



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE QUÍMICA
EPI-INFORMÁTICA EN LA SELECCIÓN DE PRODUCTOS
NATURALES COMO INHIBIDORES POTENCIALES DE
METILTRANSFERASAS DE ADN

TESIS

QUE PARA OBTENER EL TÍTULO DE
QUÍMICA FARMACÉUTICA BIÓLOGA

PRESENTA

JOCELYN SALAZAR VALENCIA

ASESOR: DR. JOSÉ LUIS MEDINA FRANCO

CIUDAD UNIVERSITARIA, CD.MX 2022





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO:

PRESIDENTE:	Profesor: López López Marisol
VOCAL:	Profesor: Plasencia de la Parra Francisco Javier
SECRETARIO:	Profesor: Medina Franco José Luis (Asesor)
1er. SUPLENTE:	Profesor: Rivero Cruz Isabel del Carmen
2° SUPLENTE:	Profesor: Del Valle Pérez Paulina

SITIO DONDE SE DESARROLLÓ EL TEMA:

EL PRESENTE TRABAJO SE DESARROLLÓ EN EL GRUPO DE INVESTIGACIÓN DIFACQUIM (DISEÑO DE FÁRMACOS ASISTIDO POR COMPUTADORA) DEL DEPARTAMENTO DE FARMACIA, UBICADO EN EL CUBÍCULO 117 DEL EDIFICIO F DE LA FACULTAD DE QUÍMICA DE LA UNAM.

ASESOR DEL TEMA:

DR. JOSÉ LUIS MEDINA FRANCO

SUPERVISOR TÉCNICO:

DR. NORBERTO SÁNCHEZ CRUZ

SUSTENTANTE:

JOCELYN SALAZAR VALENCIA

AGRADECIMIENTOS

Al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la Dirección General de Asuntos del Personal Académico de la UNAM, por la beca otorgada en el proyecto PAPIIT IN201321, lo que permitió el desarrollo del presente trabajo.

Al Dr. José Luis Medina Franco, al Dr. Norberto Sánchez Cruz y al grupo de investigación DIFACQUIM de la Facultad de Química, por su apoyo y aportaciones en el desarrollo del proyecto.

A los miembros del jurado el Dr. Francisco Javier Plasencia, Dra. Marisol López López, Dra. Paulina del Valle y Dra. Isabel del Carmen Rivero por sus valiosas aportaciones durante la revisión del trabajo escrito.

ÍNDICE

ABREVIATURAS	5
1. Resumen	6
2. Antecedentes	7
3. Hipótesis	23
4. Objetivos	23
5. Materiales y métodos	24
6. Resultados y discusión	28
7. Conclusiones	52
8. Perspectivas	54
9. Referencias	55
ANEXOS	58

ABREVIATURAS

COCONUT: *Collection of Open Natural Products*.

DIFAC: Diseño de fármacos asistido por computadora.

VS: cribado virtual (del inglés, *virtual screening*).

vHTS: cribado virtual de alto rendimiento (del inglés, *Virtual High- Throughput Screening*).

QSAR: relación cuantitativa estructura- actividad (del inglés *Quantitative Structure-Activity Relationship*).

DNMTs: ADN metiltransferasas (del inglés *DNA methyltransferases*).

ETP: *Epigenetic Target Profiler*.

FP: huella digital (del inglés *fingerprint*).

SVM: máquinas de soporte vectorial (del inglés *Support Vector Machines*).

ECIF: *Extended Connectivity Interaction Features*.

SMILES: *Simplified Molecular Input Line Entry Specification*.

ID: identificador.

PDB: *Protein Data Bank*.

LMC: Leucemia mieloide crónica

LLA: Leucemia linfoblástica aguda

VIH: Virus de Inmunodeficiencia Humana

SAM: S-adenosilmetionina

SAH: S- adenosil L-homocisteína

HDAC: Histona Deacetilasa

AR: Artritis Reumatoide

ARNm: Acido Desoxirribonucléico mensajero.

Kd: Constante de disociación

1. Resumen

En las últimas décadas se ha impulsado la utilización de métodos computacionales para el desarrollo de nuevos fármacos, el modelado molecular y los enfoques quimoinformáticos han contribuido a grandes avances en el campo ya que han ayudado, por ejemplo, a explorar el mecanismo de acción de fármacos a nivel molecular. Dentro de los campos que se han explorado utilizando métodos informáticos se encuentra la epigenética, que se define como el estudio de los cambios en la función de los genes que son hereditarios y que no se pueden atribuir a alteraciones en la secuencia de ADN. Por lo tanto, se ha introducido la subdisciplina epi-informática para resumir los avances en el descubrimiento de fármacos epigenéticos y sondas químicas impulsados por métodos computacionales. Dentro de las alteraciones o aberraciones epigenéticas que alteran la transcripción de la secuencia de ADN, se encuentra la metilación. Cuando existe una hipermetilación o hipometilación del mismo puede producirse el silenciamiento de ciertos genes supresores de enfermedades como el cáncer. La metilación del ADN se lleva a cabo mediante enzimas conocidas como ADN metiltransferasas (DNMT, por sus siglas en inglés), las cuales han sido objeto de estudio debido a que su inhibición podría revertir la hipermetilación de los genes antes mencionados y utilizarse como tratamiento en ciertas enfermedades.

En el presente trabajo se describe el análisis de diferentes bases de datos de productos naturales de acceso libre a partir de las cuales se determinaron los mejores compuestos candidatos a inhibidores de metiltransferasas de ADN, utilizando métodos computacionales basados en similitud y en estructura. Se utilizó la herramienta *Epigenetic Target Profiler*, con la cual se predijeron los posibles blancos epigenéticos de cada uno de los compuestos en las bases de datos. Con este método no se encontraron compuestos que pudieran tener cierta actividad con DNMT1 por lo que se optó por utilizar un método computacional basado en estructura. Para ello se utilizaron programas de acoplamiento molecular, en especial AutoDock VINA que permitió calcular la afinidad de unión del sitio catalítico de DNMT1 con cada uno de los compuestos de estudio. Mediante acoplamiento molecular se identificaron algunos compuestos con potencial actividad para DNMT1 y de los cuales pudieran realizarse posteriormente estudios de inhibición enzimática.

2. Antecedentes

2.1. Diseño de fármacos asistido por computadora

El diseño de fármacos asistido por computadora (DIFAC) tiene como objetivo la predicción, 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.¹ El DIFAC es uno de los enfoques fundamentales para el descubrimiento contemporáneo de fármacos en la fase preclínica, en el cual se utilizan diversas técnicas y programas computacionales. Generalmente se usan en combinación, en un intento por lograr el resultado deseado.²

Desde el inicio y la aplicación de DIFAC en el desarrollo de fármacos, se han descubierto y aprobado más de 70 fármacos y este número aumenta constantemente con el tiempo.² Algunos fármacos desarrollados con métodos computacionales han sido imatinib (un inhibidor de la quinasa utilizado para tratar ciertos tipos de cáncer como leucemia mieloide crónica (LMC) y leucemia linfoblástica aguda (LLA)), dorzolamida (un inhibidor de la carboanidrasa II utilizado para tratar la presión alta dentro del ojo, incluido el glaucoma), enfuvirtida (la primera en su clase fármaco antirretroviral utilizado en terapia combinada para el tratamiento del VIH-1), oseltamivir (un inhibidor de la neuraminidasa utilizado para tratar la gripe), entre otros. Sin embargo, a pesar de los grandes avances que ha tenido DIFAC, aún quedan retos importantes por afrontar.³

El modelado por homología, acoplamiento molecular, cribado virtual (VS, por las siglas en inglés) o cribado virtual de alto rendimiento (vHTS, por las siglas en inglés), relación cuantitativa estructura-actividad (QSAR, por las siglas en inglés) y mapeo farmacóforo tridimensional (3D) son algunas de las principales técnicas utilizadas en DIFAC.²

Para el desarrollo de esta tesis se utilizaron técnicas de VS y acoplamiento molecular para identificar compuestos de productos naturales con potencial actividad sobre dianas epigenéticas.

El acoplamiento molecular automatizado tiene como finalidad buscar la conformación y posición óptima de un ligando dentro de un blanco molecular o la posición y conformación más favorable de dos macromoléculas.¹

El cribado virtual consiste en filtrar series de compuestos, normalmente bases de datos moleculares grandes, para seleccionar un subconjunto de moléculas que se sometan a ensayos biológicos. El filtrado se hace empleando uno o más métodos computacionales tales como acoplamiento molecular o similitud molecular. El cribado virtual debe estar integrado con la validación experimental y normalmente se hace en forma de dos o más ciclos de refinamiento de los resultados.¹

2.2. Epi-informática en el desarrollo de nuevos fármacos

En los últimos años el aumento del número de datos relacionados con epigenética ha impulsado el desarrollo de métodos informáticos con aplicación en este campo. El término epi-informática se introdujo en 2016 para resumir los avances en el descubrimiento de fármacos epigenéticos y sondas químicas impulsados por métodos computacionales.⁴ Desde entonces el modelado molecular y los enfoques quimoinformáticos han contribuido a grandes avances en el campo ya que han ayudado a explorar el mecanismo de acción de fármacos a nivel molecular y guían a la optimización de programas.⁵

2.3. Epigenética y metilación del ADN

La epigenética es el estudio de los cambios en la función de los genes que son hereditarios y que no se pueden atribuir a alteraciones en la secuencia de ADN.⁶ Los cambios o alteraciones epigenéticas incluyen la metilación de ADN, modificación de histonas y alteración de microARN (miRNAs).

En las últimas décadas, la metilación del ADN se ha establecido como un mecanismo dominante en la regulación de la expresión génica. La evidencia más convincente se ha reunido a través de varios estudios que comparan los genes de gemelos monocigóticos. A pesar de tener genomas idénticos, las parejas de gemelos experimentan con frecuencia una susceptibilidad diferente a una serie de enfermedades, incluyendo cánceres, enfermedades autoinmunes y trastornos psicológicos. Estas diferencias se han atribuido a que cada gemelo tiene un entorno diferente durante su vida. A nivel molecular, estas diferencias de susceptibilidad se correlacionan con diferencias en el

silenciamiento de los genes entre gemelos, y más importante, el silenciamiento coincide con la metilación del ADN en regiones promotoras de genes específicos.⁷

El hecho de que algunas modificaciones epigenéticas, incluida la metilación del ADN, parezcan ser reversibles, las convierte en un área atractiva en la que buscar nuevos medicamentos con el potencial de detener o revertir la progresión de la enfermedad.⁷

2.3.1 Metilación del ADN

La metilación de ADN se produce en la posición C-5 de citosina dentro de las secuencias de ADN que contienen el motivo CpG y es catalizada por un grupo de enzimas llamadas ADN metiltransferasas (DNMTs, por sus siglas en inglés). En los mamíferos la familia de DNMTs tiene cuatro miembros: DNMT1, DNMT3A, DNMT3B y DNMT3L de las cuales las tres primeras son catalíticamente activas y utilizan la S-adenosin metionina (SAM) como cofactor para llevar a cabo su función.^{7, 8}

Las isoformas DNMT3A y DNMT3B se conocen como metiltransferasas “de novo” y desempeñan papeles similares pero distintos en el desarrollo. La DNMT3a está asociada a la impronta, durante la gametogénesis, en donde las marcas epigenéticas se establecen mediante la metilación diferencial de los alelos paternos y maternos de los genes esenciales en el crecimiento del embrión. Por otra parte, la DNMT3B tiene un papel clave en el desarrollo somático temprano. Tanto la DNMT3A y DNMT3B requieren la presencia de la DNMT3L, catalíticamente inactiva, como coactivador.⁷

Durante la replicación y división posterior de las células somáticas, la metilación de ADN se replica por acción de DNMT1, que copia las marcas de metilación del ADN de la hebra parental en la hebra hija recién sintetizada durante la fase S del ciclo celular. La DNMT1 se conoce como la metiltransferasa de ADN de “mantenimiento”, y proporciona una forma de transmitir la información epigenética entre generaciones celulares. El porcentaje de error en el mantenimiento de la metilación del ADN se produce con una frecuencia de aproximadamente 5% por sitio CpG por división celular, de manera que después de muchas generaciones celulares, la metilación de loci específicos puede estar en un nivel mucho más bajo que en las células originarias.⁷

La metilación del ADN se asocia normalmente con el silenciamiento de los genes, aunque la función precisa parece depender del contexto y muchos aspectos siguen siendo poco conocidos.⁷

La importancia de la metilación del ADN en el silenciamiento génico se puso de manifiesto en los primeros estudios que demostraron que, en el caso de algunos genes supresores de tumores cuyos promotores están hipermetilados, el silenciamiento se revierte con fármacos que revierten la hipermetilación pero no con fármacos que revierten las marcas de histoplasma asociadas a la represión, como los inhibidores de HDAC. Sin embargo, estudios realizados por otros grupos que examinan genes supresores de tumores alternativos han observado que la reexpresión puede ser inducida por el tratamiento con inhibidores de la HDAC sin afectar a los niveles de metilación del ADN.⁷

El estado de las histonas determina el potencial de expresión génica inmediata, mientras que la metilación del ADN proporciona una memoria a largo plazo para el silenciamiento de los genes que puede ser anulada temporalmente.⁷ Este modelo se apoya en la observación de que la histona metiltransferasa EZH2, que media la represión transcripcional por metilación en H3K27, recluta metiltransferasas de ADN en el contexto de los complejos Polycomb 2 y 3, introduciendo la metilación del ADN en los promotores de los genes diana de EZH2 y provocando el silenciamiento a largo plazo.

2.3.2 Metilación aberrante del ADN en el cáncer y en otras enfermedades.

La metilación aberrante del ADN es un rasgo distintivo de muchos tipos de cáncer, aunque no siempre puede ser causa del silenciamiento de los genes. Se observa comúnmente una hipometilación generalizada, consistente con la activación de genes necesarios para la proliferación celular, la migración, la adhesión, etc.; mientras que la hipermetilación se produce en las islas CpG de las regiones promotoras de supresores tumorales bien caracterizados (Figura 1).

En general, aproximadamente 70% de los promotores de genes contienen islas CpG y éstas no suelen estar metiladas, en cambio, la metilación es más frecuente en las islas CpG dentro de los cuerpos de los genes. La metilación aberrante de las islas CpG como

la hipermetilación o la hipometilación se han asociado con diferentes mecanismos celulares implicados en cáncer. Existe evidencia de que la hipermetilación de las islas CpG en regiones promotoras de genes supresores de tumores induce a su inactivación lo cual se relaciona con el desarrollo de los mismos. Por el contrario, la hipometilación en el ADN promueve vías de invasión y metástasis, una etapa avanzada de cáncer.⁸

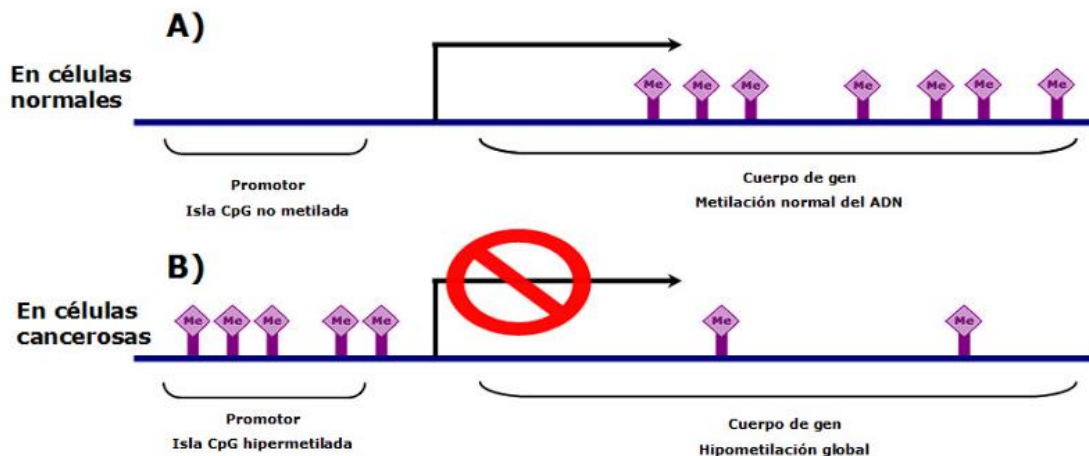


Figura 1. Representación de una isla CpG en un gen. A) En células normales, el promotor se encuentra en un estado no metilado lo que permite la transcripción del gen. B) En células cancerosas, se observa una hipermetilación en el promotor del gen lo que impide la transcripción y conduce a una hipometilación en el resto del genoma.

El ADN hipometilado también puede activar secuencias virales latentes incorporadas al genoma. Por ejemplo, la metilación del ADN reprime la expresión del virus del papiloma humano genital y proteínas del virus de Epstein-Barr, que están asociadas con el cáncer de cuello uterino y el carcinoma nasofaríngeo, respectivamente.⁸

En la tabla 1 se muestran algunos de los genes implicados en diferentes tipos de cáncer, el porcentaje de incidencia y el proceso que se ve afectado debido a la hipermetilación del gen.

Tabla 1. Ejemplos de genes supresores de tumores metilados de manera aberrante en cáncer. Adaptación de la Tabla 4.1 en el Capítulo 4 en Referencia 7.

Gen	Tejido tumoral (incidencia de hipermetilación)	Papel en la transformación
APC	Colon (17–28 %), metástasis hepáticas de colon (42 %), mama, pulmón	Proliferación celular, migración, adhesión, estabilidad cromosómica
BRCA1	Mama (13–15%), cáncer metaplásico de mama (63%), ovario (21%)	Reparación de ADN, punto de control del ciclo celular, estabilidad genómica
CDKN2A (p16)	Colon (30–40 %), cabeza y cuello (27–67 %), mama (15–31 %), pulmón (25–31 %)	Pérdida del punto de control del ciclo celular
CDKN2B (p15)	Leucemia mieloide aguda (71–94 %), leucemia mieloide crónica (24 %)	Pérdida del punto de control del ciclo celular
E-cadherina	Tiroides (83%), mama, próstata	Proliferación celular, invasión, metástasis
ER	Mama (25 % en receptores de estrógeno negativos), próstata (80–95 %), pulmón, leucemia	Resistencia hormonal
MLH1	Endometrial (91%), gástrico (95-100%), colon (9-15%), tumores secundarios de ovario (56%)	Reparación de desajustes de ADN, mutaciones genéticas, inestabilidad de microsatélites
RASSF1A	Pulmón (34–70 %), mama (9 %), ovario (10–50 %), riñón (23 %)	Proliferación celular
Rb	Retinoblastoma (10–13 %), tumores del sistema nervioso (19 %)	Proliferación celular
VHL	Renal (10–19%)	Estabilidad de las proteínas oncogénicas

También se han observado perfiles de metilación aberrante en el ADN en una amplia variedad de enfermedades más allá del cáncer, en particular en enfermedades inmunológicas/ inflamatorias, psiquiátricas, neurológicas y metabólicas. En muchas de estas enfermedades aún no está claro si los cambios en la metilación del ADN son causales o simplemente una respuesta adaptativa al estado de la enfermedad.⁷

El papel de la metilación del ADN es especialmente evidente en las enfermedades autoinmunes e inflamatorias, como la artritis reumatoide (AR), el asma y el lupus. Los mononúcleos similares a los fibroblastos obtenidos en pacientes con AR y osteoartritis presentan una disminución significativa en los niveles de ARNm de DNMT1, 3A y 3B en respuesta a niveles bajos de IL-1. La disminución de los niveles de DNMTs da lugar a una desmetilación del locus CHI3L1 que está hipometilado en los fibroblastos de AR.

Las células T de los pacientes con lupus eritematoso sistémico muestran hipometilación global (reducción del 15-20%) y una disminución de la expresión de DNMT1, que se cree impulsa la producción de anticuerpos autoreactivos.⁷

2.3.3 Estructura y mecanismo de DNMT1

Actualmente se ha resuelto la estructura de la DNMT1 humana la cual contiene, además de un dominio de metiltransferasa C-terminal, una gran región reguladora N-terminal que se compone de un dominio RFTS (secuencia dirigida a focos de replicación), un dominio de dedo de zinc CXXC y un par de dominios BAH (bromo homología adyacente). Los dominios reguladores de DNMT1 median una red de interacciones proteína-proteína y proteína-ADN para controlar el reclutamiento y la actividad enzimática de DNMT1.⁹

En la Figura 2, se observa la estructura tridimensional de la DNMT1, la cual puede encontrarse como 4WXX en *Protein Data Bank* (PDB).



Figura 2. Estructura cristalográfica de DNMT1 humana, PDB 4WXX. Resolución: 2.62 Å.

El mecanismo de metilación de la citosina en la posición C5 del ADN se representa en la Figura 3. La DNMT forma un complejo con el ADN y la citosina que será metilada se voltea fuera del ADN. El tiol de la cisteína catalítica actúa como un nucleófilo que ataca en la posición 6 de la citosina objetivo para generar un enlace covalente intermedio. La posición 5 de la citosina se activa y lleva a cabo un ataque nucleofílico sobre el cofactor S-adenosil-L-metionina (SAM) para formar el 5-metil aducto covalente y S-adenosil-L-homocisteína (SAH). El ataque a la posición 6 es asistido por una protonación transitoria del anillo de citosina en el nitrógeno endocíclico (posición 3), que está estabilizado por un residuo de glutamato. Los residuos de arginina pueden participar en la estabilización del intermediario que hace una interacción de enlace de hidrógeno con el oxígeno carbonílico de la citosina. El complejo covalente entre la base metilada y el ADN se resuelve por desprotonación en la posición 5 para generar la citosina metilada y la enzima libre.¹⁰

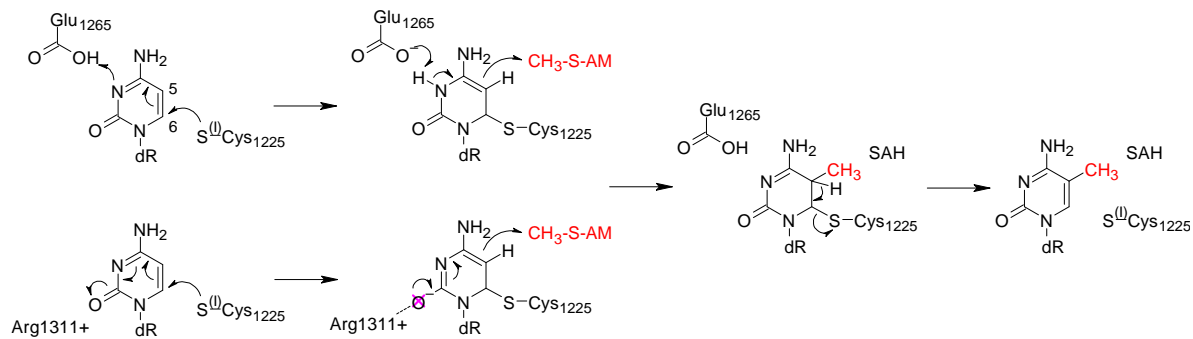


Figura 3. Representación del mecanismo de acción de DNMTs.

2.4 Inhibidores de ADN metiltransferasas

2.4.1 Análogos de nucleósidos

A principios de los 60's se descubrieron dos inhibidores de DNMTs, 5-azacitidina (azacitidina, AZA, Vidaza) y 5-aza-2'-deoxicitidina (decitabina, DAC, Dacogen), los cuales son agentes muy utilizados en cáncer y entran en la clasificación de inhibidores conocidos como análogos de nucleósidos.⁸

El mecanismo de acción de estos inhibidores es el bloqueo irreversible de DNMTs para ello luego de la captación celular tanto la decitabina como la azacitidina se convierten en desoxiacitidina trifosfato a través de distintas vías metabólicas.⁷

La decitabina es fosforilada sucesivamente por la deoxicitidina quinasa (DCK, por sus siglas en inglés) para generar el monofosfato y luego por las quinasas de monofosfato y difosfato de nucleótidos para generar los di y trifosfatos (Figura 4a).⁷

La azacitidina es monofosforilada por la uridina-citidina-quinasa y luego convertida en difosfato por la enzima nucleótido monofosfato quinasa. En este punto, la mayor parte del difosfato se convierte en trifosfato para incorporarse al ARN y una fracción mínima se convierte en difosfato de decitabinma por la ribonucleotido reductasa y sigue la ruta antes mencionada para decitabina. (Figura 4-a).

La formación de desoxiacitidina trifosfato a partir de cualquiera de los dos fármacos conduce a su incorporación al ADN, donde se convierte en un sustrato suicida para las DNMTs. La adición de la cisteína catalítica a la posición 6 de la desoxiacitidina ocurre según el mecanismo para la desoxicitidina natural seguida de la transferencia del grupo metilo del cofactor SAM al nitrógeno en la posición 5 del núcleo base de azacitidina (Figura 4b).⁷

El intermediario resultante es estable: forma un complejo covalente irreversible con DNMT1 la cual permanecerá atrapada en la cadena de ADN. Esto conduce a la degradación de la DNMT1 mediada por el proteosoma.⁷

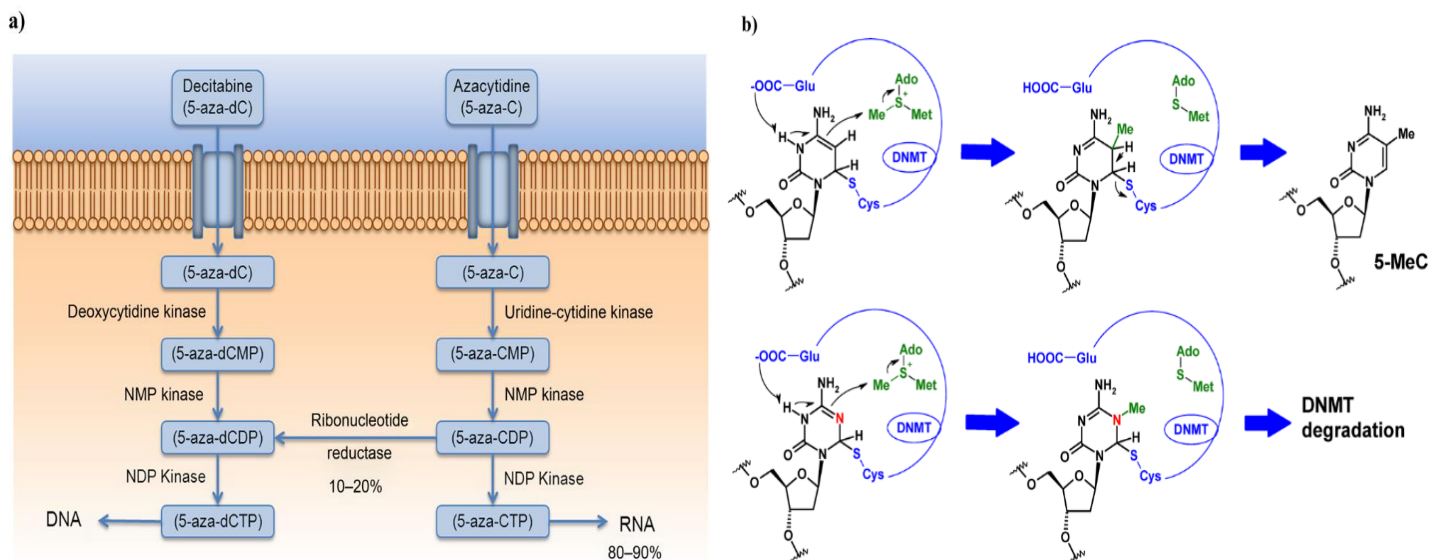


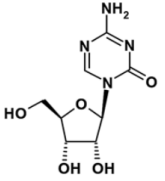
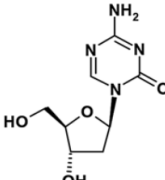
Figura 4. Representación del mecanismo de acción de los inhibidores de DNMT1 análogos de nucleósidos azacitidina y decitabina. a) Formación de desoxiacitidina trifosfato a partir de azacitidina y decitabina y enzimas implicadas. b) Mecanismo mediante el cual se forma el complejo covalente de desoxiacitidina con DNMT1. Tomado de la Figura 4.4 en el Capítulo 4 de (7).

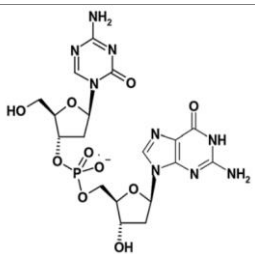
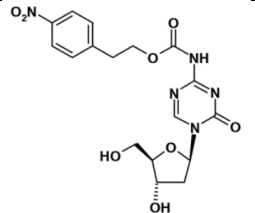
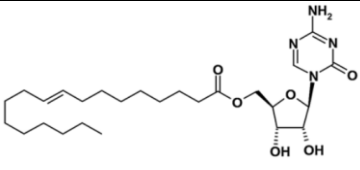
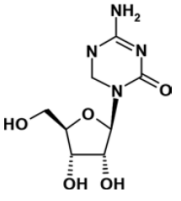
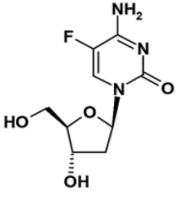
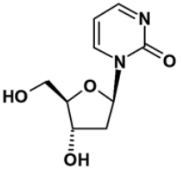
Los inhibidores de DNMTs en dosis altas facilitan la formación de aductos voluminosos, lo que lleva al estancamiento de la horquilla de replicación y la inhibición de la replicación del ADN lo que causa muerte celular. Cuando se usa una dosis baja de inhibidor, éste puede incorporarse al ADN y unirse a DNMTs lo que llevaría a la degradación de las enzimas. Sin DNMT para mantener la metilación del ADN, los sitios CpG pierden su metilación después de la replicación celular y la transcripción de genes previamente silenciados por metilación se restaura.⁸

La Decitabina puede disminuir la expresión de DNMT1 y DNMT3A, mientras que la azacitidina solo se dirige a DNMT1. Sin embargo, la azitadina puede incorporarse al ADN y ARN mientras que la decitabina solo puede incorporarse al ADN. Estos fármacos están aprobados para su uso clínico por la FDA en el tratamiento de mielodisplasia.⁸

La Tabla 2 resume algunos inhibidores análogos de nucleósidos que se han desarrollado, su estructura y fase de estudios clínicos en la que se encuentran. Sin embargo, la mayoría de los análogos de nucleósidos tienen problemas de toxicidad, inestabilidad química en condiciones fisiológicas y baja biodisponibilidad. Esto ha promovido la búsqueda de otras moléculas que actúen como inhibidores de DNMTs pero que no se incorporen a DNA, es decir, compuestos no nucleósidos.⁸

Tabla 2. Ejemplos de inhibidores análogos de nucleósidos, estructura y estatus en estudios clínicos. Adaptación de la Tabla 4.3 en el Capítulo 4 de (7).

Inhibidor	Estructura	Fase de estudio clínico
Azacitidina		Aprobado (SMD, LMC)
Decitabina		Aprobado (SMD, LMC)

SGI-110		Fase II
NPEOC-DAC		Preclínica
CP-4200		Preclínica
5,6-di-hidroazacitidina		Preclínica
5-fluoro-2'-deoxi-citidina		Preclínica
Zebularina		Preclínica

2.4.2 Análogos no nucleósidos

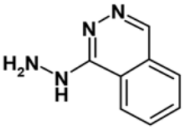
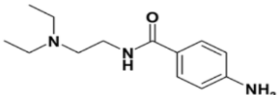
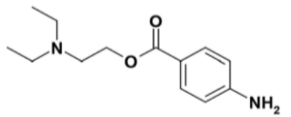
Las estructuras de los análogos o inhibidores no nucleósidos son variadas al igual que sus mecanismos de acción. Algunos de estos agentes muestran afinidad por las regiones de ADN ricas en CpG, bloqueando la actividad de DNMTs y re-activando algunos genes supresores de tumores. Un ejemplo es la procainamida la cual se ha reposicionado después de observar sus efectos desmetilantes. Otros compuestos actúan uniéndose directamente al sitio catalítico de la enzima.⁸

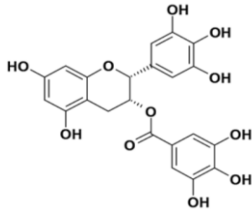
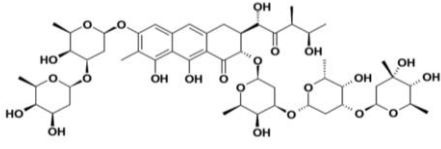
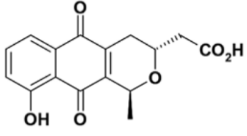
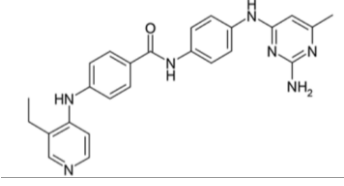
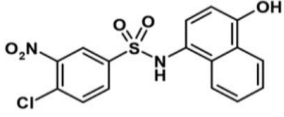
SGI-1027 es un compuesto a base de quinolina que se sintetizó y describió contra DNMT1, DNMT3A y DNMT3B. Actúa inhibiendo la actividad de DNMTs e induce la degradación de DNMT1, reactivando genes supresores de tumores. SGI-1027 también puede perjudicar la propagación de células de cáncer de cuello uterino y de células de carcinoma hepatocelular, aumentando drásticamente la muerte celular apoptótica y la detención del ciclo celular. Se han descrito dos análogos de este compuesto con una mejor selectividad y potencia: MC3343 y MC3353, este último demostró tener actividad antiproliferativa en varios tipos de líneas celulares.⁸ El N-ftaloil-L-triptófano (RG108), un inhibidor de DNMT1, se dirige a la unión del cofactor SAM de DNMT1.

Dentro de esta categoría se encuentran también los productos naturales o derivados, el ejemplo principal de este grupo es la epigalocatequina-3-galato (EGCG) el principal compuesto del té verde, la cual se ha propuesto para inhibir DNMT1 bloqueando el sitio catalítico de la enzima y reactivando los genes silenciados por metilación en células cancerosas. La psammaplina A, inhibidor de DNMTs y de desacetilasas de histonas fue aislado de la esponja marina *Pseudoceratina purpurea* perteneciente a la familia Psammaplysilla. Otros polifenoles del té como catequina, epicatequina o bioflavonoides al igual que la curcumina han sido descritos como inhibidores de DNMTs.¹¹

La Tabla 3 resume algunos compuestos que se han desarrollado como inhibidores no nucleótidos de DNMTs.

Tabla 3. Ejemplos de inhibidores no nucleótidos de DNMTs, estructura y fase de estudios clínicos en la que se encuentran. Adaptación de la Tabla 4.4 en el Capítulo 4 de (7).

Inhibidor	Estructura	Fase de estudio clínico
Hidralazina		Fase II
Procainamida		Preclínica
Procaina		Preclínica

<p>(-) Epigallocatequina-3-galato</p>		<p>Preclínica</p>
<p>Mitramicina A</p>		<p>Preclínica</p>
<p>Nanaomicina</p>		<p>Preclínica</p>
<p>SGL-1027</p>		<p>Preclínica</p>
<p>SW155246</p>		<p>Preclínica</p>



2.5. Productos naturales en el diseño de fármacos

Los productos naturales son una excelente fuente de estructuras químicas para identificar moduladores de DNMTs que puedan tener aplicación clínica como epifármacos.¹² A la fecha, se han descrito algunos productos naturales como EGCG, cúrcuma o polifenoles del té y otros compuestos de origen sintético como inhibidores de DNMTs. Algunos de ellos se han encontrado con la ayuda de métodos de la epinformática.^{5, 13}

2.6. Epigenetic Target Profiler

Como se mencionó en la sección 2.2 el crecimiento de datos quimiogenómicos dentro del campo de la epigenética ha facilitado la aplicación y desarrollo de métodos computacionales bien establecidos. Un ejemplo de este desarrollo es *Epigenetic Target Profiler* (ETP). ETP es una aplicación web gratuita (Figura 5), desarrollada dentro del grupo de investigación DIFACQUIM de la Facultad de Química de la UNAM. Esta

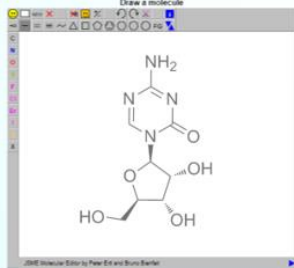
herramienta realiza la predicción de posibles dianas epigenéticas para un compuesto o un grupo de compuestos mediante la utilización de modelos de aprendizaje automático. Para ello, la aplicación realiza la estandarización del compuesto de entrada y la generación de las huellas digitales moleculares o *fingerprints* (FP, por las siglas en inglés) de Morgan y RDKit del mismo. Si el compuesto de entrada se encuentra dentro de las moléculas de entrenamiento de la herramienta, ésta lo reconoce y devuelve la leyenda *Known*. De lo contrario el servidor mide la distancia de Jaccard del compuesto con respecto al conjunto de datos de entrenamiento y realiza la predicción utilizando dos modelos de máquinas de soporte vectorial (SVM, por las siglas en inglés) utilizando las FP previamente calculadas (Morgan::SVM y RDKit::SVM, respectivamente), y se obtienen como resultado las dianas epigenéticas predichas con el servidor mediante el consenso de ambos modelos. ETP puede realizar la predicción de los compuestos para 55 dianas epigenéticas.^{14, 15}

a)  **DIFACQUIM**
COMPUTER-AIDED DRUG DESIGN AT UNAM 

Epigenetic Target Profiler v1.0
Home Help Contact Disclaimer Acknowledgements Publications

This website allows you to estimate the bioactivity profile of a small molecule over a panel of 55 human epigenetic targets. Predictions are based on the consensus prediction of two machine learning models relying on support vector machines (SVM) for each target: one built on Morgan fingerprints (Morgan::SVM) and the other built on RDKit fingerprints (RDKit::SVM).



Draw a molecule



Or paste a SMILES in the box below

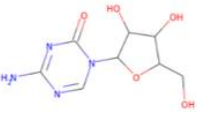
C1=NC(=NC(=O)N(C1)C2=O)C3=CC(=O)N(C=C3)O

Predict Targets

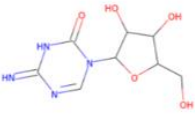
b)  **DIFACQUIM**
COMPUTER-AIDED DRUG DESIGN AT UNAM 

Epigenetic Target Profiler v1.0
Home Help Contact Disclaimer Acknowledgements Publications

Query Molecule



Processed Molecule



Name	CHEMBL ID	Gene	Status	Quartile
DNA (cytosine-5)-methyltransferase 1	CHEMBL1593	DNMT1	Known	
Serine protein kinase ATM	CHEMBL3737	ATM	Predicted	Q4
Bromodomain-containing protein 2	CHEMBL1293289	BRD2	Predicted	Q4
Histone deacetylase 1	CHEMBL325	HDAC1	Predicted	Q4
Lysine-specific demethylase 5A	CHEMBL2424584	KDMA5	Predicted	Q4

[Download CSV](#)

Figura 5. Servidor web de la herramienta *Epigenetic Target Profiler*. a) Para utilizar la herramienta mediante la página de Internet, se dibuja la estructura o se inserta el código SMILES de la misma en la sección correspondiente y se selecciona la opción Predict Targets. b) El servidor mostrará los resultados con las dianas epigenéticas predichas para el compuesto, ID de ChEMBL y cuartil. Además, devuelve la leyenda Know en caso de que el resultado sea conocido.

2.7. Acoplamiento molecular

El acoplamiento molecular es un método computacional basado en estructura utilizado para predecir la interacción entre dos moléculas generando un modelo de unión. En el descubrimiento de fármacos, el acoplamiento generalmente se realiza entre una molécula pequeña y una macromolécula, por ejemplo, acoplamiento proteína-ligando, aunque también puede utilizarse para predecir el modo de unión de dos macromoléculas, por ejemplo, acoplamiento proteína-proteína.¹⁶

Existen varios servidores en internet, *suites* y programas disponibles para realizar acoplamiento molecular. Cada herramienta utiliza diferentes algoritmos para la generación, refinamiento y cálculo de los modelos de interacción ligando-receptor.¹⁶ Algunos de ellos se describen en la Tabla 4.

Tabla 4. Ejemplos de programas computacionales disponibles para acoplamiento proteína-ligando y sus algoritmos de búsqueda. Adaptación de Tabla I. en Referencia 16.

Nombre	Algoritmo de búsqueda	Tipo
VINA	Optimización local	Académico
AUTODOCK 4	Algoritmo genético Lamarckiano	Académico
PLANTS	Optimización de colonia de hormigas	Académico
LEDOCK	Recocido simulado	Académico
MOE	Hibrido	Comercial
GOLD	Algoritmo genético	Comercial
DOCK	Coincidencia de formas	Académico
IGEMDOCK	Algoritmo evolutivo	Académico
FLEKSY	Basado en conjuntos	Comercial

Además, estos algoritmos ejecutan predicciones cuantitativas de la energía de unión, proporcionando clasificaciones o puntuaciones según la afinidad de unión de los

complejos ligando-receptor.¹⁷ Los programas de acoplamiento molecular utilizan funciones de puntuación para estimar la energía de unión de los complejos ligando-receptor predichos. La variación de energía, debido a la formación de la estructura ligando-receptor, viene dada por la constante de disociación (K_d) o la energía libre de Gibbs (ΔG_L). La predicción de la energía de unión se hace por medio de la evaluación de los fenómenos físico-químicos implicados en la unión ligando-receptor, incluidas las interacciones intermoleculares, la desolvatación y los efectos entrópicos. Cuanto mayor sea el número y calidad de parámetros físico-químicos evaluados, mayor será la precisión de la función de puntuación.¹⁷

Durante la realización del presente trabajo se utilizó la herramienta AutoDock Vina, programa de código abierto para realizar acoplamiento molecular. AutoDock Vina fue diseñado e implementado originalmente por el Dr. Oleg Trott en el Instituto de investigación Scripps.¹⁸ Este programa ha sido validado ampliamente para la identificación de interacciones proteína-ligando, así como sus modos de unión. Sin embargo, la energía de unión predicha correlaciona pobremente con los datos experimentales, por lo que a menudo se emplean funciones de puntuación para reevaluar los modos de unión predichos por este programa. Dentro de estas funciones de puntuación, las basadas en aprendizaje automático han mostrado ser particularmente útiles.

2.8. ECIF

Extended Connectivity Interaction Features (ECIF) hace referencia a un conjunto de descriptores que toman en cuenta la conectividad de cada átomo para definir tipos de pares proteína-ligando. A partir de estos descriptores se construyen funciones de puntuación mediante inteligencia artificial que mejoran la predicción de afinidad de unión.¹⁹

En este trabajo de tesis se utilizó ECIF para reevaluar las mejores poses (modos de unión) de acoplamiento molecular ligando-receptor entre los compuestos naturales evaluados y DNMT1. Así, se obtuvo una puntuación que permite elegir los mejores candidatos a fármacos entre los compuestos evaluados.

3. Hipótesis

Los productos naturales tienen estructuras químicas diversas con probabilidad de ser inhibidores de DNMTs. Estas características, aunadas al conocimiento de relaciones estructura-actividad de inhibidores de DNMTs conocidos y a la información disponible de las estructuras tridimensionales de DNMTs, permitirán identificar a nuevos inhibidores potenciales de estas enzimas mediante la aplicación de métodos quimioinformáticos y bioinformáticos aplicados a epigenética.

4. Objetivos

4.1. General

Identificar inhibidores potenciales de DNMTs u otros blancos moleculares epigenéticos de interés terapéutico en bases de datos de productos naturales.

4.2. Específicos

- A. Seleccionar y preparar para su uso, bases de datos moleculares o quimiotecas de productos naturales públicas.
- B. Identificar productos naturales con probabilidad de mostrar actividad inhibitoria de DNMT1.
- C. Identificar productos naturales con potencial actividad biológica contra otros blancos moleculares epigenéticos.
- D. Identificar productos naturales con alta afinidad calculada con la estructura cristalográfica de DNMT1.
- E. Determinar el perfil de selectividad de los compuestos de productos naturales en bases de datos públicas.

5. Materiales y métodos

5.1. Bases de datos de productos naturales

Se seleccionaron siete bases de datos de productos naturales disponibles públicamente en internet. Se utilizaron las quimiotecas: BIOFACQUIM, MEGx, NuBBE, CYANOBACTERIA, MARINES, FUNGI y COCONUT. Algunas características como tamaño y contenido se resumen en la Tabla 5.

Tabla 5. Tamaño, contenido y referencias de las quimiotecas de productos naturales utilizadas.

Base de datos	Número de compuestos	Compuestos procesados ^a	Contenido	Referencia
BIOFACQUIM	530	530	Plantas, hongos y propóleos. México	[20]
CYANOBACTERIA	473	456	Metabolitos de cianobacterias	[21]
FUNGI	207	195	Metabolitos fúngicos aprobados y no aprobados	[22]
MARINES	6152	5814	Compuestos marinos.	[23]
MEGx	~ 5000	3992	Plantas y microorganismos. Comercial.	[24]
NuBBE	2218	2015	Plantas, organismos marinos y hongos. Brasil.	[21]
COCONUT	432,706	350,070	Mayor recopilación de productos naturales disponibles de manera libre.	[25]

^a Número de compuestos totales después del proceso de curado

5.2. Preparación o curado de las bases de datos

El proceso de curado o preparación de los compuestos consiste en generar las cadenas en nomenclatura SMILES (del inglés *Simplified Molecular Input Line Entry Specification*), sin información estereoquímica, ya que no todos los compuestos de los conjuntos de datos tienen una estereoquímica definida. Posteriormente, se elimina cualquier elemento que no sea H, B, C, N, O, F, Si, P, S, Cl, Se, Br y I, así como compuestos con errores de valencias. Se neutralizan y re-ionizan los compuestos remanentes; se genera el tautómero canónico para cada compuesto. Por último, se remueven las estructuras duplicadas. El proceso se observa a continuación en la Figura 6.

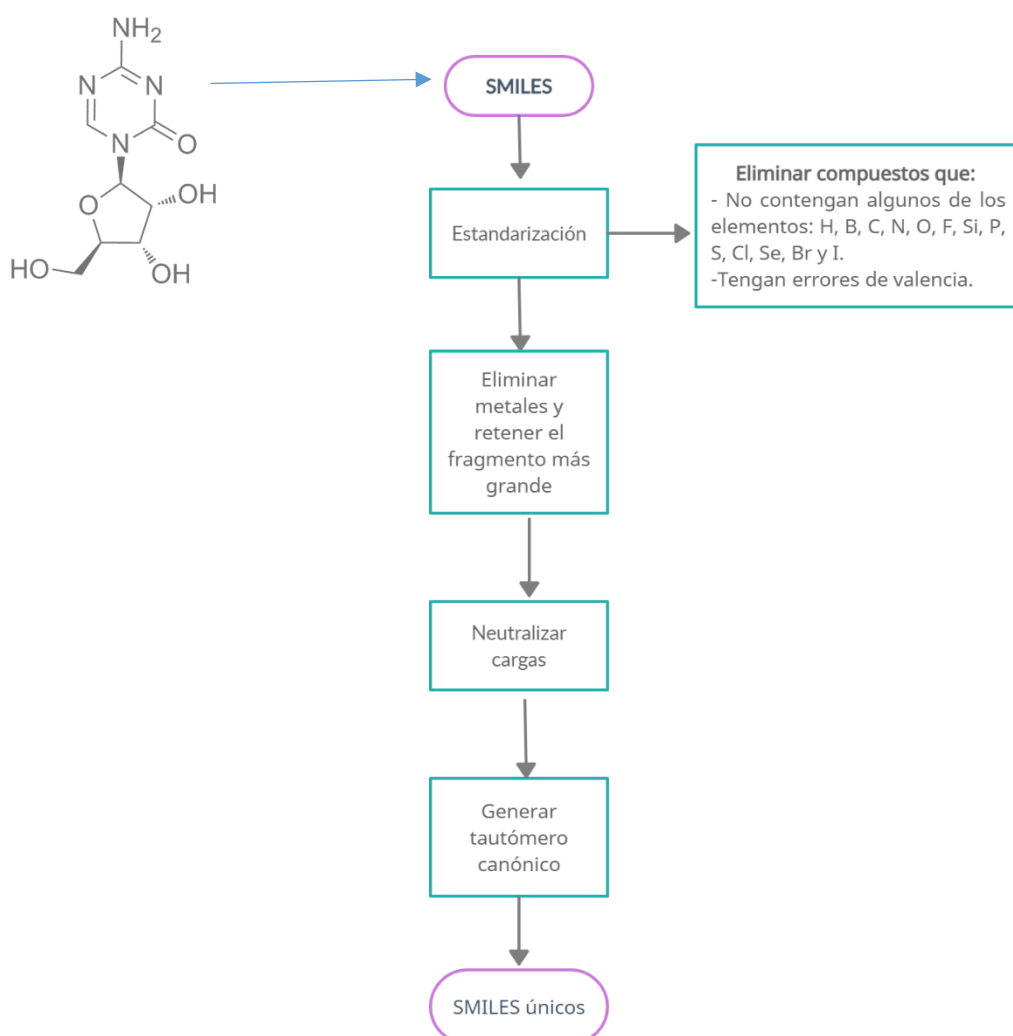


Figura 6. Esquema del proceso de curado descrito en (26) y utilizado para las bases de datos de productos naturales.

Las bases de datos utilizadas para este trabajo pasaron por el proceso de curado anteriormente como se describe en un trabajo anterior del grupo de investigación DIFACQUIM (Ver Referencia 26), utilizando el grupo (*kit*) de herramientas de quiminformática de código abierto RDKit, (versión 2020.03.2.0, RDKit, San Francisco, California, EUA). Se emplearon las funciones: Standardizer, LargestFragmentChoser, Uncharger, Reionizer y TautomerCanonicalizer implementados en la herramienta de estandarización y validación de moléculas MolVS. El objetivo de cada una de las funciones empleadas se puede encontrar en la Tabla A3, en la sección de Anexos.

5.3. Predicción de posibles dianas epigenéticas con *Epigenetic Target Profiler*

Con las bases de datos previamente curadas o preparadas se realizó VS para identificar potenciales inhibidores de DNMT1 u otras dianas epigenéticas, utilizando la herramienta ETP, la cual predice posibles dianas epigenéticas para cada compuesto en la base de datos. Se puede obtener la predicción de más de una diana epigenética por cada compuesto, con diferente o igual precisión de resultado. Así mismo, para algunos compuestos pueden no encontrarse predicciones u obtener datos conocidos. En ETP la precisión de cada resultado se define mediante distancia de Jaccard. Se pueden obtener precisiones desde 0,952 para compuestos clasificados por el programa como Q1, hasta precisiones de 0,773 para compuestos clasificados como Q4.

En la sección 2.6 de Antecedentes se observa el funcionamiento del servidor web de ETP, sin embargo, durante la realización de este trabajo se empleó la herramienta utilizando el lenguaje de programación Python. Los archivos para utilizarlo de esta manera pueden encontrarse en la página de Internet del grupo de investigación DIFACQUIM de la Facultad de Química de la UNAM.

5.4. Identificación de productos naturales con alta afinidad calculada con estructuras cristalográficas de DNMT1

Una vez obtenidas las predicciones de ETP y debido a que es un estudio basado en ligando, se realizó un estudio con base en estructura (e.g., estructura tridimensional de la proteína) para comparar los resultados obtenidos con ETP y los obtenidos con

modelado molecular. Se utilizó la estructura cristalográfica de DNMT1 humana, cuyas coordenadas se encuentran en *Protein Data Bank*, identificador (ID) PDB ID: 4WXX.

Se realizó VS para las bases de datos excepto COCONUT y Marines utilizando AutoDock Vina obteniendo como resultados el modo de unión predicho más favorable.

5.5. Re-evaluación de los resultados de acoplamiento molecular con *Extended Connectivity Interaction Features*

Para las mejores poses de acoplamiento proteína-ligando obtenidas con AutoDock Vina se realizó una re-evaluación utilizando ECIF. Se obtuvo una puntuación (definida en pKd) para cada uno de los ligandos. Al compararse con la puntuación o *score* obtenido en Vina fue útil para seleccionar los compuestos que pudieran ser inhibidores de DNMT1.

6. Resultados y discusión

6.1. Preparación de las bases de datos

Las bases de datos utilizadas en este trabajo fueron curadas como se describe en la sección 5. Materiales y métodos. La cantidad de compuestos antes y después del curado se resumen en la Tabla 5. La base de datos COCONUT, a diferencia de las otras seis bases utilizadas, se filtró para eliminar los compuestos cuyo peso molecular fuera menor a 68 Da o mayor a 800 Da. Esto se hizo para agilizar el proceso de cribado debido a la gran cantidad de información que contiene (432,076 moléculas) y a que éstas son las características de las moléculas con las que fue entrenado ETP.

6.2. Procesamiento de las bases de datos en ETP para identificación de compuestos activos para DNMT1

Se analizaron 363,072 compuestos disponibles en las bases de datos BIOFACQUIM, MEGx, NuBBE, Fungi, Marines, Cyanobacterias y COCONUT utilizando ETP. Los resultados de cada base de datos se analizaron individualmente. Se obtuvieron los resultados de posibles dianas epigenéticas para cada compuesto y el cuartil en que caen de acuerdo a la distancia de Jaccard. Posteriormente se generaron, utilizando el mismo programa ETP, archivos que contuvieran una recopilación de los compuestos que tuvieron actividad sobre cada una de las dianas epigenéticas.

En las quimiotecas de productos naturales analizadas no se encontraron compuestos con actividad biológica predicha para la diana epigenética DNMT1. Sin embargo, debido a que COCONUT se filtró por peso molecular, es posible que algunas moléculas estuvieran fuera de los parámetros antes mencionados (menor a 68 Da o mayor a 800 Da) y no fueran procesadas correctamente. Para confirmar que estos compuestos no tendrían alguna actividad sobre la diana de interés tendría que buscarse otro servidor que pueda procesar moléculas de mayor peso o utilizar otro método computacional. Al igual considerando que ETP toma en cuenta una concentración máxima de 10 micromolar (μM) de los compuestos, es posible que algunos de ellos sí tengan actividad a una concentración mayor.

6.3. Análisis de compuestos activos para todas las dianas epigenéticas disponibles en ETP

Con las predicciones de ETP no se encontraron compuestos nuevos que pudieran tener actividad inhibitoria para DNMT1. Se identificaron a cuatro compuestos con actividad conocida en la base de datos COCONUT. Sin embargo, se realizó un análisis de los resultados obtenidos para todas las dianas epigenéticas en cada una de las quimiotecas.

Se contaron los compuestos activos para cada uno de los blancos moleculares. Los resultados están en la Tabla 6. En esta tabla se reporta el número de compuestos activos considerando compuestos conocidos, además de la cantidad de compuestos disponibles (procesados) en cada base de datos. De los resultados se obtuvo que, para algunos blancos casi la mitad de los compuestos disponibles por base son activos mientras que para otros no se predicen compuestos activos, como es el caso de DNMT1. Esto puede deberse a que no hay suficiente información de compuestos activos contra DNMT1 en el servidor actual de ETP para realizar una predicción confiable de compuestos en las bases de datos de productos naturales.

Tabla 6. Conteo de los compuestos predichos activos por ETP para cada una de las dianas epigenéticas. Se consideran en el conteo compuestos conocidos y predicciones de acuerdo al modelo consenso Morgan::SVM – RDK::SVM.

DIANA	BASE DE DATOS (compuestos)						
	BIOFACQUIM (530)	MEGx (3992)	NuBBE (2015)	FUNGI (195)	MARINES (5814)	CYANOBACTERIA (456)	COCONUT (350,070)
APEX1	433	3405	1614	139	4109	151	237,341
ATM	322	2812	1255	116	4233	203	258,867
AURKA	75	712	338	21	717	7	60,780
AURKB	137	850	535	44	1301	68	97,705
BRD2	310	2414	1143	119	3407	141	209,233
BRD4	117	1129	416	63	1515	141	100,391
BRPF1	0	15	19	0	12	0	2113
CARM1	38	294	171	4	403	31	64,177
CDK1	264	1519	871	81	2071	103	127,126
CDK2	44	401	177	20	894	37	40,350
CDK5	32	170	87	18	242	3	11,396

CDK7	3	16	5	1	90	3	7850
CHEK1	3	49	15	6	144	3	8989
CHUK	0	0	1	0	4	0	336
CREBBP	26	181	83	18	415	59	50,678
DAPK3	1	6	16	0	17	1	1559
DNMT1	0	0	0	0	0	0	4 (KNOWN)
DOT1L	0	4	0	0	5	15	3199
EHMT2	4	17	9	9	40	13	6828
EP300	132	1009	451	88	2080	98	119,920
EZH2	24	200	84	19	396	61	34,732
HDAC1	100	1236	352	57	1365	113	98,671
HDAC10	8	151	39	9	302	85	21,772
HDAC11	27	281	130	17	483	60	28,158
HDAC2	47	302	212	37	661	72	47,019
HDAC3	27	184	132	21	516	54	48,494
HDAC4	28	178	76	15	435	44	40,606
HDAC5	7	86	30	7	163	20	21,118
HDAC6	51	453	202	32	812	66	49,533
HDAC7	44	252	123	25	490	25	38,319
HDAC8	103	453	405	37	874	22	53,706
HDAC9	0	20	5	0	48	0	5223
JAK2	14	112	95	13	355	10	28,058
KAT2B	4	20	7	3	71	10	8507
KDM1A	55	244	223	16	410	72	47,451
KDM4A	1	15	9	0	13	0	2343
KDM4C	54	549	248	28	855	51	85,910
KDM4E	266	1654	818	77	1994	40	115,675
KDM5A	64	409	160	21	770	79	64,053
KDM6B	0	0	0	0	0	0	9
L3MBTL1	0	0	0	0	0	0	336
PARG	0	0	0	0	0	0	197
PARP1	144	1434	610	85	2469	171	168,852
PKN1	50	267	129	12	220	27	28,310
PRKAA1	15	11	16	0	28	0	3600
PRKCB	77	570	218	63	1432	51	56,473
PRKCD	182	1799	685	106	2700	87	112,135
PRKDC	107	820	352	39	1193	65	90,455

PRMT3	0	2	0	0	0	0	797
RPS6KA5	17	87	86	1	105	1	6038
SIRT1	5	52	27	4	108	13	8878
SIRT2	13	13	9	0	20	11	2190
SIRT3	0	0	0	0	0	0	72
TOP2A	3	2	5	0	16	0	1026
USP7	30	350	114	11	592	17	29,020

La Figura 7 muestra los gráficos de acuerdo al porcentaje de compuestos activos por base de datos para cada diana. Se observa que más del 70% de los compuestos disponibles en cada base tienen como diana a la enzima APEX1 (DNA liasa) seguida de ATM (codifica para proteína ATM serina/treonina) y BRD2 (proteína 2 con bromodominio).

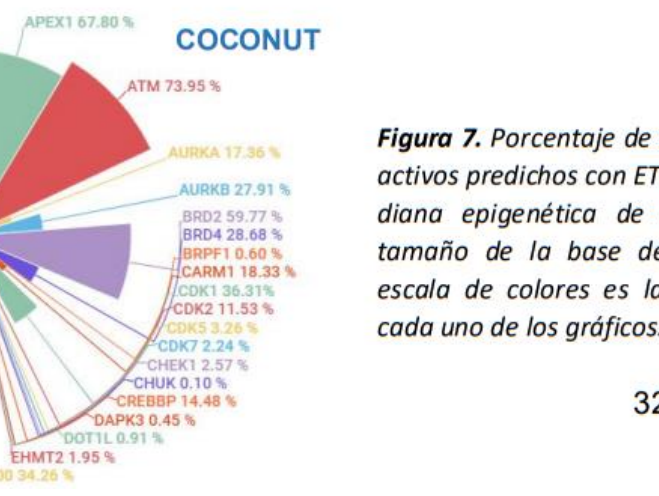
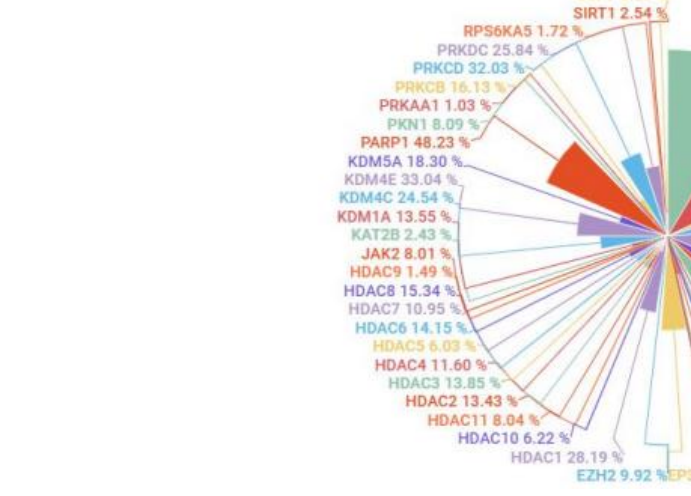
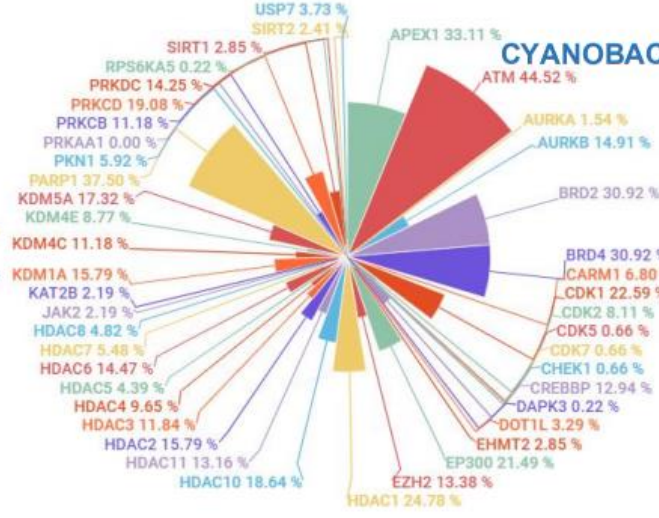
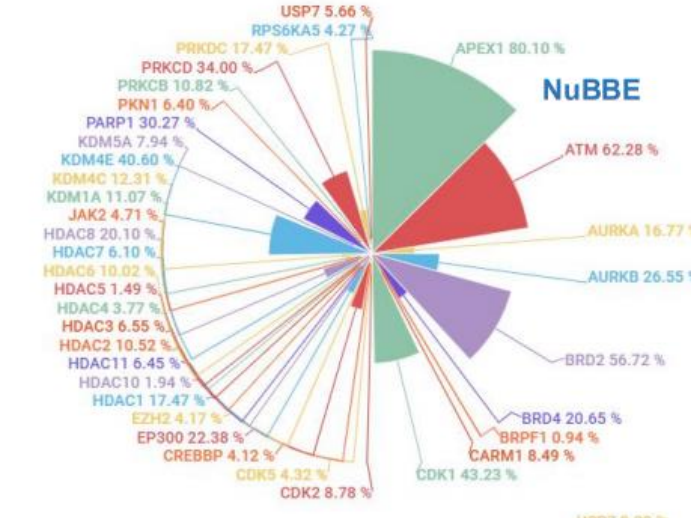
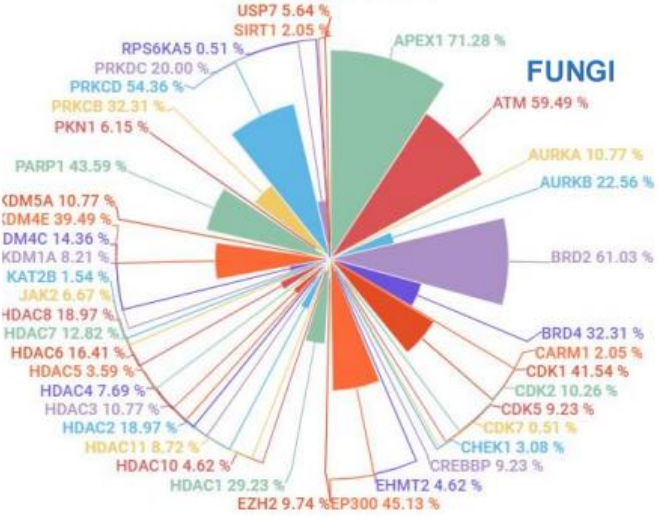
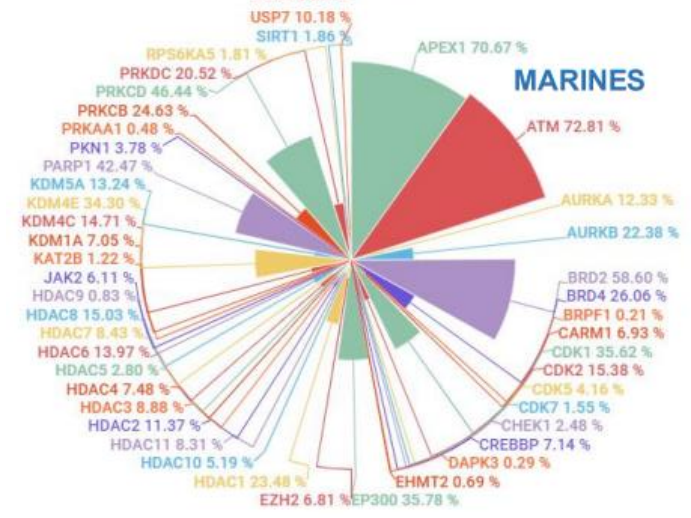
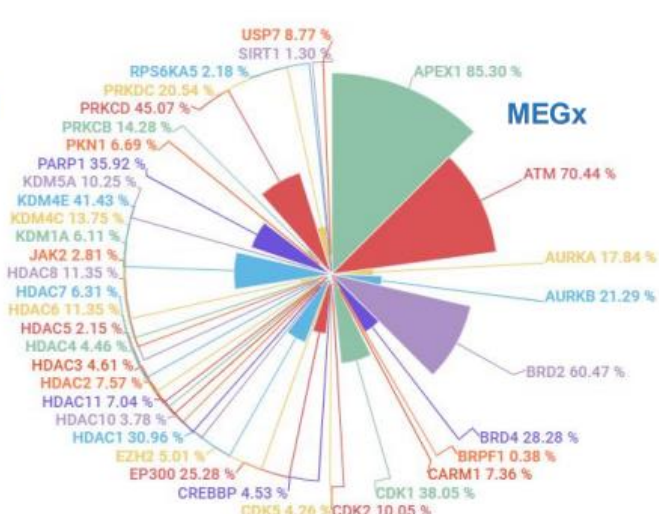
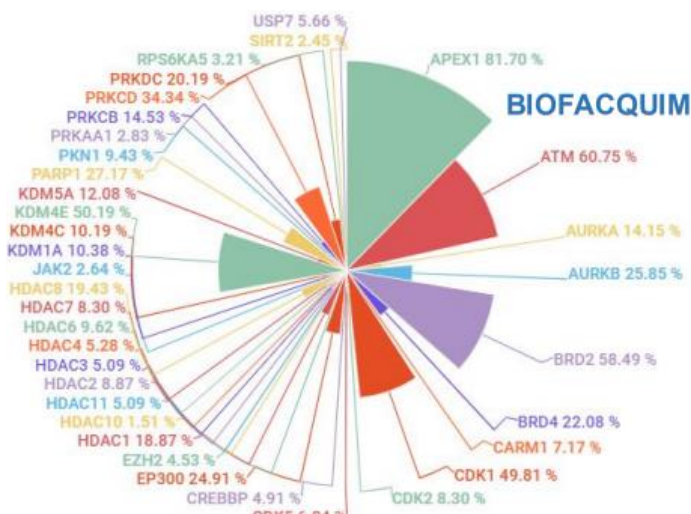


Figura 7. Porcentaje de compuestos activos predichos con ETP para cada diana epigenética de acuerdo al tamaño de la base de datos. La escala de colores es la misma en cada uno de los gráficos.

Otros blancos moleculares epigenéticos que tuvieron una gran cantidad de compuestos posiblemente activos fueron: HDAC3 (histona deacetilasa 3), EZH2 (histona-lisina metiltransferasa), PRKCD (proteína quinasa tipo C).

Después de contar el número de compuestos activos para cada diana, se determinó si dentro de estos resultados existían compuestos selectivos para un solo blanco molecular. Para ello se recopilaron todos los datos positivos obtenidos en cada base de datos y se filtró el número de veces que se repetía un compuesto de acuerdo a los 55 blancos disponibles. Los resultados se discuten en la siguiente sección.

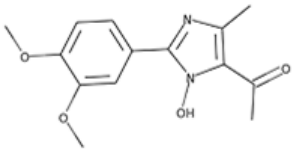
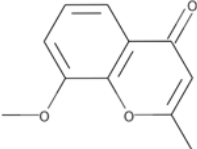
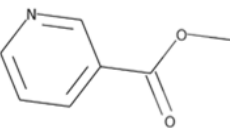
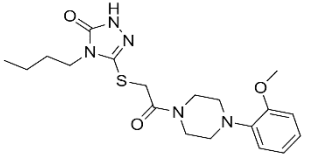
6.4. Cálculo de compuestos selectivos para todas las dianas epigenéticas

Se realizó un análisis a los resultados de cada uno de las dianas biológicas disponibles en ETP para cada una de las bases de datos en estudio. Este análisis tuvo como objetivo determinar si había compuestos altamente selectivos para alguno de las 55 dianas aún si no eran DNMT1.

Se encontró que la mayoría de los compuestos en cada una de las bases de datos son poco selectivos ya que tienden a tener afinidad hasta con 20 o más blancos moleculares. No obstante, se lograron encontrar compuestos selectivos (solo 1 diana). Los resultados para las bases de datos excepto COCONUT se muestran en la Tabla A1 en la sección de Anexos. En la tabla se observa que la mayoría de los compuestos selectivos tienen afinidad con dianas epigenéticas muy estudiadas como APEX1. Al igual todos los compuestos caen en un cuartil Q4 de acuerdo a la distancia de Jaccard para ETP, determinando que la precisión de estos resultados no es mayor a 0.773.

Para COCONUT se obtuvieron un total de 2,316 compuestos selectivos. Sin embargo, se seleccionaron los compuestos con una mejor precisión en la predicción (Q2 y Q3) para su análisis. Se muestran en la Tabla A2 en la sección de Anexos. Al igual que para las otras seis bases de datos se observa que la mayoría de estos compuestos tienen interacción con dianas muy estudiadas lo que aumenta la probabilidad de similitud estructural con compuestos activos conocidos para estos blancos. La Tabla 7 resume los resultados con la mayor precisión encontrados dentro de todas las bases de datos.

Tabla 7. Compuestos selectivos con mayor precisión en ETP.

ID	DIANA	BASE DE DATOS	ESTRUCTURA	CUARTIL
CNP0009801	BRD2	COCONUT		Q2
CNP0134517	PRKDC	COCONUT		Q2
CNP0284518	KDM4A	COCONUT		Q2
CNP0345397	PARP1	COCONUT		Q2

Se determinó si las bases de datos seguían un patrón de acuerdo a los resultados de compuestos selectivos identificados. Para ello se obtuvo el porcentaje de compuestos selectivos en cada una de las bases encontrando que para todas es un promedio muy bajo no superior a 1.5% (Tabla 8).

Tabla 8. Porcentaje de compuestos selectivos en cada base de datos.

Base de datos	Tamaño	No. Compuestos selectivos	% Compuestos selectivos
FUNGI	195	1	0.51
CYANOBACTERIA	456	3	0.66
BIOFACQUIM	530	2	0.38
NuBBE	2015	26	1.29
MEGx	3992	31	0.78
MARINES	5814	19	0.33
COCONUT	350070	2316	0.66
		Promedio	0.66
		Media	0.60

Debido a que mediante el cribado virtual con ETP se obtuvieron resultados de posibles compuestos activos para las diferentes dianas epigenéticas y estos resultados

se basan en ligandos con actividad conocida para cada diana se pretende realizar un análisis basado en blanco molecular para las mismas siete bases de datos mediante acoplamiento o *docking* esperando obtener resultados para DNMT1. Al igual se pretende realizar una re-evaluación utilizando la herramienta ECIF. Los resultados se presentan y discuten en la siguiente sección.

6.5. Preparación de proteína (DNMT1) y bases de datos para acoplamiento molecular

Para poder obtener un análisis basado en blanco molecular, en este caso DNMT1, se utilizó el programa AutoDock Vina para realizar acoplamiento molecular, posteriormente se llevó a cabo una re-evaluación con ECIF. Se utilizó la estructura cristalográfica de la enzima DNMT1 PDB ID: 4WXX, la cual se encuentra en complejo con SAH (S-adenosil-1-homocisteína) a una resolución de 2.62 Å. Se eliminó una de las cadenas de la proteína y la molécula en complejo SAH quedando únicamente la cadena A en donde se encuentra el sitio activo de interés. La proteína se preparó utilizando MOE 2014 con las herramientas pre-establecidas de la función “Preparar estructura”.

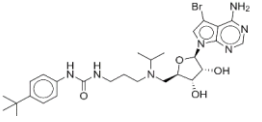
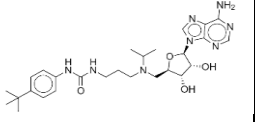
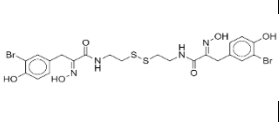
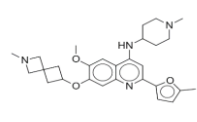
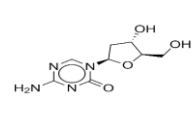
Las bases de datos se prepararon realizando un filtro de peso molecular: se recuperaron únicamente los compuestos con un peso menor a 800 Da para agilizar el proceso ya que pesos moleculares muy grandes son difíciles de procesar y sus propiedades farmacológicas se afectan. Tanto la proteína como los compuestos se prepararon antes de realizar el acoplamiento utilizando Openbabel para añadir hidrógenos y a pH = 7. Al igual se utilizó la herramienta `prepare_ligand(receptor)4.py` de MGLTools. Una vez preparados se hizo el acoplamiento con AutoDock VINA.

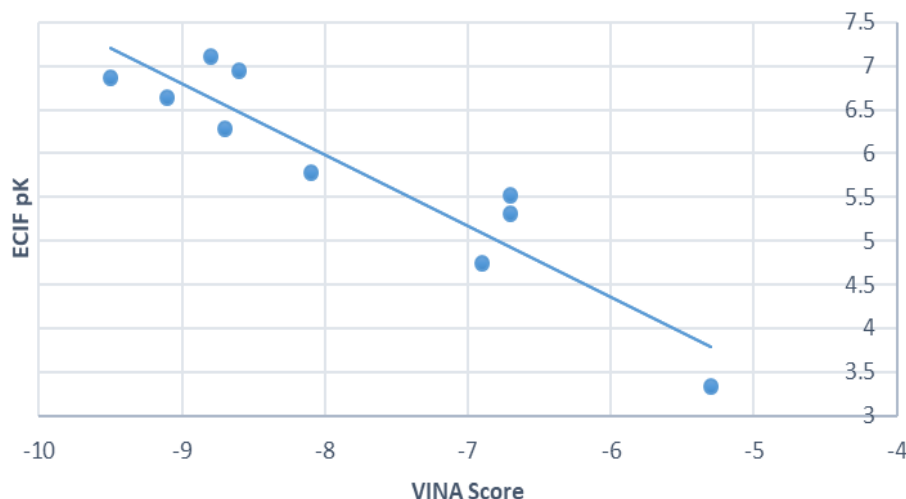
Para hacer la re-evaluación con ECIF se prepararon los archivos de salida del acoplamiento anterior utilizando X-Tools y Standardizer. Una vez preparados, se utilizó el *script* de ECIF para Python3.

Primero se realizó el acoplamiento molecular y re- evaluación con un grupo de 10 compuestos de los cuales 5 tenían actividad biológica conocida con DNMT1 y 5 eran compuestos inactivos. En la Tabla 9 se pueden encontrar los nombres, estructura, referencias y resultados de los compuestos activos. Al igual que en la Grafica 1, se observa la relación de los resultados o puntaje obtenido en el acoplamiento con VINA y

en la re- evaluación con ECIF. Se observa una relación lineal entre los resultados por lo que se espera que en las bases de datos analizadas se obtengan resultados mayores de pK en ECIF para los mejores puntajes obtenidos con VINA.

Tabla 9. Datos y resultados de acoplamiento molecular para compuestos con actividad biológica conocida con DNMT1.

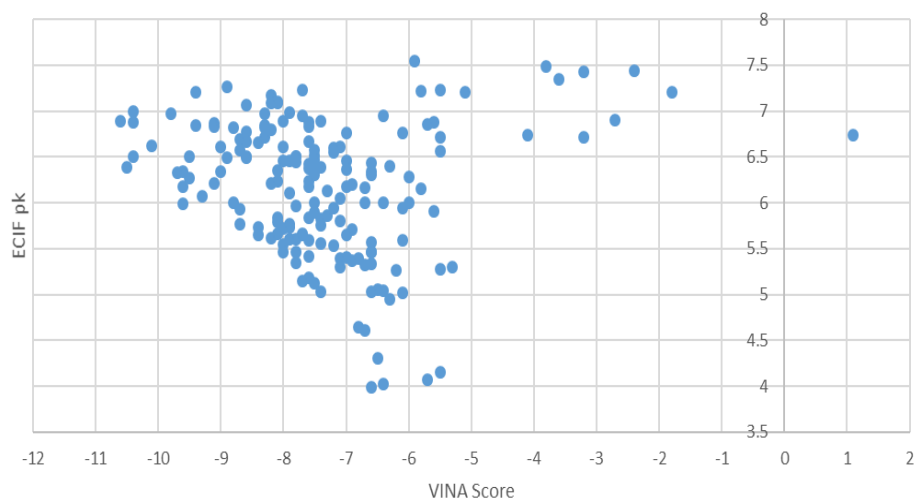
	Ligando 1	Ligando 2	Ligando 3	Ligando 4	Ligando 5
Estructura					
Nombre	SGC 0946	5'-[(3-[(4-tert-butylphenyl)carbamoyl]amino)propyl](propan-2-yl)amino]-5'-deoxyadenosine	Psammaplin A	6-methoxy-7-((2-methyl-2-azaspiro[3.3]heptan-6-yl)oxy)-2-(5-methylfuran-2-yl)-N-(1-methylpiperidin-4-yl)quinolin-4-amine / BDBM5045772 2	Decitabina
Fórmula	C ₂₈ H ₄₀ BrN ₇ O ₄	C ₂₇ H ₄₀ N ₈ O ₄	C ₂₂ H ₂₄ Br ₂ N ₄ O ₆ S ₂	C ₂₈ H ₃₆ N ₄ O ₃	C ₈ H ₁₂ N ₄ O ₄
ID ChEMBL	CHEMBL3087498	CHEMBL2171169	SCHEMBL364511	CHEMBL4215326	CHEMBL1201129
Peso Molecular [Da]	617.23	540.32	661.95	476.28	228.09
pIC₅₀	9.52	8.54	7.73	7.68	7.52
VINA Score	-8.5	-8.7	-7.9	-9.4	-6.7
ECIF Score (Predicted Pk)	7.45	6.90	5.89	7.19	5.23
Referencia	[27]	[28]	[29]	[30]	[31]



Gráfica 1. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF.

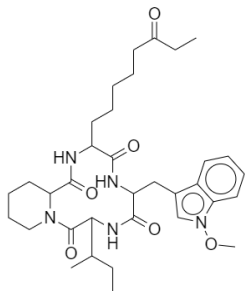
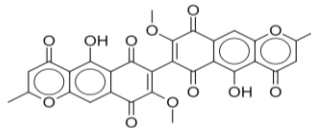
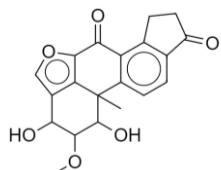
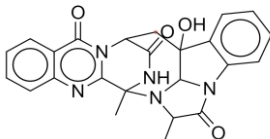
Como segundo paso se analizaron las bases de datos de menor tamaño: FUNGI, CYANOBACTERIA y BIOFACQUIM. El tratamiento de las bases de datos antes del acoplamiento fue el mismo que se realizó para el grupo de compuestos de prueba. Una vez obtenidos los resultados de VINA y ECIF para cada base, se realizó un primer filtro de acuerdo a los resultados de VINA y uno más basado en los resultados de ECIF, esto para disminuir el número de compuestos candidatos a inhibidores de los cuales tendrían que realizarse pruebas experimentales.

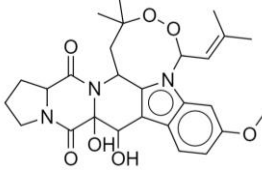
Para la base de datos FUNGI se seleccionaron los compuestos con un puntaje de Vina mayor a -8.5 ya que es el valor correspondiente al compuesto activo con un mayor pIC_{50} del grupo de prueba. Una vez seleccionados los compuestos, se hizo un segundo filtrado de acuerdo a los resultados de pK obtenidos en la re-evaluación con ECIF. Al no haber compuestos con un pK mayor a 7.4, correspondiente al valor ECIF del ligando con mayor pIC_{50} , se tomó como punto de corte el valor 6.9 correspondiente al segundo ligando de prueba con mayor actividad biológica. Los resultados se muestran en la Tabla 10. En esta tabla se observa que los compuestos tienen partes parecidas en su estructura principalmente ciclos aromáticos que contienen nitrógeno u oxígeno. La relación de los resultados de acoplamiento molecular y re-evaluación de la misma se presenta en la Gráfica 2.



Gráfica 2. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF para la base de datos FUNGI.

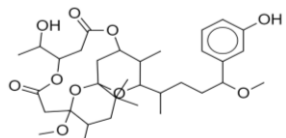
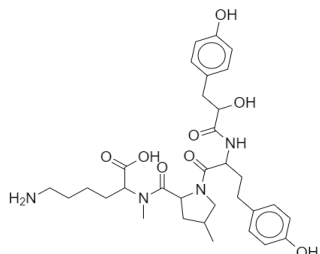
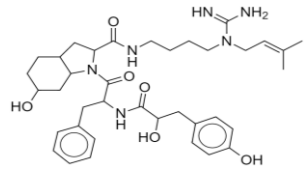
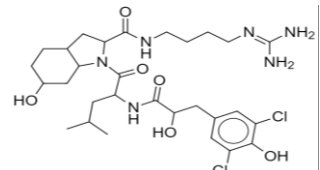
Tabla 10. Compuestos seleccionados para la base de datos FUNGI con un mejor resultado de acoplamiento molecular y re-evaluación utilizando AutoDock VINA y ECIF.

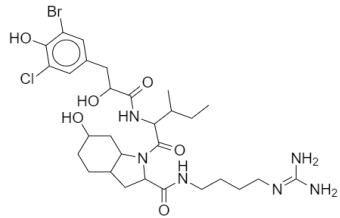
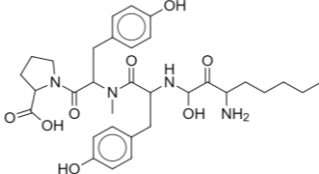
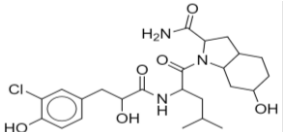
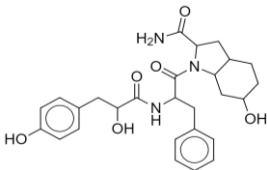
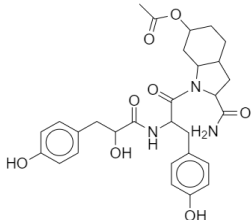
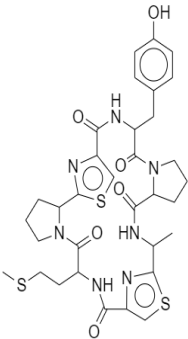
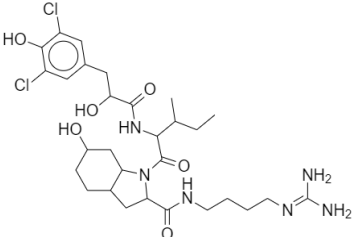
ID	VINA SCORE	ECIF Score (Predicted Pk)	Peso molecular (Da)	Estructura
FUNGI 98	-8.9	7.27	623.37	
FUNGI 174	-9.4	7.21	570.08	
FUNGI 164	-8.6	7.06	354.11	
FUNGI 179	-10.4	6.99	443.16	

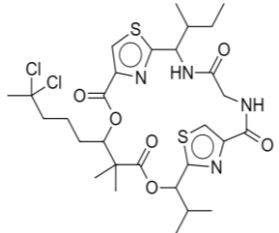
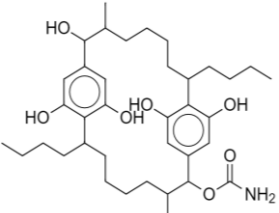
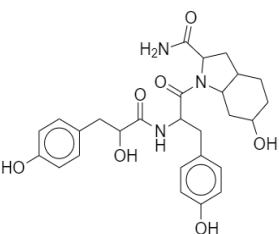
FUNGI 177	-9.8	6.98	511.23	
-----------	------	------	--------	---

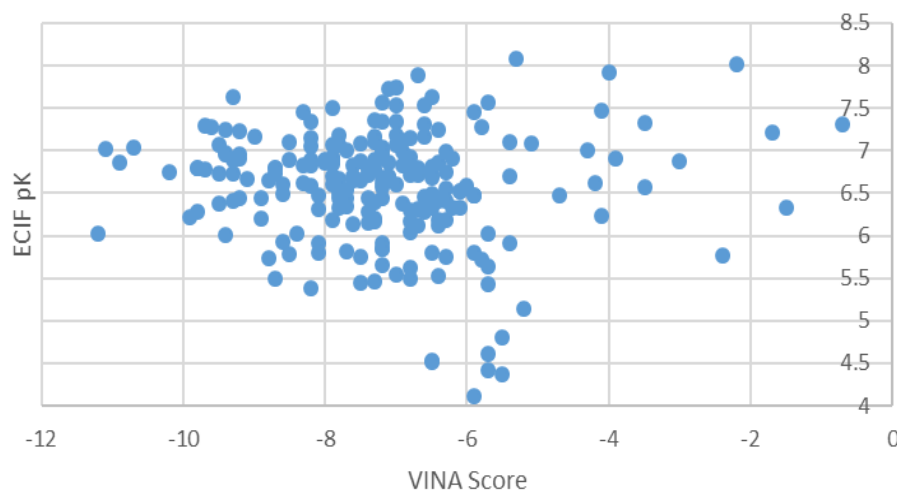
Para la base de datos CYNOBACTERIA se tomaron en cuenta los mismos criterios que para FUNGI, un punto de corte en los resultados de Vina de -8.5 y para ECIF a un valor de 6.9. Los resultados se muestran en la Tabla 11, en donde se observan compuestos con varios ciclos que incluyen en su estructura nitrógeno y azufre; la relación entre la puntuación de Vina y ECIF se observa en la Gráfica 3.

Tabla 11. Compuestos seleccionados para la base de datos CYANO BACTERIA con un mejor resultado de acoplamiento molecular utilizando AutoDock VINA.

ID	VINA SCORE	ECIF Score (Predicted Pk)	Peso molecular (Da)	Estructura
CYANO 121	-9.3	7.64	606.34	
CYANO 126	-9.7	7.29	612.31	
CYANO 165	-9.6	7.27	676.39	
CYANO 143	-9.4	7.24	642.27	

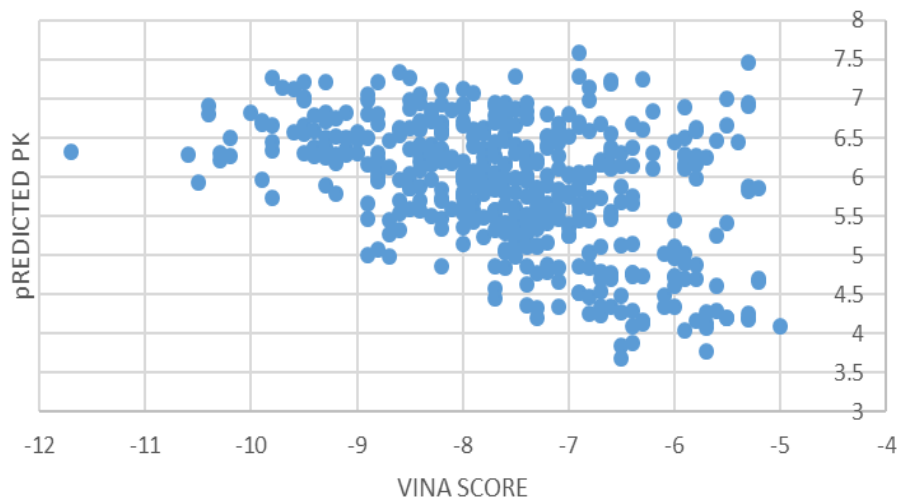
CYANO 174	-9.2	7.23	686.22	
CYANO 127	-9	7.16	612.31	
CYANO 83	-8.5	7.10	495.21	
CYANO 84	-9.5	7.06	495.24	
CYANO 99	-10.7	7.03	553.24	
CYANO 195	-11.1	7.027	725.21	
CYANO 144	-9.4	6.97	642.27	

CYANO 161	-9.2	6.95	674.18	
CYANO 136	-9.4	6.95	627.41	
CYANO 88	-9.2	6.91	511.23	



Gráfica 3. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF para la base de datos CYANOBACTERIA.

Para la base de datos BIOFACQUIM se tomaron en cuenta los mismos parámetros anteriores para realizar los filtros que permitieran seleccionar los mejores candidatos a compuestos activos. La Gráfica 4 muestra la relación de los resultados obtenidos para el acoplamiento molecular y la re-evaluación.



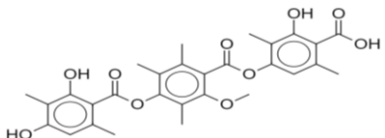
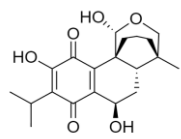
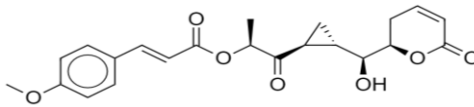
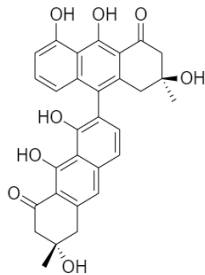
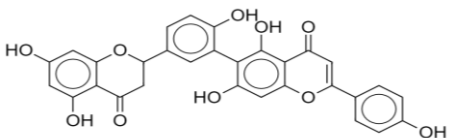
Gráfica 4. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF para la base de datos BIOFACQUIM.

Se tomaron como punto de corte un valor de -8.5 para VINA y 6.9 para ECIF. Los resultados se muestran en la Tabla 12. Al igual que en las bases de datos anteriores, se observan compuestos largos y con compuestos aromáticos cíclicos.

Tabla 12. Compuestos seleccionados para la base de datos BIOFACQUIM con un mejor resultado de acoplamiento molecular y re-evaluación utilizando AutoDock VINA y ECIF.

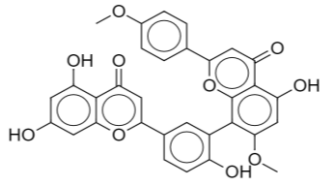
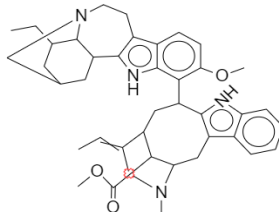
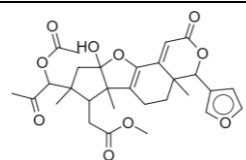
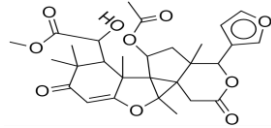
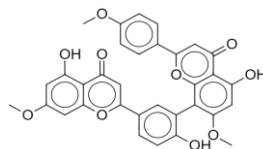
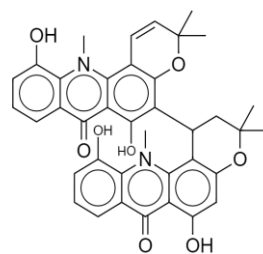
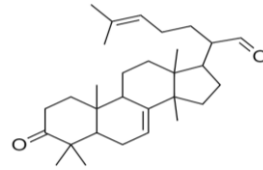
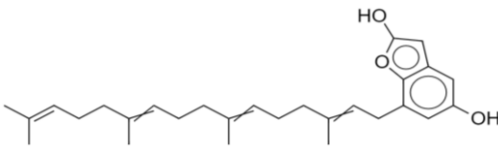
ID	VINA SCORE	ECIF Score (Predicted pK)	Peso molecular (Da)	Estructura
FQNP152	-8.6	7.34	454.34	
FQNP204	-8.5	7.27	356.13	
FQNP19	-9.8	7.26	492.24	

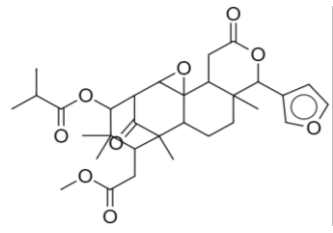
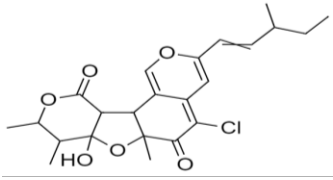
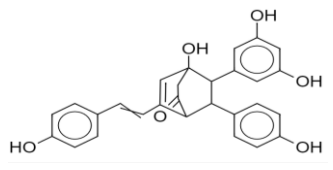
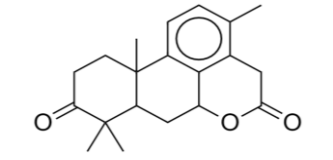
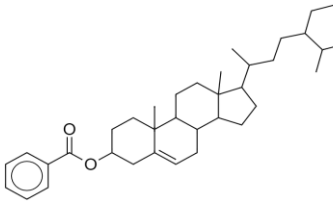
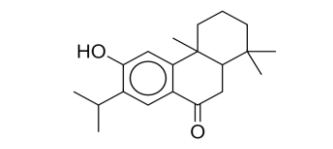
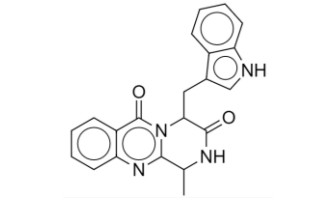
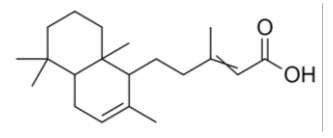
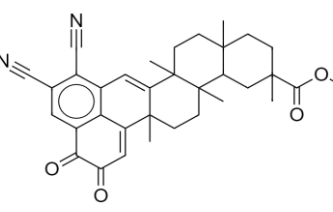
FQNP15	-9.5	7.22	468.23	
FQNP596	-9.3	7.21	544.21	
FQNP18	-8.8	7.21	538.24	
FQNP17	-9.7	7.15	538.24	
FQNP16	-9.6	7.13	508.23	
FQNP206	-8.9	7.05	360.16	
FQNP594	-9.5	7.02	574.22	

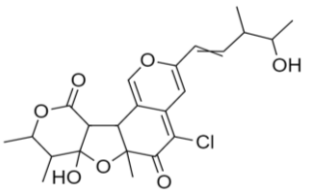
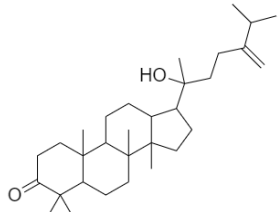
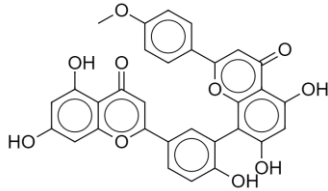
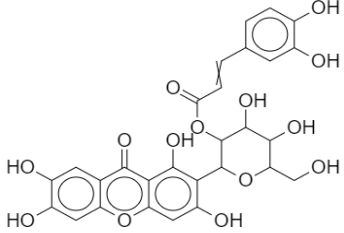
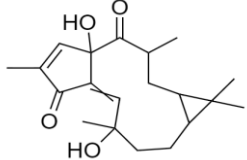
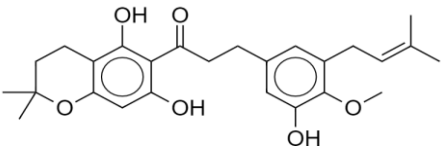
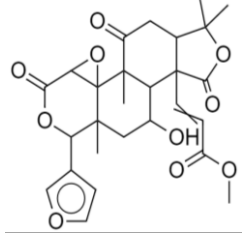
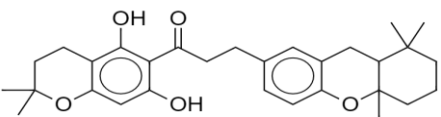
FQNP135	-9.5	6.99	538.18	
FQNP212	-8.9	6.98	362.17	
FQNP156	-8.5	6.98	400.15	
FQNP599	-9.5	6.98	514.16	
FQNP119	-10.4	6.91	540.10	

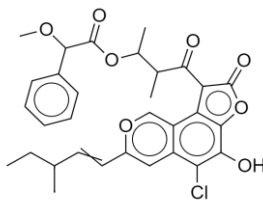
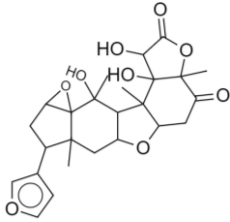
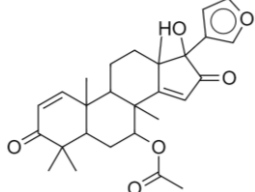
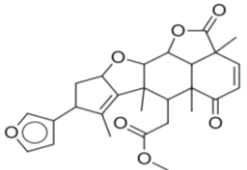
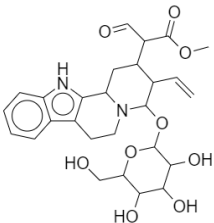
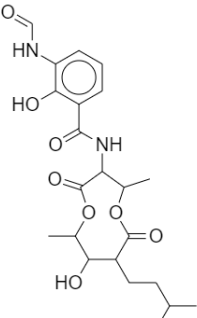
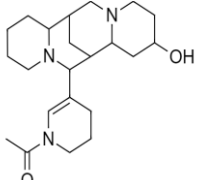
Para la base de datos NuBBE se tomaron como punto de corte un valor de -8.5 para VINA y 6.9 para ECIF. Los resultados se muestran en la Tabla 13.

Tabla 13. Compuestos seleccionados para la base de datos NuBBE con un mejor resultado de acoplamiento molecular y re-evaluación utilizando AutoDock VINA y ECIF.

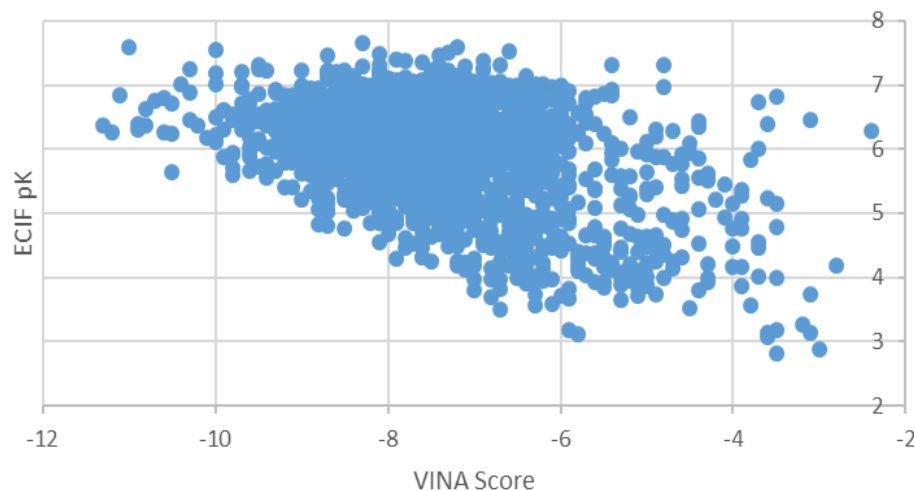
ID (NuBBE)	VINA SCORE	ECIF Score (pK)	Peso Molecular (Da)	Estructura
1006	-11	7.58	566.12	
1427	-10	7.53	646.39	
392	-8.7	7.46	542.21	
480	-9.5	7.32	542.21	
995	-10.3	7.25	580.14	
494	-9.5	7.24	646.23	
79	-8.5	7.23	438.35	
655	-8.7	7.23	422.28	

1982	-9.4	7.22	556.27	
1891	-9	7.22	434.15	
994	-8.7	7.21	456.16	
1221	-8.6	7.20	312.17	
567	-9.7	7.20	518.41	
1405	-8.7	7.18	300.21	
515	-10	7.17	358.14	
1410	-8.5	7.17	304.24	
1963	-8.7	7.14	536.27	

1893	-8.7	7.09	450.14	
1284	-8.5	7.07	456.39	
1017	-10	7.02	552.10	
1637	-10.4	7.02	584.12	
1801	-8.6	7.00	332.19	
1353	-9	6.99	440.22	
354	-8.6	6.99	514.18	
219	-9.7	6.99	478.27	

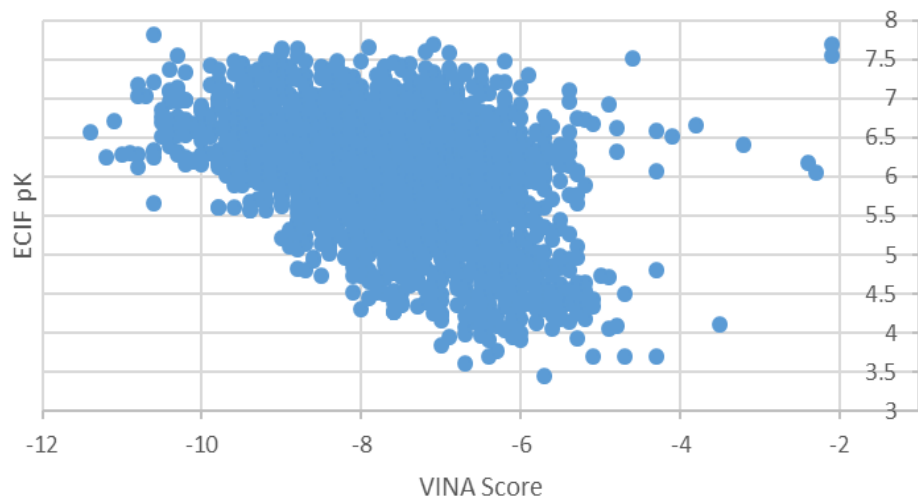
1883	-9.7	6.98	566.17	
456	-9.7	6.97	474.19	
1924	-8.9	6.95	466.23	
307	-8.8	6.94	466.19	
2015	-8.5	6.93	530.23	
1950	-8.9	6.92	450.20	
383	-9.3	6.91	373.27	

La relación de las puntuaciones para Vina y ECIF para la base de datos NuBBE se muestra en la Gráfica 5.



Gráfica 5. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF para la base de datos NuBBE.

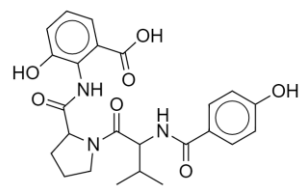
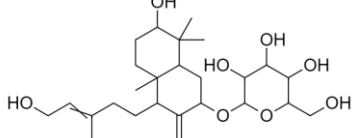
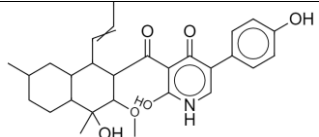
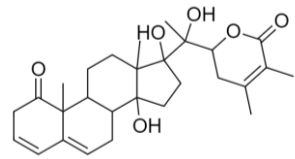
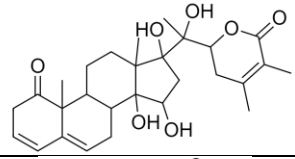
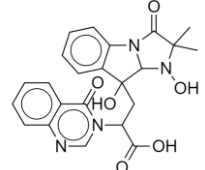
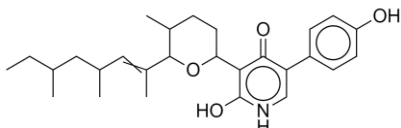
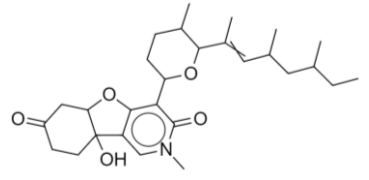
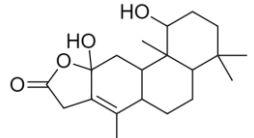
Por último, se obtuvieron los resultados score de VINA y re-evaluación del acoplamiento molecular con ECIF para la base de datos MEGx. La relación entre los resultados se muestra en la Gráfica 6.

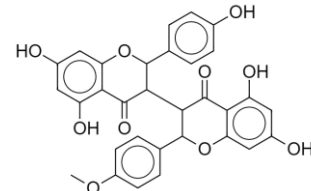
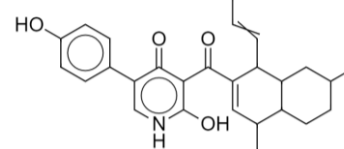
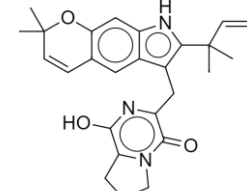
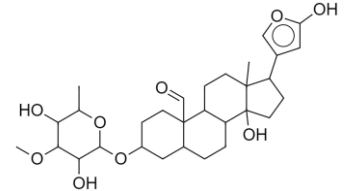
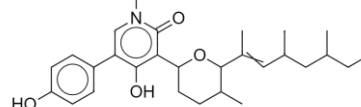
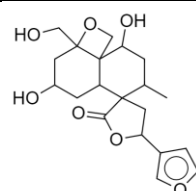
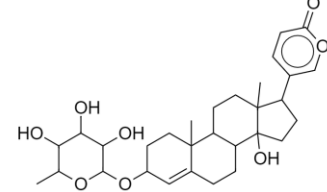


Gráfica 6. Relación de los resultados de acoplamiento molecular con VINA y re-evaluación utilizando ECIF para la base de datos MEGx.

Debido a que se obtuvieron más de 30 compuestos candidatos al realizar los filtros de puntuación como se realizó para las bases de datos anteriores, se determinaron como puntos de corte para Vina -8.5 y para ECIF 7.4 que corresponden a los valores para el compuesto de mayor actividad biológica en el grupo de prueba. Los resultados se resumen en la Tabla 14.

Tabla 14. Compuestos seleccionados para la base de datos MEGx con un mejor resultado de acoplamiento molecular y re-evaluación utilizando AutoDock VINA y ECIF.

ID (MEGx)	Peso molecular (Da)	VINA SCORE	ECIF Score (Precited pK)	Estructura
1666	468.18	-10.6	7.82	
2058	484.30	-8.8	7.65	
2843	480.24	-9	7.63	
3559	470.27	-9	7.60	
3455	486.26	-8.8	7.55	
3309	449.15	-10.3	7.55	
1701	438.26	-9.2	7.50	
2966	471.29	-9.1	7.49	
2259	334.21	-8.7	7.48	

3026	556.14	-9.6	7.48	
2394	432.22	-9	7.45	
2699	431.22	-9.5	7.45	
3275	548.29	-9.5	7.44	
1919	452.28	-9.9	7.42	
3373	378.17	-8.9	7.41	
3300	530.29	-8.5	7.40	

7. Conclusiones

Se realizó el cribado virtual de quimiotecas de productos naturales accesibles públicamente, como potenciales inhibidores de DNMT1 y de otros blancos epigenéticos. Se identificaron posibles compuestos inhibidores de dianas epigenéticas diferentes a DNMTs mediante diseño de fármacos basado en ligando, específicamente utilizando ETP, una aplicación libre desarrollada en la Facultad de Química de la UNAM, desarrollado a partir de modelos de aprendizaje automático (*machine learning*).

La mayoría de los compuestos identificados por métodos computacionales mostraron una predicción de selectividad baja, es decir, se predice que cada producto natural interaccione con más de una diana epigenética. Además, aproximadamente el 70% de los compuestos se identificaron como activos para dianas epigenéticas mayormente estudiadas como APEX1 y BRD2.

Con los resultados de esta investigación también se encontró que, en promedio, solo el 0.66 % de los productos naturales en cada quimioteca son selectivos para una sola diana epigenética. De estos compuestos selectivos (Tablas A1 y A2, Sección de Anexos), aunque presentan una precisión de resultados baja (Q3 y Q4) se destaca que:

- * La mayoría de los compuestos con afinidad por APEX1 tienen grupos hidroxilo y carbonilo en su estructura, además de que algunos tienen cadenas de carbonos y dobles enlaces. Es el mismo caso para ATM y AURKB.
- * Compuestos que mostraron afinidad por HDAC8 y HDAC11 contienen nitrógeno, grupo amino, en su estructura, al igual que para BRD2.

Las estructuras para las que se obtuvo una precisión y selectividad alta de predicción con ETP (Q2) son cuatro y se encuentran en la base de datos COCONUT. Sus dianas epigenéticas son BRD2, PRKCD, KDM4A y PARP1, las cuales se asocian a enfermedades como epilepsia fotosensible, síndrome linfoproliferativo autoinmune, síndrome de Meier- Gorlin, difteria y xeroderma pigmentoso. Debido a esto resultados, se proponen como productos naturales candidato para realizar pruebas experimentales de inhibición enzimática.

Mediante diseño de fármacos basado en estructura utilizando acoplamiento molecular y una re- evaluación con ECIF, se lograron identificar al menos cinco compuestos con alta afinidad calculada para DNMT1 provenientes de cada base de datos de productos naturales. Los compuestos seleccionados provienen de las bases de

datos FUNGI, CYANOBACTERIA, BIOFACQUIM, MEGx y NuBBE y se muestran en la sección de Resultados. Estos compuestos pueden ser candidatos a pruebas experimentales mediante inhibición enzimática debido a que se seleccionaron a partir de los resultados de acoplamiento molecular obtenidos para compuestos con actividad biológica conocida.

De los resultados y análisis realizados en esta investigación también se concluye que las técnicas computacionales empleadas fueron útiles para identificar productos naturales con posible actividad biológica con dianas epigenéticas. ETP es una herramienta que facilita y agiliza la identificación de compuestos debido a que no se necesita demasiada preparación de las bases de datos. Sin embargo, como se observó durante la realización de este trabajo es difícil obtener resultados con alta precisión y selectividad. Esto está vinculado con la información que está disponible en bases de datos quimiogenómicas públicas como ChEMBL, que es la fuente de información experimental para el desarrollo de los modelos predictivos de ETP, al contrario, el acoplamiento molecular automatizado, es una técnica basada en el receptor que requiere mayor preparación de los datos, pero permite obtener resultados con mayor precisión o afinidad con la diana de interés.

Los resultados de esta tesis son una contribución de la epi-informática aplicada a productos naturales para la identificación de potenciales candidatos a epi-fármacos. Es parte de un programa de investigación multidisciplinaria enfocada en la búsqueda y desarrollo de compuestos bioactivos contra blancos epigenéticos de interés terapéuticos. Los pasos siguientes de esta investigación se describen en la sección 8, Perspectivas.

8. Perspectivas

Las etapas que se derivan de esta investigación son las siguientes:

- Determinar la relación estructural entre los compuestos con alta afinidad calculada para DNMT1.
- Determinar la afinidad calculada de los productos naturales predichos como activos obtenidos con ETP y la diana epigenética para la cual se predijo la actividad.
- Adquirir muestras físicas de 1-5 mg de cada producto natural seleccionado a evaluar mediante inhibición enzimática. Analizar la pureza de cada compuesto a evaluarse.
- Evaluar experimentalmente los compuestos más prometedores y de los cuales se puede obtener una muestra física, con los blancos epigenéticos predichos con ETP y acoplamiento molecular. La evaluación experimental puede realizarse mediante pruebas bioquímicas de inhibición enzimática.

9. Referencias

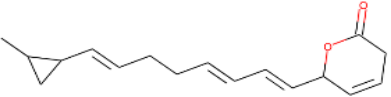
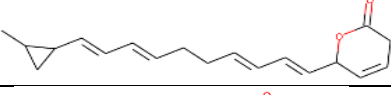
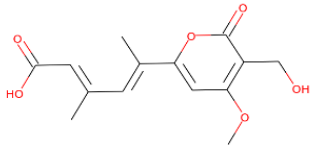
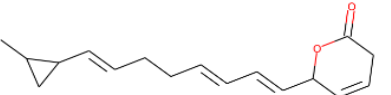
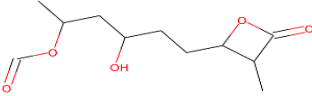
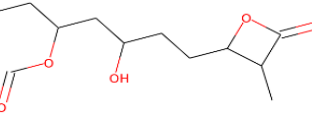
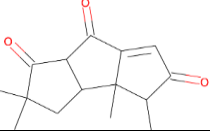
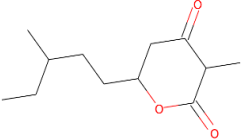
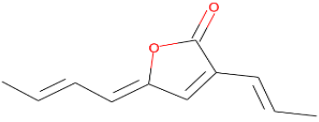
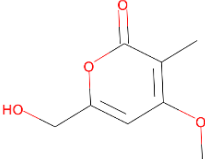
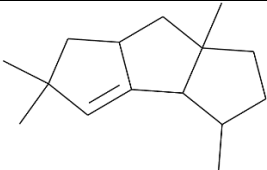
1. Medina, J.L; Fernandez,E.; Naveja,J. Avances en el diseño de fármacos asistido por computadora, Educ. Quím. (2015) 26, 180-186.
2. Sabe, V., Ntombela, T., Jhamba, L., Maguire, G., Govender, T., Naicker,T.; Kruger,H.; Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review, Eur. J. Med. Chem (2021) 224, 113705.
3. Medina, J.L. Grand challenges of computer-aided drug design: The road ahead. Front. Drug. Discov . (2021) 1. 728551.
4. Dueñas, A.; Naveja, J.; Medina, J.L. Chapter 1 - Introduction of epigenetic targets in drug discovery and current status of epi-drugs and epi-probes, Epi-Informatics, Academic Press (2016) 1-20.
5. Sessions, Z.; Sánchez-Cruz, N.; Prieto-Martínez, F. D.; Alves, V. M.; Santos Jr. H. S., Muratov, E.; Tropsha, A.; Medina-Franco, J. L. Recent progress on cheminformatics approaches to epigenetic drug discovery. Drug Discovery Today. (2020) 25, 2268-2276.
6. Elnitski, L; Epigenética, National Human Research Institute, Consultado el 20 de Septiembre de 2021 en: <https://www.genome.gov/es/genetics-glossary/Epigenetica>
7. Nessa Carey. (2015). *Epigenetics for Drug Discovery*. Royal Society of Chemistry.
8. Hu, C., Liu, X., Zeng, Y, DNA methyltransferase inhibitors combination therapy for the treatment of solid tumor: mechanism and clinical application. Clin Epigenet (2021) 13, 166.
9. Zhang ZM, Liu S, Lin K, Luo Y, Perry JJ, Wang Y, Song J. Crystal Structure of Human DNA Methyltransferase 1. J Mol Biol. 2015 Jul 31; 427(15):2520-2531.
10. Yoo, J; Medina J.L. Trimethylaurintricarboxylic acid inhibits human DNA methyltransferase 1: insights from enzymatic and molecular modeling studies, J Mol Model (2012) 18:1583–1589.
11. García, P; Síntesis total y elucidación de los Peyssoneninos y análogos, Inhibidores de DNA metiltransferasa 1 (DNMT1), Tesis de Doctorado, Universidad de Vigo, España, 2012.

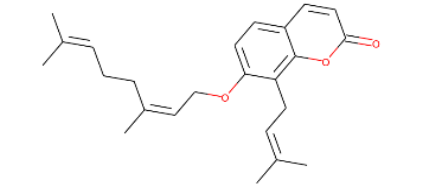
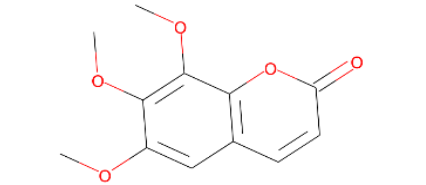
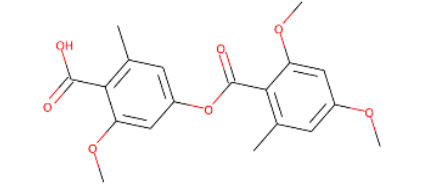
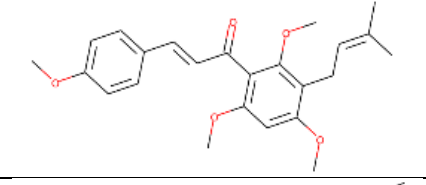
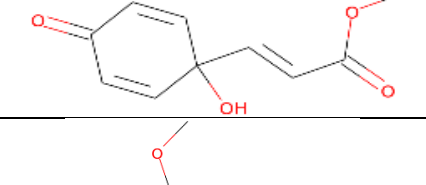
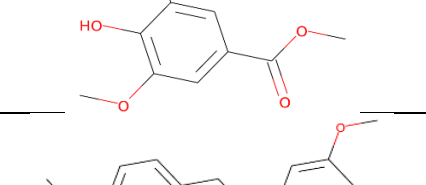
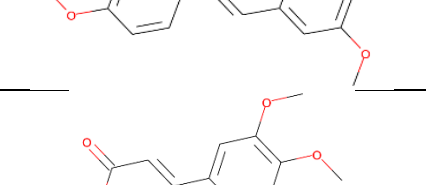
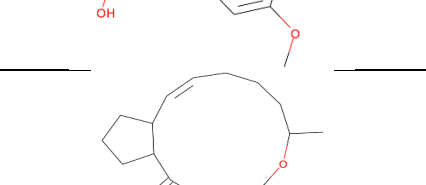
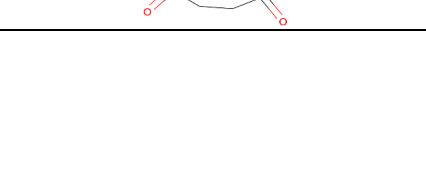
12. Saldívar, F.I.; Gómez, A.; Chávez, D.E.; Sánchez, N.; Ruiz, J.; Pílon, B.A. Medina, J.L. Inhibitors of DNA methyltransferases from natural sources: A computational perspective. *Front. Pharmacol.*, 2018, 9, 1144.
13. Prieto, F.; Peña, A.; Méndez, O.; Fernández, E.; Medina, J.L. Molecular modeling and chemoinformatics to advance the development of modulators of epigenetic Targets: A focus on DNA methyltransferases. *Adv Prot. Chem. Struct. Biol.* (2016) 105, 1-26.
14. Sánchez, N.; Medina, J.L. Epigenetic Target Profiler: A web server to predict epigenetic targets of small molecules. *J. Chem. Inf. Model.* (2021) 61 (4), 1550-1554.
15. Sánchez, N.; Medina, J.L. Epigenetic target fishing with accurate machine learning models. *J. Chem. Inf. Model* (2021) 64 (12), 8208-8220.
16. Prieto, F.D; Arciniega, M; Medina, J.L. Molecular docking: current advances and challenges, *TIP, Rev. Espec. Cienc. Quim. - Biol.* (2018) 21 (Supl. 1): 65-87.
17. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular docking and structure-based drug design strategies. *Molecules* (2015) 20, 13384-13421.
18. AutoDock Vina, Consultado el 15 de Agosto de 2021 en: <https://vina.scripps.edu/>.
19. Sánchez.N.;Medina, J.L.; Mestres, J.; Barri, X. Extended connectivity interaction features: improving binding affinity prediction through chemical description, *Bioinformatics* (2021) 37, 1376–1382.
20. Pílon, B. A., Saldívar, F. I., Díaz, B. I., & Medina, J. L. BIOFACQUIM: A Mexican compound database of natural products. *Biomolecules*, (2019) 9(1), 31.
21. Saldívar F.I; Valli M; Andricopulo A.D; da Bolzani Vd; Medina-Franco JL. Chemical space and diversity of NuBBE Database: A chemoinformatic characterization. *J. Chem. Inf. Model.* (2019) 59, 74–85.
22. González, M.; Prieto, F.D.; Naveja, J.; Méndez, O.; El-Elimat. T.; Pearce, C.; Oberlies, N.; Figueroa, M.; Medina, J.L. Chemoinformatic expedition of the chemical space of fungal products *Future Med. Chem.* (2016) 8:12, 1399-1412.
23. Pye, C.R.; Bertin, M.J.; Lokey, R.S.; Gerwick, W.H.; Linington, R. G; Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U. S. A.* (2017) 114 (22) 5601-5606.

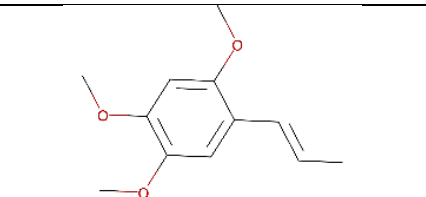
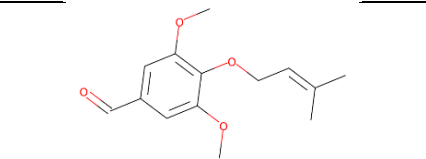
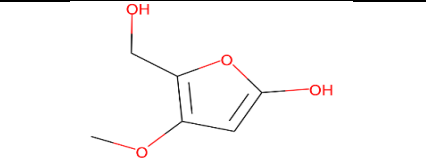
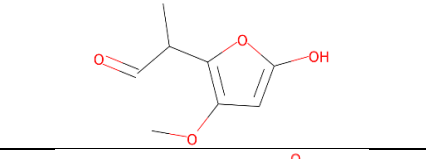
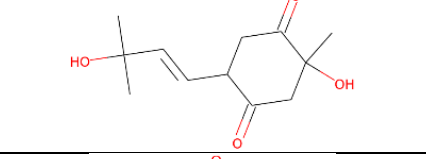
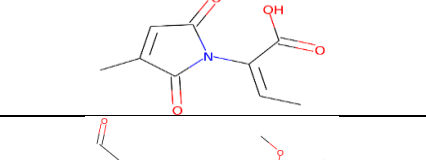
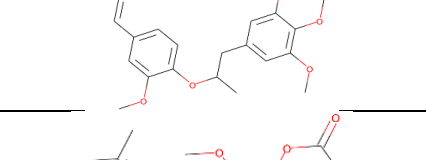
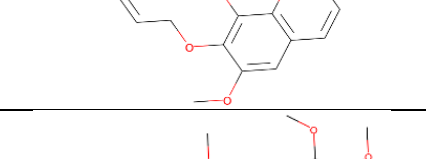
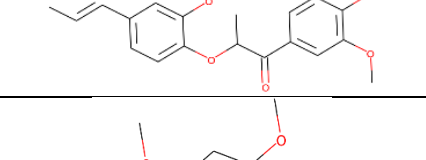
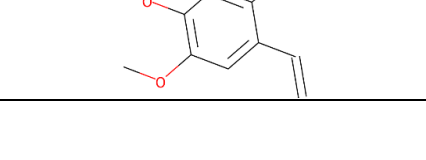
24. Screening Libraries and Collections, Edible plant material: Compounds, Fractions, Extracts, MEGx, Consultado el 15 de enero de 2021 en: <https://ac-discovery.com/screening-libraries/>.
25. Sorokina, M., Merseburger, P., Rajan, K. COCONUT online: Collection of Open Natural Products database. *J. Cheminf.* (2021). 13, 2.
26. Chávez, A.L.; Sánchez, N.; Medina, J.L.; Fragment library of natural products and compound databases for drug discovery. *Biomolecules* (2020) 10 (11), 1518 .
27. Sustancia CHEMBL3087498, Consultado el 25 de Febrero de 2022 en: <https://pubchem.ncbi.nlm.nih.gov/substance/194165848#section=Modify-Date> .
28. AW1, *Protein Data Bank Japan*, Consultado el 25 de Febrero de 2022 en: <https://pdj.org/chemie/summary/AW1>.
29. Compuesto Psammaplin A, Consultado el 25 de Febrero de 2022 en: https://pubchem.ncbi.nlm.nih.gov/compound/Psammaplin_A.
30. Compuesto 6-methoxy-7-[(2-methyl-2-azaspiro[3.3]heptan-6-yl)oxy]-2-(5-methylfuran-2-yl)-N-(1-methylpiperidin-4-yl)quinolin-4-amine, Consultado el 25 de Febrero de 2022 en: <https://pubchem.ncbi.nlm.nih.gov/compound/129072680>.
31. 28 Decitabina, Consultado el 25 de Febrero de 2022 en: <https://pubchem.ncbi.nlm.nih.gov/compound/Decitabine>.

ANEXOS

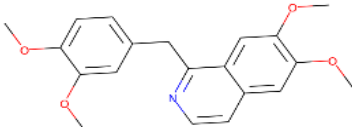
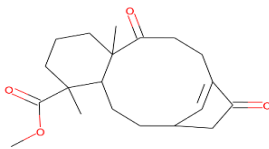
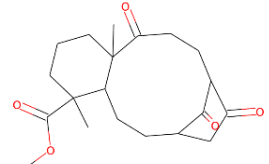
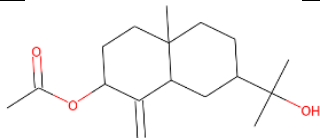
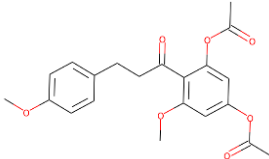
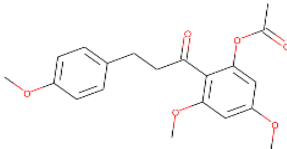
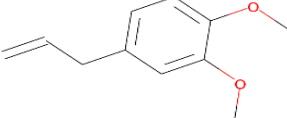
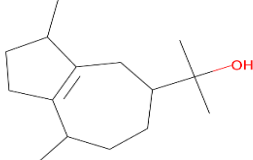
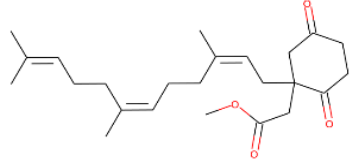
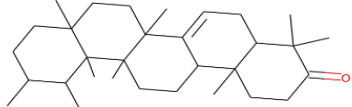
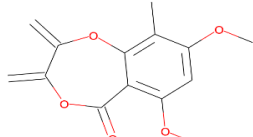
Tabla A1. Compuestos selectivos identificados en las bases de datos BIOFACQUIM, CYANOBACTERIA, FUNGI, MARINES, MEGx y NuBBE. Se observan en la tabla el identificador del compuesto, diana epigenética, base de datos a la que pertenece, estructura molecular procesada por ETP, status referente a si el resultado del compuesto es conocido o predicción y cuartil de acuerdo a la distancia de Jaccard para los datos de ETP.

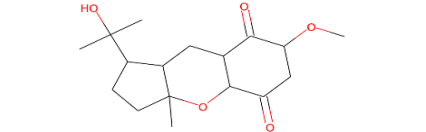
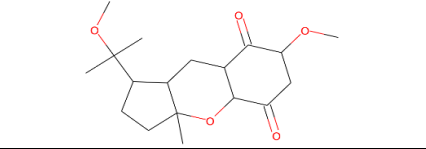
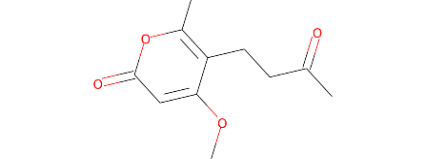
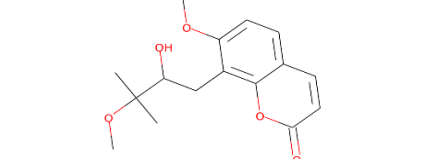
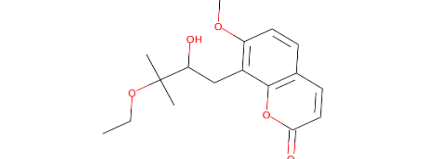
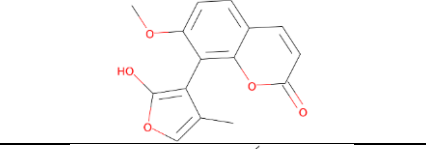
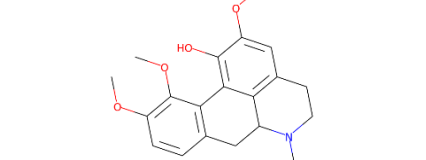
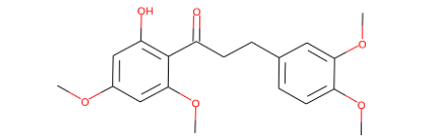
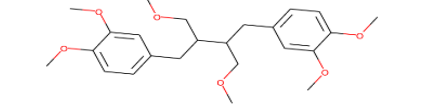
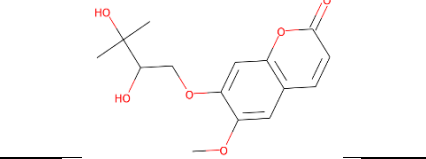
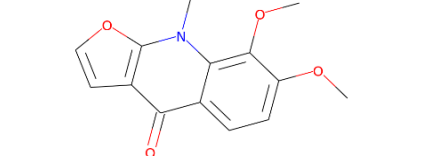
ID	Diana	Base de datos	Estructura (procesada)	Status	Cuartil
CYAN17	APEX1	CYANOBACTERIAS		PREDICTED	Q4
CYAN19	APEX1	CYANOBACTERIAS		PREDICTED	Q4
M1019	APEX1	MARINES		PREDICTED	Q4
M143	APEX1	MARINES		PREDICTED	Q4
M207	APEX1	MARINES		PREDICTED	Q4
M229	APEX1	MARINES		PREDICTED	Q4
M3030	APEX1	MARINES		PREDICTED	Q4
M338	APEX1	MARINES		PREDICTED	Q4
M473	APEX1	MARINES		PREDICTED	Q4
M697	APEX1	MARINES		PREDICTED	Q4
M923	APEX1	MARINES		PREDICTED	Q4

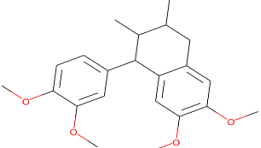
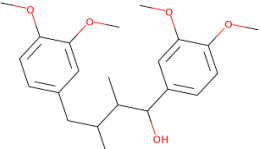
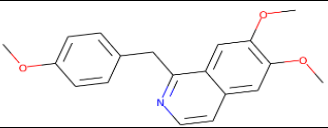
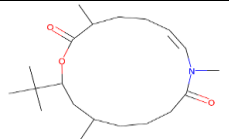
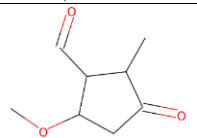
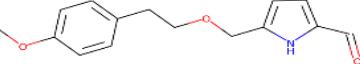
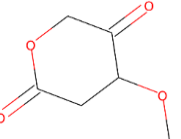
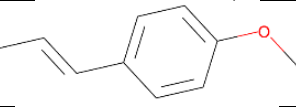
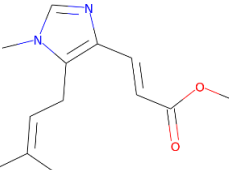
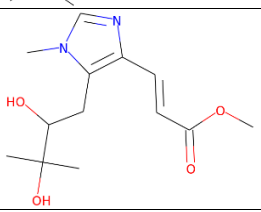
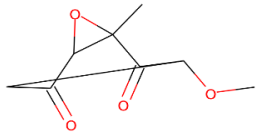
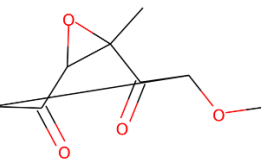
MEGx1157	APEX1	MEGx		PREDICTED	Q4
MEGx1261	APEX1	MEGx		PREDICTED	Q4
MEGx1310	APEX1	MEGx		PREDICTED	Q4
MEGx1336	APEX1	MEGx		PREDICTED	Q4
MEGx191	APEX1	MEGx		PREDICTED	Q4
MEGx363	APEX1	MEGx		PREDICTED	Q4
MEGx372	APEX1	MEGx		PREDICTED	Q4
MEGx384	APEX1	MEGx		PREDICTED	Q4
MEGx389	APEX1	MEGx		PREDICTED	Q4

MEGx461	APEX1	MEGx		PREDICTED	Q4
MEGx465	APEX1	MEGx		PREDICTED	Q4
MEGx510	APEX1	MEGx		PREDICTED	Q4
MEGx541	APEX1	MEGx		PREDICTED	Q4
MEGx549	APEX1	MEGx		PREDICTED	Q4
MEGx616	APEX1	MEGx		PREDICTED	Q4
MEGx806	APEX1	MEGx		PREDICTED	Q4
NuBBE1093	APEX1	NuBBE		PREDICTED	Q4
NuBBE1256	APEX1	NuBBE		PREDICTED	Q4
NuBBE1261	APEX1	NuBBE		PREDICTED	Q4

NuBBE127	APEX1	NuBBE		PREDICTED	Q4
NuBBE1414	APEX1	NuBBE		PREDICTED	Q4
NuBBE1454	APEX1	NuBBE		PREDICTED	Q4
NuBBE151	APEX1	NuBBE		PREDICTED	Q4
NuBBE1591	APEX1	NuBBE		PREDICTED	Q4
NuBBE1669	APEX1	NuBBE		PREDICTED	Q4
NuBBE1735	APEX1	NuBBE		PREDICTED	Q4
NuBBE267	APEX1	NuBBE		PREDICTED	Q4
M3402	ATM	MARINES		PREDICTED	Q4
M4208	ATM	MARINES		PREDICTED	Q4

MEGx1198	ATM	MEGx		PREDICTED	Q4
MEGx1830	ATM	MEGx		PREDICTED	Q4
MEGx2029	ATM	MEGx		PREDICTED	Q4
MEGx856	ATM	MEGx		PREDICTED	Q4
NuBBE1069	ATM	NuBBE		PREDICTED	Q4
NuBBE1071	ATM	NuBBE		PREDICTED	Q4
NuBBE1901	ATM	NuBBE		PREDICTED	Q4
NuBBE372	ATM	NuBBE		PREDICTED	Q4
NuBBE638	ATM	NuBBE		PREDICTED	Q4
NuBBE854	ATM	NuBBE		PREDICTED	Q4
M1867	BRD2	MARINES		PREDICTED	Q4

M2830	BRD2	MARINES		PREDICTED	Q4
M2994	BRD2	MARINES		PREDICTED	Q4
M591	BRD2	MARINES		PREDICTED	Q4
MEGx1231	BRD2	MEGx		PREDICTED	Q4
MEGx1253	BRD2	MEGx		PREDICTED	Q4
MEGx2032	BRD2	MEGx		PREDICTED	Q4
MEGx3173	BRD2	MEGx		PREDICTED	Q4
MEGx755	BRD2	MEGx		PREDICTED	Q4
MEGx921	BRD2	MEGx		PREDICTED	Q4
NuBBE1092	BRD2	NuBBE		PREDICTED	Q4
NuBBE113	BRD2	NuBBE		PREDICTED	Q4

NuBBE669	BRD2	NuBBE		PREDICTED	Q4
NuBBE681	BRD2	NuBBE		PREDICTED	Q4
NuBBE1443	CARM1	NuBBE		PREDICTED	Q4
CYAN31	CDK1	CYANOBACTERIAS		PREDICTED	Q4
M1089	CDK1	MARINES		PREDICTED	Q4
M216	CDK1	MARINES		PREDICTED	Q4
MEGx493	CDK1	MEGx		PREDICTED	Q4
NuBBE1015	CDK1	NuBBE		PREDICTED	Q4
MEGx365	HDAC3	MEGx		PREDICTED	Q4
MEGx496	HDAC3	MEGx		PREDICTED	Q4
FQNP44	HDAC8	BIOFACQUIM		PREDICTED	Q4
FQNP7	HDAC8	BIOFACQUIM		PREDICTED	Q4

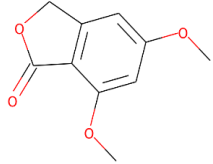
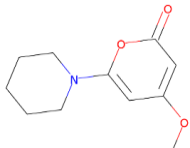
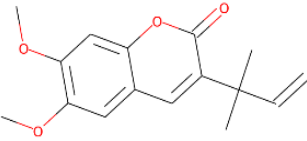
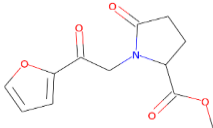
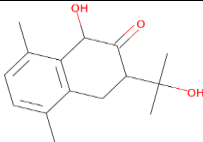
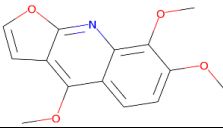
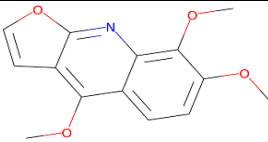
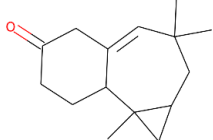
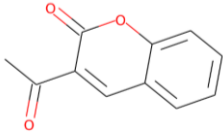
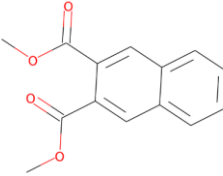
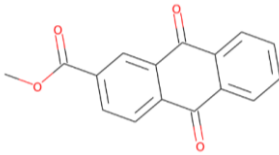
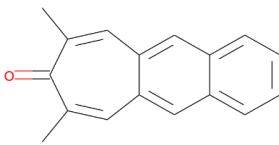
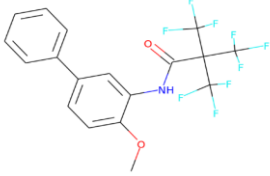
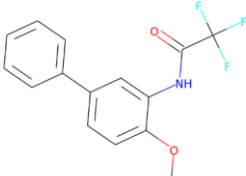
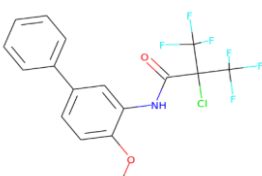
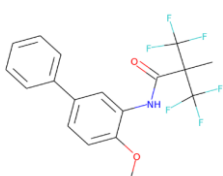
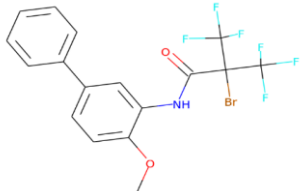
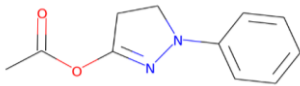
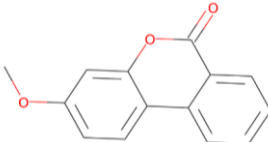
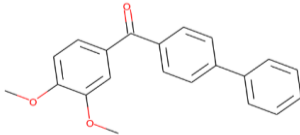
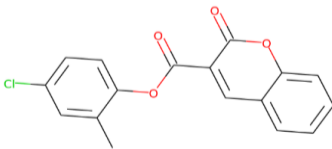
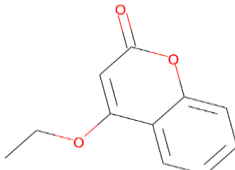
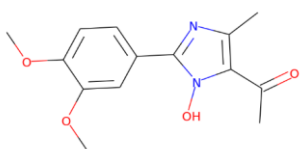
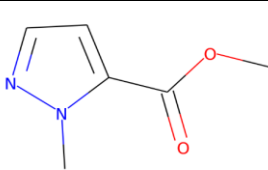
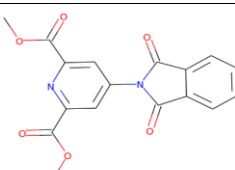
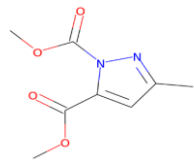
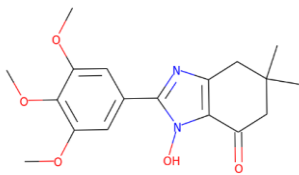
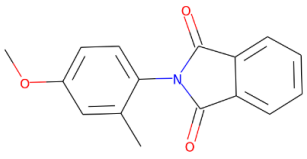
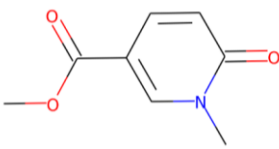
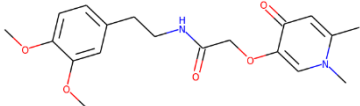
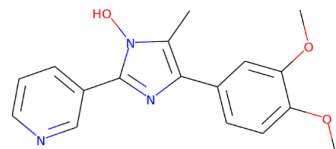
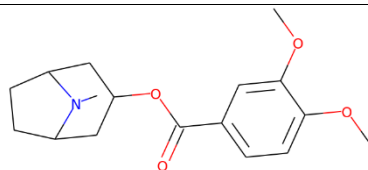
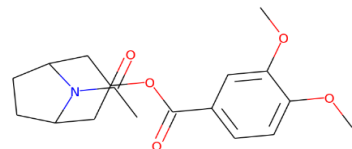
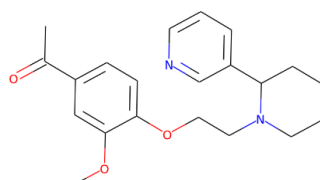
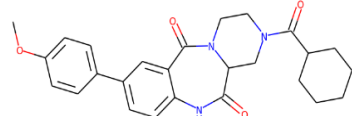
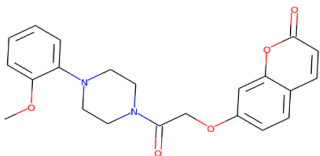
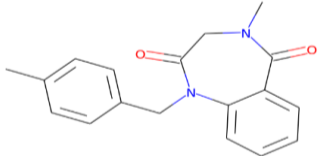
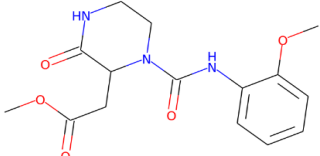
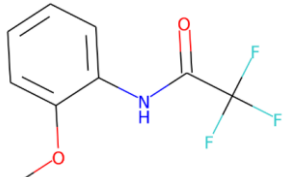
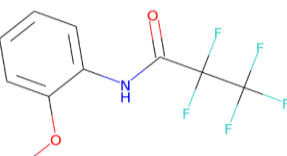
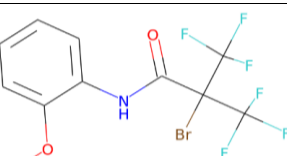
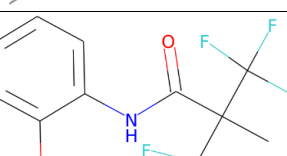
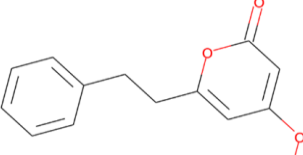
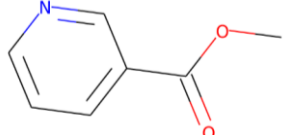
FUNGI 25	HDAC8	FUNGI		PREDICTED	Q4
NuBBE1277	HDAC8	NuBBE		PREDICTED	Q4
MEGx1080	KDM4E	MEGx		PREDICTED	Q4
M864	PARP1	MARINES		PREDICTED	Q3
NuBBE1290	PRKCD	NuBBE		PREDICTED	Q4
MEGx1800	PRKDC	MEGx		PREDICTED	Q4
NuBBE1819	PRKDC	NuBBE		PREDICTED	Q4
M1307	RPS6K A5	MARINES		PREDICTED	Q4

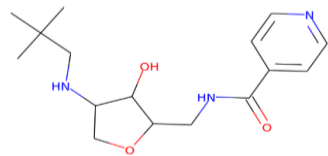
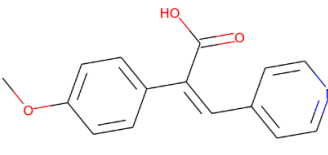
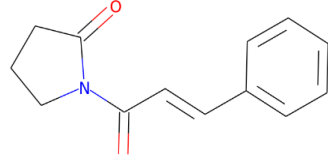
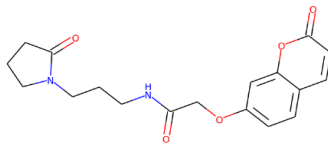
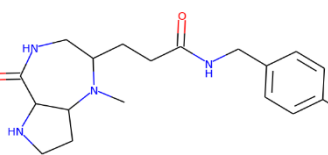
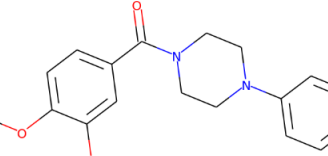
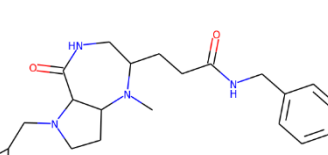
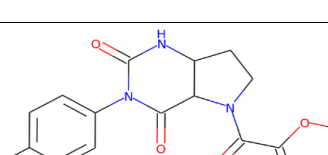
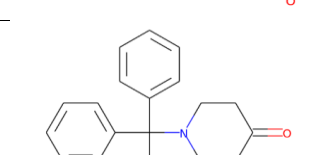
Tabla A2. Compuestos selectivos identificados en la base de datos COCONUT. Se observan en la tabla el identificador del compuesto, diana epigenética, base de datos a la que pertenece, estructura molecular procesada por ETP, status referente a si el resultado del compuesto es conocido o predicción y cuartil de acuerdo a la distancia de Jaccard para los datos de ETP.

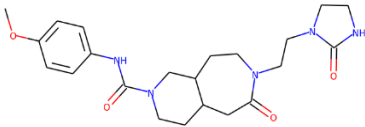
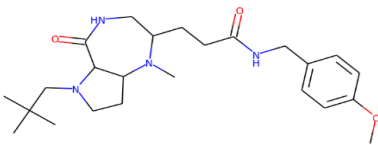
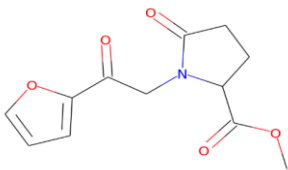
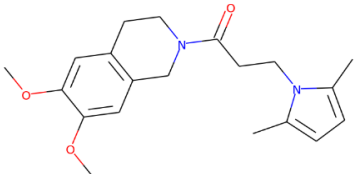
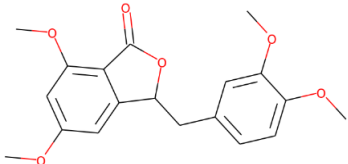
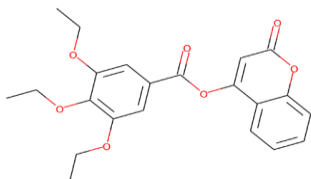
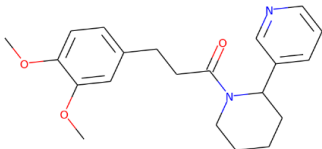
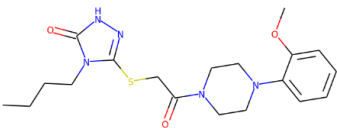
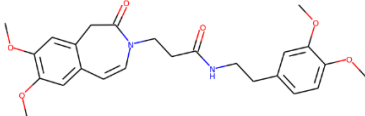
ID	Diana	Base de datos	Estructura (procesada)	Status	Cuartil
CNP0045457	APEX1	COCONUT		Predicted	Q3
CNP0065497	APEX1	COCONUT		Predicted	Q3
CNP0384131	APEX1	COCONUT		Predicted	Q3
CNP0401439	APEX1	COCONUT		Predicted	Q3
CNP0030052	ATM	COCONUT		Predicted	Q3
CNP0053115	ATM	COCONUT		Predicted	Q3
CNP0062872	ATM	COCONUT		Predicted	Q3
CNP0097363	ATM	COCONUT		Predicted	Q3

CNP0137146	ATM	COCONUT		Predicted	Q3
CNP0150521	ATM	COCONUT		Predicted	Q3
CNP0229638	ATM	COCONUT		Predicted	Q3
CNP0337786	ATM	COCONUT		Predicted	Q3
CNP0369055	ATM	COCONUT		Predicted	Q3
CNP0400045	ATM	COCONUT		Predicted	Q3
CNP0009801	BRD2	COCONUT		Predicted	Q2
CNP0027219	BRD2	COCONUT		Predicted	Q3
CNP0051704	BRD2	COCONUT		Predicted	Q3

CNP0057200	BRD2	COCONUT		Predicted	Q3
CNP0145859	BRD2	COCONUT		Predicted	Q3
CNP0193945	BRD2	COCONUT		Predicted	Q3
CNP0334743	BRD2	COCONUT		Predicted	Q3
CNP0399848	BRD2	COCONUT		Predicted	Q3
CNP0413560	BRD2	COCONUT		Predicted	Q3
CNP0186395	BRD4	COCONUT		Predicted	Q3
CNP0309475	BRD4	COCONUT		Predicted	Q3
CNP0060958	CARM1	COCONUT		Predicted	Q3
CNP0397356	CARM1	COCONUT		Predicted	Q3
CNP0046220	CREBBP	COCONUT		Predicted	Q3

CNP0119183	CREBBP	COCONUT		Predicted	Q3
CNP0392928	EP300	COCONUT		Predicted	Q3
CNP0371075	HDAC2	COCONUT		Predicted	Q3
CNP0064789	HDAC3	COCONUT		Predicted	Q3
CNP0195938	HDAC3	COCONUT		Predicted	Q3
CNP0213989	HDAC3	COCONUT		Predicted	Q3
CNP0220666	HDAC3	COCONUT		Predicted	Q3
CNP0191534	HDAC8	COCONUT		Predicted	Q3
CNP0284518	KDM4A	COCONUT		Predicted	Q2

CNP0065187	KDM4E	COCONUT		Predicted	Q3
CNP0065190	KDM4E	COCONUT		Predicted	Q3
CNP0277712	KDM4E	COCONUT		Predicted	Q3
CNP0313542	KDM4E	COCONUT		Predicted	Q3
CNP0015601	PARP1	COCONUT		Predicted	Q3
CNP0016880	PARP1	COCONUT		Predicted	Q3
CNP0041239	PARP1	COCONUT		Predicted	Q3
CNP0041398	PARP1	COCONUT		Predicted	Q3
CNP0096598	PARP1	COCONUT		Predicted	Q3

CNP0158456	PARP1	COCONUT		Predicted	Q3
CNP0174199	PARP1	COCONUT		Predicted	Q3
CNP0179342	PARP1	COCONUT		Predicted	Q3
CNP0185651	PARP1	COCONUT		Predicted	Q3
CNP0265651	PARP1	COCONUT		Predicted	Q3
CNP0287571	PARP1	COCONUT		Predicted	Q3
CNP0289509	PARP1	COCONUT		Predicted	Q3
CNP0307462	PARP1	COCONUT		Predicted	Q3
CNP0345397	PARP1	COCONUT		Predicted	Q2
CNP0408106	PARP1	COCONUT		Predicted	Q3

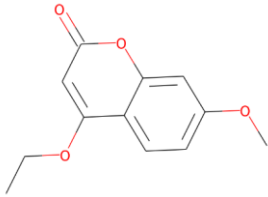
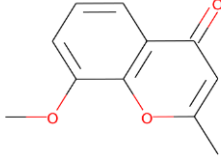
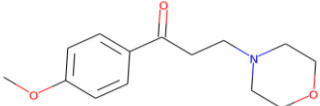
CNP0051033	PRKDC	COCONUT		Predicted	Q3
CNP0134517	PRKDC	COCONUT		Predicted	Q2
CNP0191840	PRKDC	COCONUT		Predicted	Q3

Tabla A3. Descripción de las funciones de la herramienta MolVS utilizadas en el proceso de curado de las bases de datos.

Proceso	Función de MolVS	Descripción
Estandarización	Standardizer	La función satandardizer_ smile proporciona una forma rápida y sencilla de obtener la versión estandarizada de una cadena SMILES.
Eliminación de metales	Standardizer (Disconnect Metals)	Se rompen los enlaces covalentes entre metales y átomos orgánicos en determinadas condiciones. Primero, se desconecta N, O, F de cualquier metal. Luego, se desconectan otros no metales de los metales de transición. Por cada enlace roto se ajustan las cargas de los átomos.
Retención del fragmento más grande	Largest Fragment Choser	El término fragmento se refiere a unidades dentro de una molécula unidas covalentemente. La función selecciona el fragmento más largo en cada molécula.
Neutralizar cargas	Uncharger / Reionizer	Uncharge: neutraliza la molécula añadiendo o eliminando hidrógenos. Reionizer: Aplica cargas a ciertos átomos y luego realiza una reionización competitiva. Primero se aplican correcciones de carga para garantizar, por ejemplo, que los metales libres estén correctamente ionizados. Luego, si una molécula con múltiples grupos ácidos está parcialmente ionizada, se debe asegurar de que los ácidos más fuertes se ionicen primero.
Generación del tautómero canónico	Tautomer Canonicalizer	Los tautómeros son conjuntos de moléculas que se interconvierten fácilmente entre sí mediante el movimiento de un átomo de hidrógeno. Los tautómeros tienen la misma fórmula molecular y carga neta, pero difieren en términos de las posiciones de los hidrógenos y los cambios asociados en los enlaces dobles y sencillos adyacentes. La función enumera todos los tautómeros posibles utilizando reglas de transformación y se utiliza un sistema de puntuación para determinar el tautómero canónico.