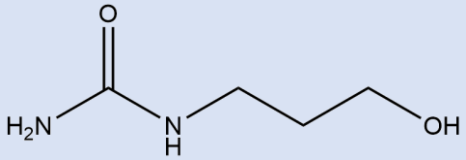
Se obtuvieron siete *NME* de I-DNMT3A (ver **Tabla 1**) mediante el método de construcción del ligando *exploring strategy*, el cual diseña los ligandos a partir de una biblioteca de fragmentos predeterminados por el mismo programa. Esta estrategia no requiere información estructural de ligandos previamente conocidos (plantillas o “semillas” moleculares). La construcción y selección de las moléculas se guió a través de un algoritmo genético considerando un número de generaciones y poblaciones de siete y doscientos, respectivamente, hasta un máximo de 10000 moléculas por generación.

### **7.1.1.3 Análisis de accesibilidad y rutas sintéticas de las *NME* de I-DNMT3A**

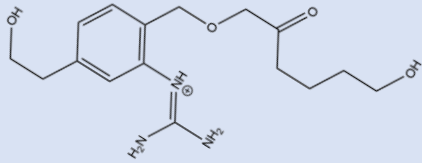
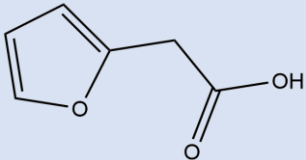
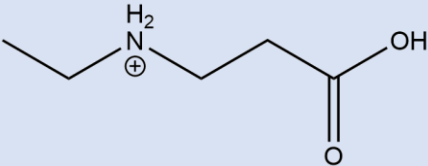
La accesibilidad sintética de los compuestos diseñados está basada en un análisis retrosintético que permite generar moléculas que sean fáciles de sintetizar. Las moléculas propuestas son transformadas en precursores más simples. El método finaliza hasta que los precursores sean moléculas muy simples o que sean moléculas comercialmente disponibles (ver **Fig. 4**) (Yuan, Pei, y Lai, 2011).



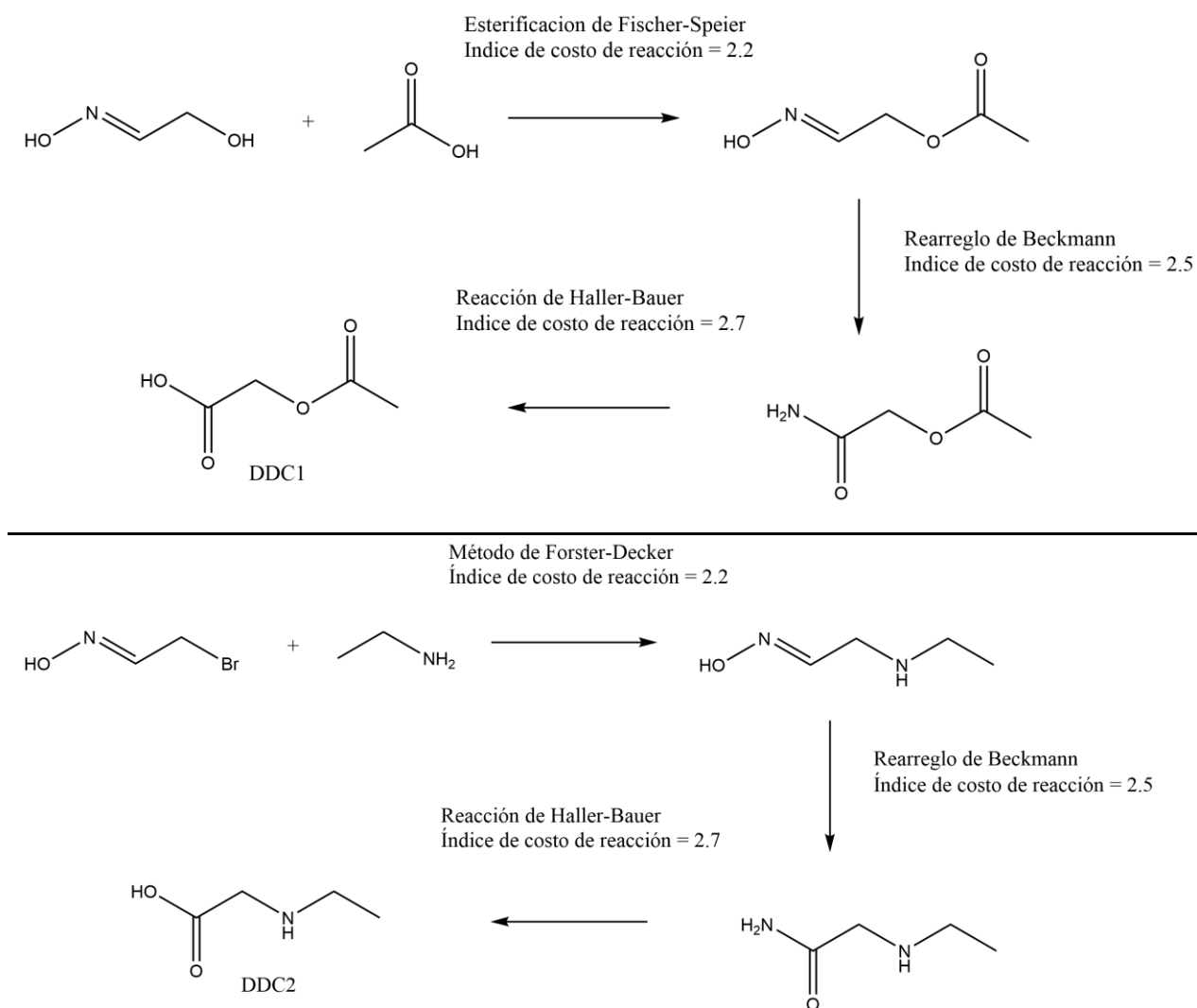
Tabla 1. Compuestos generados por el programa computacional *Ligbuilder v2.0* para la enzima DNMT3A.\*

ID (nombre)	Estructuras 2D	Propiedades del ligando
DDC1		Puntaje: -4.55 Kcal/mol
		Peso molecular: 118 Da
		XLogP: -0.38
		Átomos pesados: 8
		Enlaces rotables: 3
Estabilidad del ligando: 66%		
DDC2		Puntaje: -4.38 Kcal/mol
		Peso molecular: 104 Da
		XLogP: -0.37
		Átomos pesados: 7
		Enlaces rotables: 3
Estabilidad del ligando: 68%		
DDC3		Puntaje: -4.45 Kcal/mol
		Peso molecular: 118 Da
		XLogP: -1.79
		Átomos pesados: 8
		Enlaces rotables: 2
Estabilidad del ligando: 65%		
DDC4		Puntaje: -4.65 Kcal/mol
		Peso molecular: 118 Da
		XLogP: -1.52
		Átomos pesados: 8
		Enlaces rotables: 3
Estabilidad del ligando: 61%		

(Tabla 1 continuación)

<b>DDC5</b>		<b>Puntaje:</b> -5.71 Kcal/mol
		<b>Peso molecular:</b> 324 Da
		<b>XLogP:</b> -0.73
		<b>Átomos pesados:</b> 23
		<b>Enlaces rotables:</b> 10
		<b>Estabilidad del ligando:</b> 49 %
<b>DDC6</b>		<b>Puntaje:</b> -4.32 Kcal/mol
		<b>Peso molecular:</b> 126 Da
		<b>XLogP:</b> 0.15
		<b>Átomos pesados:</b> 9
		<b>Enlaces rotables:</b> 2
		<b>Estabilidad del ligando:</b> 61 %
<b>DDC7</b>		<b>Puntaje:</b> -4.13 Kcal/mol
		<b>Peso molecular:</b> 118 Da
		<b>XLogP:</b> -0.21
		<b>Átomos pesados:</b> 8
		<b>Enlaces rotables:</b> 4
		<b>Estabilidad del ligando:</b> 61%

\* La selección de los compuestos generados por el algoritmo genético se hizo por medio del cálculo de un puntaje (*scoring function*) empírico, para estimar la afinidad de unión del complejo ligando-enzima. El puntaje fue construido y entrenado con 800 complejos ligando-proteína que contenían afinidades de unión experimentales conocidas, provenientes de la base de datos PDBbind. El coeficiente de correlación y la desviación estándar para este puntaje fue de 0.68 y 1.8 Kcal/mol, respectivamente. Las propiedades de los ligandos se manejaron en los siguientes rangos: log P: -0.4 a 5.6, peso molecular: 100 a 500 Da y átomos pesados: 20 a 70 (Yuan, Pei, y Lai, 2011).



**Fig. 4. Rutas sintéticas de los compuestos generados por el programa Ligbuilder v2.0.**

En la imagen se ilustran las rutas sintéticas con los menores índices de coste de reacción (donde un valor de 1 es muy fácil de sintetizar y un valor de 10 es muy difícil de sintetizar). Este índice contempla variables como temperatura, número de reactivos y condiciones especiales de reacción. En la imagen no se muestra el compuesto DDC5, ya que el programa solo planea las rutas sintéticas para aquellos compuestos que alcancen una estabilidad del ligando mayor al 50 %.

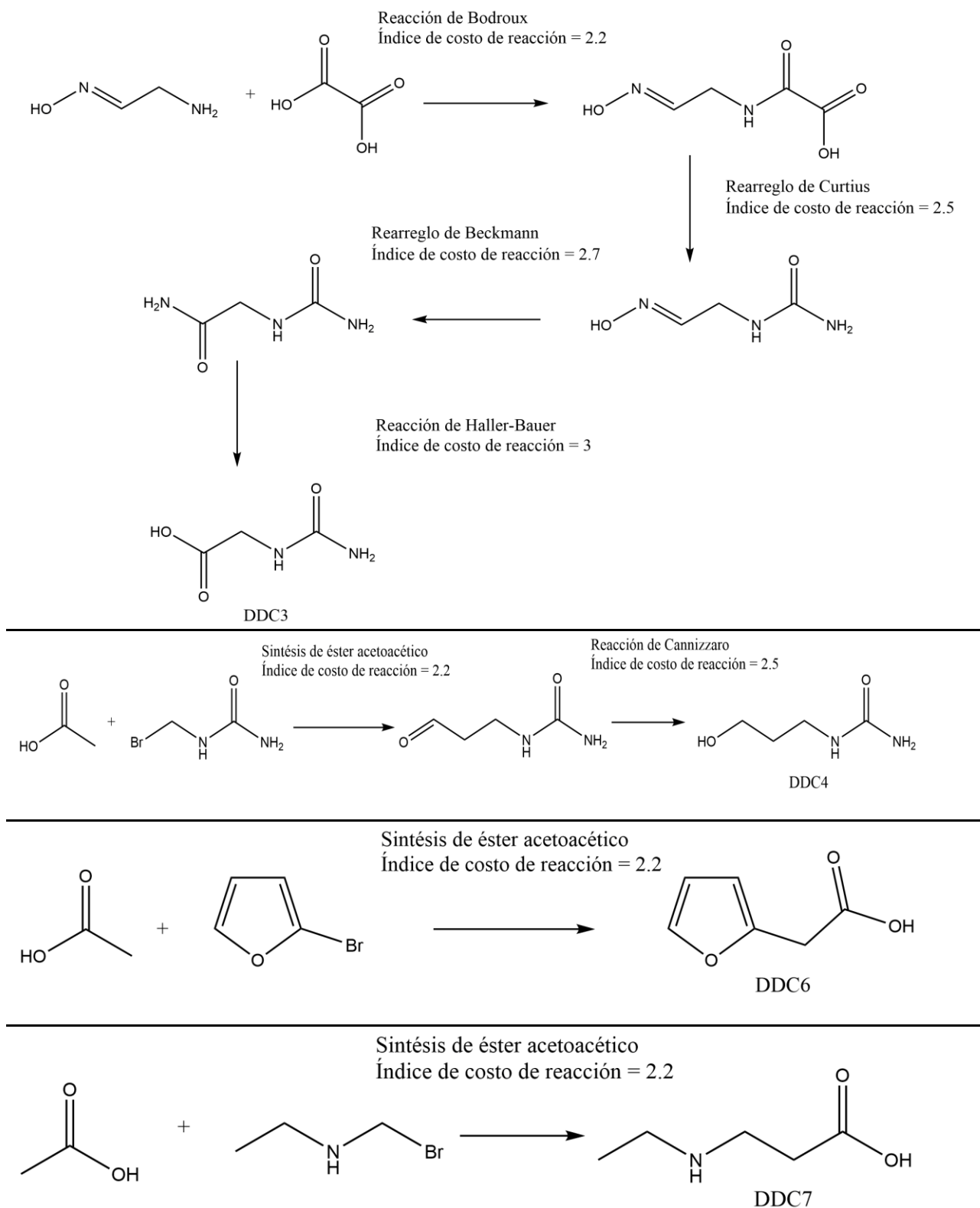


Fig. 4 (continuación)

## 7.2 DMDN-L de I-DNMT

### 7.2.1 Uso de bases de datos de acceso libre para obtener plantillas moleculares

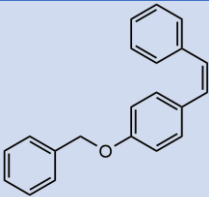
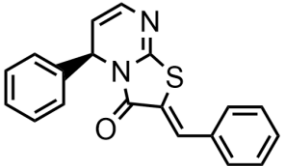
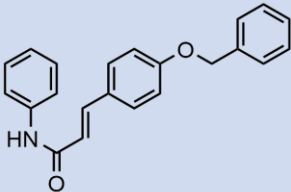
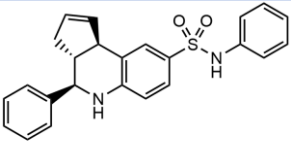
#### 7.2.1.1 DMDN-L por cambio de núcleo base (*scaffold replacement*)

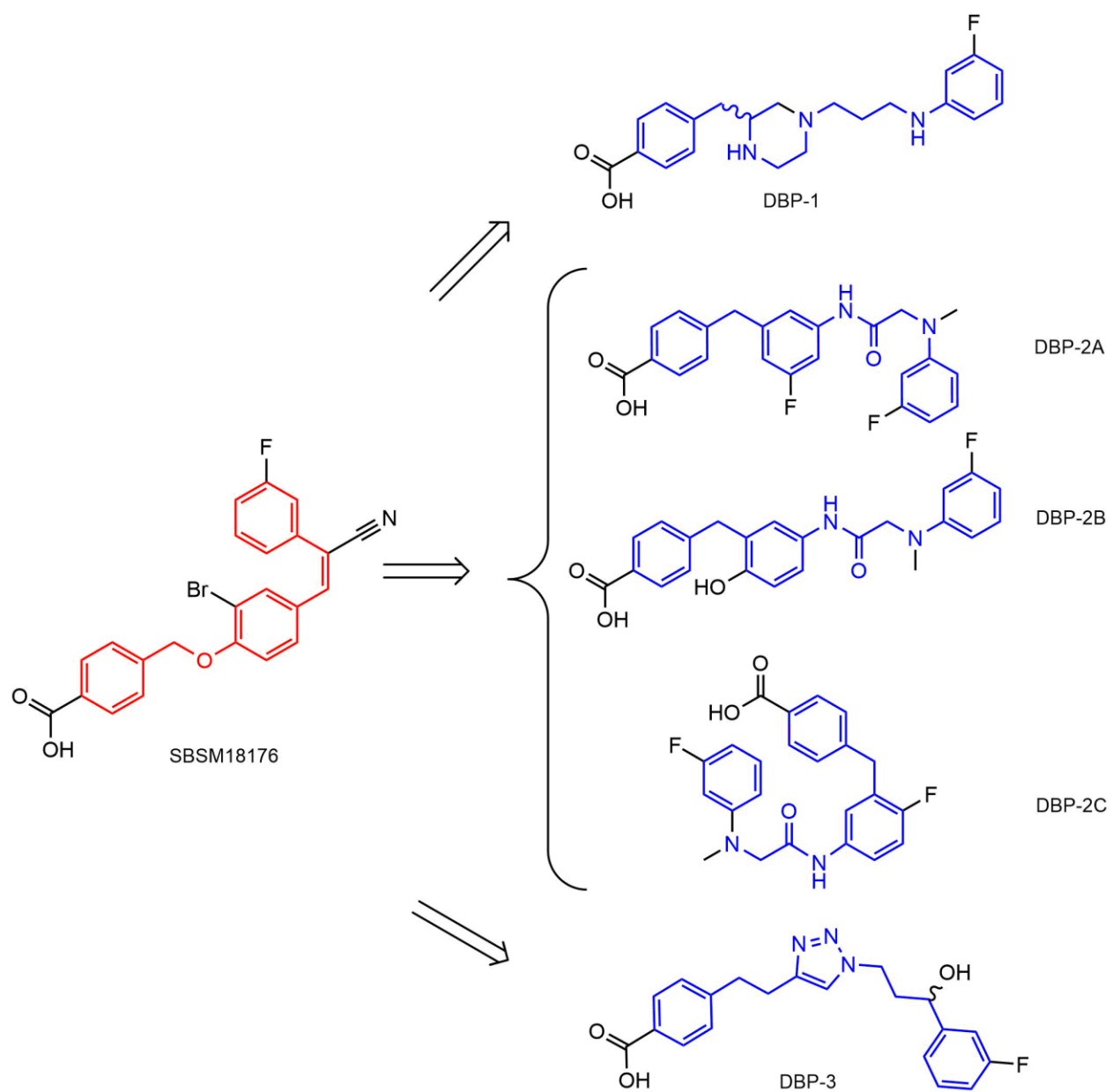
Se obtuvieron un total de 137 compuestos y 41 nuevos núcleos base provenientes de la BD de las 13,468 moléculas candidatas (ver metodología, **página 20**). Los resultados están resumidos en la **Tabla 2**. Para los compuestos de referencia SBSM18176 y SBSM534522, los compuestos obtenidos presentaron un puntaje más favorable que SAH (-7.563 Kcal/mol). Por otro lado, para el compuesto SBSM16556 los compuestos obtenidos sólo mostraron un puntaje más favorable a PS-3114 (inhibidor previamente identificado, -5.6489 Kcal/mol). Para el compuesto SBSM534514 se obtuvieron moléculas con un puntaje más favorable tanto para SAH como para PS-3114.

Para los compuestos generados a partir de SBSM18176, SBSM534514 y SBSM534522 se calculó un rango porcentual de accesibilidad sintética mayor al 50 %. Para los compuestos obtenidos a partir de SBSM534514 se calculó un rango de hasta el 45 % (**Tabla 2**).

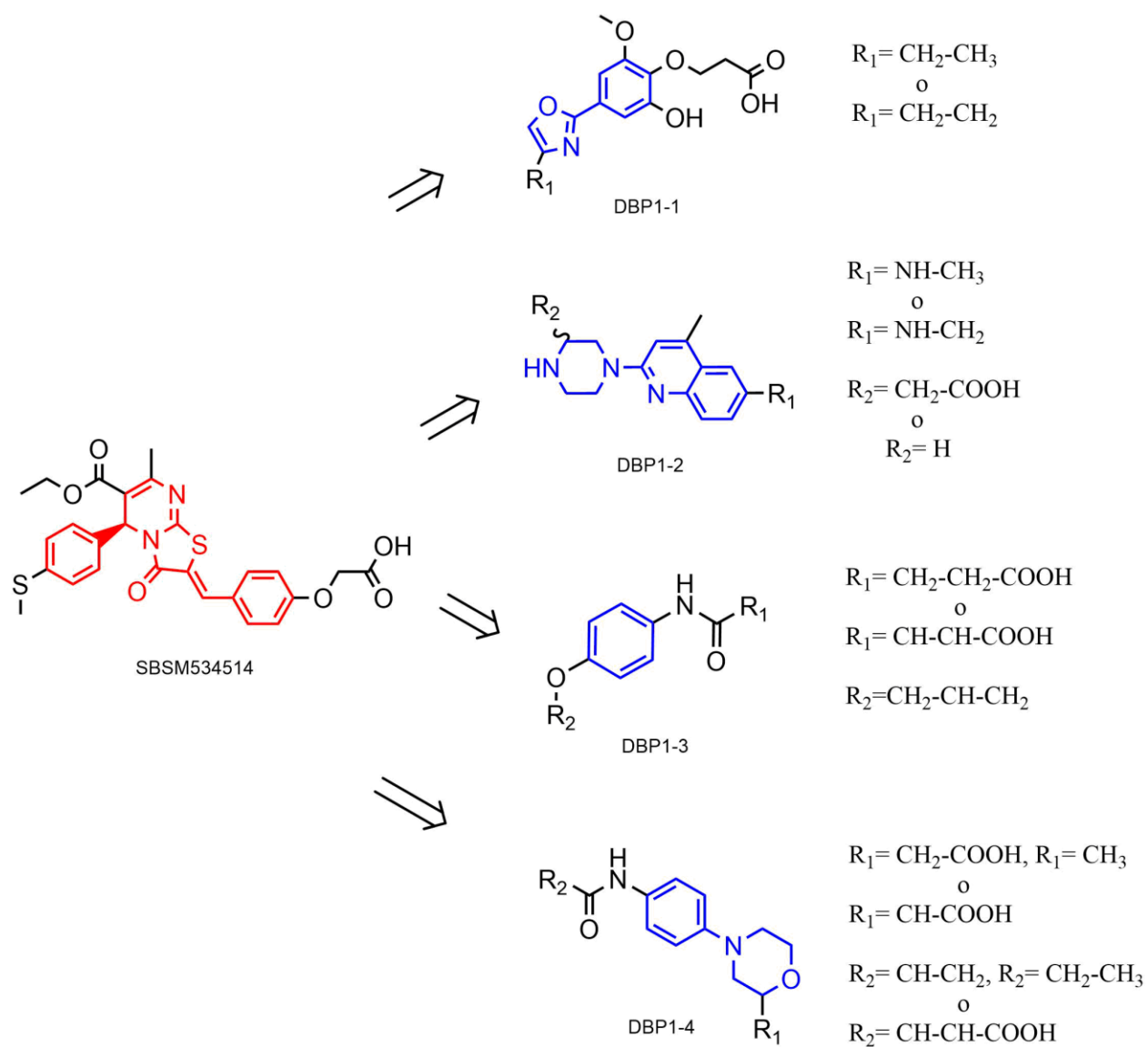
Las **Figuras 5-8** muestran ejemplos de las estructuras obtenidas para cada compuesto de referencia. Cabe mencionar que sólo para el compuesto SBSM18176 se muestran todos los compuestos obtenidos.

**Tabla 2. Resultados del cribado virtual basado en la diana biológica de la BD de I-DNMT3A.**

Núcleos base de los compuestos de referencia	Número de compuestos obtenidos	Número de nuevos núcleos base obtenidos	Puntaje más favorable que	Accesibilidad sintética (%)
 SBSM18176	5	3	SAH	52-89
 SBSM534514	16	5	PS-3114 y SAH	52-73
 SBSM16556	102	22	PS-3114	Hasta el 45
 SBSM534522	14	11	SAH	50-60
<b>Total</b>	<b>137</b>	<b>41</b>		



**Fig. 5.** Compuestos generados por la función *scaffold replacement* para el compuesto de referencia SBSM18176.



**Fig. 6. Compuestos obtenidos por la función *scaffold replacement* para el compuesto de referencia SBSM534514.**



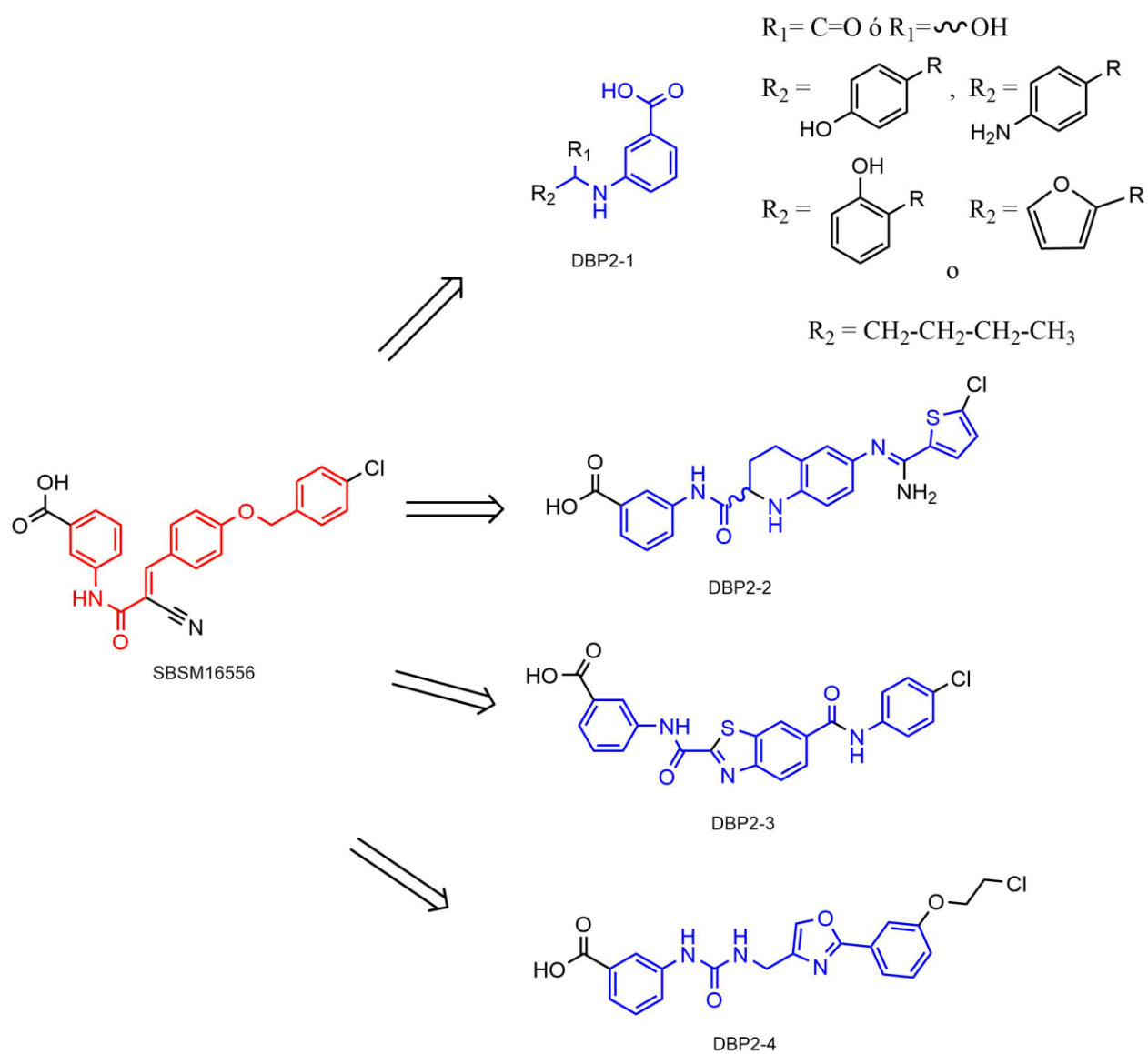
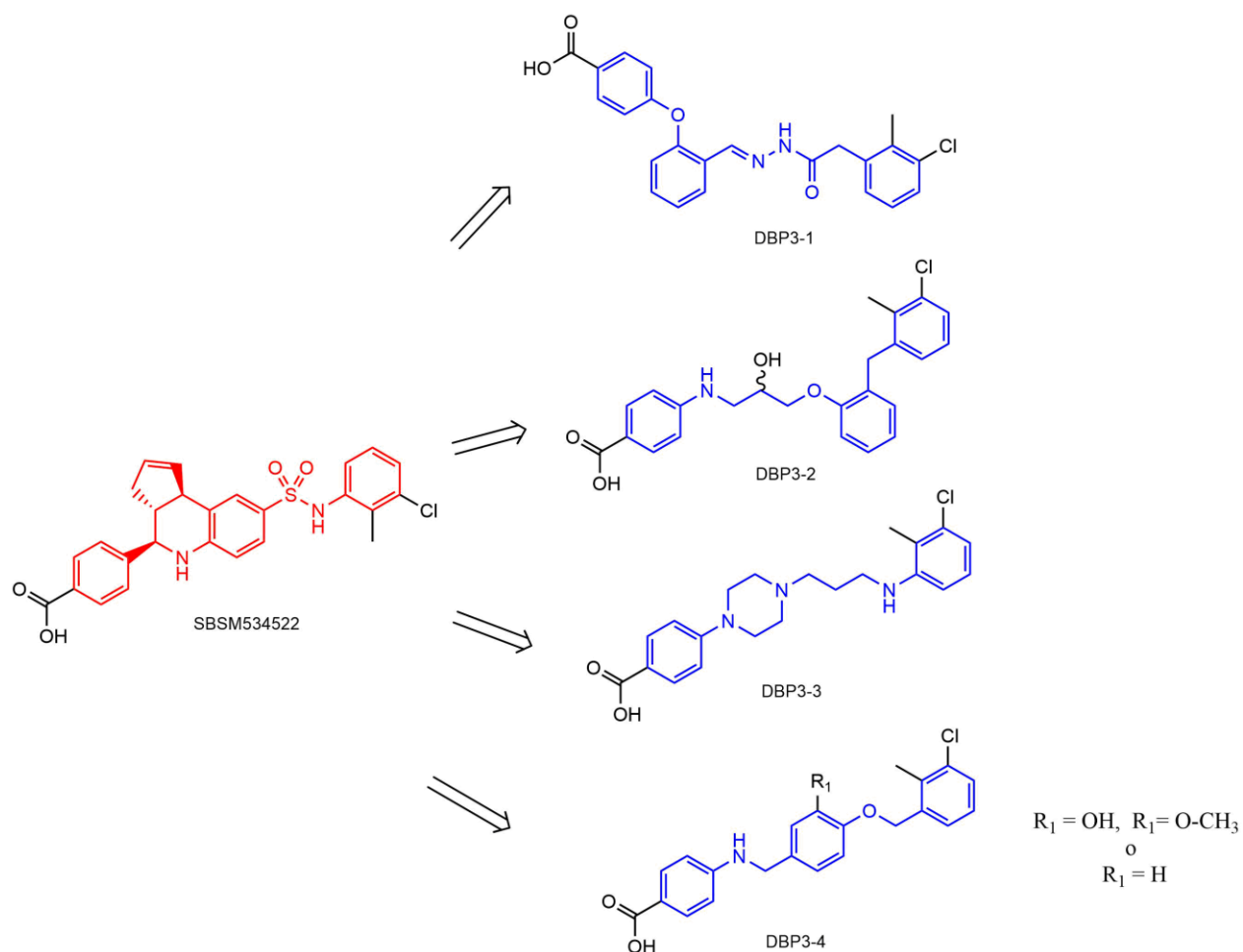


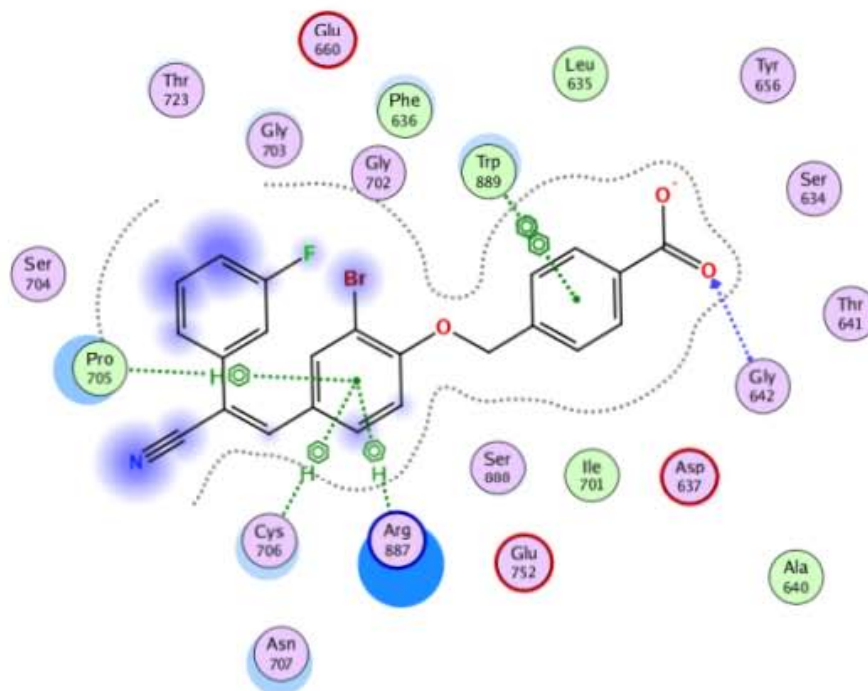
Fig. 7. Compuestos obtenidos por la función *scaffold replacement* para el compuesto de referencia SBSM16556.



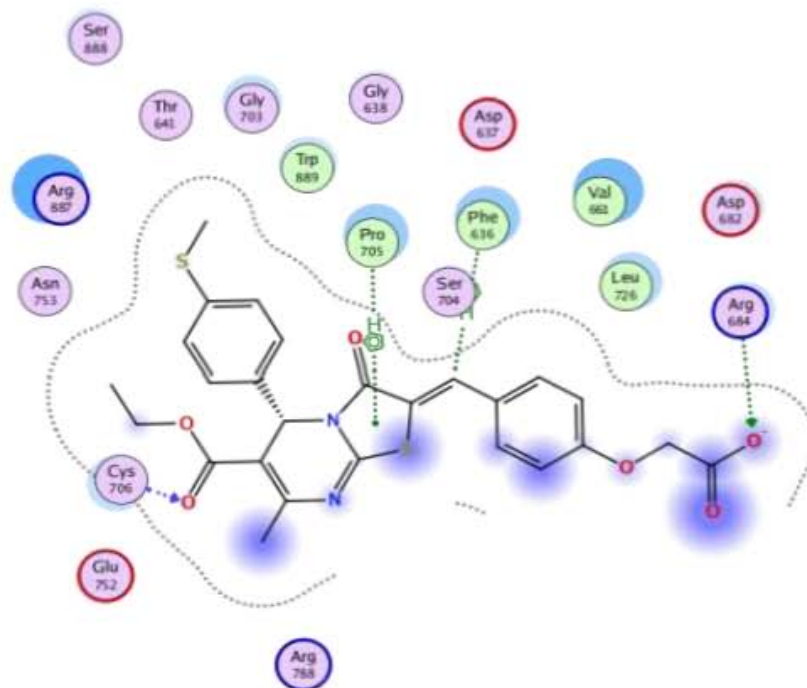
**Fig. 8.** Compuestos obtenidos por la función *scaffold replacement* para el compuesto de referencia SBSM534522.

En las cuatro moléculas de referencia el carboxilato se retuvo y conservó para el diseño de las *NME*. Esta porción molecular presentó interacciones moleculares relevantes para la formación de puentes de hidrógeno. Cabe destacar que en los modelos de acoplamiento se predice que la Gly 642, en el sitio catalítico, forma estas interacciones en todos los compuestos con excepción del compuesto SBSM534514 (ver **Fig. 9**).

a)

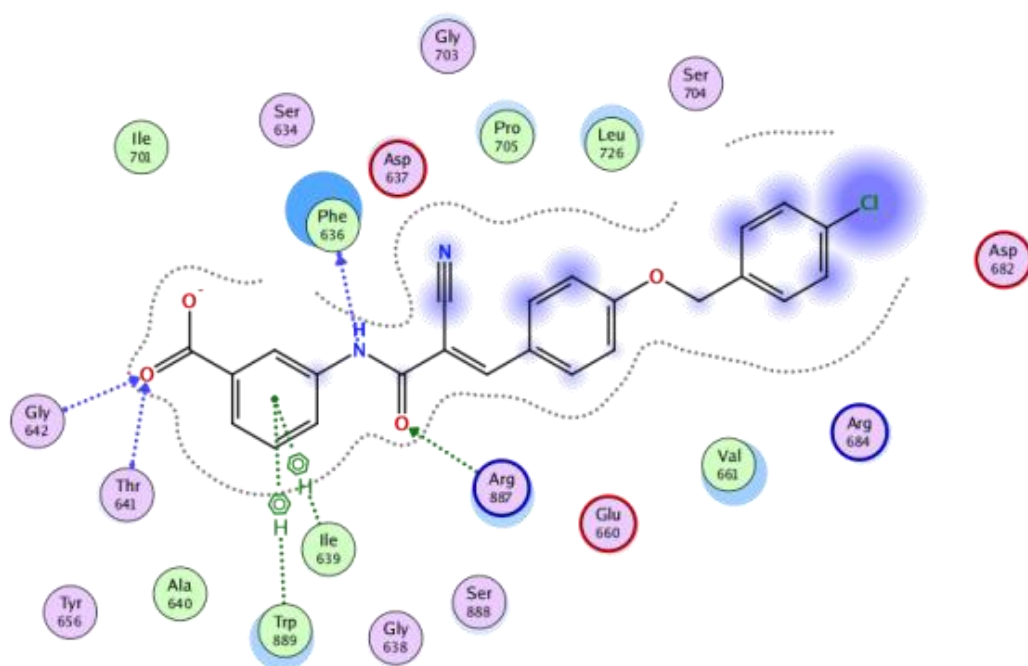


b)

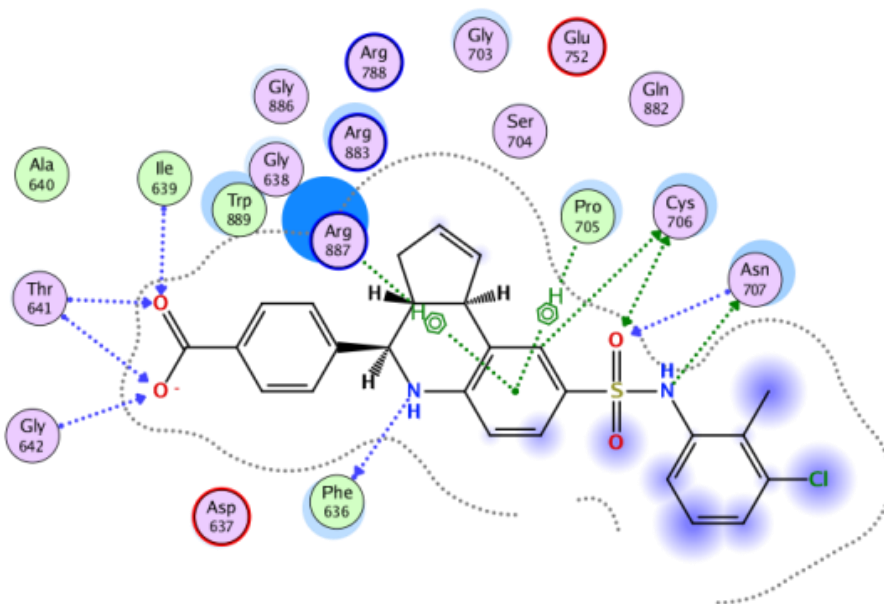


**Fig. 9. Acoplamiento molecular de los compuestos de referencia.** Las regiones farmacofóricas identificadas en los modelos de acoplamiento se emplearon para la función *scaffold replacement*. a) SBSM18176: formación de un puente de hidrógeno con el carboxilato de Gly 642; interacción hidrófoba con el anillo fenílico de Trp 889; b) SBSM534514: formación de dos puentes de hidrógeno con Arg 684 y Cys 706 en la porción del carboxilato y carbonilo del éster, respectivamente.

c)



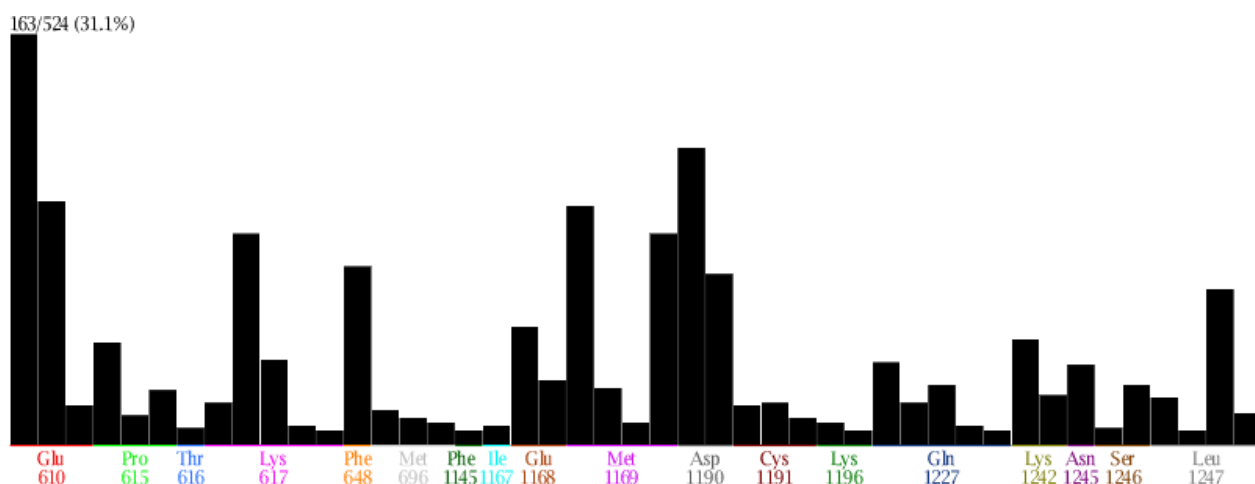
d)



**Fig. 9. (continuación).** C) SBSM16556: interacciones tipo puente de hidrógeno con el carboxilato del ácido benzoico, carbonilo y nitrógeno de la amida con los aminoácidos Gly 642, Thr 641, Phe 636 y Arg 887, respectivamente y d) SBSM16556: formación de puentes de hidrógeno en la Gly 642, Thr 641, Ile 639, Phe 636, Cys 706 y Asn 707.

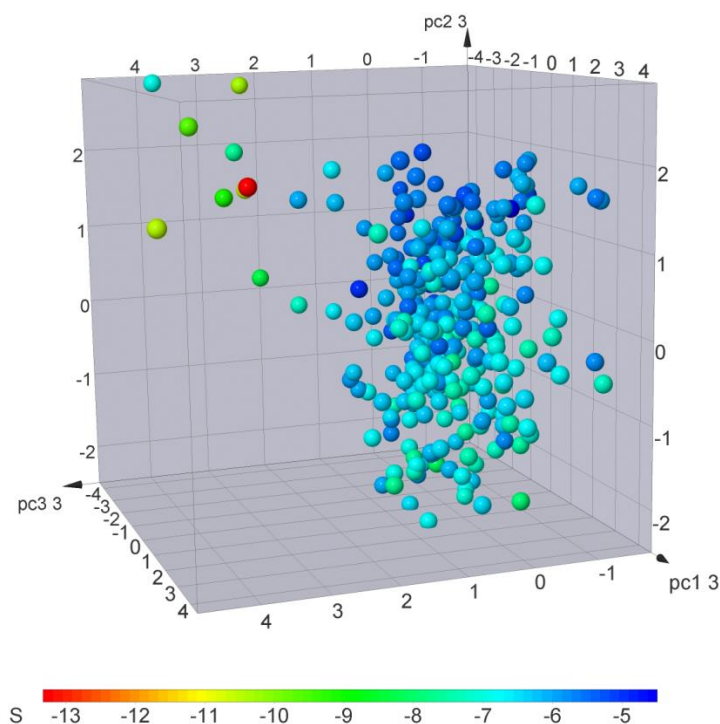
### 7.2.1.2 DMDN-L por fragmentos moleculares generados a través de reglas basadas en análisis retrosintéticos estadísticos (*RECAP analysis* y *RECAP synthesis*)

Se obtuvieron 1000 *NME* para la isoenzima DNMT1 a través del uso de la función *RECAP synthesis* (ver metodología, **página 21**). De éstas, se estima que 362 *NME* podrían unirse al sitio catalítico. Para estas *NME* se obtuvo un total de 524 modos de unión. De acuerdo con la **Figura 10**, que muestra la frecuencia de interacción de las 362 *NME* con aminoácidos del sitio de unión, 31.1 % de las interacciones ocurren con Glu 610. Este resultado sugiere que este aminoácido es importante para formación del complejo ligando-enzima en las *NME*.



**Fig. 10.** Frecuencia de interacción de las 362 *NME* diseñadas con aminoácidos del sitio catalítico de DNMT1. 31.1% de los 524 modos de unión dentro del sitio catalítico de la DNMT1 tienen interacción con Glu 610 seguido por Asp 1190.

La **Fig. 11** muestra que las propiedades fisicoquímicas para las *NME* podrían influir en la probable actividad de las *NME*. En el esquema se puede ver la formación de agrupaciones moleculares (*clusters*) en la visualización del espacio químico, a razón del puntaje obtenido que está representado con color, en una escala continua.

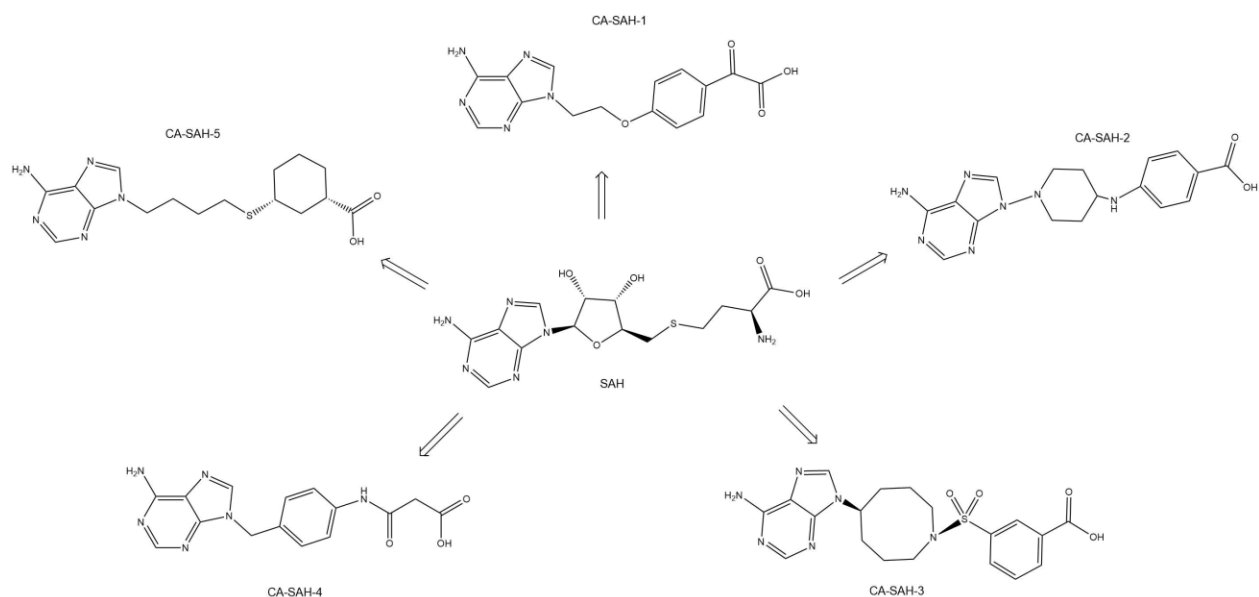


**Fig. 11. Visualización del espacio químico de las 362 NME obtenidas por la función *RECAP synthesis* estimadas con unión al sitio catalítico de DNMT1.** La visualización del espacio químico se obtuvo por análisis de componentes principales (PCA). Con un porcentaje de varianza recuperada de: PC1 = 52.694, PC2 = 22.311 y PC3 = 15.165. Considerando los descriptores: peso molecular, log P, número de aceptores de puentes de hidrógeno (*HBA*), número de donadores de puente de hidrógeno (*HBD*) y área superficial polar topológica (*TPSA*). El puntaje “S” obtenido por acoplamiento molecular con MOE.2018.1 está representado con una escala continua de color.

## 7.2.2 Uso de la información estructural de la coenzima S-adenosilhomocisteína (SAH) co-cristalizada como plantilla molecular

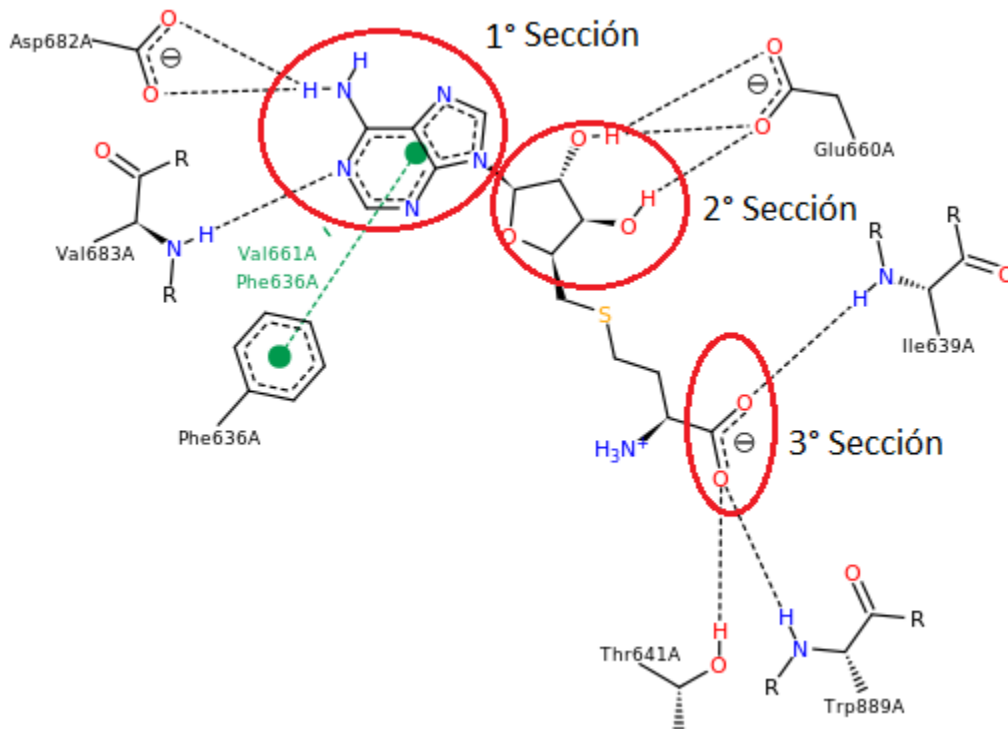
### 7.2.2.1 Análogos de S-adenosilhomocisteína (SAH)

Tal como se describió en la sección de metodología (**página 21**), se tomó como plantilla molecular el núcleo base de la coenzima SAH. Utilizando la función *scaffold replacement* se generó una BD de 426 moléculas análogas a SAH. Al realizar un cribado virtual de esta BD, se obtuvieron cinco moléculas análogas a SAH (**Fig. 12**) con una energía de unión que se sugiere es más favorable a PS-3114 (inhibidor previamente identificado).



**Fig. 12. Análogos de SAH generados por la función *scaffold replacement* eliminando la segunda sección molecular ligada al efecto colateral de citotoxicidad presente en los fármacos nucleósidos.** En la figura, las moléculas están nombradas y numeradas en orden descendiente con respecto a su energía de unión calculada.

La **Figura 13** indica que SAH tiene tres secciones moleculares principales que están asociadas a interacciones clave del complejo ligando-enzima endógeno, tanto en DNMT1 como en DNMT3A: 1) base nitrogenada (adenina); 2) pentosa y 3) carboxilato terminal (correspondiente al aminoácido metionina). La generación de los análogos de SAH se realizó conservando la primera y tercera sección con el objeto de que se conservaran el mayor número de interacciones clave y para pretender que la afinidad fuese similar al de SAH. La eliminación de la segunda sección se realizó para contrarrestar y descartar el segundo mecanismo de acción asociado a los efectos colaterales de citotoxicidad de los fármacos nucleósidos disponibles hasta la fecha. De esta manera, se impediría la metabolización de las *NME* mediadas por la trifosforilación en el hidroxilo contiguo al carbono 5 de la pentosa y su posterior internalización al núcleo celular para incorporarse al RNA o DNA (ver Introducción, **página 14**).



**Fig. 13. Secciones moleculares ligadas a las interacciones clave en el complejo ligando-enzima entre SAH y DNMT3A.** Imagen modificada de la original obtenida de RCSB, PDB ID: 2QRV.

### 7.2.2.2 Propiedades fisicoquímicas, toxicológicas y químico farmacéuticas de los análogos de SAH

Las **Tablas 3-5** resumen los resultados del cálculo de las propiedades ADME/T obtenidas a través de las herramientas quimioinformáticas descritas en la sección de Metodología.

**Tabla 3. Propiedades toxicológicas y farmacocinéticas relevantes calculadas con admetSAR.\***

Compuesto	admetSAR			
	Solubilidad en agua (log S)	Carcinogénesis (en roedor)	Mutagénesis (prueba de Ames)	Inhibidor de CYP 450
CA-SAH-1	-3.34	no	no	no
CA-SAH-2	-2.8	no	no	no
CA-SAH-3	-3.66	no	no	2C9, 2C19
CA-SAH-4	-3.44	no	no	no
CA-SAH-5	-3.8	no	no	no

\*Cada propiedad es estimada a través de modelos de QSAR de alta calidad. (Wang et al., 2007; Lagunin et al., 2009; Hansen et al., 2009; Cheng et al., 2011 y Cheng et al., 2012).



**Tabla 4. Propiedades farmacocinéticas y químico farmacéuticas de interés calculadas con SwissADME.\***

SwissADME								
Compuesto	Lipofilicidad (log P)	Solubilidad en agua (log S)	Accesibilidad sintética	lipinski	Veber	PAINS	Alertas estructurales	Inhibidor de CYP 450
CA-SAH-1	0.37	-3.07	2.32	sí	sí	no	grupo diceto	no
CA-SAH-2	-1.57	-2.28	3.42	sí	Sí	no	no	no
CA-SAH-3	0.68	-3.44	3.51	sí	no	no	no	no
CA-SAH-4	0.29	-3.31	2.2	sí	sí	no	beta cetona	no
CA-SAH-5	1.69	-3.69	3.75	sí	sí	no	no	2C19

\* En donde el log P se calcula con base a un consenso de cinco diferentes métodos; log S se obtiene de un consenso de tres métodos; la accesibilidad sintética se valora por medio de un modelo de regresión lineal que emplea 12,782,590 moléculas de prueba y 40 de validación externa con una  $R^2=0.94$ , donde un valor de 1 es fácil y un valor de 10 muy difícil de sintetizar y la inhibición de CYP 450 se estima por medio de modelos QSAR (Daina y Michielin, 2017).

**Tabla 5. Propiedades toxicológicas y metabolitos más probables calculadas por medio de la aplicación Toxtree.\***

Toxtree				
Compuesto	Carcinogénesis	Mutagénesis (prueba Ames)	Alertas estructurales	Reacción para metabolitos
CA-SAH-1	Poco probable	Poco Probable	Amina primaria aromática	N-hidroxilación
CA-SAH-2	Poco probable	Poco Probable	Amina primaria aromática	N-desalquilación
CA-SAH-3	Poco probable	Poco Probable	Amina primaria aromática	N-hidroxilación
CA-SAH-4	Poco probable	Poco Probable	Amina primaria aromática	N-hidroxilación
CA-SAH-5	Poco probable	Poco Probable	Amina primaria aromática	S-oxidación

\* Estimaciones realizadas por medio de algoritmos de árboles de decisión (Patlewicz et al., 2008).

En general, para las cinco moléculas propuestas con base en los cálculos realizados de los descriptores fisicoquímicos log P y log S (**Tablas 3 y 4**), la permeabilidad de éstos a través de las membranas del epitelio intestinal podría llevarse a cabo eficientemente, ocasionando un transporte transcelular y con ello un buen proceso de absorción. De manera semejante, la solubilidad en agua calculada para todas las moléculas es buena, favoreciendo el proceso de disolución.

De acuerdo con los resultados de la plataforma SwissADME (**Tabla 4**) sintetizar las moléculas candidatas no resultaría inaccesible, ya que éstas fueron valoradas en un rango de 2.2 a 3.75, en una escala de 10, siendo este valor el de mayor complejidad sintética.

En cuanto a las propiedades químico farmacéuticas (**Tabla 4**) todas las moléculas, a excepción del compuesto CA-SAH-3, cumplen los criterios de las reglas empíricas de Lipinski y Veber, en donde CA-SAH-3 sólo sobrepasa ligeramente el límite de 140 Å<sup>2</sup>

correspondiente al área superficial polar topológica (TPSA) con un valor de 155.5 Å<sup>2</sup>. Sin embargo, en ninguno de los compuestos propuestos se detectó la presencia de compuestos que aparentan mostrarse como activos en ensayos HTS, es decir, compuestos falsos positivos (por sus siglas en inglés PAINS, *Pan-assay interference compounds*).

En el servidor admeSAR se evaluó que los cinco compuestos propuestos no presentarían efectos tóxicos, tales como carcinogénesis y mutagénesis (**Tabla 3**). Sin embargo, en *Toxtree* se determinó que en dichas moléculas habría una baja probabilidad de presentarse eventos tóxicos (**Tabla 5**), debido a que se detectó como alerta estructural una amina primaria aromática, perteneciente a la primera sección de todas las moléculas candidatas. Sólo hasta la evaluación experimental de estas moléculas se determinaría la veracidad de esta estimación, ya que al ser análogos de SAH se esperaría semejanza en las propiedades toxicológicas de SAM, el cual se comercializa en E.E.U.U. como suplemento alimenticio y como fármaco para el tratamiento de la depresión, osteoartritis y enfermedades hepáticas sin efectos tóxicos graves (Hardy et al., 2002). En SwissADME se encontraron como alertas estructurales a los grupos diceto y beta ceto, y por tanto como posibles grupos toxicofóros de los compuestos CA-SAH-1 y CA-SAH-3, respectivamente (**Tabla 4**). Por su parte, con admeSAR se predijo a CA-SAH-3 como posible inhibidor de CYP 2C9 y CYP 2C19, mientras que el compuesto CA-SAH-5 se predijo como inhibidor de CYP 2C19.

### **7.2.3 Material educativo y de difusión para la aplicación del diseño *de novo* y dinámica molecular**

#### **7.2.3.1 Diseño *de novo* basado en fragmentos y dinámica molecular (Segunda edición manual de Quimioinformática)**

Como parte de la aportación social hacia el ámbito educativo, a partir del conocimiento adquirido en este trabajo, se elaboró la segunda edición de un manual de Quimioinformática (Saldívar, 2017). Se incorporaron dos nuevos capítulos en los cuales, de manera didáctica y a forma de tutorial, se introduce al lector en la implementación del diseño *de novo* y dinámica molecular, respectivamente, usando el programa computacional MOE 2018.1. Esto se hizo porque estos métodos

computacionales han mostrado ser de utilidad para descubrimiento y desarrollo de nuevos fármacos (Merk et al., 2018) y no habían sido abordados en la primera edición del manual.

Cabe destacar, que al final de cada capítulo se incluyen ejercicios a modo casos de estudio. En ellos, el lector puede practicar y poner a prueba los conocimientos adquiridos, además de lecturas complementarias donde se puede profundizar más sobre los temas abordados. También se desarrolló un glosario, en el cual el lector puede aclarar dudas conceptuales, ya que el manual al tratar temas poco estudiados en México, de primera instancia, el contenido de este podría resultar difícil para su íntegra comprensión. Además, el glosario que se redactó puede ser empleado como material de estudio complementario por alumnos de licenciatura de la Facultad de Química que cursan la materia de Química Farmacéutica. Resulta importante recalcar que, hasta el momento, no existe material educativo de este tipo escrito en español. De hecho, la disponibilidad de manuales enfocados a la enseñanza de la quimioinformática es escasa en otros idiomas, incluyendo el idioma inglés. Por tanto, de esta manera se evidencia la relevancia de la aportación.

#### **7.2.3.2 Diseño *de novo* para el desarrollo de inhibidores de DNMT enfocado a productos naturales (coautoría en artículo de revisión)**

La divulgación científica es una actividad que resulta crucial para el desarrollo científico y tecnológico que culmina en el bienestar general de la sociedad. A través de ésta, se da el medio por el cual la comunidad científica puede compartir los conocimientos recién generados, entrelazar lazos con diferentes áreas y generar nuevo conocimiento.

Con base en la premisa anterior, se participó en un proyecto de divulgación científica. En éste se publicó un artículo de revisión como coautor en la revista *Frontiers in Pharmacology*. En este trabajo se aborda cómo puede implementarse el diseño *de novo* para el desarrollo de inhibidores de DNMT enfocado a productos naturales.

## 8. Conclusiones

- A través de la aplicación del diseño molecular *de novo* (DMDN) se obtuvieron 511 nuevas entidades químicas (*NME*) de inhibidores de DNA metiltransferasa no nucleósidos (I-DNMT-NN) dirigidas al sitio catalítico.
- Se diseñaron 149 *NME* de I-DNMT-NN para la isoforma DNMT3A.
- Se diseñaron 362 *NME* de I-DNMT-NN para la isoforma DNMT1.
- Recursos quimioinformáticos adicionales al DMDN, tales como acoplamiento molecular, visualización del espacio químico, filtrado y curado de bases de datos, así como uso de servidores externos para evaluar propiedades ADME/T, complementaron los resultados obtenidos.
- Se generó material educativo y tutorial con la adición de dos capítulos enfocados en diseño *de novo* y dinámica molecular, en la segunda edición de un manual de quimioinformática.

## 9. Perspectivas

- Realizar estudios de dinámica molecular como método de selección de las moléculas a sintetizar.
- Sintetizar las mejores moléculas propuestas.
- Realizar los ensayos biológicos de las moléculas sintetizadas.
- Realizar un estudio de DMDN-D más extenso en tiempo de cálculo (meses) que permitan generar *NME* de I-DNMT3A más diversas y novedosas.
- Realizar DMDN-D en la isoforma DNMT1.
- Realizar DMDN-D con una biblioteca de fragmentos privilegiados para I-DNMT que provengan de productos naturales e inhibidores sintéticos bioactivos previamente identificados.

## 10. Referencias

- Agrawal, K., Das, V., Vyas, P., & Hajdúch, M. (2018). Nucleosidic DNA demethylating epigenetic drugs – A comprehensive review from discovery to clinic. *Pharmacol. Therapeut.*, 188, 45–79.
- Auclair, G., & Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*, 94, 2202–2211.
- Bajorath, J. (2018). Improving the utility of molecular scaffolds for medicinal and computational chemistry. *Future Med. Chem.* 10, 1645–1648.
- Bennett, R. L., & Licht, J. D. (2018). Targeting Epigenetics in Cancer. *Annu. Rev. pharmacol.*, 58, 187–207.
- Bernardini, S., Tiezzi, A., Laghezza Masci, V., & Ovidi, E. (2018). Natural products for human health: an historical overview of the drug discovery approaches. *Nat. Prod. Res.*, 32, 1926–1950.
- Best, J. D., & Carey, N. (2010). Epigenetic therapies for non-oncology indications. *Drug Discovery Today*, 15, 1008–1014.
- Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular signals of epigenetic states. *Science*, 330, 612–616.
- Brown, N., Cambuzzi, J., Cox, P. J., Davies, M., Dunbar, J., Plumbley, D., Sheppard, D. W. (2018). *Big Data in Drug Discovery. Progress in Medicinal Chemistry (1st ed., Vol. 57)*. Elsevier B.V.
- Brueckner, B., & Lyko, F. (2004). DNA methyltransferase inhibitors: old and new drugs for an epigenetic cancer therapy, *Trends Pharmacol. Sci.* 25, 551-554.
- Bunnage, M. E. (2011). Getting pharmaceutical R&D back on target. *Nat. Chem. Biol.*, 7, 335–339.
- Chen, H., Kogej, T., & Engkvist, O. (2018). Cheminformatics in Drug Discovery, an Industrial Perspective. *Mol. Inf.*, 37, 1–7.
- Cheng et al. (2011). Insights into molecular basis of cytochrome P450 inhibitory promiscuity of compounds. *J. Chem. Inf. Model.*, 51, 2482–2495.

- Cheng, Li, Zhou, Shen, Wu, Liu, Lee & Tang. (2012) admetSAR: a comprehensive source and free tool for evaluating chemical ADMET properties. *J. Chem. Inf. Model.*, 52, 3099-3105.
- Daina, Michielin & Zoete. (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7:42717.
- Danziger, D. J. & Dean, P. M. (1989). Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces. *Proc. R. Soc. Lond. B* 236, 101–113,
- DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: New estimates of drug development costs. *J. Health Econ.* 22, 151–185.
- Erdmann, A., Arimondo, P. B., & Guianvarc, D. (2016). Structure-Guided Optimization of DNA Methyltransferase Inhibitors. *Epi-Informatics*. Elsevier Inc.
- Erlanson, D.A. (2012) Introduction to fragment-based drug discovery. *Top. Curr. Chem.*, 317, 1–32.
- Erlanson, D.A. et al. (2004) Fragment-based drug discovery. *J. Med. Chem.*, 47, 3463–3482.
- Fernanda Saldívar González, Francisco Hernández Luis, Alfonso Lira Rocha, José Luis Medina Franco. *Manual de Quimioinformática*. Universidad Nacional Autónoma de México, 2017. 1a edición. Ciudad de México, México. ISBN: 978-607-02-9220-0.
- Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., & Jacob, S. T. (2005). 5-Aza-Deoxycytidine Induces Selective Degradation of DNA Methyltransferase 1 by a Proteasomal Pathway That Requires the KEN Box, Bromo-Adjacent Homology Domain, and Nuclear Localization Signal. *Mol. Cell. Biol.*, 25, 4727–4741.
- Gillet, V.J., Myatt, G., Zsoldos, Z. et al. (1995). *Perspectives in Drug Discovery and Design* 3, 34.
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Bestor, T. H. (2006). Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science*, 311, 395–398.

- Granda, J. M., Donina, L., Dragone, V., Long, D. L., & Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature*, 559, 377–381.
- Hajduk, P. J., & Greer, J. (2007). A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.*, 6, 211–219.
- Hansen, K., Mika, S., Schroeter, T., Sutter, A., Laak, A. Ter, Thomas, S. H., Müller, K. R. (2009). Benchmark data set for in silico prediction of Ames mutagenicity. *J. Chem. Inf. Model.*, 49, 2077–2081.
- Hardy et al. (2002). S-Adenosyl-L-Methionine for Treatment of Depression, Osteoarthritis, and Liver Disease. Agency for Healthcare Research and Quality (Dept Health and Human Services). Sitio web: <https://web.archive.org/web/20090927133628/http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1a.chapter.2159>
- Hartenfeller M., Schneider G. (2010) De Novo Drug Design. In: Bajorath J. (eds) Chemoinformatics and Computational Chemical Biology. *Methods in Molecular Biology (Methods and Protocols)*, vol 672. Humana Press, Totowa, NJ
- Hartenfeller, M., & Schneider, G. (2011). Enabling future drug discovery by de novo design, *Comput. Mol. Sci.* 1, 742–759.
- Herman, J. G., & Baylin, S. B. (2003). Gene Silencing in Cancer in Association with Promoter Hypermethylation. *N. Engl. J. Med.*, 349, 2042–2054.
- Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, 449, 248–251.
- Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, 12, 206–222.
- Keating, G. M. (2009). Azacitidine Syndromes / Acute Myeloid Leukaemia. *Drugs*, 69, 2501–2518.
- Khanna, I. (2012). Drug discovery in pharmaceutical industry: Productivity challenges and trends. *Drug Discovery Today*, 17, 1088–1102.

- Krueger, B. A., Dietrich, A., Baringhaus, K., & Schneider, G. (2009). Scaffold-Hopping Potential of Fragment-Based De Novo Design: The Chances and Limits of Variation, *Comb. Chem. High Throughput screening* 12, 383–396.
- Lagunin, A., Filimonov, D., Zakharov, A., Xie, W., Huang, Y., Zhu, F., Poroikov, V. (2009). Computer-aided prediction of rodent carcinogenicity by PASS and CISOC-PSCT. *QSAR Comb Sci.*, 28, 806–810.
- Li, J. W., & Vederas, J. C. (2009). Drug Discovery and Natural Products: Science, 325(5935), 161–165.
- Luechtefeld, T., Marsh, D., Rowlands, C., & Hartung, T. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol. Sci.*, 165, 198–212.
- Lyko, F. (2018). The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.*, 19, 81–92.
- Martens, U. M. (2010). *Small Molecules in Oncology*. Springer (Vol. 184).
- Matthew A Sellwood, Mohamed Ahmed, M. H. S. & N. B. (2018). Artificial Intelligence in Drug Discovery and Development. *J. Pharmacovigilance*, 06, 2025–2028.
- Merk, D., Friedrich, L., Grisoni, F., & Schneider, G. (2018). De Novo Design of Bioactive Small Molecules by Artificial Intelligence, *Mol. Inf.* 1700153, 3–6.
- Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11, 29–34.
- Miyao, T., Reker, D., Schneider, P., Funatsu, K., and Schneider, G. (2015). Chemography of natural product space. *Planta Med.* 81, 429–435.
- Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, 3, 651–657.
- Mullard, A. (2015). 2014 FDA drug approvals. *Nat. Rev. Drug Discov.*, 14, 77–81.
- Naveja, J. J., & Medina-franco, J. L. (2018). Insights from pharmacological similarity of epigenetic targets in epipolypharmacology. *Drug Discovery Today*, 23, 141–150.



- Neema Bisht and B. K. Singh. (2018). Role Of Computer-aided Drug Design in drug development and discovery, *Int J Pharm Sci Res.*, 9, 1405–1415.
- Ning, B., Li, W., Zhao, W., & Wang, R. (2015). Targeting epigenetic regulations in cancer *Acta Biochim. Biophys. Sin*, 48, 97–109.
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development, *Cell*. 99, 247–257.
- Patlewicz G, Jeliaskova N, Safford RJ, Worth AP, Aleksiev B. (2008) An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ Res.*, 19, 495-524.
- Prieto-Martínez, F. D., Arciniega, M., & Medina-Franco, J. L. (2018). Acoplamiento Molecular: Avances Recientes y Retos. *TIP Revista Especializada En Ciencias Químico-Biológicas*, 21, 65–87.
- Ramachandran, S., lent, J., Göttgen, E. L., Krieg, A. J., & Hammond, E. M. (2015). Epigenetic therapy for solid tumors: Highlighting the impact of tumor hypoxia. *Genes*, 6, 935–956.
- Renaissance, M. (2016). Biologic Variability to Drug Response: Sex Differences in Clinical Trials. *Duke-Margolis Center Meeting Summary*. Duke University.
- Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016). Counting on natural products for drug design. *Nat. Chem.* 8, 531–541.
- Ruddigkeit, L., van Deursen, R., Blum, L.C., and Reymond, J.L. (2012) Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.*, 52, 2864–2875.
- Saba, H. I., & Wijermans, P. W. (2005). Decitabine in myelodysplastic syndromes. *Semin. Hematol.*, 42, 23-31.
- Saldívar-gonzález, F. I., Gómez-garcía, A., Chávez-Ponce de León, D. E., Sánchez-cruz, N., Ruiz-rios, J., Pilon-jiménez, B. A., & Medina-franco, J. L. (2018). Inhibitors of DNA Methyltransferases From Natural Sources: A Computational Perspective. *Front. Pharmacol.*, 9, 1–10.

- Saldívar-González, F., Prieto-Martínez, F. D., & Medina-Franco, J. L. (2017). Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química*, 28, 51–58.
- Schneider, G. (2014). *De novo Molecular Design*. First edition. Wiley-VCH.
- Schneider, G., & Fechner, U. (2005). Computer-Based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.*, 4, 649–663.
- Schneider, G., & Lloyd, D. (2013). De novo design – hop(p)ing against hope. *Drug Discovery Today: Technologies*, 10, 453–460.
- Sinha, S., & Vohora, D. (2017). *Drug Discovery and Development: An Overview*. Pharmaceutical Medicine and Translational Clinical Research. Elsevier Inc.
- Sliwoski, Gregory. Kothiwale, Sandeepkumar. Meiler, Jens. Lowe Jr, E. W. (2014). Computational Methods in Drug Discovery. *Pharmacolo. Rev.*, 66, 334–395.
- Sneader, W. (2006). *Drug Discovery: A History*. Drug Discovery: A History.
- Stresemann, C., & Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine *Int. J. Cancer*, 123, 8–13.
- Suryanarayanan, V., Panwar, U., & Chandra, I. (2018). De Novo Design of Ligands Using Computational Methods, *Methods Mol. Biol.* 1762, 71-86.
- Victoria-Acosta, G., Vazquez-Santillan, K., Jimenez-Hernandez, L., Muñoz-Galindo, L., Maldonado, V., Martinez-Ruiz, G. U., & Melendez-Zajgla, J. (2015). Epigenetic silencing of the XAF1 gene is mediated by the loss of CTCF binding. *Sci. Rep.*, 5, 1–15.
- Wang, J., Krudy, G., Hou, T., Zhang, W., Holland, G., & Xu, X. (2007). Development of reliable aqueous solubility models and their application in druglike analysis. *J. Chem. Inf. Model.*, 47, 1395–1404.
- Wermuth, C., Aldous, D., Raboisson, P., & Rognan, D. (2015). *The practice of medicinal chemistry*.
- Xu, J., & Hagler, A. (2002). Chemoinformatics and drug discovery. *Molecules*, 7, 566–600.

- Xu, P., Hu, G., Luo, C., & Liang, Z. (2016). DNA methyltransferase inhibitors: an updated patent review (2012-2015), *Expert. Opin. Ther. Pat.* 3776, 1017-30.
- Yuan, Y., Pei, J., & Lai, L. (2011). LigBuilder 2: A practical de novo drug design approach. *J. Chem. Inf. Model.*, 51, 1083–1091.
- Zhu, L. M., Shi, D. M., Dai, Q., Cheng, X. J., Yao, W. Y., Sun, P. H., Tu, S. P. (2014). Tumor suppressor XAF1 induces apoptosis, inhibits angiogenesis and inhibits tumor growth in hepatocellular carcinoma. *Oncotarget*, 5, 5403-15.

## Anexos

### A.1 Glosario

1. **Acantilados de actividad (*Activity Cliff*):** Se refiere a aquellos compuestos que poseen similitud estructural entre sí, pero presentan diferente potencia en la actividad biológica hacia una diana biológica.
2. **Acoplamiento molecular automatizado (*Docking*):** Método computacional que busca el óptimo modo de unión tridimensional de un ligando con su respectiva diana biológica.
3. **Análisis de componentes principales (PCA):** Es un método de reducción de variables que transforma un conjunto variables originales a otro conjunto de variables ortogonales (no correlacionales entre sí), mediante el uso de una matriz de correlación.
4. **Archivo PDB:** Archivo de texto que almacena las coordenadas atómicas de una macromolécula, generalmente una proteína o ácido nucleico junto con el solvente y iones del medio.
5. **Cabeza de serie (*Lead*):** Compuesto que durante el proceso de descubrimiento de fármacos ha demostrado ser activo, y que además cumple con ser selectivo, novedoso, accesible para su obtención y se pueden obtener análogos de él fácilmente.
6. **Cambios de actividad (*Activity Switch*):** Son compuestos estructuralmente similares, pero que presentan una acción opuesta.

7. **Campo de fuerza:** Conjunto de funciones matemáticas que asocia los parámetros necesarios de mecánica o dinámica molecular para realizar los cálculos de conformación, flexibilidad e interacciones moleculares.

8. **Compuestos Scaffold hopping:** Del inglés, se refiere a compuestos que presenta un mismo tipo de actividad hacia una cierta diana biológica, pero que poseen diferente núcleo base o, en general, estructuras químicas muy diferentes.

9. **Coefficiente de Tanimoto:** Relación matemática para medir la similitud estructural  $T_{ij}$  de las moléculas  $i$  y  $j$  mediante la siguiente expresión:

$$S_{Tan} = \frac{\sum x_{1j}x_{2j}}{\sum x_{1j}^2 + \sum x_{2j}^2 - \sum x_{1j}x_{2j}}$$

10. **Cribado virtual (*Virtual screening*):** Es una técnica que consiste en realizar un filtrado computacional de una biblioteca de compuestos en donde se pronóstica la capacidad de unión de una molécula a una determinada diana biológica, y por ende su probable bioactividad.

11. **Curado (de compuestos):** Preparación de las estructuras químicas a analizar, que consiste en precisar de manera correcta las estructuras químicas a emplear, asignar la geometría, cargas y quiralidad adecuadas de los compuestos de interés. Los compuestos pueden proceder de bases de datos públicas, mismos que deben depurarse, ya que suelen tener errores. La alternativa es usar compuestos identificados por un grupo de investigación o hacer la propuesta teórica de nuevas estructuras.

12. **Base de datos:** Es una colección de datos relacionados entre sí que se encuentran almacenados sistemáticamente. En este sentido en el DIFAC se refiere a una compilación de estructuras químicas (pueden estar en diversos formatos como: sdf, mol2 etc.) enfocadas a una cierta diana biológica, que sirve como punto de partida para realizar estudios de acoplamiento molecular o cribado virtual.

13. **Big data:** Del inglés, es un conjunto de datos enormes, en donde al realizar un análisis computacional de éste, se pueden obtener patrones, tendencias o información relevante relativa a un cierto tópico.

14. **Descriptor molecular:** Es una propiedad intrínseca cualitativa o cuantitativa de una molécula que representa la propiedad de una molécula. Se representa con un número.

15. **Diana biológica:** Estructural macromolecular endógena a la que es capaz de unirse un ligando para ocasionar una respuesta fisiológica dentro del organismo.

16. **DIFAC (Diseño de fármacos asistido por computadora):** Es un enfoque de la química farmacéutica que tiene como objetivo: Identificar, diseñar, optimizar y seleccionar compuestos bioactivos, utilizando como herramientas a la Quimioinformática, Química computacional, Química teórica y Modelado molecular.

17. **Dinámica molecular:** Es una técnica computacional que permite simular el movimiento e interacciones de las moléculas en un sistema dado.

18. **Diseño de novo:** Es un método computacional que emplea diversos algoritmos con el objeto de generar nuevos compuestos bioactivos que posean un determinado perfil de propiedades deseables; los cuales se pueden obtener mediante la construcción creciente de un ligando dentro un modelo de receptor generado a partir información cristalográfica o proveniente de RMN.

19. **Diseño basado en el ligando:** Es el enfoque de diseño de fármacos en el que se utilizan las propiedades estructurales y fisicoquímicas de los ligandos conocidos para generar nuevos compuestos que puedan tener la misma actividad biológica.

20. **Diseño basado en el receptor:** Enfoque de diseño de fármacos donde es necesario conocer la información estructural tridimensional de la diana biológica para generar posibles compuestos bioactivos.

21. **Espacio químico:** Es la representación visual de la distribución de un conjunto de moléculas, con el objeto de observar su similitud molecular, mediante la implementación de coordenadas libres basadas en descriptores moleculares.

22. **Farmacóforo:** Es la concepción abstracta del conjunto de características electrónicas y estéricas que son responsables de las óptimas interacciones supramoleculares del ligando con su receptor.

23. **Fármaco:** Sustancia pura, químicamente definida que puede ser natural, sintética o biotecnológica, que se caracterice por tener alguna actividad terapéutica y que se identifique por sus propiedades físicas, química o biológicas; además no debe presentarse en una forma farmacéutica definida.

24. **Filtrado (de compuestos):** Consiste en la selección de compuestos que mejor cumplan con los criterios de interés.

25. **Huellas digitales moleculares (*Molecular fingerprints*):** Es una descripción computacional de la estructura de un compuesto químico, con base en la ausencia o presencia de ciertas subestructuras codificadas por bits (dígitos binarios representados por los números 0 y 1).

26. **Lead-like:** Término en inglés que se le da a las moléculas pequeñas, que se caracterizan por cumplir criterios más reales y más estrictos a las reglas de Lipinski.

27. **Ligando:** Compuesto químico, generalmente de pequeño peso molecular, puede ser endógeno o exógeno que tiene afinidad de unión con una determinada diana biológica.

28. **Minería de datos (*Data Mining*):** Se refiere a la extracción de datos útiles provenientes de bases de datos diversas.

29. **Minimización energética:** Operación en la que se optimiza la geometría de una molécula.

30. **Modelado molecular:** Es la representación o mimetización aproximada del comportamiento molecular, considerando sus características tridimensionales.

31. **Núcleo base (*Scaffold*):** Es la subestructura principal de un compuesto químico, a partir de la cual se extienden el resto de los grupos funcionales o subestructuras secundarias.

32. **Protein Data Bank (PDB):** Base de datos de acceso libre que contiene una vasta cantidad de información estructural tridimensional de proteínas, ácidos nucleicos y complejos macromoleculares.

33. **Prototipo (*hit*):** Compuesto puro que ha sido identificado durante el proceso de descubrimiento de fármacos, con presumible actividad biológica como mínimo a una concentración de 1  $\mu\text{M}$ .

34. **Puntaje (*Scoring*):** Es una estimación de la afinidad de unión de un ligando a su diana biológica, basada en el complejo formado y que generalmente se expresa en términos energéticos ( $\Delta\text{G}$  de unión).

35. **Química Farmacéutica:** Ciencia interdisciplinaria que se encarga del diseño, descubrimiento, identificación, desarrollo y reposicionamiento de compuestos bioactivos que posteriormente pueden ser utilizados para la invención de nuevos medicamentos.

36. **Quimioinformática:** Ciencia que se encarga del almacenamiento, organización, búsqueda, obtención, manejo, visualización y análisis de información química.

37. **QSAR:** Método cuantitativo por el cual se busca establecer una relación matemática entre las estructuras químicas y la propiedad de interés (actividad biológica por lo general). Frecuentemente se asume que existe una relación directa entre la similitud de la actividad y la similitud estructural de los compuestos.

38. **Reglas de Lipinski:** Es un conjunto de cuatro reglas empíricas cualitativas que permiten estimar si un compuesto pudiese tener buenas propiedades farmacocinéticas (administración vía oral) que incluye los procesos ADME. 1)  $\log P$  menor a 5; 2) Peso molecular menor a 500; 3) No contener más de 5 donadores de puente de hidrógeno y 4) No contener más de 10 aceptores de puentes de hidrógeno.

39. **Reposicionamiento (de fármacos):** Buscar un uso terapéutico alternativo a los fármacos aprobados para uso clínico o en fases avanzadas del desarrollo clínico. El fármaco puede tener dosis diferentes según la enfermedad para la cual esté prescrito.

40. **Restricciones primarias del receptor:** Toda la información relacionada con las interacciones del complejo ligando-receptor, se asocia con la afinidad de unión de un ligando a una cierta diana biológica.

41. **Restricciones secundarias del receptor:** Propiedades esenciales del ligando que lo hace adecuadamente farmacológico, diferente de la afinidad de unión con la diana biológica (propiedades ADME/T, accesibilidad sintética y selectividad).

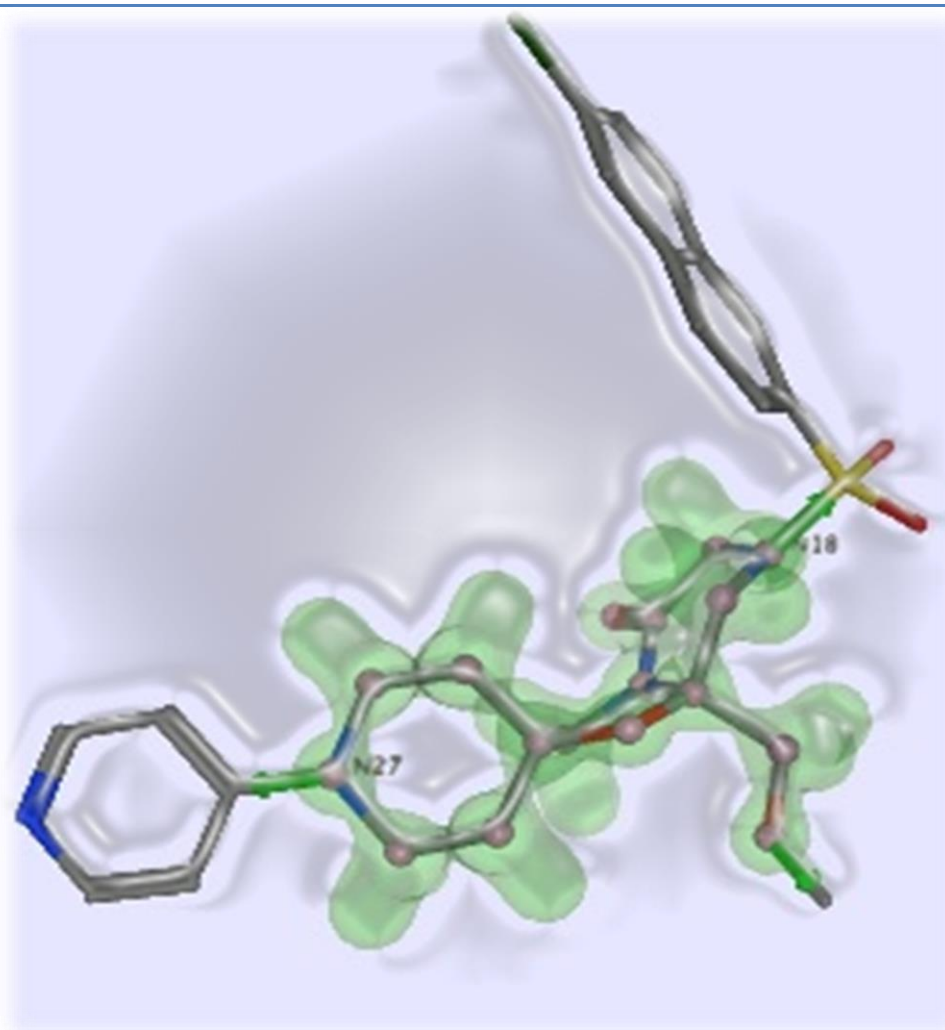
42. **SMILES (*simplified molecular-input line-entry system*):** Es una forma de representación computacional de estructuras químicas en la que se utiliza caracteres del código ACSII, para describir la conectividad, tipo de enlace, grupos funcionales y estereoquímica de la molécula a representar.

43. **Similitud estructural:** Grado en el que dos moléculas se asemejan entre sí, el cual se calcula a partir de sus características 2D o 3D.

44. **Sitio de unión:** Es la fracción estructural de la macromolécula, donde se lleva a cabo la interacción con el ligando que resulta en la respuesta biológica.



## Diseño *de novo* basado en fragmentos



MOE 2018.1

# Reemplazo del núcleo base (*Scaffold replacement*)

## Objetivo

- ❖ Generar nuevas moléculas con núcleos bases diferentes que posean propiedades fisicoquímicas similares y que mantengan una respuesta biológica similar que las moléculas originales o plantilla (*scaffold hopping*).

## Objetivo específico

- ❖ Reforzar las habilidades aprendidas en los capítulos anteriores incluyendo, la preparación de las estructuras de la diana biológica y ligando.

## Introducción

Las bases del diseño molecular *de novo* se asentaron desde hace más de 15 años (Schneider y Fechner, 2005). Esta metodología originalmente tenía un enfoque de diseño basado en el receptor y solo incluía criterios estéricos como método de búsqueda y generación de fragmentos. Actualmente se tienen nuevas ideas y enfoques que no sólo se limitan a la estructura del receptor, sino que también se puede utilizar la información estructural de ligandos activos ya conocidos, que sirvan como plantillas moleculares para el diseño de nuevas moléculas.

El diseño *de novo* con este enfoque ha adquirido una perspectiva diferente, en la cual mediante la consideración simultánea de múltiples propiedades en párelo (Schneider, 2013) abre la posibilidad de generar compuestos novedosos que tengan características tales como, ser bioactivos, perfil farmacocinético favorable y ser sintéticamente accesibles.

## Recursos materiales

La estructura de la enzima PDB ID: 1IOE utilizada para este ejemplo se obtuvo de la base de datos de *Protein Data Bank* (PDB). Ésta estructura es una hidrolasa que corresponde al factor de coagulación Xa. Funciona como activadora de la trombina, cuya inhibición selectiva conllevaría a una supresión de la cascada de coagulación y por tanto sería de interés terapéutico la generación de nuevos agentes anticoagulantes. Al igual que la preparación de las estructuras de la diana biológica y el ligando, la generación de las molecular se realizó utilizando el programa MOE.2018.1.

## Procedimiento (utilizando MOE)

1. Obtenga la estructura de la proteína en formato PDB.

- a) Ingrese a la página de *Protein Data Bank* (ver URL en la **Tabla 5.1**) y en el buscador escriba el nombre de la proteína. En su defecto escriba el código PDB.
- b) Descargue la estructura en la opción **Download Files** (para este ejemplo se trabajó con la proteína PDB ID: 1IOE).

2. Abrir la estructura de la proteína en MOE.

**File | Open |** seleccionar macromolécula en formato PDB| **OK**

3. Preparación de proteína.

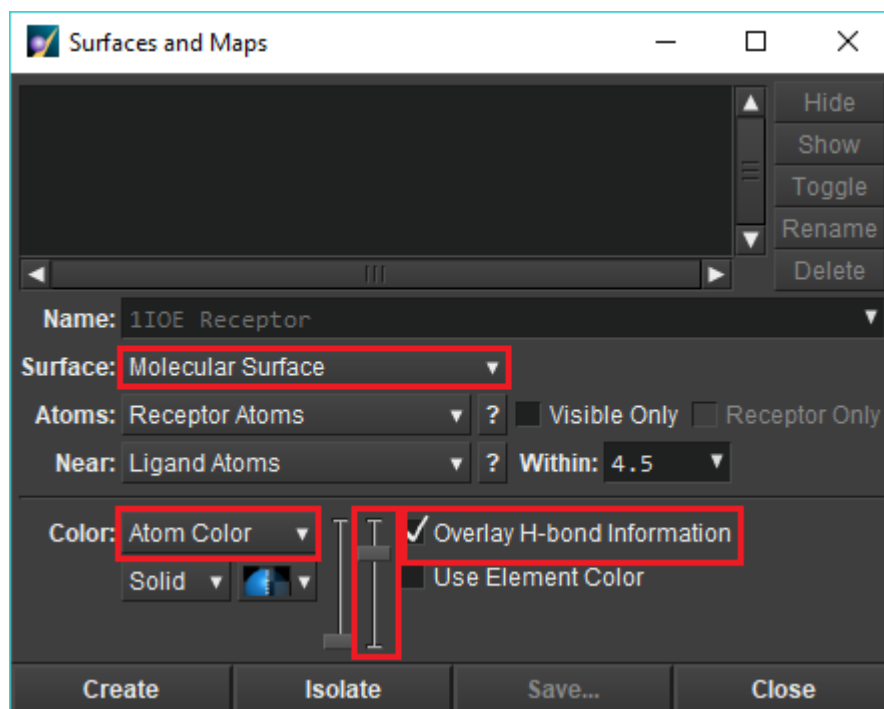
La preparación de la diana biológica se puede realizar de dos formas distintas:

- a) En la barra de herramientas superior hacer click en **Compute | Prepare | Protonate 3D** escoja los parámetros fisiológicos que a usted convenga o en su defecto deje los parámetros predeterminados |**OK**.

b) Seleccione en la barra derecha la opción **Quickprep (en versiones anteriores de MOE puede aparecer como LigX)** y asegúrese de utilizar la conformación protonada |OK. (Para más detalle consultar la sección de acoplamiento molecular **página 77**).

4. Seleccione el sitio de unión.

Haga click en | **RHS** | **Surface** | **Surface and Maps** elija **Color: Atom color**, **Surface: Molecular Surface**, **Atoms: Receptor Atoms**, **TB: 75%** | **Create**.



**Nota:** Activar la función **Overlay H-bond information**, ya que ésta permite visualizar las regiones donde la geometría sería óptima para generar puentes de hidrógeno.

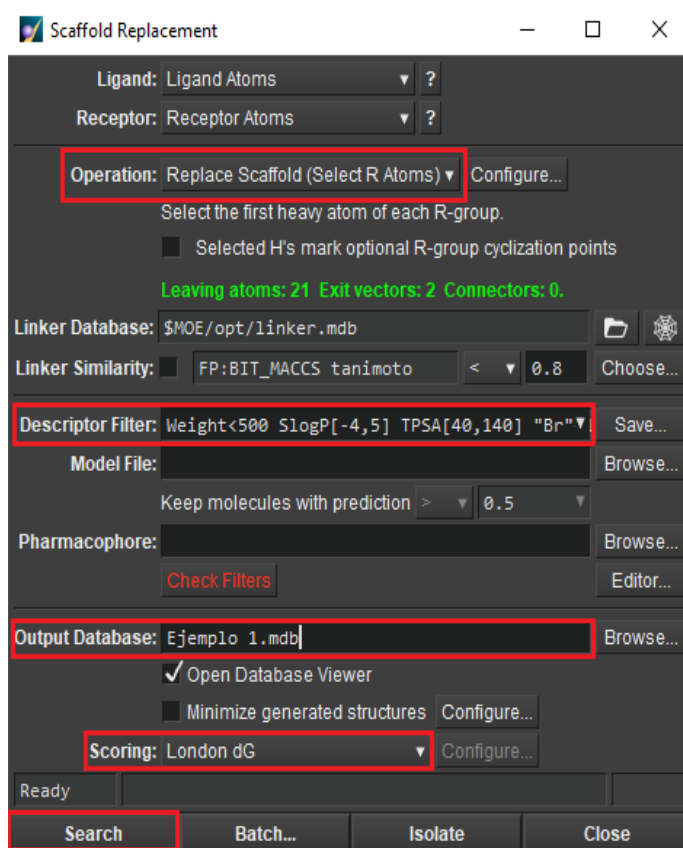
5. Análisis del sitio de unión.

Es importante saber qué tipo de interacciones se están dando en el sitio de unión, ya que una pérdida de éstas podría conllevar a una disminución o incluso la pérdida de la actividad biológica en los nuevos compuestos. Para este fin es útil generar un mapa electrostático que nos indique las regiones hidrofóbicas, aceptoras y donadoras de puentes de hidrógeno (HBA y HBD) dentro del sitio de unión, a modo de identificar las porciones de molécula plantilla que se superponen en dichas regiones y de esta forma se conserven. En este ejemplo hay dos grupos hidrofóbicos y un grupo aceptor de puente de hidrógeno.

Para esto seleccione | **RHS** | **Surface** | **Surfaces and Maps** | **Electrostatic Map** | **Create**.



A continuación, en el recuadro de la función *Scaffold replacement* deberemos de seleccionar los parámetros de filtrado para la búsqueda y selección de fragmentos que podrán ser ensamblados a las nuevas moléculas. Elija **Operation: Replace Scaffold (Select R atoms)**; **Descriptor Filter: Weight<500 SlogP [-4,5] TPSA[40,140] "Br"<1**; **Scoring: London dG**; **Output database: Ejemplo 1.mdb**; **Search** |



**Notas:** 1) Para tener una búsqueda más refinada y precisa es posible añadir un modelo farmacofórico previamente generado, así como habilitar la sección **Linker Similarity**, la cual mediante huellas digitales (*fingerprints*) MACCS keys buscará fragmentos que posean una similitud estructural menor a 0.8 usando la métrica del coeficiente de Tanimoto. 2) Cabe mencionar que los descriptores que se recomiendan utilizar corresponden en ciertos puntos con las reglas de Lipinski y Veber, además de la eliminación átomos toxicofóros como el bromo; 3) En la sección de **Operation** es posible escoger otro método de operación distinta para generar las estructuras, tales como *Medchem transformation*, *BREED*, *Combinatorial builder*, entre otros.

7. Obtención y análisis de resultados.

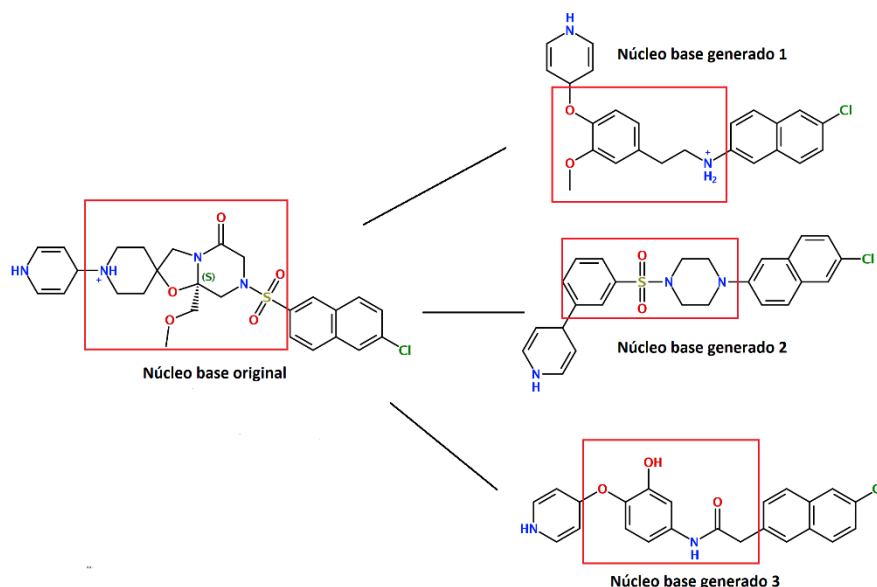
Al finalizar el cálculo se generará una base de datos automáticamente en la cual se podrán observar:

**Tabla 7.1** Parametros de la base de datos de resultados *Scaffold replacement*.

Parámetro	Descripción
<b>RMSD</b>	Indica la diferencia posicional entre la molécula original y la generada.
<b>AcidBase</b>	Indica la presencia de grupos funcionales ácidos/básicos que pueden cambiar su estado de protonación, cuando esto suceda se establece un número uno.
<b>rsynth</b>	Indica la accesibilidad sintética que tendría la molécula generada. Un número uno representa que la molécula tiene una alta probabilidad de ser sintetizada.
<b>linker</b>	Muestra el fragmento que fue usado para generar la nueva molécula.
<b>Ligand efficiency</b>	Indica la contribución energética en la unión del complejo ligando/receptor por átomo que constituye la molécula.
<b>London dG</b>	Energía de unión del complejo ligando/receptor calculada.
<b>Weight</b>	Masa molecular del compuesto generado.
<b>TPSA</b>	Área polar superficial topológica

En este ejemplo se puede observar que bajo con los parámetros establecidos se construyeron automáticamente a 2388 moléculas.

**Figura 7.1** Ejemplos de cambios en el núcleo base de las moléculas generadas por *Scaffold replacement*.



## Ejercicio

A partir del siguiente archivo PDB describir lo que se pide a continuación:

### **PDB ID: 1O86**

<i>Tipo de diana biológica</i>	
<i>Función fisiológica</i>	
<i>Padecimiento asociado</i>	
<i>Interacciones claves</i>	
<i>Grupos funcionales por preservar</i>	
<i>Número de moléculas generadas</i>	



## Evaluación

De acuerdo con los datos obtenidos de un estudio de reemplazo de núcleo base determine si los valores arrojados son adecuados o inadecuados, colocando en la columna de consideración una la letra A e I respectivamente.

<b>Parámetro</b>	<b>Valor</b>	<b>Justificación</b>
<b>rsynth</b>	0.9234	
<b>SlogP</b>	2.4789	
<b>Weight</b>	1545	
<b>London dG</b>	3.7832	

### **Para saber más:**

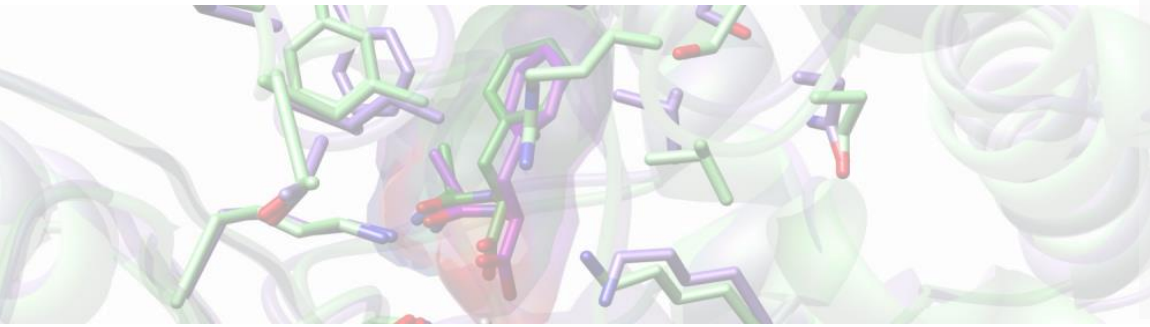
Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8), 649–663.

Schneider, G., & Lloyd, D. (2013). De novo design – hop(p)ing against hope. *Drug Discovery Today: Technologies*, 10, e453–e460.

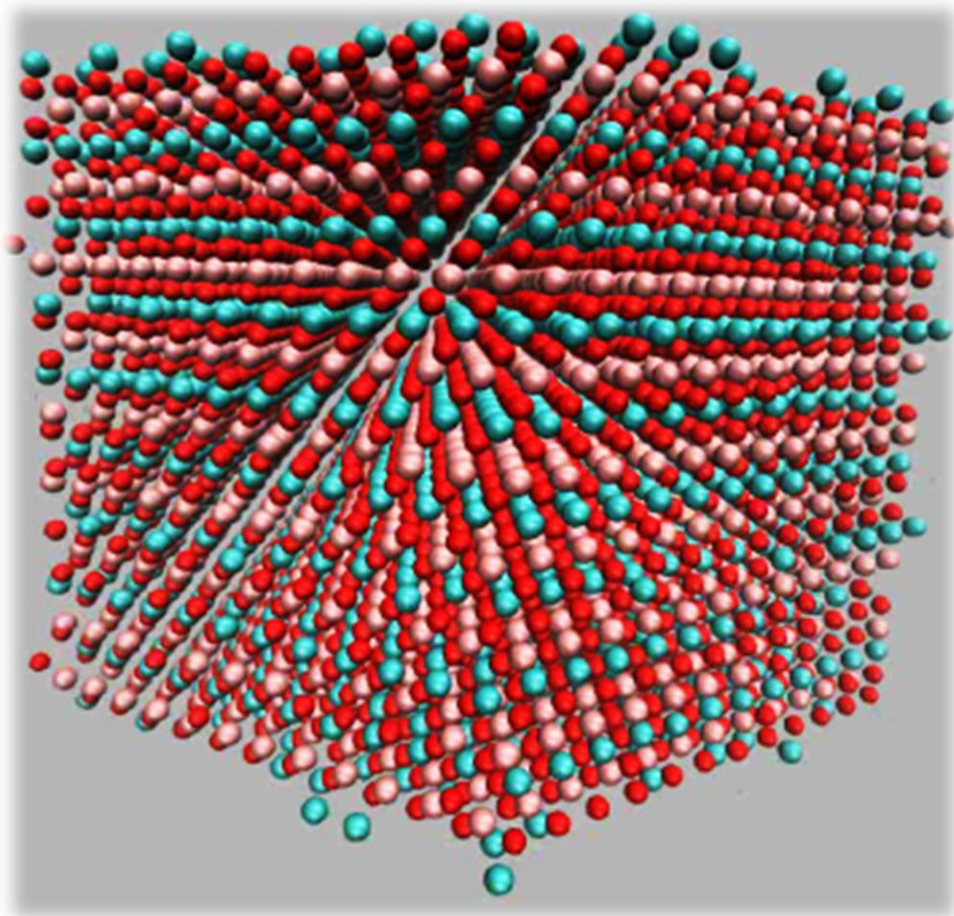
Kawai, K., Nagata, N., & Takahashi, Y. (n.d.). De Novo Design of Drug-Like Molecules by a Fragment-Based Molecular Evolutionary Approach.

Gisbert Schneider, De novo molecular design, Wiley-VCH, primera edición, 2013





## Dinámica molecular



<https://www.cimat.mx/es/SolucionesTecnologicasInnovacion/AulaCIMAT-CIMNE>

# Preparación y simulación de la dinámica molecular

## Objetivo

- ❖ Realizar una simulación de un sistema dinámico con el objeto de identificar las interacciones y su estabilidad en un complejo ligando-receptor.

## Objetivo específico

- ❖ Plantear los pasos básicos para realizar una simulación de dinámica molecular.
- ❖ Complementar y reforzar las habilidades aprendidas en los capítulos 5 y 7, incluyendo el acoplamiento molecular automatizado, así como la preparación de las estructuras de la diana biológica y ligando.

## Introducción

Gracias a la extensa cantidad de información biológica (estructuras tridimensionales y secuencias génicas proteicas) que se encuentra disponibles en bases de datos públicas, tales como *Protein Data Bank* (PDB) o UniprotKB. Es que el diseño de fármacos asistido por computadora (DIFAC), bajo la estrategia de diseño basado en el receptor, ha resultado efectivo en el proceso de identificación de compuestos prototipos (*hits*) (Ganesan et al, 2017).

Unos de los métodos más representativos del diseño basado en el receptor es el acoplamiento molecular automatizado (*docking*). Como se vio en el capítulo 5 (página 74), el acoplamiento molecular consiste en predecir el modo de unión más favorable dentro de un complejo ligando-receptor. Sin embargo, una limitante trascendental de este método es su enfoque estático, donde no se considera los cambios conformacionales del receptor al tiempo del reconocimiento del ligando durante su unión a éste, es decir, no se toma en cuenta el efecto de ajuste inducido. Esto tiene como consecuencia que el acoplamiento molecular se adecue más al paradigma de unión de llave cerradura de Emil Fischer (Sledz y Caflisch, 2018). Por tanto, no se consideran con suficiente exactitud aspectos clave para la predicción de la unión, tales como los factores entrópicos y cinéticos propios del sistema (Fernando 2018).

La dinámica molecular es una herramienta que permite generar modelos del comportamiento molecular de un sistema a lo largo de un intervalo de tiempo: de los milisegundos hasta los femtosegundos. Basándose en la mecánica clásica busca estudiar las interacciones y movimientos de las moléculas considerando las fuerzas enlazantes y no enlazantes dentro del sistema a simular. (De vivo Et al., 2016).

## Recursos materiales

La estructura del receptor de estrógeno con código PDB ID: 1ERR utilizada para este ejemplo se extrajo de *Protein Data Bank* (PDB).

Al igual que la preparación de la estructura del receptor y el ligando (ver capítulo 5 página **74**), la simulación de la dinámica molecular se realizó utilizando el programa MOE.2018.1.

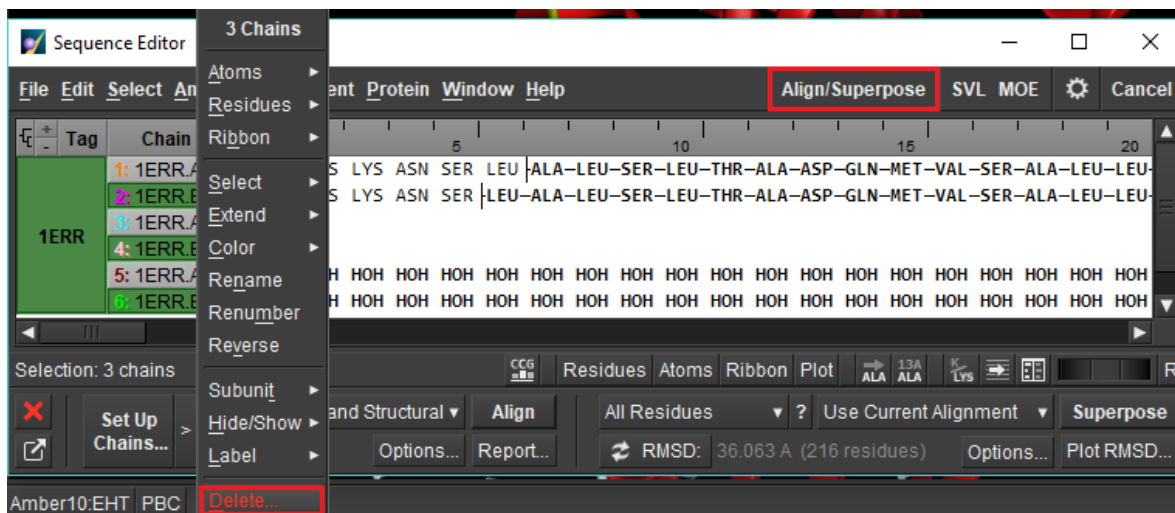
Los receptores de estrógenos se encargan de la regulación metabólica ósea y lipídica en las mujeres. Durante la postmenopausia, el desbalance hormonal aumenta la probabilidad de desarrollar osteoporosis y cáncer de mama. El consumo de raloxifeno como tratamiento preventivo es el más común (Moen y Keating, 2008), ya que funciona como modulador selectivo hacia dichos receptores, debido a que en tejido mamario y uterino actúa como antagonista, mientras que en tejido óseo y graso actúa como agonista (Heringa, 2003). Por tanto, es importante identificar la estabilidad de las interacciones clave dentro del receptor, ya que, de esta manera, sería posible extraer información útil para partir del diseño de nuevas moléculas moduladoras.

**Nota:** Para un tiempo de cálculo óptimo se recomienda contar como mínimo con las siguientes características en hardware: Procesador Intel CORE i5 o AMD A8, 3.3 GHz, 12 GB RAM y un sistema operativo de 64 bits (se recomienda el uso de MOE en Windows sobre Linux).

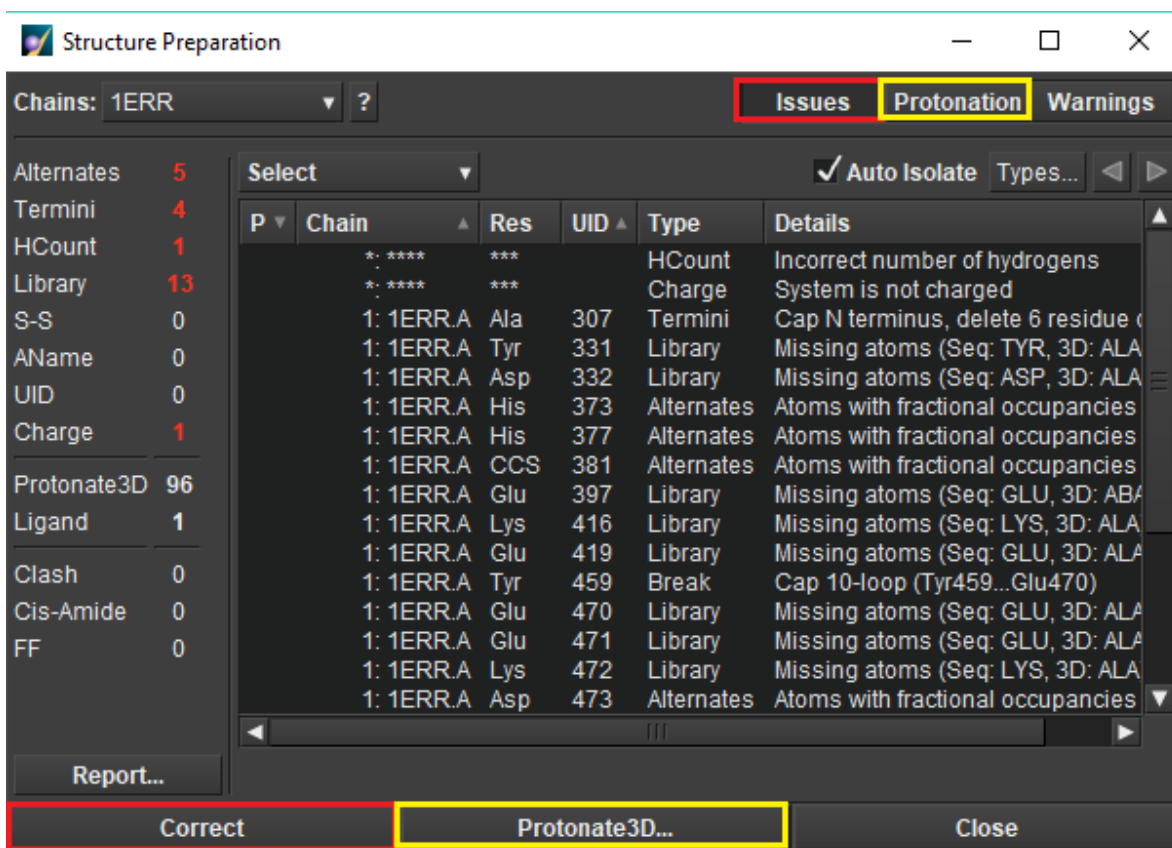
## Procedimiento (utilizando MOE)

1. Obtenga la estructura de la proteína en formato PDB.  
Ingrese al portal web de la base de datos PDB (ver URL en la **Tabla 5.1**) y descargue la estructura del receptor de estrógenos con ID PDB: 1ERR (para más información sobre descargas de estructuras consultar la sección de acoplamiento molecular automatizado o Diseño *de novo* basado en fragmentos en las páginas **74** o **106** respectivamente).
2. Abrir la estructura de la proteína en MOE.  
**File | Open |** seleccionar macromolécula en formato PDB| **OK**
3. Preparación de proteína.  
La preparación de la estructura de la diana biológica y la obtención de los estados de protonación se pueden realizar siguiendo los pasos que se exponen el capítulo 7 (páginas **108-110**). En forma alterna también se puede hacer de la siguiente manera:
  - a) Hacer clic en **SEQ | Align/Superpose** | posteriormente seleccionar los elementos de la cadena B y proceder a eliminarlos haciendo **clic derecho | Delete**

- b) Dar clic en **Compute | Prepare | Structure preparation | Issues | Correct** De esta forma MOE corregirá automáticamente problemas estructurales.



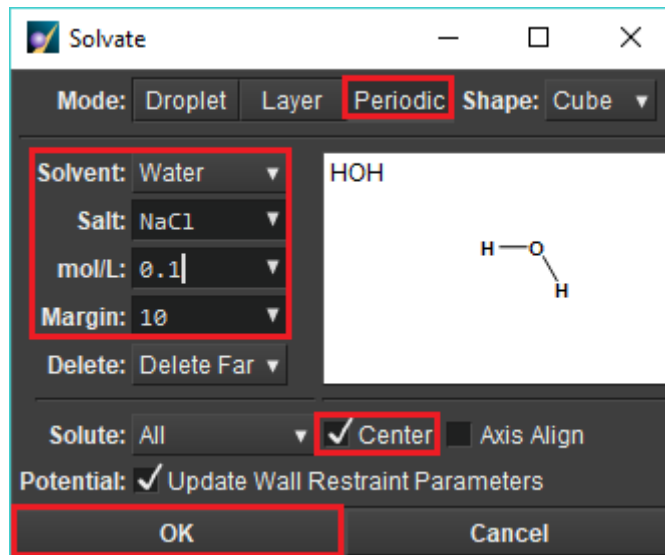
- c) Después Seleccionar en el mismo panel **Protonation | Protonate 3D | OK**. Seleccionar los parámetros fisiológicos predeterminados (o escoger las que sean relevantes a su sistema en estudio).
- d) Finalmente visualizar únicamente el ligando con la macromolécula (para más información sobre visualización consultar Capítulo 7 página 110).



#### 4. Selección y solvatación del sistema.

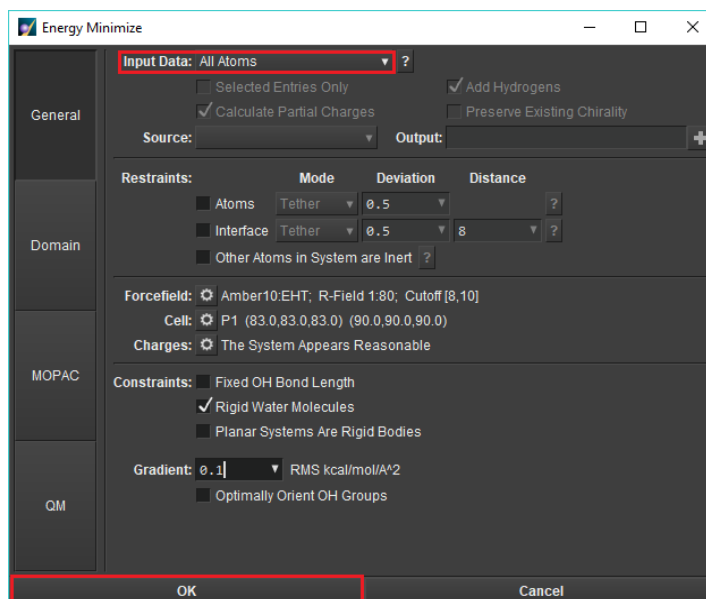
Cuando se simula un sistema periódico, se generan réplicas de celda en todas las direcciones del sistema, es decir, se simula un sistema infinito. Es importante que el sistema se encuentre neutralizado, por ello es conveniente adicionar iones  $\text{Na}^+$  y  $\text{Cl}^-$ . Se debe de seleccionar una celda lo suficientemente grande para evitar choques o contactos intermoleculares. Las restricciones de frontera ocasionan que haya un comportamiento distinto de las moléculas entre el centro y los límites del sistema, ya que las moléculas del centro son las que interaccionan con el complejo ligando-receptor.

Para considerar todas estas características seleccionar **Compute | Simulations | Dynamics | Solvent | Setup | Mode: Periodic, Shape: cube, Solvent: water, Salt: NaCl, mol/L: 0.1, Margin: 10** y hacer clic en **Center** y en **Potential | OK**



5. Minimización de la energía del sistema.

**Compute | Energy Minimize | Input Data: All atoms | OK**



6. Seleccionar los parámetros y el protocolo de la simulación. Realizar la simulación.

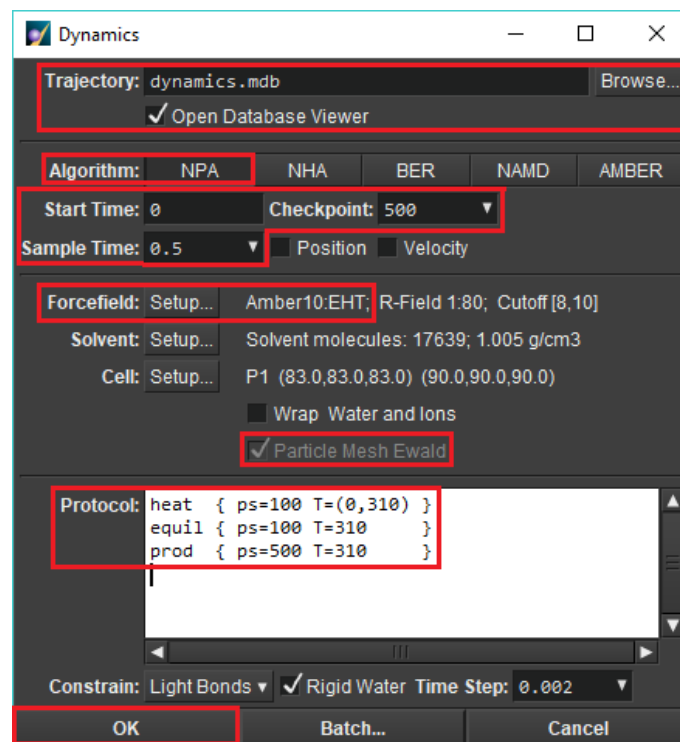
**Algorithm: NPA**, este algoritmo utiliza la ecuación de movimiento de **Nosé-Poincaré-Andersen**, la cual considera un sistema isocórico, isobárico e isotérmico, condiciones similares a las de un ambiente fisiológico real. **Start time: 0, Checkpoint: 500, Sample time: 0.5**, en **Protocol** escribir lo siguiente:

```
heat {ps=100 T= (0,310)}
equil {ps=100 T=310}
prod {ps=500 T=310}
```

Este protocolo de operación indica tres etapas básicas, las cuales consisten en una fase de calentamiento, equilibrio y simulación del sistema.

**Forcefield:** Asegurarse de seleccionar **AMBER 10 EHT** o **CHARMM27**, ya que se ha demostrado buen desempeño usando estos dos campos de fuerza. Esto debido a las consideraciones que hacen en lo que respecta a las interacciones enlazantes y no enlazantes del sistema (Guvench y MacKerell, 2014). Al configurar al sistema como periódico automáticamente se activará la opción de **Particle Mesh Ewald**, la cual es un método de sumas que establece que una partícula cargada interacciona no sólo con las partículas en la celda sino también con sus imágenes.

Finalmente, en la sección **Trajectory**, seleccionar el lugar donde se guardará la base de datos de la trayectoria de la dinámica molecular, además activar la opción **Open Database Viewer**. Hacer clic en | **OK**



## 7. Análisis de resultados.

Durante la simulación de la dinámica molecular se abrirá automáticamente una base de datos en la cual, de acuerdo con el tiempo de muestro, se podrá dar seguimiento en tiempo real a los resultados de las trayectorias de las etapas conforme se vayan generando. Los datos que se podrán ver son:

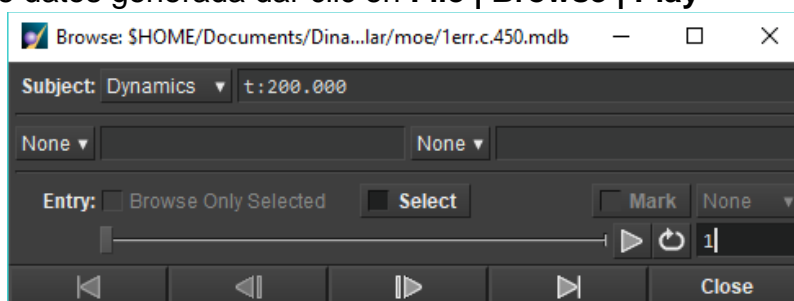
**Tabla 8.1** Parametros de la base de datos de resultados dinámica molecular.

Parámetro	Descripción
<b>stage</b>	Indica la etapa en la que se encuentra la simulación.
<b>t</b>	Indica el tiempo en picosegundos de muestreo.
<b>H</b>	Valor total del Hamiltoniano al tiempo t de muestreo.
<b>U</b>	Muestra el valor de la energía potencial del sistema en Kcal/mol al tiempo t.
<b>K</b>	Muestra el valor de la energía cinética del sistema en Kcal/mol al tiempo t.
<b>T, P, V</b>	Temperatura, presión y volumen del sistema simulado al tiempo t en Kelvin, kPa, y Angstrom cúbicos respectivamente.

	mol	stage	t	H	U	K	T	P	V	Ua
468	1ERR	prod	433.5000	1.3568	-126467.17	26533.9707	310.3138	-121.7514	448570.125	-126464.78
469	1ERR	prod	434.0000	1.2916	-126471.10	26637.9258	311.5295	-102.5152	448570.125	-126468.84
470	1ERR	prod	434.5000	1.3737	-126243.37	26220.3809	306.6463	-94.9973	448570.125	-126240.59
471	1ERR	prod	435.0000	1.4084	-126036.25	26435.0625	309.1570	-89.5735	448570.125	-126031.32
472	1ERR	prod	435.5000	1.0955	-126522.16	26379.4531	308.5067	-104.7527	448570.125	-126518.15
473	1ERR	prod	436.0000	1.3173	-126349.39	26327.7734	307.9023	-88.0900	448570.125	-126347.71
474	1ERR	prod	436.5000	1.3246	-126232.81	26258.5859	307.0931	-106.5827	448570.125	-126229.40
475	1ERR	prod	437.0000	1.1633	-126406.91	26434.0820	309.1456	-100.7553	448570.125	-126403.19
476	1ERR	prod	437.5000	1.3847	-126202.43	26335.1641	307.9887	-109.9345	448570.125	-126198.71
477	1ERR	prod	438.0000	1.3646	-126386.60	26463.1035	309.4850	-109.2501	448570.125	-126383.55
478	1ERR	prod	438.5000	1.3104	-126578.60	26414.1543	308.9125	-142.4513	448570.125	-126576.28
479	1ERR	prod	439.0000	1.2400	-126215.78	26447.7949	309.3059	-111.5674	448570.125	-126210.64

## 8. Visualización de la simulación de la dinámica molecular.

En la base de datos generada dar clic en **File | Browse | Play**



**Nota:** El número uno indica la muestra que se está visualizando.



## Ejercicio

A partir de la siguiente estructura en PDB realizar un estudio de acoplamiento molecular automatizado (ver capítulo 5, página 74) y posteriormente una simulación de dinámica molecular con los mismos parámetros establecidos en el ejemplo anterior. Describir lo que se pide a continuación:

### **PDB ID: 1 HWK**

<i>Tipo de diana biológica</i>	
<i>Función fisiológica</i>	
<i>Padecimiento asociado</i>	
<i>Interacciones claves (Docking)</i>	
<i>Interacciones claves (dinámica molecular)</i>	
<i>Interacciones claves diferentes o en común (Docking y dinámica molecular)</i>	
<i>Número de interacciones donde intervengan moléculas de agua</i>	
<i>Fragmentos de la molécula donde se dan las interacciones clave</i>	

### **Para saber más:**

Ganesan, A., Coote, M. L., & Barakat, K. (2017). Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*, 22(2), 249–269.

Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology*, 48, 93–102.

De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry*, 59(9), 4035–4061.

Moen MD, Keating GM: Raloxifene: a review of its use in the prevention of invasive breast cancer. *Drugs*. 2008;68(14):2059-83.

Heringa M: Review on raloxifene: profile of a selective estrogen receptor modulator. *Int J Clin Pharmacol Ther*. 2003 Aug;41(8):331-45

Prieto-Martínez, F. D., Arciniega, M., & Medina-Franco, J. L. (2018). Acoplamiento Molecular: Avances Recientes y Retos. *TIP Revista Especializada En Ciencias Químico-Biológicas*, 21, 65–87.

Guvench, O., & MacKerell, A. D. (2014). Current status of protein force fields for molecular dynamics simulations. *Molecular Modeling of Proteins: Second Edition*, 443, 47–71.



# Inhibitors of DNA Methyltransferases From Natural Sources: A Computational Perspective

Fernanda I. Saldívar-González, Alejandro Gómez-García, David E. Chávez-Ponce de León, Norberto Sánchez-Cruz, Javier Ruiz-Ríos, B. Angélica Pílon-Jiménez and José L. Medina-Franco\*

Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City, Mexico

## OPEN ACCESS

### Edited by:

Shibashish Giri,  
University of Leipzig, Germany

### Reviewed by:

Jaigopal Sharma,  
Delhi Technological University, India  
Sujata Mohanty,  
All India Institute of Medical Sciences,  
India  
Flup Lal,  
University of Delhi, India

### \*Correspondence:

José L. Medina-Franco  
medinaj@unam.mx;  
jose.medina.franco@gmail.com

### Specialty section:

This article was submitted to  
Pharmacogenetics  
and Pharmacogenomics,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 08 August 2018

**Accepted:** 21 September 2018

**Published:** 10 October 2018

### Citation:

Saldívar-González FI,  
Gómez-García A,  
Chávez-Ponce de León DE,  
Sánchez-Cruz N, Ruiz-Ríos J,  
Pílon-Jiménez BA and  
Medina-Franco JL (2018) Inhibitors  
of DNA Methyltransferases From  
Natural Sources: A Computational  
Perspective.  
Front. Pharmacol. 9:1144.  
doi: 10.3389/fphar.2018.01144

Naturally occurring small molecules include a large variety of natural products from different sources that have confirmed activity against epigenetic targets. In this work we review chemoinformatic, molecular modeling, and other computational approaches that have been used to uncover natural products as inhibitors of DNA methyltransferases, a major family of epigenetic targets with therapeutic interest. Examples of computational approaches surveyed in this work are docking, similarity-based virtual screening, and pharmacophore modeling. It is also discussed the chemoinformatic-guided exploration of the chemical space of naturally occurring compounds as epigenetic modulators which may have significant implications in epigenetic drug discovery and nutriepigenetics.

**Keywords:** chemical space, chemoinformatics, databases, DNMT inhibitors, drug discovery, molecular modeling, similarity searching, virtual screening

## SECTION 1: INTRODUCTION

Epigenetics has been defined as a change in phenotype without an underlying change in genotype (Berger et al., 2009). In the 1940s Waddington suggested the term “epigenetics” trying to describe “the interactions of genes with their environment, which brings the phenotype into being” (Waddington, 2012). Alterations in epigenetic modifications have been related to several diseases including cancer, diabetes, neurodegenerative disorders, and immune-mediated diseases (Dueñas-González et al., 2016; Tough et al., 2016; Hwang et al., 2017; Lu et al., 2018). Moreover, epigenetic targets are also attractive for the treatment of antiparasitic infections (Sacconay et al., 2014).

In epigenetic drug discovery, epigenetic targets have been classified into three main groups (Ganesan, 2018). “Writers” are enzymes that catalyze the addition of a functional group to a protein or nucleic acid; “readers” are macromolecules that function as recognition units that can distinguish a native macromolecule vs. the modified one; and “erasers” that are enzymes that aid in the removal of chemical modifications introduced by the writers. Thus far, several targets from these three major families have reached different stages of drug discovery, ranging from lead discovery, preclinical development, clinical trials and approval. Currently, there are seven compounds approved for clinical use (Ganesan, 2018).

DNA methyltransferases (DNMTs) are a family of “writer” enzymes responsible for DNA methylation that is the addition of a methyl group to the carbon atom number five (C5) of cytosine. As surveyed in this work, since DNA methylation has an essential role for cell differentiation and

development, alterations in the function of DNMTs have been associated with cancer (Castillo-Aguilera et al., 2017) and other diseases (Lyko, 2017).

Several natural products have been identified as inhibitors of epigenetic targets including DNMTs. Most of these compounds have been uncovered fortuitously. However, there are recent efforts to screen systematically natural products as DNMT inhibitors. The vastness of the chemical space of natural products led to the hypothesis that many more active compounds could potentially be identified. Indeed, it has been estimated that more than 95% of the biodiversity in nature remains to be explored to identify potential bioactive molecules (Ho et al., 2018).

The aim of this work is to discuss a broad range of computational methods to identify novel inhibitors of DNMTs from natural products. The manuscript also discusses the chemical space of natural products as inhibitors of DNMTs. The manuscript is organized into nine sections. After this introduction, Section 2 reviews briefly the structure of DNMTs including different isoforms. The next section covers major aspects of the function of DNMTs including the mechanism of methylation. Section 4 reviews currently known inhibitors of DNMTs from natural sources including food chemicals. Section 5 discusses the epigenetic relevant chemical space of natural products comparing the chemical space of DNMT inhibitors from natural sources vs. other compounds. The next section reviews computational strategies that are used to identify natural compounds as potential epi-hits or epi-leads targeting DNMTs. Sections 7 and 8 presents Summary conclusions and Perspectives, respectively.

## SECTION 2: STRUCTURE OF DNMTs

The human genome encodes DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. While DNMT1, DNMT3A, and DNMT3B have catalytic activity, DNMT2 and DNMT3L do not (Lyko, 2017). DNMT1 is a maintenance methyltransferase, responsible for duplicating the pattern of DNA methylation during replication. DNMT1 is essential for proper mammalian development and it has been proposed as the most interesting target for experimental cancer therapies (Dueñas-González et al., 2016). DNMT3A and DNMT3B are *de novo* methyltransferases. Human DNMT1 has 1616 amino acids whose structure can be divided into an N-terminal regulatory domain and a C-terminal catalytic domain (Jeltsch, 2002; Jurkowska et al., 2011). The N-terminal domain contains a replication foci-targeting domain, a DNA-binding CXXC domain, and a pair of bromo-adjacent homology domains. The C-terminal catalytic domain has 10 amino acid motifs. The cofactor and substrate binding sites in the C-terminal catalytic domain are comprised of motif I and X and motif IV, VI, and VIII, respectively (Lan et al., 2010). The target recognition domain which is maintained by motif IX and involved in DNA recognition, is not conserved between the DNMT family. **Figure 1A** shows a three-dimensional (3D) model of

a DNMT1 (PDB ID: 4WXX) (Zhang et al., 2015). **Figure 1B** shows a schematic diagram of human of DNMT1, 2, 3A, 3B, and L.

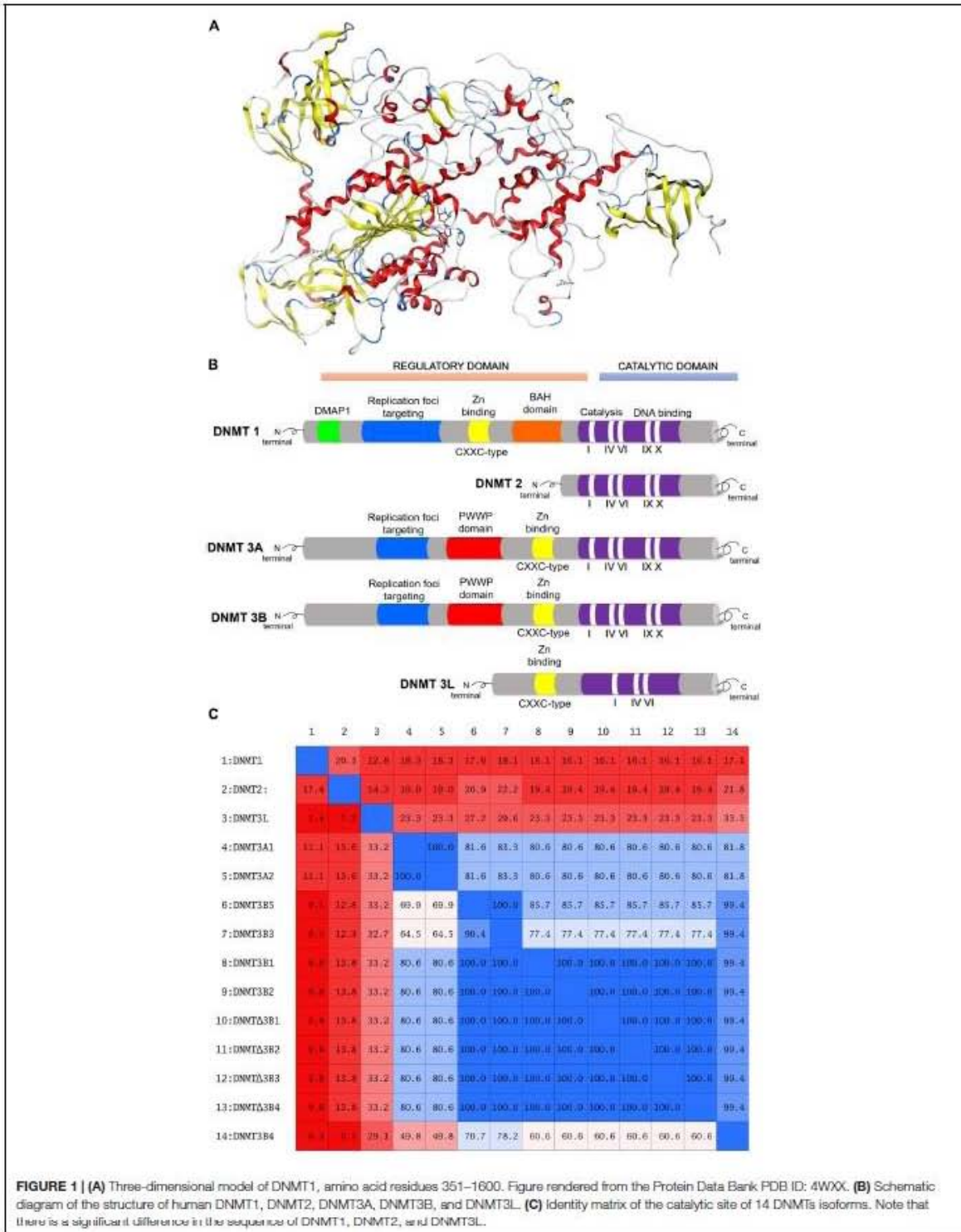
### Section 2.1: Isoforms

Two isoforms of DNMT3A have been identified, DNMT3A1 and DNMT3A2. At the N-terminal domain both isoforms have a PWWP (Pro-Trp-Trp-Pro) and an ADD (ATRX-DNMT3-DNMT3L) domains (Jurkowska et al., 2011). The C-terminal domain is identical in the two isoforms (Choi et al., 2011).

There are more than 30 isoforms of DNMT3B, however, only DNMT3B1 and DNMT3B2 are catalytically active (Ostler et al., 2007). Similar to DNMT3A, DNMT3B1, and DNMT3B2 have a PWWP and ADD domains at the N-terminal region (Lyko, 2017). The rest of the isoforms are not catalytically active. Some of these such as DNMT3B3, DNMT3B4, and DNMT3B7 are overexpressed in many tumor cell lines (Gordon et al., 2013).  $\Delta$ DNMT3B has seven isoforms and lacks 200 amino acids from the N-terminal region of DNMT3B (Wang et al., 2006).  $\Delta$ DNMT3B1–4 possess catalytic activity whereas  $\Delta$ DNMT3B5–7 lacks the catalytic domain (Wang et al., 2006).  $\Delta$ DNMT3B is mainly expressed in non-small cell lung cancer (Wang et al., 2006; Ostler et al., 2007). **Figure 1C** shows the identity matrix of 14 DNMTs isoforms. The identity matrix indicates that the amino acid sequence at the catalytic site of DNMT3A1 and DNMT3A2 isoforms is identical. In the same manner, the amino acid sequence at the C-terminal domain of the catalytically active isoforms DNMT3B1, DNMT3B2, and  $\Delta$ DNMT3B1–4 are identical. DNMT1, DNMT2, and DNMT3L show a significant difference in the sequence of the catalytic site with respect to the rest of the isoforms. Therefore, it can be anticipated that is possible to identify or design selective inhibitors for these isoforms.

## SECTION 3: FUNCTION AND MECHANISM OF DNMTs

As outlined in Section 2, cytosine-5 DNMTs catalyze the addition of methylation marks to genomic DNA. All DNMTs have a related catalytic mechanism that is featured by the formation of a covalent adduct intermediate between the enzyme and the substrate base. All DNMTs use S-adenosyl-L-methionine (SAM) as the methyl group donor (Vilkaitis et al., 2001; Du et al., 2016). DNMT forms a complex with DNA and the cytosine which will be methylated flips out from the DNA (Klimasauskas et al., 1994). A conserved cysteine performs a nucleophilic attack to the six-position of the target cytosine yielding a covalent intermediate. The five-position of the cytosine is activated and conducts a nucleophilic attack on the cofactor SAM to form the 5-methyl covalent adduct and S-adenosyl-L-homocysteine (SAH). The attack on the six-position is aided by a transient protonation of the cytosine ring at the endocyclic nitrogen atom N3, which can be stabilized by a glutamate and arginine residues. The covalent complex between the methylated base and the DNA is resolved by deprotonation at the five-position to generate the methylated cytosine and the free enzyme.



## SECTION 4: KNOWN INHIBITORS OF DNMTs FROM NATURAL SOURCES

Thus far more than 500 compounds have been tested as inhibitors of DNMTs. The structural diversity and coverage in chemical space has been analyzed using chemoinformatic methods (Fernandez-de Gortari and Medina-Franco, 2015). The chemical space of DNMT inhibitors has been compared with inhibitors of other epigenetic targets (Naveja and Medina-Franco, 2018). Furthermore, the structure-activity relationships (SAR) of DNMT inhibitors using the concept of activity landscape has been documented (Naveja and Medina-Franco, 2015).

DNA methyltransferase inhibitors have been obtained from a broad number of different strategies including organic synthesis, virtual, and high-throughput screening (Medina-Franco et al., 2015). Organic synthesis has been employed in several instances for lead optimization (Castellano et al., 2008; Kabro et al., 2013; Davide et al., 2016). Natural products and food chemicals have also been a major source of active compounds. Natural products that are known to act as DNMT inhibitors or demethylating agents have been extensively reviewed by Zwergel et al. (2016). These natural products are of the type polyphenols, flavonoids, anthraquinones, and other classes. Some of the first natural products described were curcumin, (-)-epigallocatechin-3-gallate (EGCG), mahanine, genistein, and quercetin. Other natural products that have been described as inhibitors of DNMT or demethylating agents are silibinin, luteolin, kazinol Q, laccaic acid, hypericin, boswellic acid, and lycopene. **Figure 2** shows the chemical structure of representative DNMT inhibitors with emphasis on compounds from natural origin.

The bioactivity profile and potency in enzymatic and/or cell-based assays of these natural products have been discussed in detail by Zwergel et al. (2016). Of note, it will be valuable if all natural products could have been screened under the same conditions. For few natural products the selectivity has been characterized being nanaomycin A an exception (*vide infra*). Indeed, for about eight natural products the  $IC_{50}$  has been measured in enzymatic based assays. Despite the fact that the potency of the natural products with DNMTs is not very high in enzymatic-based assays, e.g.,  $IC_{50}$  between 0.5 and 10  $\mu$ M, several natural products have shown promising activity in cell-based assays. Notably, natural products have distinct chemical scaffolds that could be used as a starting point in lead optimization efforts. Moreover, quercetin in combination with green tea extract has advanced into phase I clinical trials for the treatment of prostate cancer.

Most of the natural products with demethylating activity or ability to inhibit DNA methyltransferases in enzymatic assays have been identified fortuitously. However, as discussed in this work, there are efforts toward the identification of bioactive demethylating agents using systematic approaches such as a virtual screening. Indeed, the natural product nanaomycin A (**Figure 2**) was identified from a virtual screening campaign initially focused on the identification of inhibitors of DNMT1. The quinone-based antibiotic isolated from *Streptomyces* showed antiproliferative effects in three human tumor cell lines, HCT116, A549, and HL60 after 72 h of treatment. Moreover, nanaomycin A showed

reduced global methylation levels in all three cell lines when tested at concentrations ranging from 0.5 to 5  $\mu$ M. Nanaomycin A reactivated the transcription of the RASSF1A tumor suppressor gene inducing its expression up to 18-fold at 5  $\mu$ M, higher than the reference drug 5-azacytidine (sixfold at 25  $\mu$ M). In an enzymatic inhibitory assay, nanaomycin A was selective toward DNMT3B with an  $IC_{50} = 0.50 \mu$ M.

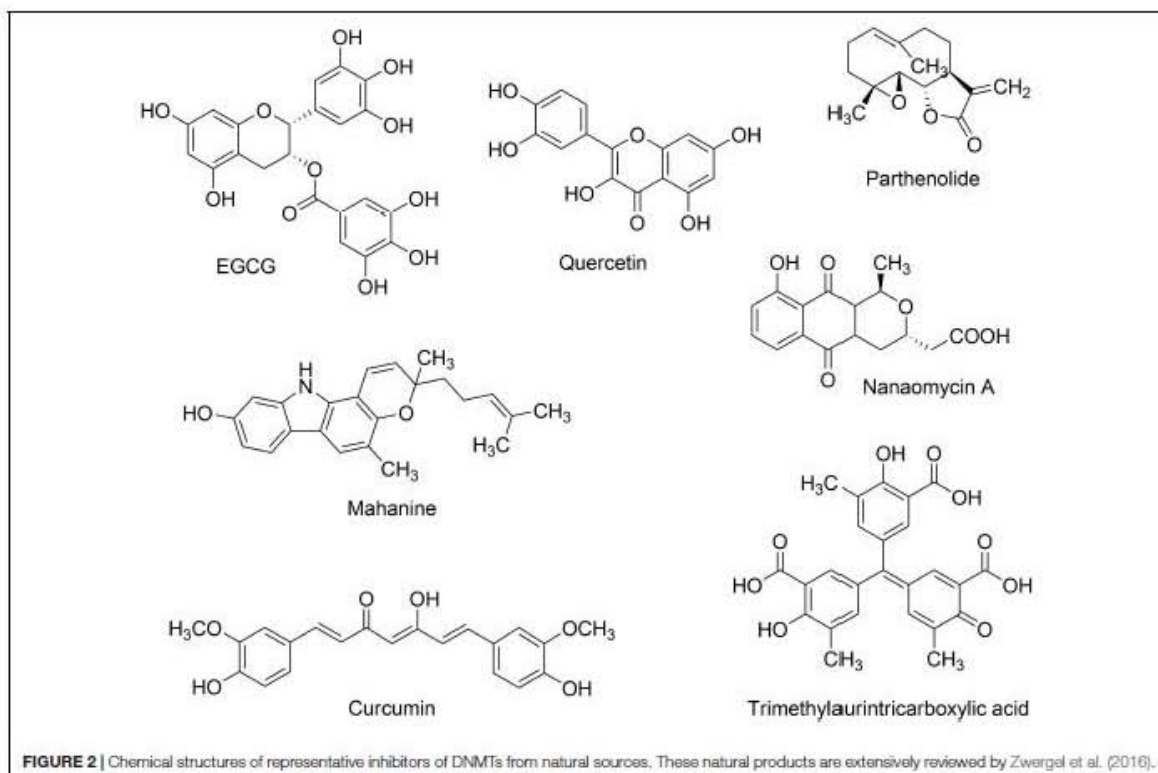
### Section 4.1: Natural Products and Food Chemicals

It is remarkable that several natural products are used as dietary sources such as curcumin, caffeic acid and chlorogenic acid found in *Coffea arabica*, genistein found in soybean, quercetin found in fruits, vegetables, and beverages. Of course, there is a large overlap between the chemical space of food chemicals and natural products (Naveja et al., 2018). This has given rise to systematically screen food chemical databases for potential regulators of epigenetic targets.

## SECTION 5: EPIGENETIC RELEVANT CHEMICAL SPACE OF NATURAL PRODUCTS: FOCUS ON DNMT INHIBITORS

In drug discovery it is generally accepted that a major benefit of natural products vs. purely synthetic organic molecules is, overall, the feasibility of the former to exert a biological activity and increased chemical diversity (Ho et al., 2018). The chemical space of natural products is vast and its molecular diversity has been quantified over the past few years (López-Vallejo et al., 2012; Olmedo et al., 2017; Shang et al., 2018). A major contribution to these studies has been the increasing availability of natural products collections in the public domain (Medina-Franco, 2015). Examples of major compound collections are the Traditional Chinese Medicine (Chen, 2011), natural products from Brazil – NuBBE (Pilon et al., 2017), AfroDb (Ntie-Kang et al., 2013) or collections available for screening in a medium to high-throughput screening mode. The large importance of natural products in drug discovery has boosted the development of open access applications to mine these rich repositories. Few examples are ChemGPS-NP, TCMAAnalyzer, and other resources described elsewhere (Rosen et al., 2009; Chen et al., 2017; Gonzalez-Medina et al., 2017; Liu et al., 2018).

The chemical space of natural products from different sources has been compared to several other collections including the chemical space of drugs approved for clinical use and synthetic compounds (Olmedo et al., 2017; Shang et al., 2018). These studies demonstrate that the chemical space of natural products is vast, that there is a notable overlap with the chemical space of drugs, and that natural products cover novel regions of the chemical space. The overlap with the chemical space of approved drugs is not that surprising since there are a large percentage of drugs from natural origin. **Figure 3** shows a visual representation of the chemical space of 15 representative DNMT inhibitors from natural sources vs. 4103 compounds from a commercial



vendor library of natural products, 206 fungi metabolites, and 6253 marine natural products (Krishna et al., 2017). The visual representation was generated with principal component analysis of six physicochemical properties of pharmaceutical relevance, namely molecular weight (MW), topological surface area (TPSA), number of hydrogen bond donors and acceptors (HBD/HBA), number of rotatable bonds (RB), and octanol/water partition coefficient (logP). The first two principal components capture about 90% of the total variance. The visual representation of the chemical space in this figure indicates that marine natural products (data points in blue) cover a broader area of the chemical space followed by natural products in the vendor collection (orange) and by fungi metabolites (green). DNMT inhibitors from natural origin (purple) are, in general, inside the subspace of the DNMT1 inhibitors (red). This visualization of the chemical space indicates that there would be expected to identify more DNMT1 inhibitors in the marine and vendor collections, as well as in the data set of fungi metabolites.

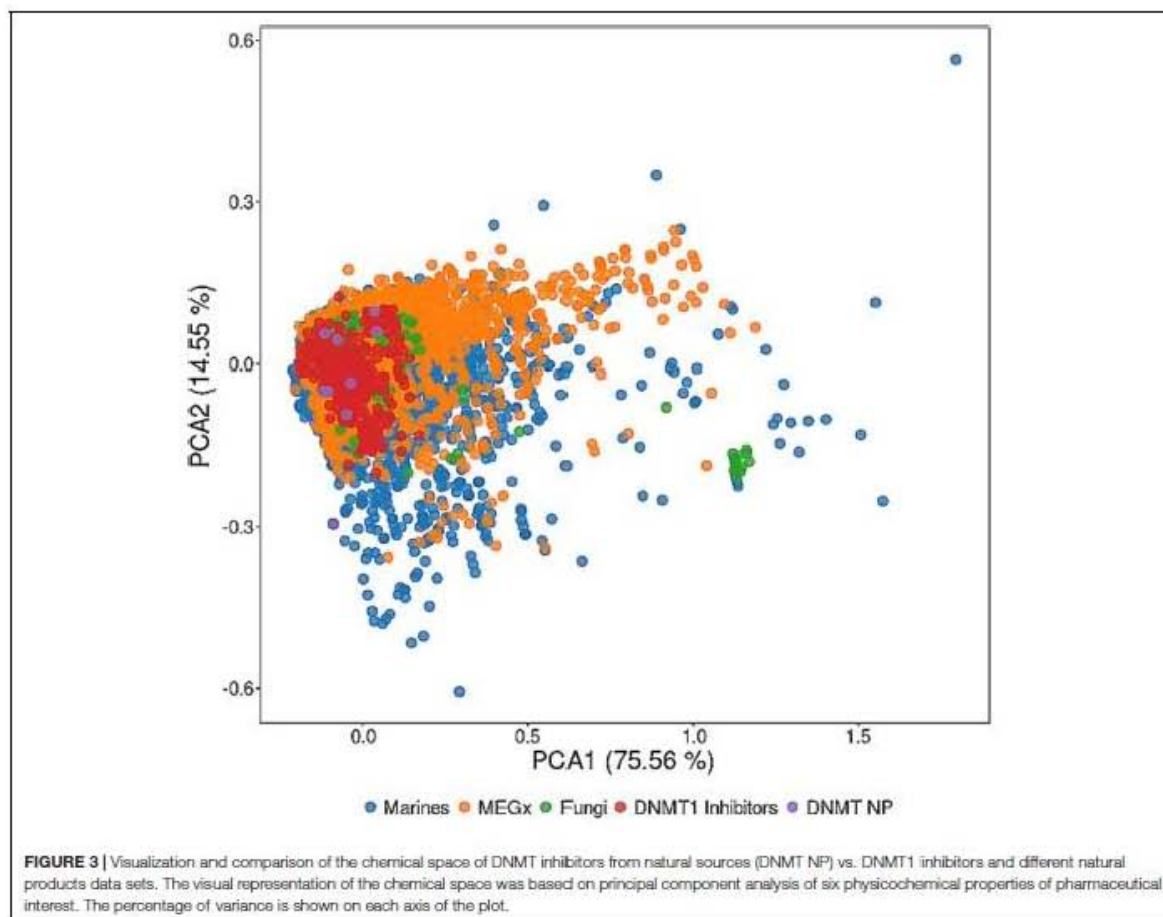
## SECTION 6: OPPORTUNITIES FOR SEARCHING FOR NATURAL PRODUCTS AS DNMT INHIBITORS

Most of the DNMT inhibitors from natural sources have been identified by serendipity. As discussed in Section 5, the chemical

space of natural products and food chemicals can be explored in a systematic manner using computational approaches. A classical and general approach is using virtual screening. The main aim of virtual screening is filtering compound data sets to select a reduced number of compounds with increased probability to show biological activity. Virtual screening has proven to be useful to identify hit compounds (Clark, 2008; Lavecchia and Di Giovanni, 2013). **Table 1** summarizes representative case studies where virtual screening has led to the identification of active compounds with novel scaffolds. In other studies, virtual screening has uncovered potential active compounds but experimental validation still needs to be conducted. Examples of these studies are further discussed in the following sections.

There are several published studies of virtual screening of natural products to identify DNMT inhibitors and/or demethylating agents. In an early work, Medina-Franco et al. (2011) reported the screening of a lead-like subset of natural products available in ZINC. Authors of that work implemented a multistep virtual screening approach selecting consensus hits identified from three different docking programs. One computational hit showed DNMT1 activity in a previous study. Other candidate compounds were identified for later experimental validation (Medina-Franco et al., 2011).

In a separate work, Maldonado-Rojas et al. (2015) developed a QSAR model based on linear discriminant analysis to screen 800 natural products. Hits selected were further

**TABLE 1** | Summary of virtual screening hits as inhibitors of DNMTs.

Study	<i>In silico</i> approach	Major outcome	Reference
Structure-based screening of a lead-like subset of NP from ZINC	Cascade docking followed by a consensus approach	One computational had reported activity. Additional natural products were identified for screening.	Medina-Franco et al., 2011
Ligand- and structure-based screening of 800 NP	QSAR model based on linear discriminant analysis and consensus docking.	Six consensus hits were identified as potential inhibitors.	Maldonado-Rojas et al., 2015
Structure-based screening of 111,121 molecules.	Docking-based screening of synthetic screening compounds.	Identification of a low micromolar hit with a novel scaffold. Further similarity searching led to the identification of two more potent hits.	Chen et al., 2014
Ligand-based screening of 500 compounds.	Pharmacophore-based virtual screening.	Identification of one inhibitor of DNMT1 with activity in the low micromolar range. The hit showed some selectivity vs. DNMT3B.	Hassanzadeh et al., 2017
Structure- and ligand-based screening of 53,000 synthetic compounds.	Pharmacophore model, a Naïve Bayesian classification model, and ensemble docking.	Two compounds showed DNMT1 inhibitory activity at single but low concentration of 1 $\mu$ M.	Krishna et al., 2017

NP: natural products.

docked with two crystallographic structures of human DNMT employing two docking programs. Six consensus hits were identified as potential inhibitors (Maldonado-Rojas et al., 2015).

Virtual screening of synthetic libraries has also been reported to identify active compounds with novel scaffolds and suitable for lead optimization. For instance, Chen et al. (2014) reported a docking-based virtual screening of the commercial screening compound library SPECS with 111,121 molecules (after filtering compounds with undesirable physicochemical properties). Results of that work led to the identification of a compound with a novel scaffold with low micromolar  $IC_{50}$  (10.3  $\mu$ M). Starting from the computational hit, similarity searching led to the identification of two more potent compounds.

Hassanzadeh et al. (2017) recently reported a pharmacophore-based virtual screening of a compound database with 500 compounds. The pharmacophore was generated using a ligand-based approach by superimposing a group of active nucleoside analogs. Selected hits, which are structurally related to the barbituric acid, were docked into the substrate binding site of DNMT1. One compound was identified with a novel chemical scaffold that inhibits DNMT1 in the low micromolar range ( $IC_{50} = 4.1 \mu$ M). The compound also showed some selectivity on DNMT1 over DNMT3 enzymes (Hassanzadeh et al., 2017).

Krishna et al. (2017) implemented a virtual screening protocol using several structure- and ligand-based approaches. Methods included a pharmacophore model, a Naïve Bayesian classification model, and ensemble docking. Three out of ten selected compounds from a commercial library of synthetic molecules (e.g., Maybridge with 53,000 small drug-like compounds), showed DNMT1 inhibitory activity at compound concentration of 20  $\mu$ M. Two of these molecules showed activity at 1  $\mu$ M (Krishna et al., 2017).

In addition to the studies discussed above and summarized in Table 1, the next subsections discuss other approaches that can be explored. Case studies for each strategy are outlined briefly.

### Section 6.1: Similarity-Based Virtual Screening of Natural Products

Similarity searching is a commonly used approach for identifying new hit compounds. Major goals are identifying starting points for later optimization or expand the SAR of analog series. Since similarity searching is fast it can be used to filter large chemical databases and it can be used in combination with other computational approaches such as molecular docking.

Similarity searching involves two major components: a molecular representation and a similarity coefficient. In practice, one of the most common molecular representations are two-dimensional (2D) fingerprints. A fingerprint is generally a string of zeros and ones that indicate the presence or absence of molecular features, respectively. In turn, one of the most common similarity coefficients is Tanimoto's (Bajusz et al., 2015). Full discussion of molecular representations and similarity coefficients are published elsewhere (Willett et al., 1998; Maggiora et al., 2014).

A novel approach to encode the chemical structures of data sets is the database fingerprint (DBFP) (Fernández-de Gortari et al., 2017). The rationale of DBFP is account for the most structural features encoded in bit positions of an entire data set. In principle, virtually any data set can be represented. For instance, it can be a small or large chemical database of screening compounds or a group of active compounds. DBFP can be used in visual representation of the chemical space (Naveja and Medina-Franco, 2018) and similarity searching (Fernández-de Gortari et al., 2017). More recently, this approach was further refined into the so-called statistical based database fingerprint (SB-DFP). This approach has the same underlying idea and application of DBFP. A key improvement is the approach to account for the most relevant structural features that are derived from a statistical comparison between the structural features of a data set of interest vs. a database of reference.

### Section 6.2: Pharmacophore-Based

Thus far, several pharmacophore modeling studies have been conducted for inhibitors of DNMT1. Different approaches and input molecules have been used to develop these models. Most of the pharmacophore models have been employed to virtually screen chemical databases and identify novel hit compounds.

Yoo and Medina-Franco (2011) reported one of the first pharmacophore models for inhibitors of DNMT1. The model was generated based on the docking poses of 14 known inhibitors available at that time. The docking was conducted with a homology model of the catalytic domain of DNMT1. Of note, at the time of that study the crystallographic structure of human DNMT1 was not available. Known DNMT inhibitors used to develop the pharmacophore model included the natural products curcumin, parthenolide, EGCG and mahanine (Yoo and Medina-Franco, 2011). A year later was reported that trimethylaurintricarboxylic acid (Figure 2) showed a good agreement with this structure-based pharmacophore model. This compound is structurally related to 5,5'-methylenedisalicylic acid that has an inhibition of DNMT1 in a low micromolar range ( $IC_{50} = 4.79 \mu$ M) (Yoo and Medina-Franco, 2012; Yoo et al., 2012).

More recently, as described in the first part of Section 6, Hassanzadeh et al. (2017) developed a pharmacophore model based on a ligand-based approach by 3D superimposition of active nucleoside analogs. That model was used to do virtual screening (*vide supra*). In the same year, with the aid of the Hypogen module of the software DS4.1, Krishna et al. (2017) developed a ligand-based pharmacophore model using the structures of 20 compounds obtained from the literature. The model was validated with the classification of an external set with known active and inactive compounds. The validated pharmacophore models were employed as part of a combined strategy to identify novel active molecules (Krishna et al., 2017).

### Section 6.3: De novo Design

*De novo* design is a technique currently explored for DNMT inhibitors on a limited basis. Here we briefly outline two promising perspectives related to natural product research.



The first one is a strategy that provides a structural diversity classification of natural products scaffolds through generative topographic map algorithm implementation often so-called chemographies. Chemographies allow the visualization of the landscape distribution of the chemical space of natural products and their synthetic mimetic compounds (Miyao et al., 2015). Since chemographies could be generated from pharmacophoric features and molecular descriptors, it would be feasible to do scaffold hopping based on the structures of natural products (Rodrigues et al., 2016). The second approach is based on scaffold simplification that could be adapted to generate fragment-like natural products focused on DNMT inhibitors. This strategy reduces the molecular framework of natural products through the implementation of a scaffold tree algorithm based on rule-based decomposition of ring systems (Bajorath, 2018).

## SECTION 7: CONCLUSION

Epigenetic targets are attractive to develop therapeutic strategies. DNA methyltransferases are the major enzyme family being one of the first epigenetic targets studied, in particular for the treatment of cancer. However, over the past few years, more therapeutic opportunities related to the modulation of DNMTs are emerging. Therefore, there is a growing interest in the scientific community to identify and develop small molecules that can be used as epi-drugs or epi-probes targeting DNMTs. Virtual screening has become more used in recent years to uncover natural products as inhibitors of DNMTs and/or demethylating agents. To this end, well established structure- and ligand-based virtual screening approaches are being used such as automated docking, QSAR and similarity searching. Also, novel chemoinformatic approaches are being developed. Of course, the computational methods should be validated with rigorous experiments *in vitro* and *in vivo* experiments to support their application.

Natural products have a well established history as inhibitors of DNMTs and demethylating molecules. However, most of the active natural products have been identified by serendipity. The knowledge of the three-dimensional structures of DNMTs in combination with increased *in silico* approaches and better computational resources are boosting the systematic search of bioactive molecules from natural origin. In addition, the increasing availability of natural product databases facilitates the discovery of epi-drugs and epi-probes targeting DNMTs.

## SECTION 8: PERSPECTIVES

Natural products inside or outside of the traditional drug-like chemical space represent a large promise to develop novel compounds with DNMT inhibitory activity or demethylating properties. This is because the traditional chemical space is highly represented by small molecules

that over the past few years have not been very successful. A notable example in this direction is the reemergence of peptide-based drug discovery. Indeed, linear, cyclic peptides and peptidomimetics are regaining interest in drug discovery (Fosgerau and Hoffmann, 2015; Henninot et al., 2018).

Other promising an emerging avenue are the modulators of protein-protein interactions (PPIs) (Díaz-Eufracio et al., 2018). DNMTs are known to be involved in several PPIs (Díaz-Eufracio et al., 2018). Modulation of such interactions can be conveniently achieved with natural products. This is because PPIs are “difficult targets” not easily addressed by small molecules from the traditional chemical space (Villoutreix et al., 2014). In other words, since PPIs have unique features these can be approached with novel chemical libraries. Natural products collections represent excellent candidates for this purpose.

We foresee an augmented hit and lead identification efforts based on natural products combining approaches such as high-throughput screening, structure-, ligand-based *in silico* screening, structure-based optimization, similarity searching, and scaffold hopping (Schneider et al., 1999). As part of the search for novel and more potent compounds is crucial to consider potential toxicity since toxicity issues play a major part in the lack of success of drug discovery projects.

## DISCLAIMER

A similar version of this manuscript was deposited in a Pre-Print server on July 6, 2018. The reference is: Saldívar-González, F. I.; Gómez-García, A.; Sánchez-Cruz, N.; Ruiz-Ríos, J.; Pilón-Jiménez, B. A.; Medina-Franco, J. L. Computational Approaches to Identify Natural Products as Inhibitors of DNA Methyltransferases. *Preprints* 2018, 2018070116 (doi: 10.20944/preprints201807.0116.v1).

## AUTHOR CONTRIBUTIONS

All authors contributed to methodology and formal analysis. FS-G, JR-R, and BP-J contributed to data curation. AG-G, FS-G, DC-PdL, and JM-F contributed to writing-original draft preparation. AG-G, FS-G, NS-G, and JM-F contributed to writing-review and editing. AG-G, FS-G, and BP-J contributed to visualization. JM-F contributed to project administration.

## FUNDING

This research was funded by *Consejo Nacional de Ciencia y Tecnología* (CONACYT, Mexico) grant number 282785, the *Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica* (PAPIIT) grant IA203718, and by *Programa de Apoyo a Proyectos para*

la Innovación y Mejoramiento de la Enseñanza (PAPIME) grant PE200118, UNAM.

## ACKNOWLEDGMENTS

FS-G, AG-G, and NS-C acknowledge *Consejo Nacional de Ciencia y Tecnología* (CONACyT, Mexico) for the graduate

scholarships. DC-PdI and JR-R thanks the *Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza* (PAPIME) for the undergraduate scholarship. The authors also thank Chanachai Sae-Lee for providing the sequences used in **Figure 1**. They also acknowledge all current and past members of the DIFACQUIM research group for their comments and discussions that enriched this manuscript.

## REFERENCES

- Bajorath, J. (2018). Improving the utility of molecular scaffolds for medicinal and computational chemistry. *Future Med. Chem.* 10, 1645–1648. doi: 10.4155/fmc-2018-0106
- Bajusz, D., Rácz, A., and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* 7:20. doi: 10.1186/s13321-015-0069-3
- Berger, S. L., Kouzarides, T., Shikhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783. doi: 10.1101/gad.1787609
- Castellano, S., Kuck, D., Sala, M., Novellino, E., Lyko, F., and Sbardella, G. (2008). Constrained analogues of procaine as novel small molecule inhibitors of DNA methyltransferase-1. *J. Med. Chem.* 51, 2321–2325. doi: 10.1021/jm7015705
- Castillo-Aguilera, O., Depreux, P., Halby, L., Arimondo, P., and Goossens, L. (2017). DNA methylation targeting: the DNMT/HMT crosstalk challenge. *Biomolecules* 7:3. doi: 10.3390/biom7010003
- Chen, C. Y.-C. (2011). TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS One* 6:e15939. doi: 10.1371/journal.pone.0015939
- Chen, S. J., Wang, Y. L., Zhou, W., Li, S. S., Peng, J. L., Shi, Z., et al. (2014). Identifying novel selective non-nucleoside DNA methyltransferase 1 inhibitors through docking-based virtual screening. *J. Med. Chem.* 57, 9028–9041. doi: 10.1021/jm501134e
- Chen, Y., de Bruyn Kops, C., and Kirchmair, J. (2017). Data resources for the computer-guided discovery of bioactive natural products. *J. Chem. Inf. Model.* 57, 2099–2111. doi: 10.1021/acs.jcim.7b00341
- Choi, S. H., Heo, K., Byun, H.-M., An, W., Lu, W., and Yang, A. S. (2011). Identification of preferential target sites for human DNA methyltransferases. *Nucleic Acids Res.* 39, 104–118. doi: 10.1093/nar/gkq774
- Clark, D. E. (2008). What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discov.* 3, 841–851. doi: 10.1517/17460440802281978
- Davide, G., Sandra, A., Emily, B., Mattia, C., Marta, G., Annalisa, C., et al. (2016). Design and synthesis of N-benzoyl amino acid derivatives as DNA methylation inhibitors. *Chem. Biol. Drug Des.* 88, 664–676. doi: 10.1111/cbdd.12794
- Díaz-Eufracio, B. I., Naveja, J. J., and Medina-Franco, J. L. (2018). Chapter three - protein-protein interaction modulators for epigenetic therapies. *Adv. Protein Chem. Struct. Biol.* 110, 65–84. doi: 10.1016/bs.apcsb.2017.06.002
- Du, Q., Wang, Z., and Schramm, V. L. (2016). Human DNMT1 transition state structure. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2916–2921. doi: 10.1073/pnas.1522491113
- Dueñas-González, A., Jesús Naveja, J., and Medina-Franco, J. L. (2016). Chapter 1 - Introduction of Epigenetic Targets in Drug Discovery and Current Status of Epi-Drugs and Epi-Probes, in *Epi-Informatics*. Boston, MA: Academic Press, 1–20. doi: 10.1016/B978-0-12-802808-7.00001-0
- Fernández-de Gortari, E., García-Jacas, C. R., Martínez-Mayorga, K., and Medina-Franco, J. L. (2017). Database fingerprint (DFP): an approach to represent molecular databases. *J. Cheminform.* 9:9. doi: 10.1186/s13321-017-0195-1
- Fernandez-de Gortari, E., and Medina-Franco, J. L. (2015). Epigenetic relevant chemical space: a chemoinformatic characterization of inhibitors of DNA methyltransferases. *RSC Adv.* 5, 87465–87476. doi: 10.1039/C5RA19611F
- Fosgerau, K., and Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discov. Today* 20, 122–128. doi: 10.1016/j.drudis.2014.10.003
- Ganesan, A. (2018). Epigenetic drug discovery: a success story for cofactor interference. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20170069. doi: 10.1098/rstb.2017.0069
- Gonzalez-Medina, M., Naveja, J. J., Sanchez-Cruz, N., and Medina-Franco, J. L. (2017). Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* 7, 54153–54163. doi: 10.1039/C7RA11831G
- Gordon, C. A., Hartono, S. R., and Chédin, F. (2013). Inactive DNMT3B splice variants modulate de novo DNA methylation. *PLoS One* 8:e69486. doi: 10.1371/journal.pone.0069486
- Henninot, A., Collins, J. C., and Nuss, J. M. (2018). The current state of peptide drug discovery: back to the future? *J. Med. Chem.* 61, 1382–1414. doi: 10.1021/acs.jmedchem.7b00318
- Ho, T. T., Tran, Q. T. N., and Chai, C. L. L. (2018). The polypharmacology of natural products. *Future Med. Chem.* 10, 1361–1368. doi: 10.4155/fmc-2017-0294
- Hwang, J.-Y., Aromolaran, K. A., and Zukin, R. S. (2017). The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat. Rev. Neurosci.* 18, 347–361. doi: 10.1038/nrn.2017.46
- Jeltsch, A. (2002). Beyond Watson and crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem* 3, 274–293. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S
- Jurkowska, R. Z., Jurkowski, T. P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. *ChemBioChem* 12, 206–222. doi: 10.1002/cbic.201000195
- Kabro, A., Lachance, H., Marcoux-Archambault, L., Perrier, V., Dore, V., Gros, C., et al. (2013). Preparation of phenylethylbenzamide derivatives as modulators of DNMT3 activity. *MedChemComm* 4, 1562–1570. doi: 10.1039/c3md00214d
- Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. D. (1994). HHA1 methyltransferase flips its target base out of the DNA helix. *Cell* 76, 357–369. doi: 10.1016/0092-8674(94)90342-5
- Krishna, S., Shukla, S., Lakra, A. D., Meeran, S. M., and Siddiqi, M. I. (2017). Identification of potent inhibitors of DNA methyltransferase 1 (DNMT1) through a pharmacophore-based virtual screening approach. *J. Mol. Graph. Model.* 75, 174–188. doi: 10.1016/j.jmgs.2017.05.014
- Lan, J., Hua, S., He, X. N., and Zhang, Y. (2010). DNA methyltransferases and methyl-binding proteins of mammals. *Acta Biochim. Biophys. Sin.* 42, 243–252. doi: 10.1093/abbs/gmq015
- Lavecchia, A., and Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* 20, 2839–2860. doi: 10.2174/09298673113209990001
- Liu, Z., Du, J., Yan, X., Zhong, J., Cui, L., Lin, J., et al. (2018). TCMAnalyzer: a chemo- and bioinformatics web service for analyzing traditional Chinese medicine. *J. Chem. Inf. Model.* 58, 550–555. doi: 10.1021/acs.jcim.7b00549
- López-Vallejo, F., Giulianotti, M. A., Houghten, R. A., and Medina-Franco, J. L. (2012). Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov. Today* 17, 718–726. doi: 10.1016/j.drudis.2012.04.001
- Lu, W., Zhang, R., Jiang, H., Zhang, H., and Luo, C. (2018). Computer-aided drug design in epigenetics. *Front. Chem.* 6:57. doi: 10.3389/fchem.2018.00057
- Lyko, F. (2017). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* 19, 81–92. doi: 10.1038/nrg.2017.80
- Maggiara, G., Vogt, M., Stumpfe, D., and Bajorath, J. (2014). Molecular similarity in medicinal chemistry. *J. Med. Chem.* 57, 3186–3204. doi: 10.1021/jm401411z
- Maldonado-Rojas, W., Olivero-Verbel, J., and Marrero-Ponce, Y. (2015). Computational fishing of new DNA methyltransferase inhibitors from natural products. *J. Mol. Graph. Model.* 60, 43–54. doi: 10.1016/j.jmgs.2015.04.010

- Hassanzadeh, M., Kasymov, R., Mahernia, S., Adib, M., Emperle Michael Dukatz, M., Bashtrykov, P., et al. (2017). Discovery of novel and selective DNA methyltransferase 1 inhibitors by pharmacophore and docking-based virtual screening. *ChemistrySelect* 2, 8383–8392. doi: 10.1002/slct.201701734
- Medina-Franco, J. L. (2015). "Discovery and development of lead compounds from natural sources using computational approaches," in *Evidence-Based Validation of Herbal Medicine*, ed. P. Mukherjee (New York, NY: Elsevier), 455–475. doi: 10.1016/B978-0-12-800874-4.00021-0
- Medina-Franco, J. L., López-Vallejo, F., Kuck, D., and Lyko, F. (2011). Natural products as DNA methyltransferase inhibitors: a computer-aided discovery approach. *Mol. Divers.* 15, 293–304. doi: 10.1007/s11030-010-9262-5
- Medina-Franco, J. L., Méndez-Lucio, O., Yoo, J., and Dueñas, A. (2015). Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discov. Today* 20, 569–577. doi: 10.1016/j.drudis.2014.12.007
- Miyao, T., Reker, D., Schneider, P., Funatsu, K., and Schneider, G. (2015). Chemography of natural product space. *Planta Med.* 81, 429–435. doi: 10.1055/s-0034-1396322
- Naveja, J. J., and Medina-Franco, J. L. (2015). Activity landscape sweeping: insights into the mechanism of inhibition and optimization of DNMT1 inhibitors. *RSC Adv.* 5, 63882–63895. doi: 10.1039/C5RA12339A
- Naveja, J. J., and Medina-Franco, J. L. (2018). Insights from pharmacological similarity of epigenetic targets in epipolypharmacology. *Drug Discov. Today* 23, 141–150. doi: 10.1016/j.drudis.2017.10.006
- Naveja, J. J., Rico-Hidalgo, M. P., and Medina-Franco, J. L. (2018). Analysis of a large food chemical database: chemical space, diversity, and complexity. *FI000Res.* 7:993. doi: 10.12688/fi000research.15440.2
- Ntie-Kang, F., Zofou, D., Babiaka, S. B., Meudom, R., Scharfe, M., Lifongo, L. L., et al. (2013). AfroDb: a select highly potent and diverse natural product library from African medicinal plants. *PLoS One* 8:e78085. doi: 10.1371/journal.pone.0078085
- Olmedo, D. A., González-Medina, M., Gupta, M. P., and Medina-Franco, J. L. (2017). Cheminformatic characterization of natural products from panama. *Mol. Divers.* 21, 779–789. doi: 10.1007/s11030-017-9781-4
- Ostler, K. R., Davis, E. M., Payne, S. L., Gosalia, B. B., Expósito-Céspedes, J., Beau, M. M. L., et al. (2007). Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. *Oncogene* 26, 5553–5563. doi: 10.1038/sj.onc.1210351
- Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., et al. (2017). NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* 7:7215. doi: 10.1038/s41598-017-07451-x
- Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016). Counting on natural products for drug design. *Nat. Chem.* 8, 531–541. doi: 10.1038/nchem.2479
- Rosen, J., Lovgren, A., Kogej, T., Muresan, S., Gottfries, J., and Backlund, A. (2009). ChemGPS-NPWeb: chemical space navigation online. *J. Comput. Aided Mol. Des.* 23, 253–259. doi: 10.1007/s10822-008-9255-y
- Sacconay, L., Angleviel, M., Randazzo, G. M., Queiroz, M. M., Queiroz, E. F., Wolfender, J. L., et al. (2014). Computational studies on sirtuins from *Trypanosoma cruzi*: structures, conformations and interactions with phytochemicals. *PLoS Negl. Trop. Dis.* 8:e2689. doi: 10.1371/journal.pntd.0002689
- Schneider, G., Neidhart, W., Giller, T., and Schmid, G. (1999). Scaffold-hopping by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.* 38, 2894–2896. doi: 10.1002/(SICI)1521-3773(19991004)38:19<2894::AID-ANIE2894>3.0.CO;2-F
- Shang, J., Hu, B., Wang, J., Zhu, F., Kang, Y., Li, D., et al. (2018). Cheminformatic insight into the differences between terrestrial and marine originated natural products. *J. Chem. Inf. Model.* 58, 1182–1193. doi: 10.1021/acs.jcim.8b00125
- Tough, D. F., Tak, P. P., Tarakhovskiy, A., and Prinjha, R. K. (2016). Epigenetic drug discovery: breaking through the immune barrier. *Nat. Rev. Drug Discov.* 15, 835–853. doi: 10.1038/nrd.2016.185
- Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E., and Klimasauskas, S. (2001). The mechanism of DNA cytosine-5 methylation - kinetic and mutational dissection of HhaI methyltransferase. *J. Biol. Chem.* 276, 20924–20934. doi: 10.1074/jbc.M101429200
- Villoutreix, B. O., Kuenemann, M. A., Poyet, J. L., et al. (2014). Drug-like protein-protein interaction modulators: challenges and opportunities for drug discovery and chemical biology. *Mol. Inf.* 33, 414–437. doi: 10.1002/minf.201400040
- Waddington, C. H. (2012). The epigenotype. *Int. J. Epidemiol.* 41, 10–13. doi: 10.1093/ije/dyr184
- Wang, L., Wang, J., Sun, S., Rodriguez, M., Yue, P., Jang, S. J., et al. (2006). A novel DNMT3B subfamily,  $\Delta$ DNMT3B, is the predominant form of DNMT3B in Non-small cell lung cancer. *Int. J. Oncol.* 29, 201–207. doi: 10.3892/ijo.29.1.201
- Willett, P., Barnard, J., and Downs, G. (1998). Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996. doi: 10.1021/ci9800211
- Yoo, J., Kim, J. H., Robertson, K. D., and Medina-Franco, J. L. (2012). Molecular modeling of inhibitors of human DNA methyltransferase with a crystal structure: discovery of a novel DNMT1 inhibitor. *Adv. Protein Chem. Struct. Biol.* 87, 219–247. doi: 10.1016/B978-0-12-398312-1.00008-1
- Yoo, J., and Medina-Franco, J. L. (2011). Homology modeling, docking, and structure-based pharmacophore of inhibitors of DNA methyltransferase. *J. Comp. Aided Mol. Des.* 25, 555–567. doi: 10.1007/s10822-011-9441-1
- Yoo, J., and Medina-Franco, J. L. (2012). Trimethylaurintricarboxylic acid inhibits human DNA methyltransferase 1: insights from enzymatic and molecular modeling studies. *J. Mol. Model.* 18, 1583–1589. doi: 10.1007/s00894-011-1191-4
- Zhang, Z.-M., Liu, S., Lin, K., Luo, Y., Perry, J. J., Wang, Y., et al. (2015). Crystal structure of human DNA methyltransferase 1. *J. Mol. Biol.* 427, 2520–2531. doi: 10.1016/j.jmb.2015.06.001
- Zwergel, C., Valente, S., and Mai, A. (2016). DNA methyltransferases inhibitors from natural sources. *Curr. Top. Med. Chem.* 16, 680–696. doi: 10.2174/1568026615666150825141505

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Saldívar-González, Gómez-García, Chávez-Ponce de León, Sánchez-Cruz, Ruiz-Rios, Pilon-Jiménez and Medina-Franco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.