



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

**Identificación de moduladores de DNA metiltransferasas de
quimiotecas enfocadas**

TESIS

PARA OPTAR POR EL GRADO DE

MAESTRA EN CIENCIAS

PRESENTA

Ing. Liliam Parvati Martínez Fernández

Dr. José Luis Medina Franco
Facultad de Química, UNAM

Ciudad Universitaria. CDMX. Enero 2023



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.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

**Identificación de moduladores de DNA metiltransferasas de
quimiotecas enfocadas**

**T E S I S
PARA OPTAR POR EL GRADO DE**

MAESTRA EN CIENCIAS

P R E S E N T A

Ing. Liliam Parvati Martínez Fernández



Ciudad de México, Enero 2023

Los seres humanos no nacen para siempre
el día en que sus madres los alumbran,
sino que la vida los obliga a parirse
a sí mismos una y otra vez.

Gabriel García Marquez

Dedicatoria

Quiero dedicarle esta tesis a mi madre, Zenaida Liliam Fernández Sanchez, por su infinito amor, apoyo y aliento. Gracias por enseñarme y transmitirme tu amor por la medicina.

Agradecimientos

0.1 Personales

Quiero agradecerle a mi pareja, Nissim Hernández, por su apoyo durante el desarrollo de esta investigación. También quiero agradecerle a mis amigos y compañeros por el acompañamiento en esta etapa de mi vida, en especial a Bárbara Morales. Finalmente, le agradezco a mi gata, Cat, por acompañarme todas las noches en las que desarrollé esta tesis.

0.2 Institucionales

Agradezco a mi tutor Jose Luis Medina Franco por su apoyo en el desarrollo de este trabajo. También, agradezco a los integrantes del grupo de investigación de DIFACQUIM, en especial a Euridice Juárez Mercado y a Bárbara Diaz Eufracio.

Agradezco al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la Dirección General de Asuntos del Personal Académico de la UNAM, proyecto PAPIIT IN201321. También agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca 1012509 otorgada para la realización de este trabajo.

Glosario de abreviaturas

5 mC = 5-metilcitosina

AID = enzimas deaminasa de citosina inducida por activación

BAH = Homología adyacente de bromo

BL = Basados en ligandos

BS = Basados en estructuras

CXXC = Cys-X-X-Cys

DIFAC = Diseño de fármacos asistido por computadora

DIFBL = Diseño de fármacos basados en ligando

DIFBS = diseño de fármacos basados en estructuras

DNMT = ADN metiltransferasas

ETP = *Epigenetic Target Profiler*

FDA = Administración de Alimentos y Medicamentos de los Estados Unidos

HBA = aceptores de enlaces de hidrógeno

HBD = donadores de enlaces de hidrógeno

IA = Inteligencia artificial

kNN = *k-Nearest neighbors*

LMA = leucemia mieloide aguda

MDB = proteínas con dominio de unión a metil-CpG

ML = Aprendizaje automático

MOE = *Molecular Operating Environment*

MW = peso molecular

PDB = Protein Data Bank

PUMA = *Platform for Unified Molecular Analysis*

RB = enlaces rotables

RF = *Random forest*

RFTS = secuencia dirigida a focos de replicación

RMSD = Desviación cuadrática media

SlogP = coeficiente de partición octanol/agua

SVM = *Support-vector machines*

TET = enzimas de translocación diez-once

TPSA = área topológica superficial

UHRF = proteínas similares a la ubiquitina

Resumen

La metilación del ADN es un proceso epigenético clave involucrado en el control de la expresión génica y la regulación de la impronta parental, entre otras cosas. No obstante, los patrones de metilación aberrantes están asociados con diversas enfermedades incluidas las enfermedades psiquiátricas, las enfermedades del sistema inmunológico y contribuyen tanto al inicio como a la progresión de varios tipos de cáncer.

La metilación está mediada por una familia de enzimas llamadas ADN metiltransferasas (DNMTs, por sus siglas en inglés, *DNA methyltransferases*), entre ellas la DNMT1. Debido al rol de esta enzima, se busca modular su actividad para poder contrarrestar los patrones de metilación aberrantes. No obstante, a pesar de la importancia de la DNMT1, todavía no se han podido desarrollar fármacos que permitan modular su actividad sin fuertes efectos citotóxicos.

Debido al presente interés en el campo de la epigenética se han desarrollado quimiotecas enfocadas centradas en moduladores de objetivos epigenéticos. Al estar enriquecidas, estas bibliotecas permiten disminuir la cantidad de datos a analizar. Además, en los últimos años ha aumentado el desarrollo de herramientas computacionales como las basadas en acoplamiento molecular y aprendizaje automático.

El objetivo de este trabajo fue encontrar moduladores de DNMT1 en quimiotecas enfocadas. Para alcanzar el objetivo primero se realizó una búsqueda de quimiotecas enfocadas cuyos compuestos estuviesen disponibles de manera comercial. Posteriormente, se analizó el espacio químico y se realizó un filtrado a partir de subestructuras presentes en compuestos con actividad reportada. Se estudiaron los compuestos filtrados mediante acoplamiento molecular (LeDock, Autodock Vina y MOE) y modelos de aprendizaje automático (Orange3 y *Epigenetic Target Profiler*). Finalmente, se evaluaron los compuestos seleccionados mediante pruebas de inhibición enzimática. Se identificaron once compuestos activadores de la DNMT1 de los cuales siete aumentaron la actividad enzimática para que fuera mayor de 150%. Estos once compuestos resultan ser los primeros activadores históricamente reportados de este blanco. Este descubrimiento da pie a la posibilidad de que estos sean utilizados para el tratamiento de enfermedades como el cáncer de próstata, donde hay una menor cantidad de metilación (sección 1.1.3).

Índice

0.1	Personales	iii
0.2	Institucionales	iii
1	Introducción	1
1.1	Antecedentes	1
1.1.1	Epigenética	1
1.1.2	Metilación	2
1.1.3	Relevancia terapéutica de la metilación	7
1.1.4	DNMT1	9
1.1.5	Moduladores ya conocidos y sus subestructuras	11
1.1.6	Bibliotecas moleculares (quimiotecas) enfocadas	14
1.1.7	Diseño de fármacos asistido por computadora	15
1.2	Planteamiento del problema	22
1.3	Hipótesis	23
1.4	Objetivos	23
1.4.1	Objetivo general	23
1.5	Contribución	24
2	Metodología	25
2.1	Formación y curado de las bases de datos moleculares	26
2.2	Cuantificación de la diversidad estructural	27
2.3	Espacio químico	27
2.4	Núcleo estructural base	27
2.5	Métodos basados en ligando	28
2.5.1	Aprendizaje supervisado para la predicción de actividad in- hbitoria	28
2.6	Métodos basados en estructura	30
2.6.1	Acoplamiento molecular	30
2.7	Ensayos de inhibición enzimática	31
3	Resultados y discusión	33
3.1	Diversidad estructural de la base de datos	33
3.2	Núcleo estructural base	35
3.3	Selección de compuestos mediante subestructuras	35

3.4	Métodos basados en ligandos	37
3.4.1	<i>Epigenetic Target Profiler</i> v1.0	37
3.4.2	Orange 3	37
3.5	Acoplamiento molecular	43
3.6	Selección de compuestos para evaluar actividad	45
3.7	Ensayos enzimáticos	45
4	Conclusiones	49
5	Perspectivas	50
	Referencias	50
	Apéndice	60
A	Acoplamiento molecular	61
A.1	Muy activadores	61
A.2	Activadores	65
B	Pósteres científicos y publicaciones	67

Lista de figuras

1.1	Inactivación de transcripción por metilación de DNMT1.	2
1.2	Metilación del ADN. Una citosina se convierte a 5-metilcitosina catalizada por DNMTs con la donación del grupo metilo por SAM la cual se convierte a SAH (S-adenosil L-homocisteína).	4
1.3	Metilación <i>de novo</i> donde se lleva a cabo la metilación sin que ninguna de las hebras esté metilada y metilación de mantenimiento donde el ADN esta hemimetilado, es decir que solo una de las hebras esta metilada.	4
1.4	Mecanismo de desmetilación del ADN donde tras ser metilada por DNMT1 se lleva a cabo un proceso de desmetilación por las entimzas TET y TDG.	6
1.5	Publicaciones de PubMed con la palabra clave epigenética. Accesado el 10 de agosto de 2022.	7
1.6	Dominios de DNMT1 (PDB ID: 4WXX) donde: A corresponde a RFTS (AA351-600), B a CXXC (AA651-697), C al enlazador autoinhibidor (AA697-728), D a BAH1 (AA728-900), E es BAH2 (AA906-1100), F es el dominio de metiltransferasa (AA1139-1616) G es el sitio catalítico (AA1224-1245) y H corresponde a los enlazadores entre los dominios.	9
1.7	Regulación alostérica de DNMT1. El dominio RFTS (amarillo) tiene una función de autoinhibición, donde se une al dominio catalítico de la enzima (turquesa) esto bloquea el acceso del ADN a la hendidura de unión al ADN y al centro catalítico (rojo).	10
1.8	Mecanismo de los inhibidores de DNMT1.	12
1.9	Ejemplos de inhibidores de DNMT1 donde a corresponde a la categoría de los nucleosídicos y b a la categoría de los no nucleosídicos.	13
1.10	Formación de núcleo estructural base de Murcko de GSK-3484862.	17
1.11	Metodología de un método de aprendizaje automático Orange3. Un conjunto de datos esta unido a una herramienta de un modelo de RF el cual genera un modelo que permite predecir la actividad de los compuestos mediante una herramienta de predicciones.	18

1.12	Métodos de aprendizaje automático: <i>k-nearest neighbors</i> (kNN), <i>Random forest</i> (RF) y <i>Support vector machines</i> (SVM). Los distintos colores en los círculos representan distintas clasificaciones de los datos, para kNN se utiliza la proximidad de los datos, para RF se forman árboles de decisión y para SVM se generan hiperplanos.	19
2.1	Esquema de la metodología seguida.	25
2.2	Interfaz del programa <i>Epigenetic Target Profiler</i> v1.0.	28
2.3	Flujo de trabajo en Orange donde se abren los archivos de DNMT1-IC ₅₀ -MORGAN y DNMT1-IC ₅₀ -RDK, se someten a los métodos de SVM, <i>Random forest</i> y kNN.	29
2.4	Reacción de S-adenosil-L-[metil- ³ H]metionina + ADN = S-adenosil-L-homocisteína + ADN 5-[metil- ³ H]-citosina.	32
3.1	Diagramas de caja y bigotes de las propiedades fisicoquímicas estudiadas donde: A es Asinex, B es Axon, C es ChemDiv, D es Enamine, E es Enzo, F es Life Chemicals, G es OTAVA_DNMT1, H es OTAVA_DNMT3, I es Selleck y J es TargetMol.	34
3.2	Subestructuras a estudiar donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol.	36
3.3	Visualización del espacio químico mediante un análisis de componentes principales donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas, D es carbazol y X es el resto de los compuestos.	36
3.4	Visualización del espacio químico mediante un análisis de componentes principales para la selección de compuestos por subestructura. En a se clasifican por subestructura donde: A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol. En b se clasifican entre ChemDiv y el resto de las bases de datos.	37
3.5	Resultados del método de aprendizaje automático para cada subestructura donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol.	39
3.6	Visualización del espacio químico mediante un análisis de componentes principales donde se incluyen las bases de datos de la tabla 3.6. En a se clasifican por el tipo de base de datos. En b se expande la clasificación para mostrar los compuestos de las bases de datos que tienen alguna de las subestructuras estudiadas donde: A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas, D es carbazol, X es el resto de los compuestos de las bases de datos enfocadas y Y los compuestos de las bases de datos de la tabla 3.6. En c se muestran las bases de datos de la tabla 3.6. En d se añadió a la figura c los compuestos de las bases de datos enfocadas.	42
3.7	Distribución de la energía libre de Gibbs (ΔG) de unión por programa empleado.	43

3.8	Función de densidad de probabilidad del ΔG de LeDock, Autodock Vina y MOE de acuerdo a la subestructura encontrada.	44
3.9	Función de densidad de probabilidad de la energía de Gibbs de LeDock, Autodock Vina y MOE de acuerdo a la actividad predicha por el método de aprendizaje automático.	44
3.10	Selección de compuestos para evaluación de inhibición enzimática.	45
3.11	Porcentaje de los compuestos clasificados por su energía libre y su actividad donde: I es inactivos, A es activadores y MA es muy activadores.	47
3.12	Porcentaje de compuestos que tienen algún contacto con cada dominio de la DNMT1.	47
3.13	Dominios de DNMT1 donde: A corresponde a RFTS (AA351-600), B a CXXC (AA651-697), C al enlazador autoinhibidor (AA697-728), D a BAH1 (AA728-900), E es BAH2 (AA906-1100), F es el dominio de metiltransferasa (AA1139-1616) G es el compuesto SAM y H corresponde a los enlazadores entre los dominios, I es el compuesto C350-0522.	48
A.1	I es el compuesto C066-2421.	61
A.2	I es el compuesto C350-0522.	62
A.3	I es el compuesto L112-0668.	62
A.4	I es el compuesto L112-0725.	63
A.5	I es el compuesto L112-0728.	63
A.6	I es el compuesto L112-0731.	64
A.7	I es el compuesto L485-2669.	64
A.8	I es el compuesto L112-0688.	65
A.9	I es el compuesto L483-0323.	65
A.10	I es el compuesto L485-2761.	66
A.11	I es el compuesto M056-0726.	66

Lista de tablas

1.1	DNMTs en distintas enfermedades.	8
2.1	Descripción de las quimiotecas enfocadas.	26
3.1	Porcentaje de compuestos con reglas cumplidas de Lipinski y Veber por base de datos.	33
3.2	Núcleos estructurales por base de datos donde: M es el número de compuestos, N el número de estructuras de Murcko, Ns los <i>singleton scaffolds</i> y Pn el porcentaje de núcleos estructurales que representan n% de compuestos.	35
3.3	Resultados de la validación interna de Morgan – IC ₅₀	38
3.4	Resultados de la validación interna de RDK – IC ₅₀	38
3.5	Resultados de los modelos de aprendizaje automático para todos los compuestos de las bases de datos.	39
3.6	Descripción de las bases de datos.	40
3.7	Resultados del método de aprendizaje automático para todos los compuestos de las bases de datos generales.	40
3.8	Núcleos estructurales por base de datos donde: M es el número de compuestos, N el número de estructuras de Murcko, Ns los <i>singleton scaffolds</i> y Pn el porcentaje de núcleos estructurales que representan n% de compuestos.	41
3.9	Resultados de la energía libre de los controles utilizados para el acoplamiento molecular.	43
3.10	Resultados de la actividad enzimática relativa con DNMT1 después de la administración de 100 μ M de cada compuesto donde MA corresponde a los compuestos muy activadores, A a los compuestos activadores y I a los compuestos inactivos.	46
3.11	Comparación entre actividad del compuesto y subestructura.	46

Capítulo 1

Introducción

1.1 Antecedentes

1.1.1 Epigenética

La epigenética ha sido descrita en la literatura con múltiples definiciones. La primera de estas fue propuesta en 1942 por Conrad Waddington buscando describir la conexión entre genotipo y fenotipo.[1–3] Después, el grupo de investigación de Arthur Riggs la definió como las diferencias heredadas entre mitosis y meiosis, las cuales permiten explicar los cambios en el fenotipo.[3] Ambos estaban tratando de encontrar el vínculo entre el genotipo y el fenotipo. Es por esto que el término epigenética significa: genética fuera de lo convencional.[3, 4]

En la actualidad, la epigenética define el estudio de los cambios hereditarios en la actividad o función de los genes que no está asociada a un cambio en la secuencia del ácido desoxirribonucleico (ADN). Dentro de esta categoría se incluyen un conjunto de mecanismos reversibles que modulan la función del genoma en diferentes tipos de células lo cual conlleva a un cambio en la expresión génica.[4–6] Estos cambios son heredables y afectan una amplia gama de actividades celulares como: el crecimiento celular, la diferenciación y el desarrollo de enfermedades, entre otras.[7] Las señales epigenéticas incluyen: metilación del ADN, modificaciones covalentes de las proteínas histonas (tales como acetilación, fosforilación y metilación), incorporación de variantes de histonas y ARN no codificantes.[3] Todas ellas funcionan sinérgicamente para modular la estructura de la cromatina. Esta se encuentra de dos maneras: la eucromatina que es una conformación abierta en la que se encuentran la mayoría de los genes transcripcionalmente activos y la heterocromatina el cual es un estado más compacto e inactivo transcripcionalmente.[8, 9] Esta modificación epigenética regula el cambio dinámico entre los dos estados, que es un proceso reversible y regulado dinámicamente.[10] Por lo tanto, permite determinar la actividad transcripcional del genoma. Entre los eventos epigenéticos, la metilación del ADN ha sido la más ampliamente estudiada.[4, 6, 11]

1.1.2 Metilación

La metilación del ADN, descubierta por primera vez por Hotchkiss en 1948[6, 12], es un proceso epigenético clave involucrado en el control de la expresión génica, la regulación de la impronta parental, la estabilización de la inactivación del cromosoma X y el mantenimiento de la integridad del genoma.[12] También está implicada en el desarrollo del sistema inmunológico y en la reprogramación celular, así como en la función y el comportamiento del cerebro.[12]

Los patrones de metilación del ADN aberrantes están asociados con varias enfermedades humanas, incluidas las enfermedades psiquiátricas, las enfermedades del sistema inmunológico y contribuyen tanto al inicio como a la progresión de varios cánceres como el cáncer de mama, de pulmón y de próstata.[13–15]

La hipometilación surge principalmente de la pérdida de metilación en elementos repetidos normalmente metilados, lo que lleva a la inestabilidad genómica y la activación de oncogenes.[4] Por otro lado, la hipermetilación específica del locus generalmente ocurre en las islas CpG intragénicas (CGI, por sus siglas en inglés *CpG islands*) que son promotoras de los genes supresores de tumores.[4] Los sitios CpG o sitios CG son regiones de ADN donde un nucleótido de citosina es seguido por un nucleótido de guanina en la secuencia lineal de bases a lo largo de su dirección 5' → 3'; las regiones donde estos se repiten son llamadas islas CpG.[16] Por ende, el silenciamiento de estas islas mediante la metilación resulta en un silenciamiento de genes supresores de tumores.[4]

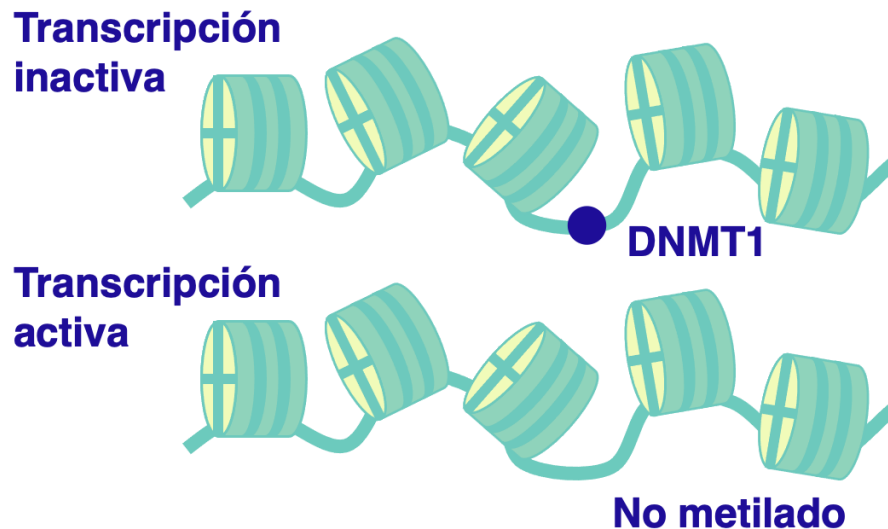


Figura 1.1: Inactivación de transcripción por metilación de DNMT1.

Al ser reconocida como un regulador clave de la estabilidad transcripcional, la metilación del ADN establece un estado silencioso de la cromatina (figura 1.1).[4] No obstante, la integración de la metilación del ADN con otras modificaciones epi-

genéticas es un proceso complejo que depende de la colaboración de numerosos componentes, muchos de los cuales quedan por dilucidar.[17, 18] Por lo tanto, el análisis de los perfiles de metilación del ADN mejorará la comprensión de todo el epigenoma.

Mecanismo de metilación

Los elementos involucrados en diferentes patrones de modificación se dividen en tres grupos principales: escritores, lectores y borradores.[19] Los escritores son enzimas que catalizan una modificación estructural de la proteína o del ADN y en este caso sería la adición del grupo metilo. En segundo lugar, los lectores son unidades de reconocimiento que pueden discriminar entre la macromolécula nativa y su versión modificada. Finalmente, los borradores son enzimas que catalizan la eliminación de la señal química introducida por los escritores.[3]

Gracias a los muchos años de investigación dedicados a comprender cómo se borra y remodela el panorama epigenético durante el desarrollo embrionario, se han identificado muchas de las proteínas y los mecanismos implicados en la metilación del ADN.[6]

Rol de las metiltransferasas en la metilación del ADN

La metilación del ADN es uno de los mecanismos epigenéticos mejor estudiados en células eucariotas y consiste en la adición covalente de un grupo metilo al carbono 5' de un anillo de citosina; principalmente dentro de un dinucleótido CpG, dando lugar a una nueva base de ADN, 5-metilcitosina (5 mC).[20] Este proceso es catalizado por ADN metiltransferasas (DNMTs, por sus siglas en inglés, *DNA methyltransferases*), una familia de enzimas escritoras que utilizan S-adenosil-L-metionina (SAM) como donante del grupo metilo (figura 1.2).[18]

La metilación de la citosina no interfiere con la separación de las bases nucleotídicas de Watson-Crick ya que el grupo se inserta en el surco mayor del ADN, donde es reconocido por las proteínas que interactúan con el ADN.[20] Esto causa que se inactive la transcripción y se silencie la actividad de la cromatina, por lo que pasa de eucromatina a heterocromatina.

La DNMT1, DNMT3A y DNMT3B tienen una región reguladora N-terminal y una región catalítica C-terminal; no obstante, tienen funciones y patrones de expresión únicos.[20] Se pensaba que la DNMT1 actuaba principalmente como una metiltransferasa de mantenimiento durante la síntesis de ADN y que DNMT3A y DNMT3B actuaban como enzimas que metilan *de novo* (figura 1.3). Sin embargo, actualmente existe evidencia de que la DNMT1 también puede ser necesaria para la metilación *de novo* del ADN y que DNMT3A y DNMT3B contribuyen al mantenimiento de la metilación durante la replicación.[21, 22]

Estudios en ratones indican que la desactivación de DNMT1 produce como resultado letalidad embrionaria. Sorprendentemente, las células madre embrionarias

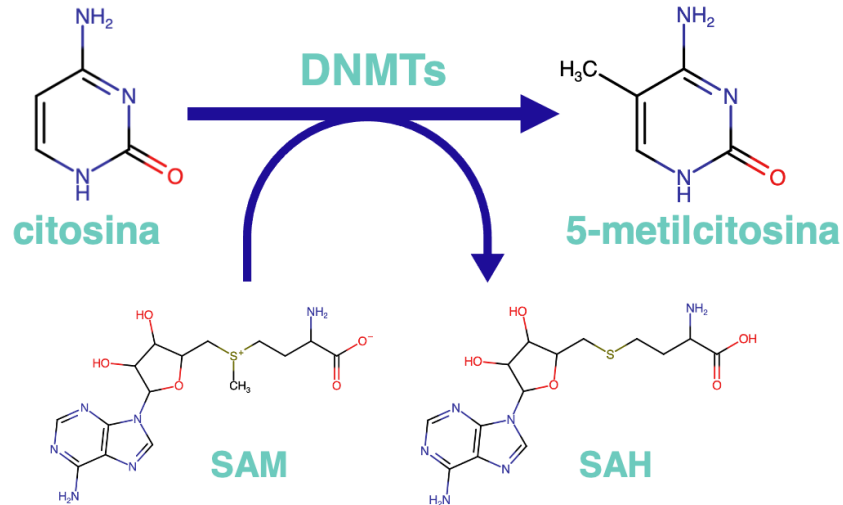


Figura 1.2: Metilación del ADN. Una citosina se convierte a 5-metilcitosina catalizada por DNMTs con la donación del grupo metilo por SAM la cual se convierte a SAH (S-adenosil L-homocisteína).

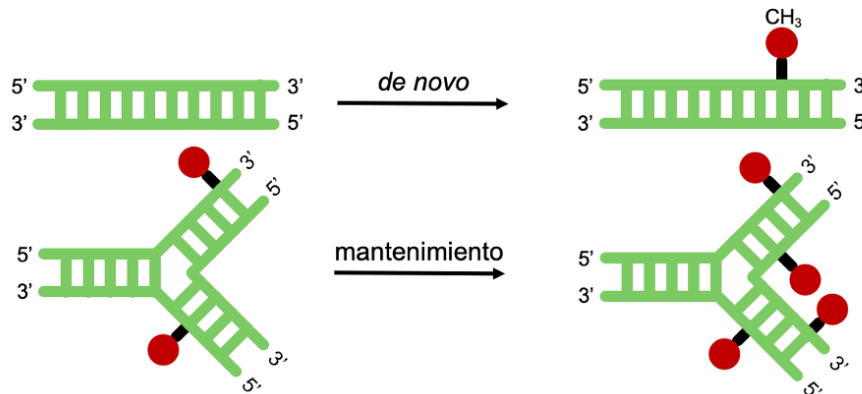


Figura 1.3: Metilación *de novo* donde se lleva a cabo la metilación sin que ninguna de las hebras esté metilada y metilación de mantenimiento donde el ADN está hemimetilado, es decir que solo una de las hebras está metilada.

de ratón que carecen de DNMT1 siguen siendo viables. Es durante la diferenciación *in vitro* que hay una muerte celular masiva. Estos hallazgos establecen firmemente que DNMT1 desempeña un papel fundamental en la diferenciación celular, así como en la división de las células. [6]

Por otro lado, DNMT3A y DNMT3B son extremadamente similares en estructura y función, lo que las distingue es su patrón de expresión génica. Aunque DNMT3A se expresa de manera relativamente ubicua, DNMT3B se expresa de manera deficiente en la mayoría de los tejidos diferenciados, con la excepción de la

tiroides, los testículos y la médula ósea. Similar a DNMT1, la desactivación de DNMT3B en ratones causa letalidad embrionaria. Por otro lado, los ratones con DNMT3A desactivado sobreviven hasta cuatro semanas después del nacimiento. Estos resultados indican que se requiere DNMT3B durante el desarrollo temprano, mientras que se requiere la DNMT3A para la diferenciación celular normal. [6]

Mecanismos de desmetilación del ADN

La desmetilación del ADN puede ser pasiva o activa. La desmetilación pasiva sucede cuando hay una inhibición de la actividad de la DNMT1 lo que permite que durante la replicación celular la citosina recién incorporada permanezca sin metilar disminuyendo así el nivel de metilación general. [6]

La desmetilación activa del ADN puede ocurrir tanto en células en división como en células que no se dividen, pero el proceso requiere reacciones enzimáticas para convertir las 5 mC a citosinas desnudas. [6] Se han propuesto varios mecanismos de desmetilación activa del ADN. En estos principalmente se involucran las enzimas deaminasa de citosina inducida por activación (AID o AICDA, por sus siglas en inglés, *activation induced cytosine deaminase*) o enzimas de translocación diez-once (TET, por sus siglas en inglés, *ten–eleven translocation*) (figura 1.4). Sin embargo, se ha demostrado que de las enzimas partícipes en el proceso de desmetilación, solamente en la ausencia de la timina ADN glicosilasa (TDG, por sus siglas en inglés, *thymine DNA glycosylase*) existe letalidad embrionaria. Esto indica la relevancia de los distintos métodos de desmetilación tanto activos como pasivos. [6]

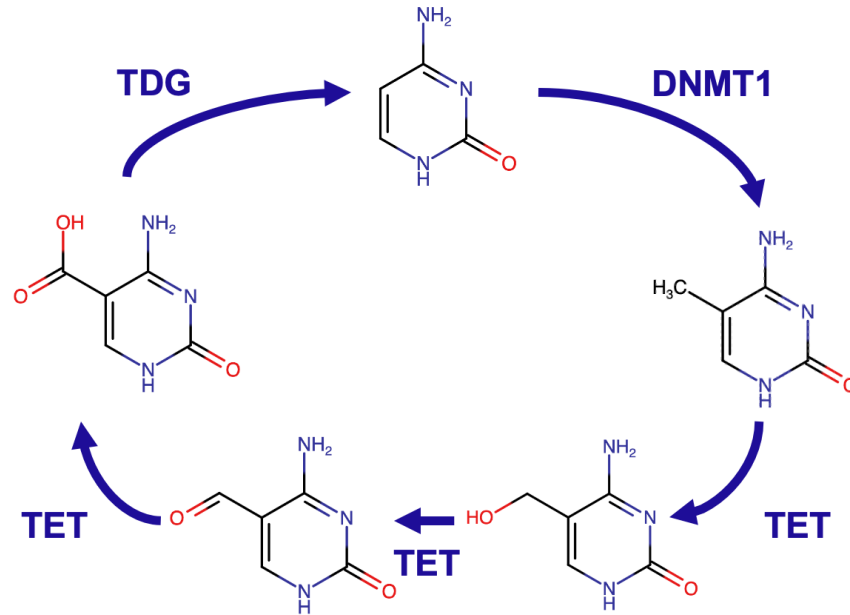


Figura 1.4: Mecanismo de desmetilación del ADN donde tras ser metilada por DNMT1 se lleva a cabo un proceso de desmetilación por las entimzas TET y TDG.

Lectores de la metilación

Mientras que las DNMTs pueden por sí mismas reducir la expresión génica al alterar la unión de los activadores transcripcionales, una segunda clase de proteínas con una alta afinidad por 5 mC inhibe la unión del factor de transcripción. La metilación del ADN es reconocida por tres familias de proteínas: las proteínas con dominio de unión a metil-CpG (MBD, por sus siglas en inglés, *methyl-CpG-binding domain*), las proteínas similares a la ubiquitina (UHRF, por sus siglas en inglés, *Ubiquitin-like, containing PHD and RING finger domains*) y las proteínas con dedos de zinc. De estas familias, la MBD fue la primera en ser identificada.[6]

Las proteínas MBD contienen un dominio de unión a metil-CpG (MBD) conservado que confiere una mayor afinidad por los sitios CpG metilados individuales. Esta familia incluye MeCP2, la primera proteína de unión a metilo identificada, junto con MBD1, MBD2, MBD3 y MBD4. Además de su papel como represor transcripcional, MeCP2 parece tener un papel único en el mantenimiento de la secuencia dirigida a focos de replicación (RFTS, por sus siglas en inglés, *replication focus targeting sequence*) y puede conducir la DNMT1 al ADN hemimetilado para realizar la metilación.[6]

Por otro lado, se encuentran las proteínas UHRF, incluidas UHRF1 y UHRF2, las cuales a diferencia de la mayoría de las proteínas MBD, su función principal no es unirse al ADN y reprimir la transcripción. Estas primero se unen al dominio de RFTS de la DNMT1 y luego la dirigen al ADN hemimetilado para mantener la

metilación del ADN.[23, 24] UHRF1 interactúa tan estrechamente con DNMT1 que su eliminación, así como la eliminación de DNMT1, conduce a la letalidad embrionaria. Por otro lado, la última familia de proteínas de dedos de zinc, compuesta por Kaiso, ZBTB4 y ZBTB38, se unen al ADN metilado mediante su dominio de dedo de zinc.[6]

1.1.3 Relevancia terapéutica de la metilación

Las modificaciones epigenéticas son esenciales para el desarrollo y la proliferación celular, pero se pueden ver afectados por factores aleatorios o por influencias ambientales. La interrupción de los procesos epigenéticos da como resultado estados transcripcionales alterados que pueden culminar en una transformación celular maligna. Los estados epigenéticos pueden verse alterados por factores ambientales, lo que probablemente conduzca al desarrollo de fenotipos anormales o patológicos.[6]

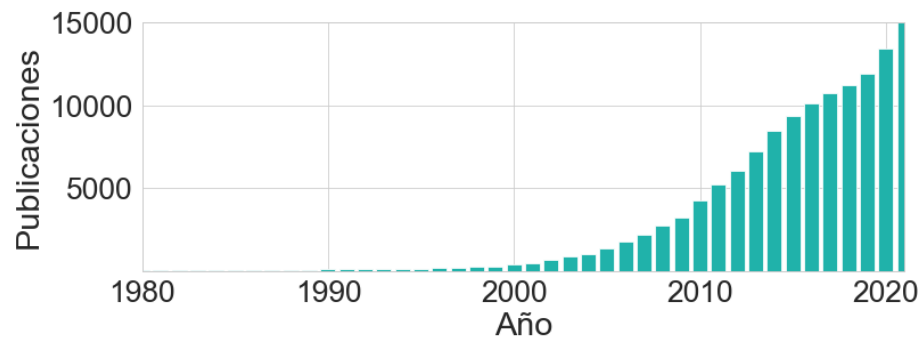


Figura 1.5: Publicaciones de PubMed con la palabra clave epigenética. Accesado el 10 de agosto de 2022.

La metilación del ADN tiene un papel importante en la comprensión de la herencia, el desarrollo y la progresión de las enfermedades. Algunos ejemplos de las enfermedades involucradas incluyen cáncer, diabetes, trastornos neurodegenerativos y enfermedades inmunomediadas.[3, 14, 25] Consecuentemente, el descubrimiento y desarrollo de epi-fármacos, también conocidos como fármacos epigenéticos, se ha convertido en un foco importante de investigación para múltiples empresas biotecnológicas. Esta importancia se refleja en las tendencias y el número creciente de publicaciones científicas sobre el tema en la plataforma de PubMed.[26] En los últimos 25 años se han publicado más de 100,000 artículos relacionados al tema; de los cuales 85,000 fueron publicados en los últimos 10 años (figura 1.5). El término coexistente más frecuentemente utilizado en la bibliografía es “cáncer”, el interés se debe a que a diferencia de los orígenes genéticos del cáncer, las aberraciones epigenéticas son eventos reversibles que ocurren en etapas tempranas de la génesis del tumor. En la última década, se han detectado muchas interacciones y conexiones entre los cambios genéticos y epigenéticos. Esto permite resaltar la compleja y multifactorial naturaleza de tal enfermedad. Además, poco se sabe ac-

erca del mecanismo que tienen las DNMTs en otros tipos de enfermedades lo que deja muchas áreas de oportunidad para realizar investigación en el tema.

Tabla 1.1: DNMTs en distintas enfermedades.

Enzima	Rol	Enfermedad
DNMT1	Hipermetilación	Cáncer Leucemia mieloide aguda (LMA), Leucemia mieloide crónica (LMC) y glioma, cáncer de: cuello uterino, colorrectal, pulmón, páncreas, gástrico, mama, esófago, vejiga, próstata, tiroides, ovario y carcinoma hepatocelular. [3, 14, 25] Sistema nervioso Neuropatía autonómica sensorial hereditaria con demencia y pérdida auditiva, ataxia cerebelosa, sordera y narcolepsia.[27] Autoinmune Artritis reumatoide y osteoartritis lupus.[11]
	Hipometilación	Cáncer de próstata y cáncer cervical.[3]
DNMT3A	Hipermetilación	Cáncer LMC, cáncer de: cuello uterino, mama, gástrico, próstata, ovario, huesos y testicular.[3] Autoinmune Artritis reumatoide y osteoartritis.[11]
	Hipometilación	Linfoma, LMA, cáncer de: mama, colorrectal y pulmón [3]
DNMT3B	Hipermetilación	Cáncer LMC, LMA, glioma, cáncer de: pulmón, mama, gástrico, colorrectal, próstata, páncreas, vejiga y cuello uterino.[3] Sistema nervioso Inestabilidad centromérica-anomalías faciales de inmunodeficiencia (ICF).[27]
	Hipometilación	LMA y cáncer de vejiga[3]

Debido a esto, las DNMTs se han convertido en un blanco de estudio importante en los últimos años.[11] Las enfermedades involucradas se ejemplifican en la tabla 1.1. Se ha observado que la expresión de DNMT1, DNMT3A y DNMT3B está elevada en la leucemia mieloide aguda (LMA) y varios cánceres sólidos. Por otro lado, en algunos cánceres solamente una de las DNMTs está involucrada.

Paralelamente, la disminución de las DNMT también puede conducir a la tumorigénesis.[3] En las células normales, la heterocromatina está hipermetilada alrededor del punto central; sin embargo, en muchos tumores este mecanismo se interrumpe y se pierde la metilación del ADN en regiones normalmente inactivas.[3, 28] Esto causa que elementos transponibles se reactiven y se integren en sitios arbitrarios del genoma, lo que lleva a la mutación y la inestabilidad genómica.[28] En particular, existen casos como el cáncer de próstata en donde se ha probado que una menor metilación acelera el proceso de metástasis.[13, 29]

Es por esto que es necesario un enfoque individual para cada enfermedad a tratar tomando en consideración las necesidades del paciente. Esto deja un amplio campo de investigación para el desarrollo de distintos fármacos. La DNMT1 demuestra ser un blanco con múltiples aplicaciones terapéuticas debido a su rol

en las distintas enfermedades mencionadas. Es por esto que será el objetivo de la investigación.

1.1.4 DNMT1

DNMT1 se compone de una región N-terminal reguladora grande (1139 AA) y una región C-terminal catalítica más pequeña (476 AA) (figura 1.6).[15, 27]

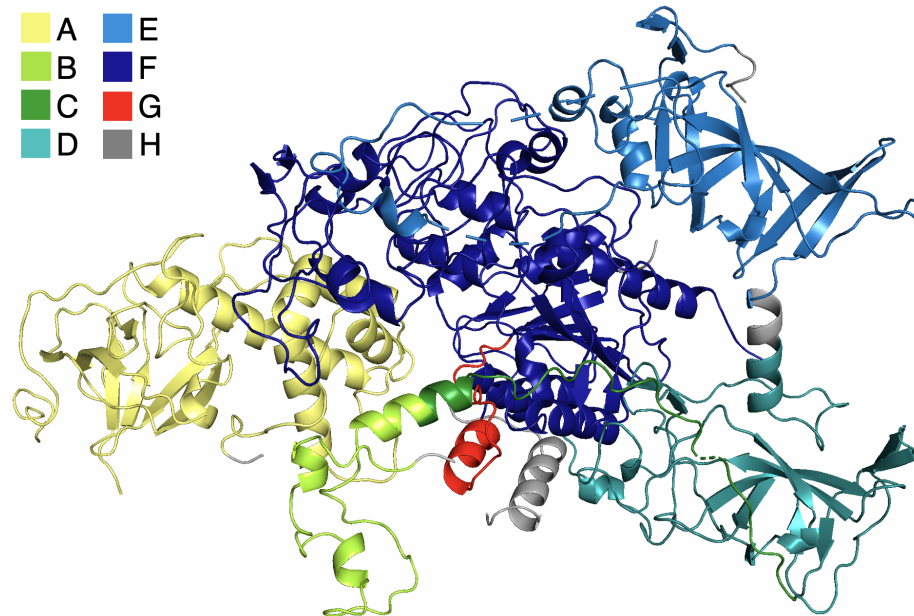


Figura 1.6: Dominios de DNMT1 (PDB ID: 4WXX) donde: A corresponde a RFTS (AA351-600), B a CXXC (AA651-697), C al enlazador autoinhibidor (AA697-728), D a BAH1 (AA728-900), E es BAH2 (AA906-1100), F es el dominio de metiltransferasa (AA1139-1616) G es el sitio catalítico (AA1224-1245) y H corresponde a los enlazadores entre los dominios.

Región N-Terminal

La región N-terminal de DNMT1 tiene tres funciones: (i) regulación alostérica de la actividad catalítica[30, 31], (ii) regulación de la actividad catalítica de DNMT1 mediante múltiples sitios de fosforilación y metilación[32–34] y (iii) la interacción entre DNMT1 y otras moléculas que afectan la metilación del ADN[35]. Esta región está compuesta por un dominio RFTS, un dominio de dedo de zinc CXXC (Cys-X-X-Cys) y un par de dominios de homología adyacente de bromo (BAH, por sus siglas en inglés, *bromo-adjacent homology*).[36]

En esta región terminal N se encuentran los sitios de unión al ADN. El primer sitio de unión al ADN está ubicado en la enzima dentro de la región 261–356; específicamente en los aminoácidos Lys 284, Lys 285, His 286 y Arg 287. El segundo

sitio de unión al ADN está ubicado dentro dominio del dedo zinc CXXC, en el segmento de aminoácidos 580–697.[31] Estos sitios de unión se ven inhibidos por el dominio RFTS el cual regula de manera alostérica la metilación por DNMT1 (figura 1.7). Bashtrykov y colaboradores realizaron una mutagénesis de sitio dirigida en donde se mutaron los aminoácidos que tenían interacción con el sitio catalítico y encontraron que esta mutación aumentaba la tasa de metilación de la DNMT1.[30] De manera paralela, Syeda y colaboradores encontraron que una variante DNMT1 humana (621–1616) sin el dominio RFTS el cual mostró una unión más fuerte al ADN que un fragmento de DNMT1 más largo (351–1616) que incluía el dominio RFTS.[24] Al comparar la actividad de metilación del ADN de estas dos variantes encontraron que la actividad de la DNMT1 sin RFTS (621–1616) era aproximadamente 40 veces mayor que el de la DNMT1 con RFTS (351–1616), lo que permite concluir que el dominio RFTS inhibe la actividad de la DNMT1.[24] Además, el enlazador de los dominios CXXC-BAH1 juega un papel importante en dos mecanismos autoinhibidores diferentes de DNMT1. El análisis estructural y los ensayos bioquímicos de hDNMT1 351–1600 revelan que el conector helicoidal entre el dominio CXXC y el dominio BAH1 constituye un elemento regulador importante en la autoinhibición mediada por el dominio RFTS.[36]

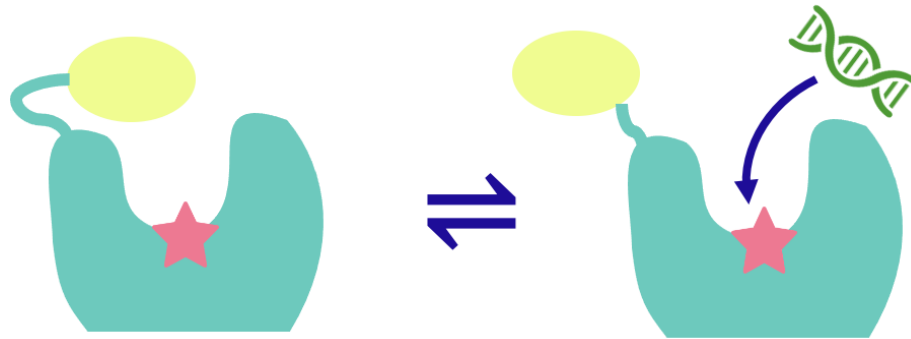


Figura 1.7: Regulación alostérica de DNMT1. El dominio RFTS (amarillo) tiene una función de autoinhibición, donde se une al dominio catalítico de la enzima (turquesa) esto bloquea el acceso del ADN a la hendidura de unión al ADN y al centro catalítico (rojo).

Por otro lado, la selectividad de la DNMT1 hacia la metilación *de novo* es mediada por el dominio CXXC.[36] En el caso de los sitios hemimetilados, el dominio CXXC no se une al ADN, por lo que el sitio activo está abierto para la unión al ADN y así llevar a cabo la metilación de mantenimiento. Por el contrario, en el caso del sustrato no metilado, el dominio CXXC de dedo de Zn se une al sitio CpG no metilado de modo que el bucle entre BAH1 y el dominio CXXC esté en una posición en la que pueda bloquear la unión del ADN en el sitio activo.[31]

En segundo lugar, se han encontrado más de sesenta sitios de fosforilación en la proteína DNMT1 humana y de ratón, aunque solo unos pocos de ellos se han

caracterizado funcionalmente. Se sabe que se requiere el S515 fosforilado ubicado dentro del dominio RFTS para preservar la interacción entre las regiones N-terminal y C-terminal de DNMT1 que es crucial para la actividad enzimática. También, la fosforilación de S146 reduce la afinidad de unión al ADN de DNMT1.[33]

Además, se encuentran las interacciones entre DNMT1 y otras moléculas; un ejemplo sería la interacción con la enzima UHRF1. Como se mencionó previamente (sección 1.1.2), la enzima UHRF1 se une al ADN hemimetilado lo que causa que el dominio RFTS se mueva y permita el acceso del ADN hemimetilado al sitio activo.[37] Otras de las enzimas que interactúan con esta región terminal son PCNA y MeCP2.[38, 39] Finalmente, una secuencia rica en lisina y glicina conecta la región N-terminal con la región catalítica C-terminal.[40] Esta secuencia interactúa con el dominio RFTS mediante puentes de hidrógeno formando interacciones entre R582, K586 con D700 y entre R582 y E703. También interactúa con el dominio de metiltransferasa mediante los residuos de D702 y N1233. Esto permite incrementar la asociación entre los dominios RFTS y de metiltransferasa.[36] La interacción alostérica adecuada de la región reguladora N-terminal con la región C-terminal catalítica es necesaria para la función enzimática y controla su preferencia por el ADN hemimetilado.[27]

Región C-Terminal

La región C-Terminal contiene el dominio de la metiltransferasa. DNMT1 necesita al menos tres elementos estructurales para la actividad catalítica: (i) un dominio de unión a SAM, (ii) una cavidad de unión y (iii) un dominio de reconocimiento de secuencia objetivo.[31]

En la DNMT1, el sitio de unión de SAM y la cavidad de unión se pueden encontrar dentro de la región C-terminal después de las repeticiones de Gly-Lys. El dominio de reconocimiento de secuencia objetivo no está definido con precisión por los motivos de secuencia conservados, y no se puede asignar con certeza en función de la secuencia de ADN. Históricamente, el dominio catalítico se define como la secuencia C-terminal después de las repeticiones de Gly-Lys. Este dominio catalítico no tiene actividad catalítica ya que se ha demostrado que si a DNMT1 le faltan los primeros 672 aminoácidos esta no tendrá actividad catalítica.[31] Sin embargo, en este dominio se encuentra la cisteína C1226 la cual interactúa con la citosina en el proceso de metilación.[36]

1.1.5 Moduladores ya conocidos y sus subestructuras

Debido al papel tan importante que juega la DNMT1 en el desarrollo de múltiples enfermedades se han buscado fármacos que puedan modular a la misma (figura 1.8). Los inhibidores de DNMT1 se clasifican principalmente en dos categorías: inhibidores nucleósidos y no nucleósidos (figura 1.9). Según su estructura y/o mecanismo de acción, los inhibidores también pueden dividirse en diferentes

grupos: (i) análogos de citosina, (ii) análogos de SAM, (iii) hidrazinas, (iv) ftalimidas, (v) polifenoles, (vi) derivados aminobenzoicos, (vii) oligonucleótidos modificados, (viii) oligos antisentido, y (ix) análogos de disulfuro.[31]



Figura 1.8: Mecanismo de los inhibidores de DNMT1.

Los inhibidores de nucleósidos generalmente se derivan de la citosina por lo que son los más conocidos. Estos análogos no pueden inhibir la DNMT1 directamente, sino que deben incorporarse al ADN. Su mecanismo consiste en atrapar a la DNMT1 en el proceso de metilación del ADN al formar un enlace covalente con la misma. Esto da como resultado una “inhibición suicida”, pero también un daño excesivo al ADN y en consecuencia una alta toxicidad. Los análogos de citosina son más efectivos a una concentración baja, lo cual representa un equilibrio entre la interferencia deseada con la DNMT y una toxicidad aceptable.[31, 41] La 5-azacitidina y la 5-aza-2'-desoxicitidina (decitabina) son análogos de la citosina con un átomo de nitrógeno en lugar de carbono en la quinta posición del anillo. El uso de esos compuestos para la terapia del cáncer es valioso para las neoplasias hematológicas, y tanto la decitabina como la 5-azacitidina recibieron la aprobación de la Administración de Drogas y Alimentos de los Estados Unidos (FDA, por sus siglas en inglés, *Food and Drug Administration*) para el tratamiento de los síndromes mielodisplásicos.[42]

Entre los inhibidores descritos se encuentra SGI-1027 el cual posee una gran actividad inhibitoria contra DNMT1 ($CI_{50} = 6,000$ nM). Además, también tiene actividad inhibitoria contra DNMT3A y DNMT3B así como la detención de la proliferación en células cancerosas.[11, 20, 41, 43, 44] Es por este caso de éxito que se ha buscado encontrar compuestos basados en la quinolina que tengan actividad inhibitoria contra DNMT1.

Con una estructura similar se encuentran las 4-quinazolaminas como UNC046,

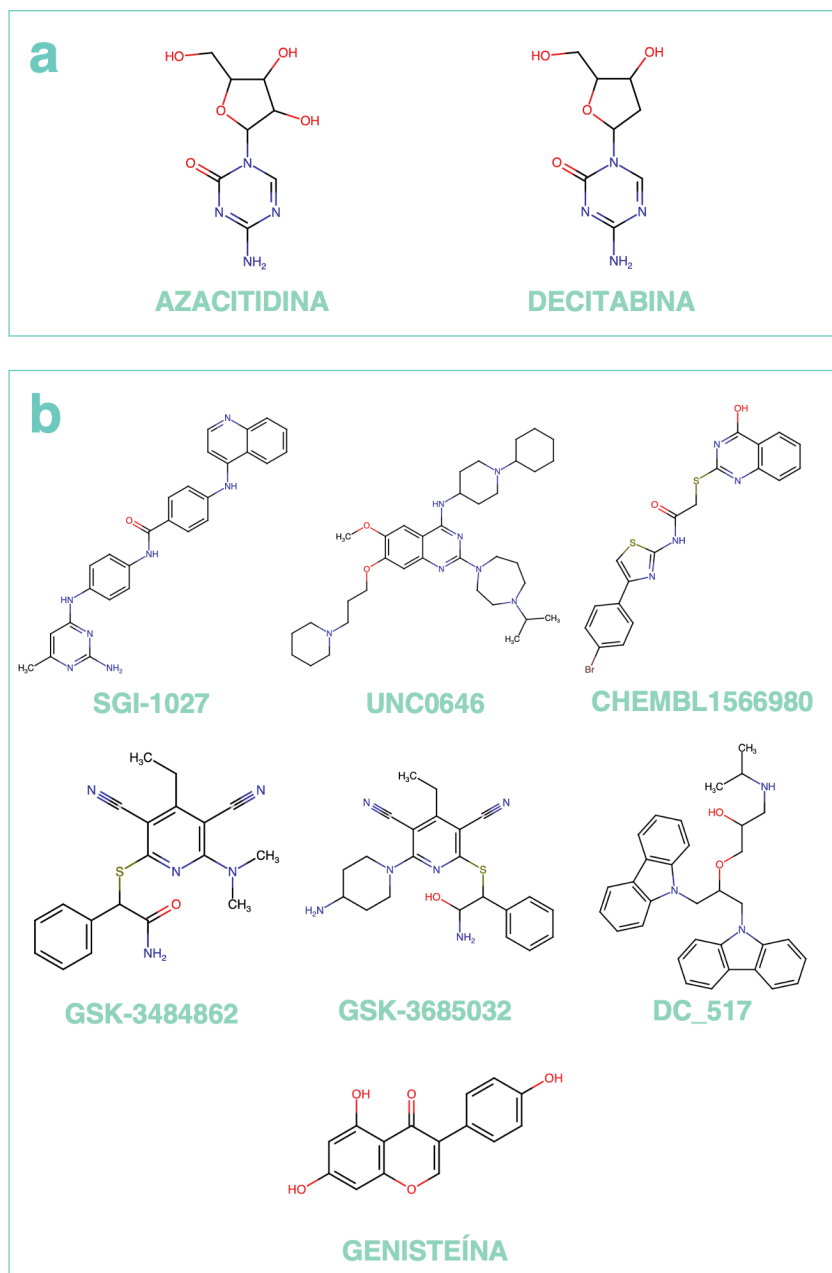


Figura 1.9: Ejemplos de inhibidores de DNMT1 donde **a** corresponde a la categoría de los nucleosídicos y **b** a la categoría de los no nucleosídicos.

la cual tiene un CI_{50} de 30nM. En estudios previos se ha reportado que este compuesto tiene interacciones con la SER 582, ASP 764 y SER1292 de la DNMT1 en presencia de su cofactor SAM.[45] Además, el compuesto CHEMBL1566980 tiene un CI_{50} de 5120nm; esto marca un precedente en el potencial de los compuestos con esta subestructura.[46]

Por otro lado, las 2-(bencilsulfanil)piridinas como GSK3484862 y GSK-3685032 han reportado tener una inhibición reversible a la DNMT1 lo que sugiere que no se unen de manera covalente a la DNMT1.[41] A su vez, ya se han estudiado en líneas celulares de leucemia y estos compuestos han logrado reducir la metilación teniendo GSK-3685032 un CI_{50} de 36 nM y GSK3484862 un CI_{50} de 400 nM. Incluso, GSK-3685032 ya se ha probado en modelos de xenotrasplante de leucemia en ratones.[41, 47] Los xenotrasplantes son el trasplante de células, tejidos u órganos entre especies filogenéticamente diferentes.[48] Por otro lado, GSK-3484862 ha logrado causar una desmetilación en las células madre embrionarias murinas con una baja toxicidad funcionando así mejor que los 5-azanucleósidos.

Finalmente, se encuentran los compuestos con carbazol como el DC_517. Este compuesto inhibe a la DNMT1 con un CI_{50} de 1,700 nM. A su vez, se ha reportado que este induce la muerte apoptótica en células HCT116.[44] Se cree que este tiene actividad debido a que el sitio de unión del SAM tiene una gran afinidad a los grupos hidrofóbicos.

Hasta la fecha, los únicos moduladores aprobados por la FDA son inhibidores que pertenecen a la familia de análogos de nucleósidos. No obstante, estos muestran efectos secundarios relevantes, incluida la toxicidad celular y clínica, así como su inestabilidad química. Por lo tanto, se busca identificar fármacos que permitan modular tanto la activación como la inhibición de la DNMT1 sin estos efectos secundarios.

1.1.6 Bibliotecas moleculares (quimiotecas) enfocadas

Además de las bibliotecas de detección general tradicionales disponibles como ZINC[49], las empresas químicas han creado bibliotecas centradas en objetivos epigenéticos también llamadas quimiotecas enfocadas o bibliotecas moleculares enfocadas. En estas las estructuras químicas están disponibles al dominio público para que puedan ser estudiadas por medio de métodos de quimioinformática y pruebas experimentales.[50]

En la anterioridad, Green y colaboradores demostraron el poder que poseen estas quimiotecas enfocadas. Estos observaron una mejora en la tasa de aciertos usando la biblioteca de fragmentos enfocada en comparación con una de compuestos disponibles comercialmente contra el bromodominio ATAD2 (13% frente a 1%, respectivamente). Esto ejemplifica que las bibliotecas de compuestos enriquecidas tienen una relevancia significativa por lo que el desarrollo y la aplicación de las quimiotecas enfocadas debe ampliarse.[51]

Por diseño, las bibliotecas enfocadas en epigenética tienen el potencial de aumentar el espacio químico epigenético relevante.[50] Debido a la utilidad de estas en los últimos años se han desarrollado esfuerzos para diseñar y analizar las bibliotecas enfocadas.

1.1.7 Diseño de fármacos asistido por computadora

El diseño de fármacos asistido por computadora (DIFAC) tiene como objetivo el diseño, optimización y selección de compuestos con actividad biológica. El DIFAC forma parte de un esfuerzo multidisciplinario y está conformado por una serie de disciplinas científicas que abarcan modelado molecular, quimioinformática, química teórica y química computacional.[52]

En 1988, Brown describió la quimioinformática como la combinación de todos los recursos de información que un científico necesita para optimizar las propiedades de un ligando para convertirse en un fármaco.[53] En los últimos años, esta se ha ampliado hasta convertirse en un campo de la tecnología de la información que utiliza computadoras y programas informáticos para facilitar la recopilación, el almacenamiento, el análisis y la manipulación de grandes cantidades de datos químicos. Los datos químicos incluyen fórmulas químicas, estructuras químicas, propiedades químicas, espectros químicos y actividades bioquímicas o biológicas. Los métodos empleados en el DIFAC pueden clasificarse como basados en ligandos (BL) y basados en estructuras (BS).[54]

El uso de computadoras en el campo de la investigación del diseño de fármacos es esencial, ya que permite el uso de técnicas novedosas y eficientes que reducen el tiempo y el costo involucrados en el descubrimiento de fármaco.[55] Por ejemplo, diferentes bibliotecas son apropiadas para identificar compuestos exitosos con estructuras químicas novedosas para un objetivo determinado, como las DNMTs. El análisis quimioinformático de bibliotecas químicas ha encontrado varias aplicaciones no solo en el área de investigación de bibliotecas relacionadas con la epigenética discutidas en este trabajo, sino también en el descubrimiento de fármacos basados en productos naturales, bibliotecas combinatorias y productos químicos relacionados con los alimentos.[56]

Espacio químico

La caracterización quimioinformática de las bibliotecas químicas es un paso crucial hacia el cribado virtual o experimental para identificar nuevas moléculas con actividad biológica.[56] El análisis de la diversidad estructural y el análisis de la distribución en el espacio químico de las bases de datos de compuestos es un paso crucial en la detección de compuestos activos.[56] Las modificaciones químicas de las estructuras moleculares pueden modular sus propiedades químicas y biológicas y su distribución en el espacio químico. Algunas de las cosas a estudiar para el espacio químico son: las propiedades fisicoquímicas, los conteos de elementos estructurales, las huellas digitales moleculares y el núcleo estructural base.[57]

Las propiedades fisicoquímicas definen el comportamiento de la molécula frente a condiciones externas.[58] Estas propiedades buscan describir aspectos de las estructuras moleculares mediante el uso de variables numéricas.[59] Las seis propiedades de interés farmacéutico más utilizadas para las bases de datos son: coeficiente de

partición octanol/agua (SlogP), enlaces rotables (RB), donadores de enlaces de hidrógeno (HBD, por sus siglas en inglés, *Hydrogen bond donor*), aceptores de enlaces de hidrógeno (HBA, por sus siglas en inglés, *Hydrogen bond acceptor*), área superficial topológica (TPSA) y peso molecular (MW, por sus siglas en inglés, *Molecular weight*). Este conjunto de propiedades se usa comúnmente para comparar colecciones de compuestos para el descubrimiento de fármacos.[56]

Una de las herramientas que se utilizó en este proyecto es *Platform for Unified Molecular Analysis* (PUMA) ya que permite realizar el cálculo de estas propiedades. PUMA es un servidor en línea que integra, en una sola plataforma, varios métodos y análisis para caracterizar las bases de datos químicas proporcionadas por el usuario en términos de propiedades moleculares de interés farmacéutico, contenido y diversidad de andamios, diversidad basada en huellas digitales moleculares y representación visual del espacio químico.[60]

Por otro lado, los *fingerprints* o las huellas digitales moleculares, codifican las características estructurales de las moléculas.[61] Estas son herramientas quimiinformáticas esenciales para el cribado virtual y el mapeo del espacio químico. Entre los diferentes tipos de huellas digitales moleculares, las huellas digitales moleculares de la subestructura son principalmente usadas para moléculas pequeñas como los fármacos, mientras que las huellas digitales moleculares de pares de átomos son preferibles para moléculas grandes como los péptidos. Sin embargo, ninguna huella digital molecular disponible logra un buen desempeño en ambas clases de moléculas.[61]

Una de las huellas digitales moleculares más utilizadas en la actualidad es la huella digital de Morgan; esta es también conocida como huella digital de conectividad extendida (ECFP, por sus siglas en inglés, *Extended-Connectivity Fingerprint*).[62] ECFP pertenece a las huellas digitales de mejor rendimiento en el cribado virtual de moléculas pequeñas y los puntos de referencia de predicción de blancos moleculares.[59] Esta percibe la presencia de subestructuras circulares específicas alrededor de cada átomo en una molécula; sin embargo, tiene una mala percepción de las características globales de las moléculas, como el tamaño y la forma.[61] Los ECFP son huellas digitales circulares con una serie de cualidades útiles.[59] Algunas de estas son:

1. Se pueden calcular muy rápidamente.
2. No están predefinidas y pueden representar una gran cantidad de características moleculares diferentes (incluida la información estereoquímica).
3. Sus características representan la presencia de subestructuras particulares, lo que permite una interpretación más fácil de los resultados del análisis.
4. Están diseñadas para representar tanto la presencia como la ausencia de funcionalidad, ya que ambas son cruciales para analizar la actividad molecular.
5. El algoritmo ECFP se puede adaptar para generar diferentes tipos de huellas

digitales circulares, optimizadas para diferentes usos.

Finalmente se encuentran los núcleos estructural base también conocidos como *scaffolds* o andamios estructurales, la definición de estos varía de acuerdo al problema a resolver. En el caso de las bibliotecas moleculares enfocadas frecuentemente se encuentra una representación densa de un pequeño número de núcleos. [63]

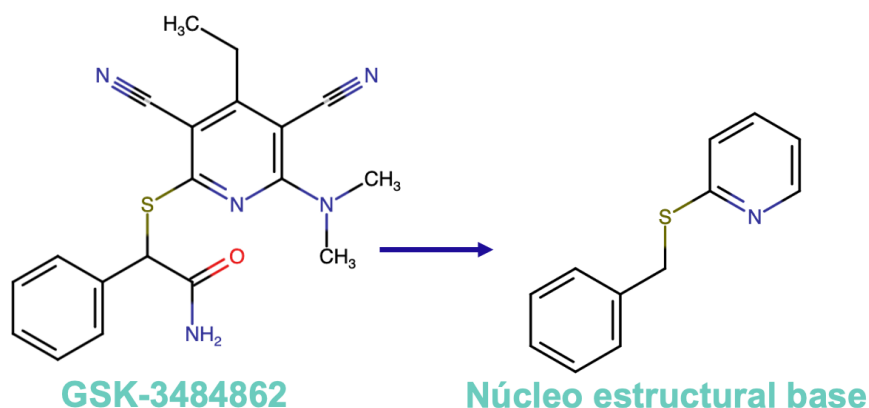


Figura 1.10: Formación de núcleo estructural base de Murcko de GSK-3484862.

Uno de estos métodos es el núcleo estructural base de Murcko (figura 1.10), propuesto por Bemis y Murcko en 1996, que se ha utilizado para analizar las estructuras de fármacos conocidos. [63] El método divide las moléculas en sistemas de anillos, enlazadores, átomos de cadenas laterales y el marco, que es la unión de sistemas de anillos y enlazadores en una molécula. Un núcleo estructural base Murcko retiene información sobre el tipo de átomo, mientras que un marco gráfico reduce todos los átomos a carbono y todos los enlaces a enlaces simples. [63, 64]

Métodos basados en ligando

Algunos métodos de diseño de fármacos basados en ligando (DIFBL) miden la similitud de los compuestos de la biblioteca con los compuestos de referencia que son activos hacia un blanco de interés o muestran las propiedades deseadas. [54] Estos modelos de relación estructura-actividad (SAR, por sus siglas en inglés, *structure-activity relationship*) están basados en el principio de similitud el cual establece que los compuestos similares tienen propiedades similares. [65, 66] Esto implica que compuestos con una gran similitud estructural con los compuestos de referencia se comportarán de manera similar. Ha habido un aumento significativo en el desarrollo de modelos QSAR por medio de DIFBL que evalúan la actividad terapéutica de moléculas pequeñas, uno de estos es por medio del aprendizaje automático. [67]

El aprendizaje automático (ML por sus siglas en inglés, *Machine learning*) es una rama de la inteligencia artificial (IA) que tiene como objetivo desarrollar y

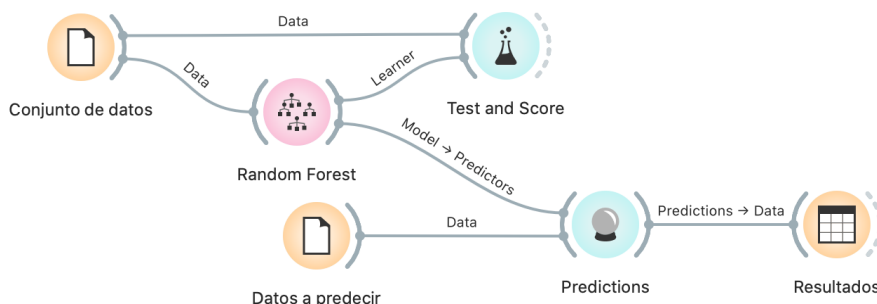


Figura 1.11: Metodología de un método de aprendizaje automático Orange3. Un conjunto de datos está unido a una herramienta de un modelo de RF el cual genera un modelo que permite predecir la actividad de los compuestos mediante una herramienta de predicciones.

aplicar algoritmos informáticos que aprenden de datos sin procesar, para luego realizar una tarea específica.[59, 68] Esta rama se basa en la idea de que las computadoras pueden aprender a predecir a partir de datos sin programación manual. Estos métodos se basan en algoritmos computacionales que relacionan todos o algunos elementos de un conjunto de variables predictoras con un resultado.[69] Los algoritmos generados aprenden de los datos de entrenamiento para hacer predicciones o decisiones y son capaces de mejorarse a sí mismos independientemente de la intervención humana, teniendo como propósito extraer conocimiento de los datos (figura 1.11).[70–72]

Las principales tareas realizadas por los algoritmos de IA son la clasificación, la regresión, la agrupación o el reconocimiento de patrones dentro de un gran conjunto de datos.[59] La tarea a realizar por los modelos depende de los tipos de variables que están disponibles para la fase de aprendizaje. Si el resultado de interés es una variable cuantitativa, se conoce como modelo de regresión, mientras que, para tareas cualitativas o categóricas, se conoce como modelo de clasificación.[69]

Existe una gran variedad de métodos de aprendizaje automático que se han utilizado en la industria farmacéutica para la predicción de nuevas características moleculares, actividades biológicas, interacciones y efectos adversos de fármacos (figura 1.12). Algunos ejemplos de estos métodos son *k-Nearest neighbors*, *Support vector machines*, *Random forest* y, más recientemente, *Deep neural networks*. [59] Sin embargo, el más utilizado es *Random forest*.

Random forest (RF) es un método caracterizado por tener una combinación de árboles de decisión. En RF se utiliza el método de *ensemble* de *bagging* en donde se ajustan múltiples modelos, cada uno con un subconjunto distinto de los datos de entrenamiento. Para predecir, todos los modelos que forman el agregado participan aportando su predicción. Como valor final, se toma la media de todas

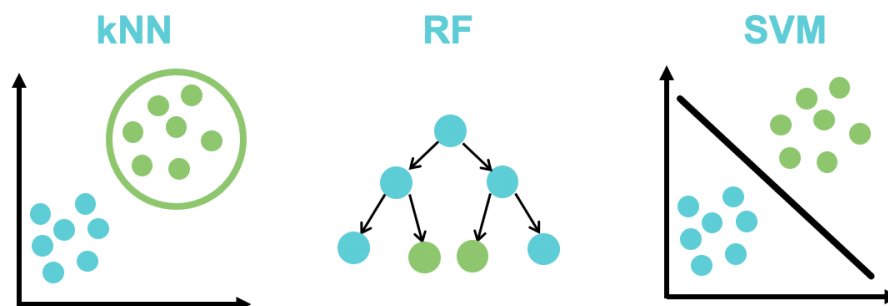


Figura 1.12: Métodos de aprendizaje automático: *k-nearest neighbors* (kNN), *Random forest* (RF) y *Support vector machines* (SVM). Los distintos colores en los círculos representan distintas clasificaciones de los datos, para kNN se utiliza la proximidad de los datos, para RF se forman árboles de decisión y para SVM se generan hiperplanos.

las predicciones o la clase más frecuente.[72] Independientemente del problema a resolver, RF es uno de los mejores modelos en términos de rendimiento y velocidad es por esto que este se ha convertido en uno de los algoritmos más utilizados en ML.[59]

La utilidad de *Random forest* ha sido reportada en la literatura en múltiples ocasiones; una de ellas fue el proyecto DREAM. A partir de experimentos realizados por AstraZeneca de 910 combinaciones de medicamentos en 85 líneas celulares de cáncer caracterizadas molecularmente, 160 equipos internacionales intentaron predecir las mejores sinergias entre pares de fármacos y biomarcadores para los que se utilizaron diferentes modelos: SVM, MKL, RF, árboles de decisión o ANN. El equipo ganador, dirigido por Yuanfang Guan, utilizó una combinación de modelos de RF.[59, 73]

Como se refleja previamente, el aprendizaje automático ha ganado una inmensa popularidad en la industria farmacéutica desde su aparición.[70] Esto se debe a que brinda la posibilidad de automatizar procesos que de otro modo consumirían una cantidad considerable de tiempo y recursos, lo que hace que la atención médica accesible y eficiente sea un objetivo más realista. La aplicación de estos métodos comenzó en 1964 cuando Hansch y colaboradores propusieron la ecuación de Hansch.[59, 74] Era un modelo de regresión lineal que utilizaba descriptores fisicoquímicos (como el parámetro de hidrofobicidad, el parámetro electrónico y el parámetro estérico), utilizados para describir la relación estructura-actividad 2D. Así inició la aplicación del aprendizaje automático al descubrimiento de fármacos.

En el campo del descubrimiento de fármacos, los métodos computacionales se han aplicado en muchas áreas, incluida la evaluación de la similitud de compuestos, la clasificación de moléculas y la predicción de bioactividad.[75, 76] Me-

dante la integración del aprendizaje automático se han explorado otros intentos de producir compuestos similares a fármacos seguros y eficaces de forma indirecta, por ejemplo, mediante la identificación de blancos en lugar de concentrarse en las propiedades de las moléculas pequeñas.[70] Numerosos estudios han permitido la clasificación de blancos macromoleculares prometedores, incluso ayudando en el descubrimiento de medicamentos ahora aprobados por la FDA.[77, 78] Uno de estos estudios involucró el uso de modelos para identificar diferentes tipos de sitios de metilación de ácido ribonucleico (ARN) en diferentes especies.[79] No obstante, la aplicación de métodos computacionales depende ampliamente de la información obtenida a través de métodos experimentales. Es por esto que es importante la integración entre estos dos tipos de métodos.

Las soluciones basadas en el aprendizaje automático pueden apoyar la toma de decisiones sobre la aplicación de los recursos, asegurando la eficiencia y la rentabilidad.[69] Estas características permiten que el principal potencial en la industria farmacéutica sea reducir costos y aumentar la eficiencia.[80] Esto se debe a que la aplicación de ML reduce significativamente la carga experimental y los plazos que se dedican actualmente a la caracterización de la respuesta farmacológica *in vitro* e *in vivo*. [67] Algunas de las herramientas utilizadas para generar estos modelos son *Epigenetic Target Profiler* y *Orange 3*. [75, 81]

Epigenetic Target Profiler

Epigenetic Target Profiler (ETP) es una aplicación web gratuita y fácil de usar que se puede acceder desde <http://www.epigenetictargetprofiler.com/>. Este programa se utiliza para la predicción del perfil de bioactividad de moléculas pequeñas en un panel de 55 blancos epigenéticos. ETP implementa el modelo de mejor rendimiento para la predicción de objetivos epigenéticos, según se identifica a partir de una comparación sistemática de modelos de aprendizaje automático basados en huellas digitales moleculares. [75]

Orange 3

Orange 3 es un conjunto de herramientas de visualización de datos, aprendizaje automático y minería de datos de código abierto. Cuenta con una interfaz de programación visual para el análisis exploratorio rápido de datos cualitativos y la visualización interactiva de datos. [81]

Métodos basados en estructura

En el marco de DIFAC, los métodos de diseño de fármacos basados en estructuras (DIFBS) aprovechan la abundancia de estructuras en *Protein Data Bank* (PDB), que posiblemente también se pueden usar como plantillas para modelos de homología si falta la estructura de interés. [54]

El DIFBS se basa en la premisa de que el conocimiento de la estructura diana puede ayudar a racionalizar y optimizar la unión, ya que las interacciones ligando-

diana están mediadas por su complementariedad. Con la evolución de los modelos de unión, está claro que hablar de “estructura diana” es una aproximación, dado que las proteínas fluctúan entre un conjunto de estructuras.[54] Un ejemplo de esto serían las distintas estructuras que tiene la DNMT1 durante la inhibición del RFTS (sección 1.1.4).

Desde su primera aparición a mediados de la década de 1980, el acoplamiento ha demostrado ser un componente importante de los proyectos de descubrimiento de fármacos, ya sea en la academia o en la industria. Las tres razones principales son: los avances en biología estructural, el desarrollo de las computadoras y la creciente disponibilidad de estructuras 3D de proteínas y quimiotecas.[82]

El acoplamiento molecular es una de las herramientas más utilizadas en el diseño de fármacos asistido por computadora que simula las interacciones moleculares entre dos moléculas y, por lo general, predice el modo de unión y la afinidad entre ellas. En otras palabras, el acoplamiento describe el ajuste de dos moléculas en un espacio tridimensional (3D).[55, 82] Las técnicas de acoplamiento molecular consisten en la generación de una serie de posibles conformaciones, también llamados modos de unión, del ligando dentro del sitio de unión a la proteína. Por ello, la disponibilidad de la estructura tridimensional de la diana molecular es una condición necesaria; puede ser una estructura resuelta experimentalmente (cristalografía de rayos X o RMN) o una estructura obtenida por técnicas computacionales (como el modelado por homología).

El acoplamiento molecular se compone principalmente de dos etapas: un programa para el muestreo de conformaciones/orientaciones y una función de puntuación, que asocia una puntuación a cada modo de unión predicho. El proceso de muestreo debe buscar de manera efectiva el espacio conformacional descrito por el panorama de energía libre, donde la energía en acoplamiento se aproxima mediante la función de puntuación.[54]

El acoplamiento molecular también se ha utilizado para generar farmacóforos de inhibidores de DNMT utilizando 14 inhibidores de DNMT1 bien establecidos, entre los que se encuentran 5-azacitidina, decitabina y partenolida.[55] Los resultados sugirieron que las interacciones con GLU 1265, ARG 1311, ARG 1461, SER 1229 y GLY 1230 son importantes, y que son necesarias las interacciones con una carga negativa, un aceptor de enlaces de hidrógeno, un anillo aromático y dos donantes de enlaces de hidrógeno. En consecuencia, el acoplamiento molecular permite dilucidar los mecanismos de acción de los moduladores de DNMT permitiendo así establecer las relaciones estructura-actividad (SAR).[54, 55]

Algunos de los programas utilizados para realizar estudios de acoplamiento molecular son: AutoDock Vina, LeDock y MOE.[62, 83]

AutoDock Vina

AutoDock Vina es un programa de código abierto para hacer acoplamiento molecular. Originalmente fue diseñado e implementado por el Dr. Oleg Trott en el

Laboratorio de Gráficos Moleculares en el Instituto de Investigación Scripps.[83] Es una nueva generación de software de acoplamiento para realizar el acoplamiento o la detección virtual. Es más rápido y mejoró significativamente la precisión de la predicción del modo de enlace en comparación con AutoDock. AutoDock Vina implementa el algoritmo de búsqueda basado en Monte Carlo.[84]

LeDock

LeDock es un programa de código abierto por LePhar.[62] Este programa exhibe un rendimiento sobresaliente en una evaluación exhaustiva reciente de los programas de acoplamiento en un conjunto diverso de complejos de proteína-ligando. Las ventajas de este software es la rapidez y la exactitud.[85]

Molecular Operating Environment

Molecular Operating Environment (MOE) es un programa comercial que permite entre muchas cosas realizar acoplamiento molecular de grandes conjuntos de datos. Una de las modalidades de este programa es el uso del acoplamiento molecular de ajuste inducido.

La aplicación de enfoques computacionales en el descubrimiento de fármacos epigenéticos es todavía reciente y hasta ahora no se han descubierto o desarrollado epi-fármacos puramente con estrategias computacionales. Por lo tanto, el desarrollo y la aplicación de nuevas metodologías de epi-informática es un área de investigación en constante crecimiento. Consecuentemente, se han desarrollado un gran número de herramientas computacionales aplicadas al análisis de los espacios químicos y biológicos.[60, 76]

1.2 Planteamiento del problema

Hasta la fecha los fármacos aprobados como moduladores de DNMT1, la 5-azacitidina y la 5-aza-2'-desoxicitidina, tienen una alta toxicidad debido a su unión al ADN.[86] De esto deriva el gran interés en la investigación de inhibidores que puedan emplear la misma actividad sin los efectos adversos. No obstante, se presentan grandes dificultades en el desarrollo de moduladores ya que cada enfermedad está asociada a distintas dianas. Aunado a esto, existen dificultades en la investigación de moduladores de DNMT1. Como se mencionó previamente, las enfermedades pueden estar asociadas a una o varias de las DNMT. También, se ha encontrado que dependiendo de la etapa de la enfermedad se requiere una distinta modulación de la actividad. Por ejemplo, en el caso del cáncer se deben tener en cuenta tres aspectos: (1) el mismo gen puede desempeñar diferentes funciones en la carcinogénesis en diferentes neoplasias, (2) la misma modificación en un locus de gen específico puede tener distintas implicaciones en diferentes tejidos, y (3) dentro del mismo modelo de neoplasia, diferentes alteraciones epigenéticas pueden conducir a diferentes subtipos de cáncer.[20] Por otro lado, en las enfermedades neurológicas y autoinmunes todavía hay muchos mecanismos que no son

entendidos lo cual dificulta el tratamiento.

Finalmente, todavía se presentan múltiples desafíos en cuanto a la aplicación de la quimioinformática en el desarrollo de fármacos. Algunos de los problemas que se presentan son la dificultad del manejo de datos debido a las distintas representaciones que se les puede dar a las moléculas.[80] Esto ocasiona que hasta la fecha no haya una metodología homogénea que permita comparar los resultados de distintas investigaciones. Esto dificulta el entendimiento de los datos experimentales y su integración en métodos computacionales como el aprendizaje automático. En consecuencia, esto provoca que haya escepticismo sobre los resultados obtenidos mediante el aprendizaje automático en la industria farmacéutica y que no haya financiamiento para su desarrollo.

Se considera que el problema a resolver con esta investigación involucra los siguientes aspectos: la búsqueda de un modulador de DNMT1 que no tenga los efectos secundarios que tienen sus predecesores (sección 1.1.5). La implementación de una metodología que cubra distintos aspectos tanto los métodos basados en ligando como los métodos basados en estructura. Esto se logrará mediante la integración de resultados experimentales en el desarrollo de los métodos computacionales.

1.3 Hipótesis

Las quimiotecas enfocadas en blancos epigenéticos están enriquecidas con compuestos potencialmente activos, por lo que se teoriza que los compuestos encontrados en estas serán moduladores de DNMTs.

1.4 Objetivos

1.4.1 Objetivo general

Identificar moduladores de DNMT1 mediante el uso de quimiotecas enfocadas y herramientas quimioinformáticas.

Objetivos específicos

Cuantificar la diversidad estructural de las quimiotecas enfocadas en las DNMTs disponibles de manera gratuita.

Para llevar esto a cabo se realizó una búsqueda exhaustiva de quimiotecas enfocadas en compuestos que sean posibles inhibidores de DNMT1. Posteriormente, se cuantificaron las características de: peso molecular, átomos aceptores de hidrógeno, átomos donadores de hidrógeno, coeficiente de partición, área topológica superficial y la cantidad de enlaces rotables. Finalmente, se utilizaron estas propiedades para estudiar el espacio químico cubierto por dichas bases de datos.

Proponer una metodología completa que logre abarcar métodos basados en ligando y en estructura.

Para llevar a cabo esto se emplearon metodologías de aprendizaje automático incluyendo la aplicación web: *Epigenetic Target Profiler* así como un modelo generado en Orange 3. Asimismo, se utilizaron distintos programas para realizar estudios de acoplamiento molecular. Este objetivo busca delimitar la cantidad de compuestos a estudiar para las pruebas enzimáticas.

Identificar compuestos con actividad moduladora de DNMT1 mediante ensayos de inhibición enzimáticas.

Se realizaron pruebas enzimáticas con SAM marcado con tritio. Esto permitió evaluar la capacidad de los compuestos de actuar como moduladores de la metilación por DNMT1.

1.5 Contribución

Se encontró que de los diecinueve compuestos evaluados en ensayos bioquímicos, once aumentaron la actividad de metilación de la DNMT1. Estos once compuestos resultan ser los primeros activadores históricamente reportados de este blanco. Este descubrimiento da pie a la posibilidad de que estos compuestos se continúen investigando para el posible tratamiento de enfermedades como el cáncer de próstata, donde hay una menor cantidad de metilación (sección 1.1.3). Aunado a esto, los resultados permiten aportar nuevo conocimiento del espacio químico lo que permitiría desarrollar nuevos modelos de aprendizaje automático de las relaciones de estructura-actividad para futuras investigaciones.

La metodología implementada utiliza distintos métodos de cribado virtual como los basados en ligando y basados en estructura que han sido de gran utilidad en el desarrollo de fármacos, por lo que, se utilizó para lograr los objetivos planteados.[55, 59, 73] Debido a que la mayoría de los programas utilizados son de acceso libre se abre la posibilidad de que se planteen metodologías que sean replicables de manera gratuita. El que haya este tipo de metodologías permite que los resultados puedan utilizarse de manera integral y facilita las interpretaciones de los mismos.

Finalmente, se utilizaron quimiotecas enfocadas en blancos epigenéticos las cuales ya han tenido un papel en el descubrimiento de fármacos líderes.[51] El uso de estas quimiotecas enfocadas permite facilitar el cálculo realizado al poder trabajar con un menor número de compuestos, disminuyendo así el costo computacional.

Capítulo 2

Metodología

Para realizar la metodología se siguió el esquema descrito en la figura 2.1.

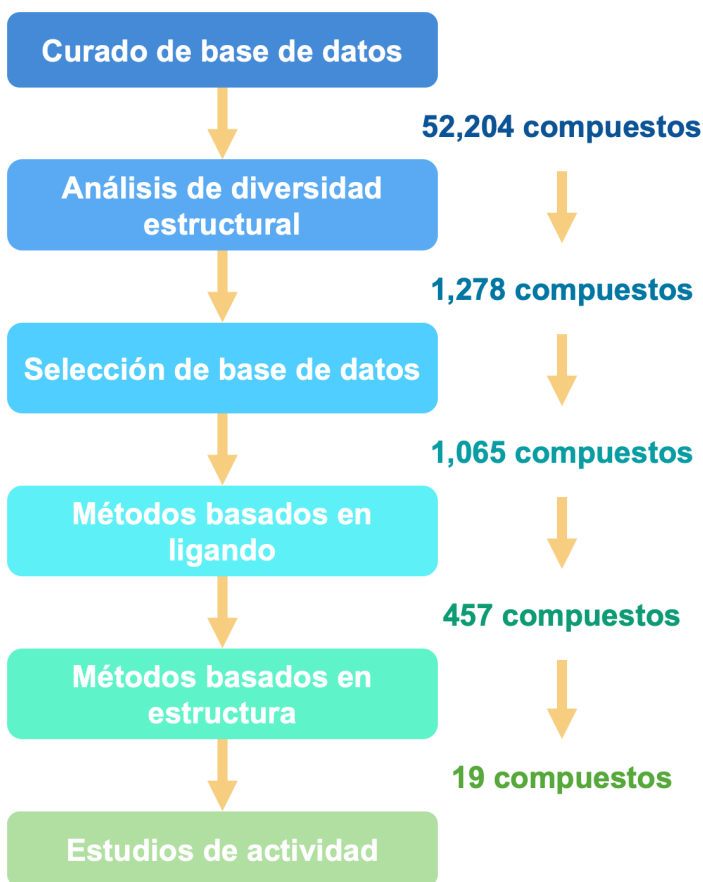


Figura 2.1: Esquema de la metodología seguida.

2.1 Formación y curado de las bases de datos moleculares

Se realizó una búsqueda de base de datos enfocadas a blancos epigenéticos que el acceso a los archivos con las estructuras estuviesen disponibles de forma gratuita para después poder hacer la compra de los compuestos. Posterior a eso se seleccionaron diez bases de datos enfocadas: Asinex, Axon, Chemdiv, Enamine, Enzo, Life Chemicals, OTAVA_DNMT1, OTAVA_DNMT3, Selleck y Targetmol. Después del curado de la base de datos se determinó la cantidad de compuestos (Tabla 2.1).

Tabla 2.1: Descripción de las quimiotecas enfocadas.

Base de datos	Descripción	Compuestos	Referencia
Asinex	Compuestos epigenéticos disponibles seleccionados de nuestras bibliotecas Signature & BioDesign.	5391	[87]
Axon	Colección única de 125 compuestos de alto perfil para la investigación epigenética.	182	[88]
ChemDiv	Biblioteca diseñada para la modulación de las tres clases de epiojetivos (es decir, lectores, escritores y borradores).	25883	[89]
Enamine	Biblioteca diseñada mediante una combinación métodos basados en ligando y estructura.	9352	[90]
Enzo	Compuestos con actividad contra enzimas epigenéticas.	43	[91]
Life Chemicals	Biblioteca centrada en las epienzimas relacionadas con la metilación.	8091	[92]
OTAVA_DNMT1	Biblioteca obtenida con un enfoque basado en la estructura hacia DNMT1.	401	[93]
OTAVA_DNMT3	Biblioteca obtenida con un enfoque basado en la estructura hacia DNMT3B.	1230	[94]
Selleck	Biblioteca enfocada en blancos epigenéticos.	699	[95]
TargetMol	Conjunto de epirreguladores cuyo foco principal es la optimización de leads y HTS.	932	[96]

Posteriormente, se realizó una selección de estas y se llevó a cabo un proceso de curado de las bases utilizando una metodología anteriormente creada por el grupo de investigación DIFACQUIM [97]. Este proceso se llevó a cabo mediante el uso de un código escrito en Python 3, el cual tuvo como objetivo estandarizar los SMILES, que son representaciones lineales de las moléculas. Primero el código descarta a las moléculas que tengan errores en la valencia o que tengan elementos diferentes a H, B, C, N, O, F, Si, P, S, Cl, Se, Br y I. Después, los compuestos se neutralizan y se reionizan para obtener el tautómetro apropiado.

Este proceso se lleva a cabo utilizando las funciones obtenidas de la paquetería de RDKit versión 2022.03.1: “Standardizer”, “LargestFragmentChoser”, “Uncharger”, “Reionizer”, “TautomerCanonicalizer” y “RemoveStereochemistry”. Finalmente se eliminaron los compuestos duplicados de la base de datos. Estos pasos permitieron obtener una base de datos unificada para poder continuar con el cálculo de las propiedades moleculares.

2.2 Cuantificación de la diversidad estructural

Para realizar un análisis de la diversidad estructural se utilizó un código en Python 3 el cual mediante la documentación de RDKit permite calcular los descriptores moleculares de: MV, TPSA, RB, SlogP, HBD y HBA. El módulo utilizado de RDKit es `RDKit.Chem.rdMolDescriptors` el cual contiene funciones que permiten calcular los descriptores moleculares utilizados: “CalcCrippenDescriptors”, “CalcTPSA”, “CalcExactMolWt”, “CalcNumRotatableBonds”, “CalcNumLipinskiHBD” y “CalcNumLipinskiHBA”.

Se analizó que los compuestos siguieran las reglas de Lipinski[98] y Veber[99] las cuales indican que para que un compuesto sea adecuado para consumo oral tiene que cumplir las siguientes características: menos de cinco HBD, menos de diez HBA, MW a 500 uma, SlogP menor a 5.3, menos de diez RB y un TPSA menor a 140 Å².

2.3 Espacio químico

Para visualizar el espacio químico se realizó un análisis de componentes principales (PCA, por sus siglas en inglés, *principal component analysis*) utilizando el programa de data warrior versión 5.5.0 y el programa Orange 3 versión 3.30.2. El PCA se realizó con los seis descriptores moleculares calculados en el paso previo.

2.4 Núcleo estructural base

Se llevó a cabo un análisis de los núcleos estructurales en la base de datos. Para llevar esto a cabo se estudiaron los *frameworks* de Bemis & Murcko. Un *framework* se define como la unión de anillos y los átomos que los conectan (figura 1.10).[63] Para obtener el núcleo estructural base se escribió un código en Python 3 en donde mediante el módulo utilizado de RDKit es `RDKit.Chem.Scaffolds.MurckoScaffold` a partir del cual se utilizó la función `GetScaffoldForMol`.

2.5 Métodos basados en ligando

2.5.1 Aprendizaje supervisado para la predicción de actividad inhibitoria

Para llevar esto a cabo se siguieron dos pasos: el uso del programa *Epigenetic Target Profiler* v1.0 y el desarrollo de un método de aprendizaje supervisado (figura 2.2).

Epigenetic Target Profiler v1.0



Figura 2.2: Interfaz del programa *Epigenetic Target Profiler* v1.0.

Para lograr el primer paso se utilizaron los SMILES en el programa de *Epigenetic Target Profiler* v1.0.

Aprendizaje supervisado con Orange 3.30.2

Se descargó la base de datos de compuestos con actividad en DNMT1 en humanos de ChEMBL 29. Se seleccionaron los compuestos que tuviesen información del CI_{50} por lo que quedaron 419 compuestos con CI_{50} reportado.

Para crear el método se corrió un código en Python 3 para lograr el cálculo de las huellas digitales moleculares de RDK (2048 bits) y Morgan (2048 bits), mismas que fueron utilizadas en el *Epigenetic Target Profiler*. Para realizar el cálculo del fingerprint de Morgan se usó la función “GetMorganFingerprintAsBitVect” del módulo

de RDKit.Chem y para realizar el cálculo del fingerprint de RDK se utilizó la función “RDKFingerprint” del módulo RDKit.Chem.AllChem.

Para realizar los modelos de clasificación binaria se definió el corte de actividad a un CI_{50} de 10,000nM. Se obtuvo que de los 419 compuestos 281 eran activos y 138 eran inactivos.

Posteriormente, utilizando el programa de Orange 3 versión 3.30.2 se realizó un flujo de trabajo (figura 2.3).

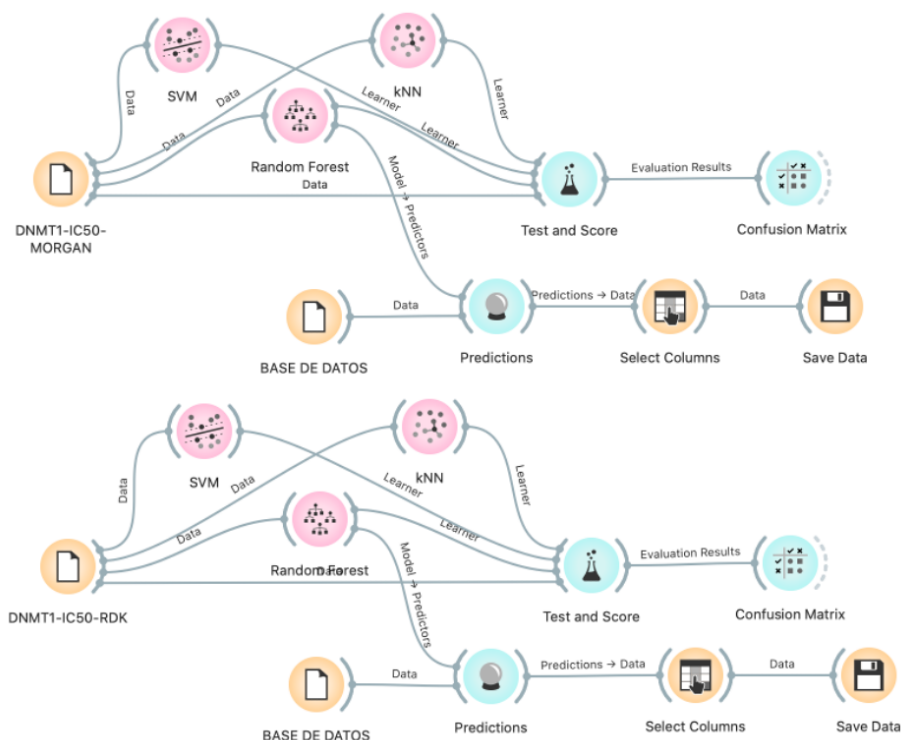


Figura 2.3: Flujo de trabajo en Orange donde se abren los archivos de DNMT1-IC₅₀-MORGAN y DNMT1-IC₅₀-RDK, se someten a los métodos de SVM, *Random forest* y kNN.

Este flujo de trabajo utilizó los métodos de *Random forest* (RF), kNN y SVM. Para el modelo de RF se utilizaron 32 árboles, 5 predictores considerados en cada decisión y 5 como profundidad máxima. Para el modelo kNN se utilizaron 2 vecinos, como métrica de distancia se utilizó Euclidean y los pesos eran uniformes. Finalmente para el modelo de SVM se utilizó una función de base radial con un costo y una gamma de 1.

Para la selección del método se realizó un análisis de los resultados del área bajo la curva (AUC, por sus siglas en inglés, *Area under the curve*), exactitud (CA, por sus siglas en inglés, *classification accuracy*), valor F1, precisión y exhaustividad (*recall*). En las ecuaciones 2.1 a 2.4 se encuentran las ecuaciones donde: TP es

verdadero positivo, FP es falso positivo, TN es verdadero negativo y FN es falso negativo.

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

$$F1 = 2 * \frac{Precisión * Exhaustividad}{Precisión + Exhaustividad} \quad (2.2)$$

$$Precisión = \frac{TP}{TP + FP} \quad (2.3)$$

$$Exhaustividad = \frac{TP}{TP + FN} \quad (2.4)$$

Para la validación interna se utilizó el método de *leave one out*. En este tipo de validación el algoritmo de aprendizaje se aplica una vez para cada instancia, usando todas las demás instancias como un conjunto de entrenamiento y usando la instancia seleccionada como un conjunto de prueba de un solo elemento.

2.6 Métodos basados en estructura

2.6.1 Acoplamiento molecular

Para realizar los estudios de acoplamiento molecular se utilizó la estructura cristalográfica de DNMT1 (PDB ID: 4WXX). Se utilizaron como control para el acoplamiento molecular SAH, SAM y los compuestos no nucleosídicos de la figura 1.9.

Autodock Vina

Utilizando Autodock Vina versión 1.2.0 se realizó un acoplamiento molecular con los compuestos seleccionados.[83] Primero se prepararon los ligandos utilizando iBabel en donde se generó un archivo individual con las estructuras 3D de cada molécula. Posteriormente, se utilizó un script para iniciar el acoplamiento molecular ciego en donde se utilizó una caja que permitiera abarcar toda la superficie de la proteína.

Para el análisis de resultados se generó un código en Python 3 para evaluar las puntuaciones de los modos de unión. Finalmente, se utilizó el servidor de Protein-Ligand Interaction Profiler (PLIP) para determinar las interacciones entre proteína y ligando.[100]

Primero fue necesario tener los archivos out.pdbqt en una misma carpeta en donde fueron transformados a pdb. Posteriormente, utilizando iBabel versión 5.0 se

selecciona el modo de unión de menor energía y la transforma a pdb. Finalmente, se generó un .pdb que contenía el ligando y el DNMT preparado el cual permite utilizar PLIP para analizar las interacciones entre el ligando y el DNMT1.

LeDock

Para realizar el acoplamiento molecular utilizando LeDock fue necesario instalar LePro para preparar la proteína.[101] Para llevar esto a cabo se eliminaron las moléculas de agua del archivo 4WXX y desde la terminal se corrió el siguiente comando: \$RutaLePro 4WXX.pdb. El siguiente paso fue utilizar iBabel versión 5.0 para preparar los ligandos en donde se centraron las coordenadas, se agregaron los hidrógenos polares y se generó la estructura 3D. Finalmente a partir de una lista de los ligandos, se realizó el acoplamiento molecular dirigido al sitio de SAH utilizando LeDock.

Molecular Operating Environment (MOE)

Rígido

Para realizar el acoplamiento molecular utilizando MOE primero se protonó la DNMT1 utilizando la función de Protonate 3D.[102] Se utilizó SiteView para ver el espacio del sitio de unión y mediante las funciones de System y Surface se generó un mapa del espacio del sitio de unión de SAH.

Para realizar el acoplamiento molecular se utilizó “Triangle Matcher”, método en donde los modos de unión se generan superponiendo tripletes de átomos de ligandos y tripletes de puntos de sitios receptores. Los puntos del sitio receptor son centros de esferas alfa que representan ubicaciones de empaquetamiento apretado. Se generaron treinta complejos para cada ligando probado. Posteriormente se eliminaron los modos de unión duplicadas y para los modos restantes se utilizó la función de puntuación de London dG para estimar la energía libre de unión del ligando.[103]

Ajuste inducido

Para realizar el acoplamiento molecular con ajuste inducido se siguieron los mismos pasos que para el acoplamiento rígido pero en el modo de ajuste inducido.

2.7 Ensayos de inhibición enzimática

Los compuestos se evaluaron en los laboratorios de servicio de Reaction Biology (USA) donde se llevaron a cabo estudios de actividad enzimática para los candidatos seleccionados.

En estos ensayos se lleva a cabo la reacción de la figura 2.4. La DNMT1 utiliza SAM marcada con tritio como donante de metilo. Esta se convierte en SAH durante la transferencia del grupo metilo radiactivo al sustrato de histona.

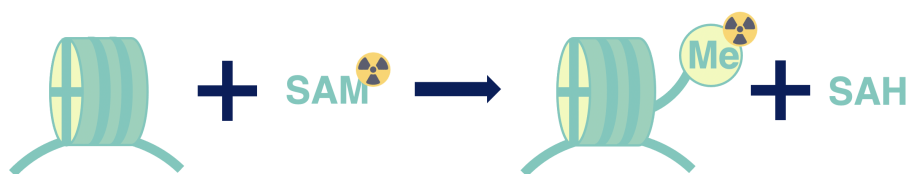


Figura 2.4: Reacción de S-adenosil-L-[metil-³H]metionina + ADN = S-adenosil-L-homocisteína + ADN 5-[metil-³H]-citosina.

Capítulo 3

Resultados y discusión

3.1 Diversidad estructural de la base de datos

La figura 3.1 muestra la distribución de: peso molecular, área topológica superficial, enlaces rotables, coeficiente de partición y donadores y aceptores de hidrógeno. Se encontró que las bases de datos de Axon, Enzo, Selleck y TargetMol fueron las que tenían una mayor diversidad estructural de los parámetros estudiados. En la tabla 3.1 se puede apreciar como estas mismas bases de datos fueron las que rompieron la mayor cantidad de reglas de Lipinski y Veber. Aunado a esto, se encontró que la base de datos ChemDiv fue de las que también tuvo una menor proporción de reglas cumplidas. No obstante, al tomar en cuenta la cantidad de compuestos que cumplen cinco o seis reglas se encuentra que para ChemDiv es el 98% de los compuestos mientras que para Axon, Enzo Selleck y TargetMol es solo el 92%, 89%, 86% y 91% respectivamente.

Tabla 3.1: Porcentaje de compuestos con reglas cumplidas de Lipinski y Veber por base de datos.

Base de datos	Reglas cumplidas de Lipinski y Veber							Compuestos
	0	1	2	3	4	5	6	
Asinex	0%	0%	1%	2%	3%	3%	92%	5391
Axon	0%	0%	1%	2%	6%	17%	75%	182
ChemDiv	0%	0%	0%	0%	1%	20%	78%	25883
Enamine	0%	0%	0%	0%	0%	3%	96%	9352
Enzo	2%	0%	0%	7%	2%	19%	70%	43
Life Chemicals	0%	0%	0%	0%	0%	4%	95%	8091
OTAVA.DNMT1	0%	0%	0%	0%	1%	2%	97%	401
OTAVA.DNMT3	0%	0%	0%	0%	0%	2%	98%	1230
Selleck	1%	2%	2%	2%	6%	14%	72%	699
TargetMol	0%	1%	1%	2%	5%	12%	79%	932

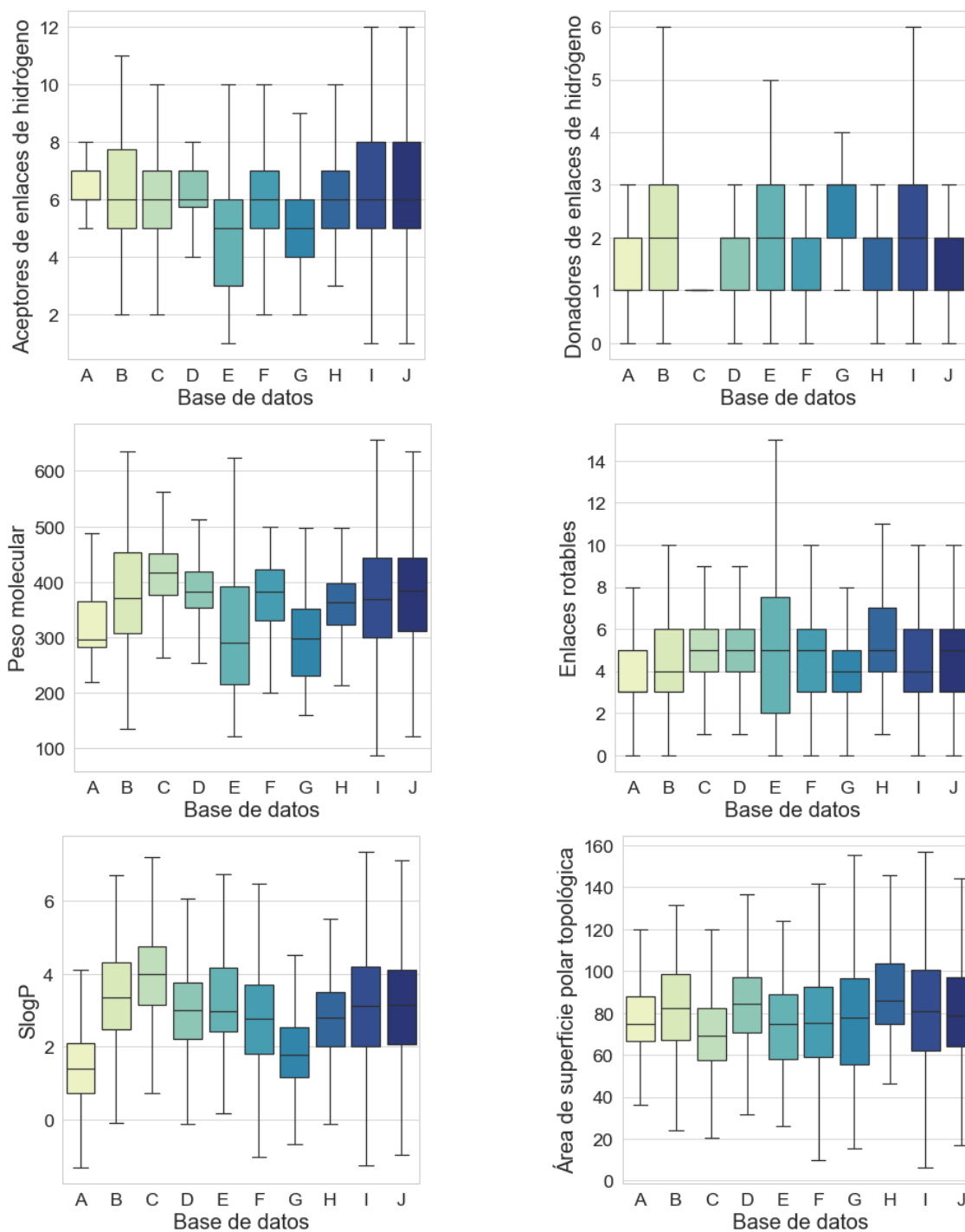


Figura 3.1: Diagramas de caja y bigotes de las propiedades fisicoquímicas estudiadas donde: A es Asinex, B es Axon, C es ChemDiv, D es Enamine, E es Enzo, F es Life Chemicals, G es OTAVA.DNMT1, H es OTAVA.DNMT3, I es Selleck y J es TargetMol.

3.2 Núcleo estructural base

En la tabla 3.2 se reportan métricas de los núcleos estructurales base encontrados por base de datos.

Tabla 3.2: Núcleos estructurales por base de datos donde: M es el número de compuestos, N el número de estructuras de Murcko, Ns los *singleton scaffolds* y Pn el porcentaje de núcleos estructurales que representan n% de compuestos.

Base de datos	M	N	Ns	N/M	Ns/N	P25	P50	P75
Asinex	5313	3448	2606	65%	76%	6%	24%	61%
Axon	178	165	157	93%	95%	19%	46%	73%
ChemDiv	25883	4421	2333	17%	53%	1%	3%	15%
Enamine	9352	7663	6995	82%	91%	9%	39%	69%
Enzo	43	29	23	67%	79%	7%	28%	64%
Life Chemicals	8091	5338	4367	66%	82%	5%	24%	62%
OTAVA_DNMT1	399	226	187	57%	83%	1%	15%	56%
OTAVA_DNMT3	1230	576	419	47%	73%	2%	12%	47%
Selleck	677	573	528	85%	92%	11%	41%	71%
TargetMol	859	755	706	88%	94%	15%	43%	72%

Se encontró que, con excepción de ChemDiv, las bases de datos estudiadas son muy diversas ya que la menor proporción de núcleos estructurales a compuestos (N/M) de estas fue de 47% lo que indica que hay aproximadamente un núcleo estructural por cada dos moléculas. En especial, se detectó que las bases de datos con una mayor proporción de N/M fueron Axon, TargetMol y Selleck, mismas que tuvieron la mayor diversidad en cuanto a las propiedades fisicoquímicas en la figura 3.1. Paralelamente, se encontró que las bases que tenían una mayor cantidad de compuestos tienen una menor proporción de *singleton scaffolds*, los cuales se definen como núcleos estructurales que solo ocurren una vez en la base de datos. Esto es parte de los resultados esperados ya que las bases de datos como ChemDiv tienden a tener una mayor cantidad de análogos, lo cual se ve reflejado en la proporción de N/M así como de Ns/N. No obstante, debido a la gran cantidad de compuestos la base de ChemDiv termina cubriendo la mayor cantidad de espacio químico.

3.3 Selección de compuestos mediante subestructuras

Con base a los compuestos activos reportados en la figura 1.9 se seleccionaron los compuestos que tuvieran las subestructuras en la figura 3.2. Posterior al filtrado de la base de datos por subestructura se encontró que para la subestructura A había 369 compuestos, para la B había 708 compuestos, para la C había 151 compuestos y para la D había 50 compuestos; resultando en un total de 1,278 compuestos.

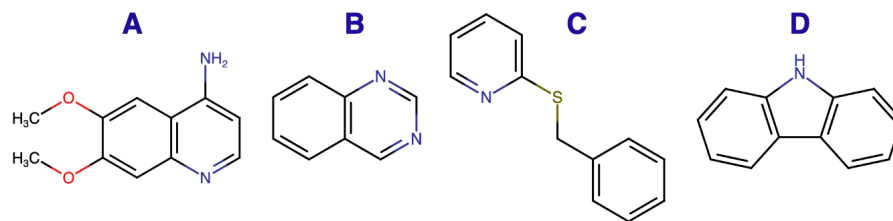


Figura 3.2: Subestructuras a estudiar donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol.

Se realizó un PCA utilizando los parámetros de MV, TPSA, RB, SlogP, HBD y HBA. Se obtuvo una varianza para el componente principal 1 de 45.8% y para el componente principal 2 de 30.3%. Se concluye que, a pesar de la selección de los compuestos que tuviesen esas subestructuras, se mantiene una diversidad en el espacio químico la cual se ve reflejada en la figura 3.3.

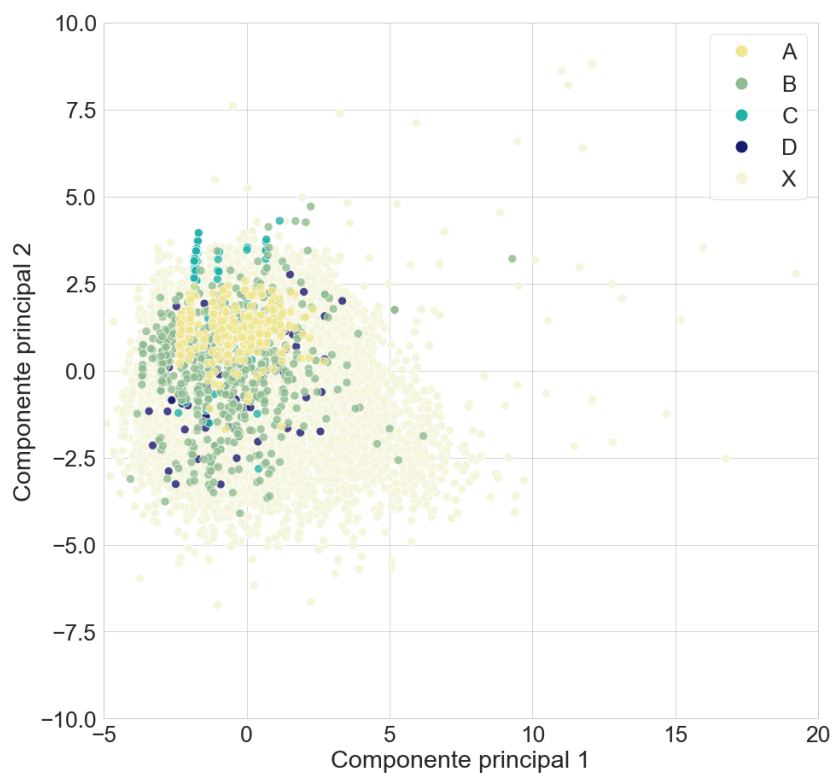


Figura 3.3: Visualización del espacio químico mediante un análisis de componentes principales donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas, D es carbazol y X es el resto de los compuestos.

Se analizó la distribución de los compuestos en el espacio tridimensional (3D) de la figura 3.4a y 3.4b y se encontró que estos ocupan un espacio químico diverso

por subestructura. Aunado a esto, se halló que 1,065 de los 1,278 compuestos venían de la base de datos ChemDiv.

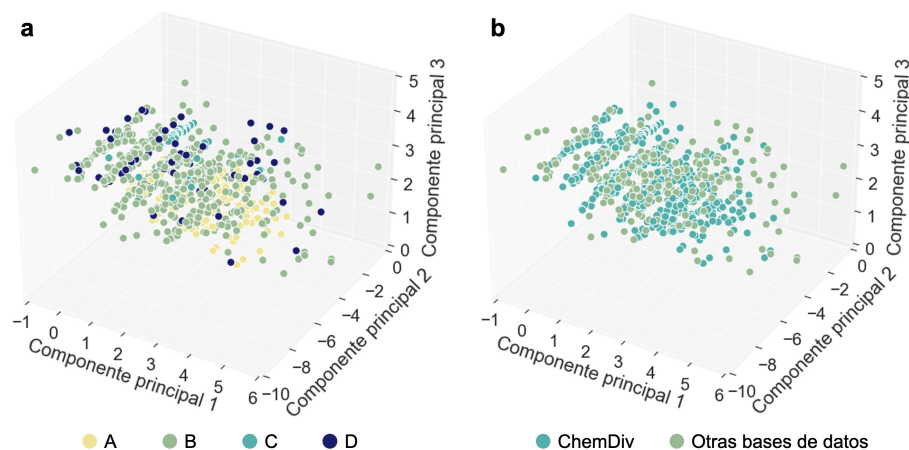


Figura 3.4: Visualización del espacio químico mediante un análisis de componentes principales para la selección de compuestos por subestructura. En **a** se clasifican por subestructura donde: A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol. En **b** se clasifican entre ChemDiv y el resto de las bases de datos.

3.4 Métodos basados en ligandos

3.4.1 *Epigenetic Target Profiler v1.0*

Al utilizar el *Epigenetic Target Profiler* utilizando la metodología reportada en la sección 2.5.1 se predijo que los compuestos no tendrían actividad. Se cree que esto se debe a que el programa utiliza una versión anterior de la datos de ChEMBL. Es por esto que se decidió hacer un modelo de aprendizaje automático con los datos más recientes de ChEMBL v29 para obtener predicciones más actualizadas.

3.4.2 *Orange 3*

Para cada una de las huellas digitales moleculares utilizadas, Morgan y RDK, se realizó el flujo de trabajo reportado en la sección 2.5.1. A partir de los modelos generados para cada una de las huellas, kNN, SVM y RF se realizó el proceso de selección mediante los resultados de la validación interna.

La validación interna es el proceso de primero dividir los datos de entrada en un conjunto de entrenamiento y un conjunto de prueba. El conjunto de entrenamiento se usa para entrenar el modelo, mientras que el conjunto de prueba se usa para verificar la precisión del modelo después de haberlo entrenado. Al realizar el proceso de validación interna se obtuvieron los valores de las tablas 3.3

y 3.4 donde: AUC es el área bajo la curva, CA es exactitud, F1 es valor F1, P es precisión y R es exhaustividad. La métrica más usada para medir la validez de los modelos generados es la AUC y se ha encontrado que la mayoría de los modelos generados por RF tienen un valor mayor a 0.7.[104] Los modelos desarrollados por medio de Orange 3 tuvieron un AUC que excedía el valor esperado para un método de RF.

Tabla 3.3: Resultados de la validación interna de Morgan – IC₅₀.

Modelo	AUC	CA	F1	P	R
RF	0.81	0.76	0.84	0.77	0.93
SVM	0.77	0.71	0.82	0.71	0.96
kNN	0.75	0.71	0.77	0.82	0.73

Tabla 3.4: Resultados de la validación interna de RDK – IC₅₀.

Modelo	AUC	CA	F1	P	R
RF	0.80	0.76	0.84	0.77	0.92
SVM	0.78	0.71	0.81	0.74	0.88
kNN	0.71	0.66	0.71	0.82	0.63

Se escogió el método de RF para predecir la actividad de los compuestos tomando como criterio que estos fueran positivos en ambos modelos para continuar con la selección.

Primero se predijo la actividad de los compuestos control (figura 1.9) y se identificó como inactivo: la genisteína en ambos modelos, el DC_517 en el modelo de Morgan y UNC0646 en RDK. El resto de los compuestos fueron exitosamente predichos como activos.

Al analizar los resultados de los compuestos se encontró que las bases de datos tenían entre el 64% y 71% de los compuestos predichos como activos para ambos modelos (tabla 3.5).

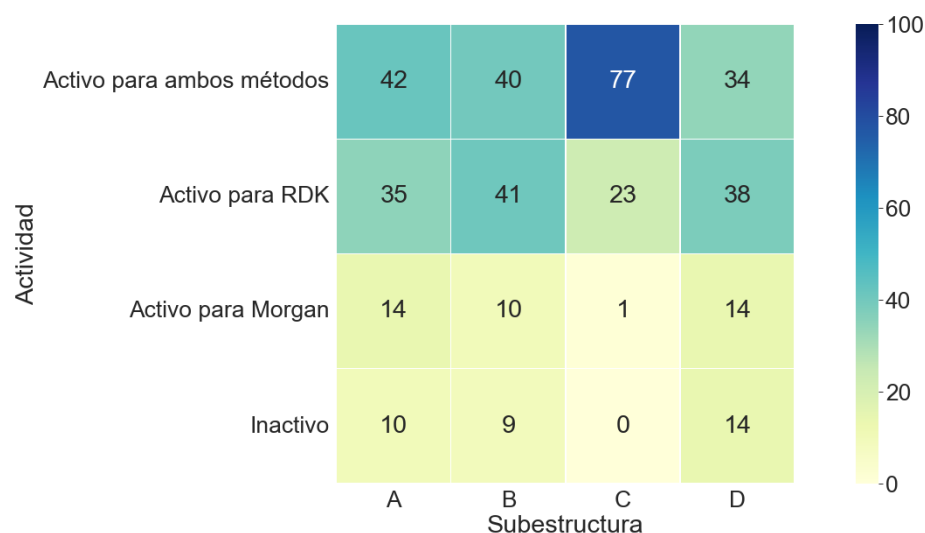
Posteriormente, se analizaron los compuestos que contenían las subestructuras buscadas y se encontró que el 45.9% de los compuestos fueron predichos activos para ambos modelos, 36.3% de los compuestos fueron predichos activos para RDK, 9.5% de los compuestos fueron predichos activos para Morgan, y 8.3% fueron predichos inactivos. A pesar de que se encontró que el porcentaje de compuestos predichos activos para ambos modelos (45.9%) es menor que el predicho para toda la base de datos (67%), el 82.2% de los compuestos fueron activos para RDK para la base de datos con las subestructuras contra el 85% de la base de datos. Aunado a esto, se encontró que el 55.4% de los compuestos de la base de subestructuras fueron activos para Morgan contra el 78% de la base de datos. Esto indica que posiblemente no haya suficiente información de compuestos que contengan una estructura similar a la buscada (figura 3.2).

Finalmente, se realizó un análisis por subestructura de la base de datos y se mostró en la figura 3.5 que hay una gran disparidad en la actividad predicha por

Tabla 3.5: Resultados de los modelos de aprendizaje automático para todos los compuestos de las bases de datos.

Base de datos	Activos en ambos modelos	Activos solo para RDK	Activos solo para Morgan	Inactivos en ambos modelos
Axon	65%	13%	15%	7%
ChemDiv	67%	18%	11%	4%
Asinex	67%	21%	8%	4%
Enamine	69%	16%	11%	4%
Enzo	65%	9%	21%	5%
LifeChem	64%	22%	10%	4%
OTAVA_DNMT1	65%	13%	17%	5%
OTAVA_DNMT3b	65%	18%	14%	4%
Selleck	67%	16%	12%	5%
Targetmol	71%	14%	10%	4%
Total	67%	18%	11%	4%

los distintos modelos. Esta disparidad se ve principalmente reflejada con las 2-(bencilsulfanil) piridinas; se encontró una mayor proporción de compuestos con actividad predicha para ambos modelos (77%), RDK y Morgan, a comparación de las otras subestructuras (42% a 34%). Esto permite reiterar la importancia del estudio de compuestos que ocupen un mayor espacio químico.

**Figura 3.5:** Resultados del método de aprendizaje automático para cada subestructura donde A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas y D es carbazol.

Se encontró que de los 1,278 compuestos que tenían alguna de las subestructuras buscadas, 1,065 compuestos venían de ChemDiv; en donde: 81% de los com-

puestos predichos activos para ambos modelos, 89% de los compuestos predichos para Morgan, 76% de los compuestos predichos activos para RDK y 81% de los compuestos predichos inactivos eran de ChemDiv. En consecuencia, se decidió utilizar los compuestos de la base de datos de ChemDiv para los pasos siguientes.

Además, para poder verificar la efectividad de las quimiotecas enfocadas se decidió aplicar el método a bases de datos que no fueran enfocadas a compuestos epigenéticos para poder realizar una comparación (tabla 3.6).

Tabla 3.6: Descripción de las bases de datos.

Base de datos	Descripción	Compuestos	Referencia
BIOFACQUIM	Base de datos de compuestos con productos naturales aislados y caracterizados en México	423	[105]
ChEBI	Base de datos que contiene información sobre compuestos químicos de interés biológico	50996	[106]
ChemTastesDB	Base de datos que incluye información seleccionada de 2944 saborizantes moleculares	2944	[107]
DrugBank	Biblioteca de fármacos aprobados	1713	[108]
T3DB	T3DB es una base de datos de toxinas comunes y sus objetivos de toxinas asociados.	3533	[109]

Se encontró que una menor cantidad de los compuestos fueron predichos como activos (tabla 3.7). Existen dos razones para esto; primeramente que al no estar enfocadas se pierde la posibilidad de que se tenga conocimiento sobre la actividad en ese espacio químico. En segundo lugar, que esto podría deberse a que al no ser enfocadas para blancos epigenéticos no se encuentren estructuras que tengan actividad.

Tabla 3.7: Resultados del método de aprendizaje automático para todos los compuestos de las bases de datos generales.

Base de datos	Activos en ambos modelos	Activos solo para RDK	Activos solo para Morgan	Inactivos en ambos modelos
BIOFACQUIM	52%	21%	19%	9%
ChEBI	57%	16%	18%	9%
ChemTastes	54%	20%	17%	10%
DrugBank	38%	25%	22%	15%
T3DB	68%	12%	14%	6%

Se realizó un análisis de los núcleos estructurales base de estas quimiotecas. Se detectó que la mayoría de los compuestos contenían una alta frecuencia de los núcleos encontrados en la tabla 3.8. Al comparar estas con las bases de datos enfocadas que se utilizaron se encontró que hay una menor proporción de N/M lo

que indica que hay una mayor cantidad de análogos. Posteriormente, se estudió la cantidad de compuestos que no tenían ningún anillo así como los que solamente tenían un anillo de benceno. Se observó que estas dos subestructuras correspondían al núcleo del: 9% de los compuestos de BIOFACQUIM, 37% de los compuestos de ChEBI, 32% de los compuestos de ChemTastesDB, 14% de los compuestos de DrugBank y 45% de los compuestos de T3DB. Esto indica que las bases de datos están principalmente compuestas por moléculas pequeñas. Finalmente, se encontró que 160 de los compuestos de la base de datos de T3DB tenían la subestructura de carbazol lo que explica la mayor cantidad de compuestos predichos como activos.

Tabla 3.8: Núcleos estructurales por base de datos donde: M es el número de compuestos, N el número de estructuras de Murcko, Ns los *singleton scaffolds* y Pn el porcentaje de núcleos estructurales que representan n% de compuestos.

Base de datos	M	N	Ns	N/M	Ns/N	P25	P50	P75
BIOFACQUIM	423	195	125	46%	64%	1%	17%	46%
ChEBI	50996	9898	3749	19%	38%	1%	1%	10%
ChemTastes	2944	757	256	26%	34%	1%	2%	18%
DrugBank	1713	709	431	41%	61%	2%	12%	39%
T3DB	3519	875	684	25%	78%	1%	1%	10%

Con base a los resultados, se decidió estudiar el espacio químico mediante un PCA utilizando las propiedades fisicoquímicas que habían sido utilizadas (sección 3.1) y se encontró que tenían varianzas donde $PC1 = 75.4\%$, $PC2 = 18.7\%$ equivaliendo a un total de 94.1% (figura 3.6). El primer hallazgo fue que las bases de datos que no son enfocadas cubren un mayor espacio químico. A primera vista, esto indica que la diferencia entre la cantidad de compuestos predichos como positivos es menor debido a que se abarca un espacio químico distinto. Sin embargo, en la imagen 3.6b se muestra un análisis de la distribución entre las subestructuras y la base de datos lo que muestra que las subestructuras ocupan un espacio todavía más reducido. Se realizó una comparación entre la actividad predicha por las bases de datos (53.8% en ambos modelos), las bases de datos enfocadas (67% en ambos modelos) y los compuestos con subestructuras (48.25% en ambos modelos). Esta comparación permite elucidar que la diferencia entre la actividad predicha no se debe necesariamente a que haya una diferencia en el espacio químico estudiado.

Además, al estudiar el espacio químico entre las distintas bases de datos (figura 3.6c y 3.6d) se encontró que la base de datos de ChEBI era la que ocupaba un mayor espacio. Es por esto que al traslapar estas bases con las enfocadas se encontró que la mayoría de las otras cubrían un espacio químico similar. Tomando en consideración lo mencionado, se refleja la utilidad de las bases de datos enfocadas.

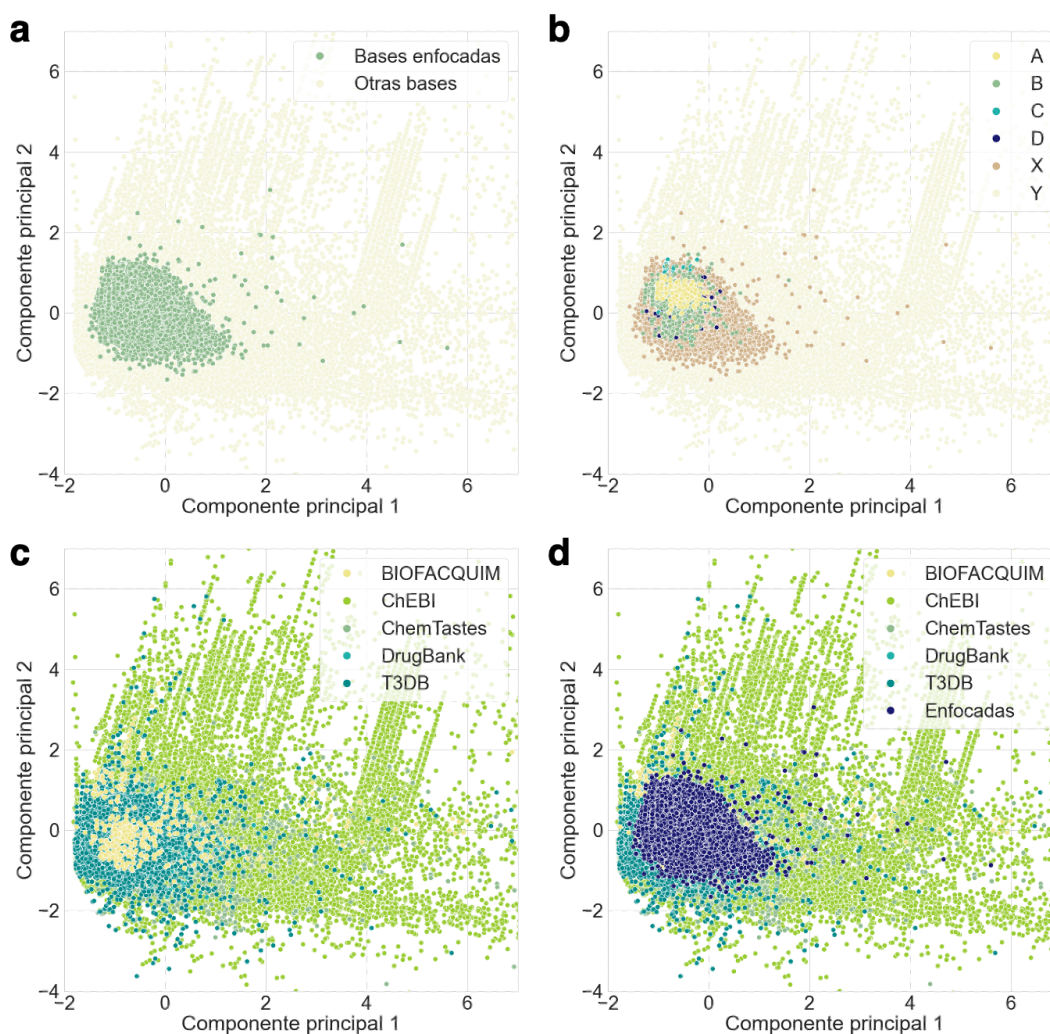


Figura 3.6: Visualización del espacio químico mediante un análisis de componentes principales donde se incluyen las bases de datos de la tabla 3.6. En **a** se clasifican por el tipo de base de datos. En **b** se expande la clasificación para mostrar los compuestos de las bases de datos que tienen alguna de las subestructuras estudiadas donde: A es amiquinsina, B es quinazolina, C es 2-(bencilsulfanil) piridinas, D es carbazol, X es el resto de los compuestos de las bases de datos enfocadas y Y los compuestos de las bases de datos de la tabla 3.6. En **c** se muestran las bases de datos de la tabla 3.6. En **d** se añadió a la figura c los compuestos de las bases de datos enfocadas.

3.5 Acoplamiento molecular

Para los compuestos seleccionados se realizó el acoplamiento molecular utilizando los distintos programas mencionados en la metodología (sección 2.6.1).

Tabla 3.9: Resultados de la energía libre de los controles utilizados para el acoplamiento molecular.

Control	Vina	LeDock	MOE
SAH	-7.9	-10.3	-7.6
SAM	-7.4	-9.2	-7.7
Genistein	-8.8	-6.5	-6.0
SGI-1027	-10.4	-7.6	-8.2
CHEMBL1566980	-9.2	-9.4	-8.3
UNC0646	-10.1	-6.8	-10.2
GSK-3484862	-6.9	-6.8	-8.0
GSK3685032	-7.9	-7.7	-8.1
DC_517	-9.3	-7.6	-9.6

Al realizar la comparación entre las medidas de mediana (μ), media (M) y moda (Mo) se observaron los mismos valores por lo que se puede decir que hubo una distribución simétrica.

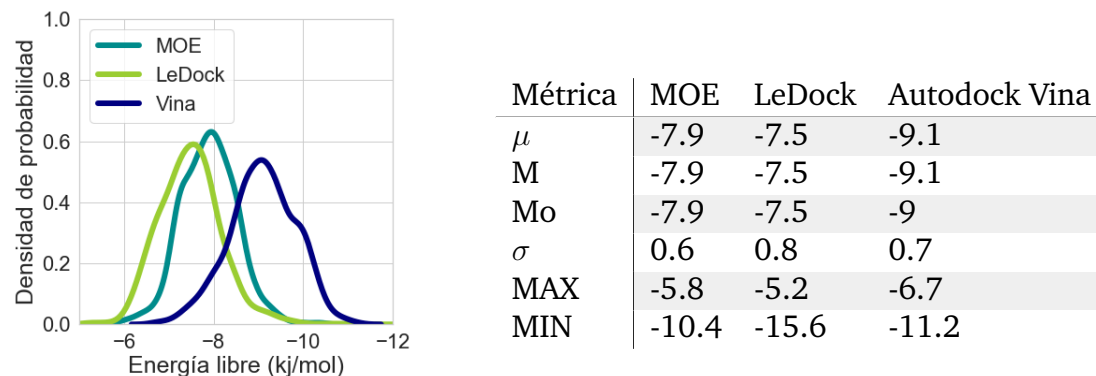


Figura 3.7: Distribución de la energía libre de Gibbs (ΔG) de unión por programa empleado.

Posteriormente, se estudiaron los patrones de acoplamiento molecular para los compuestos con distintas subestructuras (figura 3.8). Al comparar las distribuciones de cada programa se encontró que las de las quinazolininas eran simétricas. Por otro lado, se encontró que las distribuciones de las carbazol, las 2(bencilsulfanil) piridinas y las amiquinsinas eran asimétricas y tenían una menor concentración en la zona central de las distribuciones.

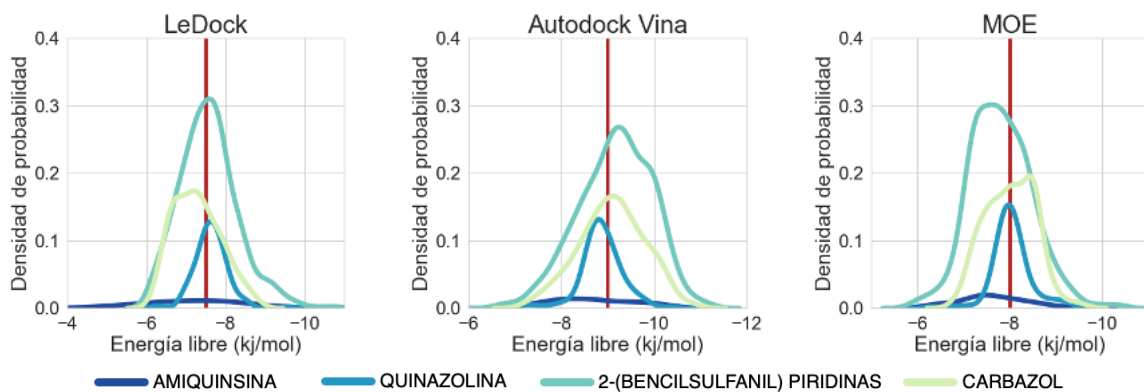


Figura 3.8: Función de densidad de probabilidad del ΔG de LeDock, Autodock Vina y MOE de acuerdo a la subestructura encontrada.

Se realizó un análisis entre la clasificación obtenida por el método de aprendizaje automático y la energía libre de Gibbs de unión obtenido por programa (figura 3.9). Debido a que no se encontró una relación entre el valor energético y la actividad predicha por los modelos se determinó que no hubo una clara relación entre los métodos de acoplamiento molecular y los de aprendizaje automático. Esto permite que sean usados de manera complementaria en la selección de compuestos.

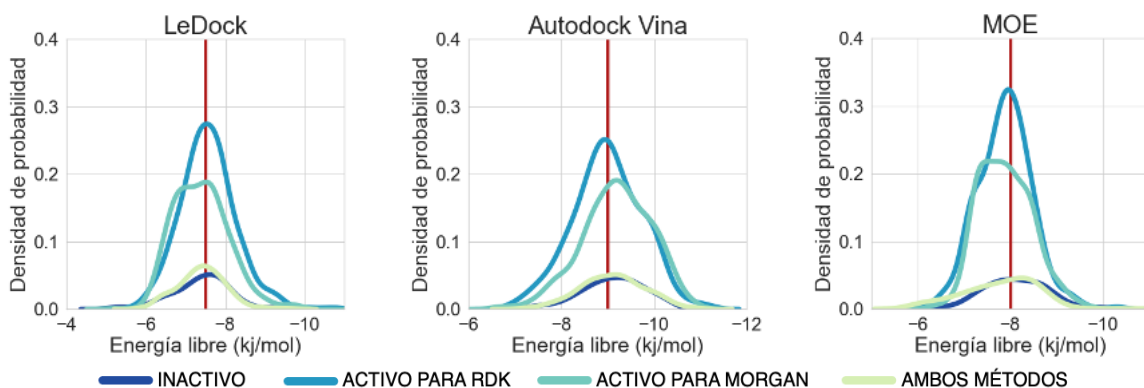


Figura 3.9: Función de densidad de probabilidad de la energía de Gibbs de LeDock, Autodock Vina y MOE de acuerdo a la actividad predicha por el método de aprendizaje automático.

3.6 Selección de compuestos para evaluar actividad

Tomando en cuenta los resultados de los métodos basados en ligando y los métodos basados en estructura se realizó la selección de compuestos que siguieran la metodología indicada en la figura 3.10. En el caso del aprendizaje automático se tomo como criterio que los compuestos seleccionados hubiesen sido predichos como activos para ambos modelos por lo que se paso de tener 1065 compuestos a 457 comuestos. Finalmente, se utilizaron los valores energéticos de los estudios de acoplamiento molecular para realizar la selección de compuestos y pasar de 457 compuestos a los 19 compuestos que se mandaron a evaluar.

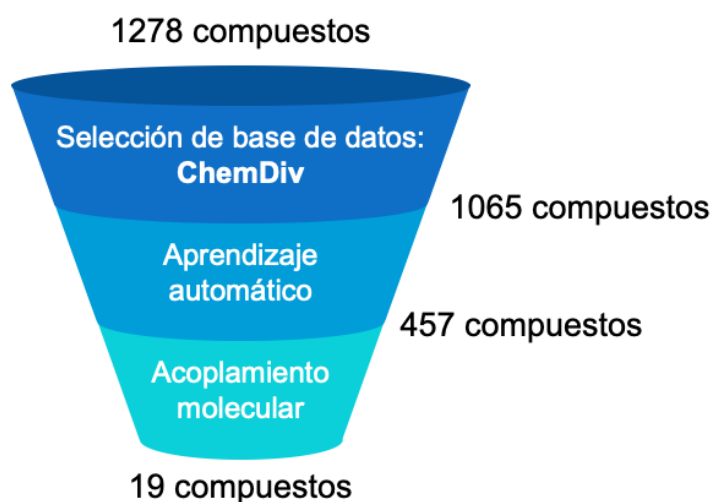


Figura 3.10: Selección de compuestos para evaluación de inhibición enzimática.

3.7 Ensayos enzimáticos

Se dividieron los compuestos en tres categorías de acuerdo a su actividad enzimática: muy activador (donde la actividad es mayor a 150%), activador (donde la actividad está entre 150% y 105%) y inactivo (donde la actividad es menor a 105%). Se encontró que de los 19 compuestos siete eran muy activadores, cuatro eran activadores y ocho eran inactivos.

De los siete compuestos clasificados como muy activadores cuatro contenían la subestructura de quinazolina, dos de amiquinsina y uno de carbazol (tabla 3.11). Por otro lado, se encontró que de los ocho compuestos que no presentaron ningún tipo de actividad cinco tenían la subestructura de 2-(bencilsulfanil) piridina. Esto difiere de los resultados predichos por el modelo de aprendizaje automático en donde el 77% de los compuestos con esta subestructura tenían predicho alguna actividad. Este resultado refuerza la importancia de tener una base de datos con un amplio espacio químico.

Tabla 3.10: Resultados de la actividad enzimática relativa con DNMT1 después de la administración de $100\mu\text{M}$ de cada compuesto donde MA corresponde a los compuestos muy activadores, A a los compuestos activadores y I a los compuestos inactivos.

Compuesto	Actividad enzimática relativa (%)		Activación
	Medición 1	Medición 2	
C066-2421	385.65	463.43	MA
C201-0209	99.24	101.63	I
C201-0214	100.45	104.63	I
C350-0522	546.63	569.62	MA
C769-1612	95.18	101.42	I
C769-2124	99.32	107.76	I
G198-0326	97.67	102.86	I
G198-0650	92.86	89.85	I
G198-0794	91.61	96.58	I
L112-0668	634.59	403.90	MA
L112-0688	113.17	129.39	A
L112-0725	586.94	463.13	MA
L112-0728	448.20	486.65	MA
L112-0731	501.15	517.05	MA
L483-0323	115.27	115.47	A
L485-2669	415.71	469.14	MA
L485-2761	158.47	144.51	A
M056-0726	130.98	132.34	A
M628-0321	93.65	97.42	I

Tabla 3.11: Comparación entre actividad del compuesto y subestructura.

Compuestos	Amiquinsina	Quinazolina	2-	Carbazol
			(bencilsulfanil) piridinas	
Inactivos	2	1	5	
Activador	2	1	1	
Muy Activador	2	4		1
Total	6	6	6	1

En la figura 3.11 se reporta el análisis que se realizó de la relación entre la actividad reportada y los resultados del acoplamiento molecular. Para los programas de MOE y Autodock Vina había un patrón en donde a mayor actividad había un menor valor de energía libre. Esto muestra que los compuestos activadores tenían una menor energía de unión a la DNMT1 lo que permite elucidar la utilidad de emplear estos métodos computacionales en la selección de compuestos para mandar a evaluar.

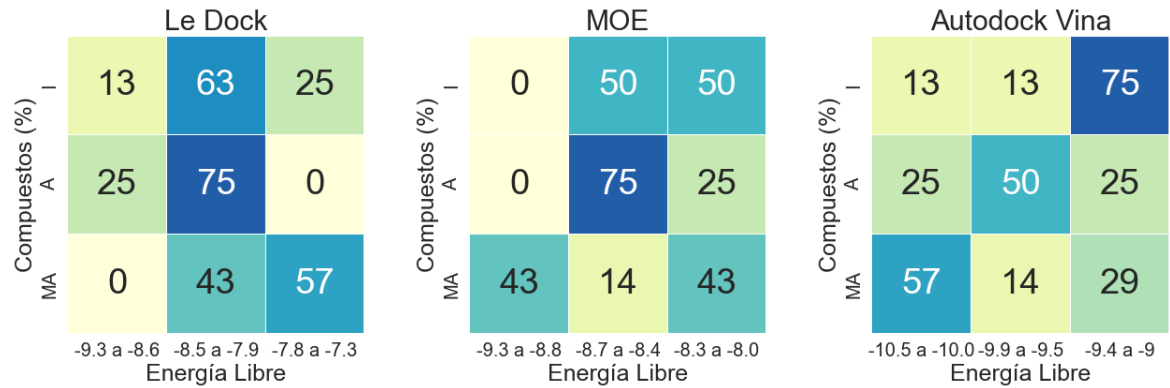


Figura 3.11: Porcentaje de los compuestos clasificados por su energía libre y su actividad donde: I es inactivos, A es activadores y MA es muy activadores.

Finalmente, se realizó un análisis de los modos de unión entre los compuestos inactivos y los que presentaron alguna actividad (figura 3.12). Se encontró que a diferencia de los compuestos inactivos la mayoría de los compuestos que presentaban actividad tenían una interacción con el dominio RFTS y CXXC (figura 3.13 y anexo 1). Se propone que los compuestos interactúan con el RFTS; esto permite que se evite la función de autoinhibición del RFTS y aumente la metilación (sección 1.1.4).

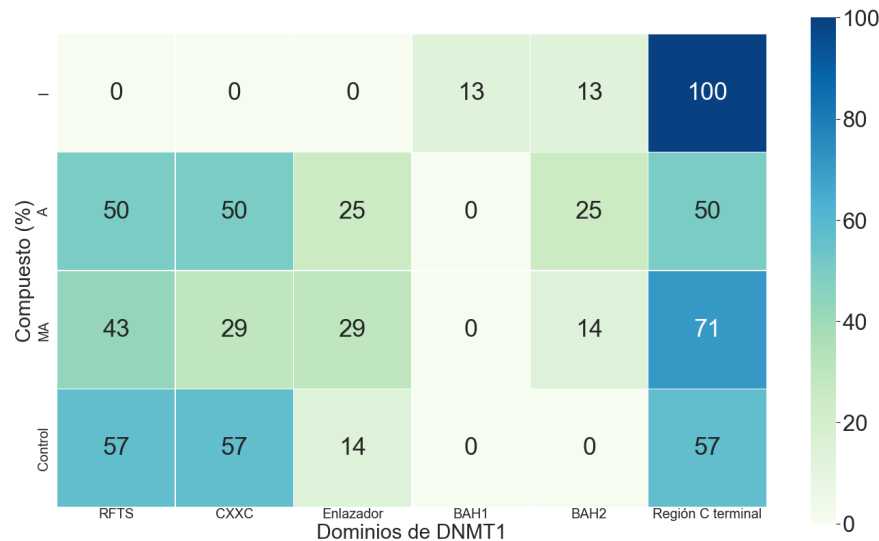


Figura 3.12: Porcentaje de compuestos que tienen algún contacto con cada dominio de la DNMT1.

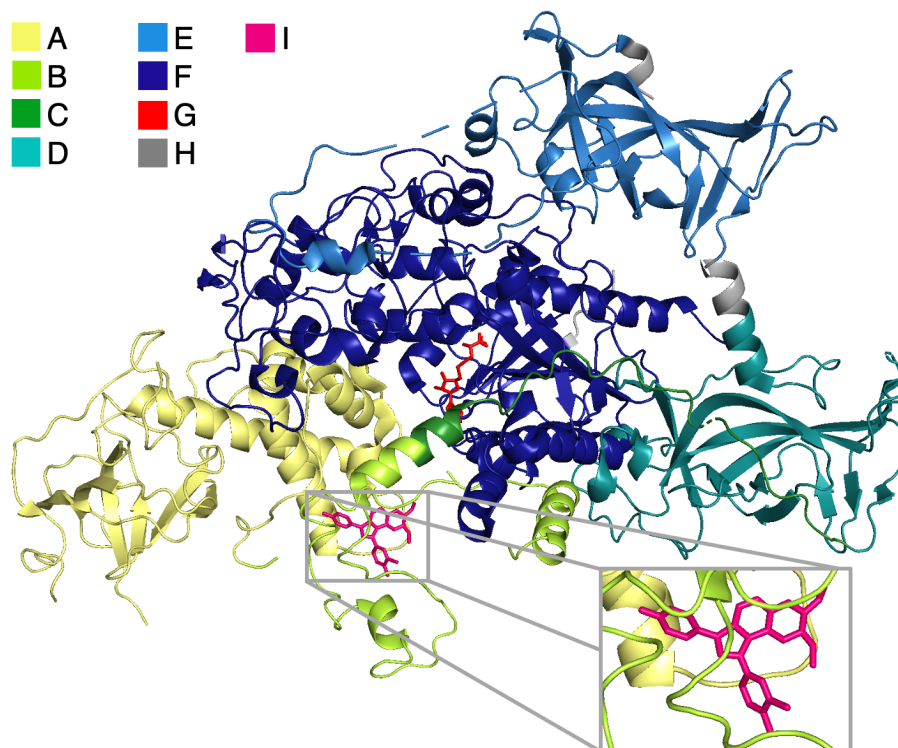


Figura 3.13: Dominios de DNMT1 donde: A corresponde a RFTS (AA351-600), B a CXXC (AA651-697), C al enlazador autoinhibidor (AA697-728), D a BAH1 (AA728-900), E es BAH2 (AA906-1100), F es el dominio de metiltransferasa (AA1139-1616) G es el compuesto SAM y H corresponde a los enlazadores entre los dominios, I es el compuesto C350-0522.

Capítulo 4

Conclusiones

Se demostró la eficiencia de las quimiotecas enfocadas ya que logran cubrir un gran espacio químico (*singleton scaffolds* y propiedades moleculares) y muestran un mejor rendimiento con los métodos de aprendizaje automático. Esto se probó al realizar la comparación entre las bases de datos que no son enfocadas en blancos epigenéticos con las que sí.

Además, se logró utilizar exitosamente una combinación de métodos basados en ligando y basados en estructura. En cuanto a los métodos basados en ligando, se encontró que fue necesario utilizar dos métodos de *Random forest* ya que ninguno de los compuestos fue identificado como activo con el *Epigenetic Target Profiler*. Se obtuvo que el 67% de los compuestos de la base de datos generada fueron identificados como activos por ambos métodos. En cuanto a los métodos basados en estructuras se encontró que algunos compuestos tenían contactos con los dominios de RFTS y CXXC. Se sabe que el dominio RFTS ejerce una función autoinhibidora de la DNMT1, evitando que el ADN interactúe con la enzima.[24, 30] Debido a esto, se plantea que los compuestos que tengan esta interacción podrían prevenir la actividad autoinhibidora lo que causaría un aumento en la metilación.

Finalmente, en los ensayos de actividad enzimática se encontró que once de los diecinueve compuestos mandados a evaluar fueron activadores de la DNMT1. La activación de la DNMT1 soporta la hipótesis de que las interacciones entre los compuestos se deben a una disrupción de la regulación autoinhibidora del dominio RFTS. Estos resultados además de complementar el espacio químico conocido de la DNMT1 también permiten plantear la posibilidad del uso de activadores de DNMT1 para tratar enfermedades, como las metástasis en cáncer de próstata (sección 1.1.3).

Capítulo 5

Perspectivas

Se plantea la posibilidad de realizar ensayos enzimáticos, como el realizado por Syeda y colaboradores, utilizando una DNMT1 sin el dominio de RFTS.[24] Esto permitiría comprobar si, en efecto, los activadores identificados actúan mediante su interacción con el dominio de RFTS, lo que evita que haya una autoinhibición (sección 1.1.4).

Por otra parte, se demostró que las quimiotecas enfocadas presentan una oportunidad para poder reducir la cantidad de compuestos a estudiar (sección 3.4.2). No obstante, es necesario exigir transparencia en cuanto a las metodologías que utilizan las empresas para seleccionar los compuestos de estas bases de datos enriquecidas. Esto también se ve reflejado en el área académica en donde existen carencias debido a los requisitos para la divulgación científica. Por esto, se propone incentivar la creación de espacios en donde los investigadores pudieran publicar sus experimentos y estudios etiquetados como no exitosos.[110] Al compartir estas experiencias, se previene que futuros estudios repitan errores y se generen investigaciones más eficientes. Por lo tanto, se plantea continuar trabajando en la divulgación científica mediante publicaciones de acceso libre y el uso de softwares gratuitos. Esto permitirá continuar conociendo el espacio químico de enzimas epigenéticas de manera accesible e integral.

Finalmente, el trabajo realizado refleja la importancia de continuar la investigación de moduladores epigenéticos, específicamente, moduladores para la DNMT1. Es necesario continuar trabajando de manera colaborativa con una combinación de métodos computacionales y experimentales para enriquecer los métodos de aprendizaje automático y tener un mejor entendimiento del espacio químico.

Referencias

- [1] Zelin Jin and Yun Liu. “DNA methylation in human diseases”. In: *Genes and Diseases* 5.1 (2018), pp. 1–8. ISSN: 23523042. DOI: 10.1016/j.gendis.2018.01.002. URL: <https://doi.org/10.1016/j.gendis.2018.01.002>.
- [2] Giacomo Cavalli and Edith Heard. “Advances in epigenetics link genetics to the environment and disease”. In: *Nature* 571.7766 (2019), pp. 489–499. ISSN: 14764687. DOI: 10.1038/s41586-019-1411-0. URL: <http://dx.doi.org/10.1038/s41586-019-1411-0>.
- [3] Yuan Cheng et al. “Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials”. In: *Signal Transduction and Targeted Therapy* 4.1 (2019). ISSN: 20593635. DOI: 10.1038/s41392-019-0095-0. URL: <http://dx.doi.org/10.1038/s41392-019-0095-0>.
- [4] Bilian Jin, Yajun Li, and Keith D. Robertson. “DNA methylation: Superior or subordinate in the epigenetic hierarchy?” In: *Genes and Cancer* 2.6 (2011), pp. 607–617. ISSN: 19476019. DOI: 10.1177/1947601910393957.
- [5] Emily N.W. Wheeler et al. “DNA methylation and brain structure and function across the life course: A systematic review”. In: *Neuroscience and Biobehavioral Reviews* 113.March (2020), pp. 133–156. ISSN: 18737528. DOI: 10.1016/j.neubiorev.2020.03.007. URL: <https://doi.org/10.1016/j.neubiorev.2020.03.007>.
- [6] Lisa D. Moore, Thuc Le, and Guoping Fan. “DNA methylation and its basic function”. In: *Neuropsychopharmacology* 38.1 (2013), pp. 23–38. ISSN: 0893133X. DOI: 10.1038/npp.2012.112.
- [7] Martin I. Lind and Foteini Spagopoulou. “Evolutionary consequences of epigenetic inheritance”. In: *Heredity* 121.3 (2018), pp. 205–209. ISSN: 13652540. DOI: 10.1038/s41437-018-0113-y. URL: <http://dx.doi.org/10.1038/s41437-018-0113-y>.
- [8] Alfonso Dueñas-González, J. Jesús Naveja, and José L. Medina-Franco. *Introduction of Epigenetic Targets in Drug Discovery and Current Status of Epi-Drugs and Epi-Probes*. Elsevier Inc., 2016, pp. 1–20. ISBN: 9780128028094. DOI: 10.1016/B978-0-12-802808-7.00001-0. URL: <http://dx.doi.org/10.1016/B978-0-12-802808-7.00001-0>.

- [9] Ganji Purnachandra Nagaraju et al. “Epigenetics and therapeutic targets in gastrointestinal malignancies”. In: *Drug Discovery Today* 26.10 (2021), pp. 2303–2314. ISSN: 18785832. DOI: 10.1016/j.drudis.2021.04.013. URL: <https://doi.org/10.1016/j.drudis.2021.04.013>.
- [10] Hao Luo et al. “Targeting the epigenetic processes to enhance antitumor immunity in small cell lung cancer”. In: *Seminars in Cancer Biology* xxxx (2022). ISSN: 10963650. DOI: 10.1016/j.semcancer.2022.02.018. URL: <https://doi.org/10.1016/j.semcancer.2022.02.018>.
- [11] Sergio Valente et al. “Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells”. In: *Journal of Medicinal Chemistry* 57.3 (2014), pp. 701–713. ISSN: 00222623. DOI: 10.1021/jm4012627.
- [12] Renata Zofia Jurkowska, Tomasz Piotr Jurkowski, and Albert Jeltsch. “Structure and Function of Mammalian DNA Methyltransferases”. In: *ChemBioChem* 12.2 (2011), pp. 206–222. ISSN: 14394227. DOI: 10.1002/cbic.201000195.
- [13] Vincenza Conteduca et al. “Epigenetics in prostate cancer: Clinical implications”. In: *Translational Andrology and Urology* 10.7 (2021), pp. 3104–3116. ISSN: 22234691. DOI: 10.21037/tau-20-1339.
- [14] Layla A. Al-Kharashi et al. “The DNA methyl-transferase protein DNMT1 enhances tumorpromoting properties of breast stromal fibroblasts”. In: *Oncotarget* 9.2 (2018), pp. 2329–2343. ISSN: 19492553. DOI: 10.18632/oncotarget.23411.
- [15] Minnatallah Al-Yozbaki et al. “Targeting DNA methyltransferases in non-small-cell lung cancer”. In: *Seminars in Cancer Biology* August 2020 (2021). ISSN: 10963650. DOI: 10.1016/j.semcancer.2021.01.005. URL: <https://doi.org/10.1016/j.semcancer.2021.01.005>.
- [16] Elena G. Pasyukova et al. “Epigenetic enzymes: A role in aging and prospects for pharmacological targeting”. In: *Ageing Research Reviews* 67. February (2021), p. 101312. ISSN: 18729649. DOI: 10.1016/j.arr.2021.101312. URL: <https://doi.org/10.1016/j.arr.2021.101312>.
- [17] Hymie Anisman and Alexander W. Kusnecov. *Genetic and epigenetic processes linked to cancer*. 2022, pp. 93–134. ISBN: 9780323919043. DOI: 10.1016/b978-0-323-91904-3.00013-6.
- [18] Stuart M. Raleigh. *Epigenetic processes-An overview*. Elsevier Inc., 2021, pp. 23–36. ISBN: 9780128206829. DOI: 10.1016/B978-0-12-820682-9.00008-6. URL: <https://doi.org/10.1016/B978-0-12-820682-9.00008-6>.
- [19] Geoffrey A. Holdgate et al. “Drug discovery for epigenetics targets”. In: *Drug Discovery Today* xxx.xx (2021). ISSN: 18785832. DOI: 10.1016/j.drudis.2021.10.020. URL: <https://doi.org/10.1016/j.drudis.2021.10.020>.

- [20] Ângela Marques-Magalhães et al. “Targeting DNA methyltransferases in urological tumors”. In: *Frontiers in Pharmacology* 9.APR (2018). ISSN: 16639812. DOI: 10.3389/fphar.2018.00366.
- [21] John R. Edwards et al. “DNA methylation and DNA methyltransferases”. In: *Epigenetics and Chromatin* 10.1 (2017), pp. 1–10. ISSN: 17568935. DOI: 10.1186/s13072-017-0130-8.
- [22] Sunzeyu Huang et al. “A novel class of selective non-nucleoside inhibitors of human DNA methyltransferase 3A”. In: *Bioorganic and Medicinal Chemistry Letters* 40.February (2021), p. 127908. ISSN: 14643405. DOI: 10.1016/j.bmcl.2021.127908. URL: <https://doi.org/10.1016/j.bmcl.2021.127908>.
- [23] Pavel Bashtrykov et al. “The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism”. In: *Journal of Biological Chemistry* 289.7 (2014), pp. 4106–4115. ISSN: 00219258. DOI: 10.1074/jbc.M113.528893.
- [24] Farisa Syeda et al. “The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1”. In: *Journal of Biological Chemistry* 286.17 (2011), pp. 15344–15351. ISSN: 00219258. DOI: 10.1074/jbc.M110.209882.
- [25] Kah Keng Wong. “DNMT1: A key drug target in triple-negative breast cancer”. In: *Seminars in Cancer Biology* 72 (2021), pp. 198–213. ISSN: 10963650. DOI: 10.1016/j.semcancer.2020.05.010. URL: <https://doi.org/10.1016/j.semcancer.2020.05.010>.
- [26] National Library of Medicine. *PubChem*. URL: <https://pubchem.ncbi.nlm.nih.gov/>.
- [27] Jonathan Baets et al. “Defects of mutant DNMT1 are linked to a spectrum of neurological disorders”. In: *Brain* 138.4 (2015), pp. 845–861. ISSN: 14602156. DOI: 10.1093/brain/awv010.
- [28] Tingshan Tan et al. “Epigenetic modification regulates tumor progression and metastasis through EMT (Review)”. In: *International Journal of Oncology* 60.6 (2022), pp. 1–17. ISSN: 17912423. DOI: 10.3892/ijo.2022.5360.
- [29] Eunsohl Lee et al. “DNMT1 Regulates Epithelial-Mesenchymal Transition and Cancer Stem Cells, Which Promotes Prostate Cancer Metastasis”. In: *Neoplasia (United States)* 18.9 (2016), pp. 553–566. ISSN: 14765586. DOI: 10.1016/j.neo.2016.07.007. URL: <http://dx.doi.org/10.1016/j.neo.2016.07.007>.
- [30] Pavel Bashtrykov et al. “Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain”. In: *ChemBioChem* 15.5 (2014), pp. 743–748. ISSN: 14397633. DOI: 10.1002/cbic.201300740.

- [31] Željko M. Svedružić. *Dnmt1: Structure and function*. Vol. 101. 2011, pp. 221–254. ISBN: 9780123876850. DOI: 10.1016/B978-0-12-387685-0.00006-8.
- [32] Chengyao Chiang et al. “The Epigenetic Regulation of Nonhistone Proteins by SETD7: New Targets in Cancer”. In: *Frontiers in Genetics* 13.June (2022), pp. 1–11. DOI: 10.3389/fgene.2022.918509.
- [33] Mangala Hegde and Manjunath B. Joshi. *Comprehensive analysis of regulation of DNA methyltransferase isoforms in human breast tumors*. Vol. 147. 4. Springer Berlin Heidelberg, 2021, pp. 937–971. ISBN: 0123456789. DOI: 10.1007/s00432-021-03519-4. URL: <https://doi.org/10.1007/s00432-021-03519-4>.
- [34] Shannon R. Morey Kinney and Sriharsa Pradhan. *Regulation of expression and activity of DNA (Cytosine-5) methyltransferases in mammalian cells*. 1st ed. Vol. 101. Elsevier Inc., 2011, pp. 311–333. ISBN: 9780123876850. DOI: 10.1016/B978-0-12-387685-0.00009-3. URL: <http://dx.doi.org/10.1016/B978-0-12-387685-0.00009-3>.
- [35] Takamasa Ito et al. “DNMT1 regulates the timing of DNA methylation by DNMT3 in an enzymatic activity-dependent manner in mouse embryonic stem cells”. In: *PLoS ONE* 17.1 January (2022), pp. 1–21. ISSN: 19326203. DOI: 10.1371/journal.pone.0262277. URL: <http://dx.doi.org/10.1371/journal.pone.0262277>.
- [36] Zhi-Min Zhang et al. “Crystal Structure of Human DNA Methyltransferase”. In: *Journal of Molecular Biology* 427.15 (2015), pp. 2520–2531.
- [37] Jikui Song et al. “Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation”. In: *Science* 335.6069 (2012), pp. 709–712. ISSN: 10959203. DOI: 10.1126/science.1214453.
- [38] Tetsuo Iida et al. “PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA”. In: *Genes to Cells* 7.10 (2002), pp. 997–1007. ISSN: 13569597. DOI: 10.1046/j.1365-2443.2002.00584.x.
- [39] Hiromichi Kimura and Kunio Shiota. “Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1”. In: *Journal of Biological Chemistry* 278.7 (2003), pp. 4806–4812. ISSN: 00219258. DOI: 10.1074/jbc.M209923200. URL: <http://dx.doi.org/10.1074/jbc.M209923200>.
- [40] Mary Grace Goll and Timothy H. Bestor. “Eukaryotic cytosine methyltransferases”. In: *Annual Review of Biochemistry* 74 (2005), pp. 481–514. ISSN: 00664154. DOI: 10.1146/annurev.biochem.74.010904.153721.
- [41] Melissa B. Pappalardi et al. “Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia”. In: *Nature Cancer* 2.10 (2021), pp. 1002–1017. ISSN: 26621347. DOI: 10.1038/s43018-021-00249-x.

- [42] K. Eurídice Juárez-Mercado et al. “DNA methyltransferase inhibitors with novel chemical scaffolds”. In: *bioRxiv* (2020), pp. 1–22. ISSN: 26928205.
- [43] Clemens Zwergel et al. “Identification of a novel quinoline-based DNA demethylating compound highly potent in cancer cells”. In: *Clinical Epigenetics* 11.1 (2019), pp. 1–18. ISSN: 18687083. DOI: 10.1186/s13148-019-0663-8.
- [44] Shijie Chen et al. “Identifying novel selective non-nucleoside DNA methyltransferase 1 inhibitors through docking-based virtual screening”. In: *Journal of Medicinal Chemistry* 57.21 (2014), pp. 9028–9041. ISSN: 15204804. DOI: 10.1021/jm501134e.
- [45] José L. Medina-Franco, Edgar López-López, and Liliam P. Martínez-Fernández. “7-Aminoalkoxy-Quinazolines from Epigenetic Focused Libraries Are Potent and Selective Inhibitors of DNA Methyltransferase 1”. In: *Molecules* 27.9 (2022), p. 2892. DOI: 10.3390/molecules27092892.
- [46] Burnham Center for Chemical Genomics. *Dose response confirmation of DNMT1 inhibitors in a Fluorescent Molecular Beacon assay*. 2012. URL: <https://pubchem.ncbi.nlm.nih.gov/bioassay/602386>.
- [47] Nathalia Azevedo Portilho et al. “The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells”. In: *Epigenetics and Chromatin* 14.1 (2021), pp. 1–12. ISSN: 17568935. DOI: 10.1186/s13072-021-00429-0. URL: <https://doi.org/10.1186/s13072-021-00429-0>.
- [48] Pablo Ramírez et al. “Estado actual del xenotrasplante de órganos”. In: *Cirugía Española* 72.4 (2002), pp. 222–231. ISSN: 0009739X. DOI: 10.1016/S0009-739X(02)72045-9. URL: [http://dx.doi.org/10.1016/S0009-739X\(02\)72045-9](http://dx.doi.org/10.1016/S0009-739X(02)72045-9).
- [49] John J Irwin and Brian K Shoichet. “ZINC – A Free Database of Commercially Available Compounds for Virtual Screening”. In: *Journal of chemical information and modeling* 45.1 (2005), pp. 177–182.
- [50] E. Alexis Flores-Padilla et al. “Chemoinformatic Characterization of Synthetic Screening Libraries Focused on Epigenetic Targets”. In: *Molecular Informatics* (2021), pp. 1–23. ISSN: 18681751. DOI: 10.1002/minf.202100285.
- [51] Adam I. Green and George M. Burslem. “Focused Libraries for Epigenetic Drug Discovery: The Importance of Isosteres”. In: *Journal of Medicinal Chemistry* 64.11 (2021), pp. 7231–7240. ISSN: 15204804. DOI: 10.1021/acs.jmedchem.1c00592.
- [52] José L. Medina-Franco, Eli Fernán-Dezde Gortari, and J. Jesús Naveja. “Avances en el diseño de fármacos asistido por computadora”. In: *Educacion Quimica* 26.3 (2015), pp. 180–186. ISSN: 18708404. DOI: 10.1016/j.eq.2015.05.002.

- [53] V. J. Gillet and A. R. Leach. *Chemoinformatics*. Vol. 3. December 2018. 2006, pp. 235–264. ISBN: 9780080450445. DOI: 10.1016/b978-0-12-409547-2.14327-6.
- [54] Veronica Salmaso and Stefano Moro. “Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview”. In: *Frontiers in Pharmacology* 9.AUG (2018), pp. 1–16. ISSN: 16639812. DOI: 10.3389/fphar.2018.00923.
- [55] Yudibeth Sixto-López and José Correa-Basurto. *Docking Approaches Used in Epigenetic Drug Investigations*. Elsevier Inc., 2021, pp. 367–390. ISBN: 9780128223123. DOI: 10.1016/B978-0-12-822312-3.00016-3. URL: <http://dx.doi.org/10.1016/B978-0-12-822312-3.00016-3>.
- [56] Eli Fernández De Gortari and José L. Medina-Franco. “Epigenetic relevant chemical space: a chemoinformatic characterization of inhibitors of DNA methyltransferases”. In: *RSC Advances* 5.106 (2015), pp. 87465–87476. ISSN: 20462069. DOI: 10.1039/c5ra19611f.
- [57] Bárbara I. Díaz-Eufracio et al. “D-Peptide Builder: A Web Service to Enumerate, Analyze, and Visualize the Chemical Space of Combinatorial Peptide Libraries”. In: *Molecular Informatics* 39.11 (2020), pp. 1–7. ISSN: 18681751. DOI: 10.1002/minf.202000035.
- [58] Jerzy Leszczynski et al. “Handbook of computational chemistry”. In: *Handbook of Computational Chemistry* (2017), pp. 1–2381. DOI: 10.1007/978-3-319-27282-5.
- [59] Paula Carracedo-Reboredo et al. “A review on machine learning approaches and trends in drug discovery”. In: *Computational and Structural Biotechnology Journal* 19 (2021), pp. 4538–4558. ISSN: 20010370. DOI: 10.1016/j.csbj.2021.08.011. URL: <https://doi.org/10.1016/j.csbj.2021.08.011>.
- [60] Mariana González-Medina and José L. Medina-Franco. “Platform for Unified Molecular Analysis: PUMA”. In: *Journal of Chemical Information and Modeling* 57.8 (2017), pp. 1735–1740. ISSN: 15205142. DOI: 10.1021/acs.jcim.7b00253.
- [61] Alice Capecchi, Daniel Probst, and Jean Louis Reymond. “One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome”. In: *Journal of Cheminformatics* 12.1 (2020), pp. 1–15. ISSN: 17582946. DOI: 10.1186/s13321-020-00445-4. URL: <https://doi.org/10.1186/s13321-020-00445-4>.
- [62] Na Zhang and Hongtao Zhao. “Enriching screening libraries with bioactive fragment space”. In: *Bioorganic and Medicinal Chemistry Letters* 26.15 (2016), pp. 3594–3597. ISSN: 14643405. DOI: 10.1016/j.bmcl.2016.06.013. URL: <http://dx.doi.org/10.1016/j.bmcl.2016.06.013>.

- [63] Sarah R. Langdon, Julian Blagg, and Nathan Brown. “Scaffold diversity in medicinal chemistry space”. In: *Scaffold Hopping in Medicinal Chemistry* 58 (2013), pp. 39–60. DOI: 10.1002/9783527665143.ch03.
- [64] Guy W. Bemis and Mark A. Murcko. “The properties of known drugs. 1. Molecular frameworks”. In: *Journal of Medicinal Chemistry* 39.15 (1996), pp. 2887–2893. ISSN: 00222623. DOI: 10.1021/jm9602928.
- [65] Yiqun Cao, Tao Jiang, and Thomas Girke. “A maximum common substructure-based algorithm for searching and predicting drug-like compounds”. In: *Bioinformatics* 24.13 (2008), pp. 366–374. ISSN: 13674803. DOI: 10.1093/bioinformatics/btn186.
- [66] Aleix Gimeno et al. “The light and dark sides of virtual screening: What is there to know?”. In: *International Journal of Molecular Sciences* 20.6 (2019). ISSN: 14220067. DOI: 10.3390/ijms20061375.
- [67] Nikhil Pillai et al. “Machine Learning guided early drug discovery of small molecules”. In: *Drug Discovery Today* xxx.xx (2022). ISSN: 13596446. DOI: 10.1016/j.drudis.2022.03.017. URL: <https://doi.org/10.1016/j.drudis.2022.03.017>.
- [68] Mohammad Shehab et al. “Machine learning in medical applications: A review of state-of-the-art methods”. In: *Computers in Biology and Medicine* 145.March (2022), p. 105458. ISSN: 00104825. DOI: 10.1016/j.combiomed.2022.105458. URL: <https://doi.org/10.1016/j.combiomed.2022.105458>.
- [69] F. M. Delpino et al. “Machine learning for predicting chronic diseases: a systematic review”. In: *Public Health* 205 (2022), pp. 14–25. ISSN: 14765616. DOI: 10.1016/j.puhe.2022.01.007.
- [70] Dominic D. Martinelli. “Generative machine learning for de novo drug discovery: A systematic review”. In: *Computers in Biology and Medicine* 145.March (2022), p. 105403. ISSN: 18790534. DOI: 10.1016/j.combiomed.2022.105403. URL: <https://doi.org/10.1016/j.combiomed.2022.105403>.
- [71] Mahinda Mailagaha Kumbure et al. “Machine learning techniques and data for stock market forecasting: A literature review”. In: *Expert Systems with Applications* 197.February (2022), p. 116659. ISSN: 09574174. DOI: 10.1016/j.eswa.2022.116659. URL: <https://doi.org/10.1016/j.eswa.2022.116659>.
- [72] Jie Yuan et al. “Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review”. In: *Journal of Neuroscience Methods* 368.October 2021 (2022), p. 109441. ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2021.109441. URL: <https://doi.org/10.1016/j.jneumeth.2021.109441>.

- [73] Michael P. Menden et al. “Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen”. In: *Nature Communications* 10.1 (2019), pp. 1–17. ISSN: 20411723. DOI: 10.1038/s41467-019-09799-2.
- [74] Corwin Hansch and Toshio Fujita. “ ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure”. In: *Journal of the American Chemical Society* 86.8 (1964), pp. 1616–1626. ISSN: 15205126. DOI: 10.1021/ja01062a035.
- [75] Norberto Sánchez-Cruz and José L. Medina-Franco. “Epigenetic Target Fishing with Accurate Machine Learning Models”. In: *Journal of Medicinal Chemistry* (2021). ISSN: 0022-2623. DOI: 10.1021/acs.jmedchem.1c00020.
- [76] Norberto Sánchez-Cruz and José L. Medina-Franco. “Epigenetic Target Profiler: A Web Server to Predict Epigenetic Targets of Small Molecules”. In: *Journal of Chemical Information and Modeling* (2021), pp. 1550–1554. ISSN: 15205142. DOI: 10.1021/acs.jcim.1c00045.
- [77] Feng Zhu et al. “Clinical Success of Drug Targets Prospectively Predicted by In Silico Study”. In: *Trends in Pharmacological Sciences* 39.3 (2018), pp. 229–231. ISSN: 18733735. DOI: 10.1016/j.tips.2017.12.002. URL: <http://dx.doi.org/10.1016/j.tips.2017.12.002>.
- [78] Ying Hong Li et al. “Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs”. In: *Briefings in Bioinformatics* 21.2 (2020), pp. 649–662. ISSN: 14774054. DOI: 10.1093/bib/bby130.
- [79] Hong Wang et al. “A brief review of machine learning methods for RNA methylation sites prediction”. In: *Methods* February 2022 (2022). ISSN: 10959130. DOI: 10.1016/j.ymeth.2022.03.001. URL: <https://doi.org/10.1016/j.ymeth.2022.03.001>.
- [80] Veer Patel and Manan Shah. “A comprehensive study on artificial intelligence and machine learning in drug discovery and drug development”. In: *Intelligent Medicine* October 2021 (2021), pp. 1–7. ISSN: 26671026. DOI: 10.1016/j.imed.2021.10.001. URL: <https://doi.org/10.1016/j.imed.2021.10.001>.
- [81] Florian Tscherne et al. “Orange: Data Mining Toolbox in Python”. In: *International Journal of Conservation Science* 7.SpecialIssue1 (2016), pp. 295–300. ISSN: 20678223.
- [82] Bruno Junior Neves et al. *Best Practices for Docking-Based Virtual Screening*. Elsevier Inc., 2021, pp. 75–98. ISBN: 9780128223123. DOI: 10.1016/B978-0-12-822312-3.00001-1. URL: <http://dx.doi.org/10.1016/B978-0-12-822312-3.00001-1>.
- [83] *Autodock Vina*. URL: <https://vina.scripps.edu/>.

- [84] Sailu Sarvagalla et al. *Resources for Docking-Based Virtual Screening*. Elsevier Inc., 2021, pp. 179–203. ISBN: 9780128223123. DOI: 10.1016/B978-0-12-822312-3.00002-3. URL: <http://dx.doi.org/10.1016/B978-0-12-822312-3.00002-3>.
- [85] Zhe Wang, Huiyong Sun, and Yao Xiaojun. “Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power”. In: *Physical Chemistry Chemical Physics* 18 (2016).
- [86] Ellen J.B. Derissen, Jos H. Beijnen, and Jan H.M. Schellens. “Concise Drug Review: Azacitidine and Decitabine”. In: *The Oncologist* 18.5 (2013), pp. 619–624. ISSN: 1083-7159. DOI: 10.1634/theoncologist.2012-0465.
- [87] Asinex. *Epigenetic Library*.
- [88] Axon MedChem. *Axon Ligands Epigenetic compound library*. URL: <https://www.axonmedchem.com/product/5052>.
- [89] ChemDiv. *Epigenetics Focused Set*. URL: <https://www.chemdiv.com/catalog/focused-and-targeted-libraries/epigenetics-focused-set/>.
- [90] Enamine. *Epigenetics Library*. URL: <https://enamine.net/hit-finding/focused-libraries/view-all/epigenetics-libraries>.
- [91] Enzo Life Sciences. *SCREEN-WELL® Epigenetics library*. URL: <https://www.enzolifesciences.com/BML-2836/screen-well-epigenetics-library/>.
- [92] Life Chemicals. *Epigenetic Screening Libraries*. URL: <https://lifechemicals.com/screening-libraries/targeted-and-focused-screening-libraries/epigenetic-screening-libraries>.
- [93] Otava Chemicals. *DNA (cytosine-5)-methyltransferase 1 (DNMT1) inhibitor library*. URL: <https://www.otavachemicals.com/products/targeted-libraries-and-focused-libraries/epigenetic-targets>.
- [94] Otava Chemicals. *DNA (cytosine-5)-methyltransferase 3 beta (DNMT3b) inhibitor library*.
- [95] Selleck Chemicals. *Epigenetics Compound Library*. URL: <https://www.selleckchem.com/screening/epigenetics-compound-library.html>.
- [96] Target Mols. *Epigenetic Inhibitor Library*. URL: <https://www.targetmol.com/compound-library/Epigenetics-Inhibitor-Library>.
- [97] Ana L. Chávez-Hernández, Norberto Sánchez-Cruz, and José L. Medina-Franco. “A Fragment Library of Natural Products and its Comparative Chemoinformatic Characterization”. In: *Molecular Informatics* 39.11 (2020). ISSN: 18681751. DOI: 10.1002/minf.202000050.
- [98] Christopher A. Lipinski et al. “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”. In: *Advanced Drug Delivery Reviews* 64.SUPPL. (2012), pp. 4–17. ISSN: 0169409X. DOI: 10.1016/j.addr.2012.09.019.

- [99] Daniel F. Veber et al. “Molecular properties that influence the oral bioavailability of drug candidates”. In: *Journal of Medicinal Chemistry* 45.12 (2002), pp. 2615–2623. ISSN: 00222623. DOI: 10.1021/jm020017n.
- [100] Sebastian Salentin et al. “PLIP: Fully automated protein-ligand interaction profiler”. In: *Nucleic Acids Research* 43.W1 (2015), W443–W447. ISSN: 13624962. DOI: 10.1093/nar/gkv315.
- [101] *LeDock*. URL: <http://www.lephar.com/>.
- [102] Chemical Computing Group Inc. *Molecular Operating Environment (MOE)*. Montreal, QC, Canada. URL: <http://www.chemcomp.com>.
- [103] Corrado Lodovico Galli et al. “A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites”. In: *PLoS ONE* 9.8 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0104822.
- [104] Fabio Urbina, Ana C. Puhl, and Sean Ekins. “Recent advances in drug repurposing using machine learning”. In: *Current Opinion in Chemical Biology* 65 (2021), pp. 74–84. ISSN: 18790402. DOI: 10.1016/j.cbpa.2021.06.001. URL: <https://doi.org/10.1016/j.cbpa.2021.06.001>.
- [105] B. Angélica Pilón-Jiménez et al. “BIOFACQUIM: A Mexican compound database of natural products”. In: *Biomolecules* 9.1 (2019). ISSN: 2218273X. DOI: 10.3390/biom9010031.
- [106] Janna Hastings et al. “ChEBI in 2016: Improved services and an expanding collection of metabolites”. In: *Nucleic Acids Research* 44.D1 (2016), pp. D1214–D1219. ISSN: 13624962. DOI: 10.1093/nar/gkv1031.
- [107] Cristian Rojas et al. “ChemTastesDB: A curated database of molecular tastants”. In: *Food Chemistry: Molecular Sciences* 4.February (2022), p. 100090. ISSN: 26665662. DOI: 10.1016/j.fochms.2022.100090. URL: <https://doi.org/10.1016/j.fochms.2022.100090>.
- [108] David S. Wishart et al. “DrugBank 5.0: A major update to the DrugBank database for 2018”. In: *Nucleic Acids Research* 46.D1 (2018), pp. D1074–D1082. ISSN: 13624962. DOI: 10.1093/nar/gkx1037.
- [109] David Wishart et al. “T3DB: The toxic exposome database”. In: *Nucleic Acids Research* 43.D1 (2015), pp. D928–D934. ISSN: 13624962. DOI: 10.1093/nar/gku1004.
- [110] Edgar López-López, Eli Fernández-de Gortari, and José L. Medina-Franco. “Yes SIR! On the structure–inactivity relationships in drug discovery”. In: *Drug Discovery Today* 27.8 (2022), pp. 2353–2362. ISSN: 13596446. DOI: 10.1016/j.drudis.2022.05.005.

Apéndice A

Acoplamiento molecular

En las figuras A.1 a A.11 se muestran los resultados del acoplamiento molecular para los compuestos muy activadores y activadores. En estas figuras A corresponde a RFTS (AA351-600), B a CXXC (AA651-697), C al enlazador autoinhibidor (AA697-728), D a BAH1 (AA728-900), E es BAH2 (AA906-1100), F es el dominio de metiltransferasa (AA1139-1616) G es el compuesto SAM y H corresponde a los enlazadores entre los dominios, I es el compuesto.

A.1 Muy activadores

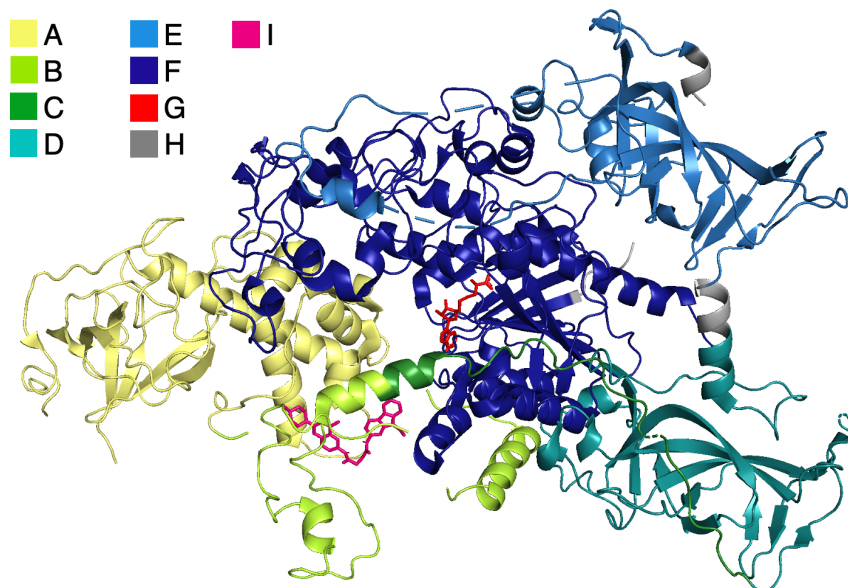


Figura A.1: I es el compuesto C066-2421.

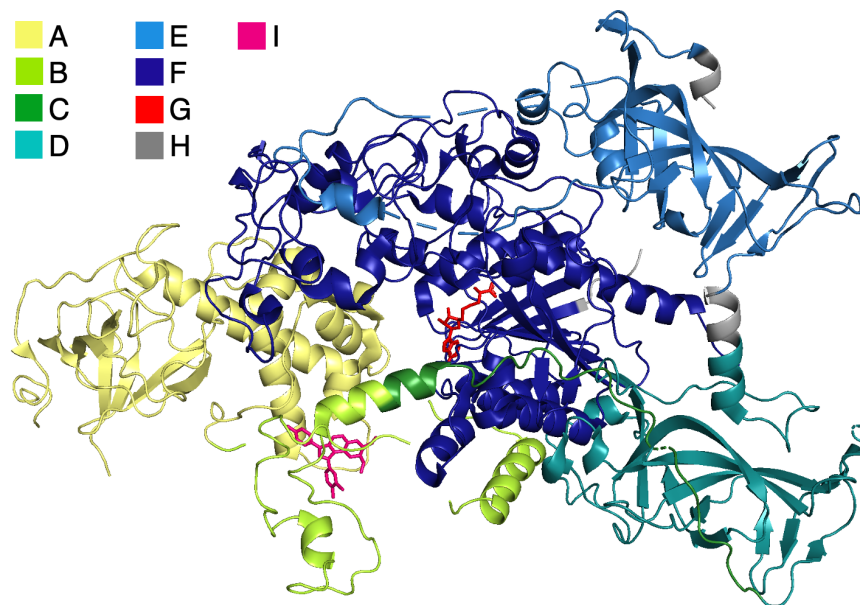


Figura A.2: I es el compuesto C350-0522.

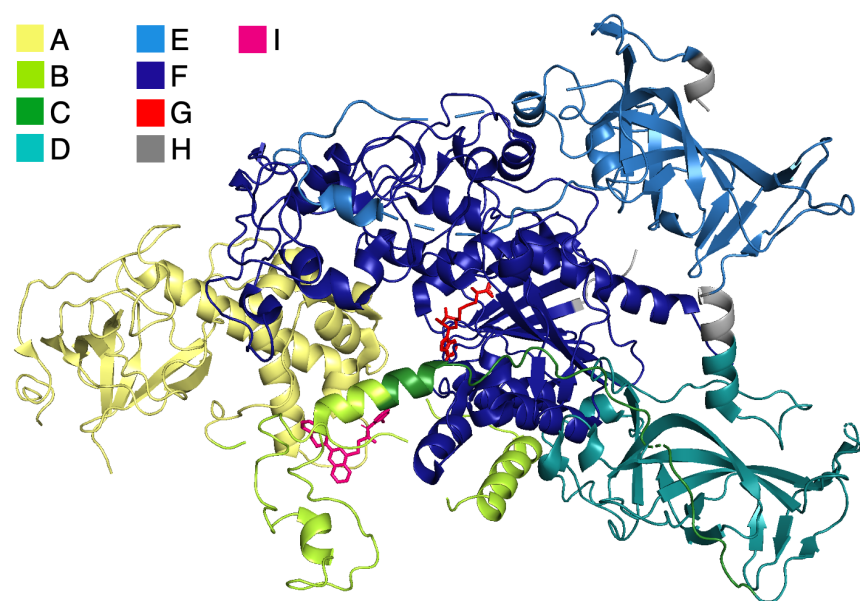


Figura A.3: I es el compuesto L112-0668.

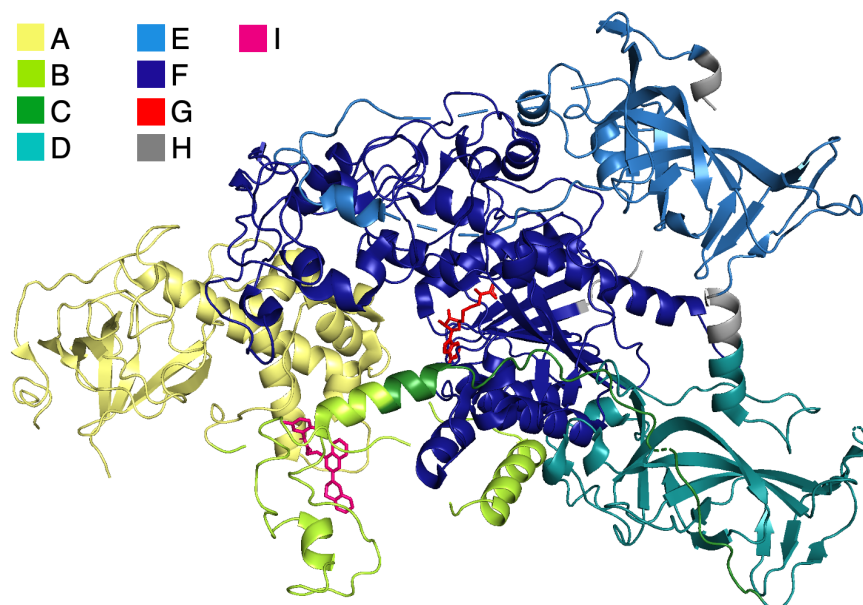


Figura A.4: I es el compuesto L112-0725.

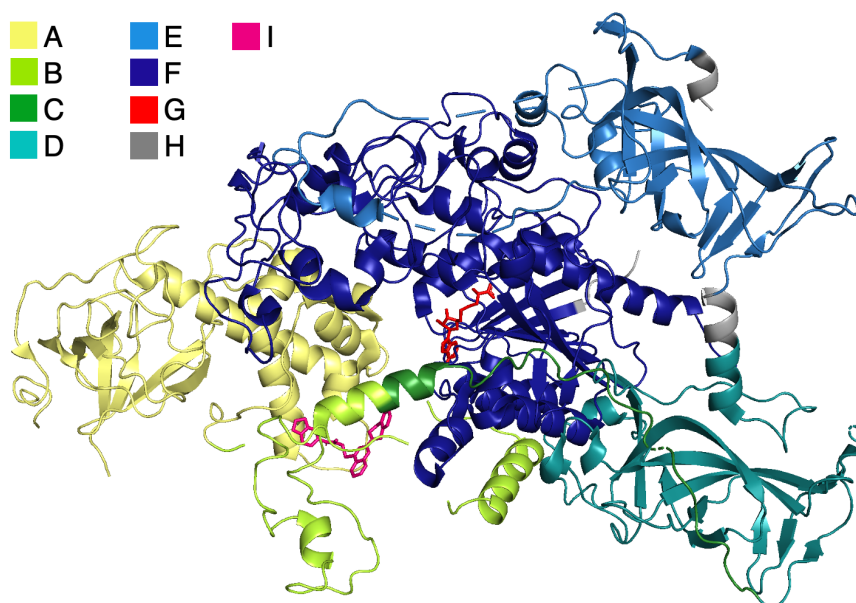


Figura A.5: I es el compuesto L112-0728.

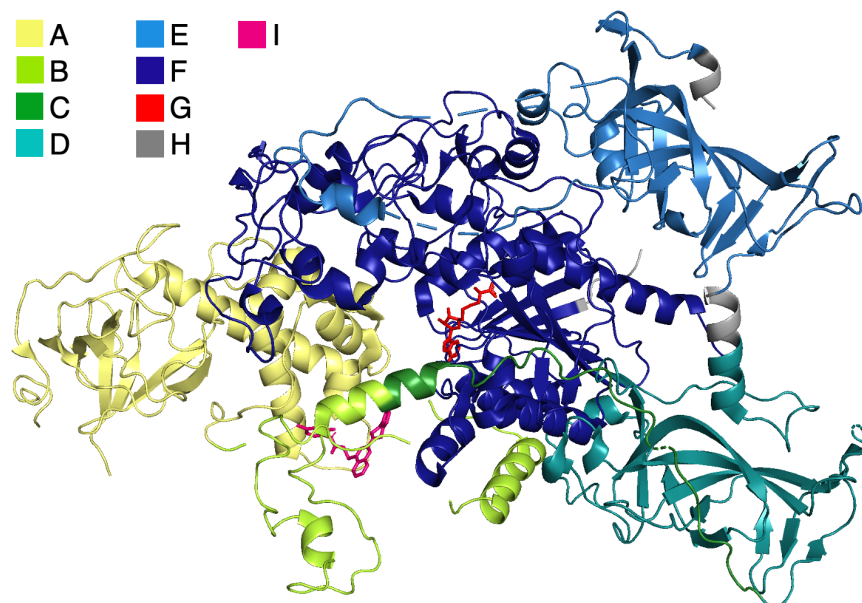


Figura A.6: I es el compuesto L112-0731.

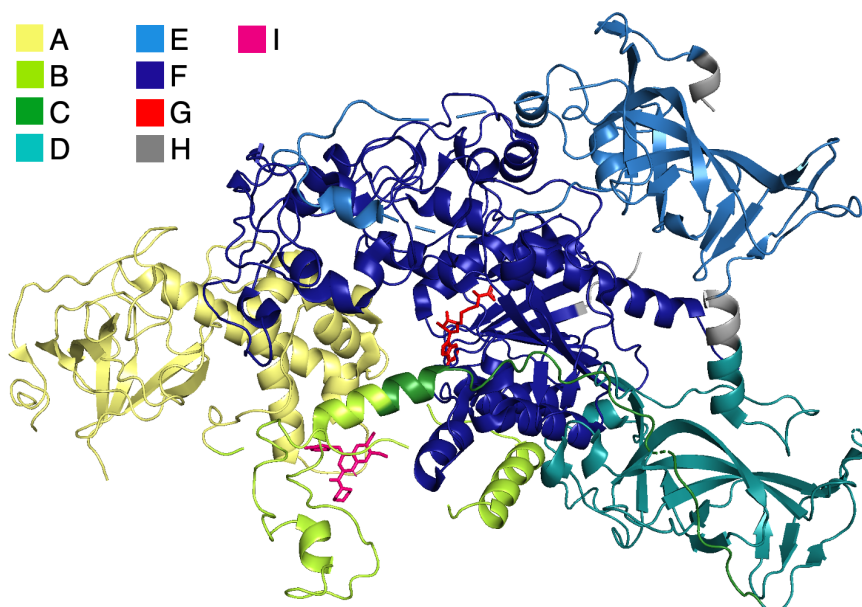


Figura A.7: I es el compuesto L485-2669.

A.2 Activadores

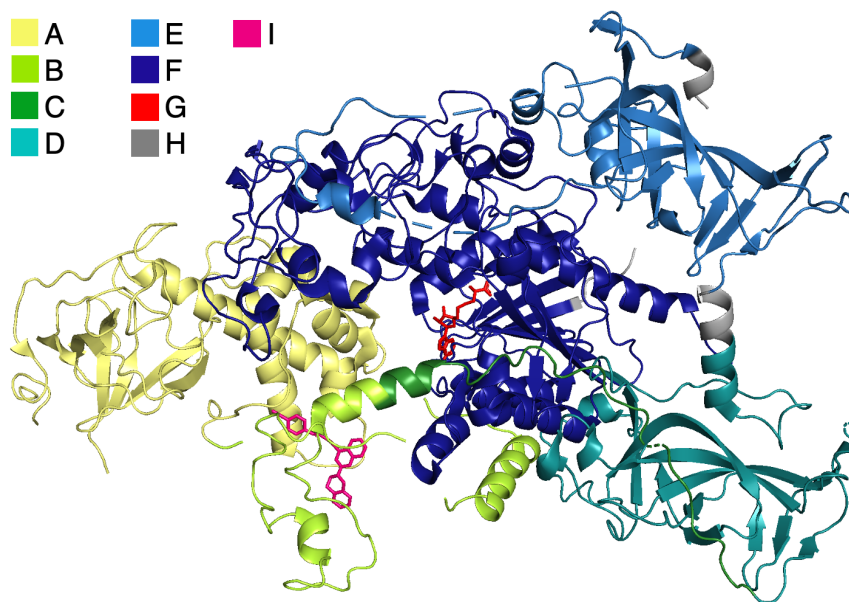


Figura A.8: I es el compuesto L112-0688.

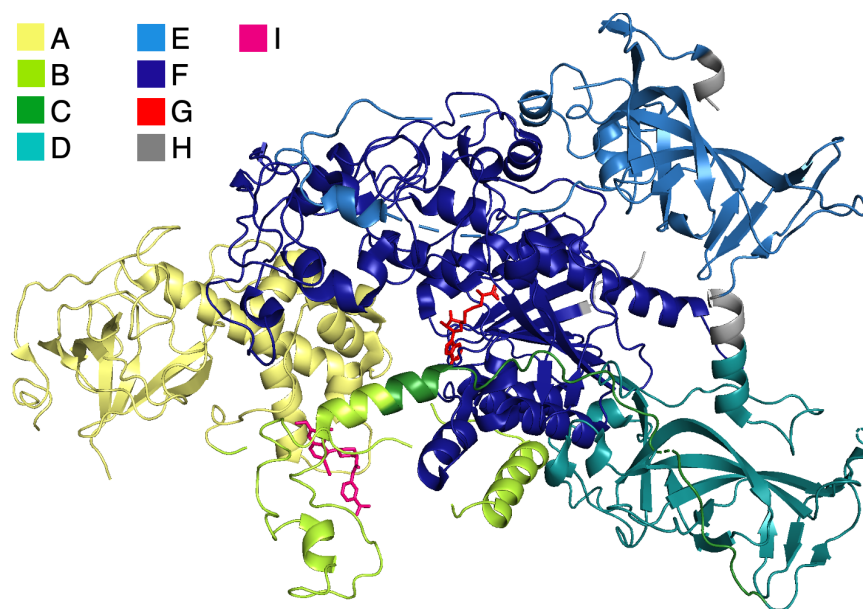


Figura A.9: I es el compuesto L483-0323.

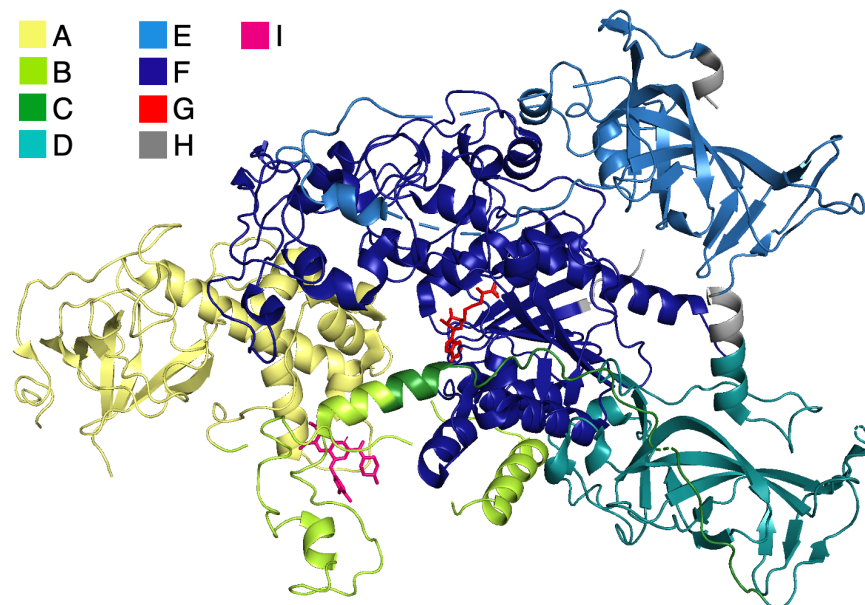


Figura A.10: I es el compuesto L485-2761.

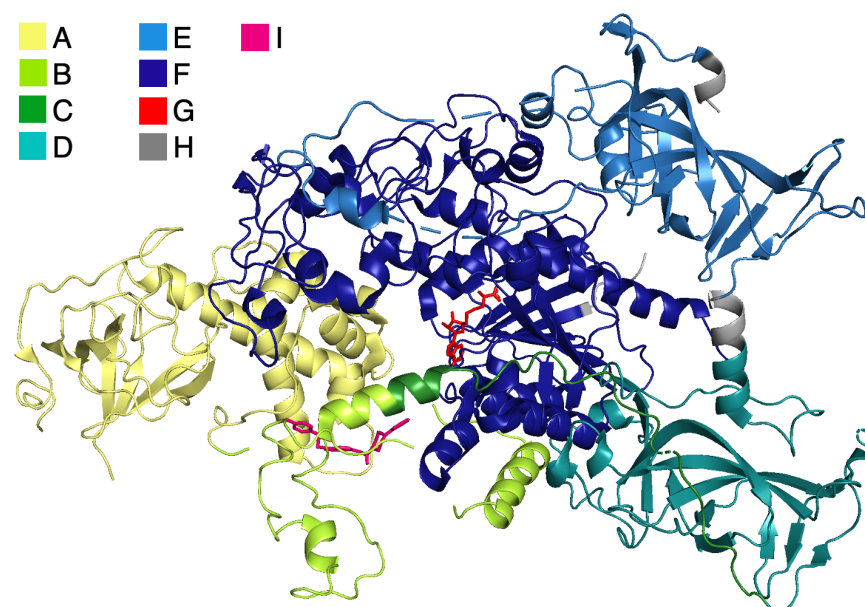


Figura A.11: I es el compuesto M056-0726.

Apéndice B

Pósteres científicos y publicaciones

Se presentó un póster científico en la séptima conferencia electrónica internacional de química médica organizado por el Instituto multidisciplinario de publicaciones digitales (MDPI, por sus siglas en inglés, *Multidisciplinary Digital Publishing Institute*) durante noviembre 1 a noviembre 30 del 2021. Además, se presentó un cartel en el congreso de la sociedad química americana (ACS, por sus siglas en inglés, *American Chemical Society*) en las sesiones del 21 y 22 de agosto de 2022.

En 2022 se publicó un artículo titulado "7-Aminoalkoxy-Quinazolines from Epigenetic Focused Libraries are Potent and Selective Inhibitors of DNA Methyltransferase 1" en la revista de *Molecules* en las páginas 2892 a 2908.

Identification of novel DNA Methyltransferase 1 (DNMT1) inhibitors from focused databases

Liliam Martínez¹, Jose Luis Medina Franco¹

¹School of Chemistry, Department of Pharmacy, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico

INTRODUCTION

DNA methylation is associated with various diseases including psychiatric diseases, diseases of the immune system, and contributes to both the initiation and progression of various cancers. This reaction is mediated by a family of enzymes called DNA methyltransferases (DNMTs), including DNMT1. However, despite the importance of DNMT1, it has not been possible to develop drugs that allow its inhibition without potent cytotoxic effects. Thus, it is crucial to use cheminformatic tools to discover DNMT1 inhibitors that do not have cytotoxic effects.

AIM

Identification of novel DNA Methyltransferase 1 (DNMT1) inhibitors from focused databases that do not have the cytotoxic side effects current approved medications.

METHODOLOGY

1. Database selection

A code was made to establish a canonical SMILES for each compound and thus obtain its physicochemical properties. Afterwards, PUMA was used to perform an analysis of chemical space, scaffold diversity, and fingerprint similarity.

Focused libraries

OTAVA database

2. Selection of compounds

Molecular docking analysis and different machine learning methods including Epigenetic Target Profiler

OTAVA database

20 compounds

3. Assessment of selected compounds

Selected compounds are currently being evaluated for their inhibitory activity against DNMT1

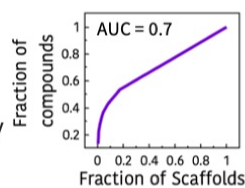
20 compounds

Active compounds

RESULTS

Database assessment

97% of the compounds comply with the 6 rules of Lipinski and Veber



M	399
N	226
Ns	187
N/M	57%
Ns/N	83%
P25	1%
P50	15%
P75	56%

Molecular Docking

DNMT1: 4WXX
Autodock vina

Score	ΣCompounds*	Compounds**
-5	401	4
-5.5	397	16
-6	381	27
-6.5	354	52
-7	302	47
-7.5	255	91
-8	164	84
-8.5	80	51
-9	29	23
-9.5	6	5
-10	1	1

* = Σcompounds > score

** = score + 0.5 ≥ compounds > score

Machine Learning using fingerprints

Fingerprints used:
Morgan by Random Forest
RDk by SVM

		RDk: Fingerprints	
		Positive	Negative
Morgan: Fingerprints	Positive	9	19
	Negative	33	340

Toxicological Assessment

Currently compounds are being assessed for their cytotoxic activity as well as inhibitory activity

All compounds were studied using Toxtree 3.1 by the following protocols:

- Cramer Class
- Start Biodegradability
- DNA Binding Alerts
- Carcinogenicity (genotox and nongenotox) and mutagenicity rulebase by ISS

REFERENCES

- [1] K. K. Wong, "DNMT1: A key drug target in triple-negative breast cancer," *Semin. Cancer Biol.*, vol. 72, pp. 198–213, 2021, doi: 10.1016/j.semcancer.2020.05.010.
- [2] and J. S. Zhi-Min Zhang1, Shuo Liu2, 3, Krystal Lin1, Youfu Luo4, John Jefferson Perry1, Yinsheng Wang2, 3, "Crystal Structure of Human DNA Methyltransferase 1," *Physiol. Behav.*, vol. 176, no. 5, pp. 139–148, 2017, doi: 10.1016/j.physbeh.2017.03.040.
- [3] L. A. Al-Kharashi, F. H. Al-Mohanna, A. Tulbah, and A. Aboussekhra, "The DNA methyl-transferase protein DNMT1 enhances tumorpromoting properties of breast stromal fibroblasts," *Oncotarget*, vol. 9, no. 2, pp. 2329–2343, 2018, doi: 10.18632/oncotarget.23411.
- [4] B. Jin, Y. Li, and K. D. Robertson, "DNA methylation: Superior or subordinate in the epigenetic hierarchy?" *Genes and Cancer*, vol. 2, no. 6, pp. 607–617, 2011, doi: 10.1177/1947601910393957.
- [5] E. J. B. Derissen, J. H. Beijnen, and J. H. M. Schellens, "Concise Drug Review: Azacitidine and Decitabine," *Oncologist*, vol. 18, no. 5, pp. 619–624, 2013, doi: 10.1634/theoncologist.2012-0465.
- [6] A. Ganesan, "Epigenetic drug discovery: A success story for cofactor interference," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 373, no. 1748, 2018, doi: 10.1098/rstb.2017.0069.



The 7th International Electronic Conference on Medicinal Chemistry
01–30 NOVEMBER 2021 | ONLINE



The 7th International Electronic Conference on Medicinal Chemistry (ECMC 2021)

01-30 NOVEMBER 2021 | ONLINE

Certificate of Participation

This certificate of attendance is given to

Liliam Parvati Martínez Fernández, Jose Luis Medina Franco

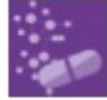
Identification of novel DNA Methyltransferase 1 (DNMT1) inhibitors from focused databases

for participating at

The 7th International Electronic Conference on Medicinal Chemistry

01-30 November 2021, online

ecmc2021.sciforum.net



pharmaceuticals

Dr. Jean Jacques Vanden Eynde
Conference Chair

Pablo Velázquez
Conference Manager

7-Aminoalkoxy-Quinazolines from Epigenetic Focused Libraries are Potent and Selective Inhibitors of DNA Methyltransferase 1

José L. Medina-Franco, Edgar López-López, Lilium P. Martínez-Fernández

Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City 04510, Mexico

Abstract

Inhibitors of epigenetic writers such as DNA methyltransferases (DNMTs) are attractive compounds for epigenetic drug and probe discovery. Herein, we report the identification of two quinazoline-based derivatives identified in focused libraries with micromolar selective inhibition of DNMT1 (30 and 81 nM). Quinazolines reported in this work are known to have low cell toxicity and be potent inhibitors of the epigenetic target G9a. Therefore, the quinazoline-based compounds presented are attractive not only as novel potent inhibitors of DNMTs but also as dual and selective epigenetic agents targeting two families of epigenetic writers.

Background

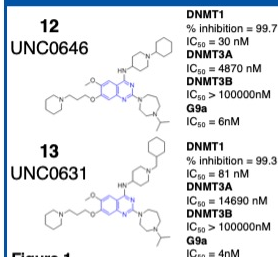
DNMT1 transfers methyl groups to cytosine nucleotides of genomic DNA. This protein is the major enzyme responsible for maintaining methylation patterns following DNA replication. G9a is an enzyme that methylates the lysine-9 position of histone H3. Both epigenetic targets participate collaboratively in the development of neoplastic hematological diseases.¹ However, the quinazoline-based derivatives reported so far have low DNMT1 potency (> 2 μ M) and mostly are inactive.

The aim of this work was to continue expanding the chemical space of potential epi-drugs, in particular DNMTi. In this work we report the experimental testing and computational evaluation of MolPort-023-277-153 (UNC0646, **12**) and MolPort-035-789-726 (UNC0631, **13**) with DNMT1, DNMT3A, and DNMT3B. The findings of this work pave the way to continue exploring quinazoline-based derivatives as inhibitors of DNMTs with multi-epigenetic target activity.

Methodology

- Biochemical Inhibition Assays**
 - Low volume radioisotope-based assay that uses tritium-labeled AdoMet (³H-SAM)
- Protein and ligand preparation**
 - DNMT1 (PDB ID: 4WXX)
 - Yasara Software
 - Ligands: MMFF94x forcefield in MOE software
- Molecular Docking**
 - MOE software
 - Grid center: GLU 1266
 - Triangle matcher method
 - RMSD < 2Å
- Molecular Dynamics**
 - Maestro – Desmond software
 - OPLS 2005 force field.
 - TIP3P water model.
 - 150 ns in presence of the cofactor(SAM)

Results and Discussion



The two quinazolines reported in this work are few of the selective inhibitors of DNMT1 over DNMT3A and DNMT3B (Figure 1).

The compounds studied tend to establish the conformation (reducing the RMSD values) of CXXC (647 - 691 aa) and autoinhibition (699 - 733 aa) domains of DNMT1 (Figure 2A3, 2B3). These domains are only present in DNMT1 and not in its 3A or 3B isoforms (Figure 3), which could explain the selective enzymatic inactivation of DNMT1.^{2,3}

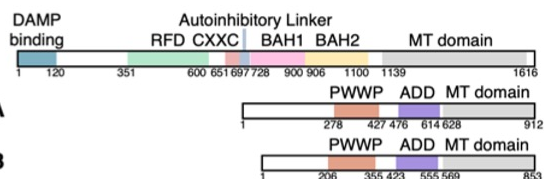
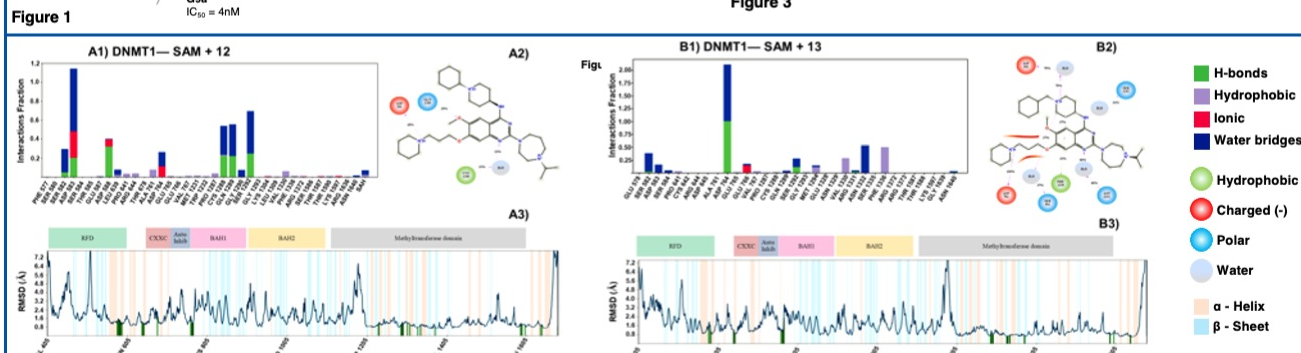


Figure 3



Conclusions

We identified two 7-aminoalkoxy-quinazolines: **12** and **13**, with nanomolar inhibition of DNMT1 (30 nM y 81 nM, respectively) in enzymatic inhibition assays.

- Both compounds showed better inhibitory activity than the positive control, SAH (340 μ M).
- Less inhibition of DNMT3A (4.87 and 14.69 μ M).
- None of them inhibited DNMT3B.
- Potent inhibitors of another epigenetic target writer, G9a.

High structural similarity to quinoline-based derivatives which are known to effectively reduce DNA methylation in cell-based assays, strongly support the hypothesis that the quinazolines reported in this work will also be inhibitors in cell based-assays.

Interactions against SER 582, ASP 764, and SER 1292 on DNMT1 in presence of their cofactor SAM. Inhibition of DNMT1 by compounds **12** and **13** is caused by the conformational changes on the CXXC and autoinhibition domains.

Enzymatic DNA—substrate competitive assays showed that compound **12** was not DNA—substrate competitive.

Acknowledgments

We thank the support of DGAPA, UNAM, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Grant No. IN201321, Edgar López-López, Lilium P. Martínez-Fernández are grateful to Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico, for the Ph.D. scholarship numbers, 762342 and 1012509, respectively.

References

- (1) Rabal, O.; Sánchez-Arias, J. A.; San José-Enériz, E.; et al. *J. Med. Chem.* **2018**, 61:6546–6573.
- (2) Medina-Franco, J. L.; López-López, E.; Martínez-Fernández, L. P. *Molecules* **2022**, 27 9:2892.
- (3) López-López, E.; Prieto-Martínez, F. D.; Medina-Franco, J. L. *Molecules* **2018**, 23 12:3282.

Contact

Speakers: M-F. L.P. liliummztfdz@gmail.com
L-L.E. elopez_lopez@cinvestav.mx
Head of Group: M-F. J.L. medinail@unam.mx
Research group website: www.difacquim.com

Check the article

ACSFALL 2022

SUSTAINABILITY IN A CHANGING WORLD

AUGUST 21-25 • CHICAGO, IL

Hybrid #ACSFall2022

THIS IS TO CERTIFY THAT

Liliam Martínez-Fernández

ATTENDED

ACSFALL 2022

AND PRESENTED THE FOLLOWING PAPER

Title: MED1 3741460: 7-Aminoalkoxy-quinazolines from epigenetic focused libraries are potent and selective Inhibitors of DNA methyltransferase 1

Session: General Posters I (In-Person)

Location: McCormick Place Convention Center Hall F2

Date and Time: 8/21/2022 7:00:00 PM

Presentation Type: Poster - In-person



THOMAS M. CONNELLY, JR.
CHIEF EXECUTIVE OFFICER



Chemistry for Life™

AMERICAN CHEMICAL SOCIETY

MEETINGS & EVENTS

ACSFALL 2022

SUSTAINABILITY IN A CHANGING WORLD

AUGUST 21-25 • CHICAGO, IL Hybrid #ACSFall2022

THIS IS TO CERTIFY THAT

Liliam Martínez-Fernández

ATTENDED

ACSFALL 2022

AND PRESENTED THE FOLLOWING PAPER

Title: MEDI 3741460: 7-Aminoalkoxy-quinazolines from epigenetic focused libraries are potent and selective Inhibitors of DNA methyltransferase 1

Session: MEDI Sci-Mix

Location: McCormick Place Convention Center Hall F2

Date and Time: 8/22/2022 8:00:00 PM

Presentation Type: Sci-mix - In-person

THOMAS M. CONNELLY, JR.
CHIEF EXECUTIVE OFFICER

Article

7-Aminoalkoxy-Quinazolines from Epigenetic Focused Libraries Are Potent and Selective Inhibitors of DNA Methyltransferase 1

José L. Medina-Franco ^{1,*} , Edgar López-López ^{1,2}  and Liliam P. Martínez-Fernández ¹

¹ DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City 04510, Mexico; elopez.lopez@cinvestav.mx (E.L.-L.); liliamtzfdz@gmail.com (L.P.M.-F.)

² Department of Pharmacology, Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Mexico City 07360, Mexico

* Correspondence: medinajl@unam.mx

Abstract: Inhibitors of epigenetic writers such as DNA methyltransferases (DNMTs) are attractive compounds for epigenetic drug and probe discovery. To advance epigenetic probes and drug discovery, chemical companies are developing focused libraries for epigenetic targets. Based on a knowledge-based approach, herein we report the identification of two quinazoline-based derivatives identified in focused libraries with sub-micromolar inhibition of DNMT1 (30 and 81 nM), more potent than *S*-adenosylhomocysteine. Also, both compounds had a low micromolar affinity of DNMT3A and did not inhibit DNMT3B. The enzymatic inhibitory activity of DNMT1 and DNMT3A was rationalized with molecular modeling. The quinazolines reported in this work are known to have low cell toxicity and be potent inhibitors of the epigenetic target G9a. Therefore, the quinazoline-based compounds presented are attractive not only as novel potent inhibitors of DNMTs but also as dual and selective epigenetic agents targeting two families of epigenetic writers.

Keywords: docking; drug discovery; enzyme inhibition; epigenetics; epi-informatics; focused library; molecular dynamics; multi-target epigenetic agent; polypharmacology; quinazoline



Citation: Medina-Franco, J.L.; López-López, E.; Martínez-Fernández, L.P. 7-Aminoalkoxy-Quinazolines from Epigenetic Focused Libraries Are Potent and Selective Inhibitors of DNA Methyltransferase 1. *Molecules* **2022**, *27*, 2892. <https://doi.org/10.3390/molecules27092892>

Academic Editors: Paola Arimondo and Frank J. Dekker

Received: 10 April 2022

Accepted: 30 April 2022

Published: 30 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The identification of drug candidates targeting epigenetic targets is of large interest for addressing several therapeutic needs [1,2]. Several epi-drugs currently approved for clinical have been reviewed elsewhere [3]. Similarly, it is attractive identifying tool compounds to understand better epigenetic processes. Among the epigenetic drug and probe candidates, small molecules targeting epigenetic writers such as DNA methyltransferases (DNMTs) and protein lysine methyltransferases (PKMTs) are promising for the treatment of various types of cancer such as colorectal, breast, lung, ovarian, pancreatic cancer and acute myeloid leukemia [4–8], neurological disorders [9,10], autoimmune diseases [11,12], and metabolic diseases [13–15].

DNA methylation is mediated by the enzyme family DNMTs that is responsible for catalyzing the covalent addition of a methyl group from *S*-adenosyl-*L*-methionine (SAM) to the 5-carbon of cytosine, mainly within CpG dinucleotides, yielding *S*-adenosyl-*L*-homocysteine (SAH). DNMT3A and DNMT3B are maintenance meanwhile DNMT1 is a de novo methyltransferase. DNMT3A is overexpressed in vulvar squamous cell carcinoma and pituitary adenoma [16] however, it is mainly associated with hematological malignancies [17]. DNMT3B is overexpressed in lung, ovarian, and breast cancer, hepatocellular carcinoma, and mild traumatic brain injury [17]. DNMT1 is associated with colorectal, pancreatic, gastric, lung, and thyroid cancer and pituitary adenoma, lupus [18], and hereditary sensory neuropathy [19].

Two drugs **1** (5-azacitidine, Vidaza) and **2** (decitabine, 5-aza-2'-deoxycytidine, Dacogen) (Figure 1A) are approved for clinical use to treat the myelodysplastic syndrome [5]. The two drugs are nucleoside analogs that inhibit all three DNMTs (1, 3A, and 3B). Because of the chemical nature of these first-generation DNMT inhibitors (DNMTi), these drugs are characterized by substantial cellular and clinical toxicity, which has driven the development of non-nucleoside, novel, and more specific drugs. Currently, more than 400 non-nucleoside compounds have been tested against at least one DNMT (mostly DNMT1) [20]. Figure 1 shows representative structures of DNMTi and compounds (**1–11**) associated with the demethylating activity of DNA. Of note, there is a limited number of selective inhibitors inhibiting DNMTs disclosed so far. For example, **6** (nanaomycin A) (Figure 1A) is a selective inhibitor of DNMT3B (do not inhibit DNMT1) and reactivates silenced tumor suppressor genes in human cancer cells [21]. Recently, **5** (GSK3685032) (Figure 1) was disclosed as the first selective inhibitor of DNMT1 that reduces global methylation, increases expression of target genes, and has antitumor efficacy in acute myeloid leukemia xenograft models [22]. Other related compounds are DNMT1-selective inhibitors that reduce global methylation and increases HbF expression, offering the potential for use in treating sickle cell disease [23]. Furthermore, dual inhibitors of DNMT and other epigenetic targets such as G9a and histone deacetylases (HDACs) are emerging as part of a current trend to develop multi-epi-target inhibitors [24–26].

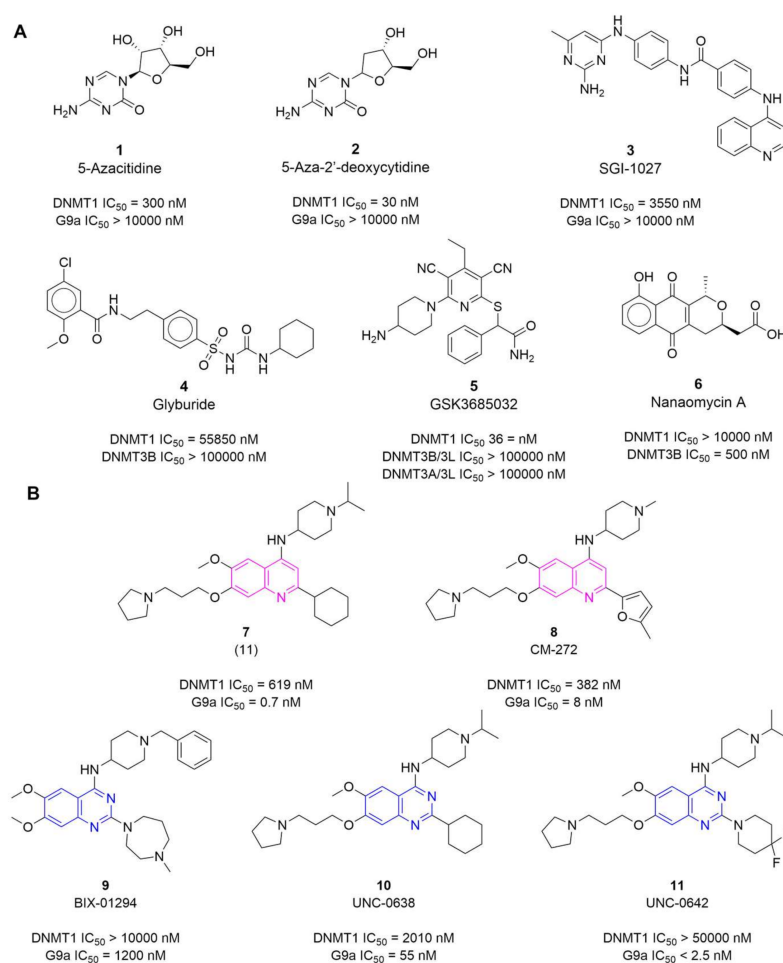


Figure 1. Chemical structures of (A) examples of known DNMT inhibitors and compounds with demethylating activity; (B) dual DNMT1/G9a inhibitors (quinoline- and quinazoline-based derivatives) reported in the literature. The dual activity profile as available in the literature for epigenetic targets is indicated [27].

Notably, DNMTi have been identified from different sources such as natural products and synthetic small molecule libraries [28,29]. Medicinal chemistry, structure- and ligand-based virtual screening, and high-throughput screening of general screening libraries have led to the identification of DNMTi [30]. There are other drug discovery approaches that are used to identify DNMTi such as de novo design [31] and screening of focused libraries. In this regard, chemical companies are developing screening libraries focused on the most therapeutically relevant epigenetic targets. The chemical samples of the libraries are commercially available for experiential testing. Chemoinformatics contents and chemical diversity analysis of such libraries support their use for drug discovery programs [32].

The most promising DNMTi developed so far are molecules with long scaffolds, for example, the 4-aminoquinoline **3** (SGI-1027) and its analogs (Figure 1A) [33]. Other quinoline-based derivatives such as compound **8** (CM-272), where the quinoline ring is the main core scaffold (Figure 1B), have been reported as potent dual inhibitors of DNMT1 and G9a with nanomolar activity [27,34]. Noteworthy, **8** has remarkable in vivo efficacy (70% tumor growth inhibition of a human acute myeloid leukemia xenograft in a mouse model) [27]. The quinolines were developed and further optimized as dual inhibitors based on 7-aminoalkoxy-quinazolines that are inhibitors of G9a such as **12** (MolPort-023-277-153) and **13** (MolPort-035-789-726) (Figure 2) that are potent inhibitors of G9a in enzymatic and in cellular-based assays [35]. Other quinazoline-based derivatives, for example, **9** (BIX-01294), **10** (UNC-0638), and **11** (UNC-0642) (Figure 1B) are potent inhibitors of G9a (at the nanomolar level, in particular **10** and **11** ($IC_{50} \leq 55$ nM) [36]. However, the quinazoline-based derivatives reported so far have low DNMT1 potency (>2 μ M and are mostly inactive, as shown in Figure 1B). As commented above, DNMTi are emerging as part of programs to develop combination therapies in drug cocktails or compounds targeting multiple epigenetic targets.

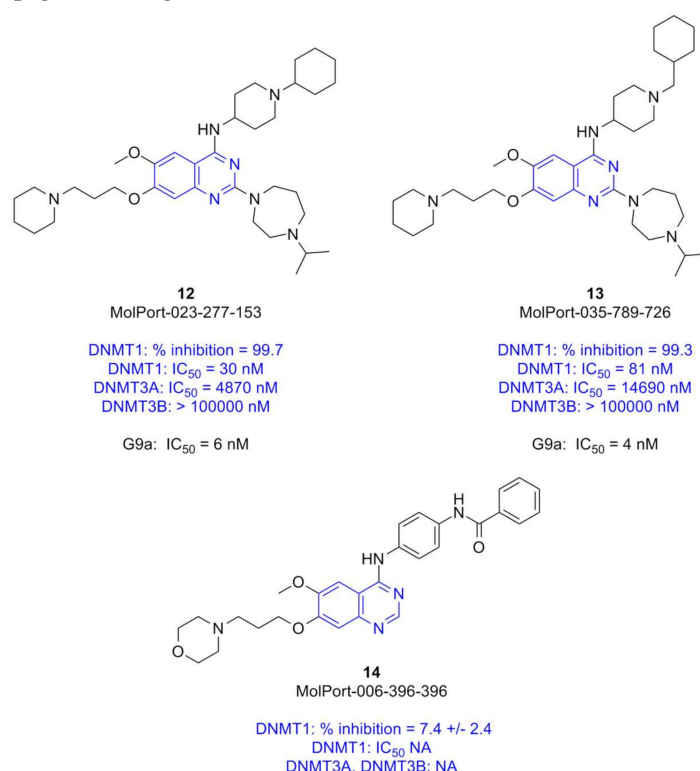


Figure 2. Chemical structures of quinazoline-based derivatives tested in enzymatic inhibition assays. The enzymatic DNMT inhibitory activity measured in this work is included in blue font. Mean value of two measurements. SAH was included as a positive control: IC_{50} (DNMT1) of 0.34 μ M; IC_{50} (DNMT3A) of 0.10 μ M; (DNMT3B) of 0.03 μ M. For reference, the enzymatic G9a inhibitory activity of **12** and **13** [35].

Chemical content analysis of the novel epigenetic-focused screening libraries (vide supra) revealed that there are several quinazoline-based derivatives similar to **10** (UNC-0638) and **11** (UNC-0642), including two compounds reported in the literature as inhibitors of G9a. Based on structure-based design reported in the literature and experimental information discussed above [27,35], herein we hypothesized that quinazoline-based derivatives available in the commercial libraries such as **12** and **13** also inhibit the enzymatic activity of DNMTs.

As part of an ongoing effort to continue expanding the chemical space of DNMTi [31,37], in this work, we report the experimental testing of **12** and **13** and another quinazoline-based derivative (cf. Figure 2) with DNMT1, DNMT3A, and DNMT3B in biochemical enzymatic inhibition assays. We also discuss the molecular dynamics (MD) simulations of the most potent compounds to rationalize their enzymatic inhibitory activity at the molecular level. The findings of this work pave the way to continue exploring quinazoline-based derivatives as inhibitors of DNMTs with multi-epigenetic target activity.

2. Methods

2.1. Compounds for Experimental Screening

Using a knowledge-based approach, for this work, we selected three quinazoline-based derivatives from epigenetic-focused libraries for experimental testing in enzymatic-based assays (Figure 2). The selection was based on the following criterion:

- (1) Knowledge of the promising activity profile of quinoline-based derivatives as dual inhibitors of DNMTs/G9a (vide supra, Figure 1B) in enzymatic and epigenetic functional cellular response;
- (2) High structural similarity of the quinoline-based compounds to the selected quinazoline-based derivatives **12** and **13**, that are known to be potent G9a inhibitors in enzymatic and cell-based assays (vide supra, Figure 1B).
- (3) Commercial availability of the physical samples from the chemical vendors (Table S3 in the Supplementary Materials).

Based on the rationale explained herein, we hypothesized that the three selected quinazolines in Figure 2 could also inhibit the enzymatic activity of DNMTs.

All three compounds were purchased from MolPort Inc. that confirmed the compound's purity (in parenthesis): **12** (100%), **13** (98%), and **14** (99.11%).

2.2. Biochemical DNMT Inhibition Assays

The inhibition of the enzymatic activity of DNMT1, DNMT3A, and DNMT3B was tested using the HotSpotSM platform for methyltransferase assays available at Reaction Biology Corporation [38]. HotSpotSM is a low volume radioisotope-based assay that uses tritium-labeled AdoMet (3H-SAM) as a methyl donor. The three test compounds diluted in dimethyl sulfoxide were added using acoustic technology (Echo550, Labcyte, San Jose, CA, USA) into an enzyme/substrate mixture in the nano-liter range. The corresponding reactions were commenced by adding 3H-SAM and incubated at 30 °C. Total final methylations on the substrate (Poly dI-dC in DNMT1, 3A, and 3B assays) were identified by a filter binding method implemented in Reaction Biology. Data analysis was done with Graphed Prism software (La Jolla, CA, USA) for curve fits. The enzymatic inhibition assays were carried out at 1 µM of SAM. In all assays, SAH was used as a standard positive control. The three compounds were tested first with DNMT1 at one 100 µM concentration in duplicate. The most active molecules were tested as DNMT1, DNMT3A, and DNMT3B inhibitors in 10-concentration IC₅₀ (effective concentration to inhibit enzymatic activity by 50%) with a threefold serial dilution starting at 100 µM. The research group has recently contracted the screening services of Reaction Biology Corporation to identify a novel DNMTi [37].

2.3. Computational Methods

2.3.1. Protein and Ligands Preparation

The crystallographic structure of human DNMT1 (PDB ID: 4WXX) was retrieved from the Protein Data Bank. Available online: <https://www.rcsb.org/> (accessed on 9 April 2022) [39]. Missing loops and side-chains were added with YASARA software [40]. The ligands were built and energy-minimized in MOE using the MMFF94x force field. The more stable protomers at physiological pH were identified [41].

2.3.2. Molecular Docking

Molecular Operating Environment (MOE) software was used to generate the dock conformation of protein-ligand complexes [42]. The grid was centered on the carbon atom of the carboxyl group of GLU 1266 (chain A) with a size of 27 \AA^3 in the presence of the native ligand (SAM). Using the “Triangle Matcher” method, the binding compounds were subjected to 50 search steps (poses) and the default values for the other parameters. The clusters with an RMSD $< 2 \text{ \AA}$ were visually explored. During the docking simulations, the receptor was considered rigid and the ligands flexible. The conformations with the lowest binding energy were selected for an additional MD analysis.

2.3.3. Molecular Dynamics

MD studies of the protein-ligand complexes were performed using Desmond (version 2021-1, Schrödinger, New York, NY, USA) with the OPLS 2005 forcefield [43]. The most representative docking pose for each ligand was used as a starting point to initiate the MD simulations. The complexes were prepared with the System Builder Utility in a buffered orthorhombic box ($10 \times 10 \times 10 \text{ \AA}$), using the transferable intermolecular potential with a 3-point model for water (TIP3P). The complexes were neutralized and NaCl was added in a 0.15 M concentration. Complexes were minimized in three stages. In the first stage, water-heavy atoms were restrained with a force constant of $1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (during 100 ps); for the second stage, backbones were constrained with a $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (during 100 ps); and for the third stage, the systems were minimized with no restraints (during 100 ps). The three minimization stages were generated using the default parameters.

The system was submitted to 300 ns of production runs, under NPT ensemble at 1 bar using the Martyna-Tuckerman-Klein (MTK) barostat and 300 K using the Nose-Hoover thermostat. Electrostatic forces were calculated with the smooth PME method using a 9 \AA cut-off, while constraints were enforced with the M-SHAKE algorithm. Integration was done every 2 fs, with a recording interval of 50 ps. All protein-ligand complexes were submitted to the “Relax model system” using the default parameters. The quality of the simulation and trajectory analyses were carried out with the tools implemented in the Maestro-GUI (Schrödinger, New York, NY, USA). SAM was used as a procedure control, 150 ns of production were generated using the same protocol described in this section.

3. Results and Discussion

3.1. Biochemical Inhibition Assays

First, we tested the three compounds **12**, **13**, and **14** (Figure 2) with DNMT1 at a single concentration (100 \mu M). The results are summarized in Figure 2 and fully detailed in Tables S1 a2 in the Supplementary Materials. Two compounds **12** and **13** showed strong inhibition of DNMT1 ($>99\%$), and **14** showed inhibition of 7.4%. The very low enzymatic inhibitory activity of **14** clearly indicated the need for substitution at position 2 of the quinazoline ring (Figure 2).

Based on the results at a single concentration, we decided to test the enzymatic inhibitory activity of DNMT1, DNMT3A, and DNMT3B of the two most active compounds **12** and **13** in concentration-response assays. Results, summarized in Figure 2 and detailed in Table S1 in the Supplementary Materials, indicated that **12** is a potent nanomolar and selective inhibitor of DNMT1 ($IC_{50} = 30 \text{ nM}$), with higher inhibition of DNMT3A ($IC_{50} = 4870 \text{ nM}$) and no enzymatic inhibition of DNMT3B ($IC_{50} > 100,000 \text{ nM}$). Notably,

under the assay conditions used in this work, **12** was about ten times more potent against DNMT1 than the positive control SAH ($IC_{50} = 340$ nM). The structural analog **13** had a similar inhibitory activity profile with similar inhibitory potency but higher selectivity towards DNMT1 ($IC_{50} = 81$ nM) over DNMT3A ($IC_{50} = 14,690$ nM), and DNMT3B ($IC_{50} > 100,000$ nM) (Figure 2 and Table S1).

The nanomolar enzymatic inhibitory potency of DNMT1 of both 7-aminoalkoxyquinazolines **12** and **13** is about ten times that of the positive control SAH. Despite the high variability across DNMT inhibitory assays and the challenge to reproduce the IC_{50} values accurately across different laboratories, there are few compounds reported in the literature with low nanomolar inhibition of DNMT1 [44]. For instance, in one of the most recent studies testing different DNMTs under the same assay conditions, **5** (GSK3685032) strongly inhibited DNMT1 ($IC_{50} = 30$ nM) and did not inhibit DNMT3A and 3B (Figure 1A). In that work, Pappalardi et al., also tested the well-known pan inhibitor **3** (SGI-1027), showing IC_{50} values of 1030 nM (DNMT1); 13,000 nM (DNMT3A), and 7000 nM (DNMT3B) [22]. In the work of Pappalardi et al., SAH (also used as a reference) had IC_{50} values of 540 nM, 100 nM, and 90 nM for DNMT1, DNMT3A, and DNMT3B, respectively. Such values for the positive control generally agree with the IC_{50} values measured under the assay conditions used in this work (340 nM, 100 nM, and 30 nM for DNMT1, DNMT3A, and DNMT3B, respectively, Table S2 in the Supplementary Materials).

In addition to the promising DNMT enzymatic inhibition profile of compounds **12** and **13**, it is remarkable the reported high inhibition of both molecules of the epigenetic target G9a ($IC_{50} = 6$, and 4 nM, respectively, Figure 2) [35]. As discussed in the Introduction, based on the structural relationship between the quinazolines tested in this work with the 4-aminoquinolines reported as dual inhibitors of G9a and DNMT1 [27], we hypothesized that quinazolines such as **12** and **13**, could inhibit DNMT1. Biochemical inhibition assays reported in this work confirmed the hypothesis. Further testing with DNMT3A and DNMT3B revealed that both molecules are potent inhibitors of DNMT1 and selective versus DNMT3A and DNMT3B. Thus far, other than **5** (GSK3685032), the two quinazolines reported in this work are the few small-molecule selective inhibitors of DNMT1 over DNMT3A and DNMT3B.

3.2. Molecular Docking and Dynamics Simulations with DNMTs

Computational methods have been used to identify novel inhibitors, optimize their activity, and/or to further understand their activity at the molecular level of compound targeting epigenetic targets, including DNMTs. These methods are collectively referred to as “epi-informatics” [30]. As previously discussed, computational methods are not used necessary to predict or identify novel inhibitors but the computational studies provide key insights to study the mechanism of action and rationalize the activity of active molecules at the molecular level. For example, we recently conducted a molecular and activity landscape modeling study to rationalize the reported enzymatic inhibitory activity of 251 G9a inhibitors [34], and 50 4-aminoquinolines as dual inhibitors of G9a and DNMT1 [45]. Results of that work yielded the establishment of a robust structural hypothesis of protein-ligand interactions associated with the dual activity or selectivity with the epigenetic targets.

In this work, we employed molecular docking and dynamics simulations to provide insights into the DNMTs enzymatic inhibitory activity of **12** and **13** at the structural level. For this purpose, we took advantage of the availability of the three-dimensional structural information of DNMT1, DNMT3A, and DNMT3B.

Figure 3 summarizes the interactions between **12** and **13** and DNMT1, according to the MD simulations. The conserved interactions with SER 582, ASP 764, and SER 1292 in both compounds. However, the generation of stable interactions with ASP 583, ASP 588, CYS 1288, and GLN 1289 (Figure 3(A1,A2)). Interestingly, **13** has been distinguished by other key interactions (GLU 766, VAL 1330, ASN 1332, and PHE 1336). Despite the differences in key interactions, the compounds studied tend to establish the conformation (reducing the RMSD values) of CXXC (647–691 aa) and autoinhibition (699–733 aa) domains

of DNMT1 (Figure 3(A3,B3)), in relationship with SAM (Figure S1C in the Supplementary Materials). We emphasize that these domains are present only in DNMT1 and are not its 3A or 3B isoforms [46], which could explain the selective enzymatic inactivation of DNMT1 [47]. Additionally, we analyze the specific conformational changes associated with the interaction of **12** and **13** against DNMT1 (Figure S2 in the Supplementary Materials).

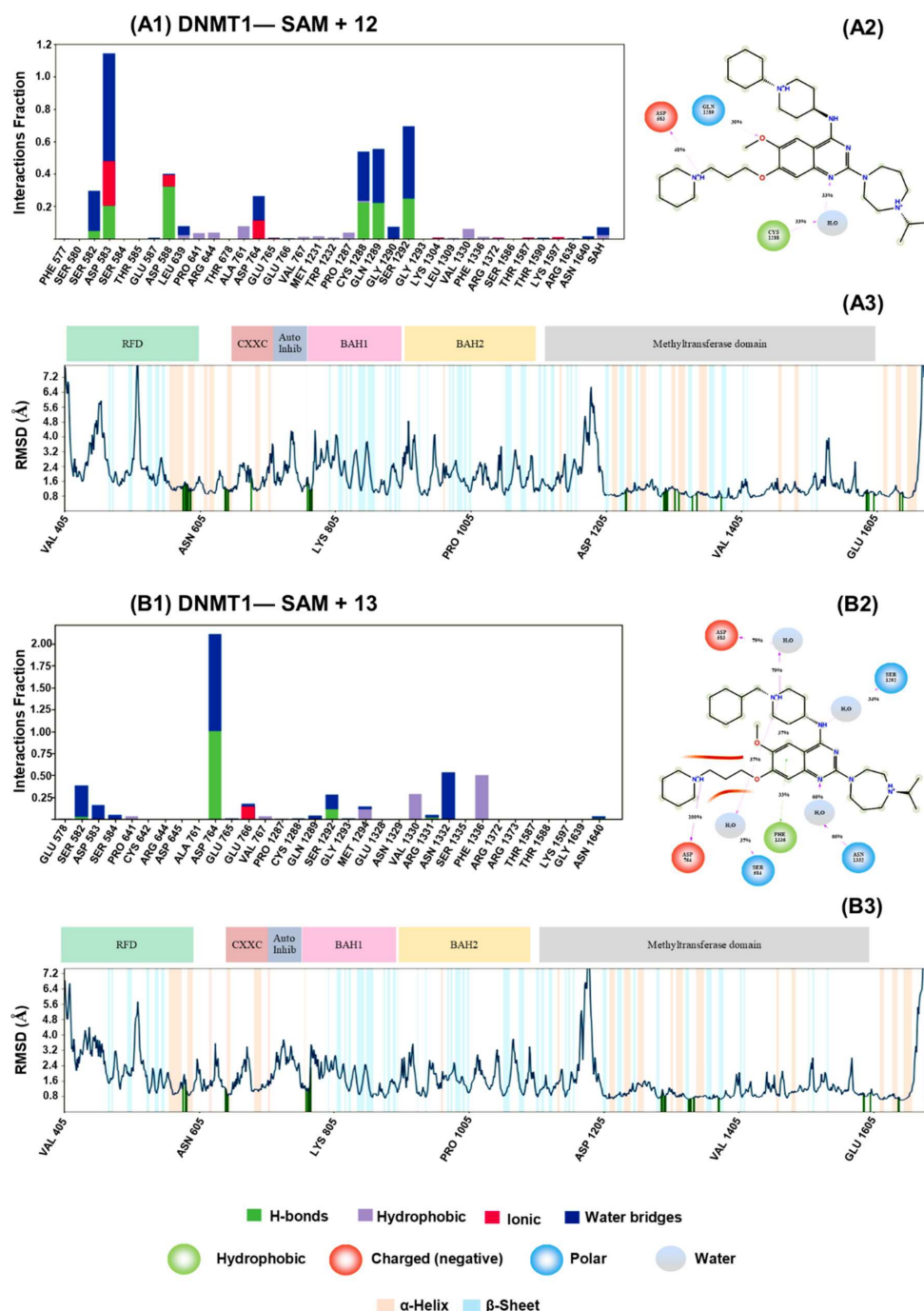


Figure 3. Molecular dynamic results of compounds **12** and **13** against DNMT1. 300 ns were produced per each compound. The (A1,B1) panels show the interactions between ligands and the DNMT1 structure in the presence of *S*-adenosyl-*L*-methionine (SAM). Additionally, panels (A2,B2) show the representative key interaction during the last 30 ns of molecular dynamic productions. Finally, panels (A3,B3) illustrate the conformational changes (RMSD values) on different key domains on DNMT1.

Recently, we described at the structural level the conformational changes generated on G9a in their SET domain by the interaction with quinoline-based compounds [34]. This work describes the conformational changes associated with the interaction of quinoline-based compounds against the CXXC and autoinhibition domain on DNMT1.

Several groups are working identifying novel potent DNMTi, including dual inhibitors of epigenetic writers such as DNMT and G9a (vide supra). For example, we recently reported DNMT1 inhibitors with novel chemical scaffolds, including two approved for clinical use. However, those compounds lack enough potency. For instance, **4** showed and $IC_{50} = 55.85 \mu\text{M}$ using the same assay conditions used in this work [37]. The knowledge of the promising enzymatic and cell-based activity profile of quinoline-based compounds such as **8** (Figure 1B) as dual inhibitors of DNMT1 and G9a; the high structural similarity of the quinoline-based derivatives to the quinazoline-based derivatives with strong inhibition profile of G9a; plus the commercial availability of quinazoline-based derivatives in epigenetic focused libraries led us to identify two 7-aminoalkoxy-quinazolines with 30 and 81 nM potency toward DNMT1 (Figure 2). Molecular modeling studies suggest that the selective inhibition of DNMT1 was carried out by the induced conformational change on their CXXC and autoinhibition domains, which is essential for enzymatic activity [47].

Notably, it is reported in the literature that **12** and **13** have high in vitro potency versus G9a ($IC_{50} = 6$ and 4 nM , respectively) and are also highly potent in reducing H3K9me2 levels in human breast adenocarcinoma (MDA-MB-231) cells (26 and 25 nM , respectively) and with low cell toxicity (EC_{50} of 3.3 and $2.8 \mu\text{M}$, respectively) [35]. Potting together, the two 7-aminoalkoxy-quinazolines **12** and **13** are promising compounds to continue developing as polypharmacological, specifically, dual-epigenetic target inhibitors as candidates compounds with potential therapeutic applications.

In silico target profiling of the three compounds tested in this work (Figure 2), with the recently developed online web server Epigenetic Target Profiler [48], suggest that all compounds could be active with additional epigenetic targets such as HDACs (Figure S3 in the Supplementary Materials). It would remain to continue exploring computationally (e.g., using structure-based methods) and then experimentally the predicted activity of the molecules.

4. Conclusions

We identified two 7-aminoalkoxy-quinazolines: **12** and **13**, with nanomolar inhibition of DNMT1 (30 nM y 80 nM , respectively) in enzymatic inhibition assays. Notably, both compounds showed better inhibitory activity than the positive control, SAH ($340 \mu\text{M}$), less inhibition of DNMT3A (4.87 and $14.69 \mu\text{M}$) and none of them inhibited DNMT3B. Also, these have been reported as potent inhibitors of another epigenetic target writer, G9a, in enzymatic and human breast adenocarcinoma (MDA-MB-231) cell-based assays with low cell toxicity. It remains to demonstrate that **12** and **13** are also able to reduce methylation levels in cell-based assays. However, their high structural similarity to the quinoline-based derivatives such as **CM-272** (Figure 1B), which are known to effectively reduce DNA methylation in cell-based assays, strongly support the hypothesis that the quinazolines reported in this work will also be inhibitors in cell based-assays. Such a hypothesis will be tested and reported in due course. The two active compounds identified in this work are structurally similar with conserving key interactions against SER 582, ASP 764, and SER1292 on DNMT1 in presence of their cofactor SAM. The molecular dynamics results suggest that the DNMT1 inhibition of **12** and **13** is caused by the conformational changes on the CXXC and autoinhibition domains. From the mechanistic point of view, one of the main perspectives of this study is to test if the quinazoline-based derivatives; for instance, **12** is DNA—substrate competitive. This will be done through enzymatic DNA-substrate competitive assays. The results will be reported in due course. In all, the results of this work are a contribution towards the further investigation and development of DNMTi as part of multiple epigenetic target therapies.

Supplementary Materials: The following Supplementary Materials can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27092892/s1>, Table S1: Number of quinazolines and 7-aminoalkoxy-quinazolines present in various epigenetic focused libraries; Table S2: Results of the relative enzymatic activity of DNMT1 as percentages^a; Table S3: Results of concentration-response assays for selected quinazolines (IC₅₀) with DNMT1, DNMT3A and DNMT3B^a; Figure S1: Analysis of 150 ns production of the molecular dynamics of production S-Adenil-Metione (SAM) against DNMT1; Figure S2: Final conformational changes on DNMT1 from molecular dynamics calculations; Figure S3: In silico profiling of the three compounds with the free online and validate server Epigenetic Target Profiler.

Author Contributions: Conceptualization, J.L.M.-F.; methodology, E.L.-L. and L.P.M.-F.; software, E.L.-L.; investigation, J.L.M.-F., E.L.-L. and L.P.M.-F.; resources, J.L.M.-F.; writing—original draft preparation, J.L.M.-F.; writing—review and editing, All; project administration, J.L.M.-F.; funding acquisition, J.L.M.-F. All authors have read and agreed to the published version of the manuscript.

Funding: We thank the support of DGAPA, UNAM, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Grant No. IN201321.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Edgar López-López, Liliam P. Martínez-Fernández are grateful to Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico, for the Ph. D. scholarship numbers, 762342 (No. CVU: 894234) and 1012509, respectively. Stimulating and rich discussions with E. Alexis Flores-Padilla, K. Eurídice Juárez-Mercado, Norberto Sánchez-Cruz, and Diana L. Prado-Romero are acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Holdgate, G.A.; Bardelle, C.; Lanne, A.; Read, J.; O'Donovan, D.; Smith, J.M.; Selmi, N.; Sheppard, R. Drug discovery for epigenetics targets. *Drug Discov. Today* **2021**, *27*, 1088–1098. [[CrossRef](#)] [[PubMed](#)]
2. Singh, M.; Kumar, V.; Sehrawat, N.; Yadav, M.; Chaudhary, M.; Upadhyay, S.K.; Kumar, S.; Sharma, V.; Kumar, S.; Dilbaghi, N.; et al. Current paradigms in epigenetic anticancer therapeutics and future challenges. *Semin. Cancer Biol.* **2021**, *13*, 3364. [[CrossRef](#)] [[PubMed](#)]
3. Ganesan, A.; Arimondo, P.B.; Rots, M.G.; Jeronimo, C.; Berdasco, M. The timeline of epigenetic drug discovery: From reality to dreams. *Clin. Epigenetics* **2019**, *11*, 174. [[CrossRef](#)] [[PubMed](#)]
4. Yu, J.; Xie, T.; Wang, Z.; Wang, X.; Zeng, S.; Kang, Y.; Hou, T. DNA methyltransferases: Emerging targets for the discovery of inhibitors as potent anticancer drugs. *Drug Discov. Today* **2019**, *24*, 2323–2331. [[CrossRef](#)]
5. Conery, A.R.; Rocnik, J.L.; Trojer, P. Small molecule targeting of chromatin writers in cancer. *Nat. Chem. Biol.* **2022**, *18*, 124–133. [[CrossRef](#)]
6. Vougiouklakis, T.; Bernard, B.J.; Nigam, N.; Burkitt, K.; Nakamura, Y.; Saloura, V. Clinicopathologic significance of protein lysine methyltransferases in cancer. *Clin. Epigenetics* **2020**, *12*, 146. [[CrossRef](#)]
7. Nepali, K.; Liou, J.P. Recent developments in epigenetic cancer therapeutics: Clinical advancement and emerging trends. *J. Biomed. Sci.* **2021**, *28*, 27. [[CrossRef](#)]
8. Zhang, J.; Yang, C.; Wu, C.; Cui, W.; Wang, L. DNA methyltransferases in cancer: Biology, paradox, aberrations, and targeted therapy. *Cancers* **2020**, *12*, 2123. [[CrossRef](#)]
9. Weng, Y.-L.; An, R.; Shin, J.; Song, H.; Ming, G.-l. DNA modifications and neurological disorders. *Neurotherapeutics* **2013**, *10*, 556–567. [[CrossRef](#)]
10. Bayraktar, G.; Kreutz, M.R. Neuronal DNA methyltransferases: Epigenetic mediators between synaptic activity and gene expression? *Neurosci. A Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **2018**, *24*, 171–185. [[CrossRef](#)]
11. Mazzone, R.; Zwergel, C.; Artico, M.; Taurone, S.; Ralli, M.; Greco, A.; Mai, A. The emerging role of epigenetics in human autoimmune disorders. *Clin. Epigenetics* **2019**, *11*, 34. [[CrossRef](#)] [[PubMed](#)]
12. Zouali, M. DNA methylation signatures of autoimmune diseases in human b lymphocytes. *Clin. Immunol.* **2021**, *222*, 108622. [[CrossRef](#)] [[PubMed](#)]
13. Arguelles, A.O.; Meruvu, S.; Bowman, J.D.; Choudhury, M. Are epigenetic drugs for diabetes and obesity at our door step? *Drug Discov. Today* **2016**, *21*, 499–509. [[CrossRef](#)] [[PubMed](#)]

14. Hishikawa, A.; Hayashi, K.; Yoshimoto, N.; Nakamichi, R.; Homma, K.; Itoh, H. DNA damage and expression of DNA methylation modulators in urine-derived cells of patients with hypertension and diabetes. *Sci. Rep.* **2020**, *10*, 3377. [[CrossRef](#)]
15. Jin, Z.; Liu, Y. DNA methylation in human diseases. *Genes Dis.* **2018**, *5*, 1–8. [[CrossRef](#)]
16. Hauser, B.M.; Lau, A.; Gupta, S.; Bi, W.L.; Dunn, I.F. The epigenomics of pituitary adenoma. *Front. Endocrinol.* **2019**, *10*, 290. [[CrossRef](#)]
17. Zhang, W.; Xu, J. DNA methyltransferases and their roles in tumorigenesis. *Biomark Res.* **2017**, *5*, 1. [[CrossRef](#)]
18. Liu, C.C.; Ou, T.T.; Wu, C.C.; Li, R.N.; Lin, Y.C.; Lin, C.H.; Tsai, W.C.; Liu, H.W.; Yen, J.H. Global DNA methylation, dnmt1, and mbd2 in patients with systemic lupus erythematosus. *Lupus* **2011**, *20*, 131–136. [[CrossRef](#)]
19. Klein, C.J.; Botuyan, M.V.; Wu, Y.; Ward, C.J.; Nicholson, G.A.; Hammans, S.; Hojo, K.; Yamanishi, H.; Karpf, A.R.; Wallace, D.C.; et al. Mutations in dnmt1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* **2011**, *43*, 595–600. [[CrossRef](#)]
20. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños María, P.; Mosquera Juan, F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucl. Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]
21. Kuck, D.; Caulfield, T.; Lyko, F.; Medina-Franco, J.L. Nanaomycin a selectively inhibits dnmt3b and reactivates silenced tumor suppressor genes in human cancer cells. *Mol. Cancer Ther.* **2010**, *9*, 3015–3023. [[CrossRef](#)] [[PubMed](#)]
22. Pappalardi, M.B.; Keenan, K.; Cockerill, M.; Kellner, W.A.; Stowell, A.; Sherk, C.; Wong, K.; Pathuri, S.; Briand, J.; Steidel, M.; et al. Discovery of a first-in-class reversible dnmt1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. *Nat. Cancer* **2021**, *2*, 1002–1017. [[CrossRef](#)] [[PubMed](#)]
23. Gilmartin, A.G.; Groy, A.; Gore, E.R.; Atkins, C.; Long, E.R.; Montoute, M.N.; Wu, Z.; Halsey, W.; McNulty, D.E.; Ennulat, D.; et al. In vitro and in vivo induction of fetal hemoglobin with a reversible and selective dnmt1 inhibitor. *Haematologica* **2021**, *106*, 1979–1987. [[CrossRef](#)] [[PubMed](#)]
24. Naveja, J.J.; Medina-Franco, J.L. Insights from pharmacological similarity of epigenetic targets in epipolypharmacology. *Drug Discov. Today* **2018**, *23*, 141–150. [[CrossRef](#)] [[PubMed](#)]
25. de Lera, A.R.; Ganesan, A. Two-hit wonders: The expanding universe of multitargeting epigenetic agents. *Curr. Opin. Chem. Biol.* **2020**, *57*, 135–154. [[CrossRef](#)]
26. Tomaselli, D.; Lucidi, A.; Rotili, D.; Mai, A. Epigenetic polypharmacology: A new frontier for epi-drug discovery. *Med. Res. Rev.* **2020**, *40*, 190–244. [[CrossRef](#)]
27. Rabal, O.; San José-Enériz, E.; Agirre, X.; Sánchez-Arias, J.A.; Vilas-Zornoza, A.; Ugarte, A.; de Miguel, I.; Miranda, E.; Garate, L.; Fraga, M.; et al. Discovery of reversible DNA methyltransferase and lysine methyltransferase g9a inhibitors with antitumoral in vivo efficacy. *J. Med. Chem.* **2018**, *61*, 6518–6545. [[CrossRef](#)]
28. Medina-Franco, J.L.; Méndez-Lucio, O.; Yoo, J.; Dueñas, A. Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discov. Today* **2015**, *20*, 569–577. [[CrossRef](#)]
29. 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. Inhibitors of DNA methyltransferases from natural sources: A computational perspective. *Front. Pharmacol.* **2018**, *9*, 1144. [[CrossRef](#)]
30. Sessions, Z.; Sánchez-Cruz, N.; Prieto-Martínez, F.D.; Alves, V.M.; Santos, H.P., Jr.; Muratov, E.; Tropsha, A.; Medina-Franco, J.L. Recent progress on cheminformatics approaches to epigenetic drug discovery. *Drug Discov. Today* **2020**, *25*, 2268–2276. [[CrossRef](#)]
31. Prado-Romero, D.L.; Medina-Franco, J.L. Advances in the exploration of the epigenetic relevant chemical space. *ACS Omega* **2021**, *6*, 22478–22486. [[CrossRef](#)] [[PubMed](#)]
32. Flores-Padilla, E.A.; Juárez-Mercado, K.E.; Naveja, J.J.; Kim, T.D.; Alain Miranda-Quintana, R.; Medina-Franco, J.L. Chemoinformatic characterization of synthetic screening libraries focused on epigenetic targets. *Mol. Inf.* **2021**, *40*, e2100285. [[CrossRef](#)] [[PubMed](#)]
33. Gros, C.; Fleury, L.; Nahoum, V.; Faux, C.; Valente, S.; Labella, D.; Cantagrel, F.; Rilova, E.; Bouhlel, M.A.; David-Cordonnier, M.-H.; et al. New insights on the mechanism of quinoline-based DNA methyltransferase inhibitors. *J. Biol. Chem.* **2015**, *290*, 6293–6302. [[CrossRef](#)] [[PubMed](#)]
34. López-López, E.; Rabal, O.; Oyarzabal, J.; Medina-Franco, J.L. Towards the understanding of the activity of g9a inhibitors: An activity landscape and molecular modeling approach. *J. Comp.-Aided Mol. Des.* **2020**, *34*, 659–669. [[CrossRef](#)] [[PubMed](#)]
35. Liu, F.; Baryte-Lovejoy, D.; Allali-Hassani, A.; He, Y.; Herold, J.M.; Chen, X.; Yates, C.M.; Frye, S.V.; Brown, P.J.; Huang, J.; et al. Optimization of cellular activity of G9a inhibitors 7-aminoalkoxy-quinazolines. *J. Med. Chem.* **2011**, *54*, 6139–6150. [[CrossRef](#)] [[PubMed](#)]
36. Herrera-Vázquez, F.S.; Hernández-Luis, F.; Medina Franco, J.L. Quinazolines as inhibitors of chromatin-associated proteins in histones. *Med. Chem. Res.* **2019**, *28*, 395–416. [[CrossRef](#)]
37. Juárez-Mercado, K.E.; Prieto-Martínez, F.D.; Sánchez-Cruz, N.; Peña-Castillo, A.; Prada-Gracia, D.; Medina-Franco, J.L. Expanding the structural diversity of DNA methyltransferase inhibitors. *Pharmaceuticals* **2020**, *14*, 17. [[CrossRef](#)]
38. Reaction Biology Corporation. Available online: <http://www.reactionbiology.com> (accessed on 5 March 2022).
39. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucl. Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]

40. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in casp8. *Proteins* **2009**, *77* (Suppl. 9), 114–122. [[CrossRef](#)]
41. Cho, A.E.; Guallar, V.; Berne, B.J.; Friesner, R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (qm/mm) approach. *J. Comput Chem* **2005**, *26*, 915–931. [[CrossRef](#)]
42. Molecular Operating Environment (MOE). Chemical Computing Group Inc.: Montreal, QC, Canada, 2021. Available online: <http://www.chemcomp.com> (accessed on 5 March 2022).
43. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on cpu and gpu architectures with namd. *J. Chem. Phys.* **2020**, *153*, 044130. [[CrossRef](#)] [[PubMed](#)]
44. Guianvarc’h, D.; Arimondo, P.B. Challenges in developing novel DNA methyltransferases inhibitors for cancer therapy. *Fut. Med. Chem.* **2014**, *6*, 1237–1240. [[CrossRef](#)] [[PubMed](#)]
45. López-López, E.; Prieto-Martínez, F.D.; Medina-Franco, J.L. Activity landscape and molecular modeling to explore the SAR of dual epigenetic inhibitors: A focus on G9a and DNMT1. *Molecules* **2018**, *23*, 3282. [[CrossRef](#)] [[PubMed](#)]
46. Jeltsch, A.; Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases—A new regulatory paradigm. *Nucleic Acids Res.* **2016**, *44*, 8556–8575. [[CrossRef](#)] [[PubMed](#)]
47. Pradhan, M.; Estève, P.O.; Chin, H.G.; Samaranayake, M.; Kim, G.D.; Pradhan, S. CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry* **2008**, *47*, 10000–10009. [[CrossRef](#)]
48. Sánchez-Cruz, N.; Medina-Franco, J.L. Epigenetic target profiler: A web server to predict epigenetic targets of small molecules. *J. Chem. Inf. Model.* **2021**, *61*, 1550–1554. [[CrossRef](#)] [[PubMed](#)]