



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral:

**BÚSQUEDA ASISTIDA POR COMPUTADORA
DE NUEVOS FÁRMACOS ANTICHAGÁSICOS
ANÁLOGOS DE POLIAMINAS**

Tesista: Lic. Lucas Nicolás Alberca

Director/a: Dr. Alan Talevi

Codirector/a: Dra. Carolina Carrillo

Año: 2018



Universidad Nacional de La Plata

Facultad de Ciencias Exactas

Departamento de Ciencias Biológicas



Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB)

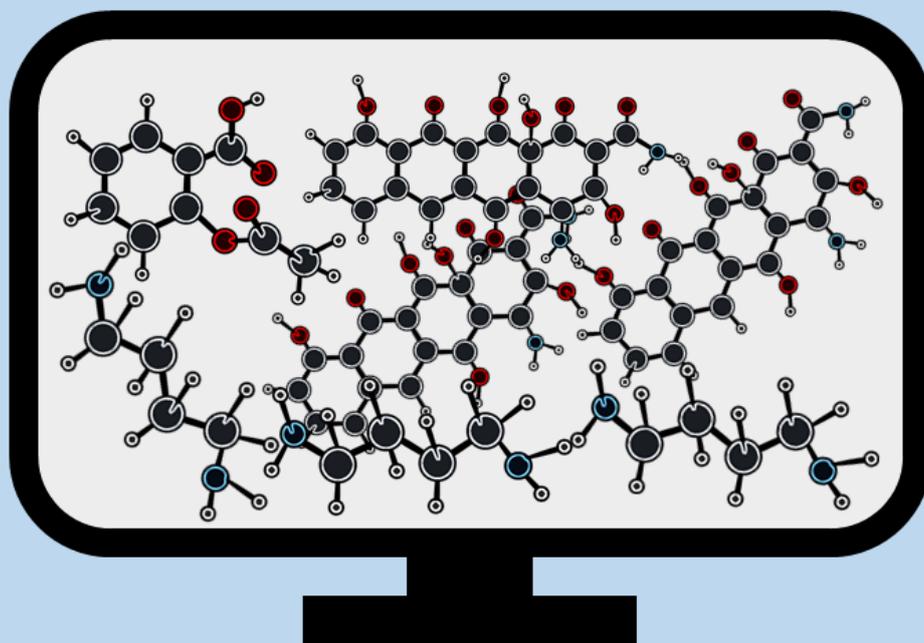
Trabajo de tesis doctoral:

BÚSQUEDA ASISTIDA POR COMPUTADORA DE NUEVOS FÁRMACOS ANTICHAGÁSICOS ANÁLOGOS DE POLIAMINAS

Tesista: Alberca, Lucas Nicolás

Director: Talevi, Alan

Codirectora: Carrillo, Carolina



AGRADECIMIENTOS

A **mi director, Dr. Alan Talevi**, porque me permitió ser parte de su grupo de trabajo y es con quien dí mis primeros pasos en el ambiente científico. Sus enseñanzas, su energía positiva, su motivación, sus consejos y su amistad fueron fundamentales para la realización de esta tesis.

A **mi codirectora, Dra. Carolina Carrillo**, por su disposición, sus consejos, su compromiso y su experiencia que fue de mucha importancia para la realización de los ensayos experimentales y la escritura de esta tesis.

Al **Dr. Luis Bruno Blanch**, por sus consejos y por la transmisión de su pasión por la investigación.

A **mis compañeros del LIDeB**, por darme una mano siempre que los necesité, por hacer del LIDeB un lugar muy agradable, y por la gran cantidad de momentos compartidos.

A la **Dra. Esperanza Ruiz** por su colaboración en los análisis estadísticos.

A la **Dra. María Laura Sbaraglini** y a **Daniela Ruiz**, por su gran ayuda y su paciencia en la realización de los ensayos experimentales, y especialmente, por la buena energía que transmiten.

A la **Dra. Catalina Alba Soto**, por poner a disposición su laboratorio para la realización de los ensayos experimentales y su colaboración en los mismos.

A la **Dra. Carolina Bellera**, por su gran ayuda durante la realización de mi tesina de grado, que fue fundamental para el comienzo de esta tesis doctoral.

A **mis amigas y amigos de Tres Arroyos** que me acompañan en todo momento a pesar de la distancia. En especial a Mati, Martu, Nicole y Emma que me han hecho parte de su familia.

A **mis amigas y amigos que conocí en La Plata** con quienes viví, aprendí, me divertí y compartí, y sigo compartiendo, increíbles momentos. En especial a Fran, Joni, Feli, Eri, Clari, Ceci, Mati, Tanke, Kelo y Jero.

A mis primos, Carli, Diego y Andrés, por el cariño que me brindaron en La Plata.

A **mi mamá, mi papá, mi hermano, Gaby, Benja, mis abuelos y el resto de mi familia**, por preocuparse por mí, por apoyarme en todo, por sus enseñanzas y consejos, y por estar siempre presentes.

A mi novia, **Viki**, por su apoyo, sus consejos, su paciencia, sus abrazos y su compañía por tantos años, y por hacerme tan feliz.

A la **Facultad de Ciencias Exactas, UNLP y CONICET**, por los medios para realizar esta tesis.

Al **pueblo argentino**, por financiar a través de sus impuestos mi educación primaria, secundaria, universitaria y de posgrado.

Muchas gracias a todos.

ÍNDICE

Objetivos	Pág. 1
Capítulo 1: INTRODUCCIÓN	Pág. 2
1.1. ENFERMEDADES TROPICALES DESATENDIDAS	Pág. 3
1.2. ENFERMEDAD DE CHAGAS	Pág. 4
1.2.1. ASPECTOS GENERALES	Pág. 4
1.2.2. BREVE REPASO HISTÓRICO	Pág. 5
1.2.3. <i>Trypanosoma cruzi</i>	Pág. 5
1.2.4. VÍAS DE TRANSMISIÓN	Pág. 6
1.2.5. CICLO DE VIDA DEL PARÁSITO	Pág. 7
1.2.6. EPIDEMIOLOGÍA Y DISTRIBUCIÓN	Pág. 10
1.2.7. MANIFESTACIONES CLÍNICAS DE LA ENFERMEDAD	Pág. 11
1.2.8. MEDICAMENTOS DISPONIBLES	Pág. 11
1.2.9. AVANCES HACIA NUEVOS MEDICAMENTOS	Pág. 13
1.3. BLANCOS MOLECULARES EN <i>T. cruzi</i>	Pág. 14
1.3.1. INHIBIDORES DE LA SÍNTESIS DE ESTEROLES	Pág. 15
1.3.2. CRUZIPAÍNA	Pág. 16
1.3.3. TRANS –SIALIDASA	Pág. 16
1.3.4. INHIBIDORES DE LA VÍA GLUCOLÍTICA	Pág. 16
1.3.5. INHIBIDORES DE LA VÍA DE LAS PENTOSAS FOSFATO	Pág. 17
1.3.6. INHIBIDORES DE LA SÍNTESIS Y METABOLISMO DE NUCLEÓTIDOS Y DE LA RECUPERACIÓN DE PURINAS	Pág. 17
1.3.7. AGLUTINANTES DE ADN	Pág. 17
1.3.8. INHIBIDORES DEL SISTEMA DE DEFENSA CONTRA EL ESTRÉS OXIDATIVO	Pág. 18
1.4. POLIAMINAS	Pág. 18
1.4.1. CARACTERÍSTICAS GENERALES	Pág. 18
1.4.2. METABOLISMO DE POLIAMINAS	Pág. 19
1.4.3. TRIPANOTIÓN	Pág. 21
1.4.4. TRANSPORTADORES DE POLIAMINAS	Pág. 21
1.5. ELECCIÓN DEL BLANCO MOLECULAR	Pág. 22
REFERENCIAS CAPÍTULO 1	Pág. 23
Capítulo 2: DESCUBRIMIENTO DE FÁRMACOS	Pág. 31
2.1. INTRODUCCIÓN	Pág. 32
2.1.1. DESCUBRIMIENTO DE FÁRMACOS PARA ETDs	Pág. 33
2.2. CRIBADO VIRTUAL	Pág. 35
2.2.1. MÉTODOS BASADOS EN EL BLANCO MOLECULAR	Pág. 37
2.2.1.1. DOCKING MOLECULAR	Pág. 38
2.2.2. MÉTODOS BASADOS EN EL LIGANDO	Pág. 39
2.2.2.1. FARMACÓFOROS	Pág. 39

2.2.2.2. SIMILITUD (O SIMILARIDAD) MOLECULAR	Pág. 40
2.2.2.3. DESCRIPTORES MOLECULARES	Pág. 41
2.2.3. MÉTODOS COMBINADOS	Pág. 43
2.3. REPOSICIONAMIENTO DE FÁRMACOS	Pág. 44
2.3.1. GENERALIDADES	Pág. 44
2.3.2. MÉTODOS SISTEMÁTICOS DE REPOSICIONAMIENTO DE FÁRMACOS	Pág. 46
2.3.2.1. ENFOQUES NO COMPUTACIONALES	Pág. 47
2.3.2.2. ENFOQUES COMPUTACIONALES	Pág. 48
2.3.2.2.1. METODOLOGÍAS QUIMIOINFORMÁTICAS	Pág. 48
2.3.2.2.2. METODOLOGÍAS BIOINFORMÁTICAS	Pág. 48
2.3.2.2.3. METODOLOGÍAS BASADAS EN LA LITERATURA	Pág. 49
2.3.3. REPOSICIONAMIENTO DE FÁRMACOS PARA ETDs	Pág. 50
2.3.3.1. CASOS PROMETEDORES Y EXITOSOS EN TRIPANOSOMÁTIDOS	Pág. 50
2.4. ESTRATEGIA UTILIZADA EN ESTE TRABAJO DE TESIS	Pág. 51
REFERENCIAS CAPÍTULO 2	Pág. 52

Capítulo 3: PRIMERA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN

EXPERIMENTAL. METODOLOGÍA	Pág. 58
3.1. INTRODUCCIÓN	Pág. 59
3.2. OBTENCIÓN DE MODELOS QSAR (FUNCIONES DISCRIMINANTES)	Pág. 60
3.2.1. COMPILADO DE LA BASE DE DATOS	Pág. 60
3.2.2. PARTICIÓN DE LA BASE DE DATOS	Pág. 61
3.2.2.1. ESTRATEGIA DE PARTICIÓN REALIZADA EN ESTE TRABAJO	Pág. 62
3.2.3. CÁLCULO DE DESCRIPTORES MOLECULARES	Pág. 63
3.2.4. MODELADO	Pág. 64
3.2.4.1. SUBESPACIOS ALEATORIOS	Pág. 64
3.2.4.2. SELECCIÓN DE DESCRIPTORES MOLECULARES Y GENERACIÓN DE LOS MODELOS CLASIFICATORIOS	Pág. 64
3.3. VALIDACIÓN DE LOS MODELOS	Pág. 66
3.3.1. MÉTODOS DE VALIDACIÓN INTERNA	Pág. 66
3.3.2. VALIDACIÓN EXTERNA	Pág. 67
3.3.3. EVALUACIÓN DEL DESEMPEÑO DE LOS MODELOS	Pág. 67
3.3.4. CAMPAÑAS SIMULADAS DE SV	Pág. 68
3.3.4.1. BIBLIOTECA PILOTO "SIMULADA"	Pág. 69
3.3.4.2. BIBLIOTECA PILOTO DUD-E	Pág. 69
3.4. COMBINACIÓN DE MODELOS	Pág. 70
3.5. SELECCIÓN DE VALOR DE CORTE PARA EL CRIBADO	Pág. 71
3.6. CRIBADO VIRTUAL	Pág. 71
3.7. DOMINIO DE APLICACIÓN	Pág. 71
3.8. ENSAYOS EXPERIMENTALES	Pág. 72
3.8.1. ACTIVIDAD BIOLÓGICA FRENTE A TRIPANOSOMÁTIDOS	Pág. 72
3.8.1.1. ENSAYOS DE VIABILIDAD – PROLIFERACIÓN DE EPIMASTIGOTES DE <i>T. cruzi</i>	Pág. 73
3.8.1.2. ENSAYOS DE VIABILIDAD – PROLIFERACIÓN DE TRIPOMASTIGOTES DE SANGUÍNEOS de <i>T. brucei brucei</i>	Pág. 73

3.8.1.3.	ENSAYOS DE VIABILIDAD – PROLIFERACIÓN DE PROMASTIGOTES DE <i>L. infantum</i>	Pág. 73
3.8.1.4.	ENSAYOS DE VIABILIDAD DE TRIPOMASTIGOTES SANGUÍNEOS DE <i>T. cruzi</i>	Pág. 74
3.8.2.	EVALUACIÓN DEL TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS	Pág. 74
3.8.3.	EVALUACIÓN DE LA ACTIVIDAD DE LA TRIPANOTIÓN SINTETASA	Pág. 75
3.8.4.	MODELO DE INFECCIÓN AGUDA EN ANIMALES	Pág. 75
	REFERENCIAS CAPÍTULO 3	Pág. 77

Capítulo 4: PRIMERA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN

	EXPERIMENTAL. RESULTADOS	Pág. 80
4.1.	COMPILACIÓN DE LA BASE DE DATOS	Pág. 81
4.2.	PARTICIÓN DE LA BASE DE DATOS	Pág. 81
4.3.	GENERACIÓN Y VALIDACIÓN DE MODELOS	Pág. 83
4.4.	SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO	Pág. 88
4.5.	CRIBADO VIRTUAL	Pág. 89
4.6.	EVALUACIÓN BIOLÓGICA	Pág. 90
4.6.1.	DETERMINACIÓN DE LA ACTIVIDAD TRIPANOCIDA	Pág. 90
4.6.2.	ENSAYOS DE TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS	Pág. 92
4.6.3.	EFFECTO SOBRE LA TRIPANOTIÓN SINTETASA	Pág. 94
4.6.4.	MODELO MURINO DE INFECCIÓN AGUDA	Pág. 94
	REFERENCIAS CAPÍTULO 4	Pág. 96

Capítulo 5: SEGUNDA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN

	EXPERIMENTAL. METODOLOGÍA	Pág. 98
5.1.	OBTENCIÓN DE FUNCIONES DISCRIMINANTES	Pág. 99
5.1.1.	CURADO Y PARTICIÓN DE LA BASE DE DATOS	Pág. 99
5.1.2.	MODELADO	Pág. 100
5.2.	CAMPAÑA SIMULADA DE SV	Pág. 101
5.3.	COMBINACIÓN DE MODELOS	Pág. 101
5.4.	SELECCIÓN DE VALOR DE CORTE PARA EL CRIBADO	Pág. 101
5.5.	CRIBADO BASADO EN EL LIGANDO	Pág. 102
5.6.	CRIBADO BASADO EN EL BLANCO MOLECULAR	Pág. 103
5.7.	SELECCIÓN FINAL DE LOS CANDIDATOS	Pág. 104
5.8.	ENSAYOS EXPERIMENTALES	Pág. 104
5.8.1.	ACTIVIDAD BIOLÓGICA SOBRE <i>T. cruzi</i>	Pág. 104
5.8.2.	EVALUACIÓN DEL TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS	Pág. 104
	REFERENCIAS CAPÍTULO 5	Pág. 105

Capítulo 6: SEGUNDA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN

	EXPERIMENTAL. RESULTADOS	Pág. 107
--	---------------------------------	----------

6.1. COMPILACIÓN Y PARTICIÓN DE LA BASE DE DATOS.....	Pág. 108
6.2. GENERACIÓN Y VALIDACIÓN DE MODELOS BASADOS EN EL LIGANDO.....	Pág. 109
6.3. SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO.....	Pág. 112
6.4. CRIBADO VIRTUAL.....	Pág. 113
6.5. CRIBADO BASADO EN EL BLANCO MOLECULAR.....	Pág. 113
6.6. EVALUACIÓN BIOLÓGICA.....	Pág. 115
6.6.1. DETERMINACIÓN DE LA ACTIVIDAD ANTI- <i>T. cruzi</i>	Pág. 115
6.6.2. ENSAYOS DE TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS.....	Pág. 117
REFERENCIAS CAPÍTULO 6.....	Pág. 119
Capítulo 7: CONCLUSIONES	Pág. 120
7.1. CONCLUSIONES GENERALES.....	Pág. 121
7.2. COMPARACIÓN CON ESTUDIOS PREVIOS.....	Pág. 122
7.3. ENSAMBLADO DE MODELOS Y CAMPAÑAS RETROSPECTIVAS DE SV.....	Pág. 123
7.4. POSIBLE APLICACIÓN DE LOS CANDIDATOS AL REPOSICIONAMIENTO.....	Pág. 124
7.5. DETALLE DE PUBLICACIONES, CONGRESOS Y BECAS.....	Pág. 126
REFERENCIAS CAPÍTULO 6.....	Pág. 127
MATERIAL SUPLEMENTARIO	Pág. 129
Material Suplementario 1.....	Pág. 130
Material Suplementario 2.....	Pág. 132
Material Suplementario 3.....	Pág. 135
Material Suplementario 4.....	Pág. 165
Material Suplementario 5.....	Pág. 168
Material Suplementario 6.....	Pág. 172
Material Suplementario 7.....	Pág. 174
REFERENCIAS MATERIAL SUPLEMENTARIO.....	Pág. 178

OBJETIVO GENERAL

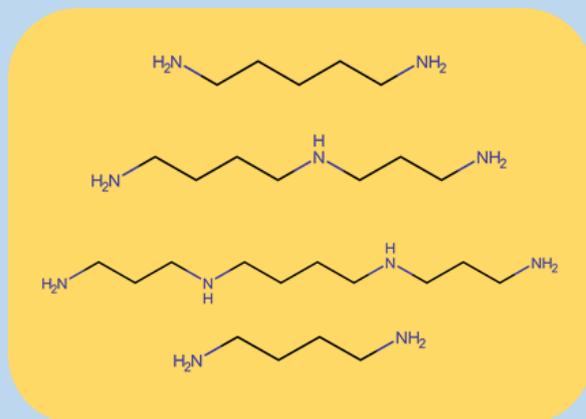
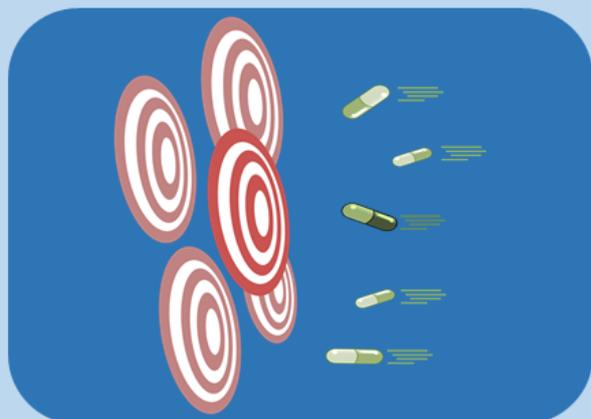
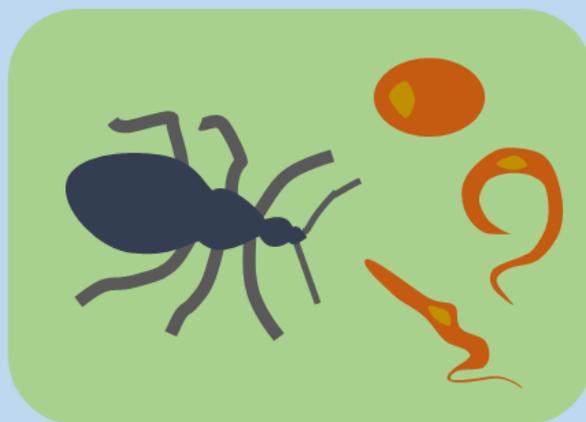
Desarrollar nuevos agentes terapéuticos aplicables en la farmacoterapia de la Enfermedad de Chagas. A tal fin se plantea desarrollar modelos computacionales capaces de identificar inhibidores del metabolismo y/o transporte de poliaminas en *Trypanosoma cruzi*. Se propone aplicar tales modelos en el cribado virtual (también conocido como tamizado o *screening virtual*) de grandes bases de datos de compuestos químicos. Se espera obtener nuevos fármacos que superen las limitaciones de los fármacos ya conocidos, tales como: falta de actividad en la fase crónica de la enfermedad y falta de selectividad.

OBJETIVO ESPECÍFICOS

- 1-** Compilar, a partir de literatura, una base de datos de análogos de poliaminas con actividad reportada sobre *T. cruzi*.
- 2-** Utilizar la base de datos compilada en el punto anterior para el desarrollo y validación de modelos computacionales basados en descriptores de baja dimensionalidad, capaces de identificar inhibidores del metabolismo y/o transporte de poliaminas en *T. cruzi*.
- 3-** Aplicar los modelos anteriores en una campaña de tamizado virtual de las bibliotecas químicas DrugBank y Sweetlead. En particular, la identificación de fármacos y drogas aprobadas por la Food and Drug Administration (FDA), permitirá detectar la actividad tripanocida en fármacos ya utilizados en clínica con otras indicaciones terapéuticas (estrategia conocida como *knowledge-based drug repurposing*, es decir, búsqueda de segundos usos médicos basada en el conocimiento).
- 4-** Adquisición y evaluación farmacológica, *in vitro*, de los candidatos seleccionados por los modelos generados.

CAPÍTULO 1

INTRODUCCIÓN



1.1. ENFERMEDADES TROPICALES DESATENDIDAS

La expresión *Enfermedades Tropicales Desatendidas* (ETDs) fue acuñada en 2003 por Hotez y colaboradores para expresar la atención política y económica que se le daba en ese entonces al HIV/SIDA, Malaria y, hasta cierto punto, Tuberculosis, en contraste con el resto de las enfermedades consideradas como objetivos prioritarios para la reducción sustentable de la pobreza por parte de la Millenium Declaration (Hotez 2011; Smith & Taylor 2016). La Organización Mundial de la Salud (OMS) propuso en 2005 englobar bajo ese rótulo a un grupo de enfermedades caracterizadas por la relativa falta de políticas, inversión e investigación en herramientas de control y prevención. En el caso de la búsqueda de nuevos tratamientos, este desinterés se debe en gran parte a que las ETDs afectan a poblaciones con limitados recursos económicos y por lo tanto las compañías farmacéuticas no encuentran un mercado favorable para el desarrollo de nuevas soluciones (World Health Organization 2010; Sbaraglini et al. 2016a).

La lista de ETDs incluye 17 enfermedades principalmente infecciosas (**Figura 1.1**) que comparten ciertas características (World Health Organization 2010):

- a) Constituyen un inconveniente importante para el desarrollo socioeconómico de los países en desarrollo, lo que provoca un aumento en la pobreza y una disminución en la calidad de vida de sus habitantes.
- b) Afectan principalmente a personas viviendo en situación de pobreza (con bajos ingresos y condiciones sanitarias inadecuadas) y personas viviendo en zonas rurales o poco pobladas.
- c) No representan una amenaza muy importante para países desarrollados ya que muchas de estas enfermedades no se propagan ampliamente y, en algunos casos, su distribución está restringida por el clima o la distribución geográfica de los vectores de transmisión (aunque, recientemente, este punto se ha modificado debido al fenómeno de globalización y las crecientes migraciones humanas).
- d) Pueden generar estigmatización debido a deformaciones y/o discapacidades propias de la enfermedad.
- e) Tienen alta tasa de mortalidad.
- f) Los diagnósticos y tratamientos disponibles son poco eficientes y/o insuficientes.
- g) Pueden ser controladas, prevenidas y posiblemente eliminadas si se tomaran las medidas necesarias.

Con respecto al último punto, la OMS indica 5 estrategias para el control y la prevención de las ETDs que deberían implementarse simultáneamente para lograr una solución eficiente: 1) utilización de quimioterapia preventiva; 2) intensificación del manejo de casos; 3) control vectorial; 4) provisión de agua segura, buenas condiciones de sanidad e higiene y; 5) incremento de la salud pública veterinaria (World Health Organization 2010). En el **Material Suplementario 1** se encuentra una breve descripción de las 17 ETDs de la OMS.

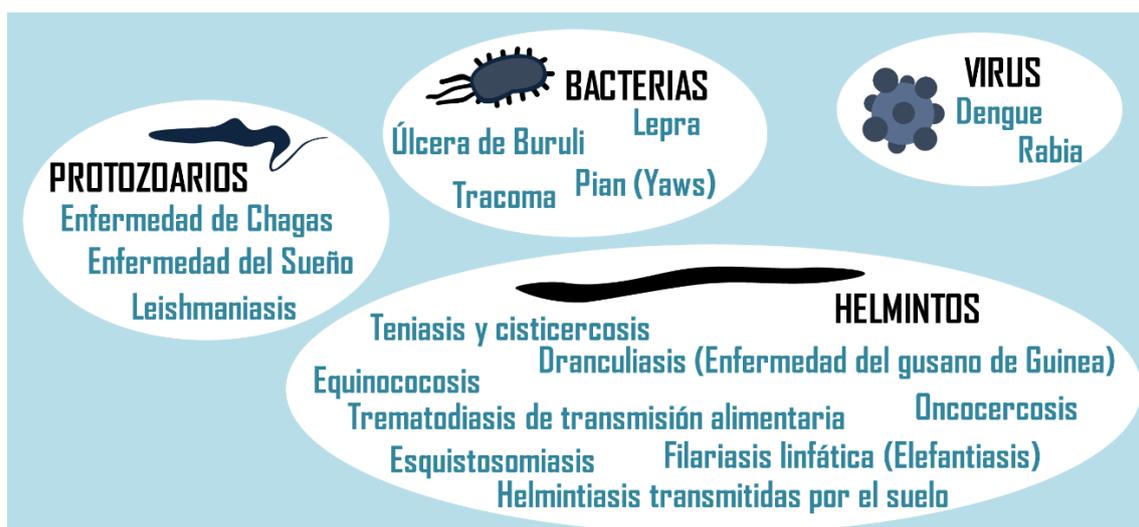


Figura 1.1: ETDs de la OMS.

Las ETDs, no solo producen mayores tasas de morbilidad, mortalidad, discapacidad y estigmas sociales, sino que imponen una gran carga social y económica en los grupos pobres y marginados que contribuyen a perpetuar el ciclo de la pobreza. La problemática de las ETDs abarca aspectos sociales, políticos y económicos que deberían trabajar conjuntamente para encontrar una solución adecuada. La vigilancia, el control, la atención rápida, el diagnóstico temprano, la vacunación, la educación son necesidades que deben atenderse y mantenerse para que las soluciones sean eficientes; por ello, es necesaria la contribución de diferentes organismos gubernamentales, no gubernamentales, laboratorios de investigación, compañías farmacéuticas, etc.

Esta tesis doctoral intenta aportar al conocimiento en cuanto a la búsqueda de nuevas quimioterapias para el tratamiento de la Enfermedad de Chagas, una de las ETDs más importantes en términos sanitarios, sociales y económicos en Latinoamérica.

1.2. ENFERMEDAD DE CHAGAS

1.2.1. ASPECTOS GENERALES

La Enfermedad de Chagas o Tripanosomiasis Americana es una parasitosis endémica de América Latina causada por el protozoario *Trypanosoma cruzi* (*T. cruzi*). La principal vía de transmisión de este parásito en zonas endémicas se produce a través de las heces del insecto vector, un hematófago de la subfamilia Triatominae, aunque existen otras vías de transmisión cuyo impacto puede ser muy alto en países no endémicos. Esta enfermedad ya no sólo afecta a poblaciones rurales de América Latina sino que debido a los movimientos migratorios de la población, la enfermedad se ha expandido a regiones urbanas y a países no endémicos y afecta a personas viviendo en ciudades tanto en Latinoamérica como en algunos países de Europa y Estados Unidos. Alrededor del 30% de las personas infectadas presentarán daños cardíacos y/o digestivos que podrían llevar a la muerte. Según datos de la OMS, se estima que existen al menos 6 millones de personas infectadas con *T. cruzi* en los 21 países de América Latina. Actualmente, existen sólo dos fármacos aprobados para el tratamiento de la enfermedad, Nifurtimox y Benznidazol, los cuales han demostrado una eficiencia inadecuada en adultos en fase crónica y además presentan efectos adversos importantes.

1.2.2. BREVE REPASO HISTÓRICO

Originalmente se creía que *T. cruzi* comenzó a infectar humanos hace 6000 años en la región andina de América cuando las personas comenzaron a adoptar el sedentarismo como forma de vida, con la domesticación de animales y la utilización de la agricultura. Según esta teoría, estos cambios provocaron un acercamiento del vector hacia los humanos, quienes se convirtieron en un nuevo hospedador para el parásito (Ferreira et al. 2011). Sin embargo, Aufderheide y colaboradores han evidenciado mediante PCR (reacción en cadena de la polimerasa), la presencia de ADN de *T. cruzi* en momias humanas pertenecientes al norte chileno y al sur peruano que datan del año 7050 A.C (Aufderheide et al. 2004). A partir de estos hallazgos ha tomado más peso la teoría de que la enfermedad es tan antigua como la llegada del hombre a América (Araújo et al. 2009; Ferreira et al. 2011).

En los años de la colonización, se han escrito numerosas crónicas y relatos de exploradores que hacen referencia a las picaduras de las vinchucas por las noches (Sanmartino 2009). Un poco más adelante en el tiempo, en 1835, Charles Darwin describía cómo durante su estadía en Mendoza unas chinches gigantes pertenecientes al género *Benchuca* picaban y succionaban la sangre de las personas. Este relato, sumado a algunos síntomas y a la causa de su muerte (insuficiencia cardíaca crónica) han colaborado a la teoría de que Darwin murió por una cardiopatía chagásica (Bernstein 1984).

A partir de 1909, y gracias a los estudios del Dr. Carlos Ribeiro Justiniano Chagas (1789-1934), comenzó a dársele importancia a los relatos sobre las vinchucas. Este investigador del Instituto Oswaldo Cruz (Brasil) estaba realizando estudios sobre la Malaria en el estado de Minas Gerais cuando comenzó a interesarse por aquellos insectos que vivían en casas precarias y que se alimentaban de la sangre humana y de los animales. Al investigar estos insectos encontró, en su tracto digestivo, gran cantidad de unos parásitos protozoarios a los que llamó "*Schizotrypanum cruzi*" (años más tarde denominado "*Trypanosoma cruzi*"). En estudios posteriores, Chagas identificó estos parásitos en sangre de un gato y en sangre de una niña de 2 años. De esta manera, descubrió cuál era el agente causante y el vector de la Enfermedad de Chagas, y además contribuyó en la descripción de la sintomatología que presenta la enfermedad (Chagas 1909). Los estudios del Dr. Chagas impulsaron a otros investigadores a direccionar sus investigaciones hacia la novedosa enfermedad. En Argentina, fue el Dr. Salvador Mazza (1886-1946) junto con otros miembros de la Misión de Estudios de Patología Regional Argentina (MEPRA) quienes retomaron las investigaciones de Carlos Chagas. Mazza y sus colaboradores investigaron y analizaron durante varios años infecciones con *T. cruzi* a lo largo del territorio argentino. En su trabajo publicado en 1949 quedó demostrada la importante contribución del MEPRA a la identificación del parásito y los vectores involucrados en la enfermedad, a la identificación de otros hospedadores naturales, y a la caracterización de las fases de la enfermedad (Mazza 1949). Esta contribución del MEPRA, sumada a resultados de otros investigadores como Romaña, Jörg, Dias y Laranja respecto a la descripción de las fases clínicas impulsó el reconocimiento de la enfermedad de Chagas como un grave problema médico de la región de América del Sur (Sanmartino 2009).

1.2.3. *Trypanosoma cruzi*

El agente etiológico de la enfermedad de Chagas, llamado *Trypanosoma cruzi*, pertenece a una diversa familia de parásitos protozoarios denominados tripanosomátidos (Figura 1.2). A esta familia pertenecen también otros parásitos peligrosos para la salud humana, tales como

distintas especies de *Trypanosoma brucei* y *Leishmania*, que causan la Enfermedad del Sueño (o Tripanosomiasis africana) y la Leishmaniasis, respectivamente (Rassi et al. 2012; Kaufer et al. 2017).

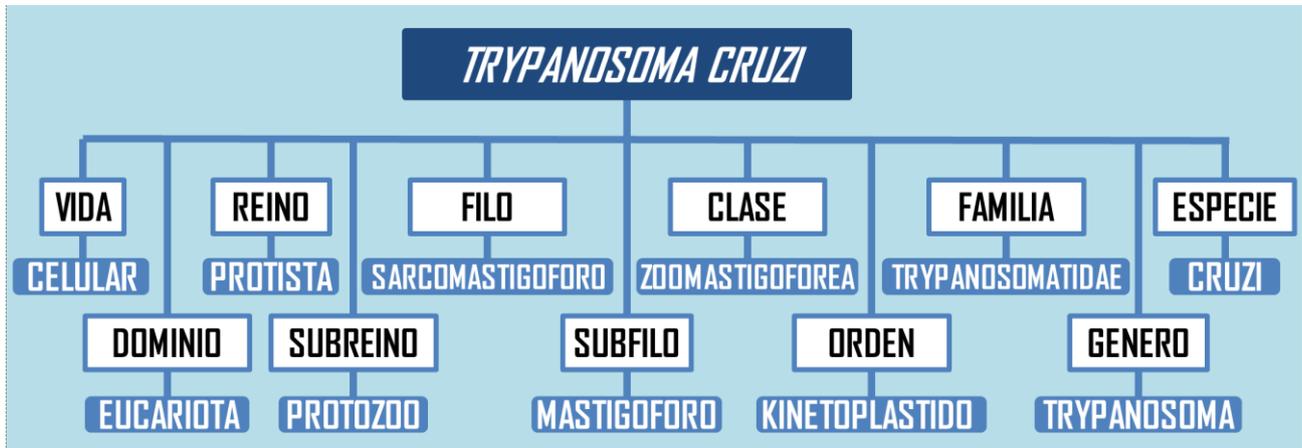


Figura 1.2: Taxonomía de *T. cruzi*.

T. cruzi presenta una alta diversidad fenotípica y genética. De acuerdo a la distribución geográfica, patogénesis, características clínicas y respuesta a la terapia, los diferentes genotipos han sido clasificados en siete Unidades Discretas de Tipificación (DTUs): desde TcI a TcVI corresponden a los DTUs que pueden infectar humanos, y TcVII corresponde al genotipo asociado a murciélagos que hasta la fecha no se ha encontrado en otros animales (Pérez-Molina & Molina 2018; Perez et al. 2014).

1.2.4. VÍAS DE TRANSMISIÓN

Históricamente, la principal vía de transmisión de *T. cruzi* en áreas endémicas se produce a través de las heces depositadas cercanamente a la picadura de ciertas especies de insectos triatominos hematófagos, conocidos popularmente en Argentina y otros países de Latinoamérica por el nombre de “vinchuca”. Esta vía de transmisión está estrechamente relacionada con las condiciones de pobreza, las viviendas precarias, las malas condiciones sanitarias y los entornos peri-domésticos que son habituales en ciertos sectores de los países endémicos (Gürtlera & Yadon 2014; Coura 2007).

Durante el día, la mayoría de las vinchucas se refugia en grietas y rincones oscuros como los que se pueden encontrar detrás de un cuadro o en los techos de paja de las viviendas. El entorno peri-domiciliario, podría ofrecer otros lugares de refugio para el insecto hasta que se realice su entrada a la casa; por otra parte, en estos entornos es común la presencia de animales como pollos, vacas, cabras y perros que ingresan a las casas y que actúan como importantes reservorios para el parásito (Leite Dias et al. 2016).

Durante la noche las vinchucas salen de sus guaridas para succionar la sangre de animales y humanos pudiendo producir la infección. Estos insectos adquieren los parásitos cuando succionan la sangre de un hospedador vertebrado infectado. Luego, los parásitos se dividen y multiplican en el intestino medio posterior, y son expulsados durante la defecación. Debido a la corta longitud del tramo digestivo de las vinchucas, cuando éstas pican a un nuevo hospedador, inmediatamente defecan en las cercanías de la zona picada. De esta manera, *T. cruzi* puede

ingresar a través de la lesión en la piel o por las microlesiones causadas con las uñas al rascarse (De Fuentes-Vicente et al. 2018).

Existen otras vías de transmisión de *T. cruzi*, que se ven acentuadas en países no endémicos. Una de ellas es la transmisión vertical o transmisión congénita, que se da en el embarazo o parto cuando la mujer es portadora del parásito. Si bien aún resulta difícil definir los factores involucrados en esta transmisión, la tasa estaría relacionada con cantidad de parásitos circulantes en la sangre de la embarazada y el DTU o los DTUs involucrados. Dadas las grandes diferencias en las estimaciones realizadas por entidades civiles, sanitarias y estatales de distintos países endémicos y no endémicos, que van de 0,8% a más del 22%, definir una tasa promedio de transmisión vertical es parte de un debate actual; en ese contexto, estimaciones recientes proponen como valor promedio un 4,7% (Pérez-Molina & Molina 2018).

Las transfusiones sanguíneas son otra forma de transmisión, representando un 10-25% del total de contagios. Menos comunes aún son las transmisiones a través de trasplantes de órganos con donantes infectados. Por último, es posible infectarse con *T. cruzi* mediante la ingesta de alimentos y bebidas contaminadas; y también como consecuencia de accidentes de laboratorio (Figura 1.3) (Pérez-Molina & Molina 2018).



Figura 1.3: Vías de transmisión de *T. cruzi*.

1.2.5. CICLO DE VIDA DEL PARÁSITO

T. cruzi presenta un ciclo de vida muy complejo ya que involucra un gran número de hospedadores vertebrados e invertebrados, así como diferentes formas dentro de cada uno de los hospedadores.

La diferenciación celular en *T. cruzi* se desencadena por factores ambientales, como la inanición, el pH y los cambios de temperatura. El ciclo de vida comienza cuando un insecto vector no infectado pica a un hospedador vertebrado infectado succionando la forma circulante de *T. cruzi* (**tripomastigotes sanguíneos**). Muchos de los tripomastigotes ingeridos serán lisados en el estómago del insecto. En unos pocos días, los parásitos sobrevivientes se transforman en la forma de **epimastigotes** y migrarán hacia el intestino donde se dividen intensamente. En las regiones posteriores del intestino y el recto, los epimastigotes alcanzan la forma de **tripomastigotes metacíclicos** (metacicloogénesis), altamente infectivos para los vertebrados, que son liberados junto con las heces y la orina de los vectores. La forma más común de infección es por inoculación directa de los tripomastigotes metacíclicos a través de la mucosa ocular y de las lesiones en la piel que produce el insecto al picar. Las formas metacíclicas invaden los fibroblastos, macrófagos y células epiteliales del sitio de inoculación comenzando con el ciclo intracelular de *T. cruzi* dentro del hospedador vertebrado. Diferentes señales dentro de las células provocan cambios morfológicos en los tripomastigotes metacíclicos, que se transforman

en **amastigotes**. Esta forma se multiplica por fisión binaria en las células infectadas. Diversos factores favorecen la conversión de los amastigotes en tripomastigotes (tripomastigogénesis), que producen la ruptura de la célula infectada e ingresan al torrente sanguíneo del hospedador vertebrado. Éstos se mantienen en circulación y pueden acumularse en otros tejidos para generar nuevos nidos de amastigotes. Cuando un insecto no infectado pica al vertebrado infectado se reinicia el ciclo. Estudios realizados por Kessler y colaboradores han demostrado que habría posibilidades de que los epimastigotes recién diferenciados también tengan capacidad infectiva (Kessler et al. 2017; de Souza et al. 2010; Gonçalves et al. 2018). En la **Figura 1.4** se encuentra un resumen del ciclo de vida del parásito.

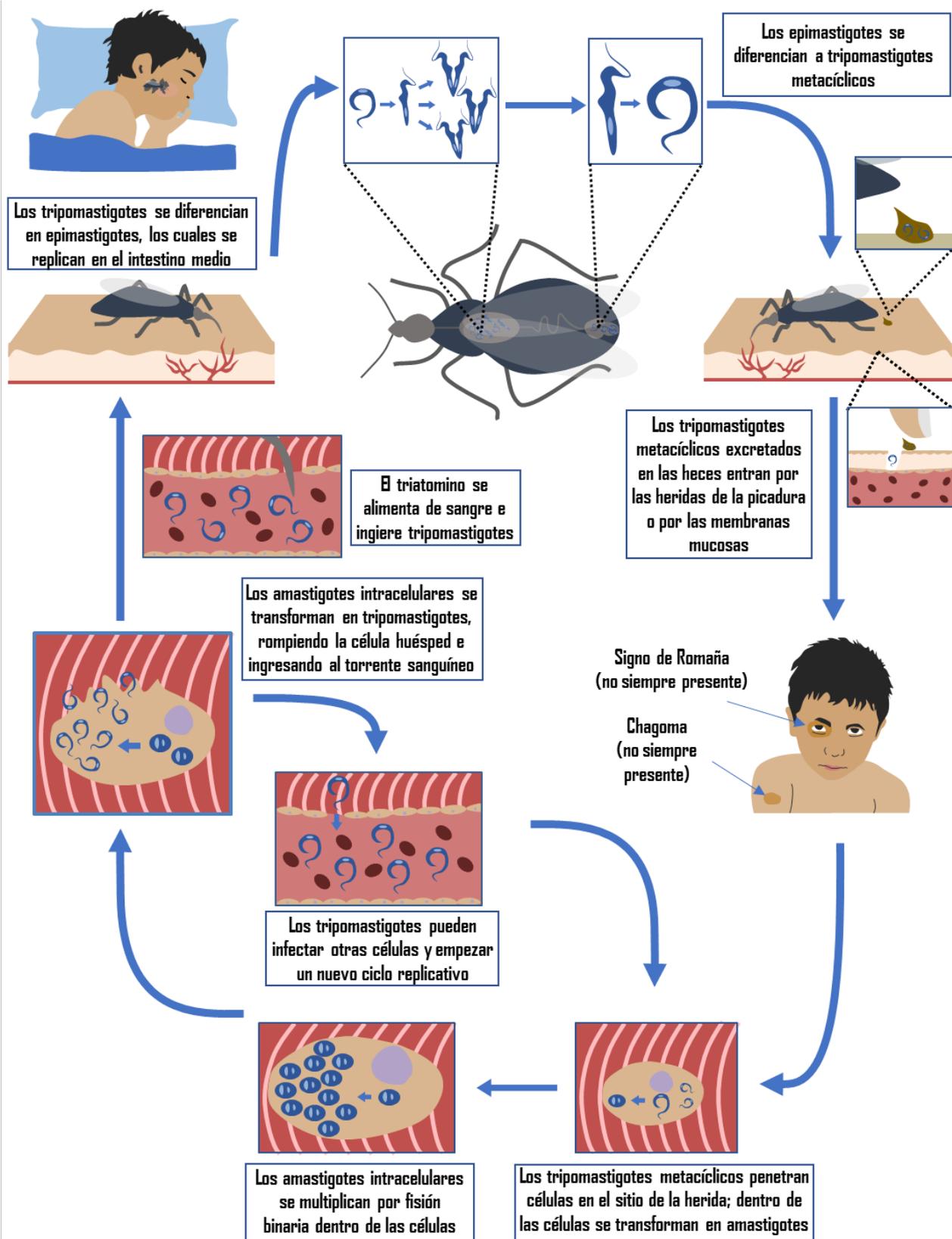


Figura 1.4: Ciclo biológico de *T. cruzi*. Adaptado de Perez-Molina 2018.

1.2.6. EPIDEMIOLOGÍA Y DISTRIBUCIÓN

Históricamente la Enfermedad de Chagas se localizaba únicamente en la parte continental de América Latina. No obstante, debido a los movimientos migratorios de la población, actualmente se reportan con mayor frecuencia casos en Estados Unidos, Canadá, muchos países europeos y algunos del Pacífico Occidental (Antinori et al. 2017; Meymandi et al. 2017; Salvador et al. 2014).

En los últimos 30 años, iniciativas nacionales e internacionales en conjunto con la OMS han logrado importantes avances en el control del vector y de *T. cruzi* en los países endémicos. Programas de rociado con insecticidas, mejoras domiciliarias, higiene del hogar, control de donantes de sangre, actividades de educación y comunicación han conducido a grandes reducciones en la transmisión del parásito por vectores domiciliarios y por transfusiones sanguíneas (World Health Organization 2010).

Según estimaciones de la OMS, para el año 2015 existirían cerca de 6 millones de personas infectadas con *T. cruzi* en Latinoamérica, y más de 70 millones de personas con riesgo de ser infectadas (World Health Organization 2015). La OPS estima, en cambio, unas 8 millones de personas infectadas (<https://www.paho.org/>, último acceso en octubre de 2018).

Argentina, con al menos 1,5 millones, presenta el mayor número de personas infectadas con *T. cruzi*; Bolivia sería el país con mayor número habitantes infectados por transmisión vectorial al año (con unos 8000 casos); y México presentaría el mayor número anual de infectados por transmisión congénita (con unos 1800) casos. La prevalencia estimada de infección cada 100 habitantes debido a la transmisión vectorial para Argentina sería de 3,640, es decir, aproximadamente 4 de cada 100 personas estarían infectadas por *T. cruzi* en Argentina (Tabla 1.1) (World Health Organization 2015).

Tabla 1.1: Datos estimados por la Organización Mundial de la Salud para los 4 países con mayor número de personas infectadas.

PAÍS	ARGENTINA	BOLIVIA	BRASIL	MÉXICO
Población	41.343.000	9.947.000	190.755.799	112.468.855
Nº estimado de personas infectadas con <i>T. cruzi</i>	1.505.235	607.186	1.156.821	876.458
Nº estimado anual de nuevos casos por vía vectorial	1078	8087	46	6.135
Nº estimado anual de infectados por transmisión congénita	1457	616	115	1788
Prevalencia estimada de infección por cada 100 habitantes	3,640	6,104	0,699	0,799
Incidencia estimada de infección debido a transmisión vectorial por 100 habitantes	0,0020	0,0810	0,084 por 100.000 habitantes	0,0050
Incidencia estimada de infección debido a transmisión congénita por 100 nacimientos	0,210	0,235	0,020	0,089
Población estimada en riesgo de infección	2.242.528	586.434	25.474.365	23.474.780

1.2.7. MANIFESTACIONES CLÍNICAS DE LA ENFERMEDAD

La evolución clínica de la enfermedad de Chagas presenta dos fases: una fase aguda y una crónica.

La infección aguda es generalmente asintomática; algunas personas infectadas pueden presentar síntomas como fiebre, inflamación en el sitio de inoculación, signo de Romaña, linfadenopatías, y hepatoesplenomegalia. Esta fase puede ocurrir a cualquier edad, aunque se presenta más a menudo en los primeros años de vida. Dura entre 4 y 8 semanas, y la parasitemia disminuye sustancialmente desde el día 90 en adelante debido a la respuesta inmune del hospedador. Entre el 1 y el 5% de las personas infectadas pueden presentar una fase aguda severa que incluyen síntomas como miocarditis aguda, derrame pericárdico y meningoencefalitis (riesgo de mortalidad: 0.2 – 5%). Los pacientes que no fueron tratados durante esta fase permanecen infectados crónicamente (Dias et al. 1956; Rassi et al. 2012).

La mayoría de las personas infectadas nunca desarrollan síntomas o problemas viscerales y permanecen en la llamada forma “indeterminada” de la Enfermedad de Chagas. Entre el 30 y 40% de los pacientes infectados crónicamente pueden presentar afecciones en ciertos órganos luego de transcurridos de 10 a 30 años de la infección aguda (Sabino et al. 2013; Pérez-Molina & Molina 2018). La afección cardíaca es el tipo más frecuente y grave de la enfermedad, ocurre en 14-45% de los pacientes crónicos y afecta principalmente el miocardio y el sistema de conducción. Las afecciones gastrointestinales (principalmente megavísceras como megaesófago, megacolon o ambas) son menos comunes (10-21%) y ocurren más frecuentemente en Sudamérica. Las neuropatías, en la forma de polineuropatía sensorial leve, pueden presentarse en más del 10% de los pacientes (Viotti et al. 2006; Coura 2007; Dias 1989; Salvador et al. 2014).

La Enfermedad de Chagas es la principal causa de ataque cardioembólico. Se estima una incidencia de ataque isquémico de 2,7 eventos por cada 100 pacientes por año, y aproximadamente 1/3 de los pacientes que experimentan un ataque pueden tener infección asintomática de *T. cruzi* (Nunes et al. 2009; Cardoso et al. 2014).

1.2.8. MEDICAMENTOS DISPONIBLES

Existen actualmente dos medicamentos para el tratamiento de la enfermedad de Chagas los cuales se encuentran en el mercado desde la década de los 70: **Benznidazol (BZN)** y **Nifurtimox (NFX)** (Figura 1.5). Ambos tratamientos comparten algunas características en común: buena tolerancia por parte de los niños, mayor efectividad durante la fase aguda de la enfermedad, baja tasa de curación en adultos en fase crónica, toxicidad en adultos y diferente susceptibilidad entre los DTUs de *T. cruzi* (Sbaraglini et al. 2016a).

El **Nifurtimox** (Figura 1.5) es un 5-nitrofurano producido discontinuamente por Bayer desde el año 1967. Aunque el mecanismo de acción no ha sido completamente dilucidado, se sabe que mediante la reducción de su grupo nitro a radicales aniónicos inestables induce a la formación de especies reactivas de oxígeno (ROS). Estas especies reactivas serían las encargadas del efecto tripanocida ya que *T. cruzi* carece de mecanismos de desintoxicación eficientes para estos metabolitos. También se han descrito mecanismos de acción adicionales que implican metabolitos de nitrilo citotóxicos (Sbaraglini et al. 2016a).

Las dosis recomendadas por la OMS para NFX son de 8 a 10 mg/kg/día en 3 dosis orales para adultos, y entre 15 y 20 mg/kg/día en 4 dosis orales para niños, durante 60 a 90 días. La tasa de curación en niños menores de 14 años en fase crónica tratados con NFX es alrededor del 86%, sin embargo, en adultos es entre 7-8% (Pérez-Molina & Molina 2018; Coura & De Castro 2002). Las afecciones gastrointestinales son los efectos adversos predominantes observados con este tratamiento, sin embargo 30% de los pacientes también pueden experimentar alteraciones del sistema nervioso central como polineuritis, confusión o convulsiones focales o generalizadas, e incluso psicosis, que pueden interrumpirse o continuar cuando se detiene el tratamiento. Algunos pacientes también pueden desarrollar erupciones en la piel, y anemia hemolítica inducida por medicamentos. Se ha observado un aumento en las aberraciones cromosómicas en los niños que reciben nifurtimox (Rajão et al. 2014; Bermudez et al. 2016). Debido a su toxicidad, el tratamiento es discontinuado en alrededor del 44% de los casos (Jackson et al. 2010; Crespillo-Andújar et al. 2018).

El **Benznidazol (Figura 1.5)** es un derivado 2-nitroimidazol que produce la formación de radicales libres y metabolitos electrofílicos que son generados cuando su grupo nitro es reducido a un grupo amino por la acción de las nitroreductasas. Estos metabolitos reducidos se unen covalentemente a las macromoléculas del parásito generando el efecto tripanocida. Fue producido originalmente por Roche y registrado en Argentina, Brasil, Bolivia, Uruguay, Perú, Nicaragua y Japón. En 2003, Roche discontinuó la producción y los derechos fueron transferidos al gobierno brasileño, quien logró producirlo recién en 2008. Sin embargo, en 2011 la Agencia Nacional de Vigilancia Sanitaria (ANVISA) impidió la comercialización del medicamento a través de la Organización Panamericana de la Salud (OPS), el principal mecanismo de adquisición para países endémicos. Un año después, mediante un convenio público-privado entre el Ministerio de Salud de la República Argentina, la Fundación Mundo Sano, la empresa elaboradora de principios activos Maprimed y el Laboratorio farmacéutico Elea, se logró producir benznidazol en la República Argentina. De esta forma, Laboratorio Elea se convirtió en el único productor de BZN capaz de ser comercializado a través del fondo de la OPS (Pinheiro et al. 2017; Bermudez et al. 2016).

El BZN se administra oralmente durante 60 días en concentraciones de 5 a 7 mg/kg para adultos, y 10 mg/kg para niños. La enfermedad congénita se cura en el 100% de los casos cuando el paciente es tratado con BZN antes del primer año de vida (Schijman et al. 2003; Russomando et al. 1998). En la fase aguda, el BZN presenta una tasa de curación del 76% (Cançado 1999), mientras que durante la fase crónica la tasa es mucho menor: 60-93% en niños menores a 13 años y solo 23-40% en adultos (aunque este dato genera controversia) (Pérez-Molina & Molina 2018). Pueden observarse variados efectos adversos como hipersensibilidad, supresión de la médula ósea y neuropatías periféricas. Se ha comprobado que BZN presenta una buena distribución por lo que se estima que la baja biodisponibilidad del compuesto sería una de las razones de la baja eficacia terapéutica durante la fase crónica de la enfermedad (Perin et al. 2017).

En Argentina el BZN es el tratamiento de primera línea, ya que presenta menor frecuencia de efectos adversos y el tratamiento es más corto con respecto al NFX.

Como se dijo, la efectividad de BZN en fase crónica es tema de debate. A fines del año 2015 fueron publicados los resultados del "Benznidazole Evaluation for Interrupting Trypanosomiasis (BENEFIT)", un estudio diseñado para evaluar el desempeño del BZN en pacientes en fase crónica de la enfermedad y sus efectos en la progresión de la misma. BENEFIT fue una prueba multicéntrica, internacional, aleatorizada, doble ciego y controlada por placebo en donde 2854

pacientes recibieron tratamiento durante 40 a 80 días con BZN (5 mg/kg/día) o placebo, y fueron seguidos durante 5 años. Se estudiaron los efectos adversos, la función del hígado, se realizaron electrocardiogramas y se determinó la carga parasitaria en sangre durante el tratamiento, al final del tratamiento, 6 meses después y luego anualmente hasta finalizar el estudio. Como resultados generales, el estudio demostró que el tratamiento con BZN reduce significativamente la carga parasitaria en sangre, pero esta reducción no se correlaciona con mejoras en el deterioro cardíaco. Además, el 23,9% de los pacientes tratados con BZN abandonaron el tratamiento debido a los efectos adversos, un porcentaje significativamente mayor que el porcentaje de abandono de pacientes tratados con placebo (9,5%) (Morillo et al. 2015). Sin embargo, un estudio anterior al BENEFIT que se realizó sobre 566 adultos en fase crónica señala que el tratamiento con BZN retarda el avance del desarrollo y la progresión de las cardiopatías (Viotti et al. 2006).

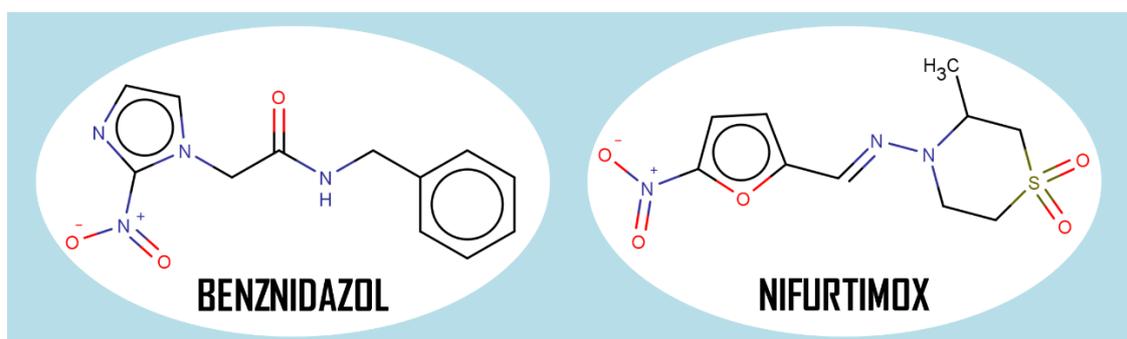


Figura 1.5: Estructuras de los fármacos tripanocidas aprobados actualmente para tratar la Enfermedad de Chagas.

De acuerdo con la OMS, una droga ideal para el tratamiento de la Enfermedad de Chagas debería cumplir ciertas características: lograr la cura parasitológica durante las fases aguda y crónica; ser efectiva en una o en unas pocas dosis; ser de bajo costo; no poseer efectos adversos o teratogénicos; y no inducir la generación de cepas resistentes. Como podemos notar, las drogas actualmente disponibles no cumplen con muchos de los requerimientos indicados y se necesitan nuevas drogas que mejoren la efectividad y sean bien toleradas (Coura & De Castro 2002).

1.2.9. AVANCES HACIA NUEVOS MEDICAMENTOS

A pesar del descubrimiento de compuestos químicos con potente actividad tripanocida *in vitro* e *in vivo*, hasta la fecha no se ha encontrado un compuesto que consiga mejores resultados terapéuticos que BZN en los ensayos clínicos.

Algunos de los candidatos más avanzados para el desarrollo de nuevos tratamientos para la Enfermedad de Chagas contienen en su estructura química grupos azoles y han sido identificados como inhibidores de la síntesis de ergosterol en *T. cruzi*:

El **Posaconazol** es un fármaco comercialmente disponible para el tratamiento de infecciones fúngicas y la candidiasis que ha demostrado una potente actividad *in vitro* e *in vivo* frente a cepas de *T. cruzi* resistentes a las drogas actuales (Urbina et al. 1998; Molina et al. 2000). Fue más activa que BZN contra cepas resistentes en modelos murinos agudo y crónico, aunque presentó poca actividad en otras cepas (Moraes et al. 2015; Molina et al. 2000). Un estudio *in vivo* reveló que el Posaconazol falla en eliminar una cepa resistente al BZN de *T. cruzi* como monoterapia y que en combinación con BZN es menos efectivo que sólo BZN en las cepas susceptibles

(Bustamante et al. 2014). En 2012 concluyó un estudio clínico de Fase II en el que participaron 78 pacientes con Chagas crónico, en el cual se administró, a las ramas del estudio: BZN (5 mg/kg/día por 60 días) y dos esquemas de Posaconazol (100 mg/12hs y 400 mg/12hs por 60 días). Durante el seguimiento se encontró que una mayor proporción de los pacientes tratados con Posaconazol mantenían seropositividad para *T. cruzi* (medida por PCR en tiempo real) (Molina et al. 2014). STOP CHAGAS fue otro estudio clínico con Posaconazol que finalizó en 2015 y demostró que la monoterapia con BZN es más eficiente que la combinación de BZN y Posaconazol (Morillo et al. 2017).

El **Ravuconazol**, otro inhibidor de la síntesis de ergosterol en *T. cruzi* ha presentado también buena actividad en ensayos *in vitro* e *in vivo* contra cepas resistentes; sin embargo, no ha demostrado buena actividad en ratones o perros infectados (Buckner & Urbina 2012; Urbina et al. 2003; Diniz et al. 2010). Debido a su perfil farmacocinético desfavorable, se ha procedido a la síntesis y evaluación de **E-1224**, una prodroga del Ravuconazol. A pesar de que E-1224 mejoró la absorción y biodisponibilidad de Ravuconazol, no logró eliminar eficientemente a *T. cruzi* en las pruebas clínicas (DNDi 2013).

Otros azoles están siendo evaluados en ensayos clínicos, tales como el **Itraconazol**, **Fluconazol**, **Fexinidazol** y **Ketaconazol** (Sales Junior et al. 2017; Sbaraglini et al. 2016a).

Otra droga que no pertenece a la familia de los azoles y está siendo evaluada en ensayos clínicos es la **Amiodarona** (ClinicalTrials.gov Código: NCT03193749). **Esta droga** es un agente antiarrítmico de clase III que se usa para el tratamiento de las arritmias que padecen pacientes con cardiopatías chagásicas. Se ha demostrado a nivel preclínico que esta droga posee una fuerte actividad tripanocida en ensayos *in vitro* e *in vivo*, y que actúa sinérgicamente con Posaconazol en células infectadas con amastigotes de *T. cruzi* (Benaïm et al. 2006). Sin embargo, la Amiodarona presenta algunos efectos secundarios debido a la presencia de yodo en su estructura, por lo tanto, se ha sintetizado y evaluado un compuesto análogo, la **Dronedarona**. Esta droga ha demostrado a nivel preclínico una actividad tripanocida más potente que la Amiodarona (Benaïm et al. 2012).

1.3. BLANCOS MOLECULARES EN *T. cruzi*

Los avances en el conocimiento de la proteómica, bioquímica y biología de *T. cruzi* ha estimulado el desarrollo de nuevas aproximaciones racionales hacia una quimioterapia novedosa y eficiente para el tratamiento de la Enfermedad de Chagas, recurriendo a diversos blancos moleculares para identificar potenciales compuestos tripanocidas (Croft et al. 2005; Urbina 2010).

La definición de un buen blanco molecular depende del área terapéutica en cuestión. Un buen blanco molecular para la búsqueda de nuevas drogas para el tratamiento de la Enfermedad de Chagas debería poseer ciertas propiedades:

- **Ser esencial:** Se debe verificar que la forma infectiva del parásito es incapaz de sobrevivir en el hospedador cuando el blanco molecular se inhibe por vía farmacológica o genética.
- **Poseer drogabilidad (*druggability*):** Debe ser sensible a la inhibición por pequeñas moléculas tipo fármaco o por productos biológicos como proteínas recombinantes.
- **Poseer selectividad:** Debe ser único para *T. cruzi*, es decir debe estar ausente o ser molecular y/o bioquímicamente diferente del homólogo presente en el hospedador.

- **Presentar baja resistencia:** Es deseable que sólo presente una copia del gen, no presente isoformas, sistemas de *backup* ni reacciones o vías de *bypass*.
- **Conocer su estructura tridimensional:** Es importante contar con la estructura RMN o el cristal del blanco molecular y/o del homólogo.
- **Poseer ensayabilidad (*assayability*):** Deben existir ensayos que permitan evaluar la actividad del blanco en presencia de los candidatos inhibidores.
- **Bioquímica favorable:** Es deseable que haya bajo contenido intracelular y baja tasa de recambio del blanco molecular; y que la vía sea un cuello de botella.

Las primeras tres propiedades planteadas son absolutamente necesarias para que una proteína sea considerada un blanco molecular adecuado, mientras que las otras características no son indispensables, pero si recomendadas para aumentar el éxito de la estrategia farmacológica. (Gashaw et al. 2011; Talevi et al. 2018).

Existen más de 30 blancos moleculares estudiados actualmente para el desarrollo de nuevos inhibidores de *T. cruzi*, la gran mayoría de ellos están asociados con enzimas involucradas en el metabolismo del parásito.

1.3.1. INHIBIDORES DE LA SÍNTESIS DE ESTEROLES

A diferencia de las células de mamífero, que sintetizan colesterol, los tripanosomátidos producen principalmente ergosterol. *T. cruzi* contiene cantidades significativas de colesterol de origen exógeno, pero mantiene una alta susceptibilidad a inhibidores de la biosíntesis de esteroides, demostrando la necesidad por esteroides específicos que no son sintetizados por el hospedador. Se ha demostrado que estos esteroides específicos son necesarios para la viabilidad celular y la proliferación en todos los estados del ciclo de vida del parásito. En particular, la síntesis de ergosterol ha sido validada *in vitro* e *in vivo* y es considerada uno de los blancos moleculares más atractivos para el tratamiento específico de la Enfermedad de Chagas (Urbina 2009; Kessler et al. 2013; Buckner & Urbina 2012), aunque los inhibidores de esta vía han conducido, a la fecha, a resultados clínicos decepcionantes.

Los azoles discutidos en la sección anterior han sido reportados como inhibidores de la enzima **C14- α -desmetilasa (CYP51)**, particularmente Ketoconazol, Miconazol, Clotrimazol, Fluconazol, Itraconazol, Posaconazol y Ravuconazol (Buckner & Urbina 2012).

Otra enzima de esta vía que se presenta como un buen blanco molecular en *T. cruzi* es la **escualeno sintasa (SQS)**, que cataliza el primer paso en la biosíntesis de esteroides. Algunos de los inhibidores de la SQS que han demostrado fuerte actividad tripanocida son las quinuclidinas E-5700 y ER-119884 y los tiocianatos de ariloxiletileno como WC-9 (Urbina 2009). La Terbinafina, un potente inhibidor de la **escualeno epoxidasa (SQO)** en hongos, ha demostrado actividad tripanocida frente a epimastigotes de *T. cruzi* (Maldonado et al. 1993).

Algunos inhibidores de la **lanosterol sintasa (OSC)** también han demostrado una fuerte actividad tripanocida *in vitro*. Además, se ha encontrado que la Amiodarona bloquea la síntesis de ergosterol mediante la inhibición de la esta enzima (Urbina 2009).

Otro potencial blanco molecular que actúa en los primeros pasos de la síntesis de esteroles es la **3-hidroxi-3-metil-glutaril-CoA reductasa (HMG-CoA reductasa)**, la cual es inhibida por las estatinas como la Lovastatina (Urbina et al. 1993; Kessler et al. 2013).

Los bifosfonatos conteniendo nitrógeno como Alendronato, Pamidronato y Risendronato han sido identificados como inhibidores de la **farnesilo pirofosfato sintasa (FPPS)** con fuerte actividad tripanocida *in vitro* e *in vivo* (Garzoni et al. 2004a; Garzoni et al. 2004b; Bouzahzah et al. 2005; Martin et al. 2001).

1.3.2. CRUZIPAÍNA

La **cruzipaína** es la principal cisteína-proteasa en *T. cruzi*. Es expresada en todas las formas del parásito y es responsable de la mayor parte de la actividad proteolítica del mismo. Esta enzima juega un rol importante en diversas funciones vitales de *T. cruzi* tales como la invasión de las células hospedadoras, la evasión de la respuesta inmune, la proliferación de epimastigotes y amastigotes y la metacicloogénesis. Además, no presenta homólogos en mamíferos por lo que es considerada uno de los blancos moleculares más atractivos para el desarrollo de nuevos inhibidores (Cazzulo 2002; Caffrey et al. 2000; McKerrow et al. 2009; Aparicio et al. 2004; Duschak & Couto 2009).

Se han desarrollado gran cantidad de inhibidores con potente actividad *in vitro* e *in vivo*, sin embargo, muy pocos han avanzado hacia los ensayos clínicos. La vinilsulfona K777, ha demostrado ser efectiva en modelos animales, pero el avance de la droga fue detenido por problemas de hepatotoxicidad (McKerrow et al. 2009).

Bellera y colaboradores han encontrado mediante estrategias computacionales varios fármacos con capacidad inhibitoria frente a la cruzipaína, entre ellos, la Benidipina y la Clofazimina (Sbaraglini et al. 2016b; Bellera et al. 2015; Bellera et al. 2013). Ambos candidatos han reducido la parasitemia en un modelo murino agudo de infección con *T. cruzi*, aunque en menor medida que el tratamiento con BZN 100 mg/kg/día. Adicionalmente, estas drogas fueron capaces de reducir la carga parasitaria en músculo cardíaco y esquelético de ratones infectados crónicamente, con respecto a los ratones tratados con BZN.

1.3.3. TRANS-SIALIDIASA

Uno de los mecanismos empleados por *T. cruzi* para evadir la respuesta inmune del hospedador es la expresión de la enzima **trans-sialidasa**, la cual cataliza la transferencia de ácidos siálicos desde los sialo-glicoconjugados del hospedador a los glicoconjugados del parásito. Además de su importante rol en la regulación del sistema inmune, la presencia del ácido siálico ayuda a los parásitos a unirse e ingresar a las células hospedadoras (Miller III & Roitberg 2013).

Estas características, sumada a la ausencia de homólogos de *trans-sialidasa* en mamíferos, han convertido a esta enzima en un interesante blanco molecular. Sin embargo, hasta la fecha no se ha encontrado ningún inhibidor que reduzca efectivamente la actividad catalítica de la enzima (Kashif et al. 2017).

1.3.4. INHIBIDORES DE LA VÍA GLUCOLÍTICA

Los tripomastigotes de *T. cruzi* no tienen un ciclo funcional del ácido tricarbóxico, siendo altamente dependiente de la glicólisis para la producción de ATP. Uno de los blancos moleculares estudiados de esta vía es la enzima **gliceraldehido-3-fosfato deshidrogenasa**

(GADPH) ya que presenta diferencias estructurales con su contraparte de mamíferos (Souza et al. 1998; Lakhdar-Ghazal et al. 2002). Se han sintetizado y evaluado varios análogos de adenosina que demostraron una alta afinidad por la enzima del parásito y buena actividad tripanocida frente a amastigotes. También derivados de flavonoides y una serie de 3-piperonilcumarinas han demostrado actividad inhibitoria frente a la enzima de *T. cruzi* (Verlinde et al. 2001; Soeiro & de Castro 2009).

La primera enzima en la glicolisis es la **hexoquinasa (HK)**, que convierte glucosa en glucosa 6-fosfato. En los tripanosomátidos está localizada en un peroxisoma modificado denominado glicosoma y se ha demostrado que a diferencia de la HK humana, la enzima es inhibida por los bisfosfonatos inorgánicos (Cáceres et al. 2003; Racagni & Machado de Domenech 1983; Hudock et al. 2006).

1.3.5. INHIBIDORES DE LA VÍA DE LAS PENTOSAS FOSFATO

Una defensa antimicrobiana de primera línea en los mamíferos hospedadores implica la producción de especies reactivas de oxígeno, las cuales son neutralizadas por reacciones dependientes de NADPH, que pueden obtenerse mediante las reacciones de la vía de las pentosas fosfato de *T. cruzi*.

La **glucosa-6-fosfato deshidrogenasa**, la **6-fosfogluconato deshidrogenasa (6PGDH)** y la **transcetolasa** han sido consideradas como potenciales blancos moleculares debido a su esencialidad y a las diferencias con su contraparte en mamíferos (Kovářová & Barrett 2016; Igoillo-Esteve et al. 2007; Soeiro & de Castro 2009).

1.3.6. INHIBIDORES DE LA SÍNTESIS Y METABOLISMO DE NUCLEÓTIDOS Y DE LA RECUPERACIÓN DE PURINAS

Mientras que los nucleótidos en mamíferos son sintetizados *de novo* o recuperados desde bases purinas recicladas, los tripanosomátidos dependen de la vía de recuperación y convierten las purinas en ribonucleótidos mediante la acción de la **hipoxantina-guanina fosforibosiltransferasa (HGPRT)** (Ullman & Carter 1995). El Alopurinol, una droga usada para el tratamiento de la gota, actúa como un análogo de purina y es incorporado, a través de HGPRT, dentro del DNA interrumpiendo la síntesis de RNA y proteínas. Esta droga fue activa en modelos murinos de infección con *T. cruzi*, sin embargo se presentaron diferentes susceptibilidades entre las cepas del parásito (Marr 1991).

La **dihidrofolato reductasa (DHFR)** cataliza la transformación de dihidrofolato en tetrahidrofolato, un paso necesario en la síntesis de timidina (Eakin et al. 1997). Una serie de compuestos han sido sintetizados y evaluados como inhibidores de la DHFR, algunos de ellos han demostrado ser mucho más activos sobre la enzima de *T. cruzi* que sobre la enzima humana (Zuccotto et al. 2001; Gilbert 2002; Pez et al. 2003).

1.3.7. AGLUTINANTES DE ADN

Los organismos kinetoplastideos presentan una masa de ADN circular extranuclear que se encuentra dentro de la doble membrana de una gran mitocondria y que contiene numerosas

copias del genoma mitocondrial. Esta estructura, denominada **kinetoplasto**, presenta gran cantidad de secuencias AT, lo que lo convierte en un blanco molecular interesante (Cavalcanti et al. 2008).

Las diamidinas aromáticas, como la Pentamidina, son ligandos del surco menor del ADN que reconocen secuencias de al menos cuatro pares de bases AT y presentan potentes efectos antimicrobianos de amplio espectro. En los tripanosomátidos se ha demostrado que el blanco principal de estos compuestos es el kinetoplasto, aunque la interacción por sí sola no explica completamente su actividad biológica (Bray et al. 2003).

A pesar de que las diamidinas han demostrado una fuerte actividad antiparasitaria, presentan importantes limitaciones en cuanto a la baja biodisponibilidad oral y a los efectos adversos, lo que ha conducido a la síntesis de nuevos candidatos (Wilson et al. 2005; Soeiro & de Castro 2009).

1.3.8. INHIBIDORES DEL SISTEMA DE DEFENSA CONTRA EL ESTRÉS OXIDATIVO

Los tripanosomátidos poseen un metabolismo tiol único basado en el tripanotión (que se verá en la siguiente sección), representando éste la principal protección del parásito frente al estrés oxidativo. Dicho compuesto se mantiene reducido por la acción de la flavoenzima específica del parásito **tripanotión reductasa (TyrR)**. La inhibición de esta vía clave aumenta considerablemente la susceptibilidad de *T. cruzi* a la acción tóxica de las drogas como NFX y BZN (Krauth-Siegel & Comini 2008).

Existe una gran cantidad de inhibidores de TyrR de diferentes clases químicas como las naftoquinonas y sus derivados, los antidepresivos tricíclicos como la clorpromazina, los derivados de poliaminas, los complejos metálicos híbridos, ciertos productos naturales y los inhibidores basados en diaril sulfuros. Sin embargo, se necesita reducir más del 90% de la actividad de TyrR para afectar la viabilidad de los parásitos, lo que explica que todavía no se hayan encontrado inhibidores de TyrR con buena actividad tripanocida (Soeiro & de Castro 2009).

Otra enzima de importancia en el mecanismo de defensa frente al estrés oxidativo es la **tripanotión sintetasa (TyrS)**, que cataliza la biosíntesis del tripanotión. Esta enzima constituye un blanco molecular interesante ya que está codificada por un gen de copia simple, se encuentra disponible la estructura tridimensional del homólogo en *Leishmania* y ha demostrado proveer control metabólico en la síntesis de tripanotión en *T. cruzi* (Benítez et al. 2016).

1.4. POLIAMINAS

1.4.1. CARACTERÍSTICAS GENERALES

Las poliaminas (Putrescina, Espermidina, Espermina y Cadaverina) (**Figura 1.6**) son un grupo de sustancias de bajo peso molecular que se encuentran prácticamente en todas las células procariotas, eucariotas y arqueobacterias (Algranati 2010). En células que proliferan

rápida como células cancerosas y organismos parasitarios la concentración de poliaminas es mucho mayor que en células en crecimiento lento. Debido a su estructura alifática policatiónica, las poliaminas son capaces de unirse a diferentes macromoléculas como ADN, ARN y proteínas mediante interacciones electrostáticas o hidrofóbicas y mediante puentes de hidrógeno (Tabor & Tabor 1976; Igarashi et al. 1974). Por lo tanto, las poliaminas pueden actuar directamente sobre la función de ácidos nucleicos y proteínas, en procesos de biosíntesis, en la conservación de las estructuras y en la transducción de señales. Además, son esenciales para varios procesos celulares como la proliferación, la supervivencia, la diferenciación, el envejecimiento, la autofagia o la tolerancia al estrés (Vanrell et al. 2013; Miller-Fleming et al. 2015; Gevrekci 2017; Handa et al. 2018).

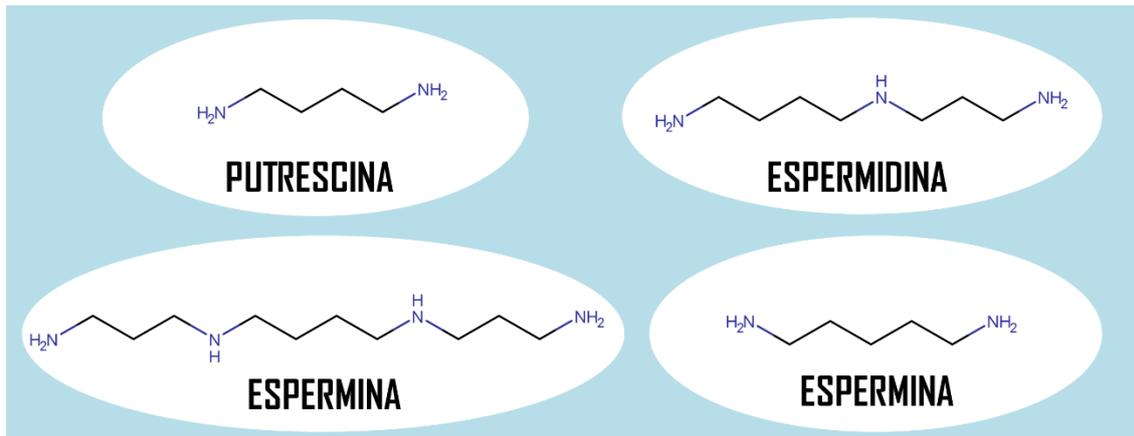


Figura 1.6: Estructura química de las poliaminas.

1.4.2. METABOLISMO DE POLIAMINAS

La forma en que las células adquieren y utilizan las poliaminas puede variar considerablemente entre los diferentes organismos.

Las células de mamífero presentan principalmente tres poliaminas: **Putrescina** (1,4-diaminobutano), **Espermidina** (N'-(3-aminopropil)butano-1,4-diamina) y **Espermina** (N,N'-bis(3-aminopropil)butano-1,4-diamina). En estas células, la biosíntesis comienza con la conversión del aminoácido L-Ornitina en Putrescina catalizado por la enzima ornitina decarboxilasa (ODC). Paralelamente, la enzima S-adenosilmetionina decarboxilasa (AdoMetDC) cataliza la formación de adenosilmetionina decarboxilada (dcAdoMet), la cual provee el grupo aminopropilo para la conversión de Putrescina en Espermidina, a cargo de la espermidina sintasa (SpdS). Finalmente, la espermina sintasa (SpmS) convierte la Espermidina en Espermina (**Figura 1.6a**). La concentración endógena de poliaminas en estas células se encuentra estrictamente regulada por una serie de mecanismos homeostáticos que comprenden su biosíntesis, interconversión y transporte desde el medio externo, su degradación y excreción (Birkholtz et al. 2011; Algranati 2010).

En células de plantas y en bacterias puede encontrarse otra vía biosintética de poliaminas que comprende dos etapas: en primer lugar, la decarboxilación de Arginina por la enzima arginina decarboxilasa (ADC) para dar Agmatina y, luego, la hidrólisis de esta molécula genera la Putrescina mediante una reacción catalizada por la enzima agmatinasa (Algranati 2010).

Por definición, los organismos parásitos viven dentro de otros organismos; debido a ello, un significativo número de metabolitos esenciales para su supervivencia y proliferación pueden ser obtenidos directamente desde las células del hospedador a través de diferentes sistemas de

transporte, sin la necesidad de sintetizarlos *de novo*. Los parásitos se han adaptado a estas condiciones y mantienen solo una limitada capacidad metabólica con respecto a las células de mamífero (Pereira & Carrillo 2013, Fairlamb 1989; Fairlamb 2002; Algranati et al. 2006). El metabolismo de poliaminas de *T. cruzi* es un claro ejemplo de esta adaptación.

Mediante diferentes estudios fisiológicos, bioquímicos y genéticos, se ha demostrado que *T. cruzi* carece de la capacidad de sintetizar poliaminas *de novo* (Figura 1.6b). Se ha observado que epimastigotes de diferentes cepas de *T. cruzi* son capaces de proliferar continuamente en medios ricos en poliaminas, pero el crecimiento se detiene luego de algunos pasajes a medios semisintéticos con solo trazas de poliaminas (Carrillo et al. 1999). Estos parásitos no proliferativos se mantienen viables por varias semanas en el medio semisintético y recuperan la capacidad de crecimiento normal cuando se adicionan poliaminas (Putrescina o Espermidina) exógenas al cultivo. El agregado de Ornitina o Arginina al medio no reinicia la proliferación, hecho que generó la sospecha inicial de que estos aminoácidos no pueden ser precursores de poliaminas en epimastigotes de *T. cruzi*.

Ensayos enzimáticos con ornitina marcada radioactivamente han intentado detectar la actividad de ornitina descarboxilasa (ODC) y la formación de Putrescina en diferentes cepas de *T. cruzi*. Ninguno de los ensayos demostró una actividad enzimática significativa (Ariyanayagam & Fairlamb 1997; Carrillo et al. 1999). Adicionalmente, los cultivos de *T. cruzi* con altas concentraciones de DFMO (el principal inhibidor de ODC) no detienen la proliferación de los parásitos consolidando la sospecha de que la actividad enzimática de ODC está ausente en *T. cruzi* (Carrillo et al. 2007). Con respecto a la síntesis de Putrescina vía la arginina descarboxilasa (ADC), se ha demostrado con Arginina marcada que diferentes cepas de *T. cruzi* no sintetizan Agmatina o Putrescina (Carrillo et al. 2003; Algranati 2010). Bacchi y colaboradores tampoco lograron detectar actividad enzimática de ODC y ADC en ninguno de los estadios del ciclo de vida del parásito concluyendo que *T. cruzi* es autótrofo para las poliaminas debido a su incapacidad de sintetizar Putrescina (Carrillo et al. 2006; Bacchi et al. 2001; Ariyanayagam & Fairlamb 1997; Carrillo et al. 2007). Por lo tanto, esta poliamina (o la Espermidina) debe ser incorporada desde el medio externo para mantener el crecimiento normal del parásito.

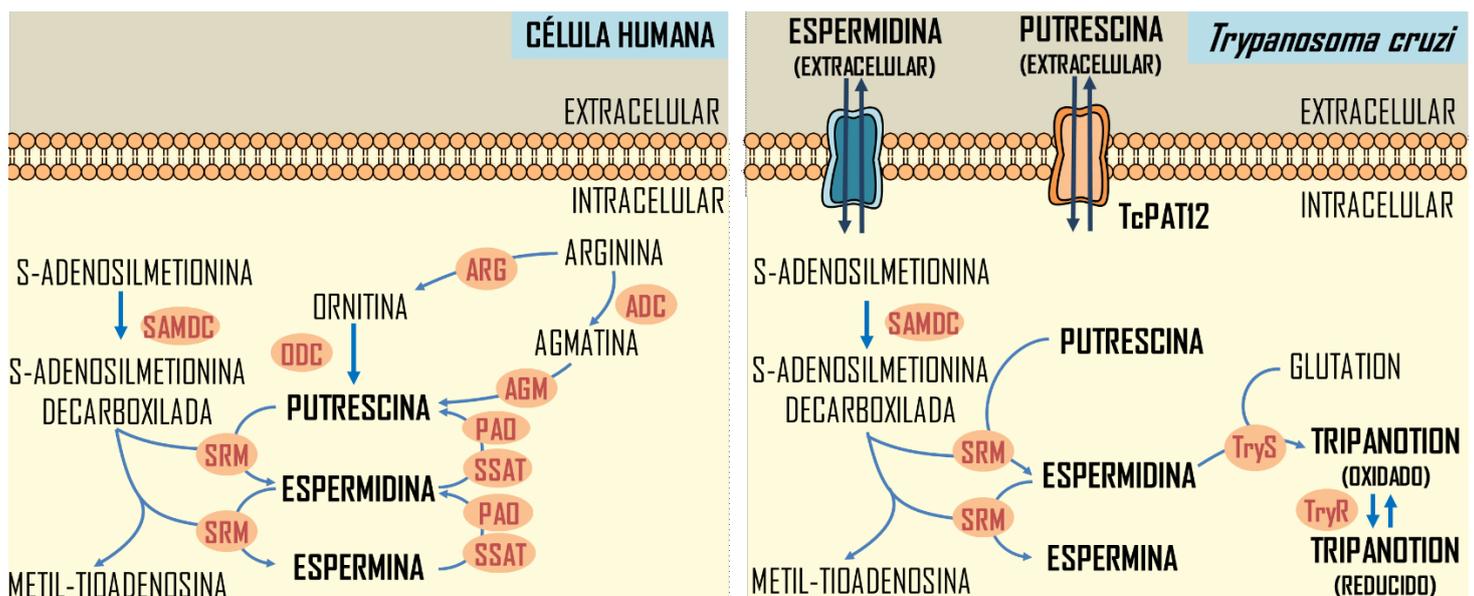


Figura 1.6: Metabolismo de poliaminas en células humanas y *T. cruzi*. SAMDC: S-adenosilmetionina descarboxilasa; ODC: Ornitina descarboxilasa; ARG: Arginasa; ADC: Arginina descarboxilasa; AGM: Agmatinasa; PAD: Poliamina oxidasa;

SRM: Espermidina sintasa; SSAT: Espermidina acetiltransferasa; TryS: Tripanotión sintetasa y TryR: Tripanotión reductasa. Adaptado de Algranati et al 2006.

1.4.3. TRIPANOTIÓN

En los mamíferos y la mayoría de los organismos vivos, el Glutatiión (GSH) cumple un rol importante en la homeostasis redox. Sin embargo, en los tripanosomátidos esta función ha sido reemplazada por el conjugado de la Espermidina con GSH denominado **Tripanotión (T(SH)₂) o N1,N8-bis(glutationil)espermidina** (Fairlamb et al. 1985; Fairlamb & Cerami 1992). Las funciones de este metabolito incluyen la reducción de enzimas involucradas en el metabolismo redox (como triparredoxina y triparredoxina peroxidasa), la reducción de proteínas implicadas en el control de la proliferación y en la síntesis de ADN (como la ribonucleótido reductasa), la desintegración de agentes oxidantes como H₂O₂, peroxinitrito, ácidos grasos de cadena corta e hidroperóxidos fosfolipídicos. Adicionalmente, T(SH)₂ reacciona con una variedad de electrófilos como cetoaldehídos, metales pesados, y compuestos como NFX y BZN (Maya et al. 2014). La biosíntesis de tripanotión involucra la formación gradual de una unión C-N entre la amina libre de la Espermidina y el grupo carboxilato de una molécula de GSH; con la consiguiente unión de la otra molécula de GSH. Ambas reacciones son catalizadas por la enzima tripanotión sintetasa (TryS) que, mediante la utilización de ATP, produce gradualmente mono-glutationileespermidina (Gsp) y luego bis-glutationileespermidina (T(SH)₂) (Talevi et al. 2018). El mantenimiento de los niveles adecuados de T(SH)₂ depende no sólo de su tasa de biosíntesis, sino principalmente de un sistema destinado a reciclar la forma oxidada del tripanotión (TS₂) nuevamente a su forma reducida (T(SH)₂). Este reciclado es llevado a cabo por la flavoenzima tripanotión reductasa (TyrR) que reduce el TS₂ a expensas del NADPH+ (Talevi et al. 2018). Debido a que en la biosíntesis de tripanotión es indispensable la utilización de poliaminas, resulta esencial la presencia de transportadores de poliaminas que incorporen poliaminas desde el hospedador al parásito.

1.4.4. TRANSPORTADORES DE POLIAMINAS

La incorporación de diaminas y poliaminas desde la célula hospedadora hacia el parásito es mediada por transportadores de membrana de alta afinidad.

Los primeros estudios bioquímicos sobre la regulación del transporte de putrescina en *T. cruzi* fueron realizados por González y colaboradores (González et al. 1992). Ellos demostraron, por primera vez, que existen propiedades distintivas entre el transporte de putrescina de *T. cruzi* y de otros tripanosomátidos. Le Quesne y Fairlamb profundizaron estos estudios demostrando que los epimastigotes de *T. cruzi* son capaces de incorporar putrescina y cadaverina rápidamente desde el medio en que se encuentran mediante un sistema de transporte de alta afinidad que puede ser regulado en respuesta a las condiciones de crecimiento y al contenido de poliaminas del medio. Esta actividad de transporte presenta una alta especificidad por Putrescina y Cadaverina y baja especificidad otras poliaminas (Le Quesne & Fairlamb 1996). Las permeasas que realizan este transporte (*TcPAT12* o *TcPOT1*) son miembros del grupo más divergente, en cuanto a la secuencia de aminoácidos, de la familia de los transportadores de aminoácidos/auxinas (AAAP). Esta familia de transportadores pertenece a la clase de transportadores conducidos por potencial electroquímico. Las proteínas de la familia AAAP no

presentan homólogos en células de mamífero: tienen longitudes entre 400 y 500 aminoácidos y se predice que poseen entre 10 y 12 α -hélices transmembrana (Bouvier et al. 2004).

El locus de *TcPAT1* consiste en dos alelos, uno para cada haplotipo de *T. cruzi*, y ambos, *TcPAT1.1* y *TcPAT1.2*, codifican un transportador de diaminas funcional que reconoce putrescina y cadaverina con alta afinidad. Ha sido demostrado que la interrupción genética de *TcPAT1.2* puede impactar en la infectividad y supervivencia de *T. cruzi* (Carrillo et al. 2006; Hasne et al. 2010; Hasne et al. 2016)

1.5. ELECCIÓN DEL BLANCO MOLECULAR

En este trabajo de tesis empleamos una estrategia computacional para la búsqueda de compuestos químicos con actividad inhibitoria sobre las diferentes formas de *T. cruzi* que puedan ser útiles para el desarrollo de nuevos medicamentos para el tratamiento de la Enfermedad de Chagas.

Considerando que varias de las enzimas del metabolismo de poliaminas, e incluso la permeasa de putrescina *TcPAT1.2*, han sido validadas experimentalmente como blancos moleculares de *T. cruzi*, hemos enfocado nuestra búsqueda a inhibidores del metabolismo y/o transporte de poliaminas con actividad tripanocida.

REFERENCIAS

- Algranati, I.D., 2010. Polyamine metabolism in *Trypanosoma cruzi*: Studies on the expression and regulation of heterologous genes involved in polyamine biosynthesis. In *Amino Acids*. pp. 645–651.
- Algranati, I.D. et al., 2006. Biología molecular del metabolismo de poliaminas en parásitos tripanosomátidos. Expresión de genes heterólogos de ornitina y arginina decarboxilasa en *Trypanosoma cruzi*. *Revista Química Viva*, 2, pp.78–94.
- Antinori, S. et al., 2017. Chagas disease in Europe: A review for the internist in the globalized world. *European Journal of Internal Medicine*, 43, pp.6–15.
- Aparicio, I.M. et al., 2004. A New Cruzipain-Mediated Pathway of Human Cell Invasion by *Trypanosoma cruzi* Requires Trypomastigote Membranes. *Infection and Immunity*, 72(10), pp.5892–5902.
- Araújo, A. et al., 2009. Paleoparasitology of Chagas disease--a review. *Memorias do Instituto Oswaldo Cruz*, 104, pp.9–16.
- Ariyanayagam, M.R. & Fairlamb, A.H., 1997. Diamine auxotrophy may be a universal feature of *Trypanosoma cruzi* epimastigotes. *Molecular and Biochemical Parasitology*, 84(1), pp.111–121.
- Aufderheide, A.C. et al., 2004. A 9,000-year record of Chagas' disease. *Proceedings of the National Academy of Sciences*, 101(7), pp.2034–2039.
- Bacchi, C.J. et al., 2001. Stage-Specific Polyamine Metabolism in *Trypanosoma cruzi*. *Journal of Eukaryotic Microbiology*, 48, pp.201–202.
- Bellera, C.L. et al., 2013. Application of computer-aided drug repurposing in the search of new cruzipain inhibitors: Discovery of amiodarone and bromocriptine inhibitory effects. *Journal of Chemical Information and Modeling*, 53(9), pp.2402–2408.
- Bellera, C.L. et al., 2015. Computer-guided drug repurposing: Identification of trypanocidal activity of clofazimine, benidipine and saquinavir. *European Journal of Medicinal Chemistry*, 93, pp.338–348.
- Benaim, G. et al., 2006. Amiodarone has intrinsic anti-*Trypanosoma cruzi* activity and acts synergistically with posaconazole. *Journal of Medicinal Chemistry*, 49(3), pp.892–899.
- Benaim, G. et al., 2012. *In vitro* anti-*Trypanosoma cruzi* activity of dronedarone, a novel amiodarone derivative with an improved safety profile. *Antimicrobial agents and chemotherapy*, 56(7), pp.3720–3725.
- Benítez, D. et al., 2016. Identification of Novel Chemical Scaffolds Inhibiting Trypanothione Synthetase from Pathogenic Trypanosomatids. *PLoS Neglected Tropical Diseases*, 10(4), pp.1–25.
- Bermudez, J. et al., 2016. Current drug therapy and pharmaceutical challenges for Chagas disease. *Acta Tropica*, 156, pp.1–16.
- Bernstein, R.E., 1984. Darwin's illness: Chagas' disease resurgens. *Journal of the Royal Society of Medicine*, 77(7), pp.608–609.

- Birkholtz, L.M. et al., 2011. Polyamine homeostasis as a drug target in pathogenic protozoa: peculiarities and possibilities. *Biochemical Journal*, 438(2), pp.229–244.
- Bouvier, L.A. et al., 2004. Post genomic analysis of permeases from the amino acid/auxin family in protozoan parasites. *Biochemical and Biophysical Research Communications*, 321(3), pp.547–556.
- Bouzahzah, B. et al., 2005. Risedronate in the treatment of Murine Chagas' disease. *Parasitology Research*, 96(3), pp.184–187.
- Bray, P.G. et al., 2003. Pentamidine uptake and resistance in pathogenic protozoa: past, present and future. *Trends in parasitology*, 19(5), pp.232–239.
- Buckner, F.S. & Urbina, J.A., 2012. Recent developments in sterol 14-demethylase inhibitors for Chagas disease. *International Journal for Parasitology: Drugs and Drug Resistance*, 2, pp.236–242.
- Bustamante, J.M. et al., 2014. New, Combined, and Reduced Dosing Treatment Protocols Cure *Trypanosoma cruzi* Infection in Mice. *Journal of Infectious Diseases*, 209(1), pp.150–162.
- Cáceres, A.J. et al., 2003. Molecular and biochemical characterization of hexokinase from *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 126(2), pp.251–262.
- Caffrey, C.R. et al., 2000. Cysteine proteinases of trypanosome parasites: novel targets for chemotherapy. *Current drug targets*, 1(2), pp.155–162.
- Cançado, J.R., 1999. Criteria of Chagas Disease Cure. *Memórias do Instituto Oswaldo Cruz*, 94(Sup. 1), pp.331–335.
- Cardoso, R.N. et al., 2014. Chagas cardiomyopathy is associated with higher incidence of stroke: A meta-analysis of observational studies. *Journal of Cardiac Failure*, 20(12), pp.931–938.
- Carrillo, C. et al., 2003. Lack of arginine decarboxylase in *Trypanosoma cruzi* epimastigotes. *Journal of Eukaryotic Microbiology*, 50(5), pp.312–316.
- Carrillo, C. et al., 2006. Molecular and functional characterization of a spermidine transporter (TcPAT12) from *Trypanosoma cruzi*. *Biochemical and Biophysical Research Communications*, 344(3), pp.936–940.
- Carrillo, C. et al., 1999. *Trypanosoma cruzi* epimastigotes lack ornithine decarboxylase but can express a foreign gene encoding this enzyme. *FEBS Letters*, 454(3), pp.192–196.
- Carrillo, C., et al., 2007. *Trypanosoma cruzi* as a model system to study the expression of exogenous genes coding for polyamine biosynthetic enzymes. Induction of DFMO resistance in transgenic parasites. *Biochimica et Biophysica Acta - General Subjects*, 1770(12), pp.1605–1611.
- Cavalcanti, D.P. et al., 2008. The kinetoplast ultrastructural organization of endosymbiont-bearing trypanosomatids as revealed by deep-etching, cytochemical and immunocytochemical analysis. *Histochemistry and Cell Biology*, 130(6), pp.1177–1185.
- Cazzulo, J., 2002. Proteinases of *Trypanosoma Cruzi*: Potential Targets for the Chemotherapy of Chagas Disease. *Current Topics in Medicinal Chemistry*, 2(11), pp.1261–1271.
- Chagas, C., 1909. Nova tripanozomíaze humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. *Memórias do Instituto Oswaldo Cruz*, 1(2), pp.159–218.

- Coura, J.R., 2007. Chagas disease: what is known and what is needed--a background article. *Memorias do Instituto Oswaldo Cruz*, 102 (Sup., agosto), pp.113–122.
- Coura, J.R. & De Castro, S.L., 2002. A critical review on chagas disease chemotherapy. *Memorias do Instituto Oswaldo Cruz*, 97(1), pp.3–24.
- Crespillo-Andújar, C. et al., 2018. Toxicity of nifurtimox as second-line treatment after benznidazole intolerance in patients with chronic Chagas disease: When available options fail. *Clinical microbiology and infection*, 0(0).
- Croft, S., et al., 2005. Chemotherapy of trypanosomiasis and leishmaniasis. *Trends in Parasitology*, 21(11), pp.508–512.
- Dias, E. et al., 1956. Chagas' disease; a clinical, epidemiologic, and pathologic study. *Circulation*, 14(6), pp.1035–1060.
- Dias, J.C.P., 1989. The indeterminate form of human chronic Chagas' disease: a clinical epidemiological review. *Revista da Sociedade Brasileira de Medicina Tropical*, 22(3), pp.147–156.
- Diniz, L. d. F. et al., 2010. Effects of Ravuconazole Treatment on Parasite Load and Immune Response in Dogs Experimentally Infected with *Trypanosoma cruzi*. *Antimicrobial Agents and Chemotherapy*, 54(7), pp.2979–2986.
- DNDi, 2013. Drug Trial for Leading Parasitic Killer of the Americas Shows Mixed Results but Provides New Evidence for Improved Therapy.
- Duschak, V.G. & Couto, A.S., 2009. Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Current medicinal chemistry*, 16(24), pp.3174–3202.
- Eakin, A.E. et al., 1997. Hypoxanthine phosphoribosyltransferase from *Trypanosoma cruzi* as a target for structure-based inhibitor design: crystallization and inhibition studies with purine analogs. *Antimicrobial agents and chemotherapy*, 41(8), pp.1686–1692.
- Fairlamb, A. et al., 1985. Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. *Science*, 227(4693), pp.1485–1487.
- Fairlamb, A.H., 2002. Metabolic pathway analysis in trypanosomes and malaria parasites. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 357(1417), pp.101–107.
- Fairlamb, A.H., 1989. Novel biochemical pathways in parasitic protozoa. *Parasitology*, 99(S1), pp.S93–S112.
- Fairlamb, A.H. & Cerami, A., 1992. Metabolism and Functions of Trypanothione in the Kinetoplastida. *Annual Review of Microbiology*, 46(1), pp.695–729.
- Ferreira, L.F., Jansen, A.M. & Araújo, A., 2011. Chagas disease in prehistory. *Anais da Academia Brasileira de Ciências*, 83(3), pp.1041–1044.
- De Fuentes-Vicente, J.A. et al., 2018. What makes an effective Chagas disease vector? Factors underlying *Trypanosoma cruzi*-triatomine interactions. *Acta Tropica*, 183, pp.23–31.
- de Souza, W. De et al., 2010. Review on *Trypanosoma cruzi*: Host Cell Interaction. *International journal of cell biology*, 2010, pp.1–19.
- Garzoni, L.R. et al., 2004a. Antiparasitic activity of risnedronate in a murine model of acute Chagas' disease. *International journal of antimicrobial agents*, 23(3), pp.286–290.

- Garzoni, L.R. et al., 2004b. Selective *in vitro* effects of the farnesyl pyrophosphate synthase inhibitor risedronate on *Trypanosoma cruzi*. *International Journal of Antimicrobial Agents*, 23(3), pp.273–285.
- Gashaw, I. et al., 2011. What makes a good drug target? *Drug Discovery Today*, 16(23–24), pp.1037–1043.
- Gevrekci, A.Ö., 2017. The roles of polyamines in microorganisms. *World Journal of Microbiology and Biotechnology*, 33: 204.
- Gilbert, I.H., 2002. Inhibitors of dihydrofolate reductase in leishmania and trypanosomes. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1587(2–3), pp.249–257.
- Gonçalves, C.S. et al., 2018. Revisiting the *Trypanosoma cruzi* metacyclogenesis: Morphological and ultrastructural analyses during cell differentiation. *Parasites and Vectors*, 11: 83.
- González, N.S. et al., 1992. Differential regulation of putrescine uptake in *Trypanosoma cruzi* and other trypanosomatids. *Biochemical and Biophysical Research Communications*, 188(1), pp.120–128.
- Gürtlera, R.E. & Yadon, Z.E., 2014. Eco-bio-social research on community-based approaches for Chagas disease vector control in Latin America. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 109(2), pp.91–98.
- Handa, A.K. et al., 2018. Polyamines: Bio-Molecules with Diverse Functions in Plant and Human Health and Disease. *Frontiers in Chemistry*, 6: 10.
- Hasne, M.P. et al., 2010. A high-affinity putrescine-cadaverine transporter from *Trypanosoma cruzi*. *Molecular Microbiology*, 76(1), pp.78–91.
- Hasne, M.P. et al., 2016. The *Trypanosoma cruzi* diamine transporter is essential for robust infection of mammalian cells. *PLoS ONE*, 11(4), e0152715.
- Hotez, P.J., 2011. *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases*.
- Hudock, M.P. et al., 2006. Inhibition of *Trypanosoma cruzi* hexokinase by bisphosphonates. *Journal of Medicinal Chemistry*, 49(1), pp.215–223.
- Igarashi, K. et al., 1974. Effect of Polyamines on Polyphenylalanine Synthesis by Escherichia coli and Rat-Liver Ribosomes. *European Journal of Biochemistry*, 48(2), pp.495–502.
- Igoillo-Esteve, M. et al., 2007. The pentose phosphate pathway in *Trypanosoma cruzi*: a potential target for the chemotherapy of Chagas disease. *An Acad Bras Cienc*, 79(4), pp.649–663.
- Jackson, Y. et al., 2010. Tolerance and Safety of Nifurtimox in Patients with Chronic Chagas Disease. *Clinical Infectious Diseases*, 51(10), pp.e69–e75
- Kashif, M. et al., 2017. Recent developments in trans-sialidase inhibitors of *Trypanosoma cruzi*. *Journal of Drug Targeting*, 25(6), pp.485–498.
- Kaufers, A. et al., 2017. The evolution of trypanosomatid taxonomy. *Parasites and Vectors*, 10(1): 287.
- Kessler, R.L. et al, 2017. Recently differentiated *Trypanosoma cruzi* epimastigotes are infective to the mammalian host. *Molecular Microbiology*, 104(5), pp.712–736.
- Kessler, R.L. et al., 2013. *Trypanosoma cruzi* Response to Sterol Biosynthesis Inhibitors: Morphophysiological Alterations Leading to Cell Death. *PLoS ONE*, 8(1).

- Kovářová, J. & Barrett, M.P., 2016. The Pentose Phosphate Pathway in Parasitic Trypanosomatids. *Trends in Parasitology*, 32(8), pp.622–634.
- Krauth-Siegel, R.L. & Comini, M.A., 2008. Redox control in trypanosomatids, parasitic protozoa with trypanothione-based thiol metabolism. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1780(11), pp.1236–1248.
- Lakhdar-Ghazal, F. et al., 2002. Glycolysis and proteases as targets for the design of new anti-trypanosome drugs. *Current topics in medicinal chemistry*, 2(5), pp.439–456.
- Leite Dias, J.V. et al., 2016. Spatial distribution of triatomines in domiciles of an urban area of the Brazilian Southeast region. *Memorias do Instituto Oswaldo Cruz*, 111(1), pp.43–50.
- Maldonado, R.A. et al., 1993. Experimental chemotherapy with combinations of ergosterol biosynthesis inhibitors in murine models of Chagas' disease. *Antimicrobial agents and chemotherapy*, 37(6), pp.1353–1359.
- Marr, J.J., 1991. Purine analogs as chemotherapeutic agents in leishmaniasis and American trypanosomiasis. *The Journal of laboratory and clinical medicine*, 118(2), pp.111–119.
- Martin, M.B. et al., 2001. Bisphosphonates inhibit the growth of *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, and *Plasmodium falciparum*: a potential route to chemotherapy. *Journal of medicinal chemistry*, 44(6), pp.909–916.
- Maya, J.D. et al., 2014. Key Proteins in the Polyamine-Trypanothione Pathway as Drug Targets Against *Trypanosoma cruzi*. *Current Medicinal Chemistry*, 21(15), pp.1757–1771.
- Mazza, S., 1949. La enfermedad de Chagas en la Republica Argentina. *Memórias do Instituto Oswaldo Cruz*, 47(1–2), pp.273–302.
- McKerrow, J.H. et al., 2009. Two approaches to discovering and developing new drugs for Chagas disease. *Memorias do Instituto Oswaldo Cruz*, 104 Suppl 1(0 1), pp.263–269.
- Meymandi, S.K. et al., 2017. Prevalence of chagas disease in the Latin American-born population of Los Angeles. *Clinical Infectious Diseases*, 64(9), pp.1182–1188.
- Miller-Fleming, L. et al., 2015. Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *Journal of Molecular Biology*, 427(21), pp.3389–3406.
- Miller III, B.R. & Roitberg, A.E., 2013. *Trypanosoma cruzi* trans -sialidase as a drug target against Chagas disease (American trypanosomiasis). *Future Medicinal Chemistry*, 5(15), pp.1889–1900.
- Molina, I. et al., 2014. Randomized Trial of Posaconazole and Benznidazole for Chronic Chagas' Disease. *New England Journal of Medicine*, 370(20), pp.1899–1908.
- Molina, J. et al., 2000. Activities of the triazole derivative SCH 56592 (posaconazole) against drug-resistant strains of the protozoan parasite *Trypanosoma (Schizotrypanum) cruzi* in immunocompetent and immunosuppressed murine hosts. *Antimicrobial agents and chemotherapy*, 44(1), pp.150–155.
- Moraes, C.B. et al., 2015. Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against *Trypanosoma cruzi*: implications for Chagas disease drug discovery and development. *Scientific Reports*, 4(1), p.4703.
- Morillo, C.A. et al., 2017. Benznidazole and Posaconazole in Eliminating Parasites in Asymptomatic *T. Cruzi* Carriers. *Journal of the American College of Cardiology*, 69(8),

- pp.939–947.
- Morillo, C.A. et al., 2015. Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy. *New England Journal of Medicine*, 373(14), pp.1295–1306.
- Nunes, M.C.P. et al., 2009. Ischemic cerebrovascular events in patients with Chagas cardiomyopathy: A prospective follow-up study. *Journal of the Neurological Sciences*, 278(1–2), pp.96–101.
- Pereira, C.A. & Carrillo, C., 2013. Transport of essential metabolites in trypanosomatids. Libro: *Parasites: Ecology, Management and Diseases*. pp.43–60.
- Pérez-Molina, J.A. & Molina, I., 2018. Chagas disease. *The Lancet*, 391(10115), pp.82–94.
- Perez, C.J., et al., 2014. Chagas disease: The challenge of polyparasitism? *Trends in Parasitology*, 30(4), pp.176–182.
- Perin, L. et al., 2017. Pharmacokinetics and Tissue Distribution of Benznidazole after Oral Administration in Mice. *Antimicrobial agents and chemotherapy*, 61(4).
- Pez, D. et al., 2003. 2,4-Diaminopyrimidines as inhibitors of Leishmanial and Trypanosomal dihydrofolate reductase. *Bioorganic & Medicinal Chemistry*, 11(22), pp.4693–4711.
- Pinheiro, E. et al., 2017. Chagas disease: Review of needs, neglect, and obstacles to treatment access in Latin America. *Revista da Sociedade Brasileira de Medicina Tropical*, 50(3), pp.296–300.
- Le Quesne, S. & Fairlamb, A., 1996. Regulation of a high-affinity diamine transport system in *Trypanosoma cruzi* epimastigotes. *The Biochemical Journal*, 316, pp.481–486.
- Racagni, G.E. & Machado de Domenech, E.E., 1983. Characterization of *Trypanosoma cruzi* hexokinase. *Molecular and Biochemical Parasitology*, 9(2), pp.181–188.
- Rajão, M.A. et al., 2014. Unveiling Benznidazole's mechanism of action through overexpression of DNA repair proteins in *Trypanosoma cruzi*. *Environmental and Molecular Mutagenesis*, 55(4), pp.309–321.
- Rassi, A. Jr., et al, 2012. American Trypanosomiasis (Chagas Disease). *Infectious Disease Clinics of North America*, 26(2), pp.275–291.
- Russomando, G. et al., 1998. Treatment of congenital chagas' disease diagnosed and followed up by the polymerase chain reaction. *American Journal of Tropical Medicine and Hygiene*, 59(3), pp.487–491.
- Sabino, E.C. et al., 2013. Ten-Year incidence of chagas cardiomyopathy among asymptomatic *Trypanosoma cruzi*-seropositive former blood donors. *Circulation*, 127(10), pp.1105–1115.
- Sales Junior, P.A. et al., 2017. Experimental and Clinical Treatment of Chagas Disease: A Review. *The American journal of tropical medicine and hygiene*, 97(5), pp.1289–1303.
- Salvador, F. et al., 2014. *Trypanosoma cruzi* infection in a non-endemic country: Epidemiological and clinical profile. *Clinical Microbiology and Infection*, 20(7), pp.706–712.
- Sanmartino, M., 2009. 100 años de Chagas (1909-2009): revisión, balance y perspectiva. *Rev Soc Entomol Argent*, 68, pp.243–252.
- Sbaraglini, M.L. et al., 2016a. Neglected Tropical Protozoan Diseases: Drug Repositioning as a

- Rational Option. *Current topics in medicinal chemistry*, 16(19), pp.2201–2222.
- Sbaraglini, M.L. et al., 2016b. Novel cruzipain inhibitors for the chemotherapy of chronic Chagas disease. *International Journal of Antimicrobial Agents*, 48(1), pp.91–95.
- Schijman, A.G. et al., 2003. Aetiological treatment of congenital Chagas' disease diagnosed and monitored by the polymerase chain reaction. *Journal of Antimicrobial Chemotherapy*, 52(3), pp.441–449.
- Smith, J. & Taylor, E.M., 2016. What Is Next for NTDs in the Era of the Sustainable Development Goals? *PLoS Neglected Tropical Diseases*, 10(7): e0004719.
- Soeiro, M.N.C. & de Castro, S.L., 2009. *Trypanosoma cruzi* targets for new chemotherapeutic approaches. *Expert Opinion on Therapeutic Targets*, 13(1), pp.105–121.
- Souza, D.H. et al., 1998. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. *FEBS Letters*, 424(3), pp.131–135.
- Tabor, C.W. & Tabor, H., 1976. 1,4-Diaminobutane (Putrescine), Spermidine, and Spermine. *Annual Review of Biochemistry*, 45(1), pp.285–306.
- Talevi, A. et al., 2018. The thiol-polyamine metabolism of *Trypanosoma cruzi*: molecular targets and drug repurposing strategies. *Current Medicinal Chemistry*, 25: 1.
- Ullman, B. & Carter, D., 1995. Hypoxanthine-guanine phosphoribosyltransferase as a therapeutic target in protozoal infections. *Infectious agents and disease*, 4(1), pp.29–40.
- Urbina, J.A. et al., 1998. Antiproliferative effects and mechanism of action of SCH 56592 against *Trypanosoma (Schizotrypanum) cruzi*: *in vitro* and *in vivo* studies. *Antimicrobial agents and chemotherapy*, 42(7), pp.1771–1777.
- Urbina, J.A., 2009. Ergosterol biosynthesis and drug development for Chagas disease. *Memorias do Instituto Oswaldo Cruz*, 104(SUPPL. 1), pp.311–318.
- Urbina, J.A. et al., 2003. *In vitro* and *in vivo* activities of ravuconazole on *Trypanosoma cruzi*, the causative agent of Chagas disease. *International journal of antimicrobial agents*, 21(1), pp.27–38.
- Urbina, J.A. et al., 1993. Mevinolin (lovastatin) potentiates the antiproliferative effects of ketoconazole and terbinafine against *Trypanosoma (Schizotrypanum) cruzi*: *in vitro* and *in vivo* studies. *Antimicrobial agents and chemotherapy*, 37(3), pp.580–591.
- Urbina, J.A., 2010. Specific chemotherapy of Chagas disease: Relevance, current limitations and new approaches. *Acta Tropica*, 115(1–2), pp.55–68.
- Vanrell, M.C. et al., 2013. Polyamine depletion inhibits the autophagic response modulating *Trypanosoma cruzi* infectivity. *Autophagy*, 9(7), pp.1080–1093.
- Verlinda, C.L.M.J. et al., 2001. Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resistance Updates*, 4(1), pp.50–65.
- Viotti, R. et al., 2006. Long-term cardiac outcomes of treating chronic Chagas disease with benznidazole versus no treatment: a nonrandomized trial. *Annals of internal medicine*, 144(10), pp.724–34.
- Wilson, W.D. et al., 2005. Dications that target the DNA minor groove: compound design and preparation, DNA interactions, cellular distribution and biological activity. *Current medicinal chemistry. Anti-cancer agents*, 5(4), pp.389–408.

World Health Organization, 2015. Chagas disease in Latin America: an epidemiological update based on 2010 estimates. *Weekly Epidemiological Record*, (6), pp.33–44.

World Health Organization, 2010. Working to overcome the global impact of neglected tropical diseases First WHO report on neglected tropical diseases. *World Health*, 86(13), p.186.

Zuccotto, F. et al., 2001. Novel inhibitors of *Trypanosoma cruzi* dihydrofolate reductase. *European Journal of Medicinal Chemistry*, 36(5), pp.395–405.

CAPÍTULO 2

DESCUBRIMIENTO DE FÁRMACOS



2.1. INTRODUCCIÓN

El descubrimiento y desarrollo de fármacos involucra la interrelación entre diferentes disciplinas como la biología, la química, el modelado matemático, las ciencias farmacéuticas y las ciencias de la computación, entre otras, para proveer drogas más efectivas y seguras. Esta tesis doctoral está enfocada en el campo de la química medicinal, una disciplina que surge de la interacción entre la química y la farmacología, y cuyo objetivo fundamental es la identificación, síntesis y desarrollo de nuevos compuestos químicos que puedan utilizarse para la obtención de nuevos productos farmacéuticos (Alagarsamy 2010; Zulfiqar et al. 2017).

Introducir una nueva droga en el mercado es un proceso que requiere varios años, es económicamente muy costoso y requiere la participación de diversos especialistas (Macalino et al. 2015). El costo total de llevar la nueva droga al mercado es difícil de estimar y varía considerablemente de acuerdo al tipo de droga; sin embargo, en todos los casos este costo es considerablemente alto. Algunos informes han reportado que introducir un fármaco novedoso puede costar hasta 2500 millones de dólares (Tufts CSDD 2014), aunque otros análisis más conservadores sugieren que ese valor está sobreestimado y fijan el valor real en unos 1800 millones de dólares (Paul et al. 2010). El tiempo promedio que se requiere desde el descubrimiento del principio activo hasta el lanzamiento oscila entre los 12 y 15 años (Macalino et al. 2015; Terry & Lesser 2015). Adicionalmente, existe un riesgo muy grande de fracaso. Las bibliotecas de compuestos de *screening* iniciales (entre 10^4 - 10^6 compuestos), conducen a un único compuesto que tiene alrededor de solo 8% de probabilidad de superar exitosamente las pruebas clínicas (Cook et al. 2014).

Como se resume en la **Figura 2.1**, cualquier proceso de desarrollo de drogas debe atravesar varias etapas para lograr un producto que sea seguro, eficaz y que supere todos los requerimientos regulatorios (Mohs & Greig 2017).

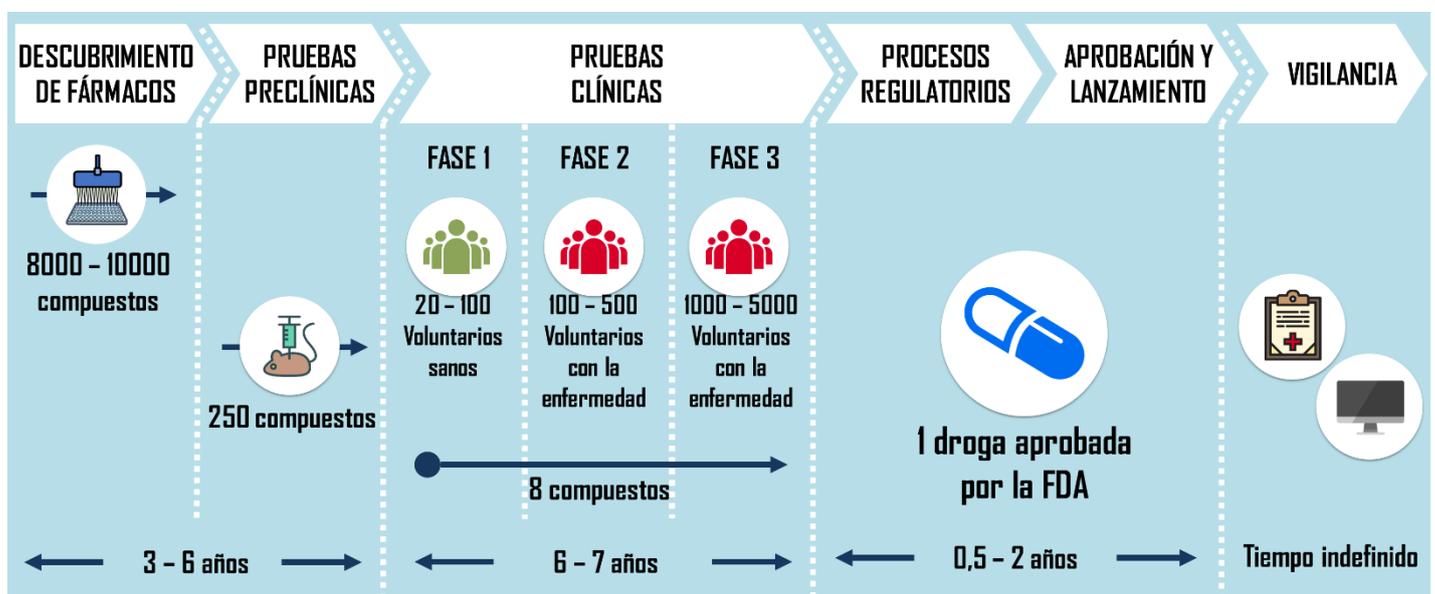


Figura 2.1: Esquema general del proceso de descubrimiento y desarrollo de un fármaco. Adaptado de GSK (<http://us.gsk.com/en-us/research/how-we-discover-new-products/how-we-discover-new-medicines/>).

La primera etapa es generalmente llamada **Descubrimiento de fármacos** (en inglés, *Drug discovery*), y comprende desde la **Identificación y Validación del Blanco Molecular**, hasta la identificación de sustancias químicas que logren inhibir el blanco seleccionado. Las técnicas de

Cribado Farmacológico de Alto Rendimiento (*High Throughput Screening*, HTS) permiten evaluar con rapidez decenas a cientos de miles de compuestos frente a un blanco molecular; adicionalmente, se pueden utilizar técnicas computacionales (*in silico*) para seleccionar de una manera más racional qué compuestos evaluar experimentalmente. Las moléculas que demuestren actividad satisfactoria frente al blanco molecular son denominadas “*hits*”. Posteriormente, los *hits* son optimizados para mejorar su potencia y otras propiedades significativas a través de ciclos de evaluación, para conducir a los “**Compuestos líderes**” (*lead compounds*) que pueden ser subsecuentemente optimizados (**Optimización del líder**). Aquí también se pueden utilizar las estrategias *in silico* para determinar qué características deberían tener los análogos del *hit* o el líder, según el caso, para mejorar su actividad. En esta última etapa es habitual interrogar distintos modelos animales (**Ensayos preclínicos**) para guiar la optimización molecular. Los estudios en animales permiten evaluar potencia y toxicidad en sistemas complejos, y comenzar la caracterización farmacocinética *in vivo*. Finalmente, alcanzados ciertos estándares de desempeño, el mejor candidato es sometido a **Pruebas clínicas** para determinar la seguridad, los efectos adversos, las dosis adecuadas, la eficacia y las propiedades farmacocinéticas y farmacológicas de la droga candidata en humanos (Macalino et al. 2015; Hughes et al. 2011).

2.1.1. DESCUBRIMIENTO DE FÁRMACOS PARA ETDs

El descubrimiento de fármacos para las ETDs se propone varios objetivos como el desarrollo de nuevos tratamientos, la reducción de la toxicidad en el hospedador, las mejoras en los regímenes de administración y el desarrollo de terapias combinadas (Field et al. 2017).

Se pueden diferenciar tres grandes enfoques para el descubrimiento de fármacos para el tratamiento de enfermedades causadas por **tripanosomátidos**:

a) **Enfoques basados en el blanco molecular**: Consiste en el cribado de una biblioteca de compuestos contra un blanco molecular validado y la posterior optimización de los compuestos para mejorar la potencia y la selectividad frente al blanco, la actividad celular y las propiedades farmacocinéticas. Las compañías farmacéuticas son capaces de evaluar rápidamente mediante HTS miles de drogas contra una proteína específica (Gilbert 2013).

b) **Enfoques fenotípicos**: Se basan en el cribado de compuestos contra el parásito intacto para encontrar directamente inhibidores del crecimiento, generalmente en un cultivo *in vitro*. En esta estrategia no es necesario conocer o postular el mecanismo de acción por el cual el compuesto ejerce su actividad; el mismo puede investigarse más adelante. También se pueden aplicar HTS para evaluar bibliotecas muy grandes de compuestos en los parásitos intactos (Gilbert 2013).

c) **Enfoques basados en el reposicionamiento de fármacos**: Implican encontrar nuevas indicaciones terapéuticas para drogas previamente desarrolladas; en el caso de los tripanosomátidos, consiste en encontrar drogas ya conocidas para utilizarlas como terapias tripanocidas (Field et al. 2017).

Estos 3 enfoques pueden complementarse con estrategias *in silico* para reducir el número de compuestos a evaluar o para guiar la optimización de los compuestos encontrados. En el siguiente punto nos centraremos en estas estrategias.

En la **Tabla 2.1** se resumen los costos y tiempos promedios que son necesarios hasta la aprobación de un medicamento por la Food and Drug Administration (FDA) cuando se recurre a las

estrategias de descubrimiento anteriormente descritas (Mignani et al. 2016). Puede observarse que la utilización del reposicionamiento de fármacos es más eficiente en cuanto al tiempo y dinero requerido.

Tabla 2.1: Promedio de costos y tiempos de aprobación de la FDA según las estrategias de descubrimiento de medicamentos.

	Enfoques basados en el blanco	Enfoques fenotípicos	Reposicionamiento de fármacos
Tiempo hasta aprobación por la FDA	≈ 10 - 12 años	≈ 10 - 12 años	≈ 2 - 3 años
Costos	≈ 1000 millones US\$	≈ 1000 millones US\$	≈ 10 millones US\$

Con respecto a la experiencia de la industria farmacéutica en el descubrimiento de fármacos mediante la utilización de los dos primeros enfoques, se han realizado estudios para analizar cuál método ha sido más exitoso en los últimos años.

Un estudio de Swinney y Anthony (Swinney & Anthony 2011), que se basó en el número de Nuevas Entidades Moleculares (NEMs) aprobadas entre 1999 y 2008, demostró que el cribado fenotípico fue la estrategia más exitosa para encontrar fármacos considerados primeros de su tipo (*“first in-class”*) mientras que el cribado basado en el blanco fue el más eficiente en encontrar fármacos *“follow-on”* durante el periodo establecido. Vale aclarar que los autores designaron como drogas *“first in-class”* a aquellas que modulaban un blanco molecular o una vía terapéutica novedosa, es decir, sin precedentes (hasta ese momento), y drogas *“follow-on”* a aquellas moléculas que actúan por el mismo mecanismo que una NEM aprobada previamente.

Un estudio posterior, que se realizó sobre las 113 drogas *“first in-class”* que fueron aprobadas por la FDA entre 1999 y 2013, demostró que la mayoría habían sido descubiertas usando un enfoque basado en el blanco (71%), en tanto aquellas identificadas mediante aproximaciones menos reduccionistas (*screening* fenotípico y aproximaciones quimio-céntricas) representan un 29% del total (Eder et al. 2014). La incidencia del cribado fenotípico en el descubrimiento de moléculas primeras en su clase se vuelve más relevante si excluimos del análisis a productos biotecnológicos y nos limitamos a moléculas pequeñas tipo fármaco.

La **Figura 2.2** representa un resumen de los análisis realizados sobre las drogas *“first in-class”* aprobadas y seguidas por la FDA entre 1999 y 2013. Puede notarse que en el periodo 1999 - 2008 el cribado fenotípico fue la estrategia más exitosa en la búsqueda de drogas *“first in-class”*, mientras que el cribado basado en el blanco fue el más eficiente para encontrar drogas *“follow-on”*. Además, desde 1999 hasta 2013 la mayoría de las NEMs han sido descubiertas usando métodos basados en el blanco.

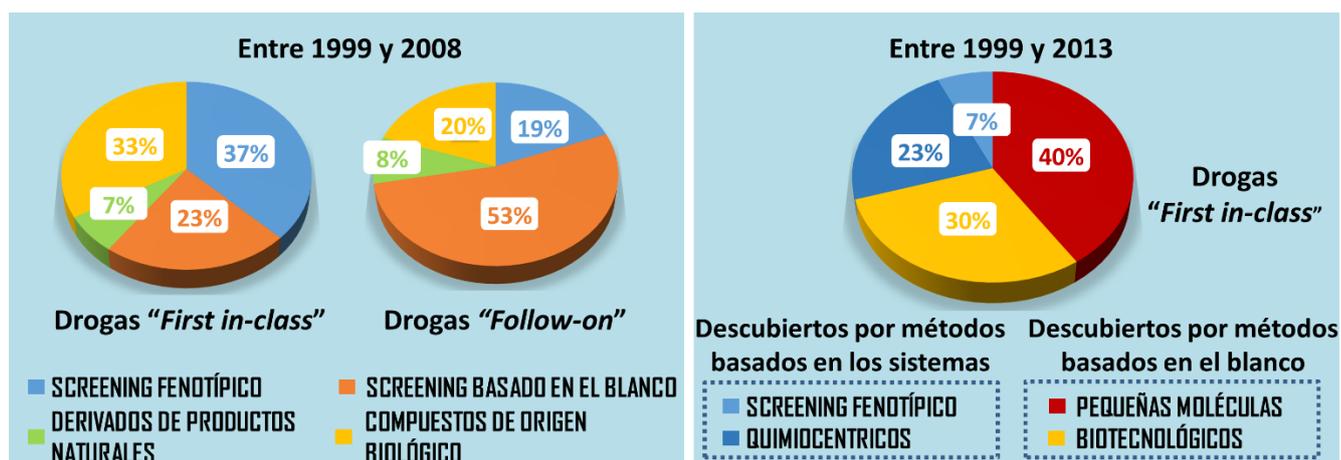


Figura 2.2: Porcentaje de evolución de las NEMs en cada categoría de las clases de drogas entre 1999 y 2013, basándose en drogas aprobadas "first in-class" y "follow-on". Los datos fueron extraídos de Swinney & Anthony 2011, y Eder et al 2014.

2.2. CRIBADO VIRTUAL

La identificación de nuevos hits contra un blanco molecular y la optimización progresiva de la potencia y de las propiedades farmacológicas de dichos compuestos son los puntos clave en el descubrimiento temprano de fármacos. Para ello, la industria farmacéutica ha adoptado por el HTS de grandes bibliotecas de compuestos químicos contra blancos terapéuticos relevantes, como medio para identificar nuevos andamios moleculares activos en tiempos breves (Lionta et al. 2014). El HTS combina la robotización, la miniaturización de ensayos y el procesamiento de datos a gran escala para identificar con rapidez y con pequeñas cantidades de muestra compuestos que modulen un blanco molecular concreto (Cronk 2012; Talevi & Bruno-Blanch 2009). Sin embargo, el elevado costo de las tecnologías necesarias para realizar HTS limita su utilización, mayormente, a compañías farmacéuticas y algunos laboratorios públicos de países desarrollados.

Para superar esta limitación, se utiliza una metodología más racional y económica, que puede encontrarse con varios nombres como lo son: **Screening Virtual** (en adelante, **SV**), **Tamizado Virtual** y **Cribado Virtual**. Walters y colaboradores describen al SV como el uso de la informática de alto rendimiento para analizar grandes bases de datos de compuestos químicos con el fin de identificar posibles candidatos a drogas (Walters et al. 1998). Talevi y Bruno-Blanch definen el SV como "conjunto de técnicas computacionales que permiten, a partir de representaciones de la estructura molecular de los compuestos químicos almacenados en grandes bases de datos, identificar compuestos potencialmente interesantes desde el punto de vista farmacológico" (Talevi & Bruno-Blanch 2009).

Esta estrategia racional e *in silico* (realizada por computadora) presenta varias ventajas con respecto al HTS:

- Menor inversión necesaria:** Existe una gran cantidad de bases de datos públicas de estructuras químicas. Entre ellas se encuentran ChEMBL (Davies et al. 2015; Gaulton et al. 2017), que compila más de 1,8 millones de compuestos químicos y más de 15 millones de datos de actividad biológica (junio 2018); PubChem (Kim et al. 2016), con más de 96 millones de compuestos y más de 230 millones de datos de bioactividad (junio

2018), DrugBank (Law et al. 2014; Wishart et al. 2018) con alrededor de 10000 compuestos indexados de los cuales cerca de 9000 son pequeñas moléculas tipo fármaco y el resto productos biotecnológicos; ZINC15 (Sterling & Irwin 2015), que contiene más de 700 millones de estructuras químicas. Adicionalmente, muchos softwares para la obtención de modelos y para el VS son de libre acceso o se pueden conseguir licencias académicas gratuitas o de costo reducido, tales como PaDEL (Yap 2011), Chimera (Pettersen et al. 2004), Open Babel (O'Boyle et al. 2011) y los diferentes paquetes de ChemAxon (<http://www.chemaxon.com>). También existen servidores de la red o webservers de libre acceso que permiten realizar campañas de SV en línea (González-Díaz et al. 2011; Schneidman-Duhovny et al. 2008; Liu et al. 2010; Konc et al. 2015). Por último, dependiendo de la estrategia *in silico* que se fuese a utilizar y de la cantidad de compuestos a tamizar, es posible realizar el SV con computadoras de uso cotidiano. Como se puede notar, realizar un SV no requiere una inversión grande y, en este sentido, supera ampliamente a la utilización de HTS, que en todos los casos requiere tecnología de costoso acceso.

- b) **Carácter teórico:** Al realizarse en una computadora, no es necesario disponer físicamente de las muestras de compuestos de la biblioteca química a analizar, lo que también disminuye los tiempos y costos del SV. Además, es posible incorporar a la biblioteca estructuras hipotéticas que no hayan sido sintetizadas aún. Por lo tanto, solo se adquirirán o sintetizarán los compuestos con mejor desempeño en los modelos (Talevi et al. 2011).
- c) **Reposicionamiento de fármacos:** Debido a que existen varias bases de datos públicas que reúnen fármacos aprobados, discontinuados o que se encuentran atravesando estudios clínicos, tales como DrugBank (Wishart et al. 2018) y Sweetlead (Novick et al. 2013), es posible cribarlas mediante SV para encontrar nuevas indicaciones terapéuticas para fármacos existentes.
- d) **Ventajas bioéticas:** La utilización de las metodologías *in silico* reduce la cantidad de compuestos a evaluar tanto *in vitro* como *in vivo* y aumenta las probabilidades de resultados positivos en esas instancias. Esto conlleva a la realización de menor cantidad de ensayos preclínicos lo que se traduce en una minimización en la cantidad de animales a utilizar (Talevi & Bruno-Blanch 2009).

En la **Figura 2.3** puede apreciarse cómo ha aumentado el uso de las metodologías de SV en los últimos 20 años. Para realizar este análisis se utilizaron los buscadores PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>) y Scopus (<https://www.scopus.com/home.uri>), buscando la palabra clave “*virtual screening*” en resúmenes y títulos de trabajos científicos en los últimos 20 años.

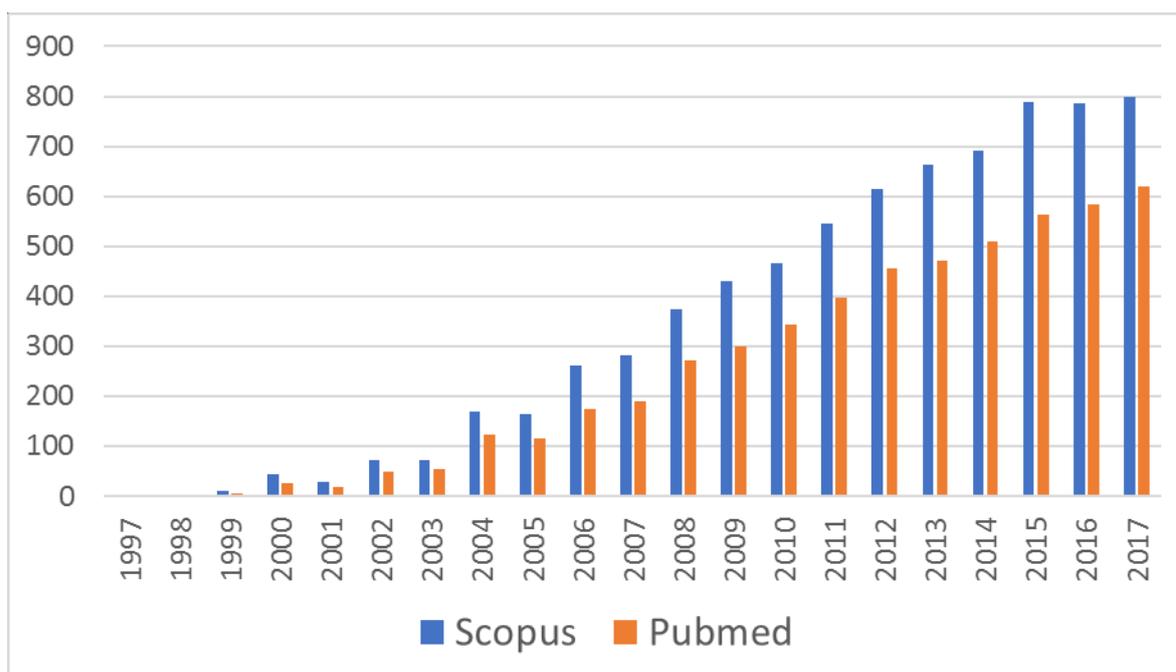


Figura 2.3: Número de artículos publicados en revistas científicas indexadas por Scopus y PubChem y que contienen el término "virtual screening" en su título y/o resumen. La búsqueda se acotó a los últimos 20 años.

En general, las estrategias para el diseño racional de fármacos asistido por computadora pueden ser clasificadas en dos categorías: **Métodos basados en el Ligando (o métodos indirectos)** y **Métodos basados en el Blanco molecular o en la Estructura (o métodos directos)**. Durante los últimos años se ha recurrido frecuentemente a la combinación de diferentes estrategias de SV basadas en el blanco molecular y en el ligando, ya que ambas aproximaciones son complementarias en términos de fortalezas y debilidades (Sotriffer 2009; Kumar & Zhang 2015).

2.2.1. MÉTODOS BASADOS EN EL BLANCO MOLECULAR

Como su nombre lo indica, en los métodos basados en el blanco molecular es indispensable conocer *a priori* (o poder modelar comparativamente) la estructura tridimensional del blanco molecular. Para obtener las estructuras tridimensionales de las proteínas pueden utilizarse diferentes métodos experimentales como cristalografía de rayos X, resonancia magnética nuclear (RMN) y microscopía electrónica (Ilari & Savino 2008; Guerry & Herrmann 2011). También es posible predecir la estructura 3D de las proteínas utilizando técnicas de modelado por homología de secuencias y de estructuras (Sotriffer 2009; Talevi et al. 2011).

Los métodos desde el blanco molecular pueden ser utilizados con dos enfoques diferentes: por un lado, para diseñar nuevos fármacos, utilizando la información disponible sobre las interacciones clave de ligandos conocidos del blanco molecular, proponiendo nuevos compuestos que permitan mejorar esas interacciones o generar interacciones nuevas. Por otro lado, se puede utilizar esta aproximación para el SV de bases de datos, analizando qué ligandos conseguirían unirse al sitio de unión elegido con energías de interacción adecuadas.

Adicionalmente, el conocimiento del sitio de unión al ligando en una estructura macromolecular 3D puede ser utilizado para buscar proteínas que contengan un sitio de unión similar,

identificando entre el conjunto de ligandos de estas últimas compuestos que potencialmente interactúen con el blanco molecular de interés (Ehrt et al. 2016).

2.2.1.1. DOCKING MOLECULAR

El método de SV basado en la estructura más ampliamente utilizado es, sin dudas, el *docking molecular* (o acoplamiento molecular). Este método permite simular y predecir las interacciones y los modos de unión de un ligando en el sitio de unión del blanco molecular (Macalino et al. 2015).

Las primeras explicaciones de los mecanismos de unión ligando-receptor se basaron en la teoría de la llave y la cerradura formulada por Fischer (Fischer 1894), en la cual el ligando se acopla dentro del receptor al igual que una llave en una cerradura y se consideran a ambas partes como cuerpos rígidos complementarios. Posteriormente, Koshland, con su teoría del ajuste inducido, expandió la teoría de Fisher, considerando que tanto el blanco molecular como el ligando inducen recíprocamente cambios conformacionales que mejoran el acople (Koshland 1963). Por lo tanto, esta teoría implica que el ligando y el receptor no deben ser tratados como cuerpos rígidos, sino como cuerpos flexibles (Hammes 2002).

El objetivo del *docking molecular* es predecir computacionalmente la estructura del complejo ligando-receptor. El *docking* involucra dos pasos: en primer lugar, la búsqueda de conformaciones del ligando en el sitio activo de la proteína; luego, el ranqueo de todas estas conformaciones mediante una función de puntaje (o *score*). Idealmente, los algoritmos de *docking* deberían ser capaces de reproducir el modo de unión experimental de un ligando, y la función de puntaje debería rankear a esa conformación como la mejor puntuada entre todas las restantes (Meng et al. 2011). En la **Figura 2.4** se muestra un esquema resumido de la utilización del *docking molecular* en campañas de SV. Básicamente, una vez que se tiene la proteína cristalizada se procede a la identificación del sitio de unión. Si la proteína fue cristalizada con algún ligando, este último se remueve. Luego, se introducen secuencialmente al sitio de unión las moléculas que se desean “dockear”. Se procede entonces a un barrido conformacional (habitualmente estocástico) y las diferentes conformaciones obtendrán un score. Luego se comparan los scores de las diferentes moléculas.

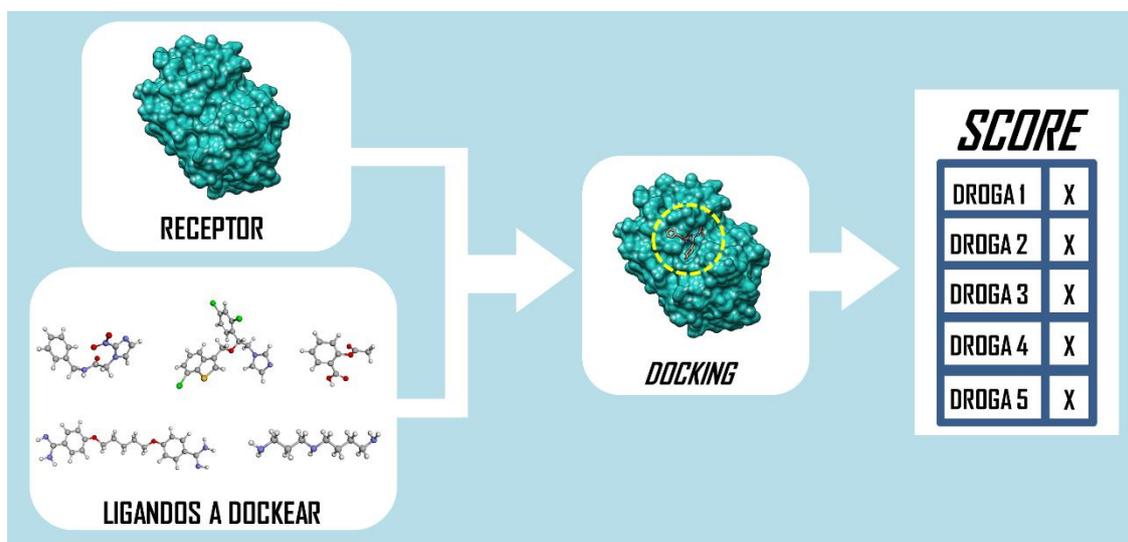


Figura 2.4: Esquema resumido de la aplicación del *docking molecular* a una campaña de SV.

Los procedimientos de *docking* habitualmente han sido realizados considerando ligandos flexibles y los blancos moleculares como cuerpos rígidos (Perola et al. 2004; Rarey et al. 1996; Moitessier et al. 2008). Sin embargo, existen estrategias de *docking* que permiten asignar flexibilidad a algunos residuos de la proteína (en especial los residuos del sitio activo involucrados en la interacción con el ligando) (Mirzaei et al. 2015; Wong 2015).

2.2.2. MÉTODOS BASADOS EN EL LIGANDO

En los casos donde no se encuentra disponible la estructura tridimensional del blanco molecular o no es posible modelarlo mediante modelado comparativo, la información tomada de un ligando activo o de un grupo de ligandos activos contra un blanco molecular puede ser usada para identificar características estructurales responsables de la actividad biológica observada (Macalino et al. 2015). La hipótesis subyacente es que moléculas similares exhibirán propiedades de unión similares con respecto a un dado blanco molecular (Sotriffer 2009).

Dependiendo de la técnica de SV que se desee utilizar, los ligandos conocidos serán caracterizados a través de la representación bidimensional (2D) y/o tridimensional (3D) de sus estructuras químicas, buscando a partir de estas características estructurales determinantes de la actividad biológica.

En los siguientes puntos de este capítulo se dará una breve descripción de los métodos de VS basados en el ligando, los cuales pueden clasificarse en 3 grandes grupos: enfoques basados en el farmacóforo, metodologías basadas en la similitud molecular y, metodologías basadas en descriptores moleculares.

2.2.2.1. FARMACÓFOROS

Un farmacóforo se puede describir como un arreglo tridimensional con características moleculares mínimas que son necesarias para asegurar interacciones con un blanco molecular específico, produciendo una determinada respuesta biológica (Talevi et al. 2011). La IUPAC lo define como un conjunto de propiedades estéricas y electrónicas para asegurar las interacciones supramoleculares óptimas con un blanco biológico específico activando o bloqueando su respuesta (Wermuth et al. 1998).

El SV basado en farmacóforos consiste en superponer un conjunto de moléculas que comparten una determinada actividad (conjunto de entrenamiento) para extraer de éstas las propiedades químicas comunes que son esenciales para producir la actividad de interés. Luego se busca en bases de datos compuestos que presenten el farmacóforo obtenido (Sotriffer 2009).

La obtención de la estructura farmacofórica se realiza generalmente mediante la extracción de las características comunes de las estructuras 3D del conjunto de entrenamiento. Se deben resolver dos cuestiones centrales para construir y aplicar hipótesis farmacofóricas:

- a) ¿Cómo manejar la flexibilidad de las moléculas?
- b) ¿Cómo realizar el alineamiento molecular o superposición? (Yang 2010).

En cuanto a la primera cuestión, pueden utilizarse 3 estrategias diferentes: la construcción de la hipótesis farmacofórica desde uno o más ligandos rígidos; la generación de múltiples

conformaciones posibles del conjunto de entrenamiento y de la base sometida al cribado; y por último, el análisis conformacional continuo (*"on the fly"*) durante el proceso de modelado farmacofórico (Talevi et al. 2011).

Con respecto a la segunda pregunta, el alineamiento molecular puede realizarse a través de enfoques basados en puntos (en los cuales las distancias entre átomos o puntos de características químicas abstractas se minimizan) o mediante aproximaciones basadas en propiedades (que usan descriptores basados en campos moleculares para generar el alineamiento) (Talevi et al. 2011). La más frecuentemente utilizada es la primera opción, prefiriéndose recurrir a un arreglo tridimensional de características moleculares abstractas en lugar de a grupos funcionales bien definidos, para lograr de esa manera la recuperación de una mayor diversidad de andamiajes moleculares activos. Tales características son, frecuentemente: grupos con capacidad de establecer enlaces de hidrógeno; grupos cargados; grupos capaces de participar en contactos hidrofóbicos o aromáticos. Adicionalmente, pueden considerarse restricciones estéricas bajo la forma de volúmenes de exclusión (Kaserer et al. 2015).

La identificación y utilización de farmacóforos también puede ser abordada desde el receptor, buscando posibles puntos de interacción entre el blanco molecular y los ligandos (Yang 2010).

2.2.2.2. SIMILITUD (O SIMILARIDAD) MOLECULAR

Cuando el número de ligandos conocidos de un blanco molecular es limitado (o, en el caso extremo, cuando se conoce un único ligando) se puede recurrir a las técnicas de SV basadas en la similitud molecular (Talevi et al. 2011). Estas técnicas consisten en cuantificar el grado de similitud entre la estructura que presenta la actividad deseada y los diferentes compuestos que integran la base de datos sometida al tamizado (Bajorath 2001).

La similitud molecular se enfoca principalmente en características estructurales de los compuestos tales como presencia, ausencia o frecuencia de determinadas subestructuras químicas o determinados grupos funcionales. Para cuantificar la similitud son necesarios tres componentes básicos: (a) una representación de la molécula cuyos componentes codifiquen las características químicas y/o moleculares relevantes para la determinación; (b) un esquema de ponderación de tales características (esto es, ¿se les asigna a todas ellas el mismo peso en la comparación o algunas características se considerarán más relevantes que otras? y; (c) un **coeficiente de similitud** que transforme la información contenida en la representación en un valor numérico. En general, este valor se encuentra entre 0 y 1, donde 1 corresponde a la identidad completa de la representación molecular (lo cual no necesariamente implica que los compuestos comparados son idénticos) (Maggiora et al. 2014).

Una forma habitual de comparar cuantitativamente dos moléculas se basa en la comparación de secuencias de bits (denominadas *molecular fingerprints* o huellas digitales moleculares) de las estructuras a comparar. Las *fingerprints* se construyen típicamente asignando cada bit a la presencia o ausencia de una propiedad estructural determinada, pudiendo ser estas propiedades de naturaleza bidimensional o tridimensional. Alternativamente, los bits pueden indicar la frecuencia con la que aparece la propiedad estructural en las moléculas comparadas. Finalmente, los bits compartidos por ambas moléculas son cuantificados mediante el coeficiente de similitud (Maggiora et al. 2014).

Uno de los coeficientes de similitud más utilizados es el coeficiente de Tanimoto o coeficiente de Jaccard (Willett 2006). El mismo se define en la **Ecuación 2.1**:

$$S = \frac{c}{a + b - c}$$

Ecuación 2.1: Definición del coeficiente de Tanimoto.

Donde a y b representan el número de tipos de subestructuras (o propiedades estructurales) presentes en cada una de las dos moléculas sometidas a comparación, y c es el número de tipos de subestructuras comunes a ambas moléculas. La **Figura 2.6** presenta un ejemplo de cómo se realiza el cálculo del coeficiente de Tanimoto. Si se consideran sólo las 8 subestructuras de la figura, el número de bits para la molécula **A** es igual a 6, el número de bits para la molécula **B** es igual a 7 y el número de subestructuras compartidas es 5; por lo tanto, el coeficiente de Tanimoto asume un valor de 0,625.

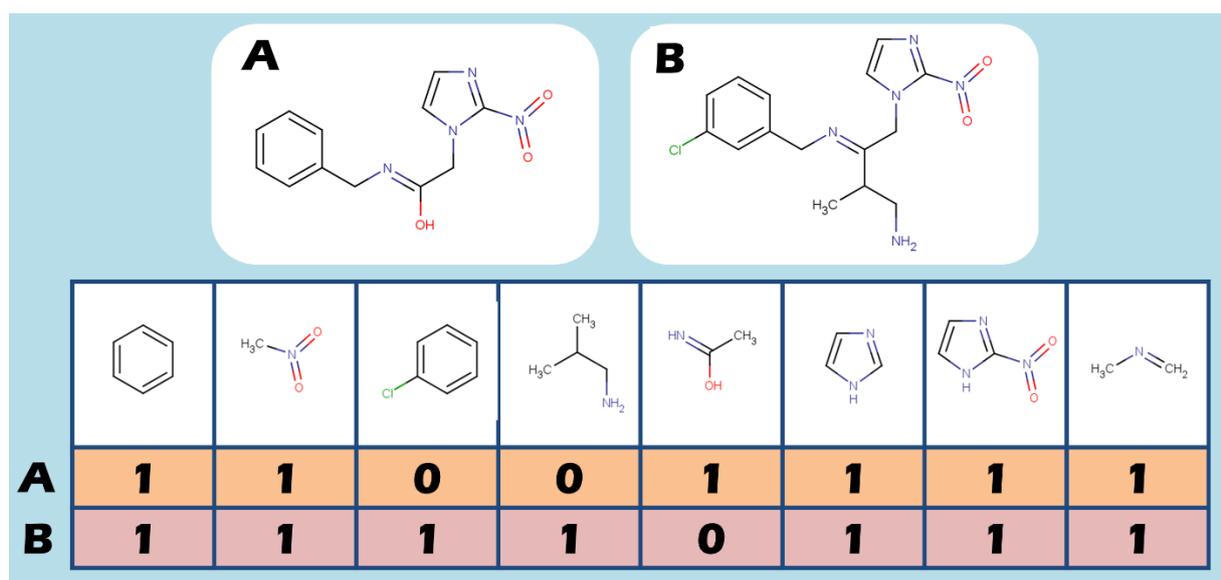


Figura 2.6: Esquema del proceso de comparación de dos moléculas por medio de un coeficiente de similitud.

Las dos ventajas principales de las técnicas de SV basadas en la similitud son el bajo costo computacional, permitiendo el tamizado de grandes bases de datos en poco tiempo, y la posibilidad de realizar búsquedas a partir de una única molécula de referencia (Di Ianni et al. 2015).

2.2.2.3. DESCRIPTORES MOLECULARES

Todeschini y Consonni definen a un descriptor molecular como: “el resultado final de un procedimiento lógico y matemático que transforma la información química codificada en una representación simbólica de una molécula en un número útil o en el resultado de algún experimento estandarizado” (Todeschini & Consonni 2009). Dicho en otras palabras, un descriptor molecular es una variable numérica que representa algún aspecto o característica de la estructura molecular como el tamaño, el volumen, la cantidad de enlaces de hidrógeno disponibles, la lipofilicidad, la forma, la distribución electrónica, etc. (Davis 2014).

La información contenida en los descriptores moleculares puede variar desde propiedades muy simples como el peso molecular, o el número de veces que aparece un cierto átomo en la molécula, hasta descriptores 3D de cálculo complejo. De acuerdo con la dimensionalidad implicada, los descriptores se pueden clasificar en:

0-Dimensionales (0D): Se denominan descriptores 0D a aquellos que pueden obtenerse de la fórmula química de la molécula. La fórmula química es la representación molecular más simple ya que sólo especifica el tipo y número de los elementos químicos constituyentes. Esta representación nada dice de la conectividad entre átomos. Estos descriptores son muy sencillos de calcular e interpretar, pero muestran bajo contenido de información y un alto grado de degeneración, es decir, pueden tener valores iguales para diferentes moléculas (Grisoni et al. 2018).

1-Dimensionales (1D): Las moléculas pueden ser entendidas como un set de subestructuras, tales como grupos funcionales. La representación 1D de la molécula estará reflejada en descriptores binarios (codifican presencia/ausencia de una dada subestructura o grupo funcional) o de naturaleza discreta, señalando la frecuencia con la que ocurren en la molécula distintas subestructuras o grupos funcionales (Grisoni et al. 2018).

2-Dimensionales (2D): Adicionan un nivel de información más a los descriptores 1D ya que consideran la conectividad de los átomos, teniendo en cuenta la presencia y naturaleza de los enlaces químicos. La molécula se representa como un grafo, cuyos vértices son los átomos y las aristas son los enlaces (Pogliani 2000). A partir de estas representaciones es posible derivar descriptores de la topología molecular, llamados índices topológicos, que codifican propiedades como el tamaño, la forma, la simetría y la ciclicidad; y también pueden reflejar la distribución en la molécula de propiedades atómicas tales como la masa atómica o la polarizabilidad (Consonni & Todeschini 2012).

3-Dimensionales (3D): Consideran a la molécula como un objeto geométrico, caracterizado por la configuración espacial de sus átomos. Estos descriptores tienen un alto contenido de información, sin embargo, para poder determinarlos hay que tener en cuenta tres puntos: (a) el método de optimización de la geometría de la molécula; (b) la presencia de más de un conformero de mínima energía; y (c) la diferencia entre la conformación bioactiva y la conformación de menor energía. Se debe analizar el costo/beneficio de usar este tipo de descriptores en cada caso ya que presentan mayor cantidad de información, pero son computacionalmente más costosos (Grisoni et al. 2018).

4-Dimensionales (4D): Los descriptores 4D son similares a los 3D, pero con la diferencia de que la información estructural se considera para un conjunto de conformeros, en vez de para una única conformación (Andrade et al. 2010).

En la **Figura 2.7** se muestra un ejemplo gráfico de las diferentes representaciones que pueden utilizarse para la molécula de ibuprofeno. Puede observarse que al aumentar la dimensionalidad de representación podemos obtener más información de la molécula, sin embargo, aumenta el costo computacional asociado a los cálculos.

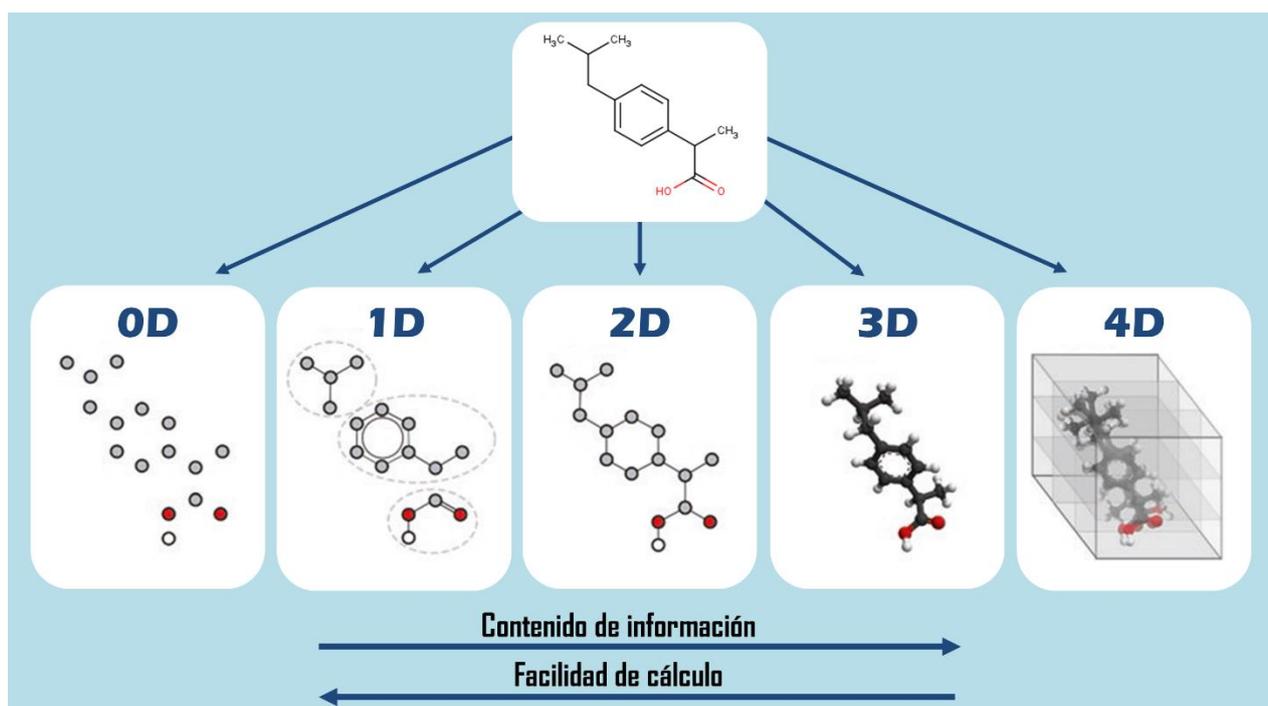


Figura 2.7: Diferentes representaciones de la molécula de ibuprofeno. Adaptado de Grisoni et al. 2018.

Para un análisis más práctico, es posible englobar a los descriptores en dos grandes categorías de acuerdo a su dimensionalidad: los de baja dimensionalidad (0D,1D,2D) o independientes de la conformación y los de alta dimensionalidad (3D, 4D) o dependientes de la conformación.

Aplicación de los descriptores moleculares a técnicas de SV:

Existen diferentes estrategias estadísticas para encontrar correlaciones entre una variable dependiente (la propiedad o actividad a modelar) y uno o más descriptores moleculares. Estas relaciones son conocidas como **Relaciones Cuantitativas Estructura-Actividad** o **QSAR** (del inglés, *Quantitative Structure-Activity Relationships*) (Talevi & Bruno-Blanch 2009). El desarrollo de este tipo de modelos se basa en que las diferentes actividades de los compuestos dependen de las variaciones estructurales entre ellos (Macalino et al. 2015).

Los modelos QSAR pueden clasificarse en **QSAR cuantitativo y cualitativo** dependiendo de la naturaleza de la variable dependiente considerada. Los primeros dan como resultado (o respuesta) una variable dependiente continua, es decir, pueden predecir el valor numérico de la propiedad (por ejemplo, K_i , IC_{50} , ED_{50} , etc.); los segundos generan una respuesta discreta (modelos clasificatorios) cuyo valor se asocia a distintas categorías (por ejemplo, “activo” e “inactivo”, “tóxico” y “no tóxico”, etc.).

Una vez que se obtiene un modelo QSAR, este puede aplicarse a campañas de SV, es decir, a la predicción de la actividad de las estructuras químicas presentes en bases de datos.

2.2.3. MÉTODOS COMBINADOS

Los métodos mixtos, híbridos o combinados conjugan metodologías basadas en el ligando y basadas en la estructura de forma secuencial o paralela para realizar el tamizado (Figura 2.8). La

forma más común de combinación de estos métodos es a través de la integración secuencial de los mismos, que es conocida como **SV jerárquico (HLVS) o en cascada**. Esta estrategia se basa en la aplicación de una serie de filtros *in silico* secuenciales para reducir una gran base de datos de pequeñas moléculas a un número de compuestos suficientemente pequeño (y promisorio) como para poder avanzar a ensayos biológicos. En contraste al HLVS, existe el **SV paralelo (PVS)** donde varios métodos complementarios son corridos en paralelo y los mejores *hits* rankeados de acuerdo a cada método son seleccionados para las pruebas biológicas (Kumar & Zhang 2015).

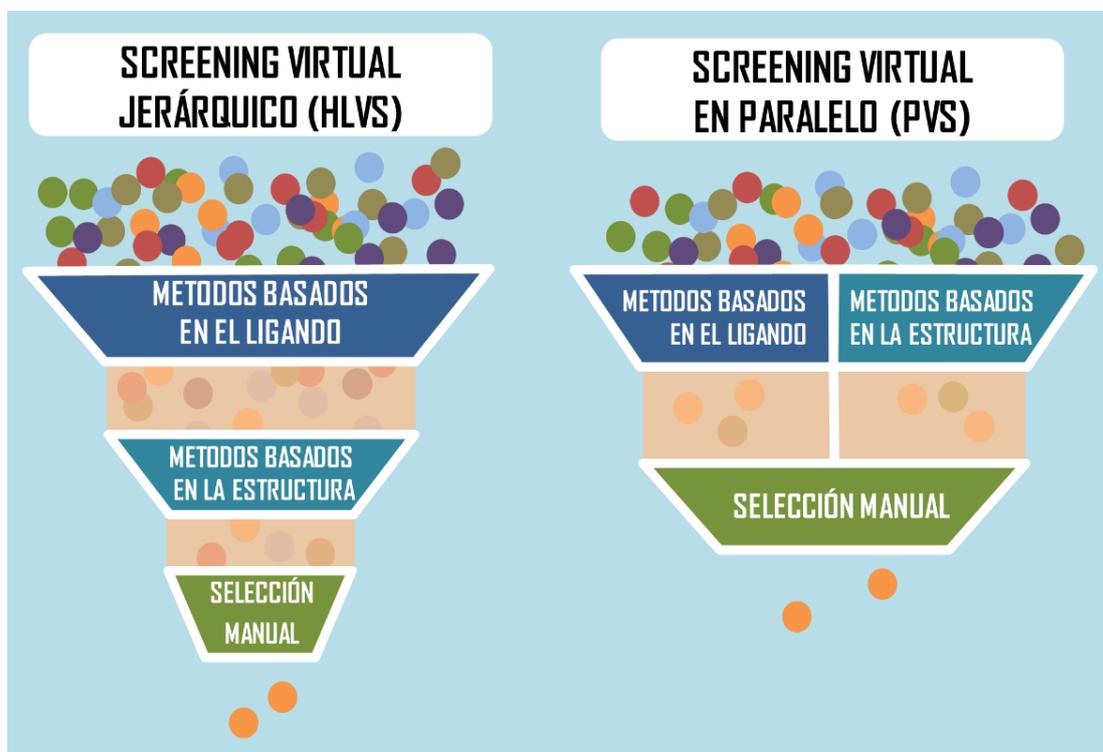


Figura 2.8: Estrategias para combinar métodos basados en el ligando y en la estructura.

2.3. REPOSICIONAMIENTO DE FÁRMACOS

2.3.1. GENERALIDADES

El **reposicionamiento de fármacos** (en inglés *“drug repositioning”* o *“drug repurposing”*) involucra encontrar nuevas indicaciones terapéuticas para drogas existentes. Los candidatos a ser reposicionados son generalmente drogas ya aprobadas, retiradas del mercado, abandonadas en pruebas clínicas por no demostrar la eficacia esperada contra determinada patología, o drogas que se encuentran en etapas avanzadas de investigación (Liu et al. 2013; Jesús Naveja et al. 2016).

Es posible utilizar drogas para más de una indicación debido a la polifarmacología (o promiscuidad) de algunas pequeñas moléculas, a la similitud entre blancos moleculares ortólogos en distintas especies y a las conexiones existentes entre vías metabólicas. La promiscuidad de las moléculas puede ser explotada para el reposicionamiento, pero debe tenerse en cuenta también que una excesiva promiscuidad se asocia a efectos adversos (Ma et al. 2013).

El reposicionamiento de fármacos es una estrategia prometedora en el campo farmacéutico debido a que reduce considerablemente los tiempos necesarios para desarrollar terapias novedosas y aumenta la probabilidad de que la droga que ingresa a estudios clínicos los supere exitosamente (Temesi et al. 2014). Además, teniendo en cuenta que las drogas reposicionadas evitan ocasionalmente la necesidad de las etapas de descubrimiento del blanco, optimización del líder y pruebas de seguridad temprana, aproximadamente 50% de los costos pueden ser reducidos para obtener una droga aprobada para la nueva indicación (Paul & Lewis-Hall 2013), aunque otras estimaciones del ahorro que produce el reposicionamiento son aún más optimistas (Mignani et al. 2016). Estas ventajas han provocado que en los últimos años aumente considerablemente el interés de las industrias farmacéuticas, organismos gubernamentales y del sector académico por el reposicionamiento de fármacos (Liu et al. 2013).

En la **Figura 2.9** puede verse el aumento en la frecuencia de aparición de los términos “*drug repositioning*” y “*drug repurposing*” en los títulos y/o resúmenes de trabajos científicos indexados en Pubmed durante los últimos 10 años.

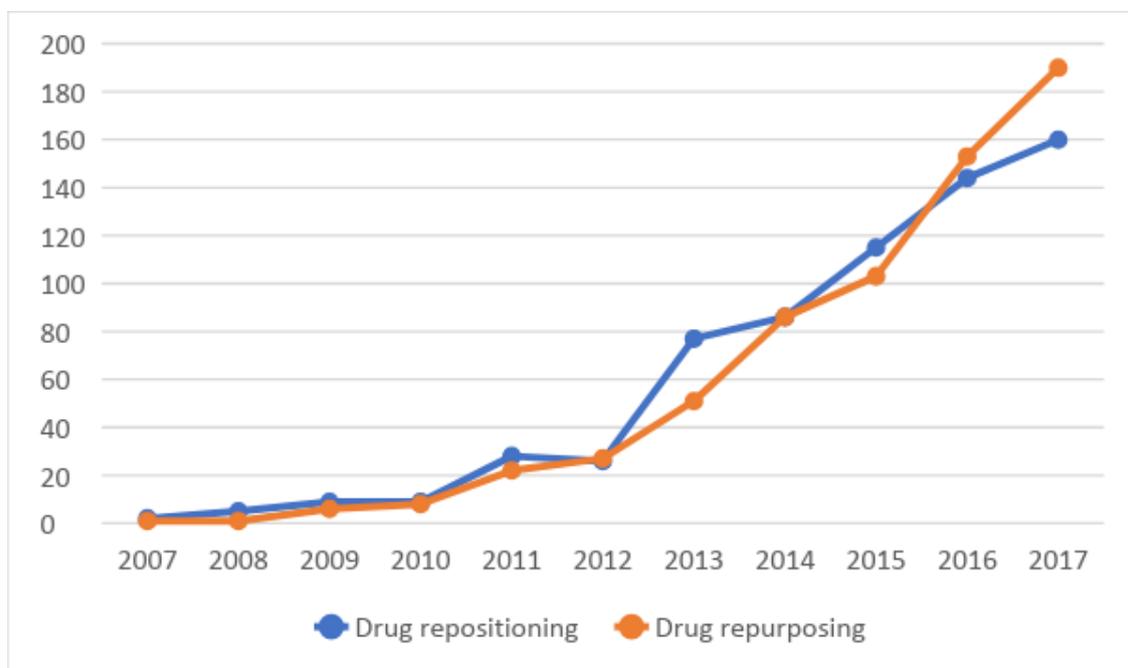


Figura 2.9: Número de artículos publicados en revistas científicas indexadas por Pubmed que contienen el término “*drug repositioning*” o “*drug repurposing*” en su título y/o resumen. La búsqueda se restringió a los últimos 10 años.

Uno de los primeros y ya clásicos casos de reposicionamiento de fármacos surgió de observaciones casuales (serendipia). El Sildenafil, que fue originalmente investigado como una droga antihipertensiva y para el tratamiento de la angina de pecho, demostró durante el desarrollo de las pruebas clínicas un efecto inesperado sobre la erección peneana. En 1998, y luego de realizar los estudios necesarios, la FDA aprobó el fármaco para el tratamiento de la insuficiencia eréctil y este fármaco se convirtió en un éxito en ventas (Ashburn & Thor 2004; Barnett & Machado 2006).

Otro caso muy conocido de reposicionamientos aprobados es el de la droga Talidomida (Vargesson 2015), que fue sintetizada en 1954 y utilizada originalmente como sedante para

calmar la ansiedad, el insomnio, las náuseas y los vómitos en mujeres embarazadas. En 1961 debió ser retirada del mercado cuando se revelaron los efectos teratogénicos en humanos que afectaron alrededor de 10.000 niños. Décadas más tarde, la Talidomida fue rescatada del olvido a través del descubrimiento fortuito de su sorprendente actividad contra complicaciones cutáneas de la lepra denominadas eritemas nudosos leprosos (ENL). A pesar de su teratogenicidad, el tratamiento fue readaptado y en 1998 la FDA aprobó el uso de la droga para el tratamiento de ENL (Teo et al. 2002). Debido a la capacidad de la Talidomida para inhibir la angiogénesis, esta droga fue estudiada y reposicionada por segunda vez para el tratamiento de mieloma múltiple (Zhou et al. 2013; Singhal et al. 1999).

Otros casos de fármacos que han ya sido aprobados para indicaciones diferentes de la original se presentan en el **Material Suplementario 2**. Además de los ya aprobados, hay muchos fármacos que están transcurriendo etapas avanzadas de investigación para una segunda indicación terapéutica. Existen bases de datos que compilan esta información como RepoDB (Brown & Patel 2017), que contiene datos de más de 1600 drogas con múltiples indicaciones terapéuticas aprobadas o en investigación.

A pesar de que muchos casos de drogas reposicionadas han surgido casualmente por el descubrimiento de efectos adversos beneficiosos para los pacientes, existen una gran cantidad de métodos basados en el conocimiento para la búsqueda de nuevos reposicionamientos (Liu et al. 2013; Jesús Naveja et al. 2016).

2.3.2. MÉTODOS SISTEMÁTICOS DE REPOSICIONAMIENTO DE FÁRMACOS

Los métodos sistemáticos utilizan la información disponible para obtener relaciones entre fármacos, proteínas y enfermedades, que puedan utilizarse para el reposicionamiento de fármacos (**Figura 2.10**). Asimismo, también podría considerarse sistemático el cribado masivo de bibliotecas de compuestos químicos. Los métodos de reposicionamiento se pueden clasificar en dos grandes grupos de acuerdo a si requieren la utilización de la computadora. Los métodos no computacionales implican, por un lado, experimentos “húmedos” como los *screening* de baja y media performance y, por otro, también métodos basados en la literatura. Los métodos computacionales se pueden dividir en tres grandes grupos de acuerdo al tipo de información que utilizan: métodos bioinformáticos, métodos quimioinformáticos y métodos basados en la literatura de alto rendimiento.

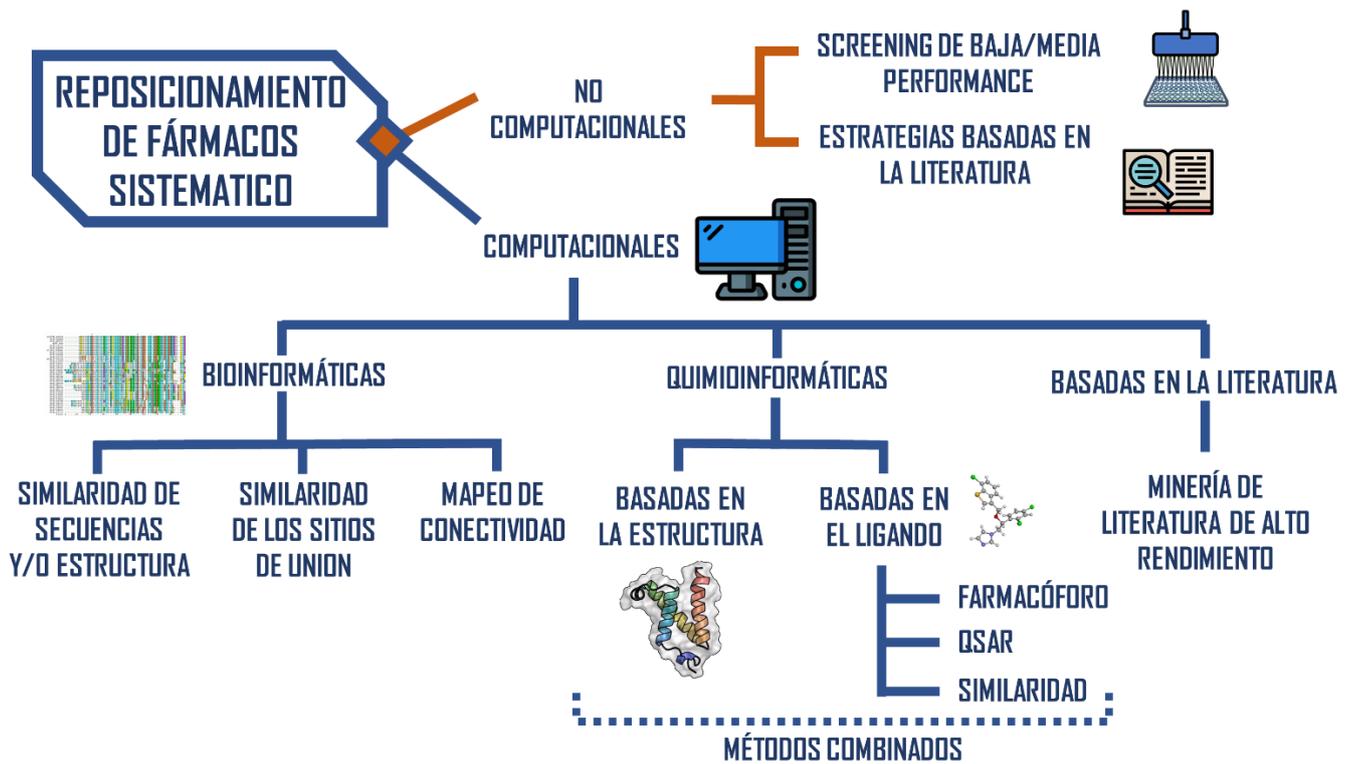


Figura 2.10: Esquema de los métodos sistemáticos que pueden ser aplicados en el reposicionamiento de fármacos. Adaptado de Talevi et al. 2018.

2.3.2.1. ENFOQUES NO COMPUTACIONALES

Desentrañar el mecanismo de acción de una droga, así como la fisiopatología de una enfermedad, podría ayudar a establecer hipótesis sobre el potencial uso de una droga conocida para una nueva indicación. Estos enfoques son los denominados **estrategias de descubrimiento basadas en la literatura o en el conocimiento** (Jesús Naveja et al. 2016).

Entre las primeras estrategias sistemáticas utilizadas para identificar posibles reposicionamientos encontramos también a las **metodologías de screening masivo experimental**. Estos métodos han ido evolucionando desde el screening de baja performance (analizando sólo un pequeño set de compuestos) hasta las ya discutidas estrategias de HTS que permiten evaluar por encima de 100.000 compuestos en tan sólo un día (Hertzberg & Pope, 2000). Estos métodos implican una manera interesante de buscar nuevos reposicionamientos ya que además de evaluar el impacto frente a un blanco molecular es posible hoy día adaptar los métodos al *screening* fenotípico y de esa manera evaluar las drogas en sistemas complejos. Como se mencionó ya, la principal limitación del HTS es el alto costo requerido para los equipamientos robóticos y el resto de la infraestructura y el costo de contar con la colección de compuestos, así como también el número de falsos positivos (Cronk 2012).

2.3.2.2. ENFOQUES COMPUTACIONALES

Las estrategias computacionales ofrecen grandes oportunidades para transformar el proceso serendípico de encontrar una nueva indicación para una droga en una exploración exhaustiva y racional de todas las oportunidades de reposicionamiento para las drogas disponibles. El reposicionamiento de fármacos *in silico* aplica varias estrategias para recuperar, integrar y analizar sistemáticamente grandes volúmenes de información, incluyendo bases de datos de fármacos existentes (Liu et al. 2013).

2.3.2.2.1. METODOLOGÍAS QUIMIOINFORMÁTICAS

Los enfoques quimioinformáticos para el reposicionamiento de fármacos incluyen principalmente a las técnicas de SV discutidas anteriormente. Actualmente existen varias bases de datos de libre acceso que contienen gran cantidad de drogas candidatas a reposicionamiento, entre ellas DrugBank (Wishart et al. 2018), DrugCentral (Ursu et al. 2017), Sweetlead (Novick et al. 2013), así como información de blancos moleculares vinculados a enfermedades humanas, tales como Therapeutic Target Database (TTD) (Li et al. 2018).

2.3.2.2.2. METODOLOGÍAS BIOINFORMÁTICAS

Los rápidos avances en la genómica han conducido a la generación de grandes volúmenes de datos genómicos y transcriptómicos de un diverso set de muestras de tejido patológico, muestras de tejidos normales, modelos animales y líneas celulares. Debido a la emergente filosofía de datos abiertos (open data), muchos de estos datos están públicamente disponibles. Junto con bases de datos que compilan información fenotípica y clínica, estos sets de datos proveen oportunidades para entender los mecanismos de las enfermedades, elucidar mecanismos de acción de las drogas e identificar nuevas indicaciones para fármacos conocidos. Entre otros, los perfiles transcriptómicos, tales como datos de expresión génica, son los datos más ampliamente usados para el reposicionamiento (Li et al. 2016).

Dos de las fuentes de información más importantes para utilizar en reposicionamiento son el proyecto Connectivity Map (CMap) (Lamb et al. 2006) y su proyecto extendido LINCS (Library of Integrated Network-Based Cellular Signatures) (Vidović et al. 2014), que han producido perfiles de expresión génica a gran escala de tejidos y tipos celulares relevantes para una diversidad de enfermedades, incluyendo cáncer, afecciones cardíacas y enfermedades neurodegenerativas, expuestos a diferentes perturbágenos (entre ellos, drogas). El objetivo de CMap es construir un mapa detallado para las asociaciones funcionales entre enfermedades, perturbaciones genéticas y acciones de drogas. Dicha base ha sido extensamente utilizada en estudios de reposicionamiento. La estrategia conocida como **reversión de la firma** (en inglés, “*signature reversión*”) utiliza estos datos para buscar relaciones droga-enfermedad inversas por comparación de los perfiles de expresión genética frente a drogas y perfiles de expresión genética de enfermedades (Li et al. 2016). Es decir, conociendo el perfil genético de las enfermedades y el que producen las drogas, es posible estimar qué drogas podrían revertir el perfil genético ocasionado por la enfermedad.

Otra estrategia, denominada **“culpable por asociación”** (“*guilt-by-association*”), busca drogas que provoquen respuestas transcripcionales similares, postulando que ellas deberían compartir similares mecanismos de acción (Iorio et al. 2013). Existen varios repositorios públicos

disponibles para realizar esta estrategia como Drug versus Drugs (DvD) (Pacini et al. 2013), The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al. 2003) y el Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005).

Otras herramientas bioinformáticas que pueden utilizarse son aquellas que comparan las secuencias genéticas o las estructuras de las proteínas. Estos métodos se basan en el postulado de que proteínas homólogas de diferentes especies pueden interactuar con los mismos ligandos (Gfeller & Zoete 2015).

Por último, proteínas sin relaciones evolutivas aparentes o incluso sin un arreglo tridimensional o una función similar, también pueden ser utilizadas para lograr reposicionamientos. Como ya se discutió anteriormente, puede encontrarse similitud entre sitios de unión de proteínas de baja similitud global, las cuales pueden interactuar con ligandos similares. Por lo tanto, conociendo el sitio de unión de una proteína, mediante diferentes herramientas bioinformáticas es posible identificar qué proteínas de una base de datos presentan sitios de unión similar, detectándose oportunidades de reposicionamiento cruzadas en sus respectivos sets de ligandos conocidos (Joachim Haupt & Schroeder 2011; Konc et al. 2015).

2.3.2.2.3. METODOLOGÍAS BASADAS EN LA LITERATURA

Otra de las estrategias que ha surgido en los últimos años es el descubrimiento basado en la literatura (LBD). Esta estrategia se basa en el procesamiento de datos de literatura para descubrir relaciones implícitas o indirectas entre entidades biológicas aparentemente desconectadas (Figura 2.II). LBD trata grandes volúmenes de literatura científica como un gran sistema de interconexiones entre parámetros de investigación; navega a través de esas interconexiones para describir los mecanismos moleculares básicos que determinan una enfermedad, para comprender mejor los efectos biológicos de una droga, para evaluar el perfil de riesgo/beneficio, y para llegar a descubrimientos novedosos (Deftereos et al. 2011; Andronis et al. 2011).

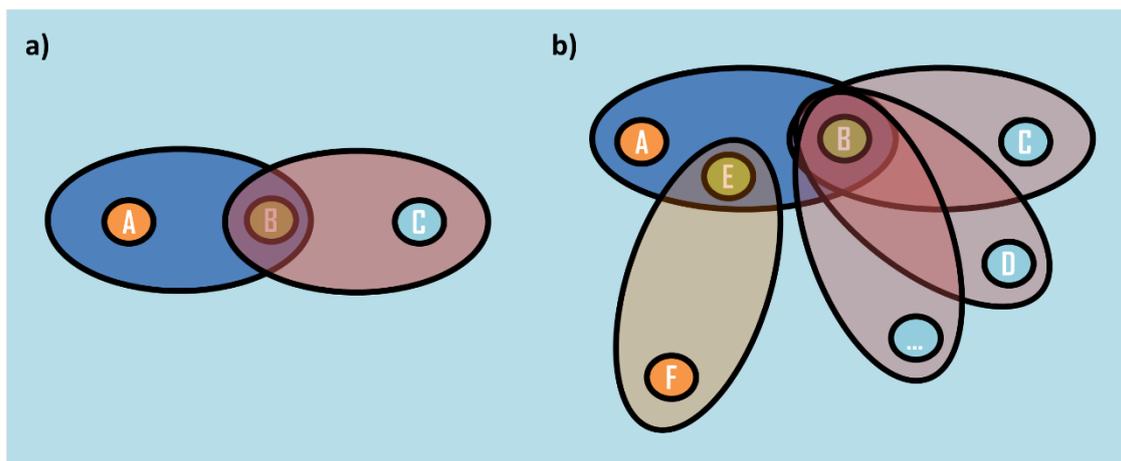


Figura 2.II: Modelos de reposicionamiento basado en la literatura. (a) El concepto A está relacionado al concepto B, el cual a su vez está relacionado con C como se informó en un set de trabajos científicos (indicados por las elipses azul y roja). A pesar de que no se conoce que A está directamente relacionado con C, se puede inferir una relación indirecta a través de B. (b) El concepto A está relacionado al concepto B, y a través de esto a C, D, etc. Además, A está relacionado a F a través de E. Adaptado de Deftereos et al. 2011.

2.3.3. REPOSICIONAMIENTO DE FÁRMACOS PARA ETDs

En comparación con otras enfermedades, y teniendo en cuenta el costo humano y económico de las ETD, el número de nuevas terapias desarrolladas y en desarrollo para las ETDs es pequeño. En el periodo entre 1975 y 1999 fueron aprobadas 1393 drogas, pero nada más 13 (0,93%) de éstas estaban dirigidas a ETDs. Entre los años 2000 y 2011, de los 850 productos terapéuticos registrados, sólo 5 (0,59%) eran para ETDs y todos ellos fueron catalogados como nuevas indicaciones o formulaciones de drogas existentes (ninguno fue una entidad química novedosa). Desde el 2000 al 2014, entraron a la fase I de ensayos clínicos 4006 productos farmacéuticos de los cuales sólo 66 (1,65%) estaban destinados al tratamiento o prevención de ETDs (Weng et al. 2018).

Gran parte del descubrimiento de fármacos para el tratamiento de las ETDs es realizado en laboratorios académicos sin el financiamiento, el personal y los recursos tecnológicos que poseen las compañías farmacéuticas. Por ello, para superar estas limitaciones, la búsqueda de nuevos tratamientos mediante el reposicionamiento de fármacos se ha adoptado ampliamente en el sector académico (Klug et al. 2016). Esta estrategia emerge como un enfoque atractivo para el desarrollo de nuevos tratamientos para ETDs. De hecho, alrededor del 40% de las drogas actualmente disponibles para el tratamiento de ETDs fueron originalmente investigadas para otras enfermedades.

2.3.3.1. CASOS PROMETEDORES Y EXITOSOS EN TRIPANOSOMÁTIDOS

Aunque la mayoría de los medicamentos utilizados para el tratamiento de la **Enfermedad del Sueño** se han desarrollado desde sus inicios con ese fin, hay algunos fármacos reposicionados aprobados y en desarrollo para mejorar el tratamiento. La Eflornitina, que ha sido desarrollada como anticancerígeno, fue introducida exitosamente para el tratamiento de la Enfermedad del Sueño en 1990 (Nwaka & Hudson 2006). El Nifurtimox, uno de los dos tratamientos disponibles para la Enfermedad de Chagas, ha demostrado también buenos resultados contra la Enfermedad del Sueño (Bouteille et al. 2003). A pesar de que esta droga no ha sido aprobada como monoterapia (debido a tasas de curación variable y a su toxicidad) es utilizada para tratar aquellos pacientes refractarios contra otras terapias, especialmente en combinación con Eflornitina (Patterson & Wyllie 2014).

El Fexinidazol estuvo en desarrollo preclínico como un agente antimicrobiano de amplio espectro. En la última década, diferentes estudios han demostrado que el Fexinidazol es un excelente candidato para curar la Enfermedad del Sueño, incluida la etapa avanzada y fatal de la enfermedad (Torreele et al. 2010), y actualmente se encuentra en etapa clínica de evaluación.

Como ya se discutió en el capítulo anterior, entre los tratamientos más prometedores para la **Enfermedad de Chagas** se encuentran drogas reposicionadas. Los agentes antifúngicos de amplio espectro Posaconazol y Ravuconazol, han sido estudiados en pruebas clínicas para el tratamiento de la Enfermedad de Chagas. El antimicrobiano Fexinidazol también ha demostrado buenos resultados en ratones infectados con *Trypanosoma cruzi*, por lo que se están llevando a cabo estudios de fase 2 para determinar la eficacia del compuesto en adultos en la fase crónica de la enfermedad. Otros dos fármacos, que están aprobados como antiarrítmicos y presentaron buenos resultados como agentes tripanocidas son la Amiodarona y Dronedarona (Sbaraglini et al. 2016).

En los últimos años, la cantidad de drogas potenciales y drogas aprobadas contra la **Leishmaniasis** ha aumentado debido al reposicionamiento de fármacos. La Anfotericina B, un compuesto antifúngico utilizado para el tratamiento de Aspergilosis, Criptococosis, Blastomycosis, Candidiasis, Histoplasmosis y Zigomicosis, ha sido aprobada por la FDA para el tratamiento de la Leishmaniasis y ha sido incluida en las medicinas esenciales publicadas por la OMS en 2011 (Meyerhoff 1999). Los antifúngicos de la familia de los triazoles, Itraconazol y Posaconazol, y los antibióticos Azitromicina y Paromomicina también han demostrado ser efectivos contra *Leishmania amazonensis* (Sbaraglini et al. 2016). La Miltefosina, un medicamento usado para el tratamiento de las metástasis cutáneas de melanoma, ha sido reposicionada para el tratamiento de Leishmaniasis visceral siendo la única droga registrada como tratamiento oral para la enfermedad (Dorlo et al. 2012).

2.4. ESTRATEGIA UTILIZADA EN ESTE TRABAJO DE TESIS

En este trabajo de tesis se han desarrollado modelos computacionales capaces de identificar inhibidores del metabolismo y/o incorporación de poliaminas de *Trypanosoma cruzi* en bases de datos de compuestos químicos. Entre los métodos ya descritos para la realización de SV, hemos utilizado principalmente el método indirecto (basado en ligandos) mediante la obtención de modelos QSAR-2D basados en descriptores moleculares, y su posterior aplicación al SV. Adicionalmente, en una segunda campaña de SV hemos implementado el *docking molecular* como filtro adicional para la selección de los candidatos.

El QSAR-2D tiene la ventaja de que no requiere de la optimización previa de las conformaciones de las representaciones moleculares por lo que se reduce considerablemente el costo computacional y el tiempo necesarios para la realización de una campaña de SV. Sin embargo, a diferencia de las técnicas que utilizan los descriptores 3D o de los métodos basados en el blanco molecular, los modelos QSAR-2D utilizan descriptores que pueden ser de difícil interpretación. Por lo tanto, es difícil extraer de un modelo QSAR-2D información significativa acerca de los mecanismos por los cuales un compuesto determinado interacciona con un blanco molecular. Por otra parte, no permiten explicar las diferencias en la actividad farmacológica observadas a menudo para isómeros de geometría.

Finalmente, teniendo en cuenta el éxito de las estrategias de reposicionamiento de fármacos para la búsqueda de nuevos tratamientos para ETDs, y considerando las ventajas que presentan las diferentes técnicas de modelado *in silico*, hemos aplicado los modelos al SV de bases de datos de fármacos existentes.

REFERENCIAS

- Alagarsamy, V., 2010. *Textbook of Medicinal Chemistry, Volume I*, Elsevier.
- Andrade, C.H. et al., 2010. 4D-QSAR: Perspectives in drug design. *Molecules*, 15(5), pp.3281–3294.
- Andronis, C. et al., 2011. Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics*, 12(4), pp.357–368.
- Ashburn, T.T. & Thor, K.B., 2004. Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8), pp.673–683.
- Bajorath, J., 2001. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *Journal of Chemical Information and Computer Sciences*, 41(2), pp.233–245.
- Barnett, C.F. & Machado, R.F., 2006. Sildenafil in the treatment of pulmonary hypertension. *Vascular Health and Risk Management*, 2, pp.411–422.
- Bouteille, B. et al., 2003. Treatment perspectives for human African trypanosomiasis. *Fundamental & clinical pharmacology*, 17(2), pp.171–181.
- Brown, A.S. & Patel, C.J., 2017. A standard database for drug repositioning. *Scientific Data*, 4, p.170029.
- Consonni, V. & Todeschini, R., 2012. Multivariate Analysis of Molecular Descriptors. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 111–147.
- Cook, D. et al., 2014. Lessons learned from the fate of AstraZeneca's drug pipeline: A five-dimensional framework. *Nature Reviews Drug Discovery*, 13(6), pp.419–431.
- Cronk, D., 2012. *High-throughput screening* Second Edi., Elsevier Ltd.
- Davies, M. et al., 2015. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, 43(W1), pp.612–620.
- Davis, A.M., 2014. CHAPTER 6. Quantitative Structure–Activity Relationships. En *The Handbook of Medicinal Chemistry*. Cambridge: Royal Society of Chemistry, pp. 154–183.
- Deftereos, S.N. et al., 2011. Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(3), pp.323–334.
- Dennis, G. et al., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(9): R60
- Dorlo, T.P.C. et al., 2012. Miltefosine: a review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *The Journal of antimicrobial chemotherapy*, 67(11), pp.2576–97.
- Eder, J. et al, 2014. The discovery of first-in-class drugs: Origins and evolution. *Nature Reviews Drug Discovery*, 13(8), pp.577–587.
- Ehrt, C. et al., 2016. Impact of Binding Site Comparisons on Medicinal Chemistry and Rational Molecular Design. *Journal of Medicinal Chemistry*, 59(9), pp.4121–4151.

- Field, M.C. et al., 2017. Anti-trypanosomatid drug discovery: An ongoing challenge and a continuing need. *Nature Reviews Microbiology*, 15(4), pp.217–231.
- Fischer, E., 1894. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3), pp.2985–2993.
- Gaulton, A. et al., 2017. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1), pp.945–954.
- Gfeller, D. & Zoete, V., 2015. Protein homology reveals new targets for bioactive small molecules. *Bioinformatics*, 31(16), pp.2721–2727.
- Gilbert, I.H., 2013. Drug Discovery for Neglected Diseases: Molecular Target- Based and Phenotypic Approaches. *Journal of medicinal chemistry*, 56, pp.7719–7726.
- González-Díaz, H. et al., 2011. MIND-BEST: Web server for drugs and target discovery; Design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from trichomonas gallinae. *Journal of Proteome Research*, 10(4), pp.1698–1718.
- Grisoni, F. et al., 2018. Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach. In Humana Press, New York, NY, pp. 3–53.
- Guerry, P. & Herrmann, T., 2011. Advances in automated NMR protein structure determination. *Quarterly Reviews of Biophysics*, 44(3), pp.257–309.
- Hammes, G.G., 2002. Multiple conformational changes in enzyme catalysis. *Biochemistry*, 41(26), pp.8221–8228.
- Hertzberg, R.P. & Pope, A.J., 2000. High-throughput screening: New technology for the 21st century. *Current Opinion in Chemical Biology*, 4(4), pp.445–451.
- Hughes, J.P. et al., 2011. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), pp.1239–1249.
- Di Ianni, M.E. et al., 2015. Systematic comparison of the performance of different 2D and 3D ligand-based virtual screening methodologies to discover anticonvulsant drugs. *Combinatorial Chemistry and High Throughput Screening*, 18(4), pp.387–398.
- Ilari, A. & Savino, C., 2008. Protein Structure Determination by X-ray Crystallography. *Methods in Molecular Biology*, 452, pp.63-87
- Iorio, F. et al., 2013. Transcriptional data: A new gateway to drug repositioning? *Drug Discovery Today*, 18(7–8), pp.350–357.
- Jesús Naveja, J. et al., 2016. Drug Repurposing for Epigenetic Targets Guided by Computational Methods. In *Epi-Informatics: Discovery and Development of Small Molecule Epigenetic Drugs and Probes*. pp. 327–357.
- Joachim Haupt, V. & Schroeder, M., 2011. Old friends in new guise: Repositioning of known drugs with structural bioinformatics. *Briefings in Bioinformatics*, 12(4), pp.312–326.
- Kaserer, T. et al., 2015. Pharmacophore models and pharmacophore-based virtual screening: Concepts and applications exemplified on hydroxysteroid dehydrogenases. *Molecules*, 20(12), pp.22799–22832.
- Kim, S. et al., 2016. PubChem Substance and Compound databases. *Nucleic acids research*, 44(D1), pp.1202-1213.
- Klug, D.M., et al., 2016. Repurposing strategies for tropical disease drug discovery. *Bioorganic*

- & *Medicinal Chemistry Letters*, 26(11), pp.2569–2576.
- Konc, J. et al., 2015. ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *Journal of Chemical Information and Modeling*, 55(11), pp.2308–2314.
- Koshland, D.E., 1963. The Role of Flexibility in Enzyme Action. *Cold Spring Harbor Symposia on Quantitative Biology*, 28(0), pp.473–482.
- Kumar, A. & Zhang, K.Y.J., 2015. Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*, 71(C), pp.26–37.
- Lamb, J. et al., 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795), pp.1929–1935.
- Law, V. et al., 2014. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, 42(D1).
- Li, J. et al., 2016. A survey of current trends in computational drug repositioning. *Briefings in Bioinformatics*, 17(1), pp.2–12.
- Li, Y.H. et al., 2018. Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Research*, 46(D1), pp.1121–1127.
- Lionta, E. et al., 2014. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, 14(16), pp.1923–1938.
- Liu, X. et al., 2010. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res*, 38(Web Server issue), pp.609–614.
- Liu, Z. et al., 2013. In silico drug repositioning: what we need to know. *Drug discovery today*, 18(3–4), pp.110–115.
- Ma, D.L., et al., 2013. Drug repositioning by structure-based virtual screening. *Chemical Society Reviews*, 42(5), pp.2130–2141.
- Macalino, S.J.Y. et al., 2015. Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research*, 38(9), pp.1686–1701.
- Maggiora, G. et al., 2014. Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 57(8), pp.3186–3204.
- Meng, X.Y. et al., 2011. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2), pp.146–157.
- Meyerhoff, A., 1999. U.S. Food and Drug Administration approval of AmBisome (liposomal amphotericin B) for treatment of visceral leishmaniasis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 28(1), pp.42–51.
- Mignani, S. et al., 2016. Why and how have drug discovery strategies in pharma changed? What are the new mindsets? *Drug Discovery Today*, 21(2), pp.239–249.
- Mirzaei, H. et al., 2015. Energy Minimization on Manifolds for Docking Flexible Molecules. *Journal of Chemical Theory and Computation*, 11(3), pp.1063–1076.
- Mohs, R.C. & Greig, N.H., 2017. Drug discovery and development: Role of basic biological

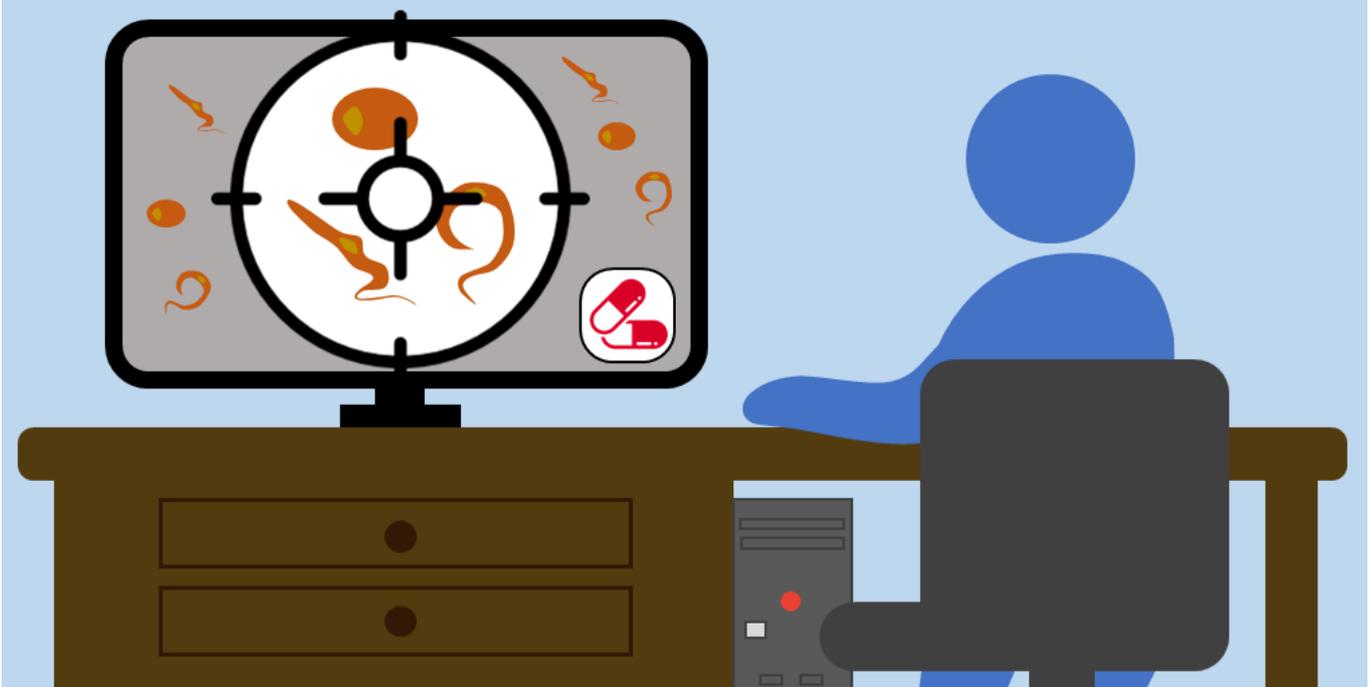
- research. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, 3(4), pp.651–657.
- Moitessier, N. et al., 2008. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. In *British Journal of Pharmacology*, 153, S7-S26.
- Novick, P.A. et al., 2013. SWEETLEAD: an In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLoS ONE*, 8(11), p.e79568.
- Nwaka, S. & Hudson, A., 2006. Innovative lead discovery strategies for tropical diseases. *Nature Reviews Drug Discovery*, 5(11), pp.941–955.
- O'Boyle, N.M. et al., 2011. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1): 33.
- Pacini, C. et al., 2013. DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics*, 29(1), pp.132–134.
- Patterson, S. & Wyllie, S., 2014. Nitro drugs for the treatment of trypanosomatid diseases: Past, present, and future prospects. *Trends in Parasitology*, 30(6), pp.289–298.
- Paul, S.M. et al., 2010. How to improve RD productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, 9(3), pp.203–214.
- Paul, S.M. & Lewis-Hall, F., 2013. Drugs in search of diseases. *Science Translational Medicine*, 5(186), pp.186fs18
- Perola, E. et al., 2004. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Structure, Function and Genetics*, 56(2), pp.235–249.
- Pettersen, E.F. et al., 2004. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem*, 25, pp.1605–1612.
- Pogliani, L., 2000. From molecular connectivity indices to semiempirical connectivity terms: Recent trends in graph theoretical descriptors. *Chemical Reviews*, 100(10), pp.3827–3858.
- Rarey, M. et al., 1996. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3), pp.470–489.
- Sbaraglini, M.L. et al., 2016. Neglected Tropical Protozoan Diseases: Drug Repositioning as a Rational Option. *Current topics in medicinal chemistry*, 16(19), pp.2201–2222.
- Schneidman-Duhovny, D. et al., 2008. PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic acids research*, 36, pp. W223–W228
- Singhal, S. et al., 1999. Antitumor activity of thalidomide in refractory multiple myeloma. *The New England journal of medicine*, 341(21), pp.1565–1571.
- Sottriffer, C., 2009. Virtual screening: principles, challenges, and practical guidelines. *Current opinion in drug discovery & development*, 12(3), p.519.
- Sterling, T. & Irwin, J.J., 2015. ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), pp.2324–2337.
- Subramanian, A. et al., 2005. Gene Set Enrichment Analysis: A Knowledge-Based Approach for

- Interpreting Genome- Wide Expression Profiles Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Source: Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp.15545–15550.
- Swinney, D.C. & Anthony, J., 2011. How were new medicines discovered? *Nature Reviews Drug Discovery*, 10(7), pp.507–519.
- Talevi, A. & Bruno-Blanch, L.E., 2009. Screening virtual: Una herramienta eficaz para el desarrollo de nuevos fármacos en Latinoamérica. *Latin American Journal of Pharmacy*, 28(1), pp.141–150.
- Talevi, A. et al., 2018. The thiol-polyamine metabolism of *Trypanosoma cruzi*: molecular targets and drug repurposing strategies. *Current Medicinal Chemistry*, 25: 1
- Talevi, A., et al., 2011. Virtual Screening: An emergent, key methodology for drug development in an emergent continent – A bridge towards patentability. In *Advanced Methods and Applications in Chemoinformatics: Research Progress and New Applications*. pp. 229–245.
- Temesi, G. et al., 2014. Early repositioning through compound set enrichment analysis: A knowledge-recycling strategy. *Future Medicinal Chemistry*, 6(5), pp.563–575.
- Teo, S.K. et al., 2002. Thalidomide in the treatment of leprosy. *Microbes and Infection*, 4(11), pp.1193–1202.
- Terry, C. & Lesser, N., 2015. Measuring the return from pharmaceutical innovation 2015 Transforming R & D returns in uncertain times. *Deloitte Centre for Health Solutions*.
- Todeschini, R. & Consonni, V., 2009. *Molecular Descriptors for Chemoinformatics*. Wiley-VCH.
- Torreele, E. et al., 2010. Fexinidazole--a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness. *PLoS neglected tropical diseases*, 4(12), p.e923.
- Tufts CSDD, 2014. Cost to Develop and Win Marketing Approval for a New Drug Is \$2.6 Billion. *Tufts CSDD*, 19(1), pp.1–7.
- Ursu, O. et al., 2017. DrugCentral: Online drug compendium. *Nucleic Acids Research*, 45(D1), pp.D932–D939.
- Vargesson, N., 2015. Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Research Part C - Embryo Today: Reviews*, 105(2), pp.140–156.
- Vidović, D. et al., 2014. Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action. *Frontiers in Genetics*, 5(342).
- Walters, W. et al., 1998. Virtual Screening – an overview. *Drug Discovery Today*, 3(4), pp.160–178.
- Weng, H.B. et al., 2018. Innovation in neglected tropical disease drug discovery and development. *Infectious Diseases of Poverty*, 7(1), p.67.
- Wermuth, C.G. et al., 1998. Glossary of terms used in medicinal chemistry (IUPAC recommendations 1997). *Annual Reports in Medicinal Chemistry*, 33, pp.385–395.
- Willett, P., 2006. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11(23–24), pp.1046–1053.

-
- Wishart, D.S. et al., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), pp.D1074–D1082.
- Wong, C.F., 2015. Flexible receptor docking for drug discovery. *Expert Opinion on Drug Discovery*, 10(11), pp.1189–1200.
- Yang, S.Y., 2010. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today*, 15(11–12), pp.444–450.
- Yap, C.W., 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, 32(7), pp.1466–1474.
- Zhou, S. et al., 2013. Thalidomide—A Notorious Sedative to a Wonder Anticancer Drug. *Current Medicinal Chemistry*, 20(33), pp.4102–4108.
- Zulfiqar, B. et al., 2017. Leishmaniasis drug discovery: recent progress and challenges in assay development. *Drug Discovery Today*, 22(10), pp.1516–1531.

CAPÍTULO 3

PRIMERA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN EXPERIMENTAL. METODOLOGÍA



3.1. INTRODUCCIÓN

Las relaciones estructura-actividad pueden ser formuladas por la **Ecuación 3.1**:

$$A = f(x_1, x_2, x_3, \dots, x_d)$$

Ecuación 3.1: Definición de las relaciones estructura-actividad.

Donde **A** es la actividad biológica de un compuesto, la cual es considerada como una función matemática de ciertas características estructurales codificadas en d descriptores moleculares (Grisoni et al. 2018).

Una vez que se ha definido la relación (f) de la ecuación anterior, la actividad (**A**) de un compuesto nuevo o no testado frente a un blanco molecular puede ser inferida desde su estructura molecular mediante el cálculo de los descriptores moleculares seleccionados. Este enfoque es generalmente conocido como **Relaciones Cuantitativas Estructura Actividad** o **QSAR** (del inglés, Quantitative Structure–Activity Relationship) (Davis 2014). Como ya se discutió en el capítulo anterior, se pueden realizar modelos QSAR cuantitativos y cualitativos (funciones discriminantes) dependiendo de la naturaleza de la variable dependiente considerada.

El desarrollo de este tipo de modelos QSAR es un proceso complejo que involucra varios pasos: 1) Compilado y curado de una base de datos de compuestos a los que se les haya medido experimentalmente la actividad **A** de interés; 2) Cálculo de los descriptores moleculares; 3) Extracción de información y generación de modelos; 4) Validación de los modelos.

Conociendo la importancia de las poliaminas en *T. cruzi* y las limitaciones del parásito para biosintetizarlas, el objetivo de este trabajo de tesis fue desarrollar funciones discriminantes que sean capaces de identificar inhibidores del transporte y/o metabolismo de poliaminas en *T. cruzi* en grandes bases de datos de fármacos existentes.

Hemos realizado 2 campañas de SV diferentes, modificando algunos métodos y formas de evaluación entre las campañas. En este capítulo se abordará la metodología empleada para llevar a cabo la primera campaña de SV.

En la **Figura 3.1** se presenta un esquema general del flujo de trabajo que se utilizó para la generación, validación y aplicación de los modelos (funciones discriminantes) en esta primera campaña de SV.



Figura 3.1: Esquema general del desarrollo de modelos

3.2. OBTENCIÓN DE MODELOS QSAR (FUNCIONES DISCRIMINANTES)

3.2.1. COMPILADO DE LA BASE DE DATOS

El primer paso para la obtención de modelos QSAR es la recolección de datos de actividad de compuestos químicos sobre el blanco molecular de interés.

Cuando comenzamos esta tesis, sabíamos *a priori* que debíamos enfrentarnos a una dificultad importante ya que había muy pocos reportes de compuestos a los que se les hubiera determinado la capacidad inhibitoria frente al transporte de poliaminas en *T. cruzi*. Por lo tanto, para resolver este problema, decidimos compilar una base de datos de análogos a las poliaminas que hayan sido evaluados frente a *T. cruzi*, aunque no hayan sido evaluados directamente frente al blanco molecular. Asimismo, también incorporamos a nuestra base de datos compuestos análogos a la pentamidina (un inhibidor reportado del transporte de putrescina) (Díaz et al. 2014) que habían sido evaluados frente a *T. cruzi*.

El compilado de la base de datos es un paso crítico y esencial para la obtención de modelos QSAR de buena calidad (Tropsha 2010; Davis 2014). La base de datos ideal para la obtención de modelos QSAR capaces de identificar inhibidores del transporte y/o metabolismo de poliaminas debería estar formada por compuestos con ciertas características:

- Que hayan sido evaluados en un ensayo *in vitro*, donde se determine la capacidad de inhibición del transporte de poliaminas en *T. cruzi*. Idealmente, se requiere que los compuestos hayan sido evaluados por el mismo método, sobre la misma forma del parásito, bajo las mismas condiciones experimentales, y que el dato de actividad se reporte de la misma manera.
- Que el método de determinación de la actividad tenga poco error.
- Que exista gran diversidad estructural entre los compuestos, para que sean aplicables a espacios químicos amplios.
- Que los valores de la actividad reportada presenten una buena distribución en un rango entre tres y cuatro órdenes logarítmicos. Lo ideal sería que la distribución de la actividad sea uniforme, aunque este requisito es difícil de cumplir y se considera adecuado que la distribución sea aproximadamente normal (Jaworska et al. 2005).

Como mencionamos antes, nuestra base de datos se construyó a partir de la búsqueda exhaustiva de análogos de poliaminas y/o pentamidina que hayan sido evaluados frente a alguna de las formas de *T. cruzi*. Esta búsqueda se realizó en repositorios de artículos científicos como Scopus (<https://www.scopus.com/home.uri>) y PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>). Los datos de actividad obtenidos provienen de diferentes trabajos científicos por lo cual los métodos para su determinación, así como las condiciones experimentales, presentan variaciones. Debido a esta heterogeneidad en los datos se eligió el método de la semicorrelación (o Análisis Lineal Discriminante, ALD) como técnica de modelado, ya que un modelo clasificador permite amortiguar, al menos parcialmente, la variabilidad experimental en la base de datos (Talevi et al. 2011).

Los compuestos de la base de datos fueron posteriormente clasificados en **ACTIVOS** o **INACTIVOS** de acuerdo con el valor de la **concentración efectiva media (EC₅₀)** contra *T. cruzi*:

- **ACTIVOS:** Compuestos con $EC_{50} \leq 20 \mu\text{M}$
- **INACTIVOS:** Compuestos con $EC_{50} > 20 \mu\text{M}$

3.2.2. PARTICIÓN DE LA BASE DE DATOS

Típicamente, para la construcción de modelos QSAR, la base de datos total es dividida en dos conjuntos: el **conjunto de entrenamiento** (con el que se calibrarán los modelos) y el **conjunto de prueba** (con el que se validarán). La ventaja de esta división es que los compuestos del conjunto de prueba son “desconocidos” para los modelos ya que ellos fueron excluidos del proceso de desarrollo de los mismos (Martin et al. 2012).

Se suele considerar que, cuanto mayor es la cantidad de casos disponibles para el entrenamiento y validación de los modelos, más robusto es el modelo obtenido (Hawkins et al. 2008). Si el conjunto de entrenamiento es demasiado pequeño, habrá mayor chance de observar correlaciones espurias y sobreajuste, que irán en detrimento de la capacidad de generalización. Se puede considerar que una situación adecuada en cuanto al número de compuestos en la base de datos está entre 150 – 300 compuestos en total, y para el caso de modelos de clasificación, como nuestro caso, debería haber aproximadamente igual cantidad de compuestos de cada clase o categoría (Tropsha 2010).

Si el número de compuestos encontrados en literatura es muy pequeño, dividir la base de datos en un conjunto de entrenamiento y uno de prueba implica sacrificar ejemplos valiosos para el entrenamiento en beneficio de una validación más rigurosa. Algunos autores han cuestionado la práctica de reservar una muestra de compuestos para la validación externa en aquellos casos en los que no se cuente con un conjunto de datos de al menos cientos de compuestos (Hawkins et al. 2003; Hawkins et al. 2008).

Para realizar la partición de los compuestos entre el conjunto de entrenamiento y el de prueba se pueden utilizar varias estrategias. Uno de los métodos más utilizados es la **asignación al azar** de los compuestos de la base de datos a cada uno de los conjuntos. Esta estrategia es adecuada para bases de datos que superan un número mínimo de estructuras que permite seleccionar conjuntos de prueba y de entrenamiento de tamaño similar (Golbraikh et al. 2003; Golbraikh & Tropsha 2000). La desventaja principal de este método es que, al no proveer un criterio racional para la asignación de las moléculas a cada conjunto de datos, no se puede garantizar que los compuestos de un conjunto estén representados en el otro conjunto en los casos en que las particiones no sean balanceadas en cuanto a número de compuestos. Esto podría ocasionar que los compuestos del conjunto de prueba no sirvan para validar los modelos generados con el conjunto de entrenamiento, o viceversa, que los compuestos del conjunto de entrenamiento no puedan obtener buenos modelos para clasificar a los compuestos del conjunto de prueba; la probabilidad de que surja este problema, lógicamente, aumenta cuanto menor es el tamaño del conjunto de prueba elegido (Hawkins et al. 2008).

Para solucionar esta situación han sido desarrollados algoritmos de partición racional que intentan dividir los datos de una forma más representativa, de manera de poder estimar con mayor confianza la capacidad predictiva de los modelos generados. En los siguientes puntos se discutirá la estrategia racional que hemos utilizado en este trabajo.

3.2.2.1. ESTRATEGIA DE PARTICIÓN REALIZADA EN ESTE TRABAJO

Para particionar la base de datos en el conjunto de entrenamiento y conjunto de prueba de una forma representativa, es decir, que los diferentes motivos estructurales de las moléculas se encuentren en ambos conjuntos, se utilizaron dos metodologías secuenciales de *clustering* (o agrupamiento).

Los métodos de *clustering* pertenecen a la familia de técnicas de clasificación no supervisadas (que no requieren de una variable dependiente), permitiendo la agrupación de datos utilizando sólo características moleculares de los compuestos. Estos métodos han sido ampliamente empleados en campañas de descubrimiento de fármacos para agrupar las moléculas según su quimiotipo. Existen numerosos métodos de *clustering* basados en *fingerprints* y en descriptores moleculares que son usualmente divididos en dos clases: los **métodos jerárquicos** como el agrupamiento de Ward (Ward 1963) y los **métodos no jerárquicos** como el de Jarvis-Patrick (Jarvis & Patrick 1973) y *k-means* (Böcker 2008). En este trabajo de tesis se utilizaron ambas clases de métodos de una forma secuencial como se explica a continuación:

Primero se realizó un ***clustering* jerárquico** utilizando la aplicación **LibraryMCS** que está integrada en el paquete de JChem 6.3.0 (ChemAxon, 2014). Este programa se basa en la máxima subestructura común (MSC) para agrupar un set de estructuras químicas sin recurrir a la comparación exhaustiva de a pares. Se entiende como MSC a la subestructura de mayor tamaño que comparten las moléculas comparadas (Kawabata 2011). En la **Figura 3.2** se puede ver un ejemplo de la MSC entre 2 moléculas.

LibraryMSC, en su primer paso, calcula la similitud química de todas las moléculas a comparar y genera una matriz con los correspondientes coeficientes de similitud. Luego selecciona el par de moléculas con el mayor coeficiente de similitud y busca la MSC entre este par. En el siguiente paso, agrupa todas las moléculas que presentan la misma MSC que el par de moléculas anterior. Con las moléculas que no fueron agrupadas se repiten los mismos 3 pasos, es decir: busca las dos moléculas más similares, se fija cuál es la MSC entre ellas y luego busca en las restantes cuáles presentan la misma MSC. Este ciclo se repite iterativamente hasta que ya no se encuentran pares de estructuras con coeficientes de similitud por encima del umbral utilizado por el algoritmo. Los compuestos que no fueron agrupados son considerados compuestos atípicos u “*outliers*” (Talevi et al. 2012). Los compuestos atípicos no comparten una MSC por encima del mínimo tamaño requerido.

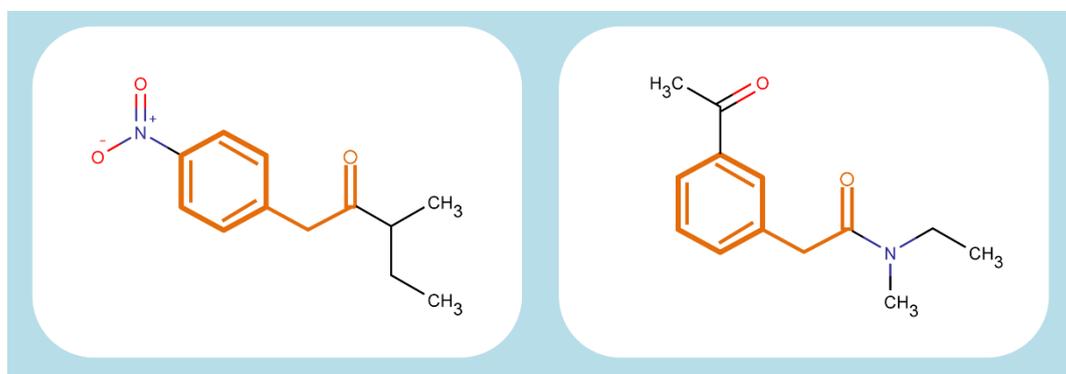


Figura 3.2: Ejemplo de Máxima Subestructura Común (MSC) entre 2 moléculas. En naranja se representa la MSC que comparten las dos moléculas representadas.

Este método se utilizó de manera independiente para los compuestos ACTIVOS e INACTIVOS de la base de datos. Se fijó que las MSC en torno a las cuales se desarrollaron los *clusters* debían tener como mínimo 7 átomos y se permitió que en la determinación de la MSC se utilicen fragmentos de anillos (sin necesidad de utilizar todos los átomos del anillo).

Posteriormente, los *clusters* resultantes para cada categoría (ACTIVOS e INACTIVOS) se optimizaron mediante un método **no jerárquico** de optimización de los *clusters* a través del algoritmo *k-means*.

Este método es comúnmente empleado para particionar automáticamente un set de datos en *k* grupos. El procedimiento se inicia por la selección de *k* centros de *clusters* iniciales y luego estos *clusters* se refinan mediante dos pasos iterativos: 1) Cada objeto *d_i* es asignado a su centro de *cluster* más cercano; 2) Cada centro de *cluster* *C_j* se actualiza para que el próximo centro de *cluster* sea la media de todos sus elementos constituyentes. El algoritmo converge cuando no hay más cambios en la asignación de los objetos a los *clusters* (puede interrumpirse antes de la convergencia si el usuario define un máximo de iteraciones) (Wagstaff et al. 2001). El criterio de agrupación para este método es la minimización de la suma de cuadrados intragrupo (Everitt et al. 2011), **Ecuación 3.2:**

$$\sum_{m=1}^k \sum_{l=1}^{nm} (x_{ml} - \underline{x}_m)^2$$

Ecuación 3.2: *Definición de la minimización de la suma de cuadrados intragrupo.*

Donde *k* representa el número de grupos en los que se particionan los objetos, *nm* es el número de objetos en el grupo *m*, *x_{ml}* denota al vector p-dimensional (considerando que se utilizan *p* variables para caracterizar cada objeto) de observaciones del objeto *l* del grupo *m*, y *x_m* es el vector que representa la media de los elementos del grupo *m* en el espacio p-dimensional.

Este proceso se llevó a cabo utilizando el módulo *Cluster Analysis* del software Statistica 10 (Statsoft, 2011).

Para caracterizar a cada compuesto y definir el espacio p-dimensional en el que se aplicará *k-means*, se seleccionaron 6 descriptores moleculares representativos de diferentes aspectos de las moléculas: peso molecular, logaritmo del coeficiente de reparto octanol-agua (logP), área de la superficie polar, número de aceptores de enlaces de hidrógeno, índice de información del contenido atómico y suma de volúmenes atómicos de van der Waals. Los valores de los descriptores fueron luego normalizados para que los diferentes descriptores se encuentren en la misma escala e influyan de la misma manera en la caracterización de los compuestos.

Los *k* centros de *clusters* iniciales fueron elegidos al azar a partir de los *clusters* obtenidos por LibraryMCS para cada categoría. Hemos configurado el programa para que se realicen 50 iteraciones.

3.2.3. CÁLCULO DE DESCRIPTORES MOLECULARES

Considerando que los modelos a desarrollar se van a aplicar al SV de grandes bases de datos, la aplicación de modelos QSAR que utilicen descriptores dependientes de la conformación sería muy demandante computacionalmente y, por lo tanto, requeriría mucho tiempo. Por ese

motivo, optamos por la generación de modelos QSAR que utilicen sólo descriptores independientes de la conformación para la predicción de la actividad.

Las moléculas de la base de datos fueron dibujadas y visualizadas utilizando el programa MarvinSketch (Versión 6.2.2, ChemAxon, 2014). Posteriormente, para el cálculo de los descriptores moleculares hemos utilizado el software Dragon 6.0 (Milano Chemometrics, 2011) que permite el cálculo de 4885 descriptores moleculares. Algunos de estos descriptores son sencillos como la cantidad de ciertos tipos de átomo, de fragmentos o de grupos funcionales, y otros más complejos como los descriptores topológicos y geométricos.

En este caso, hemos calculado 3668 descriptores moleculares independientes de la conformación por cada compuesto químico de la base de datos. Como ya se dijo, el cálculo de tales descriptores no requiere una pre-optimización de las estructuras moleculares, por lo tanto, es posible explorar grandes bibliotecas de compuestos químicos de manera eficiente.

Los descriptores que resultaron ser constantes y los que presentaban baja variabilidad (desviación estándar < 0.05) para los compuestos del conjunto de entrenamiento fueron excluidos.

3.2.4. MODELADO

Como ya se mencionó, los modelos clasificatorios pueden ser útiles para manejar datos ruidosos. Por ejemplo cuando el resultado de un experimento está asociado a una gran variabilidad o si los datos experimentales provienen de la compilación de datos de una diversidad de laboratorios (Talevi et al. 2011).

En este caso, hemos clasificado a los compuestos en dos categorías (o clases): **ACTIVOS** e **INACTIVOS**. Para la estrategia de modelado, cada categoría de los compuestos se asoció a un dado valor numérico arbitrario que sirve como etiqueta de las clases. Hemos etiquetado a los compuestos **ACTIVOS** como el valor "1" y a los compuestos **INACTIVOS** con el valor "-1". Por lo tanto, esta variable binaria arbitraria es adoptada como variable dependiente para la generación de modelos.

3.2.4.1. SUBESPACIOS ALEATORIOS

El pool total de descriptores moleculares fue dividido en 100 subconjuntos con no más de 250 descriptores cada uno; luego, a partir de cada subconjunto se generó un modelo independiente.

En esta estrategia de generación de modelos en paralelo, que corresponde a un enfoque de subespacios aleatorios, los descriptores moleculares (variables independientes) se muestrean al azar y cada modelo se entrena en un subconjunto del espacio de descriptores (Daho & Chikh 2015; Vyskovsky et al. 2016). Esto conduce a que los modelos individuales no se concentren de manera reiterada en las características que muestran una alta capacidad explicativa en el conjunto de entrenamiento, dando como resultado una mejor capacidad predictiva.

3.2.4.2. SELECCIÓN DE DESCRIPTORES MOLECULARES Y GENERACIÓN DE LOS MODELOS CLASIFICATORIOS

Para seleccionar los descriptores que mejor correlacionan con la variable dependiente en cada uno de los subconjuntos utilizamos la técnica de **semicorrelación Stepwise Foward** (por pasos y hacia adelante) (Toropova & Toropov 2017) mediante el software Statistica 10 (Statsoft, 2011).

En la **Figura 3.3** se puede apreciar cuál es la diferencia entre una correlación tradicional y el método de semicorrelación.

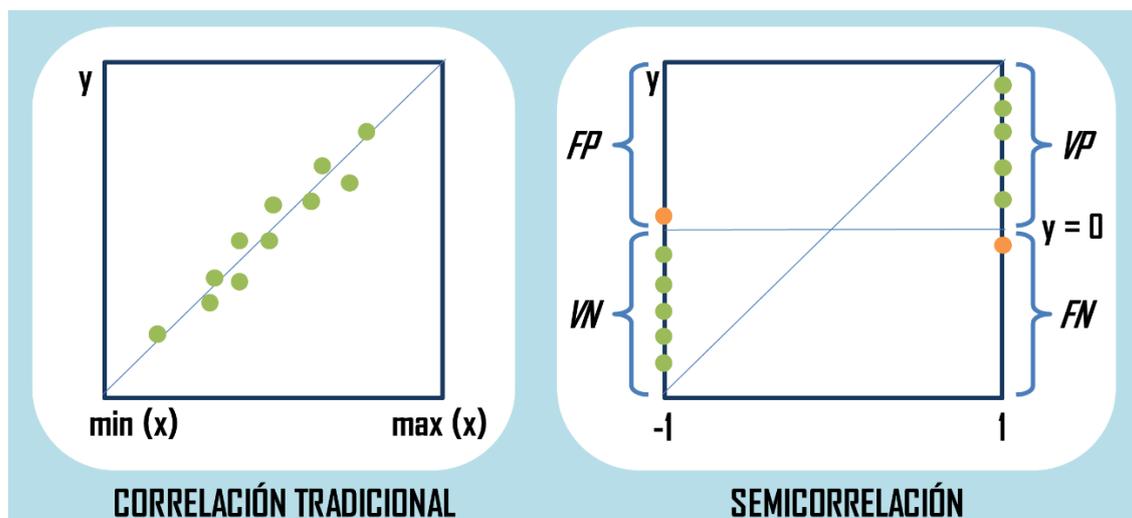


Figura 3.3: Comparación entre una correlación tradicional y una semicorrelación. *VN* indica los verdaderos negativos, *VP* los verdaderos positivos, *FN* los falsos negativos y *FP* los falsos positivos. Adaptado de Toropova & Toropov 2017

Para evitar la inclusión de descriptores altamente correlacionados a los modelos, se seleccionó un valor de **tolerancia** (β) de 0,5. La tolerancia queda definida por la **Ecuación 3.3**:

$$\beta_j = 1 - R_j^2$$

Ecuación 3.3: Definición de la tolerancia.

Donde R_j es el coeficiente de regresión entre el descriptor j y todos los demás descriptores/covariables incluidos en el modelo. La tolerancia es la inversa del Factor de Inflación de la Varianza (FIV); en general, se acepta que valores de $FIV \leq 5$ indican que no existe alta multicolinealidad (Rawlings et al. 1998). Por lo tanto, el valor de β de 0,5 indicaría como máximo un valor de FIV de 2.

También se configuró la cantidad máxima de descriptores que podían incluir los modelos para disminuir la probabilidad de **sobreajuste** (en inglés, **overfitting**). Con este término nos referimos al entrenamiento excesivo de un algoritmo, de manera que el modelo se ajuste demasiado bien a los compuestos del conjunto de entrenamiento (tiene un alto poder explicativo o descriptivo, *memoriza sus características*) a expensas de perder capacidad predictiva por fuera del conjunto de entrenamiento (Peduzzi et al. 1995; Peduzzi et al. 1996). Se recomienda una relación mínima de 15 entre el número de ejemplos de entrenamiento y el número de descriptores; es decir, no más de 1 descriptor al modelo cada 15 moléculas de conjuntos de entrenamiento.

Las **funciones discriminantes (FDs)** obtenidas corresponden a planos (o hiperplanos si son más de tres dimensiones) en el espacio k -dimensional (siendo k el número de variables independientes) que, idealmente, dejan a los ejemplos de cada categoría (ACTIVOS e INACTIVOS) a uno u otro lado del hiperplano. La **Ecuación 3.4** se representa la forma general de una **FD** lineal:

$$FD = a_o + \sum_{i=1}^k a_i d_i$$

Ecuación 3.4: Definición general de las funciones discriminantes.

Siendo a_o la ordenada al origen y a_i el coeficiente de regresión asociado al descriptor molecular d_i .

De esta forma se obtuvieron los 100 modelos lineales clasificatorios.

3.3. VALIDACIÓN DE LOS MODELOS

La parte final en el desarrollo de modelos QSAR es la validación de los mismos, el paso donde se evalúa la robustez y capacidad predictiva de los mismos y, por lo tanto, su capacidad de predecir la actividad biológica (o la clase) de compuestos aún no evaluados (Golbraikh et al. 2003). Los métodos de validación se pueden clasificar en dos categorías: *Validación interna*, que utilizan los compuestos del conjunto de entrenamiento para realizar la validación, y *Validación externa*, que utiliza para evaluar la capacidad predictiva compuestos que no fueron utilizados para la generación de los modelos. Tropsha y colaboradores describen la importancia de utilizar ambos métodos para seleccionar modelos de buena calidad (Tropsha 2010), aunque como se comentó Hawkins y colaboradores proponen omitir la validación externa cuando se trabaja con un número reducido de compuestos (Hawkins et al. 2003).

3.3.1. MÉTODOS DE VALIDACIÓN INTERNA

Los modelos generados fueron validados por dos métodos de validación interna:

Leave Group Out (LGO): Este método, también conocido como “validación cruzada”, se utilizó para determinar la robustez de los modelos generados identificando si entre los compuestos del conjunto de entrenamiento había algunos que influían en mayor medida en los resultados estadísticos del modelo. La técnica consiste en la remoción de un número (n) de compuestos del conjunto de entrenamiento, la generación del modelo con los compuestos restantes y la evaluación del nuevo modelo en el conjunto de compuestos removido. Este proceso se repite hasta que todos los compuestos hayan sido removidos y evaluados por lo menos una vez (Kunal & Supratik 2015).

Aleatorización de Fisher (también llamado y-Randomization): Este ensayo se utiliza para descartar la probabilidad de correlación azarosa entre las variables independientes (los descriptores) y la variable dependiente (la actividad). Lo que intenta este método es eliminar cualquier relación entre las estructuras de los compuestos del conjunto de entrenamiento y la actividad de los mismos, de manera de determinar si el procedimiento de modelado utilizado conduce a modelos espurios con parámetros estadísticos aceptables (Rücker et al. 2007).

Para realizar la técnica aleatorizamos el valor de la variable dependiente de los compuestos del conjunto de entrenamiento. En consecuencia, algunos de los compuestos ACTIVOS serán etiquetados como INACTIVOS y viceversa. Luego volvimos a generar los modelos. Se espera que los modelos generados de esta forma tengan un pobre desempeño estadístico y poca capacidad de predicción. Los resultados de esta validación los expresamos con el promedio de los

porcentajes de buenas clasificaciones obtenidos al aleatorizar 50 veces la variable dependiente y su correspondiente desvío estándar.

3.3.2. VALIDACIÓN EXTERNA

La validación externa consiste en evaluar la capacidad de los modelos en la predicción de la actividad de un conjunto de moléculas que no hayan sido utilizadas para la generación de los modelos. Aunque las validaciones internas hayan obtenido buenos resultados, el poder predictivo de los modelos QSAR sólo se puede establecer si el modelo se aplica con éxito para predecir los compuestos de un conjunto de prueba externo de tamaño adecuado.

Para realizar esta validación, se utilizaron los modelos obtenidos en el conjunto de entrenamiento para predecir la actividad de los compuestos del conjunto de prueba. Los resultados se representan como el porcentaje de buenas clasificaciones (%BC).

3.3.3. EVALUACIÓN DEL DESEMPEÑO DE LOS MODELOS

Dos indicadores relevantes para estimar el desempeño de un modelo QSAR son su **sensibilidad** (Se , la tasa de verdaderos positivos) y su **especificidad** (Sp , la tasa de verdaderos negativos). Estos factores se definen por las siguientes **Ecuaciones 3.5**:

$$Se = \frac{VP}{VP+FN} \quad ; \quad Sp = \frac{VN}{VN+FP}$$

Ecuación 3.5: Definición del coeficiente de Tanimoto.

Donde FN indica los falsos negativos (en nuestro caso, compuestos ACTIVOS que son clasificados por el modelo como INACTIVOS), FP indica los falsos positivos (compuestos INACTIVOS que son clasificados por el modelo como ACTIVOS), VN indica los verdaderos negativos (compuestos INACTIVOS clasificados como INACTIVOS) y VP indica los verdaderos positivos (compuestos ACTIVOS clasificados como ACTIVOS).

Cada molécula del conjunto de datos que se está evaluando obtendrá un resultado numérico único (o *score*) luego de la aplicación del modelo (volcado en el eje de las ordenadas, **Figura 3.3**). Se debe establecer un valor de corte del *score* a partir del cual se considerarán ACTIVOS e INACTIVOS los compuestos calculados. Por medio de la modificación de este valor de corte se puede observar que Se y Sp evolucionan de forma opuesta, por lo tanto, no es posible optimizar ambos parámetros de manera simultánea y se debe encontrar un balance adecuado entre los mismos. Para tal fin, se utilizan las curvas *Receiver Operating Characteristics* (ROC) (y sus métricas derivadas), que son representaciones gráficas de la Se versus $1-Sp$ con las que es posible establecer el valor de corte óptimo de las funciones discriminantes (**Figura 3.4a**).

Por otro lado, el **área bajo la curva ROC (AUROC)** constituye una métrica valiosa para evaluar si el modelo se comporta significativamente mejor que una clasificación al azar (Triballeau et al. 2005) y también permite comparar estadísticamente el desempeño de distintos modelos. Como puede observarse en la **Figura 3.4b**, los ejes del gráfico de la curva ROC adoptan valores entre 0 y 1, por lo que se delimita una superficie cuadrada de área = 1. El modelo clasificador de mínimo poder discriminante (clasificaciones al azar) obtendrá un AUROC de 0,5 y se representa gráficamente como la línea diagonal de la **Figura 3.4b**. La capacidad de predicción de los modelos

irá incrementando a medida que el AUROC aumente (se aleje hacia arriba de la línea diagonal) considerándose como un modelo clasificador perfecto a aquel que obtiene un AUROC = 1.

Por lo tanto, para evaluar y comparar los desempeños de los modelos generados se calcularon y compararon estadísticamente los AUROC mediante el software MedCalc (MedCalc Software, 2011). Para comparar los AUROC fue utilizado el método no paramétrico de DeLong et al. (DeLong et al. 1988).

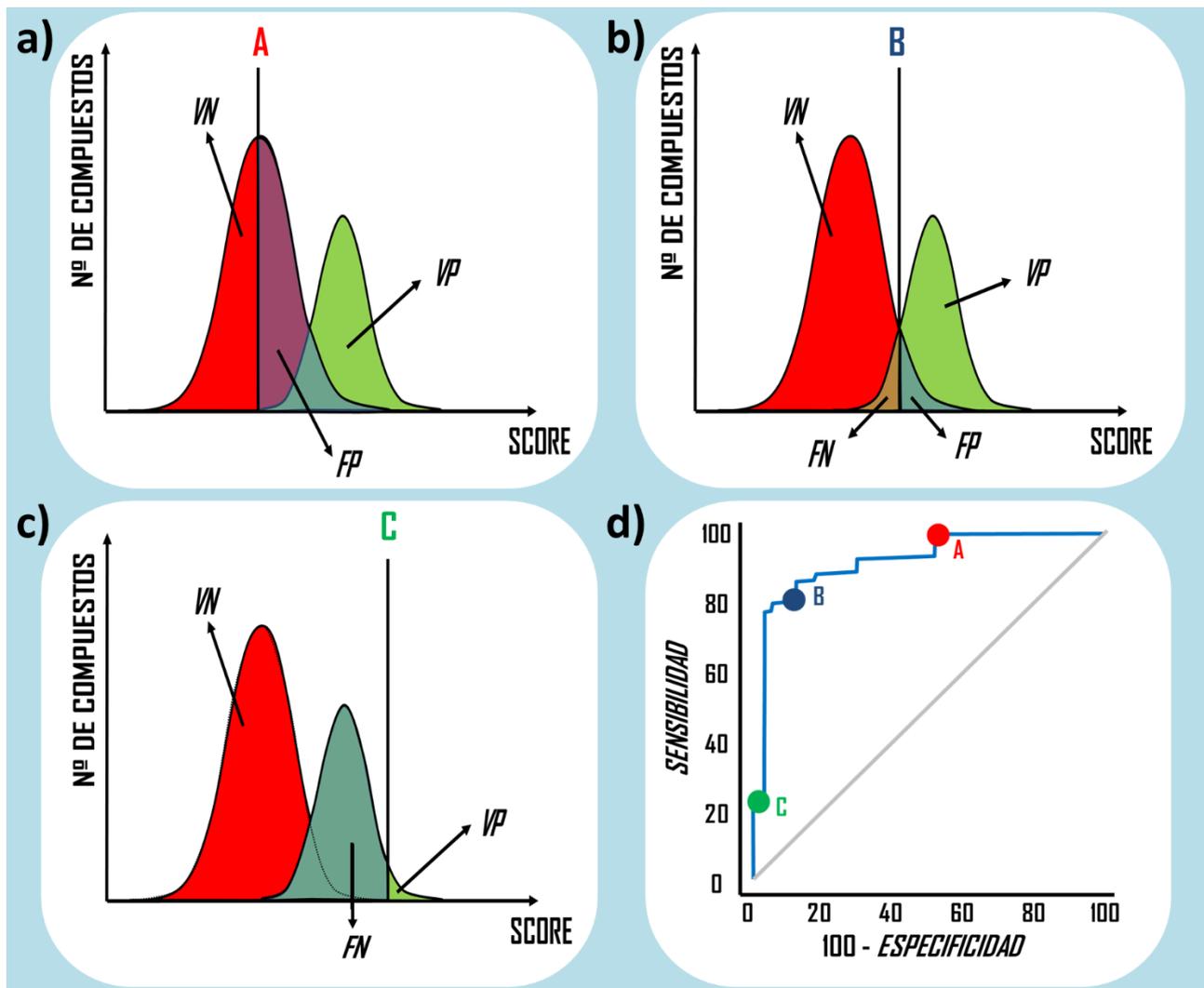


Figura 3.4: Esquema de la construcción de una curva ROC. Las figuras a), b) y c) representan gráficas de número de compuestos (eje y) versus el score obtenido por el modelo (eje x). Se puede observar como varían las proporciones de VN, VP, FN y FP cuando se modifica el valor de corte del score en los puntos A, B y C. d) Curva ROC.

3.3.4. CAMPAÑAS SIMULADAS DE SV

A través de experimentos de SV retrospectivos, Truchon y Bayly han demostrado que la métrica de enriquecimiento AUROC depende de la proporción de activos/inactivos del conjunto de datos a evaluar, y que la desviación estándar de la métrica converge a un valor constante cuando se

utilizan pequeñas proporciones de activos en el conjunto de datos (*Ya*, también llamada rendimiento de activos o prevalencia de activos, donde se ha demostrado que valores de *Ya* por debajo de 0,05 proporcionan resultados más robustos) (Truchon & Bayly 2007).

Además, cuando se evalúan conjuntos de datos con una elevada proporción de compuestos activos se produce el llamado “efecto de saturación”: una vez que los compuestos activos saturan la primera parte de la lista ordenada según el método de rankeado propuesto (el modelo), la métrica de enriquecimiento AUROC ya no puede mejorar; este efecto se atenúa cuando la relación compuestos activos/compuestos inactivos del conjunto de datos a evaluar es mucho menor a 1.

Debe señalarse que los modelos obtenidos se aplicarán a campañas de SV, en donde se espera que haya un pequeño número de compuestos activos inmersos en un número muy grande de compuestos inactivos. Entonces, es deseable que los modelos sean evaluados en una situación similar a la real. Considerando esta situación, hemos evaluado el rendimiento de los modelos en campañas piloto de SV, dispersando los compuestos activos del conjunto de prueba entre una gran cantidad de presuntos inactivos. Estos decoys fueron generados por dos métodos diferentes.

3.3.4.1. BIBLIOTECA PILOTO “SIMULADA”

El primer set de *decoys* lo generamos buscando en PubChem (Kim et al. 2016) compuestos con más de 95% de similitud química (en base al coeficiente de Tanimoto) con respecto a los compuestos inactivos del conjunto de prueba.

Hemos denominado como “**Biblioteca simulada**” a la base de moléculas que se obtiene al adicionar los *decoys* obtenidos por este método a los compuestos del conjunto de prueba.

3.3.4.2. BIBLIOTECA PILOTO DUD-E

Para obtener este segundo grupo de *decoys* hemos utilizado la herramienta de generación de *decoys* automática disponible online **Enhanced Directory of Useful Decoys (DUD-E)** (Mysinger et al. 2012). Mediante esta herramienta se pueden obtener compuestos fisicoquímicamente similares a los compuestos deseados y, a su vez, topológicamente diferentes. En nuestro caso, hemos buscado moléculas fisicoquímicamente similares a los compuestos ACTIVOS del conjunto de prueba, y al mismo tiempo, topológicamente diferentes. Las propiedades fisicoquímicas similares entre los *decoys* y los compuestos del conjunto de prueba generan un desafío más grande para los modelos, mientras que las diferencias en la topología se requieren para minimizar la probabilidad de que los *decoys* se unan al mismo sitio activo que el compuesto activo.

Brevemente, la herramienta DUD-E realiza las siguientes acciones:

- Se compara el compuesto activo deseado frente a un pool de compuestos de la base de datos ZINC (Irwin & Shoichet 2005) mediante 6 propiedades: Peso molecular (PM), una transformación logarítmica del coeficiente de partición octanol-agua estimado teóricamente por el método de Moriguchi (mLogP), el número de enlaces con libre rotación, la carga molecular neta y la cantidad de dadores y aceptores de puentes de hidrógeno. De esta manera se espera obtener entre 3000 y 6000 *decoys* por compuesto activo.
- Se calcula la similitud molecular mediante *fingerprints* ECFP4 (Rogers & Hahn 2010) entre el compuesto activo y los potenciales *decoys* encontrados.

- Se retienen sólo los *decoys* que tienen similitud menor a 0,25 con respecto al compuesto activo. Si hay más de 50 compuestos, se retienen sólo los 50 más disímiles.

3.4. COMBINACIÓN DE MODELOS

El aprendizaje por ensamblado (*ensemble learning*) utiliza múltiples algoritmos de aprendizaje para obtener mejores desempeños predictivos que los que podría obtener de forma individual cualquiera de los ellos (Li et al. 2014). Se ha demostrado que los clasificadores por ensamble (modelos combinados, o meta-clasificadores) proveen una mejor generalización y precisión que los clasificadores individuales (Daho & Chikh 2015; Min 2016; Carbonneau et al. 2016); por lo tanto, pueden ser particularmente útiles para evitar el sobreajuste que se produce al manejar conjuntos de datos que poseen un tamaño de muestra pequeño (pocos compuestos en el conjunto de entrenamiento) mientras que el pool de posibles descriptores moleculares a incluir en el modelo es grande, un escenario bastante frecuente en el campo del diseño de fármacos (Vyskovsky et al. 2016).

Para realizar estos ensambles de modelos, hemos utilizado 5 esquemas diferentes de combinación del score:

- **MÁXIMO (MAX):** El operador máximo devuelve el valor más alto de score entre los scores individuales de los modelos combinados.
- **MÍNIMO (MIN):** El operador mínimo devuelve el valor más bajo de score entre los scores individuales de los modelos combinados.
- **PROMEDIO (PROM):** Este operador devuelve el promedio de comparar todos los scores individuales de los modelos combinados.
- **PROMEDIO DEL RANKING (RANK):** El valor del ranking se obtiene al ordenar los compuestos (desde 1 hasta el número total de compuestos) según su score en cada modelo individual. Luego, para obtener el promedio del ranking (RANK) se promedian los rankings de los modelos individuales combinados.
- **PROMEDIO DEL VOTO (VOT):** El voto fue calculado de acuerdo a la ecuación previamente utilizada por Zhang y Muegge (Zhang & Muegge 2006). Para cada compuesto j , el voto para un modelo $voto_j$ se calcula mediante **Ecuaciones 3.6:**

$$voto_j = \max(0, \text{entero}(11 - \frac{rank_j}{0,02N_{BD}}))$$

Ecuación 3.6: Definición del voto.

donde $rank_j$ indica la posición en el ranking que toma el compuesto j cuando los compuestos se ordenan del mejor al peor en ese modelo y N_{BD} es el número total de compuestos en la base de datos utilizada. La función *entero* redondea un número hasta el entero inferior más próximo. Este procedimiento da 10 votos para el primer 2% de compuestos, 9 votos para el próximo 2%, y así sucesivamente hasta el 18%. A partir de allí, para los compuestos rankeados entre el 18-20%, se da sólo 1 voto. Finalmente, los compuestos en rankeados en el 80% restante no reciben votos. Para obtener el voto promedio (**VOT**) se promedian los votos de los modelos individuales.

3.5. SELECCIÓN DE VALOR DE CORTE PARA EL CRIBADO

Como ya se describió, es posible establecer un valor de corte adecuado para determinar cuáles compuestos serán considerados activos y cuales inactivos mediante la utilización de las curvas ROC.

La elección de una buena relación entre la sensibilidad y la especificidad no es una selección puramente estadística, sino que depende del contexto. En nuestro caso, debido al presupuesto limitado para la adquisición de posibles candidatos, decidimos priorizar la S_p con respecto a la S_e . Esto reduce la probabilidad de encontrarnos con un falso positivo, es decir, un compuesto que podríamos adquirir para su evaluación porque el modelo lo predice como ACTIVO, pero que en realidad es INACTIVO. Sin embargo, como se puede ver en la **Figura 3.4**, aumentar la S_p implica una disminución en la S_e ; es decir, habrá un mayor número de compuestos ACTIVOS que serán clasificados como INACTIVOS.

Para realizar esta selección se obtuvieron las curvas ROC mediante el software MedCalc (MedCalc Software, 2011). Adicionalmente se realizaron gráficos de puntos para ver de una manera más visual como se reparten los compuestos para los diferentes valores de corte.

3.6. CRIBADO VIRTUAL

Una vez que seleccionados los modelos con mejores resultados, se procedió a la aplicación de los mismos al SV de dos bases de datos con el objetivo de identificar qué compuestos podrían inhibir el transporte y/o metabolismo de poliaminas.

Para realizar esta acción, se calcularon los descriptores moleculares que utilizan los modelos y se obtuvieron los *scores* para cada uno de los compuestos de las bases. Teniendo en cuenta la estrategia de reposicionamiento de fármacos, utilizamos las siguientes bases para el cribado:

- (a) **DrugBank 3.0:** Es un repositorio online que compila fármacos aprobados por la FDA para el uso clínico, drogas en etapas de evaluación clínica y drogas en etapas tempranas de evaluación. Esta base recopila, en total, 6684 pequeñas moléculas tipo fármaco, de las cuales 1424 son fármacos ya aprobados por la FDA (Knox et al. 2011).
- (b) **Sweetlead:** Es una base de datos curada que contiene drogas aprobadas por agentes reguladores de otros países como Australia, China e India, compuestos aislados de plantas medicinales tradicionales y otros productos químicos regulados. Esta base contiene 4442 moléculas, de las cuales 2836 son fármacos aprobados (Novick et al. 2013).

3.7. DOMINIO DE APLICACIÓN

Cuando la construcción de los modelos QSAR se realiza a partir de conjuntos de entrenamiento de escasa diversidad molecular, su aplicación a bases de compuestos externas es limitada.

Definir un **dominio de aplicación (DA)**, a veces traducido del inglés como “**Dominio de aplicabilidad**”, para los modelos es esencial para determinar el subespacio de estructuras químicas que pueden ser predichas confiablemente por los modelos. En otras palabras, el grado de generalización de un modelo predictivo depende, entre otras cosas, de cuán amplio es el DA.

Si el DA es demasiado restricto, esto implica que el modelo es capaz de dar predicciones confiables sólo para un limitado número de moléculas (Dimitrov et al. 2005).

La caracterización del espacio de interpolación es muy importante para definir el DA para un dado modelo QSAR. Este espacio queda definido por la naturaleza de los compuestos químicos que conforman el conjunto de entrenamiento. Se han propuesto varios enfoques para determinar el DA, los cuales difieren en la forma en que caracterizan el espacio de interpolación definido por los descriptores utilizados (Sahigara et al. 2012).

Para definir cuáles de los compuestos del cribado caen dentro del DA hemos utilizado un método basado en las distancias, denominado **Regla de la Palanca** (en inglés, **leverage**). Este método consiste en determinar los valores de la palanca (h_i) de los compuestos de la base de datos a evaluar, siendo x_i el vector de los descriptores del compuesto i y X la matriz del modelo derivado de los valores de los descriptores del conjunto de entrenamiento (Tropsha et al. 2003), **Ecuaciones 3.7**:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i = 1, \dots, n)$$

Ecuación 3.7: Definición del valor de la palanca.

Por lo tanto, un compuesto de la base de datos del cribado caerá dentro del DA de un modelo cuando presente un valor de h_i por debajo del valor umbral (h^*), definido en nuestro caso como:

$$h^* = \frac{3f}{n}$$

donde f es el número de descriptores incluidos en el modelo y n el número de compuestos que conforman el conjunto de entrenamiento.

3.8. ENSAYOS EXPERIMENTALES

La evaluación experimental constituye la etapa definitiva para demostrar la validez de los modelos QSAR, por lo que los mejores candidatos del cribado se adquirieron para su evaluación biológica. Por un lado, se determinó el efecto de las drogas sobre la proliferación y viabilidad de los tripanosomátidos (*T. cruzi*, *T. brucei* y *L. infantum*) y, por otro lado, se evaluó la capacidad de estos fármacos de disminuir la incorporación de putrescina y aminoácidos en *T. cruzi*. Adicionalmente, se evaluó el efecto de los mismos sobre la enzima tripanotión sintetasa de *L. infantum* y *T. brucei*.

Por último, los mejores candidatos se evaluaron en un modelo murino de infección aguda.

3.8.1. ACTIVIDAD BIOLÓGICA FRENTE A TRIPANOSOMÁTIDOS

Para todos los ensayos, las soluciones stock y de trabajo de los compuestos evaluados fueron preparadas usando dimetilsulfóxido (DMSO) como solvente y las condiciones de evaluación y de control fueron ensayadas por triplicado. Los valores de EC_{50} fueron determinados a partir de las curvas dosis-respuesta ajustadas a una ecuación sigmoïdal (modelo de Boltzmann) o extrapolados a partir de gráficas de ajuste lineales (Fernández et al. 2013).

Los ensayos de viabilidad de epimastigotes de *T. cruzi*, tripomastigotes sanguíneos de *T. brucei brucei* y promastigotes de *L. infantum* fueron realizados en el Instituto de Ciencia y Tecnología César Milstein (ICT Milstein). El ensayo de viabilidad de tripomastigotes sanguíneos de *T. cruzi* se realizó en el Instituto De Microbiología Y Parasitología Médica (IMPAM).

3.8.1.1. ENSAYOS DE VIABILIDAD - PROLIFERACIÓN DE EPIMASTIGOTES DE *T. cruzi*

Epimastigotes de la cepa Y de *T. cruzi* fueron cultivadas a 28 °C en medio BHT suplementado con 20 mg/mL de hemina, 10% de suero fetal bovino inactivado por calor, 100 µg/mL de estreptomina y 100 U/mL de penicilina. La actividad antiproliferativa de los compuestos fue evaluada a concentraciones entre 3 y 100 µM en cultivos iniciados a una concentración de 10⁷ células/mL. Luego de 4 y 8 días se contó el número de parásitos viables usando una cámara de Neubauer bajo microscopio óptico (Bellera et al. 2013).

3.8.1.2. ENSAYOS DE VIABILIDAD - PROLIFERACIÓN DE TRIPOMASTIGOTES SANGUÍNEOS DE *T. brucei brucei*

La forma sanguínea de *T. brucei brucei* cepa 427 (línea celular 449) (Biebinger et al. 1997), expresando un biosensor redox fue cultivada en una incubadora humidificada a 37 °C con 5% de CO₂ en medio HMI-9 suplementado con 10% (v/v) de FBS, 10 U/mL de penicilina, 10 µg/mL de estreptomina, 0,2 µg/mL de fleomicina, 5 µg/mL de higromicina y, eventualmente, 1 µg/mL de oxitetraciclina para inducir la expresión del biosensor redox. Los tripanosomas fueron cultivados hasta la fase exponencial y diluidos a una densidad celular de 5 x 10⁵ células/mL en medio de cultivo fresco. 2 mL de la suspensión celular fue sembrada en cada pocillo de una placa de cultivo de 96 pocillos. Las drogas fueron inmediatamente agregadas a concentraciones finales en el rango de 0,1 a 25 µM. Se agregó DMSO 1% (v/v) al pocillo control (sin droga). Luego de 24 hs de incubación, los parásitos viables fueron contados usando un citómetro de flujo BD Accuri™ C6 y discriminando células muertas de parásitos vivos por tinción con yoduro de propidio (2 µg/ml).

3.8.1.3. ENSAYOS DE VIABILIDAD - PROLIFERACIÓN DE PROMASTIGOTES DE *L. infantum*

Promastigotes de *L. infantum* (MHOM MA67ITMAP263) fueron cultivados a 28°C en medio RPMI 1640 Glutamax suplementado con 10 % (v/v) de FBS, 10 U/mL de penicilina, 10 µg/mL estreptomina and 25 mM de sal de sodio HEPES pH 7,4. Los cultivos sincronizados fueron sembrados a 5x10⁵ células/mL en placas de 96 pocillos en medio RPMI completo con concentraciones de los compuestos variables (0,2 – 5 µM) en 1% de DMSO. La viabilidad de los parásitos fue determinada luego de 24 hs por recuento de células bajo microscopio óptico utilizando una cámara de Neubauer y calculada como porcentaje de viabilidad en relación con los cultivos control incubados con solo DMSO y con 0,02 – 0,03 mg/mL de tartrato de sodio y potasio trihidratado como droga control.

3.8.1.4. ENSAYOS DE VIABILIDAD DE TRIPOMASTIGOTES SANGUÍNEOS DE

T. cruzi

Tripomastigotes de *T. cruzi* se purificaron en el pico de parasitemia desde sangre periférica de ratones infectados con la cepa RA. Los tripomastigotes (1×10^4 por pocillo) fueron cultivados en una placa de 96 pocillos (volumen final 20 μ l) en medio RPMI suplementado con 10% FBS a 37°C en atmósfera 5% CO₂, y el agregado de la droga para conseguir una concentración final de 20 μ M. Luego de 24 hs los parásitos móviles fueron contados en cámara de Neubauer bajo la luz del microscopio (Miranda et al. 2015). Los resultados se expresan como % de viabilidad de tripomastigotes a 20 μ M de las drogas candidato. Los controles fueron realizados bajo las mismas condiciones de con concentraciones de DMSO iguales a las de las drogas evaluadas. El control negativo fue cultivado con PBS y el control positivo fue cultivado con Benznidazol (20 μ M).

3.8.2. EVALUACIÓN DEL TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS

Este ensayo se realizó en el Instituto de Ciencia y Tecnología César Milstein (ICT Milstein) bajo la coordinación de la Dra. Carolina Carrillo.

Un cultivo de epimastigotes de la cepa Y de *T. cruzi* (3×10^7 parásitos) fue incubado por 24 hs en un medio carente de putrescina y luego cosechado por centrifugación a 1500g por 10 minutos y lavado 3 veces con buffer fosfato salino (PBS). Luego las células fueron resuspendidas en 2 mL de PBS conteniendo 5 μ M de (¹⁴C)-putrescina (110 mCi/mmol), (¹⁴C)-arginina (346 mCi/mmol), (¹⁴C)-lisina (308 mCi/mmol) o (³H)-uridina (36 Ci/mmol), en ausencia o presencia de los compuestos a evaluar a una concentración de 50 μ M (10 veces la Km de la putrescina). Se tomaron alícuotas del cultivo a diferentes tiempos. Estas alícuotas fueron centrifugadas y lavadas dos veces con 1 mL de PBS enfriado y el correspondiente metabolito no radioactivo a 10 mM. Los pellets obtenidos fueron resuspendidos y la radioactividad fue determinada en UltimaGold XR (Díaz et al. 2014). Los candidatos evaluados fueron preparados usando DMSO como solvente y la concentración final de DMSO fue 0,1%. Los controles en ausencia de los candidatos fueron realizados a la misma concentración de DMSO. Todos los experimentos fueron realizados por triplicado.

Los resultados de este ensayo han sido plasmados en las curvas de incorporación de metabolito *versus* tiempo.

Adicionalmente, para chequear si las drogas producen algún efecto citotóxico en los parásitos a tiempos cortos (lo que podría disminuir la incorporación de determinados metabolitos sin necesidad de inhibir específicamente el transporte), fue evaluado el efecto de las drogas en la viabilidad de los parásitos mediante el ensayo de reducción metabólica del bromuro de 3-(4,5-dimetiltiazol-2-ilo)-2,5-difeniltetrazol (MTT) (Mosmann 1983). Brevemente, 10 μ L de MTT 5 mg/mL fueron agregados a un tubo eppendorf conteniendo 10⁶ parásitos en 100 μ L de BHT y las drogas candidatos a evaluar (50 μ M). Luego de la incubación por 3,5 horas a 28 °C, los tubos fueron centrifugados (3000 rpm) y el pellet con los cristales de formazan fue disuelto con 100 μ L de 10% (p/v) de SDS en 0,01 M de HCl. La densidad óptica fue determinada usando un lector de microplaca (Labsystems Multiskan MS, Finlandia) a 570 nm/695 nm. Bajo tales condiciones,

la densidad óptica es proporcional al número de células viables en cada pocillo. Todos los experimentos fueron realizados por triplicado.

3.8.3. EVALUACIÓN DE LA ACTIVIDAD DE LA TRIPANOTIÓN SINTETASA

La actividad de la tripanotión sintetasa (TryS) de *L. infantum* (LITryS) y *T. brucei* (TbTryS) fue determinada usando el reactivo verde de malaquita (BIOMOL GREEN™; Enzo Life Sciences) que permite la medida de la cantidad de fosfato inorgánico liberado desde el ATP durante la catálisis.

El ensayo fue realizado a temperatura ambiente, usando una placa de 96 pocillos y un volumen total de reacción de 50 µL por pocillo: 40 µL de una solución 100 mM de sal de potasio HEPES-K pH 7,4; 0,5 mM EDTA; 10 mM MgSO₄; 5 mM DTT (buffer de reacción); 2 mM espermidina; glutatión (250 µM para LITryS y 50 µM para TbTryS); y 150 µM ATP. Las concentraciones de espermidina y glutatión en la reacción fueron fijadas de acuerdo a los niveles intracelulares reportados en la literatura y la concentración de ATP fue ajustada para evitar interferencia en el ensayo. Se utilizaron 5 µL de LITryS (concentración final 200 nM) (Souza et al. 2014) o TbTryS (concentración final 200 nM) (Maiwald et al. 2014) y 5 µL del candidato seleccionado disuelto en DMSO o 5 µL de DMSO (control de reacción). La reacción fue iniciada por el agregado de la enzima (o buffer de reacción para el blanco) y detenida luego de 15 min con 200 µL de BIOMOL GREEN™. La señal colorimétrica se desarrolló por 20 min y luego fue determinada la absorbancia a 650 nm con un lector de placas MultiScan EX (Thermo Scientific). Todas las reacciones fueron evaluadas en cuadruplicado. Las curvas dosis-respuesta fueron ajustadas a la ecuación sigmoideal de Boltzmann. La actividad residual fue calculada como la media de tres experimentos independientes.

La evaluación de la actividad de tripanotión sintetasa se realizó en el Laboratorio de Biología Redox de Tripanosomátidos del Institut Pasteur de Montevideo, bajo la supervisión del Dr. Marcelo Comini.

3.8.4. MODELO DE INFECCIÓN AGUDA EN ANIMALES

Realizada la evaluación *in vitro*, se decidió avanzar con la evaluación en modelos de infección aguda en ratones de las drogas más promisorias.

Para el ensayo, se utilizaron ratones macho de la cepa C3H/HeN con un peso inicial de 22-24 gramos. Los animales se mantuvieron en condiciones libres de patógenos y alojados en habitaciones con temperatura controlada (22-25°C) recibiendo agua y comida *ad libitum*. Todos los procedimientos se realizaron en conformidad con las normas bioéticas del Comité Institucional para el Cuidado y Uso de Animales de Laboratorio (CICUAL) de la Facultad de Médicas de la Universidad de Buenos Aires, en el Instituto De Microbiología Y Parasitología Médica (IMPAM) bajo la dirección de la Dra. Catalina Alba-Soto. Se definieron los siguientes grupos experimentales (Figura 3.5): animales sin infectar (n=3); animales infectados sin tratamiento (n=5), animales infectados tratados con Benznidazol 100mg/kg/día (n=4), animales infectados tratados con las drogas seleccionadas (n=4 por droga). Las dosis de administración fueron determinadas basándose en las dosis que se utilizan en humanos y convirtiéndolas a dosis en ratones con la fórmula para la transformación de dosis basada en el Área Superficial del Cuerpo (BSA) (Reagan-Shaw 2007).

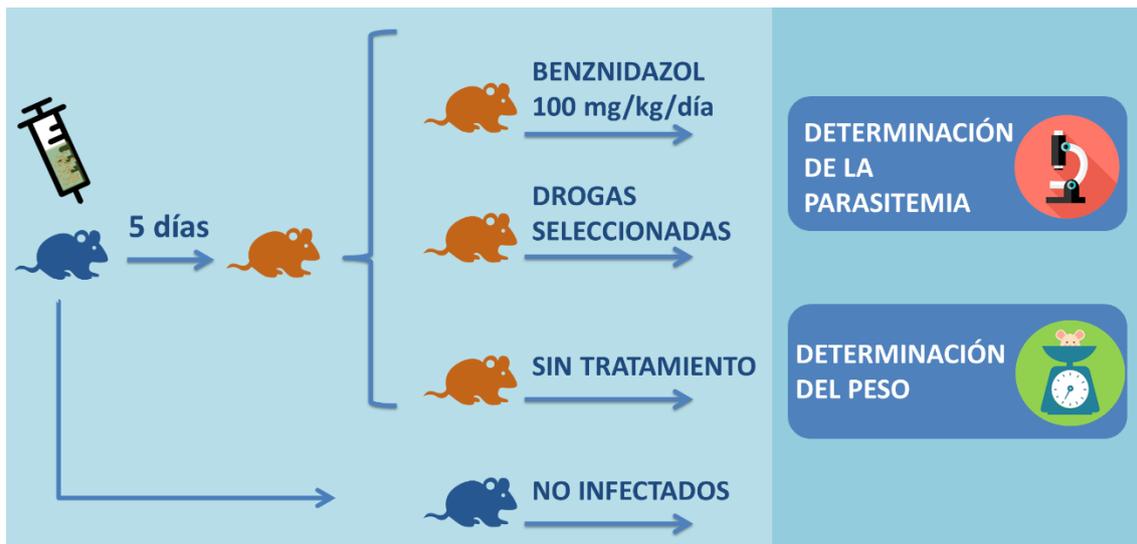


Figura 3.5: Esquema general del ensayo agudo realizado.

Los animales fueron infectados por vía intraperitoneal con 1000 tripomastigotes de *T. cruzi* (cepa RA). Luego de 5 días post infección (dpi) se inició el tratamiento vía oral, utilizando una sonda intraesofágica, durante 5 días corridos.

Se determinó la parasitemia a los 5, 12, 14 y 16 dpi, cuantificando mediante un examen microscópico en cámara de Neubauer. Los resultados se expresaron como número de tripomastigotes/ml de sangre. Adicionalmente, se realizó un seguimiento del peso de los animales cada 2-3 días durante todo el experimento.

REFERENCIAS

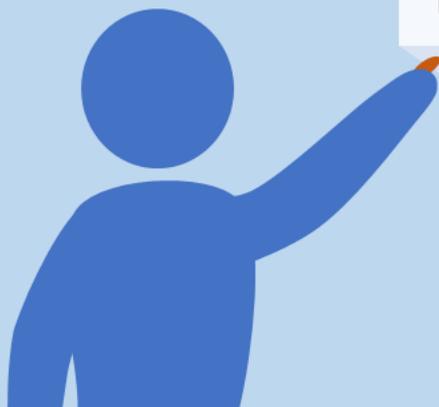
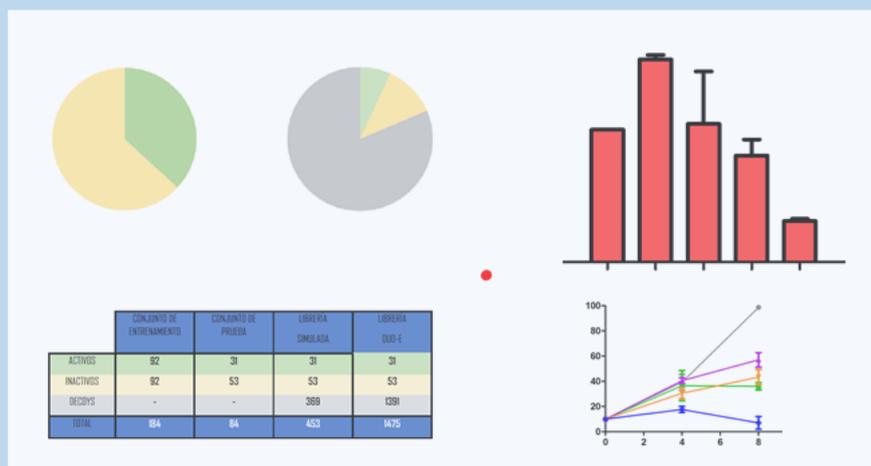
- Bellera, C.L. et al., 2013. Application of computer-aided drug repurposing in the search of new cruzipain inhibitors: Discovery of amiodarone and bromocriptine inhibitory effects. *Journal of Chemical Information and Modeling*, 53(9).
- Biebinger, S. et al., 1997. Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Molecular and biochemical parasitology*, 85(1), pp.99–112.
- Böcker, A., 2008. Toward an improved clustering of large data sets using maximum common substructures and topological fingerprints. *Journal of Chemical Information and Modeling*, 48(11), pp.2097–2107.
- Carbonneau, M.A. et al., 2016. Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognition*, 58, pp.83–99.
- Daho, M.E.H. & Chikh, M.A., 2015. Combining Bootstrapping Samples, Random Subspaces and Random Forests to Build Classifiers. *Journal of Medical Imaging and Health Informatics*, 5(3), pp.539–544.
- Davis, A.M., 2014. CHAPTER 6. Quantitative Structure–Activity Relationships. In *The Handbook of Medicinal Chemistry*. Cambridge: Royal Society of Chemistry, pp. 154–183.
- DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L., 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3), p.837.
- Díaz, M. V. et al., 2014. Pentamidine exerts *in vitro* and *in vivo* anti *Trypanosoma cruzi* activity and inhibits the polyamine transport in *Trypanosoma cruzi*. *Acta Tropica*, 134, pp.1–9.
- Dimitrov, S. et al., 2005. A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. *Journal of Chemical Information and Modeling*, 45(4), pp.839–849.
- Everitt, B.S. et al., 2011. Optimization Clustering Techniques. In Wiley-Blackwell, pp. 111–142.
- Fernández, M. et al., 2013. Oxidovanadium(IV) and dioxidovanadium(V) complexes of tridentate salicylaldehyde semicarbazones: Searching for prospective antitrypanosomal agents. *Journal of Inorganic Biochemistry*, 127, pp.150–160.
- Golbraikh, A. et al., 2003. Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*, 17(2–4), pp.241–253.
- Golbraikh, A. & Tropsha, A., 2000. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular Diversity*, 5(4), pp.231–243.
- Grisoni, F. et al., 2018. Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach. In Humana Press, New York, NY, pp. 3–53.
- Hawkins, D.M. et al., 2008. QSPR checking and validation: A case study with hydroxy radical reaction rate constant. *SAR and QSAR in Environmental Research*, 19(5–6), pp.525–539.
- Hawkins, D.M., Basak, S.C. & Mills, D., 2003. Assessing model fit by cross-validation. In *Journal of Chemical Information and Computer Sciences*. pp. 579–586.

- Irwin, J.J. & Shoichet, B.K., 2005. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model*, 45(December 2004), pp.177–182.
- Jarvis, R.A. & Patrick, E.A., 1973. Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Transactions on Computers*, C-22(11), pp.1025–1034.
- Jaworska, J., Nikolova-Jeliazkova, N. & Aldenberg, T., 2005. QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *ATLA Alternatives to Laboratory Animals*, 33(5), pp.445–459.
- Kawabata, T., 2011. Build-up algorithm for atomic correspondence between chemical structures. *Journal of Chemical Information and Modeling*, 51(8), pp.1775–1782.
- Kim, S. et al., 2016. PubChem Substance and Compound databases. *Nucleic acids research*, 44(D1), pp.D1202-13.
- Knox, C. et al., 2011. DrugBank 3.0: A comprehensive resource for “Omics” research on drugs. *Nucleic Acids Research*, 39(SUPPL. 1).
- Kunal, R. & Supratik, K., 2015. How to Judge Predictive Quality of Classification and Regression Based QSAR Models? In *Frontier in Computational Chemistry*. pp. 71–120.
- Li, L. et al., 2014. Exploration of classification confidence in ensemble learning. *Pattern Recognition*, 47(9), pp.3120–3131.
- Maiwald, F. et al., 2014. 9- and 11-substituted 4-azapallones are potent and selective inhibitors of African trypanosoma. *European Journal of Medicinal Chemistry*, 83, pp.274–283.
- Martin, T.M. et al., 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? *Journal of Chemical Information and Modeling*, 52(10), pp.2570–2578.
- Min, S.H., 2016. A genetic algorithm-based heterogeneous random subspace ensemble model for bankruptcy prediction. *International Journal of Applied Engineering Research*, 11(4), pp.2927–2931.
- Miranda, C.G. et al., 2015. A flow cytometer-based method to simultaneously assess activity and selectivity of compounds against the intracellular forms of *Trypanosoma cruzi*. *Acta Tropica*, 152, pp.8–16.
- Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods*, 65(1–2), pp.55–63.
- Mysinger, M.M. et al., 2012. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14), pp.6582–6594.
- Novick, P.A. et al., 2013. SWEETLEAD: an In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery F. Pappalardo, ed. *PLoS ONE*, 8(11), p.e79568.
- Peduzzi, P. et al., 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), pp.1373–1379.
- Peduzzi, P. et al., 1995. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical*

- Epidemiology*, 48(12), pp.1503–1510.
- Rawlings, J.O., Pantula, S.G. & Dickey, D.A., 1998. Regression Diagnostics. In *Applied regression analysis : a research tool*. Springer, pp. 341–396.
- Rogers, D. & Hahn, M., 2010. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5), pp.742–754.
- Rücker, C., Rücker, G. & Meringer, M., 2007. Y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, 47(6), pp.2345–2357.
- Sahigara, F. et al., 2012. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, 17(5), pp.4791–4810.
- Talevi, A., et al., 2011. Virtual Screening: An emergent, key methodology fro drug development in an emergent continent – A bridge towards patentability. In *Advanced Methods and Applications in Chemoinformatics: Research Progress and New Applications*. pp. 229–245.
- Talevi, A. et al, 2012. *Recent studies on similarity measures and its applications to chemoinformatics and drug design*, Bentham Science Publishers.
- Toropova, A.P. & Toropov, A.A., 2017. CORAL: Binary classifications (active/inactive) for drug-induced liver injury. *Toxicology Letters*, 268, pp.51–57.
- Triballeau, N. et al., 2005. Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of Medicinal Chemistry*, 48(7), pp.2534–2547.
- Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6–7), pp.476–488.
- Tropsha, A., Gramatica, P. & Gombar, V.K., 2003. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. In *QSAR and Combinatorial Science*. pp. 69–77.
- Truchon, J.F. & Bayly, C.I., 2007. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *Journal of Chemical Information and Modeling*, 47(2), pp.488–508.
- Vyskovsky, R. et al., 2016. Random subspace ensemble artificial neural networks for first-episode Schizophrenia classification. In *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*. Gdansk, Poland: IEEE, pp. 317–321.
- Wagstaff, K. et al., 2001. Constrained K-means Clustering with Background Knowledge. *International Conference on Machine Learning*, pp.577–584.
- Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), pp.236–244.
- Zhang, Q. & Muegge, I., 2006. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: Ranking, voting, and consensus scoring. *Journal of Medicinal Chemistry*, 49(5), pp.1536–1548.

CAPÍTULO 4

PRIMERA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN EXPERIMENTAL. RESULTADOS



4.1. COMPILACIÓN DE LA BASE DE DATOS

A partir de la búsqueda bibliográfica realizada se encontraron 20 trabajos científicos donde se reportan datos de actividad tripanocida de compuestos análogos de poliaminas o de la pentamidina (Majumder & Kierszenbaum 1993; Klenke et al. 2001; Stephens et al. 2003; Borges et al. 2004; Menezes et al. 2006; Gonzalez et al. 2007; Silva et al. 2007; Caterina et al. 2008; Da Silva et al. 2008; Daliry et al. 2009; Pacheco et al. 2009; da Silva et al. 2010; Birkholtz et al. 2011; Daliry et al. 2011; De Souza et al. 2011; Lizzi et al. 2012; Zhu et al. 2012; Liew et al. 2013; Patrick et al. 2013; Braga et al. 2014; Díaz et al. 2014).

Con esta información, compilamos una base de 268 estructuras químicas de las cuales 123 fueron consideradas como ACTIVOS y 145 como INACTIVOS. En la **Figura 4.1** se presenta un mapa de calor (*heatmap*) realizado con Gitoools v. 2.3.1 (Perez-Llamas & Lopez-Bigas 2011) donde puede apreciarse la disimilitud molecular entre los compuestos ACTIVOS e INACTIVOS (calculada como 1 menos el valor del correspondiente coeficiente de similitud entre un par cualquiera de moléculas de la base de datos). La disimilitud molecular fue calculada mediante JChem v.6 (ChemAxon, 2013), utilizando *fingerprints* ECFP_4 y el coeficiente de similitud de Tanimoto en su forma binaria. Puede notarse en la figura que hay numerosos pares de compuestos activos e inactivos que son muy similares entre sí (regiones blancas, correspondientes a una disimilitud cercana a 0) lo que representa un desafío mayor para los modelos de clasificación.

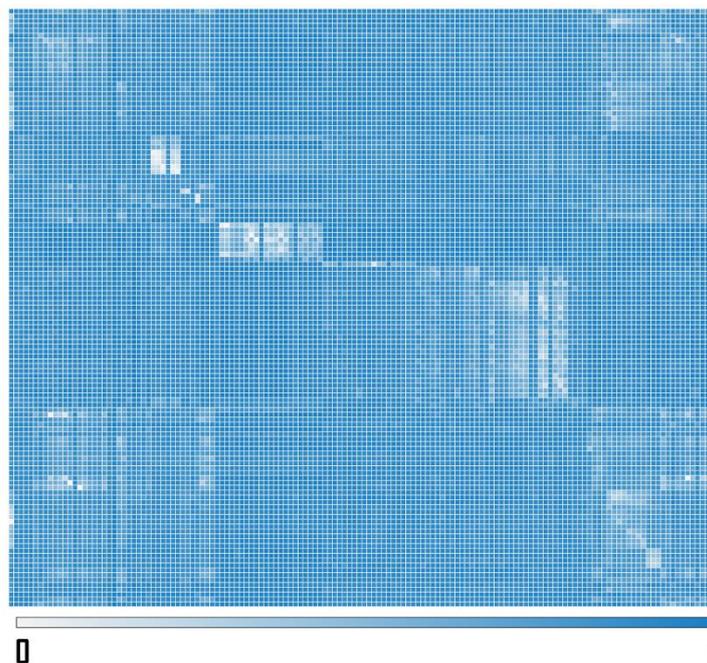


Figura 4.1: Mapa de calor de disimilitud de los compuestos ACTIVOS (eje x) versus los compuestos INACTIVOS (eje y) de la base de datos.

4.2. PARTICIÓN DE LA BASE DE DATOS

Los dos pasos de *clustering* para lograr una partición de los compuestos de la base de datos en conjuntos de entrenamiento y de prueba representativos fueron implementados de manera independiente para los compuestos ACTIVOS e INACTIVOS.

Con respecto a los 123 compuestos **ACTIVOS**, el *clustering* jerárquico resultó en 5 grupos con al menos 5 compuestos cada uno (que en total incluían 101 compuestos), y 22 compuestos atípicos. Por su parte, el *clustering* para los 145 compuestos **INACTIVOS** reveló 10 grupos con al menos 5 compuestos cada uno (que en total incluían 122 compuestos) y 23 compuestos atípicos. En ambos casos, se consideraron compuestos atípicos a todos aquellos para los que se detectaron más de 4 compuestos con la misma MSC.

Se tomó al azar un compuesto de cada uno de los *clusters* para utilizarlos como “semillas” en la optimización de *clusters* mediante k-means. Como resultado de ambas técnicas los compuestos **ACTIVOS** quedaron agrupados en 5 grupos y los compuestos **INACTIVOS** en 10.

Idealmente, el conjunto de entrenamiento debería representar una composición balanceada entre los **ACTIVOS** e **INACTIVOS** para prevenir un sesgo de los modelos hacia la predicción de una categoría particular. Por esta razón, 75% de cada *cluster* de los compuestos **ACTIVOS** (la categoría minoritaria) y 63% de cada *cluster* de los **INACTIVOS** fueron asignados al **conjunto de entrenamiento**, resultando en un conjunto de entrenamiento exactamente balanceado (92 compuestos de cada categoría). Los compuestos restantes conformaron el **conjunto de prueba**. En la **Tabla 4.1** se puede observar la composición final de los conjuntos y bibliotecas químicas que se utilizaron para generar y validar los modelos. En el **Material Suplementario 3** se pueden visualizar las estructuras químicas de los compuestos del conjunto de entrenamiento y de prueba.

Tabla 4.1: Composición de los conjuntos de entrenamiento y de prueba, y de las bibliotecas utilizadas en las campañas de tamizado retrospectivo.

	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	BIBLIOTECA SIMULADA	BIBLIOTECA DUD-E
ACTIVOS	92	31	31	31
INACTIVOS	92	53	53	53
DECOYS	-	-	369	1391
TOTAL	184	84	453	1475

Como ya se explicó en el capítulo anterior, las moléculas del conjunto de prueba se utilizaron para construir dos bibliotecas más numerosas para su utilización en las campañas de SV retrospectivo.

Utilizando los 53 compuestos **INACTIVOS** del conjunto de prueba como *query* de compuestos similares en PubChem, se encontraron 369 estructuras químicas que consideramos “presuntos inactivos” o *decoys*, ya que no se conoce el dato real de su actividad. Integrando los *decoys* al conjunto de prueba se generó la **biblioteca simulada**, donde los compuestos **ACTIVOS** solo representan un 6,84% de la base.

Por otro lado, utilizando los 31 compuestos **ACTIVOS** del conjunto de prueba y la herramienta DUD-E se encontraron 1369 *decoys*. Anexando estos *decoys* al conjunto de prueba obtuvimos una segunda biblioteca para SV retrospectivo, de mayor tamaño y diversidad química que la anterior, a la que denominamos **biblioteca DUD-E**. En ésta, los compuestos **ACTIVOS** representan solamente el 2,10% del total de compuestos.

En la **Figura 4.2** se puede comparar gráficamente la cantidad de compuestos de cada clase en los diferentes conjuntos de moléculas.

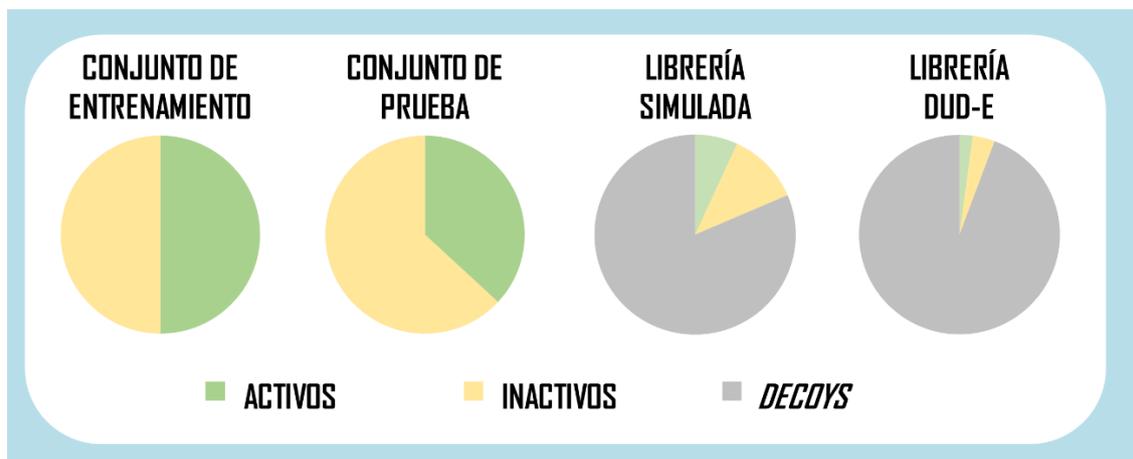


Figura 4.2: Composición de los diferentes conjuntos de moléculas utilizadas para entrenar y validar los modelos

4.3. GENERACIÓN Y VALIDACIÓN DE MODELOS

A partir de cada uno de los 100 subconjuntos de no más de 250 descriptores se generó un modelo clasificatorio capaz de detectar inhibidores del transporte y/o biosíntesis de poliaminas con actividad tripanocida sobre *T. cruzi*.

Estos modelos fueron evaluados mediante validaciones internas en el conjunto de entrenamiento y a través de la determinación del porcentaje de buenas clasificaciones (%BC) y el AUROC en cada conjunto de moléculas. En esta instancia, se consideraron como mejores modelos a aquellos que presentaron mayor AUROC en la biblioteca simulada. De esta manera, los 6 mejores modelos fueron:

Modelo 45:

$$FD = -2.96972 + 0.01213 * nCar + 0.28881 * C-034 + 0.79792 * B10[C-CI] + 0.66370 * B05[O-O] - 1.50430 * ATSC5e + 0.15263 * SM4_Dz(Z) - 0.40497 * NssNH2 - 0.03304 * CATS2D_04_LL + 0.06255 * CATS2D_09_DL + 0.33169 * F04[O-O]$$

Modelo 54:

$$FD = 0.81430 + 0.00268 * P_VSA_p_3 - 0.10301 * SM14_EA(dm) + 0.27199 * F08[C-S] - 1.19365 * C-038 + 0.71810 * B04[C-O] + 1.11877 * B10[C-S] + 0.51879 * B10[C-CI] - 1.41982 * B06[C-C]$$

Modelo 57:

$$FD = -0.2514 + 0.0040 * D/Dtr05 - 0.2842 * SM03_EA(dm) + 0.2881 * O-060 + 0.7888 * B09[C-CI] + 0.8274 * C-019 + 0.0435 * CATS2D_03_DL + 0.2156 * F05[C-S] - 24.0042 * JGI6$$

Modelo 70:

$$FD = -0.112964 - 0.059951 * SM15_EA(dm) + 0.420284 * Eig14_EA(ri) - 0.222050 * CATS2D_09_DD + 0.324251 * F01[C-S] + 0.667668 * B08[C-CI] + 0.012558 * D/Dtr09 + 0.153316 * F06[N-O] - 0.321710 * N-072$$

Modelo 86:

$$FD = -0.31008 + 0.23493*nCIC - 0.73132*B05[N-N] - 0.42672*nR=Ct - 0.31871*SM03_EA(dm) + 0.17102*F06[N-O] + 36.53540*X5Av - 1.60011*GATS1m - 0.20345*H-048$$

Modelo 98:

$$FD = 2.27734 + 0.14469*SaasC + 0.30672*C-034 - 0.86816*B06[C-N] - 0.81956*J_Dz(p) - 1.05954*GATS5s - 0.37712*C-007 - 0.59495*B09[N-N] - 1.42820*MATS2p$$

La nomenclatura de los descriptores es la utilizada por el software Dragon 6.0. En el **Material Suplementario 4** se anexa una breve descripción de los descriptores utilizados por los 6 modelos, y en **Material Suplementario 5** se anexa la explicación teórica de los descriptores del mejor modelo individual (Modelo 45).

Teniendo en cuenta que el conjunto de entrenamiento cuenta con 184 compuestos, puede observarse que estos 6 modelos presentan una buena relación entre el número de compuestos del conjunto de entrenamiento y el número de descriptores incluidos al modelo (entre 18 y 23 moléculas por descriptor), lo que indicaría una baja probabilidad de sobreajuste.

En la **Tabla 4.2** se resumen los resultados obtenidos en las validaciones internas para estos 6 modelos. Los porcentajes de buenas clasificaciones fueron calculados utilizando como valor de corte "0" para diferenciar entre activos e inactivos.

Para la implementación de la validación cruzada removimos consecutivamente 8 grupos de $n = 18$ (9 ACTIVOS y 9 INACTIVOS) y 2 grupos de $n = 20$ (10 ACTIVOS y 10 INACTIVOS) y con los conjuntos de entrenamiento restantes (de 166 o 164 compuestos) se generaron los 10 modelos y se evaluaron en el correspondiente grupo removido. Los resultados de la validación los expresamos como la media aritmética del porcentaje de buenas clasificaciones obtenido por los 10 nuevos modelos generados. Como se observa la **Tabla 4.2**, los porcentajes de buenas clasificaciones promedio son muy similares a los obtenidos por los modelos originales.

En cuanto al ensayo de aleatorización se observa que en todos los casos el rendimiento promedio de los modelos generados al azar es significativamente inferior al rendimiento de los modelos originales, lo que descarta la posibilidad de generación de buenos modelos al azar. Obsérvese que en ningún caso el intervalo obtenido calculando el promedio de buenas clasificaciones ± 3 desvíos estándar para los modelos aleatorizados incluye el porcentaje de buenas clasificaciones obtenido por el modelo original.

Tabla 4.2: Resultados de la validación interna para los 6 mejores modelos.

Nº DE MODELO	MODELO ORIGINAL	VALIDACIÓN CRUZADA		ALEATORIZACIÓN DE FISHER	
	% BC	% BC	DESVIACIÓN ESTÁNDAR	% BC	DESVIACIÓN ESTÁNDAR
45	83	79,78	11,58	64,61	3,71
54	81	81,06	7,62	63,89	4,27
57	81	79,84	9,02	63,08	4,31
70	80	80,17	13,78	62,55	3,46
86	83	82,00	8,41	64,43	3,70
98	83	81,17	10,76	64,61	3,54

La capacidad predictiva para todos los modelos fue determinada en el conjunto de prueba, la biblioteca simulada y la biblioteca DUD-E. Los resultados para los 6 mejores modelos, en término de porcentajes de buenas clasificaciones (utilizando 0 como valor de corte para clasificar en activos e inactivos) y AUROC se encuentran resumidos en la **Tabla 4.3**.

Comparando los resultados del conjunto de entrenamiento y de prueba puede considerarse que 5 de los 6 modelos tienen un cierto grado de sobreajuste ya que su capacidad de predicción disminuye en la validación externa. Sin embargo, cuando se evalúan los modelos en la biblioteca simulada, que presenta un número mayor de compuestos INACTIVOS, los modelos presentan buena capacidad predictiva, con AUROCs entre 0,876 y 0,907 y %BC mayores a 82%. En la **Figura 4.3** se pueden observar las curvas ROC obtenidas para estos 6 modelos en los diferentes conjuntos de moléculas.

Tabla 4.3: Resultados obtenidos para los 6 modelos que presentaron mejor AUROC en la base simulada.

Nº DE MODELO	CONJUNTO DE ENTRENAMIENTO		CONJUNTO DE PRUEBA		BASE SIMULADA		BASE DUD-E	
	% GC	AUROC	% GC	AUROC	% GC	AUROC	% GC	AUROC
98	83	0,880*	66,7	0,742*	90,4	0,907*	74,5	0,853*
54	81	0,862*	68,0	0,775*	84,8	0,887*	68,5	0,711*
70	80	0,878*	69,0	0,784*	82,4	0,884*	64,8	0,641*
86	83	0,865*	82,0	0,854*	88,2	0,878*	69,8	0,738*
57	81	0,864*	72,8	0,745*	85,7	0,877*	74,5	0,766*
45	83	0,884*	70,5	0,767*	83,8	0,876*	77,3	0,886*

* AUROC estadísticamente diferentes a una clasificación al azar ($p < 0,0001$).

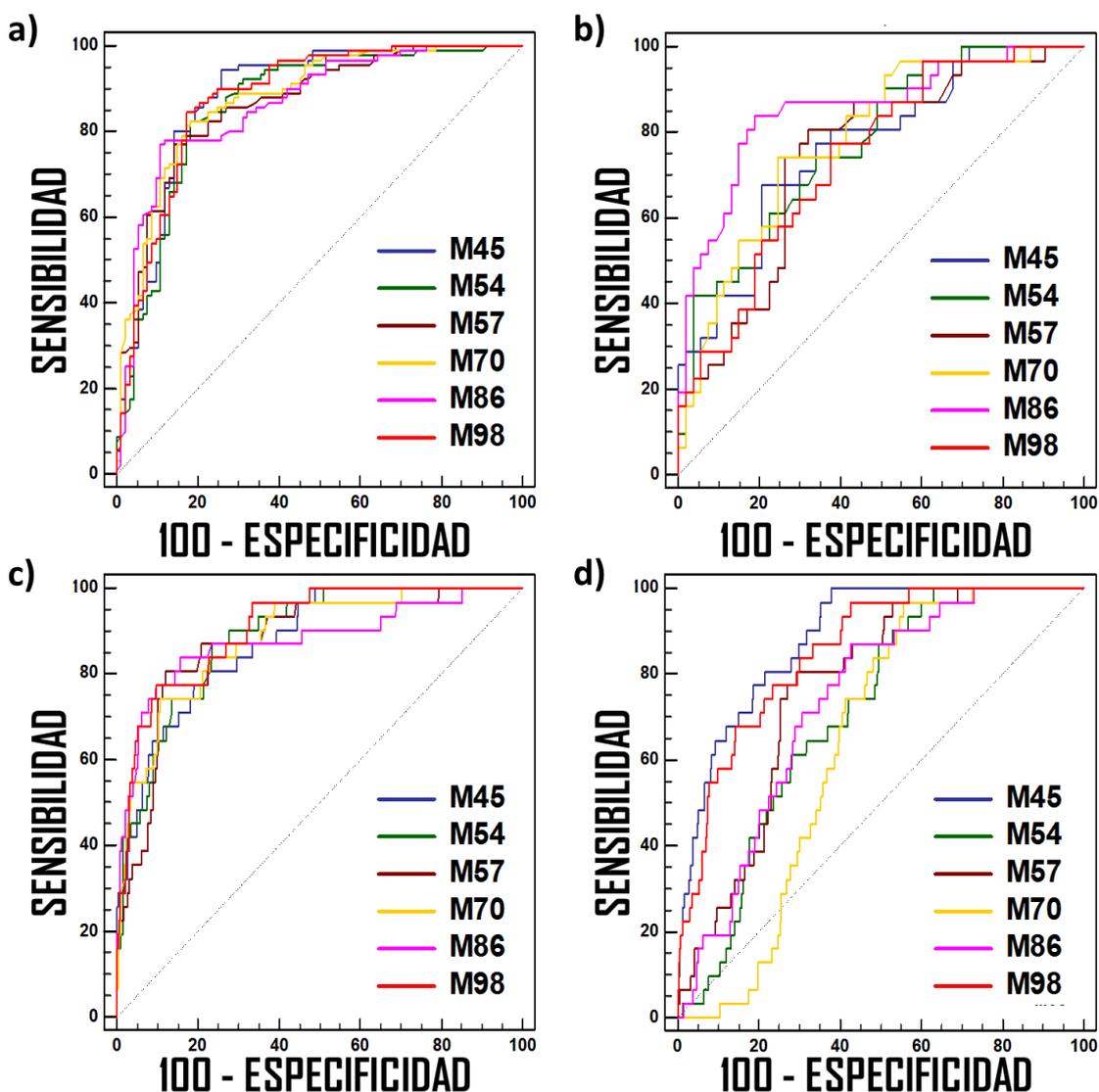


Figura 4.3: Comparación gráfica de las curvas ROC para los 6 modelos seleccionados como mejores modelos en los diferentes conjuntos de moléculas. a) Conjunto de entrenamiento; b) conjunto de prueba; c) biblioteca simulada; d) biblioteca DUD-E.

Como se muestra en la **Tabla 4.3**, los modelos demostraron un buen desempeño en la biblioteca simulada; sin embargo, la capacidad de predicción disminuye para la biblioteca DUD-E. Para mejorar esta situación se recurrieron a las metodologías de combinación de modelos. En este caso, ensamblamos mediante los diferentes esquemas descritos en el capítulo 3, los 6 modelos con mejor AUROC en la biblioteca simulada. Como puede observarse en la **Tabla 4.4**, el desempeño de los modelos mejoró considerablemente en la biblioteca simulada con respecto a los modelos individuales. Sin embargo, cuando se analiza la base DUD-E puede notarse que únicamente el operador **MIN** como criterio de combinación logró mantener el poder discriminatorio observado en la biblioteca simulada. En la **Figura 4.4** puede notarse el comportamiento de las curvas ROC para los ensambles de modelos y el mejor modelo individual.

Tabla 4.4: Valores de AUROC obtenidos para los 5 esquemas de combinación, utilizando los 6 mejores modelos.

	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	BIBLIOTECA SIMULADA	BIBLIOTECA DUD-E
MIN	0,906*	0,825*	0,930*	0,927*
MODELO 98	0,880*	0,742*	0,907*	0,853*
RANK	0,913*	0,803*	0,936*	0,851*
PROM	0,916*	0,801*	0,929*	0,794*
VOT	0,793*	0,657*	0,910*	0,728*
MAX	0,909*	0,709*	0,882*	0,530

* AUROC estadísticamente diferentes a una clasificación al azar ($p < 0,0001$).

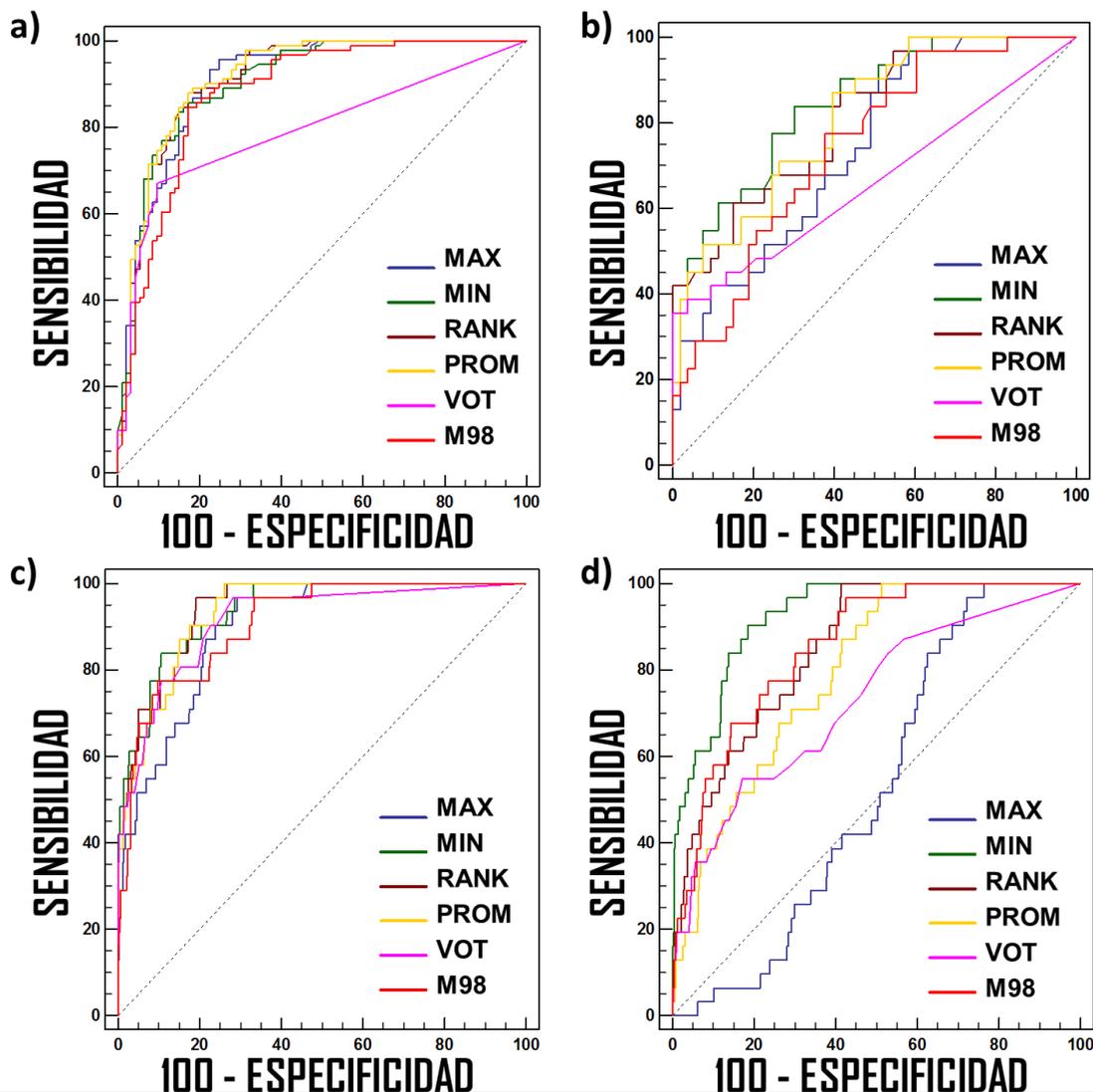


Figura 4.4: Comparación gráfica de las curvas ROC obtenidas por la combinación de los 6 mejores modelos y el mejor modelo individual (M98) en los diferentes conjuntos de datos. a) Conjunto de entrenamiento; b) conjunto de prueba; c) biblioteca simulada; d) biblioteca DUD-E.

Considerando estos resultados, decidimos avanzar con el ensamble MIN para la realización del tamizado virtual prospectivo.

4.4. SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO

La combinación de modelos mediante el operador **MIN** conduce a los mejores resultados en términos de la **Sp**. Esto es importante ya que reduce el número de falsos positivos, es decir, compuestos que podrían ser eventualmente evaluados *in vitro* y en ensayos biológicos dando resultados negativos. La optimización de la **Sp** es también importante en nuestro marco (un laboratorio público con limitados recursos para adquirir y ensayar candidatos) ya que es de esperar que los compuestos inactivos de las bibliotecas/bases de datos tamizadas superen ampliamente en número a los compuestos activos. Por lo tanto, priorizando la **Sp** se maximizan los resultados positivos en la etapa de evaluación experimental, a expensas de perder **Se** en el tamizado.

Como se mencionó en el capítulo anterior, fueron analizadas las curvas ROC para optimizar el valor de corte del score de una forma racional, balanceando la **Sp** y **Se** de una manera acorde a nuestros objetivos. El valor de corte del score escogido para diferenciar entre compuestos **ACTIVOS** e **INACTIVOS** fue **-0,1457**. De acuerdo a los datos de la curva ROC, este valor de corte corresponde a una **Se** igual a **54,84%** y una **Sp** igual a **99,05%** en la biblioteca simulada. En la **Figura 4.5** se muestra un diagrama de puntos de donde se puede visualizar cómo se distribuyen los compuestos en las dos clases cuando se utiliza el ensamble MIN de 6 modelos y el valor de corte seleccionado.

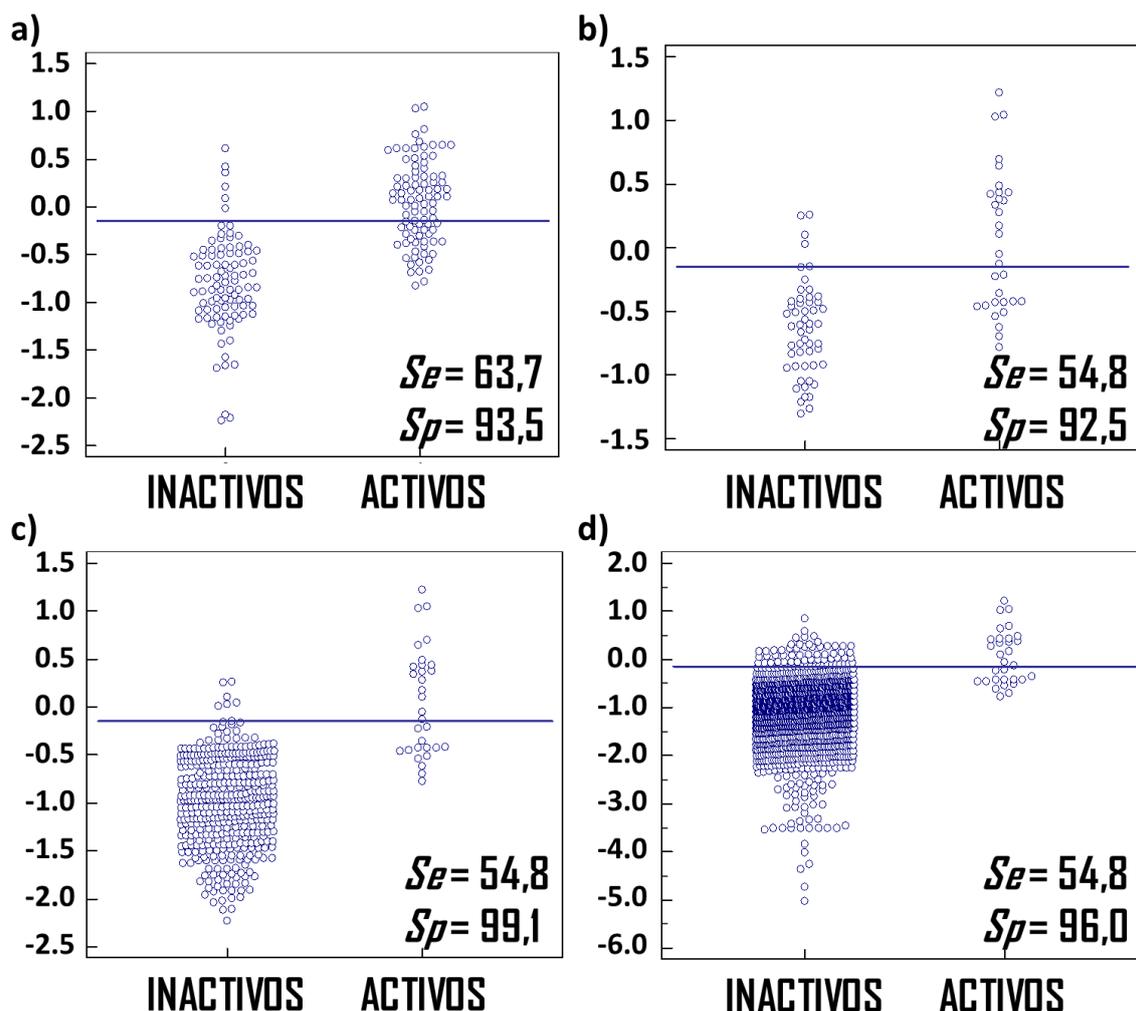


Figura 4.5: Diagrama de puntos para el operador MIN en las diferentes bases de datos. a) Conjunto de entrenamiento; b) conjunto de prueba; c) biblioteca simulada; d) biblioteca DUD-E.

4.5. CRIBADO VIRTUAL

Se aplicó la combinación de modelos MIN en el SV de las bases de datos DrugBank 3.0 y Sweetlead (más de 13000 compuestos) y 156 compuestos fueron seleccionados como ACTIVOS.

45 de estos compuestos son fármacos aprobados, 12 son productos naturales y los 99 compuestos restantes son drogas experimentales. 9 de los fármacos aprobados ya habían sido evaluados previamente contra *T. cruzi* (Amiodarona, Dronedarona, Pranlukast, Terconazol, Perfenazina, Montelukast, Raloxifeno, Carvedilol, Ketoconazol) por lo que fueron descartados para su adquisición y prueba biológica. De cualquier modo, la identificación de estos 9 fármacos aplicando nuestro cribado virtual sugiere que el ensamble de modelos aplicado tiene, verdaderamente, capacidad predictiva.

En base a su accesibilidad (costo y posibilidad de adquisición), 5 de los 45 fármacos aprobados predichos como ACTIVOS y que entraban dentro del dominio de aplicación de los modelos fueron adquiridos para su evaluación experimental: **Triclabendazol** (Sigma-Aldrich), **Sertaconazol** (Sigma-Aldrich), **Pirofosfato de Tiamina** (Sigma-Aldrich), **Mesilato de Eprosartán** (Sigma-Aldrich) y **Paroxetina** (gentilmente donado por el Instituto Nacional de Medicamentos, INAME).

Tiamina no demostraron diferencias significativas con respecto a los cultivos control (con DMSO pero sin droga).

El EC_{50} calculado al día 4 de cultivo fue $4,8 \pm 0,3 \mu\text{M}$ para Triclabendazol, $53,2 \pm 0,3 \mu\text{M}$ para Sertaconazol y $25,9 \pm 0,1 \mu\text{M}$ para Paroxetina.

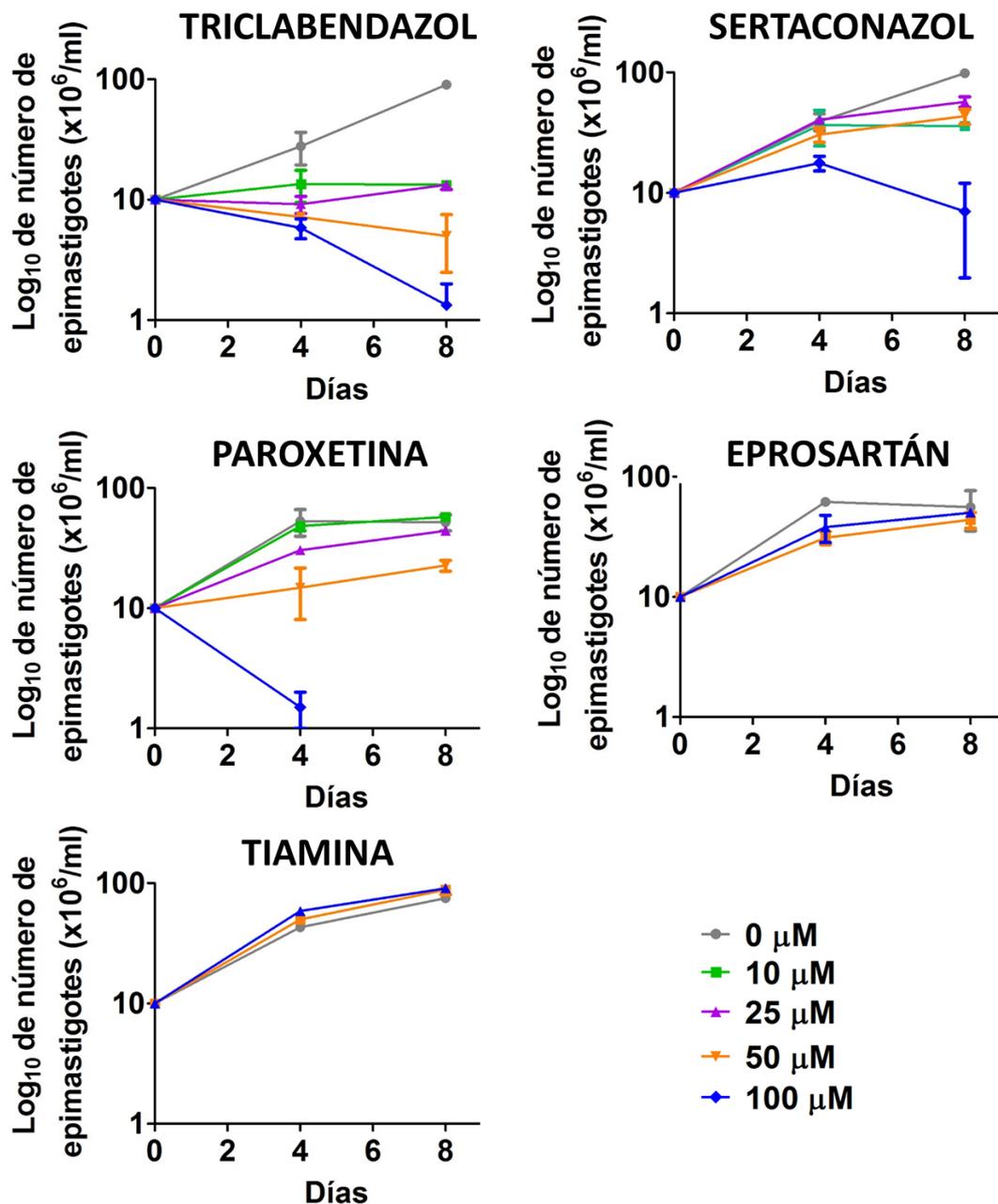


Figura 4.6: Efecto de las drogas sobre la proliferación de epimastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado.

Las dos drogas que exhibieron la actividad tripanocida más potente sobre epimastigotes fueron evaluadas sobre la forma infectiva de *T. cruzi* (los tripomastigotes), y comparadas con el efecto producido por el Benznidazol. Como se observa en la Figura 4.7, Triclabendazol eliminó los

tripomastigotes a concentraciones mayores a 30 μM ; un efecto similar se observó con Paroxetina.

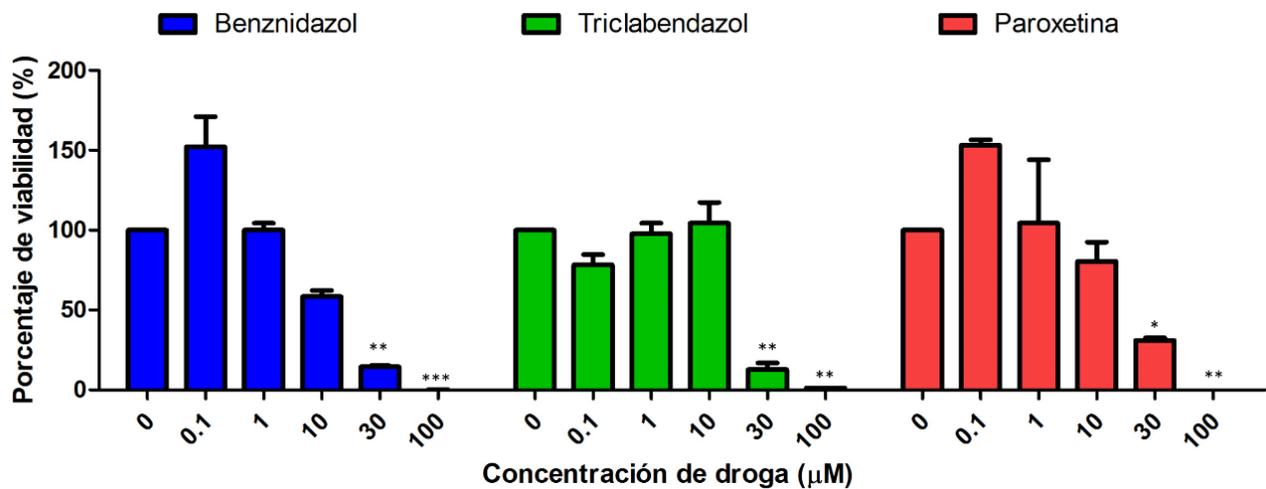


Figura 4.7: Efecto de las drogas sobre la viabilidad de tripomastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado. El análisis estadístico se realizó mediante una prueba de ANOVA de una vía seguida de una prueba de comparación múltiple post hoc de Dunnett (* $p < 0,05$; ** $p < 0,001$; *** $p < 0,0001$ con respecto al control).

Para determinar si las tres drogas que demostraron moderada y alta actividad tripanocida frente a *T. cruzi* tienen una acción tripanocida más amplia, se evaluaron sus efectos sobre la proliferación de las especies patógenas de *Trypanosoma brucei* y *Leishmania infantum*. Los resultados de estos ensayos se encuentran resumidos en la **Tabla 4.6**. Mientras que las 3 drogas demostraron una potente actividad antiproliferativa contra *T. brucei* sanguíneos (EC_{50} menores a 10 μM), solo la Paroxetina fue capaz de inhibir eficientemente la proliferación de los promastigotes de *L. infantum* (**Tabla 4.6**). Triclabendazol y Sertaconazol sólo lograron inhibir la proliferación de *L. infantum* en un 20% y un 40%, respectivamente, cuando se evaluaron a 5 μM .

Tabla 4.6: Efecto inhibitorio de las drogas sobre la viabilidad y proliferación de los tripanosomátidos patógenos.

	EPIMASTIGOTES DE <i>T. cruzi</i> ($\text{EC}_{50} = \mu\text{M}$)	TRIPOMASTIGOTES DE <i>T. cruzi</i> ($\text{EC}_{50} = \mu\text{M}$)	<i>T. brucei</i> 449 SANGUÍNEOS ($\text{EC}_{50} = \mu\text{M}$)	<i>L. infantum</i> PROMASTIGOTES ($\text{EC}_{50} = \mu\text{M}$)
TRICLABENDAOL	4,8	17,8	6,8	No determinado
SERTACONAZOL	53,2	No determinado	7,4	No determinado
PAROXETINA	25,9	7,5	1,2	2,2

4.6.2. EVALUACIÓN DEL TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS

Para confirmar si el mecanismo de acción de los candidatos se correlaciona con el predicho por los modelos, se determinó el efecto de las drogas sobre la incorporación de putrescina en

epimastigotes de la cepa Y de *T. cruzi* y de una cepa mutada (denominada PAT12), que sobreexpresa la permeasa de putrescina (Carrillo et al. 2006; Hasne et al. 2010). Como puede observarse en la **Figura 4.8**, los 3 compuestos evaluados mostraron una reducción en la incorporación de Putrescina desde el medio extracelular, tanto en los epimastigotes salvajes como en los que sobreexpresan el transportador TcPAT12.

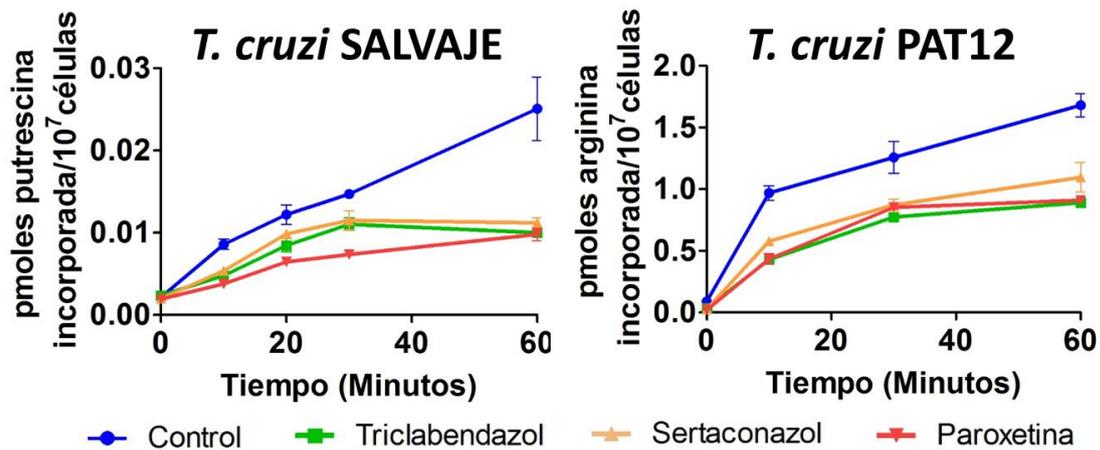


Figura 4.8: Efecto de las drogas sobre la incorporación de putrescina por epimastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado.

Adicionalmente, se evaluó el potencial efecto inhibitorio de estas drogas sobre el transporte de otros aminoácidos sustrato de transportadores de la misma familia de TcPAT12 y sobre el transporte de uridina. Como se muestra en la **Figura 4.9**, Sertaconazol y Triclabendazol inhibieron la captación de los aminoácidos Lisina y Arginina y del nucleósido Uridina, mientras que Paroxetina solo inhibió moderadamente la incorporación de Uridina y Arginina, y no inhibió la incorporación de Lisina.

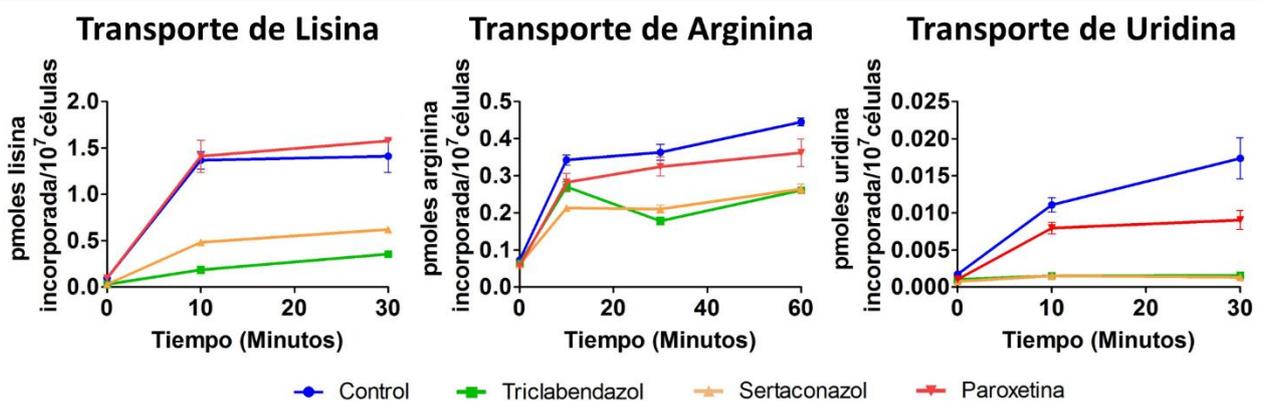


Figura 4.9: Efecto de las drogas sobre los diferentes transportadores de superficie de epimastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado.

El ensayo de MTT indicó que a los tiempos máximos del ensayo de transporte con 50 μ M de los compuestos no se ve afectado el metabolismo energético de los parásitos, por lo que la disminución observada en la incorporación de putrescina y otros metabolitos no se debe a la citotoxicidad de las drogas, sino que sería específico sobre las permeasas.

4.6.3. EFECTO SOBRE LA TRIPANOTIÓN SINTETASA

Teniendo en cuenta que *T. brucei* y *L. infantum* no dependen de la incorporación de poliaminas sino de la síntesis *de novo* para satisfacer sus necesidades biológicas de estas aminas y que las tres drogas evaluadas demostraron efecto tripanocida en estos parásitos, se hipotetizó que los candidatos, en tanto predichos como inhibidores del transporte de poliaminas, podrían estar actuando sobre algún otro punto del metabolismo de poliaminas. Se evaluó si las drogas (a una concentración final de 50 μ M) eran capaces de inhibir la actividad de la enzima tripanotión sintetasa (TryS) de los respectivos parásitos. Ninguna de las drogas inhibió significativamente la actividad de *LITryS* y *TbTryS* descartándose la hipótesis de que el efecto tripanocida se debió a la inhibición de esta enzima.

4.6.4. MODELO MURINO DE INFECCIÓN AGUDA

Los fármacos seleccionados para avanzar hacia su evaluación en el modelo animal de infección aguda fueron Triclabendazol y Paroxetina. Se decidió excluir al Sertaconazol de evaluaciones posteriores ya que el mismo sólo se utiliza, para su indicación terapéutica original, como tratamiento tópico, en tanto las terapias contra la enfermedad de Chagas deben necesariamente ejercer su acción a nivel sistémico. No es, por lo tanto, un candidato directo para reposicionamiento.

Triclabendazol y Paroxetina fueron ensayadas individualmente en un modelo murino de infección aguda de la Enfermedad de Chagas en dosis de 100 mg/kg/día y 50 mg/kg/día respectivamente. Además, se incluyeron los siguientes grupos control: animales tratados con vehículo (DMSO 1%); animales tratados con la droga de referencia, Benznidazol, (100 mg/kg/día); animales sin infectar.

En la **Figura 4.10a** puede observarse la parasitemia para las diferentes condiciones experimentales. Inesperadamente, los ratones tratados con Paroxetina tienden a valores de parasitemia mayores que los animales sin tratar, aunque no se observaron diferencias significativas entre ambos grupos. Por su parte, los tratamientos con Benznidazol y Triclabendazol, redujeron significativamente la parasitemia en comparación con el grupo sin tratar (vehículo) (**Figura 4.10a**). El análisis estadístico se realizó mediante la prueba de Kruskal Wallis sobre los datos agrupados para cada droga, para todos los tiempos.

Complementariamente, se determinaron los pesos de los animales durante todo el experimento, no encontrando diferencias significativas entre los grupos (**Figura 4.10b**).

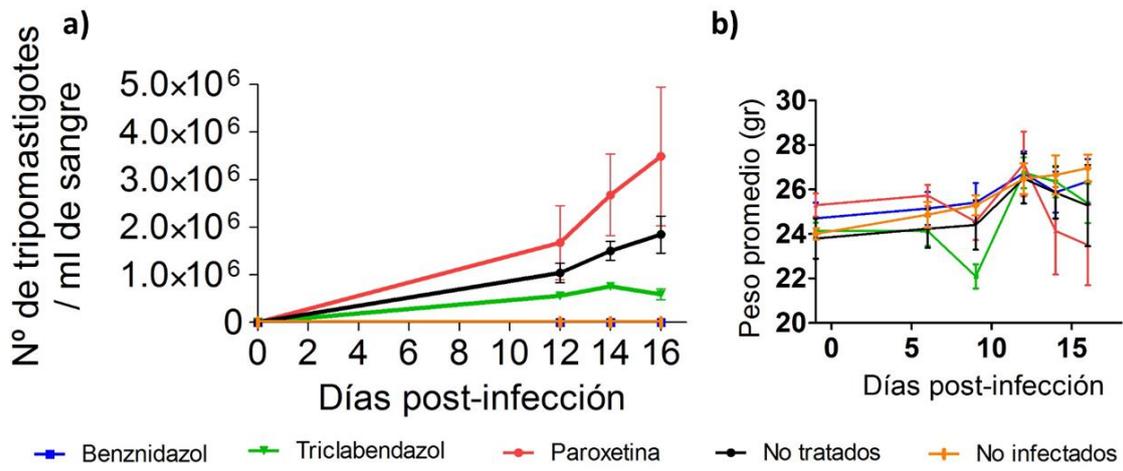


Figura 4.10: Efecto de las drogas seleccionadas en el modelo murino de infección aguda: a) efecto sobre la parasitemia; b) efecto sobre el peso de los ratones.

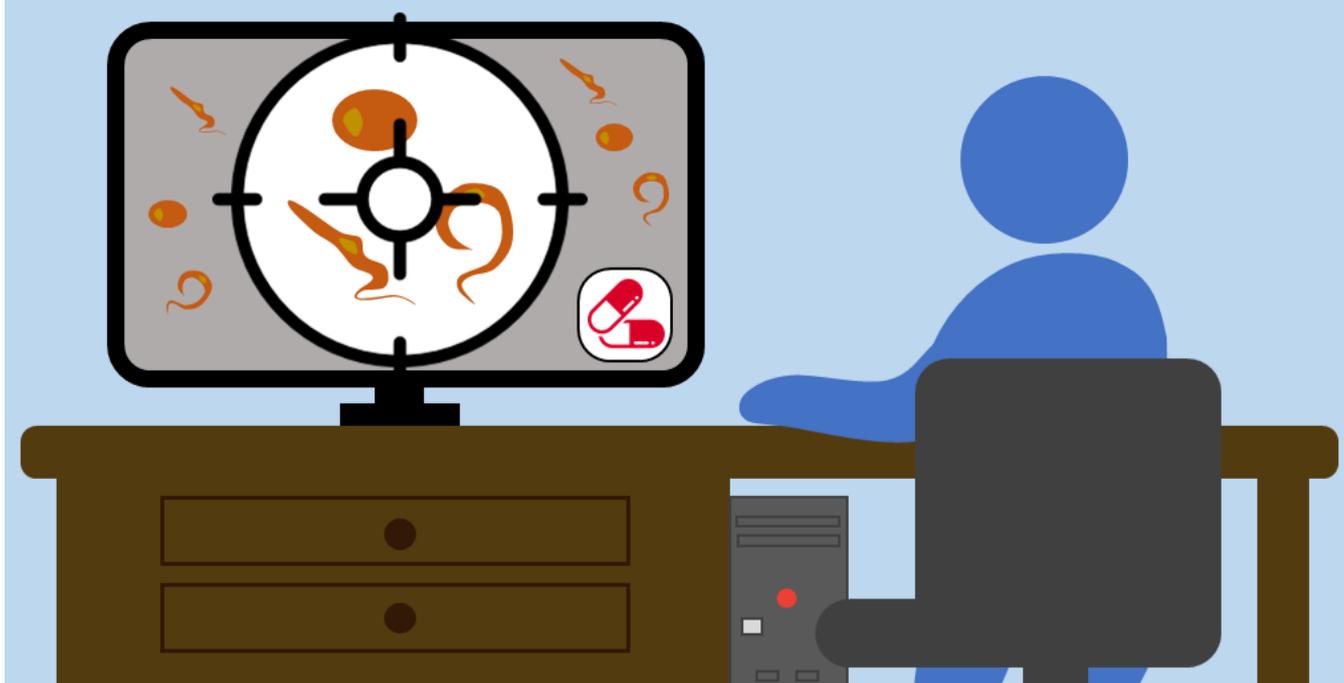
REFERENCIAS

- Birkholtz, L.M. et al., 2011. Polyamine homeostasis as a drug target in pathogenic protozoa: peculiarities and possibilities. *Biochemical Journal*, 438(2), pp.229–244.
- Borges, M.N. et al., 2004. Synthesis, anti-*Trypanosoma cruzi* activity and micelle interaction studies of bisguanilylhydrazones analogous to pentamidine. *European Journal of Medicinal Chemistry*, 39(11), pp.925–929.
- Braga, S.F.P. et al., 2014. Synthesis and evaluation of the antiparasitic activity of bis-(arylmethylidene) cycloalkanones. *European Journal of Medicinal Chemistry*, 71, pp.282–289.
- Carrillo, C. et al., 2006. Molecular and functional characterization of a spermidine transporter (TcPAT12) from *Trypanosoma cruzi*. *Biochemical and Biophysical Research Communications*, 344(3), pp.936–940.
- Caterina, M.C. et al., 2008. Imidazolidines as new anti-*Trypanosoma cruzi* agents: Biological evaluation and structure-activity relationships. *Bioorganic and Medicinal Chemistry*, 16(5), pp.2226–2234.
- Daliry, A. et al., 2009. *In vitro* analyses of the effect of aromatic diamidines upon *Trypanosoma cruzi*. *Journal of Antimicrobial Chemotherapy*, 64(4), pp.747–750.
- Daliry, A. et al., 2011. The trypanocidal activity of amidine compounds does not correlate with their binding affinity to *Trypanosoma cruzi* kinetoplast DNA. *Antimicrobial Agents and Chemotherapy*, 55(10), pp.4765–4773.
- Díaz, M. V. et al., 2014. Pentamidine exerts *in vitro* and *in vivo* anti *Trypanosoma cruzi* activity and inhibits the polyamine transport in *Trypanosoma cruzi*. *Acta Tropica*, 134, pp.1–9.
- Gonzalez, J.L. et al., 2007. Synthesis and antiparasitic evaluation of bis-2,5-[4-guanidinophenyl]thiophenes. *European Journal of Medicinal Chemistry*, 42(4), pp.552–557.
- Hasne, M.P. et al., 2010. A high-affinity putrescine-cadaverine transporter from *Trypanosoma cruzi*. *Molecular Microbiology*, 76(1), pp.78–91.
- Klenke, B. et al., 2001. Synthesis and biological evaluation of s-triazine substituted polyamines as potential new anti-*Trypanosomal* drugs. *Journal of Medicinal Chemistry*, 44(21), pp.3440–3452.
- Liew, L.P.P. et al., 2013. Discovery and preliminary structure-activity relationship analysis of 1,14-sperminediphenylacetamides as potent and selective antimalarial lead compounds. *Bioorganic and Medicinal Chemistry Letters*, 23(2), pp.452–454.
- Lizzi, F. et al., 2012. Conjugation of quinones with natural polyamines: Toward an expanded antitrypanosomatid profile. *Journal of Medicinal Chemistry*, 55(23), pp.10490–10500.
- Majumder, S. & Kierszenbaum, F., 1993. Inhibition of host cell invasion and intracellular replication of *Trypanosoma cruzi* by N,N'-bis(benzyl)-substituted polyamine analogs. *Antimicrobial Agents and Chemotherapy*, 37(10), pp.2235–2238.
- Menezes, D. et al., 2006. Putrescine analogue cytotoxicity against *Trypanosoma cruzi*. *Parasitology Research*, 98(2), pp.99–105.
- Pacheco, M.G. de O. et al., 2009. *Trypanosoma cruzi*: Activity of heterocyclic cationic molecules

- in vitro*. *Experimental Parasitology*, 123(1), pp.73–80.
- Patrick, D.A. et al., 2013. Synthesis and antiprotozoal activity of dicationic m-terphenyl and 1,3-dipyridylbenzene derivatives. *Journal of Medicinal Chemistry*, 56(13), pp.5473–5494.
- Perez-Llamas, C. & Lopez-Bigas, N., 2011. Gitoools: Analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE*, 6(5).
- Silva, C.F. et al., 2007. Activity of “reversed” diamidines against *Trypanosoma cruzi* “*in vitro*”. *Biochemical Pharmacology*, 73(12), pp.1939–1946.
- da Silva, C.F. et al., 2010. The biological *in vitro* effect and selectivity of aromatic dicationic compounds on *Trypanosoma cruzi*. *Memorias do Instituto Oswaldo Cruz*, 105(3), pp.239–245.
- da Silva, C.F. et al., 2008. *In vitro* and *in vivo* studies of the trypanocidal activity of a diarylthiophene diamidine against *Trypanosoma cruzi*. *Antimicrobial Agents and Chemotherapy*, 52(9), pp.3307–3314.
- De Souza, E.M. et al., 2011. Trypanocidal activity and selectivity *in vitro* of aromatic amidine compounds upon bloodstream and intracellular forms of *Trypanosoma cruzi*. *Experimental Parasitology*, 127(2), pp.429–435.
- Stephens, C.E. et al., 2003. The activity of diguanidino and “reversed” diamidino 2,5-diarylfurans versus *Trypanosoma cruzi* and *Leishmania donovani*. *Bioorganic and Medicinal Chemistry Letters*, 13(12), pp.2065–2069.
- Zhu, X. et al., 2012. Evaluation of arylimidamides DB1955 and DB1960 as candidates against visceral leishmaniasis and Chagas’ disease: *In vivo* efficacy, acute toxicity, pharmacokinetics, and toxicology studies. *Antimicrobial Agents and Chemotherapy*, 56(7), pp.3690–3699

CAPÍTULO 5

SEGUNDA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN EXPERIMENTAL. METODOLOGÍA



En este capítulo describimos la metodología utilizada para la realización de una segunda campaña de SV en busca de nuevos inhibidores del transporte de putrescina.

Para esta segunda campaña utilizamos una **estrategia combinada de SV**, utilizando **metodologías secuenciales basadas en el ligando y en el blanco molecular**. La metodología basada en el ligando fue similar a la utilizada en la primera campaña de SV, con algunas modificaciones que permitieron mejorar la capacidad predictiva de los modelos. Con respecto a la estrategia basada en el blanco molecular, se utilizó el *docking molecular* para identificar cuáles de los candidatos seleccionados por los modelos basados en el ligando interaccionarían mejor con el sitio activo de un modelo del transportador de putrescina TcPAT12.

En la **Figura 5.1** se presenta un esquema resumido de la estrategia descrita en este Capítulo.

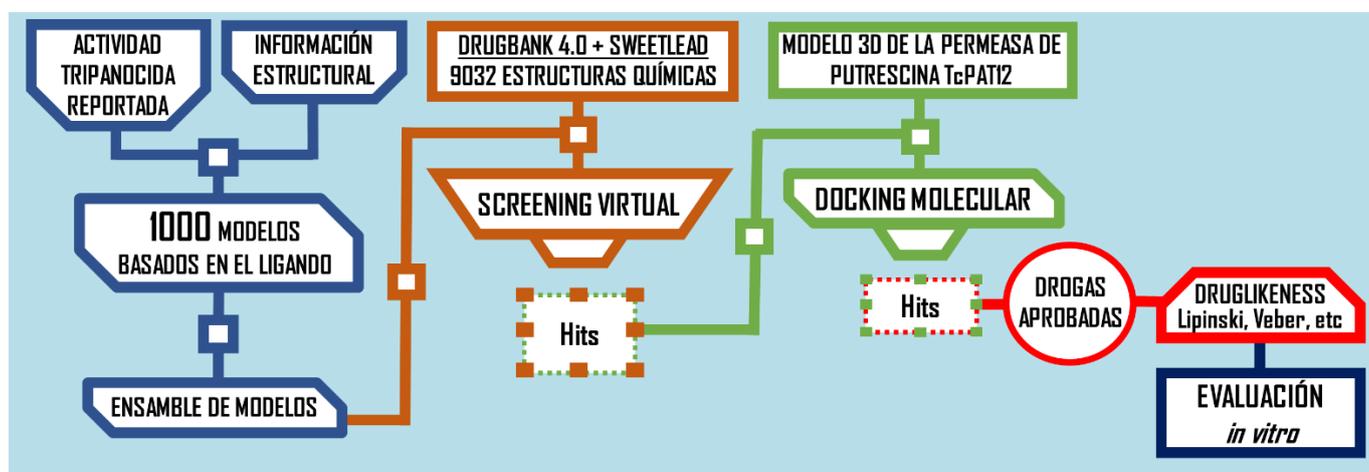


Figura 5.1: Esquema de la estrategia utilizada en la segunda campaña de SV para encontrar inhibidores de TcPAT12.

5.1. OBTENCIÓN DE FUNCIONES DISCRIMINANTES

5.1.1. CURADO Y PARTICIÓN DE LA BASE DE DATOS

La base de datos de 268 estructuras químicas que se describió en el capítulo anterior fue curada y estandarizada utilizando Instant JCHEM v17.2.6.0 (Chemaxon, 2017). Se realizaron los siguientes cambios:

- Si dos compuestos obtenían coeficiente de Tanimoto igual a 0 y ambos pertenecían a la misma categoría (ambos ACTIVOS o ambos INACTIVOS), verificada la identidad del compuesto se eliminaba uno de ellos.
- Si dos compuestos obtenían coeficiente de Tanimoto igual a 0 y ambos pertenecían a diferentes categorías, verificada su identidad se eliminaban ambos compuestos.

En síntesis, los dos pasos anteriores se orientan a eliminar estructuras duplicadas en la base de datos o estructuras que, según la fuente que se utilice, serían clasificadas de acuerdo al dato de actividad reportado y al valor de corte de actividad utilizado como activas o inactivas. Tales compuestos podrían ser considerados como “ruido de clase” (*class noise*), y es habitual cuando

el dato de actividad reportada se encuentra próximo al valor umbral arbitrariamente fijado para separar activos de inactivos (Cruz-Monteagudo et al 2014). Como se indicó en la descripción metodológica de la primera campaña de SV, se utilizó como valor de corte 20 μM en la concentración efectiva media (EC_{50}) contra *T. cruzi* para separar en ACTIVOS e INACTIVOS.

La partición de los datos en el conjunto de entrenamiento y conjunto de prueba se realizó mediante las mismas estrategias mencionadas en el Capítulo 3 con la única diferencia de que se utilizó una versión más actual del software para el clustering jerárquico (LibraryMCS, JCHEM v17.2.13.0, Chemaxon, 2017).

5.1.2. MODELADO

Nuevamente utilizamos el software Dragon 6.0 (Milano Chemometrics, 2011) para el cálculo de 3668 descriptores moleculares independientes de la conformación por cada estructura de la base de datos. Sin embargo, esta vez no fueron excluidos los descriptores con baja variabilidad (desviación estándar <0.05) para los compuestos del conjunto de entrenamiento ya que los mismos serán excluidos en la etapa de generación de los modelos.

Debido a la variabilidad asociada a la determinación experimental de las EC_{50} , nuevamente se recurrió a la obtención modelos clasificatorios. Para la nueva estrategia de modelado, y en función de las restricciones del software utilizado en esta segunda campaña, hemos modificado el valor de las etiquetas arbitrarias que le habíamos asignado a los compuestos de cada categoría. A los compuestos ACTIVOS los etiquetamos con el valor “1” y a los compuestos INACTIVOS con el valor “0”. Esta nueva variable binaria arbitraria fue adoptada como variable dependiente para la generación de modelos.

Desde este punto y en adelante se empleó el **lenguaje y entorno R** para la manipulación y análisis de los datos (R Foundation for Statistical Computing 2016).

El paquete de R “**data table**” (<https://cran.r-project.org/package=data.table>) fue utilizado para ingresar el conjunto de datos al entorno R. Se aplicó el método de los subespacios aleatorios en el conjunto de 3668 descriptores moleculares para generar 1000 subconjuntos de no más de 200 descriptores moleculares cada uno. Para la realización de estas aleatorizaciones de descriptores y particiones en subgrupos se generó un algoritmo *ad hoc* en la plataforma R. En el **Material Suplementario 6** se encuentran los *scripts* utilizados para la obtención de los modelos.

A partir de estos subgrupos se generaron 1000 modelos clasificatorios mediante la aplicación de la función “**lm**” del paquete “**stats**” integrado en R. Para evitar la incorporación de descriptores correlacionados se configuró un Factor de Inflación de la Varianza (FIV) de 2 como máximo, y para evitar el sobreajuste se indicó que se incorpore como máximo un descriptor por cada 15 compuesto del conjunto de entrenamiento.

La validación interna de los nuevos modelos se realizó utilizando las mismas dos estrategias que antes: LGO y ensayo de aleatorización. Estas dos estrategias se realizaron mediante algoritmos implementados en el paquete **caret** de R (Kuhn 2008).

La validación externa se realizó sobre los compuestos del conjunto de prueba determinando el porcentaje de buenas clasificaciones (%BC) y el AUROC mediante el paquete de R “**pROC**” (Robin et al. 2011).

5.2. CAMPAÑA SIMULADA DE SV

De los resultados obtenidos en la campaña de SV anterior puede estimarse que los *decoys* generados mediante el *Enhanced Directory of Useful Decoys* (DUD-E) (Mysinger et al. 2012) son más adecuados para la validación de los modelos. Por ello, para la validación de los nuevos modelos solo construimos una nueva base DUD-E y no recurrimos a la “base simulada”.

5.3. COMBINACIÓN DE MODELOS

Los mejores modelos individuales fueron seleccionados y combinados utilizando los operadores *MAX*, *MIN*, *PROM* y *RANK*.

Para cada uno de estos operadores se realizaron combinaciones sistemáticas de hasta 100 modelos. Se comenzó con la combinación de los dos modelos que presentaron las mejores AUROC en la base DUD-E y luego se fue agregando de a un modelo por vez (siempre agregando el mejor entre los modelos restantes) hasta llegar a la combinación de los 100 mejores modelos.

Se realizaron curvas de AUROC *versus* cantidad de modelos combinados para determinar cuáles de las combinaciones son las más convenientes. La Figura 5.2 ilustra de la metodología de combinación utilizada.

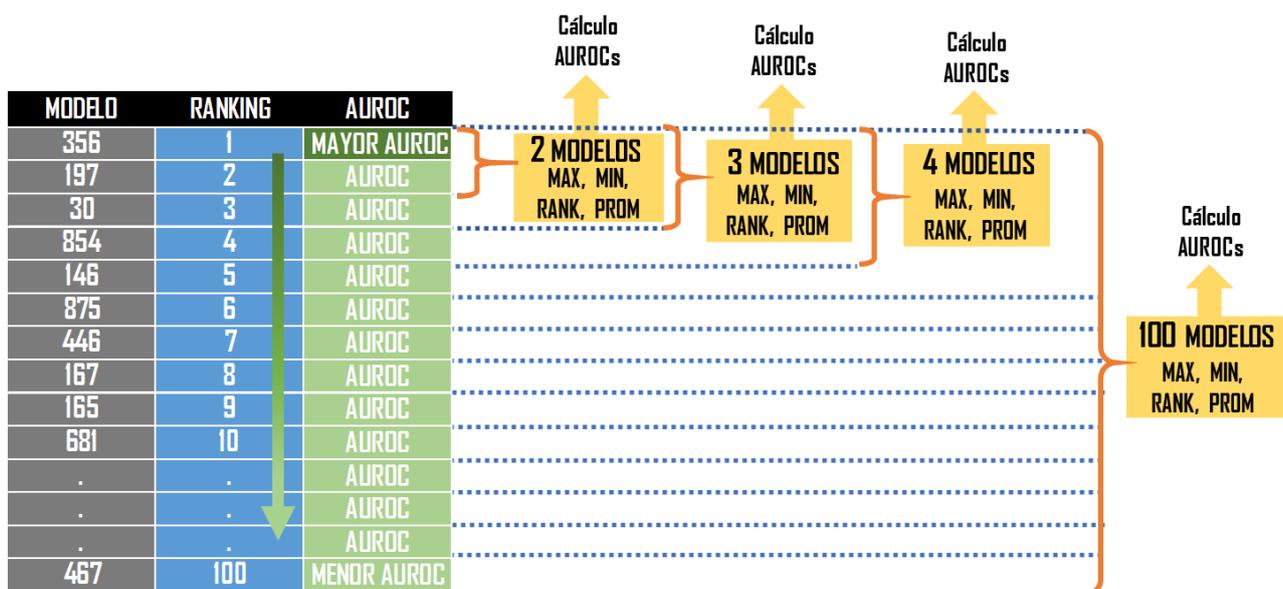


Figura 5.2: Estrategia de ensamblaje de modelos. Los modelos individuales se ordenan según el AUROC obtenido en la base DUD-E y se combinan sistemáticamente utilizando siempre los mejor rankeados.

5.4. SELECCIÓN DE VALOR DE CORTE PARA EL CRIBADO

Una preocupación importante en la realización de campañas de cribado virtual es la predicción de la probabilidad de que un candidato predicho como activo resulte realmente activo cuando se lo evalúe experimentalmente. Esta probabilidad es denominada *Positive Predictive Value*, o simplemente *PPV*. La estimación de tal probabilidad es, sin embargo, imposible de calcular para las bibliotecas sometidas a SV prospectivo debido a que en su cálculo aparece el número real de compuestos activos de la biblioteca (*Ya*, también llamado **proporción de activos o prevalencia**

de actividad o rendimiento de activos), el cual no es conocido *a priori*. El *PPV* se define por la Ecuación 5.1:

$$PPV = \frac{(Se)(Ya)}{(Se)(Ya) + (1 - Sp)(1 - Ya)}$$

Ecuación 5.1: Definición del *Positive Predictive Value (PPV)*

Donde *Se* y *Sp* representan la sensibilidad y especificidad, respectivamente, asociadas a un dado valor de corte. Esta ecuación se utilizó para construir **gráficas de superficie de *PPV*** que permitieron estimar el valor de corte del score asociado a la *PPV* deseada.

Tales gráficas de superficie 3D muestran la interacción entre el *PPV*, la relación *Se/Sp* y la prevalencia (*Ya*) para los mejores ensamblajes de modelos. Para construir las, primero se determinaron la *Se* y *Sp* en todo el rango de valores de corte del score posibles para la combinación de modelos seleccionada. A su vez, teniendo en cuenta que en las bases de cribado el valor de *Ya* es desconocido, pero invariablemente bajo, se varió su valor (teórico) entre 0,001 y 0,010; esto significa que el porcentaje de activos de la base se varió entre 0,1% y 1%. Finalmente, se calculó el *PPV* para todos los valores de corte y todos los valores de *Ya*. Se usó el paquete de R "**plotly**" (<https://CRAN.R-project.org/package=plotly>) para la realización de las gráficas de *PPV*.

Mediante el análisis visual de las superficies de *PPV* resultantes se seleccionó un valor de corte adecuado para emplear en el *SV*.

5.5. CRIBADO BASADO EN EL LIGANDO

El mejor ensamblaje de modelos se aplicó al *SV* de las bases de datos **DrugBank 4.0** (Law et al. 2014) y **Sweetlead**. La versión 4.0 de DrugBank no solo ha aumentado en cuanto al número de drogas que contiene con respecto a la versión anterior, sino que también incluye datos sobre la absorción, la distribución, el metabolismo, la excreción y la toxicidad (ADMET) de las drogas.

Ambas bases de datos fueron curadas antes de realizar el *SV* utilizando el software Standardizer v. 16.10.10.0 (ChemAxon, 2016). Aplicamos las siguientes acciones para lograr representaciones homogéneas de las estructuras moleculares de ambas bases:

- a) **Strip Salts**: Esta acción elimina los contraiones cuando el compuesto de la base aparece en forma de sal.
- b) **Remove Solvents**: Esta acción elimina moléculas de solvente incluidas en la representación (por ejemplo, aguas de hidratación).
- c) **Clear Stereo**: Esta acción elimina características estereoquímicas de molécula. Este paso se realiza teniendo en cuenta que los descriptores moleculares incluidos en nuestros modelos son "ciegos" a la conformación.
- d) **Remove Absolute Stereo**: Esta acción elimina las etiquetas de quiralidad de las moléculas.
- e) **Aromatize**: Esta acción lleva a cabo la aromatización de acuerdo con el método seleccionado. En nuestro caso utilizamos el método de Daylight.
- f) **Neutralize**: Neutraliza las moléculas cargadas agregando o eliminando hidrógenos para crear un compuesto neutro.
- g) **Add Explicit Hydrogens**: Agrega los H explícitos a las moléculas.

h) **Clean 2D**: Crea el modelo 2D de la molécula.

5.6. CRIBADO BASADO EN EL BLANCO MOLECULAR

Como hemos destacado anteriormente, la combinación de métodos basados en el ligando y en el blanco molecular se puede utilizar eficientemente para la realización de campañas de SV ya que ambas técnicas son complementarias.

En esta segunda campaña de SV recurrimos al **docking molecular** para seleccionar entre los hits que resultaron del cribado basado en el ligando, cuáles podrían interactuar mejor con los aminoácidos del sitio activo del transportador de putrescina de *T. cruzi* (TcPAT12). Como se describió en el Capítulo 2 de esta tesis, el **docking molecular** intenta predecir mediante la computadora la estructura del complejo ligando-receptor. Para simular este complejo y sus interacciones es necesario disponer de la estructura tridimensional de la proteína y de los ligandos a dockear.

Recurrimos a un modelo tridimensional de la permeasa de putrescina generado previamente en nuestro laboratorio (Dietrich et al. 2018) ya que la estructura tridimensional experimental de TcPAT12 aún no se encuentra disponible. Los ligandos a dockear fueron aquellos que obtuvieron valores de PPV mayores a 30% (para $Y\alpha = 0,010$) luego de ser sometidos al SV basado en los ligandos.

Se realizaron varios **modelos del transportador** basándose en la estructura del cotransportador (del tipo *antiporter*) de arginina/agmatina de *Escherichia coli* (Código de acceso en el Protein Data Bank: 3L1L) (Berman et al. 2000). Esta estructura molde fue identificada a través de la búsqueda de la similitud de secuencias BLAST (Madden 2013). Los modelos tridimensionales de TcPAT12 fueron desarrollados utilizando el software Modeller (Eswar et al. 2007) y el mejor modelo fue seleccionado en función del parámetro *Discrete Optimized Protein Energy* normalizado (z-DOPE) (Shen & Sali 2006) y del *score* GA341 (el cual analiza la confiabilidad de un modelo basado en potenciales estadísticos) (Melo & Sali 2007).

Para realizar el **docking** se utilizó el software Autodock4.2 (modo rígido) ya que en los estudios previos se presentó como la mejor alternativa (entre las disponibles) para discriminar entre los inhibidores conocidos de los no inhibidores de TcPAT12 (Alberca et al. 2018). El sitio activo del **docking** fue configurado de acuerdo con los estudios realizados por Soysa et al (Soysa et al. 2013). Los cálculos con Autodock2.4 se realizaron en una grilla de dimensiones (44 Å) × (58 Å) × (40 Å) en las coordenadas x, y, z, respectivamente. Adicionalmente, se realizaron estimaciones estándar para todas las variables, como las cargas parciales de Marsilli-Gasteiger. Finalmente, se realizaron 100 corridas de **docking** para cada compuesto seleccionado, utilizando el blanco molecular rígido y los ligandos flexibles (permitiendo la rotación de todos los ángulos de torsión de los candidatos).

Para determinar un valor de corte del score del **docking** a partir del cual los compuestos del cribado se considerarán buenos candidatos a inhibidores de TcPAT12 se realizaron corridas de **docking** con el sustrato natural del transportador, la putrescina.

5.7. SELECCIÓN FINAL DE LOS CANDIDATOS

Finalmente, se analizó si el pool de compuestos que superaron todas las etapas *in silico* y que se presentan como mejores candidatos a inhibir TcPAT12 satisfacían diferentes reglas de *druglikeness*, como la regla de Lipinski, la regla de Veber y los criterios de *druglikeness* adoptados por Monge et al. (Lipinski et al. 2012; Veber et al. 2002; Monge et al. 2006). Las dos primeras consideran características generales que debe reunir un fármaco para poseer buena biodisponibilidad oral, asumiendo absorción por transporte pasivo. Las reglas de Monge suman, además, algunos requisitos vinculados a la potencial toxicidad de los compuestos analizados (por ejemplo, ausencia de funciones químicas altamente reactivas).

5.8. ENSAYOS EXPERIMENTALES

5.8.1. ACTIVIDAD BIOLÓGICA SOBRE *T. cruzi*

Inicialmente se evaluó el efecto de las drogas seleccionadas sobre los diferentes estadios del ciclo de vida de *T. cruzi*: **epimastigotes, tripomastigotes sanguíneos y amastigotes**.

Los ensayos realizados sobre epimastigotes y tripomastigotes fueron desarrollados en las mismas condiciones que los descritos en el Capítulo 3 por lo que no se detallará el procedimiento aquí.

Para la **determinación del efecto de las drogas sobre amastigotes de *T. cruzi*** se emplearon células J774 infectadas con tripomastigotes sanguíneos de la cepa RA. Las mismas fueron cultivadas en una estufa a 37 °C en una atmósfera húmeda de CO₂ al 5% en placas de 24 pocillos (1,5 x 10⁵ células por pocillo; volumen final de 500 µL y por duplicado). Luego de 24 hs de incubación se agregaron diluciones frescas de las drogas a evaluar o benznidazol (control positivo) para llegar a concentraciones entre 1 y 50 µM dependiendo de la droga. Luego de 72 hs se eliminó el medio y las células fueron fijadas por 10 min con metanol frío y teñidas por 15 min con 10% v/v de una solución de Giemsa. Se registró el número de amastigotes por cada 100 células hospedadoras. Los cultivos control fueron incubados únicamente con medio o con concentraciones de DMSO iguales a las de las drogas. La citotoxicidad fue analizada en suspensiones celulares tripsinizadas, después de la adición de ioduro de propidio (Sigma, St. Louis, USA) (5ug/ml) 10 minutos antes del análisis de fluorescencia y microscopía óptica del número de células viables y muertas (Miranda et al. 2015).

5.8.2. EVALUACIÓN DEL TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS

El efecto de las drogas sobre la incorporación de putrescina en epimastigotes de *T. cruzi* se evaluó mediante el mismo método que el descrito en el Capítulo 3, por lo tanto, no será detallado en este capítulo.

En este caso, los resultados se expresaron mediante la comparación de la velocidad inicial de ingreso de putrescina por los cultivos con droga y sin droga.

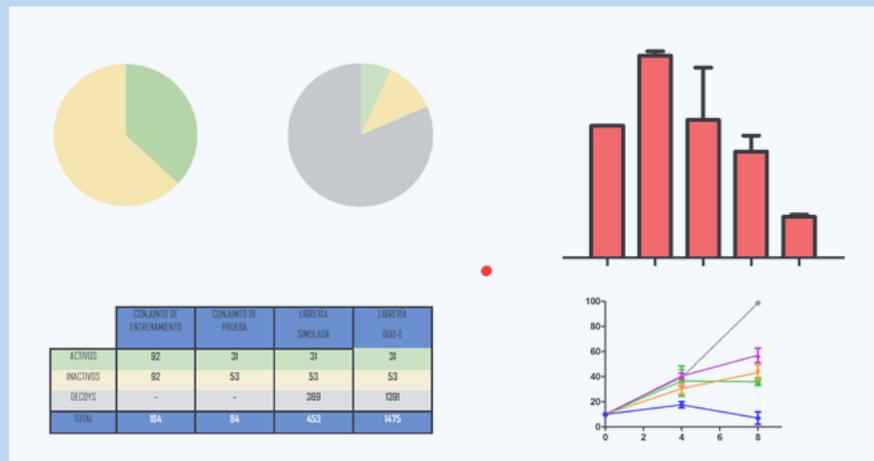
REFERENCIAS

- Alberca, L.N. et al., 2018. Cascade ligand- and structure-based virtual screening to identify new trypanocidal compounds inhibiting putrescine uptake. *Frontiers in Cellular and Infection Microbiology*, 8(MAY).
- Berman, H.M. et al., 2000. The Protein Data Bank. *Nucleic acids research*, 28(1), pp.235–242.
- Cruz-Montegudo, M. et al., 2014. Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde?. *Drug Discovery Today*, 19(8), pp.1069–1080.
- Dietrich, R.C. et al., 2018. Identification of cisapride as new inhibitor of putrescine uptake in *Trypanosoma cruzi* by combined ligand- and structure-based virtual screening. *European Journal of Medicinal Chemistry*, 149, pp.22–29.
- Eswar, N. et al., 2007. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, Chapter 2, p.Unit 2 9.
- Kuhn, M. 2007. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), pp.1-26.
- Law, V. et al., 2014. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, 42(D1).
- Lipinski, C.A. et al., 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 64(SUPPL.), pp.4–17.
- Madden, T., 2013. The BLAST sequence analysis tool. *The BLAST Sequence Analysis Tool*, pp.1–17.
- Melo, F. & Sali, A., 2007. Fold assessment for comparative protein structure modeling. *Protein Sci*, 16(11), pp.2412–2426.
- Miranda, C.G. et al., 2015. A flow cytometer-based method to simultaneously assess activity and selectivity of compounds against the intracellular forms of *Trypanosoma cruzi*. *Acta Tropica*, 152, pp. 8-16.
- Monge, A. et al., 2006. Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. *Molecular Diversity*, 10(3), pp.389–403.
- Mysinger, M.M. et al., 2012. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14), pp.6582–6594.
- R Foundation for Statistical Computing, 2016. R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*.
- Robin, X. et al., 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12.
- Shen, M. & Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11), pp.2507–2524.
- Soysa, R. et al., 2013. Structural model of a putrescine-cadaverine permease from *Trypanosoma cruzi* predicts residues vital for transport and ligand binding. *The Biochemical journal*, 452(3), pp.423–32.
- Veber, D.F. et al., 2002. Molecular properties that influence the oral bioavailability of drug

candidates. *Journal of Medicinal Chemistry*, 45(12), pp.2615–2623.

CAPÍTULO 6

SEGUNDA CAMPAÑA DE CRIBADO VIRTUAL Y VALIDACIÓN EXPERIMENTAL. RESULTADOS



	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	LIBRERÍA SIMULADA	LIBRERÍA DUS-E
ACTIVOS	92	31	31	31
INACTIVOS	92	53	53	53
DECAYS	-	-	269	1391
TOTAL	184	84	453	1475



6.1. COMPILACIÓN Y PARTICIÓN DE LA BASE DE DATOS

Luego del curado, la base de datos quedó conformada por 256 compuestos y sus correspondientes datos de actividad, de los cuales 116 fueron considerados como **ACTIVOS** y 140 como **INACTIVOS**. Se construyeron tres mapas de calor con Gitoools v.2.3.1 (Perez-Llamas & Lopez-Bigas 2011) donde puede apreciarse la disimilitud molecular (utilizando *fingerprints* ECFP4 y el coeficiente de similitud de Tanimoto en su forma binaria) entre los compuestos de una misma categoría y entre los compuestos de diferentes categorías presentes en la base de datos (Figura 6.1). En la Figura 6.1.a puede apreciarse, asimismo, que existen compuestos activos e inactivos altamente similares; esto implica, como se dijo anteriormente, un mayor desafío a la hora del modelado.

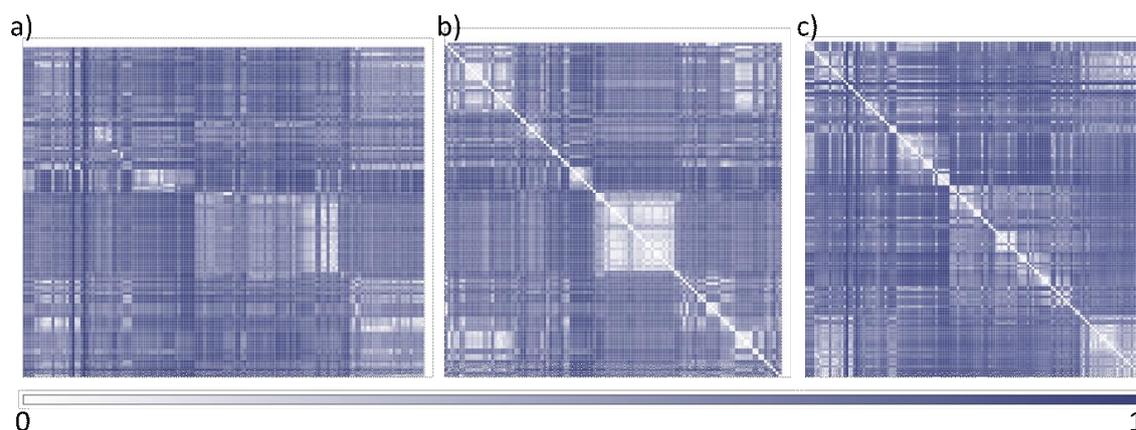


Figura 6.1: Mapas de calor de disimilitud de todo el conjunto de datos y dentro de cada categoría. a) **ACTIVOS versus INACTIVOS**; b) **ACTIVOS versus ACTIVOS**; c) **INACTIVOS versus INACTIVOS**

Al igual que en el proceso de partición realizado en los capítulos anteriores, se tomó 75% de cada *cluster* de los compuestos **ACTIVOS** (la categoría minoritaria) y se asignó la misma cantidad de compuestos **INACTIVOS** al **conjunto de entrenamiento**. El resto de las estructuras se asignaron al **conjunto de prueba** (Tabla 6.1).

Tabla 6.1: Composición de las bases de datos utilizadas para la generación y validación de los modelos desarrollados en la segunda campaña de SV

	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	BIBLIOTECA DUD-E
ACTIVOS	87	29	29
INACTIVOS	87	53	53
DECOYS	-	-	1302
TOTAL	174	82	1355

La generación de la biblioteca DUD-E se realizó a partir de los 29 compuestos **ACTIVOS** del conjunto de prueba, obteniéndose un total de 1302 señuelos que fueron adicionados al conjunto de prueba para la realización de la campaña piloto de SV.

6.2. GENERACIÓN Y VALIDACIÓN DE MODELOS BASADOS EN EL LIGANDO

A partir del conjunto de entrenamiento se realizaron 1000 modelos lineales clasificatorios en el entorno R, los cuales poseían como máximo 12 descriptores moleculares. Estos modelos fueron evaluados mediante validación interna; se evaluó, además, su capacidad predictiva mediante la determinación del %BC y el AUROC en el conjunto de prueba y la biblioteca DUD-E.

Con respecto a los AUROCs de los 1000 modelos individuales generados 82%, 57% y 25% obtuvieron AUROC mayor a 0,8 en el conjunto de entrenamiento, de prueba y la biblioteca DUD-E, respectivamente. Esto sugiere cierto grado de sobreajuste y corrobora que la campaña piloto de SV contra la biblioteca DUD-E es la tarea más desafiante para los modelos. Por lo tanto, se consideraron como mejores modelos a aquellos que poseían mayor AUROC en la biblioteca DUD-E.

El mejor modelo individual fue el modelo 348:

Modelo 348:

$$FD = 3.12777 + 0.03474*F06[C-C] + 0.20805*S-107 - 0.04291*F05[N-N] + 0.39611*C-039 - 0.34582*SM5_B(s) + 6.25705*Eta_epsilon_A + 0.53013*nSO2OH - 1.28338*SpMax_H2 + 0.44827*Eig04_AEA(ri) - 1.73390*ATSC1e + 0.05975*CATS2D_09_PL + 0.02805*SaaO$$

La nomenclatura de los descriptores es la utilizada por el software Dragon 6.0 (Milano Chemometrics, 2011). En el **Material Suplementario 7** se encuentra una breve explicación de los descriptores que utiliza este modelo.

En la **Tabla 6.2** se resumen los resultados obtenidos en la validación interna de los 8 modelos que presentaron mejor AUROC en la base DUD-E. Los porcentajes de buenas clasificaciones (%BC) fueron calculados utilizando como valor de corte "0,5" para diferenciar entre activos e inactivos. Con respecto al LGO, %BC indica el promedio de los resultados encontrados en cada subdivisión; en cuanto al ensayo de aleatorización de Fisher, %BC es el resultado promedio de las 10 aleatorizaciones. Del análisis de los resultados de la validación LGO puede observarse que el %BC es bastante estable (similar al del conjunto de entrenamiento), y que el rendimiento de los modelos generados al azar es significativamente menor al de los modelos originales.

Tabla 6.2: Resultados de la validación interna para los 6 mejores modelos.

	MODELO ORIGINAL	VALIDACIÓN CRUZADA (LGO)	VALIDACIÓN DE FISHER	
MODELO	% BC	% BC	% BC	INTERVALO DE CONFIANZA (0.95)
M 348	78,74	74,46	58,33	54,53 - 62,14
M 706	72,41	70,71	57,36	53,83 - 60,88
M 981	75,86	73,45	60,63	57,20 - 64,06
M 557	69,54	68,80	57,24	54,15 - 60,33
M 123	78,74	75,72	60,46	57,67 - 63,25
M 693	77,59	74,80	61,15	58,90 - 63,39
M 560	65,52	64,02	58,68	55,11 - 62,24
M 746	74,71	69,70	59,60	55,07 - 64,12

En la **Tabla 6.3** se observan los resultados de la validación externa para los 8 mejores modelos. Para calcular %BC se estableció en este caso como valor de corte 0,5. Como puede observarse, la capacidad explicativa de estos 8 modelos no es excelente ya que se obtienen %BC entre 65 y 78% en el conjunto de entrenamiento; sin embargo, la baja capacidad explicativa es compensada con una buena capacidad predictiva ya que cuatro de estos modelos aumentan el %BC cuando se evalúan en el conjunto de prueba. Finalmente, los resultados obtenidos en la biblioteca DUD-E demuestran el buen desempeño de estos modelos cuando se analizan bases de datos conteniendo una gran cantidad de compuestos inactivos. En otras palabras, los mejores modelos individuales generados en esta campaña, a costa de haber perdido poder explicativo, alcanzaron mayor capacidad predictiva.

Tabla 6.3: Resultados obtenidos para los 8 mejores modelos.

MODELO	CONJUNTO DE ENTRENAMIENTO		CONJUNTO DE PRUEBA		BIBLIOTECA DUD-E	
	%BC	AUROC	%BC	AUROC	%BC	AUROC
M348	78,74	0,882 ($\pm 0,0250$)	75,61	0,885 ($\pm 0,0360$)	88,08	0,934 ($\pm 0,0123$)
M706	72,41	0,809* ($\pm 0,0319$)	67,07	0,736** ($\pm 0,0565$)	91,76	0,924 ($\pm 0,0205$)
M981	75,86	0,843 ($\pm 0,0298$)	79,27	0,837 ($\pm 0,0467$)	87,36	0,922 ($\pm 0,0254$)
M557	69,54	0,778** ($\pm 0,0343$)	74,39	0,818 ($\pm 0,0482$)	90,25	0,919 ($\pm 0,0203$)
M123	78,74	0,850 ($\pm 0,0285$)	82,93	0,882 ($\pm 0,0382$)	84,39	0,918 ($\pm 0,0185$)
M693	77,59	0,860 ($\pm 0,0280$)	70,73	0,828* ($\pm 0,0459$)	86,20	0,913* ($\pm 0,0171$)
M560	65,52	0,775** ($\pm 0,0348$)	68,29	0,779* ($\pm 0,0525$)	89,02	0,911* ($\pm 0,0185$)
M746	74,71	0,844 ($\pm 0,0288$)	70,73	0,820 ($\pm 0,0456$)	84,97	0,910 ($\pm 0,0195$)

AUROC estadísticamente diferente del mejor modelo individual (M348) en el mismo conjunto/biblioteca de datos * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Teniendo como experiencia la anterior campaña de SV, recurrimos nuevamente a las estrategias de combinación de modelos para obtener mejores predicciones.

Mediante las gráficas de AUROC versus cantidad de modelos se determinó la cantidad de modelos que obtiene el mayor AUROC para cada esquema de combinación (**Figura 6.2**). El

ensamble por operador MIN de 8 modelos (en adelante, **MIN-8**) y los ensambles PROM, RANK y VOT de 15 modelos presentaron mayor AUROC en sus respectivos esquemas. Mediante la comparación estadística de estos 4 mejores ensambles se determinó que no hay diferencias significativas entre ellos y, si se tiene en cuenta el principio de parsimonia (“*En igualdad de condiciones, la explicación más sencilla suele ser la más probable*”), el ensamble MIN-8 presenta una leve ventaja sobre los demás. Además, puede advertirse que las combinaciones por encima de cierta cantidad de modelos presentan un peor comportamiento en términos del valor del AUROC y de la desviación estándar de la media, que aumenta con el número de modelos incluidos en el ensamblado (Figura 6.2).

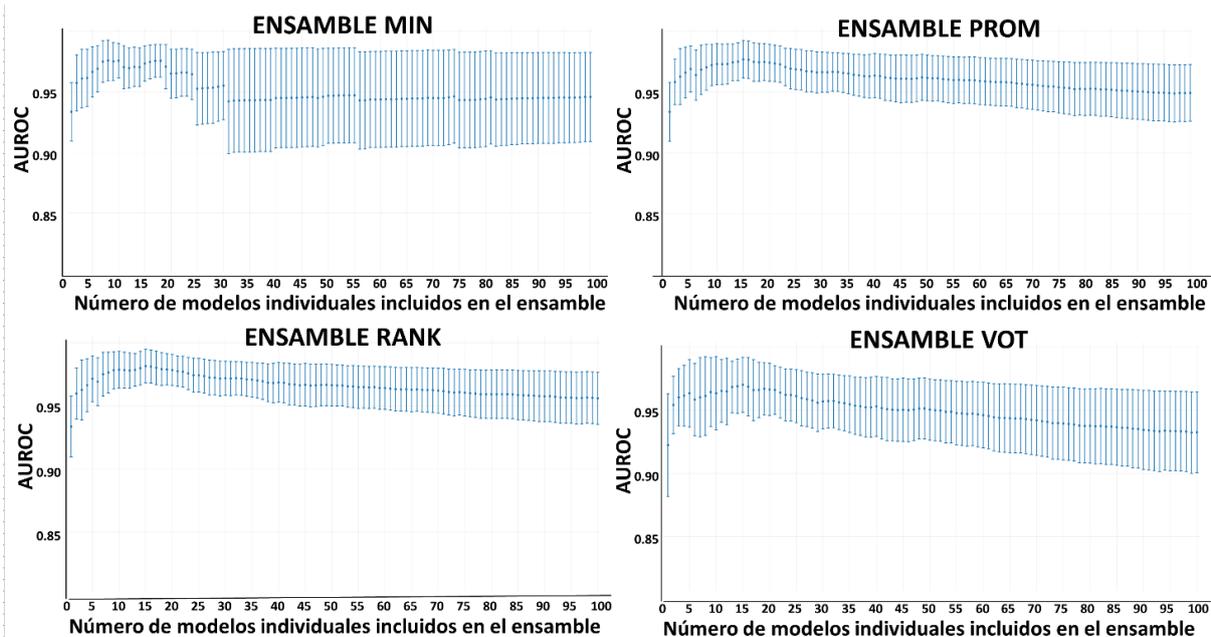


Figura 6.2: Gráficas AUROC vs cantidad de modelos involucrados en el ensamble. a) ensamble MIN; b) ensamble PROM; c) ensamble RANK; d) ensamble VOT.

En la **Tabla 6.4** se presentan los AUROCs que obtuvieron los ensambles de los 8 mejores modelos en los tres conjuntos de datos y se comparan con el mejor modelo individual (M348). Independientemente del esquema de combinación, vemos que todos se comportaron estadísticamente similar al mejor modelo individual cuando predicen los compuestos del conjunto de entrenamiento y de prueba; sin embargo, todos los ensambles superaron estadísticamente al mejor modelo individual cuando se aplicaron sobre la biblioteca DUD-E ($p < 0,001$ en tres de los casos y $< 0,05$ en el restante).

Tabla 6.4: AUROCs obtenidos para los ensambles de 8 modelos y el mejor modelo individual en la biblioteca DUD-E.

	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	BASE DUD-E
MIN-8	0.851 (± 0.0281)	0.885 (± 0.0367)	0.976 ^{***} (± 0.0085)
RANK-8	0.886 (± 0.0239)	0.878 (± 0.0375)	0.976 ^{***} (± 0.0082)
PROM-8	0.891 (± 0.0234)	0.887 (± 0.0357)	0.970 ^{***} (± 0.0096)
VOT-8	0.833 (± 0.0283)	0.810 (± 0.0516)	0.959 [*] (± 0.0173)
M348	0.882 (± 0.0250)	0.885 (± 0.0360)	0.934 (± 0.0123)

AUROC estadísticamente diferente al del mejor modelo individual (M348) en el mismo conjunto de datos

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Considerando estos resultados, optamos por pasar a la campaña de SV real con el esquema de combinación MIN-8.

6.3. SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO

Como se describió en el capítulo anterior, hemos refinado el criterio de selección del valor de corte recurriendo a gráficas de superficie de **PPV**. Con la ayuda de estas superficies, la evolución del valor de **PPV** puede optimizarse visualmente (o, eventualmente, matemáticamente) como una función de la relación **Se/Sp** en un rango de valores **Ya**. Para este análisis, hemos utilizado la relación entre **Se/Sp** obtenida para el ensamble MIN-8 para todos los valores de corte del *score* posibles en la campaña piloto de SV. Adicionalmente, se varió el **Ya** desde 0,001 hasta 0,01 con variaciones sucesivas de 0,001. En la **Figura 6.3** se representa la gráfica de superficie obtenida en estas condiciones.

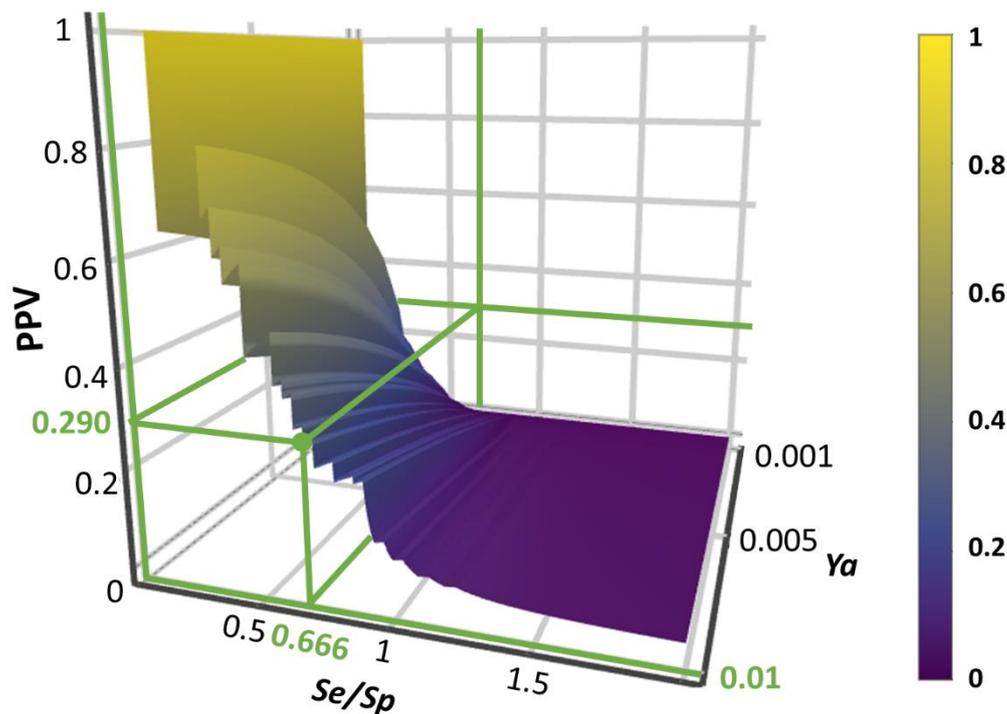


Figura 6.3: Superficie de PPV para el ensamble MIN-8.

Al analizar la superficie PPV, se observa que en el desplazamiento hacia relaciones **Se/Sp** bajas (es decir, alta **Sp**) tiene un impacto positivo en el **PPV**, lo que resulta en mayores probabilidades de confirmar hits *in silico* sometidos a ensayos *in vitro* (**Figura 6.3**). A partir de este gráfico de superficie **PPV**, elegimos **0,354** como **valor de corte del score** para ser utilizado en la segunda campaña de SV; tal puntaje está asociado a una relación **Se/Sp** igual a **0,666** (**Se =98,30%; Sp = 65,52%**).

En la **Figura 6.4** es posible realizar una comparación visual entre las superficies de PPV (teóricas) del ensamble de 6 modelos que obtuvimos en la primera campaña de cribado virtual y el ensamble de 8 modelos obtenido en esta segunda campaña. Está claro que el ensamble de 8 modelos muestra un comportamiento más robusto y predecible en el rango de **Se/Sp** que va desde 0,3 a 1.

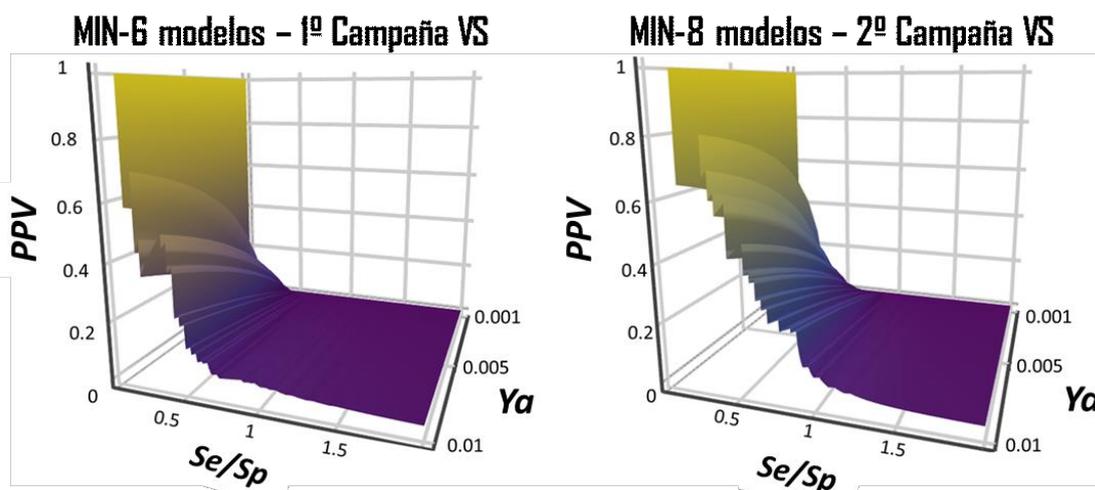


Figura 6.3: Comparación de las superficies de PPV obtenidas para los ensambles MIN-6 de la primera campaña de screening virtual y MIN-8 de la segunda campaña de screening virtual.

6.4. CRIBADO VIRTUAL

Se aplicó el ensamble MIN-8 en el SV de las bases de datos DrugBank 4.0 y Sweetlead. Utilizando el valor de corte del *score* optimizado (0,354) sólo 24 compuestos fueron seleccionados como ACTIVOS, de los cuales 15 eran fármacos aprobados (los candidatos más promisorios para el reposicionamiento) y el resto eran moléculas cursando estudios clínicos.

6.5. CRIBADO BASADO EN EL BLANCO MOLECULAR

Los 24 hits provenientes del cribado basado en el ligando fueron sometidos al cribado basado en el blanco molecular mediante la estrategia del *docking* molecular en el modelo de homología TcPAT12.

17 de los hits cribados mediante el *docking* obtuvieron un *score* inferior al valor encontrado para el ligando natural putrescina (-6.0), que se calculó previamente en las mismas condiciones de dockeo. 8 de ellos son fármacos aprobados.

La **Tabla 6.5** muestra los hits seleccionados a través de la estrategia combinada de métodos basados en el ligando y en el blanco molecular, incluyendo los valores de *PPV* en el rango de *Ya* 0,001 a 0,010, para los correspondientes *scores* del ensamble MIN-8 así como el *score* obtenido del *docking* y la indicación terapéutica original.

Tabla 6.5: Hits seleccionados a través de la estrategia combinada de SV basada en el ligando y en el receptor

DROGA	SCORE MIN-8	PPV (Ya = 0,001)	PPV (Ya = 0,010)	SCORE DEL DOCKING	INDICACIÓN TERAPÉUTICA ORIGINAL
CLOMIFENO	0,484	14,64	63,38	-9,37	Tratamiento para la infertilidad femenina, principalmente debido a la anovulación
BUTOCONAZOL	0,477	14,64	63,38	-9,87	Tratamiento local de candidiasis vulvovaginal
MECLIZINA	0,455	12,30	58,61	-8,95, -6,64^a	Cinetosis y vértigo
CLEMIZOL	0,454	12,30	58,61	-8,91	Alergias
CINARIZINA	0,427	9,20	50,56	-7,43	Cinetosis y vértigo
ORMELOXIFENO	0,380	5,67	37,76	-6,31	Principalmente usado como anticonceptivo
OXICONAZOL	0,375	5,00	34,68	-10,44	Infecciones fúngicas dérmicas
ASTEMIZOL	0,367	4,25	30,96	-6,45	Alergias
CLOFAZIMINA	0,361	4,06	29,92	-1,58	Tratamiento de lepra

^a El score del fue calculado para ambos isómeros ópticos.

Como se puede notar, en la **Tabla 6.5** se ha incluido a la Clofazimina, con un puntaje de *docking* de -1,58, que indicaría una menor afinidad por en el sitio donde se realizó el *docking* con respecto al ligando natural. En este caso particular, el fármaco ha sido anteriormente identificado como agente tripanocida dentro de nuestro grupo de investigación, durante un *SV in silico* para detectar nuevos inhibidores de cruzipaína. Posteriormente, la droga confirmó su potencial tanto en modelos de roedores agudos como crónicos de la Enfermedad de Chagas (Bellera et al. 2015; Sbaraglini et al. 2016). En esa ocasión, sin embargo, se observó que la potencia del fármaco contra el parásito era mayor que la potencia inhibitoria contra la cruzipaína, lo que sugirió la posibilidad de que existieran otros mecanismos de acción además de la inhibición de la cruzipaína.

Antes de la adquisición de los hits, se verificó si dichos compuestos obedecían las diferentes reglas de *druglikeness*. Se encontró que todos los compuestos satisfacían 3 de las 4 reglas de Lipinski (Lipinski & Hopkins 2004) (violando sólo la vinculada a su capacidad de partición,

expresada como valor clogP), la regla de Veber (Veber et al. 2002) y los criterios de *druglikeness* adoptados por Monge et al (Monge et al. 2006).

Finalmente, en base a la accesibilidad de los hits (costo y posibilidad de compra), adquirimos **Cinarizina** (Sigma Aldrich), **Dihidrocloruro de Meclizina** (en adelante, Meclizina) (Sigma Aldrich), **Clofazimina** (Sigma Aldrich) y **Nitrato de Butoconazol** (en adelante, Butoconazol) (Sigma Aldrich) para la realización de los ensayos *in vitro*. En la **Figura 6.4** se pueden visualizar las estructuras moleculares 2D de los cuatro fármacos adquiridos.

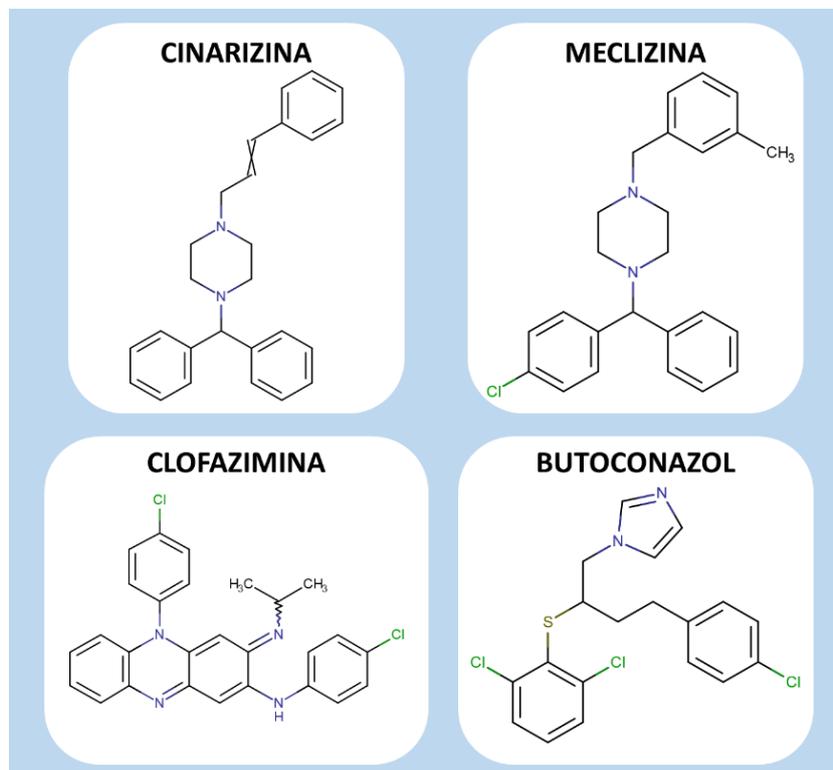


Figura 6.4: Estructura de los compuestos seleccionados para su evaluación *in vitro*.

6.6. EVALUACIÓN BIOLÓGICA

6.6.1. DETERMINACIÓN DE LA ACTIVIDAD ANTI-*T. cruzi*

Se evaluó el efecto de Cinarizina, Meclizina y Butoconazol sobre la viabilidad de diferentes formas del parásito. La Clofazimina no fue evaluada en estos ensayos ya que, como se dijo, su actividad ha sido reportada previamente (Bellera et al. 2015).

Se determinó el efecto de diferentes concentraciones de las tres drogas contra **epimastigotes de *T. cruzi*** (Figura 6.5) y se calcularon los EC_{50} correspondientes. Los tres fármacos mostraron una clara inhibición concentración-dependiente sobre la proliferación de epimastigotes, con un EC_{50} (en el día 4) de **6,05 μM** , **8,39 μM** y **3,08 μM** , para Cinarizina, Meclizina y Butoconazol, respectivamente. El fármaco de referencia Benznidazol obtuvo un EC_{50} de 2,56 μM contra epimastigotes.

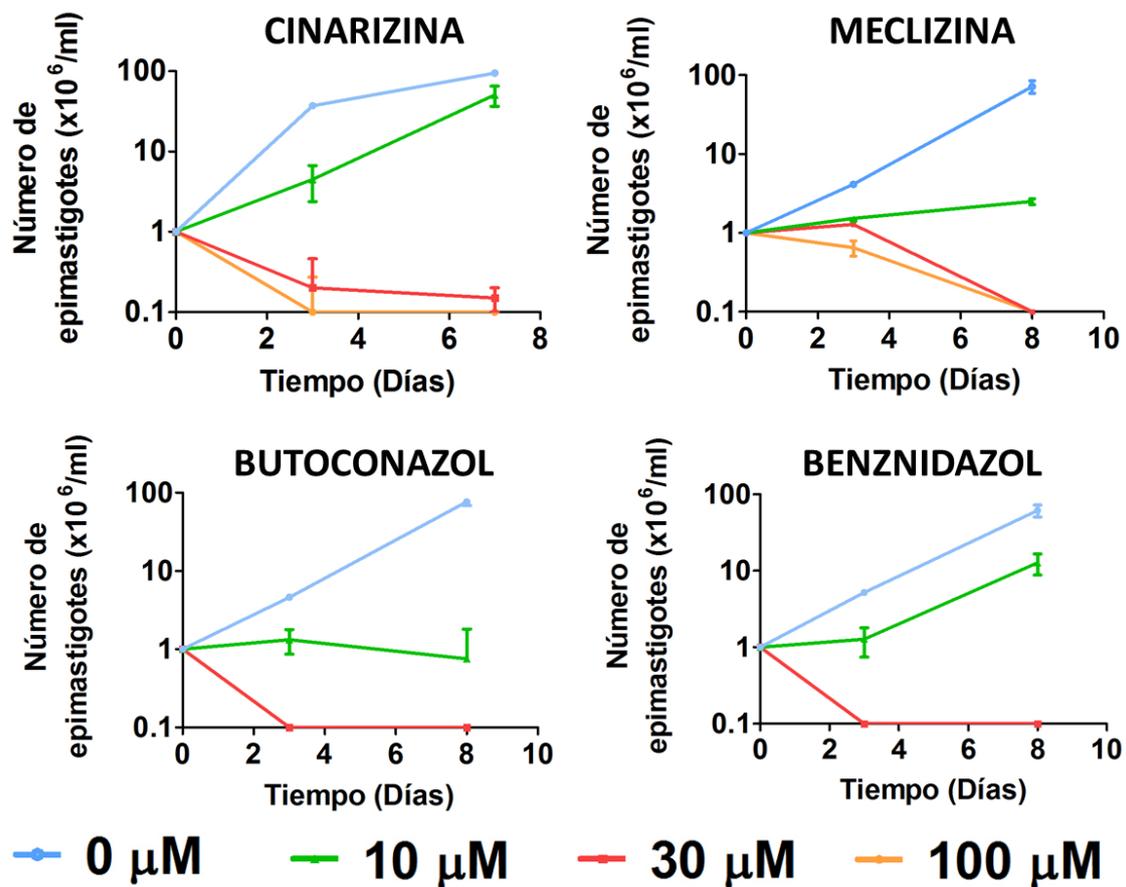


Figura 6.5: Efecto de las drogas sobre la proliferación de epimastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado.

Las tres drogas se evaluaron a una concentración de 20 μ M frente a **tripomastigotes de *T. cruzi***. Cinarizina demostró una inhibición parcial en la viabilidad a esa concentración (30% de inhibición), mientras que las otras dos drogas no presentaron inhibición en la condición ensayada.

Al evaluar sus efectos contra **amastigotes**, Butoconazol mostró una citotoxicidad considerable sobre las células hospedadoras J774 en todas las concentraciones ensayadas (100,0%, 57,5%, 48,3%, a 50 μ M; 20 μ M y 5 μ M respectivamente), por lo que se descartó para futuras investigaciones. Meclizina mostró citotoxicidad contra las células J774 a 50 μ M (13,90%), aunque no se observaron efectos tóxicos a concentraciones más bajas. Cinarizina no mostró toxicidad en ninguna de las concentraciones utilizadas.

Como se puede observar en la **Figura 6.6**, Meclizina redujo notable y significativamente la cantidad de amastigotes cada 100 células en las tres concentraciones analizadas mientras que Cinarizina mostró un efecto menor contra los amastigotes, obteniendo un efecto significativo, cercano al 50% de inhibición a 50 μ M.

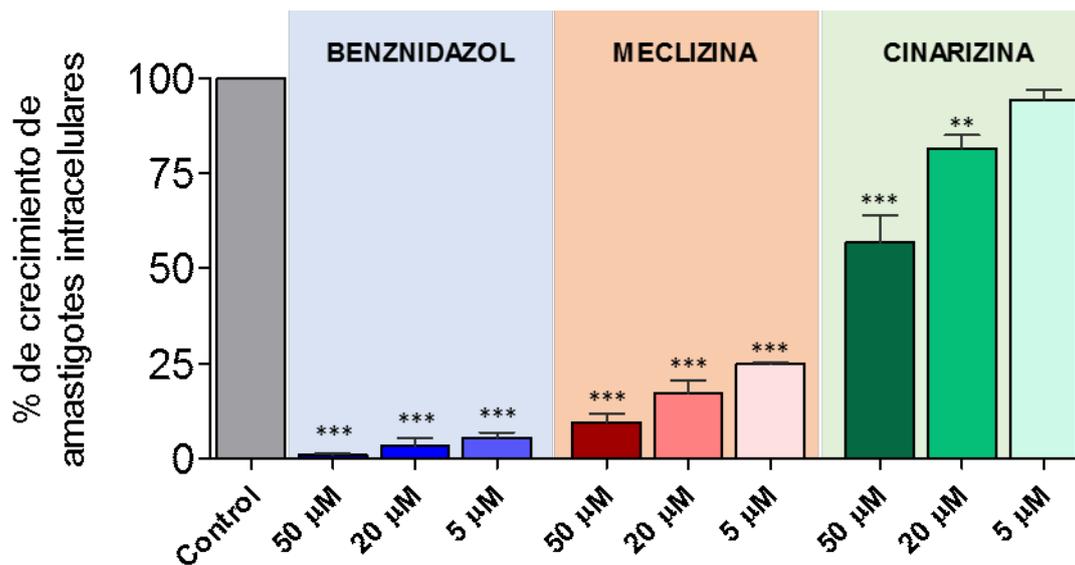


Figura 6.6: Efecto inhibitorio de las drogas evaluadas sobre amastigotes de *T. cruzi*. Los resultados son expresados como la media \pm SD de experimentos por triplicado. El análisis estadístico se realizó mediante una prueba de ANOVA de una vía seguida de una prueba de comparación múltiple post hoc de Dunnett (** $p < 0,001$; *** $p < 0,0001$).

6.6.2. ENSAYOS DE TRANSPORTE DE PUTRESCINA Y OTROS METABOLITOS

Para determinar si los candidatos realmente inhiben el transporte de putrescina como predicen nuestros modelos, se evaluó el efecto de Cinarizina, Clofazimina, Meclizina y Butoconazol sobre la incorporación de putrescina en epimastigotes de *T. cruzi*.

Cinarizina y Clofazimina mostraron un efecto claro en la captación de putrescina con una reducción significativa de la velocidad inicial a $52,56 \pm 4,84\%$ y $30,85 \pm 2,74\%$ respectivamente, en comparación con el transporte en condiciones de control (Figura 6.7a). Por el contrario, Meclizina y Butoconazol no mostraron efecto inhibitorio sobre la captación de putrescina, lo que indicaría que su efecto tripanocida se relaciona con otros mecanismos de acción.

Se comprobó la especificidad de la inhibición de la captación de putrescina por Cinarizina y Clofazimina mediante la evaluación del efecto de ambos hits en la captación de Arginina, un transportador de la misma familia de transportadores (Figura 6.7b). Como puede observarse, la inhibición del transporte de putrescina sería selectiva para ambos compuestos. Adicionalmente, se comprobó por el ensayo de MTT que la viabilidad de los parásitos con 50 μM de las drogas se mantenía intacta al tiempo máximo de realización de los ensayos de incorporación de Putrescina y Arginina, descartando que los resultados obtenidos se deban a la citotoxicidad de las drogas.

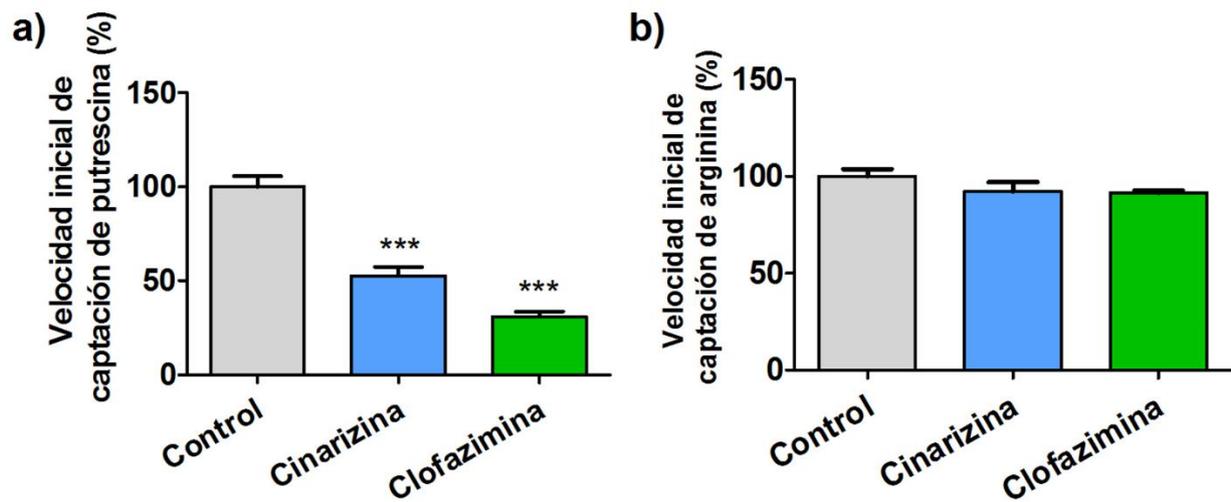


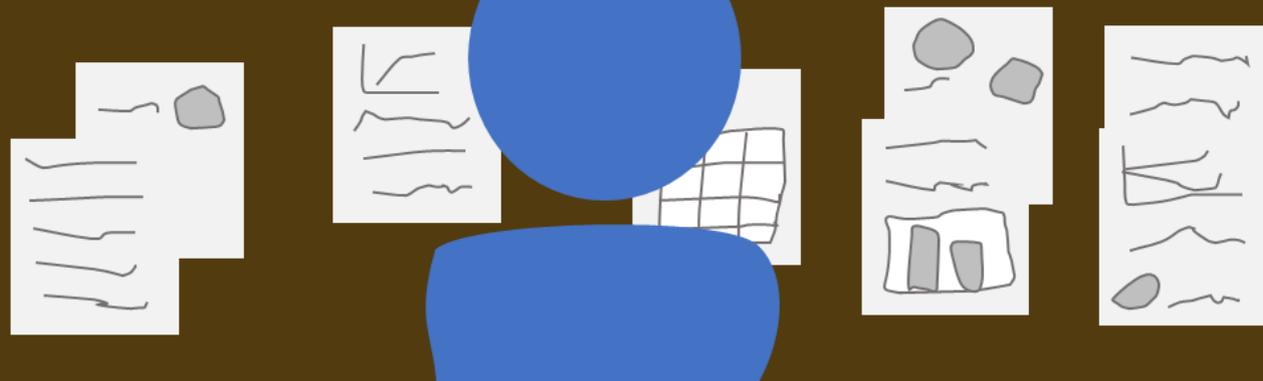
Figura 6.7: Efecto de Cinarizina (50 μ M) y Clofazimina (50 μ M) sobre a) captación de Putrescina; b) captación de Arginina. Los valores se expresan como % de la media \pm SD en comparación con el control. El análisis estadístico se realizó mediante una prueba de ANOVA de una vía seguida de una prueba de comparación múltiple post hoc de Dunnett (***) $p < 0,0001$).

REFERENCIAS

- Bellera, C.L. et al., 2015. Computer-guided drug repurposing: Identification of trypanocidal activity of clofazimine, benidipine and saquinavir. *European Journal of Medicinal Chemistry*, 93, pp.338–348.
- Lipinski, C. & Hopkins, A., 2004. Navigating chemical space for biology and medicine. *Nature*, 432(7019), pp.855–861.
- Monge, A. et al., 2006. Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. *Molecular Diversity*, 10(3), pp.389–403.
- Perez-Llamas, C. & Lopez-Bigas, N., 2011. Gitoools: Analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE*, 6(5).
- Sbaraglini, M.L. et al., 2016. Novel cruzipain inhibitors for the chemotherapy of chronic Chagas disease. *International Journal of Antimicrobial Agents*, 48(1), pp.91–95.
- Veber, D.F. et al., 2002. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 45(12), pp.2615–2623.

CAPÍTULO 7

CONCLUSIONES



7.1. CONCLUSIONES GENERALES

Durante el presente trabajo de tesis se realizaron dos campañas de cribado virtual con el objetivo de encontrar compuestos que interfirieran con la captación y/o biosíntesis de poliaminas en *T. cruzi*.

Para la realización de la **primera campaña de cribado virtual**, compilamos una base de datos de análogos de poliaminas previamente evaluados sobre *T. cruzi* y la utilizamos para generar 100 modelos clasificatorios capaces de discriminar entre análogos de poliaminas con y sin actividad inhibitoria sobre el parásito. Todos los modelos fueron validados y algunos de ellos presentaron una buena capacidad predictiva. La combinación de los mejores modelos resultó en una mejora considerable con respecto a la capacidad de predicción de los modelos individuales en la campaña piloto (retrospectiva) de cribado virtual. El cribado virtual realizado sobre las bases de datos DrugBank 3.0 y Sweetlead con la mejor combinación de modelos reveló 45 fármacos aprobados que podrían poseer actividad tripanocida; 9 de ellos ya habían sido evaluados previamente contra diferentes formas de *T. cruzi* (8 de estos obteniendo EC₅₀ menores a 15 μM).

5 de los candidatos que no habían sido ensayados previamente contra el parásito fueron ensayados en epimastigotes. Se observó que Triclabendazol inhibe potentemente la proliferación de los epimastigotes (EC₅₀ = 4,8 μM), mientras que Paroxetina y Sertaconazol presentan efecto tripanocida a concentraciones más altas.

En cuanto a la actividad tripanocida de amplio espectro, Paroxetina demostró ser el agente más potente contra tripomastigotes sanguíneos de *T. brucei* y contra promastigotes de *L. infantum* con EC₅₀ menores a 10 μM. La diferente potencia inhibitoria de las drogas evaluadas se puede atribuir a las diferencias metabólicas de las especies estudiadas; también es posible que el mecanismo de acción difiera entre los tripanosomátidos, dado que los tripanosomátidos cumplen con su requerimiento metabólico de poliaminas usando diferentes estrategias. Mientras que *T. cruzi* es auxótrofo para las poliaminas, *Leishmania spp.* puede tanto sintetizarlas a partir de precursores como incorporarlas desde el medio extracelular a través de transportadores y *T. brucei* se basa exclusivamente en la síntesis *de novo* (Heby et al. 2007). Para los epimastigotes de *T. cruzi*, pudimos demostrar que la incorporación de putrescina se vio afectada por la exposición a cada uno de los tres fármacos evaluados. Se demostró también que estas tres drogas inhiben el transporte de otros compuestos tales como los aminoácidos cargados positivamente (Lisina y Arginina) o el nucleósido Uridina, lo que hablaría de un efecto inhibitorio sobre diversos transportadores de la misma familia (Lisina y Arginina). Esta inhibición de más de un blanco molecular podría explicar la relativamente rápida acción de estas drogas sobre los epimastigotes de *T. cruzi*. De entre los hits activos, se observa una mayor selectividad hacia TcPAT12 por parte de Paroxetina. Dado que el tripanosoma podría subsistir incluso semanas en ausencia de Putrescina (Carrillo et al. 1999), de inhibirse únicamente el transporte de Putrescina se esperaría observar un efecto tripanocida más lento. Adicionalmente, mediante ensayos *in vitro* se descartó que la actividad encontrada en *T. brucei* se deba a la inhibición de la enzima tripanotión sintetasa.

La realización de la **segunda campaña de cribado virtual** involucró una metodología combinada de métodos basados en el ligando y en el receptor. Se generaron y validaron 1000 modelos lineales clasificatorios capaces de reconocer análogos de poliaminas con actividad tripanocida. Se analizaron diferentes esquemas de combinación de modelos para seleccionar la combinación que presentaba mejor capacidad predictiva en la campaña retrospectiva de cribado virtual. Mediante el análisis de gráficas de superficie del PPV optimizamos el valor de corte del *score* de

manera de disminuir considerablemente la probabilidad de encontrarnos con falsos positivos. El mejor ensamble se aplicó al *screening virtual* de las bases de datos DrugBank 4.0 y Sweetlead dando como resultados la identificación de 24 compuestos químicos (15 de estos fármacos aprobados) con posible actividad tripanocida. Los 24 candidatos fueron sometidos al *docking molecular* utilizando un modelo por homología de transportador de Putrescina TcPAT12. 17 de estos compuestos presentaron buenos resultados en el *docking*.

Se seleccionaron 4 compuestos para su evaluación *in vitro*, Cinarizina, Meclizina, Butoconazol y Clofazimina (este último sólo se evaluó en el ensayo de incorporación de Putrescina, ya que poseía reportes previos de actividad tripanocida). Los 3 fármacos evaluados frente a la proliferación de epimastigotes de *T. cruzi* demostraron una buena capacidad inhibitoria, todos ellos con EC_{50} menores a 10 μ M. Cinarizina también reveló cierta capacidad inhibitoria frente a tripomastigotes y amastigotes de *T. cruzi*, mientras que Meclizina inhibió potentemente los amastigotes, pero no presentó actividad frente a tripomastigotes. Se determinó el efecto de los candidatos sobre la incorporación de Putrescina en epimastigotes de *T. cruzi*. Sólo Clofazimina y Cinarizina evidenciaron el efecto esperado.

Cabe señalar que la probabilidad de encontrar un compuesto activo a partir de más de 13.000 compuestos que fueron sometidos al cribado virtual es muy baja. Por lo tanto, las dos campañas de tamizado virtual han dado resultados satisfactorios ya que de 8 compuestos evaluados frente a epimastigotes de *T. cruzi*, 5 presentaron actividad tripanocida; de los 6 compuestos evaluados frente a tripomastigotes de *T. cruzi*, 2 de ellos disminuyeron la viabilidad de los parásitos y; de los 3 compuestos evaluados frente a amastigotes, 2 presentaron actividad y el otro se descartó debido a su elevada citotoxicidad. Con respecto al transporte de Putrescina, de las 7 drogas evaluadas, 5 de ellas presentaron efecto inhibitorio frente a la incorporación de esta poliamina.

Estos resultados demuestran la utilidad de las metodologías computacionales (particularmente los SV de las bibliotecas de fármacos aprobados y experimentales) en la búsqueda de nuevas aplicaciones terapéuticas de fármacos conocidos. El reposicionamiento de medicamentos asistido por computadora es una estrategia moderna, racional y económica para proporcionar nuevas soluciones terapéuticas para enfermedades desatendidas.

7.2. COMPARACIÓN CON ESTUDIOS PREVIOS

Debido a que la explotación de los transportadores de poliaminas como blanco terapéutico para la realización de campañas de reposicionamiento de fármacos guiadas por computadora está en etapas tempranas, la comparación de los resultados obtenidos durante la realización de esta tesis doctoral con trabajos previos se ve obstaculizada. Además del hecho de que existen muy pocos estudios anteriores, se han aplicado en ellos diferentes enfoques de cribado virtual.

Reigada y colaboradores han utilizado una estrategia combinada de búsqueda por similitud molecular a ligandos conocidos seguida por el *docking molecular* de los candidatos para la identificación de nuevos inhibidores del transporte de Putrescina (Reigada et al. 2017). De esta forma identificaron y confirmaron experimentalmente dos nuevos inhibidores del transporte de Putrescina, la Isotreonina y Acitretina. La evaluación de la Isotreonina frente a diferentes formas del parásito demostró que la EC_{50} contra tripomastigotes (0,130 μ M) es cerca de 230 veces menor que contra epimastigotes.

En esta tesis se observó lo contrario para Triclabendazol, Meclizina, Butoconazol y Cinarizina ya que los epimastigotes fueron más sensibles a estas drogas que los tripomastigotes. Una explicación posible para este grado de variación en la inhibición al transporte de Putrescina en las diferentes formas de *T. cruzi*, es que las distintas formas de los parásitos podrían utilizar diferentes mecanismos de transporte primario de poliaminas (cada uno de ellos con diferentes especificidades por las drogas) (Seguel et al. 2016). Una posibilidad similar ha sido sugerida por Basselin y colaboradores para *Leishmania*, donde ha sido observado que los amastigotes y promastigotes usan diferentes transportadores para la captación de Putrescina (Basselin et al. 2000; Muller et al. 2001).

En otro trabajo, los compuestos resultantes del ensamble de modelos que utilizamos en nuestra primera campaña de cribado virtual fueron utilizados posteriormente por Dietrich y colaboradores para la realización de un tamizado basado en la estructura, con uno de sus tres hits validado experimentalmente (Dietrich et al. 2018).

La comparación estadística del PPV real (número de hits confirmados *in vitro* con respecto al número de hits evaluados) en las campañas de cribado virtual con el predicho en base al análisis de las superficies de PPV descritas no tiene -estadísticamente hablando- sentido debido a la pequeña cantidad de hits evaluados en cada ocasión. Sin embargo, tal PPV fue en todos los casos muy superior a la mediana de la tasa de aciertos que se obtiene en los estudios de cribado virtual (13%) informado en un análisis retrospectivo crítico de literatura científica relacionada con resultados de cribado virtual publicados entre 2007 y 2011 (Zhu et al. 2013). Adicionalmente, cabe destacar que en campañas de cribado virtual es más valioso encontrar unos pocos hits que presenten nuevos andamiajes (*scaffolds*) a encontrar varios hits que tengan andamiajes de compuestos activos conocidos (Zhu et al. 2013). En este sentido, los inhibidores del transporte de putrescina que fueron encontrados en esta tesis proveen nuevos andamiajes que pueden ser utilizados para el posterior desarrollo de inhibidores más potentes.

7.3. ENSAMBLADO DE MODELOS Y CAMPAÑAS RETROSPECTIVAS DE SV

Las bibliotecas DUD-E utilizadas como una herramienta de validación en nuestras campañas de cribado virtual retrospectivas confirmaron que el cribado retrospectivo es la prueba más desafiante (además, por supuesto, del *screening* prospectivo) para un modelo concebido para aplicaciones de *screening in silico*. Se pudo observar como en el AUROC de la mayoría de los clasificadores individuales tiende a caer abruptamente cuando se progresa desde el conjunto de entrenamiento a la Biblioteca DUD-E. Por ejemplo, mientras que un 82% (820 de los 1000 modelos generados en la segunda campaña) de los modelos individuales de la segunda campaña de SV obtuvieron AUROCs por encima de 0,8 en el conjunto de entrenamiento, el porcentaje de modelos individuales que obtuvieron tal logro en la Biblioteca DUD-E fue de solo el 25%. Además, como ya ha sido mencionado, la aproximación de ensamble de modelos condujo a resultados más robustos, mejorando la capacidad de generalización con respecto a los modelos individuales (en nuestro caso, esto se refleja en un comportamiento mejorado en el conjunto de prueba y las bibliotecas simulada y DUD-E).

7.4. POSIBLE APLICACIÓN DE LOS CANDIDATOS AL REPOSICIONAMIENTO

Una proporción importante de los hits evaluados experimentalmente nos permitieron validar los modelos ya que inhibieron el transporte de putrescina. Sin embargo, algunos de los hits obtuvieron potente efecto tripanocida a pesar de no inhibir el transporte, por lo que también, *a priori*, se los debería considerar para el posible desarrollo de nuevos tratamientos para la Enfermedad de Chagas. No obstante, deben considerarse criterios adicionales para establecer si los hits seleccionados son buenos candidatos para el reposicionamiento.

Triclabendazol, Sertaconazol, Butoconazol, Paroxetina, Cinarizina y Meclizina pertenecen a clases terapéuticas que han demostrado potencial como terapias antichagásicas en el pasado: drogas antihelmínticas (Mishina et al. 2007; Santos et al. 2015), drogas antifúngicas (Lepesheva et al. 2011), drogas antidepresivas (Doyle & Weinbach 1989; Bazán et al. 2016), bloqueantes de canales de calcio (Engel et al. 2010; Benaim & Paniz Mondolfi 2012; Planer et al. 2014) y drogas antihistamínicas (Engel et al. 2010; Planer et al. 2014; Lara-Ramirez et al. 2017). Butoconazol y Sertaconazol, sin embargo, no son candidatos directos para el reposicionamiento ya que son usados tópicamente como antifúngicos. Por ello, la mayoría de las ventajas del reposicionamiento de drogas se perderían si la segunda indicación (en este caso, quimioterapia para la Enfermedad de Chagas) requiere administración sistémica (Oprea & Overington 2015). Además, nuestra evidencia experimental de actividad citotóxica de Butoconazol y la relativamente alta EC_{50} obtenida sobre epimastigotes de *T. cruzi* para Sertaconazol desalienta proseguir el estudio de estas drogas.

Las ventajas del reposicionamiento de medicamentos también se perderán mayormente si la dosis requerida para la segunda indicación está por encima del rango de dosis utilizado en la clínica para la indicación original (Oprea & Overington 2015). En el caso de Cinarizina, la cual es administrada para su indicación original en dosis diarias bajas y mostró solo una actividad débil contra los amastigotes de *T. cruzi*, este problema se haría más evidente. El escenario se complica aún más por el hecho de que la exposición del fármaco libre se verá disminuida por la unión de la droga a proteínas plasmáticas, que en este caso es de alrededor del 90%. Con respecto a la Meclizina, a pesar de haber demostrado un efecto tripanocida potente contra amastigotes de *T. cruzi* inhibiendo 75% del crecimiento con solo 5 μM de droga, la posibilidad de reposicionamiento es baja ya que este fármaco se administra para su indicación original en una simple dosis de 25 mg con lo que se obtiene una concentración plasmática máxima de droga de 0,2 μM (Wang et al 2012).

También se debe considerar la posibilidad de que ambas drogas, Cinarizina y Meclizina, puedan ser útiles como puntos de partida para programas *hit-to-lead* pero, en cualquier caso, se arribaría a compuestos novedosos que no podrían catalogarse como reposicionamiento, perdiéndose las ventajas asociadas (por ejemplo, evitar parte de los estudios preclínicos y clínicos).

En el caso de la Paroxetina, se debe tener en cuenta que es un fármaco antidepresivo que incrementa los niveles extracelulares del neurotransmisor serotonina por lo podría generarse un desbalance en este neurotransmisor. Además, la concentración plasmática de la droga en el estado estacionario es inferior a 0,19 μM (muy por debajo de del EC_{50} obtenido) y la unión a proteínas plasmáticas se estima que es del 95%, por lo que solo 5% de la droga en plasma

tenderá a un equilibrio de concentración con la droga libre en el compartimento extravascular, siendo justamente la droga libre la que interactuará con el correspondiente blanco molecular (Kaye et al 1989). Por otro lado, ha sido reportado que esta droga revierte la disfunción cardíaca en modelos animales con infartos de miocardio, y esto podría representar un efecto adicional de interés en pacientes con Chagas (Schumacher et al. 2015). Sin embargo, los resultados obtenidos por esta droga en el modelo de infección aguda indican que no es un buen candidato para el reposicionamiento.

Por último, el Triclabendazol, podría ser un buen candidato para el reposicionamiento ya que para su indicación original (tratamiento de la fascioliasis) es administrado oralmente a altas dosis (entre 10 y 20 mg/kg/día), obtuvo una potencia moderada frente a epimastigotes y tripomastigotes y, además, disminuyó la parasitemia en el modelo de infección aguda. La realización de estudios posteriores con este fármaco podría confirmar si es un buen candidato para el reposicionamiento. Entre los estudios previstos, pueden mencionarse el estudio de los efectos de esta droga sobre otras cepas de *T. cruzi* (tanto en modelos celulares como en modelo murino agudo) y, de obtener resultados satisfactorios, su evaluación en modelo murino de infección crónica.

Como conclusión final de esta tesis doctoral se debe destacar la importancia de las estrategias *in silico* para la búsqueda de nuevos tratamientos para enfermedades desatendidas, como la Enfermedad de Chagas, ya que hemos demostrado que con una inversión de tiempo y dinero relativamente pequeña es posible identificar drogas con actividad tripanocida.

Debido a que nos hemos enfocado en bases de datos de fármacos aprobados, las drogas tripanocidas encontradas se encuentran en el mercado para su utilización en otras indicaciones terapéuticas. Esto es particularmente importante porque, como ya hemos mencionado, desarrollar un medicamento *de novo* es excesivamente costoso y la inversión de las compañías farmacéuticas en este tipo de enfermedades ha sido, históricamente, muy pequeña.

La implementación de estrategias eficientes desde el punto de vista traslacional (como la estrategia de reposicionamiento de fármacos) es fundamental para que los avances en el campo de la investigación se trasladen a los potenciales beneficiarios de los mismos; en el caso de esta tesis doctoral, que los compuestos con actividad tripanocida se transformen en un medicamento innovador disponible para los pacientes con Chagas.

DETALLE DE PUBLICACIONES, CONGRESOS Y BECAS

Durante la realización de este trabajo de tesis se han realizado 15 presentaciones a congresos y 5 publicaciones en revistas científicas con referato.

En el año 2016 se ha obtenido una beca de viaje, estadía e inscripción de parte de The World Academy of Sciences for the Developing Countries (TWAS) para asistir a Biovision 2016 – The World Life Sciences Forum en Lyon, Francia; en 2017 el Instituto de Salud Carlos III (España) ha otorgado una beca de viaje, estadía e inscripción para participar en el 6th World Congress on Leishmaniasis en Toledo, España; ese mismo año, el Comité Organizador del XXIV Congreso Latinoamericano de Parasitología otorgó una beca de inscripción al XXIV Congreso Latinoamericano de Parasitología; y en 2018 la Society for Laboratory Automation and Screening (SLAS) otorgó una beca de viaje, estadía e inscripción para participar a SLAS2018 International Congress.

En 2016 hemos recibido el premio “BIOVISION Next Award – In recognition of its contribution towards life sciences” luego de la presentación de nuestro trabajo en Biovision 2016; en 2017 obtuvimos el 3er premio a Trabajos Libres en el XXIV Congreso Latinoamericano de Parasitología; y en 2018 el premio Tony B. Academic Travel Award otorgado por SLAS.

A continuación, se detallan las publicaciones realizadas en el transcurso de la tesis doctoral:

- 1) Autores: Alberca L; Sbaraglini ML; Balcazar D; Fraccaroli L; Carrillo C; Medeiros A; Benitez D; Comini M; Talevi A. Título: Discovery of Novel Polyamine Analogs with anti-Protozoal Activity by Computer Guided Drug Repositioning. Revista: Journal of Computer-Aided Molecular Design. Vol. n° 30, Tomo 4, p 305 – 321. Año de publicación: 2016
- 2) Autores: Gantner M; Alberca L; Mercader A; Bruno-Blanch L; Talevi A. Título: Integrated Application of Enhanced Replacement Method and Ensemble Learning for the Prediction of BCRP/ABCG2 Substrates. Revista: Current Bioinformatics. Volumen: 12, n° 3, paginas 239-248(10) Año: 2017
- 3) Autores: Dietrich R; Alberca L; Ruiz D; Palestro P; Carrillo C; Talevi A; Gavernet L. Título: Identification of cisapride as new inhibitor of putrescine uptake in *Trypanosoma cruzi* by combined ligand- and structure-based virtual screening. Revista: European Journal of Medicinal Chemistry. Volumen: 149, paginas 22-29 Año: 2018
- 4) Autores: Alberca L; Sbaraglini ML; Morales JF; Dietrich RC; Ruiz D; Pino Martinez AM; Miranda CG; Fraccaroli L; Alba Soto C; Carrillo C; Palestro P; Talevi A. Título: Cascade ligand- and structure-based virtual screening to identify new trypanocidal compounds inhibiting putrescine uptake. Revista: Frontiers in Cellular and Infection Microbiology. Volumen 8, articulo 173. Año: 2018
- 5) Autores: Morales JF; Alberca LN; Di ianni M; Chuguransky S; Talevi A; Ruiz ME. Título: Molecular topology and other promiscuity determinants as predictors of therapeutic Class? A theoretical framework to guide drug repositioning? Revista: Current Topics in Medicinal Chemistry. Volumen 18, páginas 1110-1122. Año: 2018

REFERENCIAS

- Basselin, M. et al., 2000. Putrescine and spermidine transport in *Leishmania*. *Molecular and Biochemical Parasitology*, 109, pp.37–46.
- Bazán, P.C. et al., 2016. Quantitative PCR and unconventional serological methods to evaluate clomipramine treatment effectiveness in experimental *Trypanosoma cruzi* infection. *Experimental and Molecular Pathology*, 101(2), pp.274–280.
- Benaim, G. & Paniz Mondolfi, A.E., et al 2012. The emerging role of amiodarone and dronedarone in Chagas disease. *Nature Reviews. Cardiology*, 9(10), pp.605–609.
- Carrillo, C. et al., 1999. *Trypanosoma cruzi* epimastigotes lack ornithine decarboxylase but can express a foreign gene encoding this enzyme. *FEBS Letters*. 454(3), pp.192–196.
- Dietrich, R.C. et al., 2018. Identification of cisapride as new inhibitor of putrescine uptake in *Trypanosoma cruzi* by combined ligand- and structure-based virtual screening. *European Journal of Medicinal Chemistry*.149, pp.22–29.
- Doyle P.S. & Weinbach E.C., 1989. The activity of tricyclic antidepressant drugs against *Trypanosoma cruzi*. *Experimental Parasitology*, 68(2), pp.230–234.
- Engel J. et al., 2010. Image-based high-throughput drug screening targeting the intracellular stage of *Trypanosoma cruzi*, the agent of Chagas' disease. *Antimicrobial Agents and Chemotherapy*, 54(8), pp.3326–3334.
- Heby, O. et al., 2007. Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of African sleeping sickness, Chagas' disease, and leishmaniasis. *Amino Acids*, 33(2), pp.359–366.
- Kaye, C.M. et al., 1989. A review of the metabolism and pharmacokinetics of paroxetine in man. *Acta psychiatrica Scandinavica. Supplementum*, 350, pp.60–75.
- Lara-Ramirez, E.E. et al., 2017. An *in vitro* and *in vivo* evaluation of new potential sialidase inhibitors of *Trypanosoma cruzi* predicted by a computational drug repositioning method. *European Journal of Medicinal Chemistry*, 132, pp.249–261.
- Lepesheva, G.I et al., 2011. Targeting *Trypanosoma cruzi* Sterol 14 α -Demethylase (CYP51). *Advances in Parasitology*, 75, pp.65–87.
- Mishina, Y.V. et al. 2007. Artemisinins inhibit *Trypanosoma cruzi* and *Trypanosoma brucei rhodesiense* *in vitro* growth. *Antimicrobial Agents and Chemotherapy*, 51(5), pp. 1852-1854.
- Muller, S. et al., 2001. Targeting polyamines of parasitic protozoa in chemotherapy. *Trends Parasitology*. 17, pp.242–249.
- Oprea, T.I. & Overington, J.P., 2015. Computational and practical aspects of drug repositioning. *Assay and Drug Development Technologies*. 13(6), pp.299–306.
- Planer, J.D. et al., 2014. Synergy testing of FDA-approved drugs identifies potent drug combinations against *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases*, 8(7): e2977
- Reigada, C. et al., 2017. Trypanocidal effect of isotretinoin through the inhibition of polyamine and amino acid transporters in *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases*, 11(3): e0005472.
- Santos, E.C. et al., 2015. Concomitant Benznidazole and Suramin Chemotherapy in Mice

Infected with a Virulent Strain of *Trypanosoma cruzi*. *Antimicrobial Agents and Chemotherapy*, 59(10), pp.5999-6006.

Seguel, V. et al., 2016. Pentamidine antagonizes the benznidazole's effect *in vitro*, and lacks of synergy *in vivo*: Implications about the polyamine transport as an anti-*Trypanosoma cruzi* target. *Experimental Parasitology*, 171, pp.23-32.

Schumacher, S.M. et al., 2015. Paroxetine-mediated GRK2 inhibition reverses cardiac dysfunction and remodeling after myocardial infarction. *Science Translational Medicine*, 7(277): 277ra31.

Wang, Z. et al., 2012. Meclizine metabolism and pharmacokinetics: formulation on its absorption. *Journal of Clinical Pharmacology*, 52(9), pp.1343-1349.

Zhu, T. et al., 2013. Hit identification and optimization in virtual screening: practical recommendations based upon a critical literature analysis. *Journal of Medicinal Chemistry*, 56, pp.6560-6572.

MATERIAL SUPPLEMENTARIO



Material Suplementario 1: Breve descripción de las 17 ETDs de la OMS.

ENFERMEDAD	BREVE DESCRIPCIÓN	EFEECTO CLÍNICO
DENGUE	Infección vírica transmitida por la picadura de mosquitos del género <i>Aedes</i> infectados con el virus del dengue.	Fiebre elevada con síntomas similares a una gripe. Puede avanzar a dengue grave el cual es potencialmente mortal.
RABIA	Enfermedad viral prevenible transmitida a humanos a través de las mordeduras de perros infectado el virus de la rabia.	Puede adoptar dos formas con diferente sintomatología (rabia furiosa y rabia paralizante). En casi todos los casos es mortal una vez que han aparecido los síntomas
TRACOMA	Enfermedad infecciosa causada por la bacteria <i>Chlamydia trachomatis</i> , que se transmite por contacto con secreciones oculares y nasales de personas infectadas.	Puede ocasionar la formación de opacidades irreversibles que causarán discapacidad visual y ceguera.
ULCERA DE BURULI	Infección causada por la bacteria <i>Mycobacterium ulcerans</i> cuyo modo de transmisión al ser humano es desconocido.	Afecta a la piel, y a veces al hueso, y puede causar desfiguraciones permanentes y discapacidad a largo plazo.
PIAN (YAWS)	Infección bacteriana crónica causada por la bacteria <i>Treponema pallidum</i> . La transmisión se produce por contacto directo entre la piel (principalmente pequeñas lesiones) de persona a persona.	Puede causar desfiguración y discapacidad permanentes.
LEPRA	Enfermedad infecciosa crónica causada por el bacilo <i>Mycobacterium leprae</i> . La lepra se transmite por gotículas nasales y orales cuando hay un contacto estrecho y frecuente con enfermos no tratados	Afecta principalmente a la piel, los nervios periféricos, la mucosa de las vías respiratorias superiores y los ojos. Si no se trata puede causar lesiones progresivas y permanentes en la piel, los nervios, las extremidades y los ojos.
ENFERMEDAD DE CHAGAS	Enfermedad infecciosa causada por el parásito protozoo <i>Trypanosoma cruzi</i> que puede transmitirse por diferentes vías.	Los enfermos crónicos pueden presentar alteraciones cardíacas, digestivas, neurológicas o combinadas.
ENFERMEDAD DEL SUEÑO	Enfermedad infecciosa causada por el protozoo <i>Trypanosoma brucei</i> y transmitido al humano principalmente por la picadura de la mosca <i>tsetse</i> .	Los primeros síntomas incluyen cambios de comportamiento, confusión, trastornos sensoriales y falta de coordinación. En una segunda etapa se producen trastornos del ciclo del sueño. Puede ser mortal.
LEISHMANIASIS	Enfermedad causada por protozoarios parásitos del género <i>Leishmania</i> que son transmitidos a través de las picaduras de flebótomos hembra infectados.	En su forma más grave (visceral) ataca los órganos internos y en su forma más frecuente (cutánea) causa úlceras en la cara, cicatrices desfigurantes y discapacidad.

TENIASIS Y CISTICERCOSIS	Infección intestinal provocada por tres especies de cestodos, siendo <i>Taenia solium</i> la más peligrosa. La infección se produce cuando las personas consumen carne de cerdo infectada cruda o poco cocinada.	Cuando los huevos de <i>T. solium</i> alcanzan el sistema nervioso central pueden causar síntomas neurológicos (neurocisticercosis), en particular epilepsia
DRACUNCULIASIS (ENFERMEDAD DEL GUSANO DE GUINEA)	Es una parasitosis invalidante causada por <i>Dracunculus medinensis</i> que se transmite cuando las personas beben agua contaminada con pulgas de agua infectadas.	Afecta las extremidades inferiores provocando dolor e incapacidad
EQUINOCOCOSIS	Enfermedad provocada por parásitos cestodos del género <i>Echinococcus</i> . La infección se produce por la ingestión de agua y/o alimentos contaminados.	Las dos formas más importantes de la enfermedad en el ser humano son la equinococosis quística (hidatidosis) y la equinococosis alveolar, esta última, si no se trata, puede ser mortal
TREMATODIASIS DE TRANSMISIÓN ALIMENTARIA	Son causadas por gusanos trematodos; entre ellos, <i>Clonorchis</i> , <i>Opisthorchis</i> , <i>Fasciola</i> y <i>Paragonimus</i> . La infección se produce por el consumo de peces, crustáceos o verduras crudos o poco cocidos que albergan las larvas diminutas del parásito	Los síntomas dependen del órgano afectado. Pueden generar cáncer, daños en el encéfalo y la muerte.
FILARIASIS LINFÁTICA (ELEFANTIASIS)	Es causada por la infección por nematodos de la familia Filarioidea, principalmente <i>Wuchereria bancrofti</i> . La infección humana se produce por la transmisión de los parásitos por diferentes tipos de mosquitos.	La filariasis afecta al sistema linfático y puede producir hipertrofia anormal de algunas partes del cuerpo, causando dolor, discapacidad grave y estigma social.
ONCOCERCOSIS (CEGUERA DE LOS RÍOS)	Es una enfermedad parasitaria provocada por el nematodo filárico <i>Onchocerca volvulus</i> que se transmite por la picadura de moscas negras infectadas (<i>Simulium spp.</i>).	Entre sus síntomas se encuentran prurito intenso, afecciones cutáneas desfigurantes y discapacidad visual, que puede llegar a la ceguera permanente.
ESQUISTOSOMIASIS	Enfermedad parasitaria causada por trematodos del género <i>Schistosoma</i> . Las personas se infectan cuando las formas larvarias del parásito, liberadas por caracoles de agua dulce, penetran en la piel durante el contacto con aguas infestadas.	Puede afectar a la capacidad de trabajo y en algunos casos puede ser mortal.
HELMINTIASIS TRANSMITIDAS POR EL SUELO	Son transmitidas por los huevos de los parásitos eliminados con las heces fecales de las personas infestadas. Las principales especies de helmintos transmitidos por el suelo que infectan al hombre son la ascáride (<i>Ascaris lumbricoides</i>), el tricocéfalo (<i>Trichuris trichiura</i>) y el anquilostoma (<i>Necator americanus</i> y <i>Ancylostoma duodenale</i>).	Las infecciones intensas pueden causar manifestaciones intestinales (diarrea y dolor abdominal), malnutrición, malestar general y debilidad, así como disminución del crecimiento y del desarrollo físico.

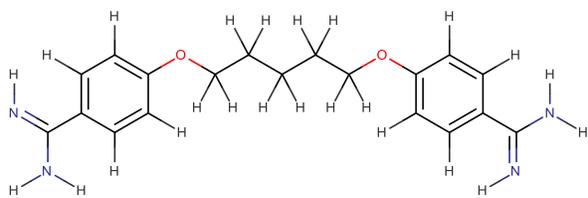
Material Suplementario 2: Ejemplos de reposicionamientos aprobados. Adaptado de Naylor & Schonfeld 2014.

FÁRMACO	INDICACIÓN ORIGINAL	NUEVA INDICACIÓN	AÑO	COMPAÑÍA
ÁCIDO RETINOICO	Acné	Leucemia mieloide aguda	1995	Hoffman La Roche
AMITRIPTILINA	Antidepresivo	Dolor Neuropático	2005	AstraZeneca
ANFOTERICINA B	Antifúngico	Leishmaniasis	1997	NeXstar Pharma
ASPIRINA	Inflamación, Dolor	Anti-plaquetas, ataque cardíaco, accidente cerebrovascular	Varios	Varias
AZATIOPRINA	Artritis Reumatoidea	Enfermedad inflamatoria intestinal, Transplante de órganos	Varios	Varias
BIMATOPROST	Glaucoma	Crecimiento de pestañas	2008	Allergan
BLEOMICINA	Antibiótico	Cáncer	1973	Kayaku/BMS
BROMOCRIPTINA	Enfermedad de Parkinson	Diabetes tipo 2	2009	Novartis
BUPRENORFINA	Dolor	Tratamiento de drogas	2002	Reckitt-Benckiser
BUPROPION	Antidepresivo	Cesación de fumar	1997	GSK
BUPROPION	Antidepresivo	Pérdida de peso	2014	Orexigen/Takeda
CANAKINUMAB	Artritis Reumatoidea	Síndrome de Muckle-Wells	2009	Novartis
CLOFAZIMA	Tuberculosis	Lepra	1986	Geigy
COLCHICINA	Gota	Fiebre mediterránea familiar	2009	URL Pharma
COLESEVELAM	Disminución LDL	Diabetes tipo 2	2008	Daiichi-Sankyo
CRIZOTINIB	Linfoma	Cáncer de pulmón de células no pequeñas	2011	Pfizer
CICLOSERINA	Tuberculosis	Desórdenes de SNC	Varios	Varias
CICLOSPORINA	Rechazo de trasplante de órganos	Psoriasis	1997	Novartis
DAPOXETINA	Antidepresivo	Eyaculación precoz	2004	J&J
DIMETIL FUMARATO	Psoriasis	Esclerosis múltiple	2013	Biogen IDEC
DONEPEZILO	Alzheimer	Demencia	2006	Eisai/Pfizer
DOXEPINA	Antidepresivo	Dermatitis atópica	2003	Varias

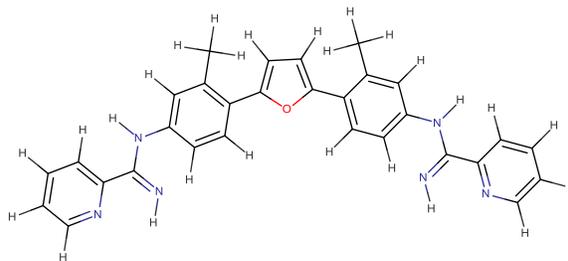
DULOXETINA	Depresión	Incontinencia urinaria de esfuerzo	2004	Lilly
DULOXETINA	Depresión	Fibromialgia	2008	Lilly
DULOXETINA	Depresión	Dolor	2010	Lilly
EFLORNITINA	Cáncer	Hirsutismo	2000	Gillette
EFLORNITINA	Cáncer	Enfermedad del Sueño	1990	Aventis
ETANERCEPT	Artritis Reumatoidea	Psoriasis	2004	Amgen/Pfizer
EVEROLIMUS	Rechazo de trasplante de órganos	Cáncer	Varios	Novartis
FINASTERIDA	Hipertensión	Hiperplasia de próstata benigna	1992	Merck
FINASTERIDA	Hipertensión	Calvicie	1997	Merck
FLUOXETINA	Hipertensión	PMDD	2002	Lilly
GABAPENTINA	Ataques	Neuralgia postherpética	2004	Parke Davis
GALANTAMINA	Síndrome de la fatiga crónica	Alzheimer	2001	Varias
GEMCITABINA	Antiviral	Cáncer	Varios	Lilly
GLICOPIRRONIO	Antiulceroso	COPD	2005	Sosei/Novartis
HISTRELINA	Cáncer de próstata	Pubertad precoz	2007	Endo Pharma
HIDROXICLOROQUINA	Malaria	Lupus	Varios	Varias
IBUPROFENO	Inflamación, Dolor	Dolor de cabeza, migraña	Varios	Varias
IMATINIB	Leucemia mieloide crónica	Tumor estromal gastrointestinal	2012	Novartis
IMATINIB	Leucemia mieloide crónica	Leucemia linfoide aguda	2013	Novartis
IMFLIXIMAB	Enfermedades autoinmunes	Enfermedad de Crohn	1998	Janssen
IPRONIAZIDA	Tuberculosis	Antidepresivo	1958	Varias
LOMITAPIDA	Hipercolesterolemia	Hipercolesterolemia familiar homocigótica	2012	Aegerion Pharma
METOTREXATO	Cáncer	Psoriasis	2001	Barr Labs
MINOXIDIL	Hipertensión	Calvicie	1988	Upjohn
MILNACIPRAN	Antidepresivo	Fibromialgia	2009	Forest Pharma
MILTEFOSINA	Cáncer	Leishmaniasis	2014	Zentaris
NALTREXONA	Adicciones	Pérdida de peso	2014	Orexigen/Takeda
ONABOTULINUMTOXINA	Espasmo facial	Distonía cervical	2000	Allergan
ONABOTULINUMTOXINA	Espasmo facial	Migraña crónica	2010	Allergan
ONABOTULINUMTOXINA	Espasmo facial	Cosméticos faciales	2012	Allergan

PACLITAXEL	Cáncer	Prevención de la reestenosis del stent	Varios	Varias
PAROXETINA	Antidepresivo	Sofocos menopáusicos	2013	GSK
PLERIXAFOR	AIDS/HIV	Linfoma y mieloma múltiple	2008	Genzyme
PRAMIPEXOL	Enfermedad de Parkinson	Síndrome de la pierna inquieta	2006	Boehringer
PREGABALINA	Anticonvulsivo	Fibromialgia	2007	Pfizer
PROPRANOLOL	Hipertensión	Migraña, angina	Varios	Varias
RALOXIFENO	Osteoporosis	Cáncer de mama	2007	Lilly
RITUXIMAB	Cáncer	Artritis Reumatoidea	2004	IDEC
ROPININOLA	Enfermedad de Parkinson	síndrome de la pierna inquieta	2005	GSK
SILDENAFIL	Angina	Disfunción eréctil	1998	Pfizer
SILDENAFIL	Angina	Hipertensión pulmonar	2005	Pfizer
SUNITINIB	GIST and RCC	Tumores pancreáticos	2010	Pfizer
TALIDOMIDA	Anti-náusea	Lepra	1998	Celgene
TALIDOMIDA	Anti-náusea	Mieloma múltiple	2006	Celgene
ZIDOVUDINE	Cáncer	HIV/AIDS	1987	Burroughs

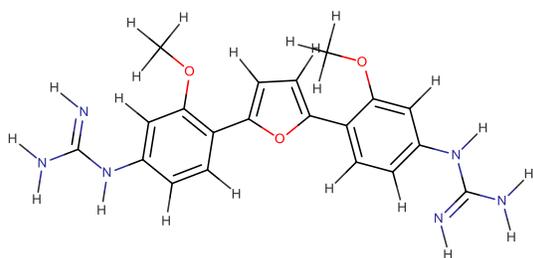
Material Suplementario 3a: Estructuras de los compuestos del conjunto de entrenamiento.



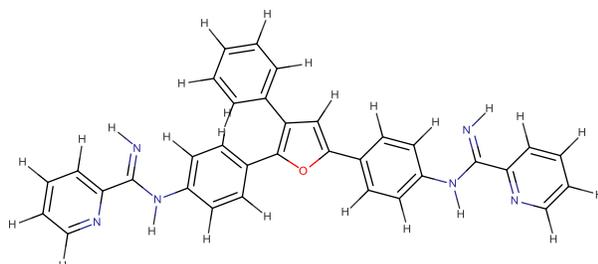
A01



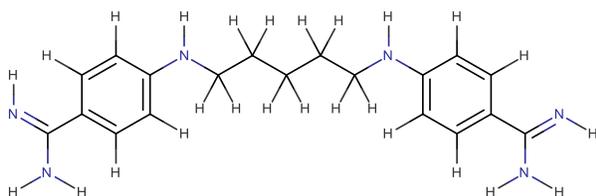
A02



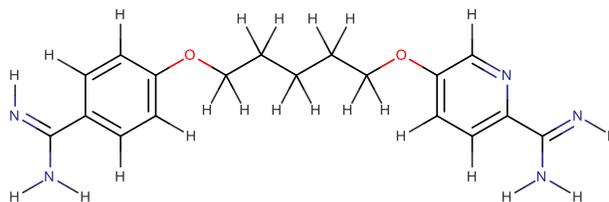
A03



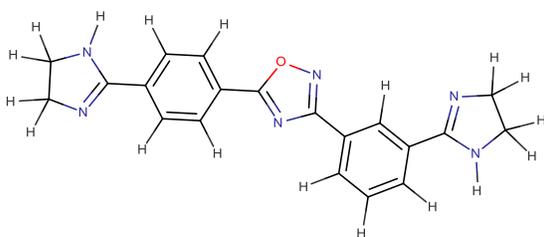
A06



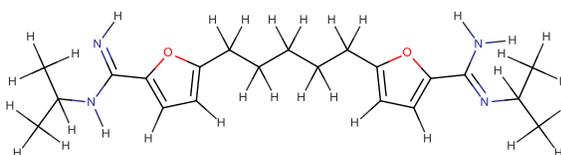
A32



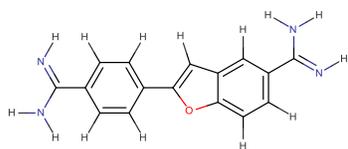
A34



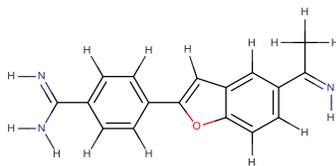
A35



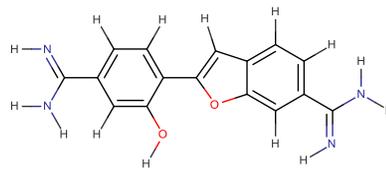
A37



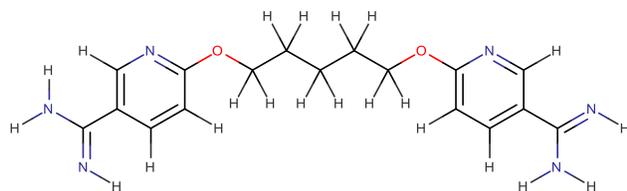
A38



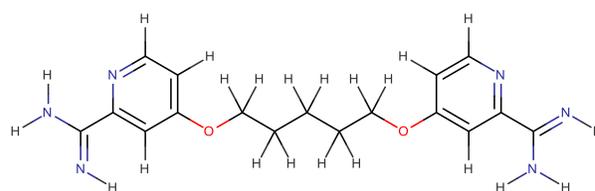
A39



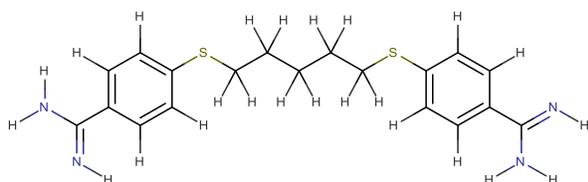
A40



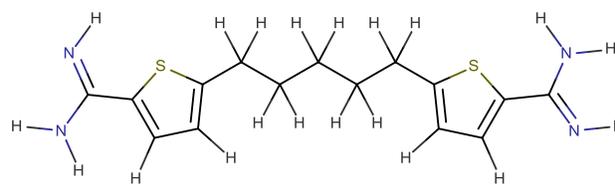
A44



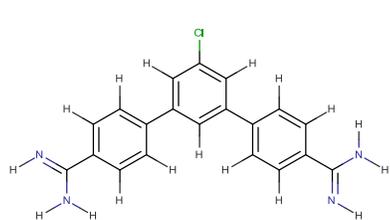
A45



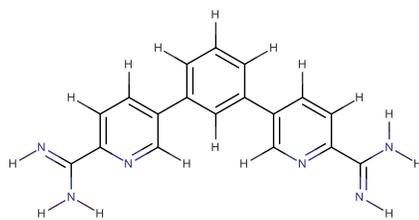
A46



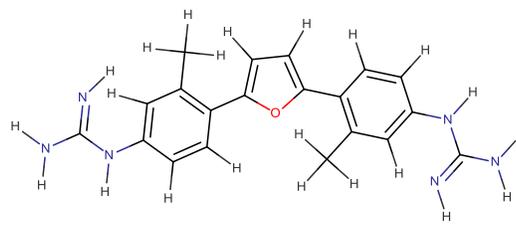
A48



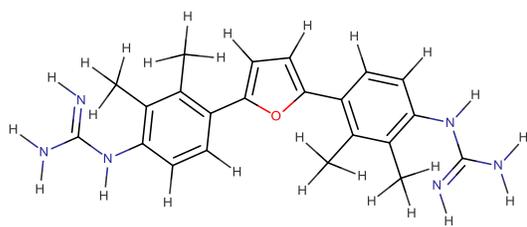
A50



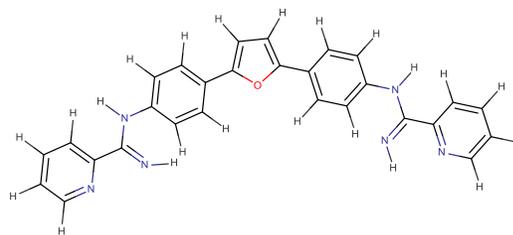
A51



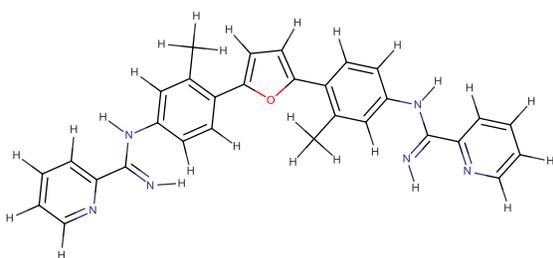
A54



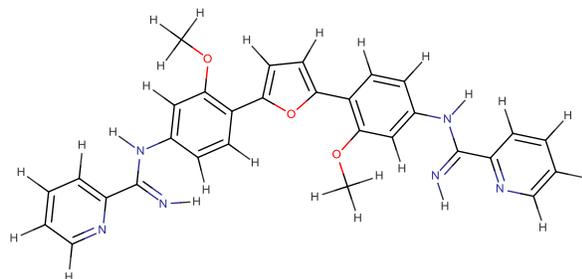
A59



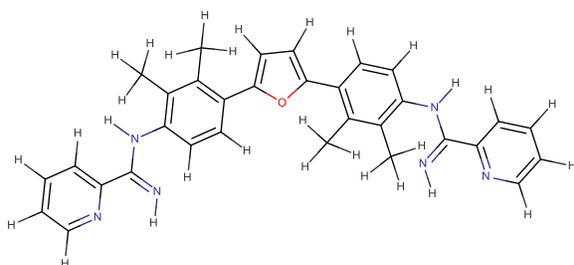
A64



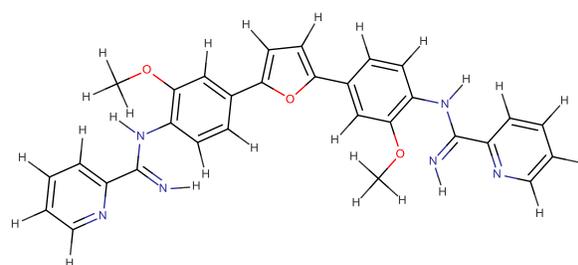
A65



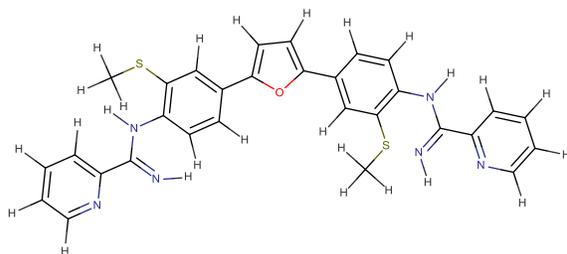
A66



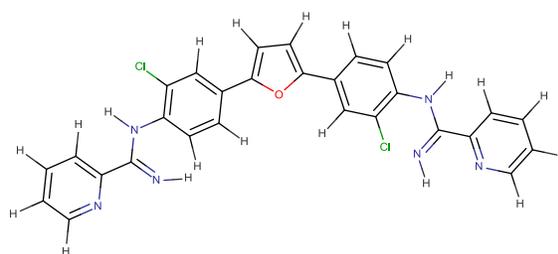
A68



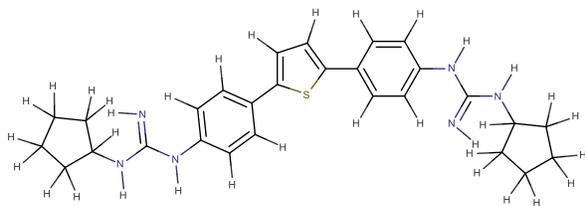
A69



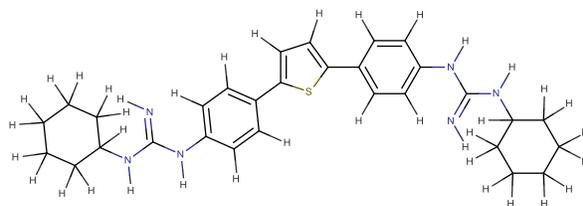
A70



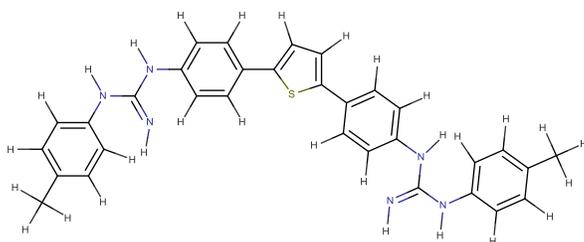
A72



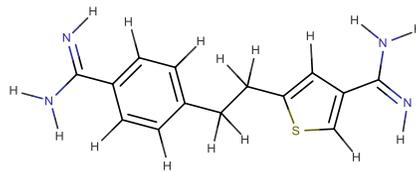
A76



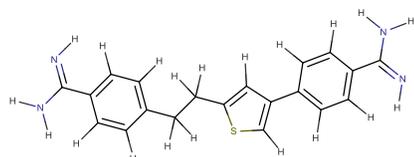
A77



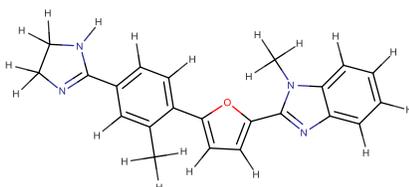
A79



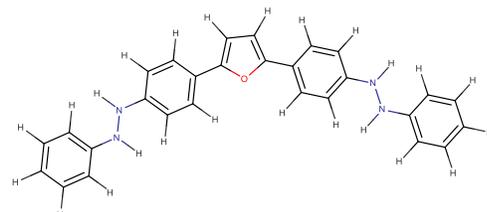
A80



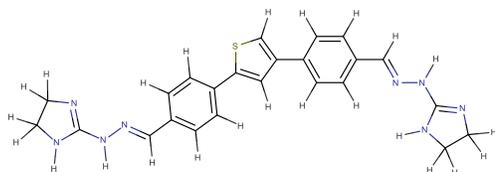
A81



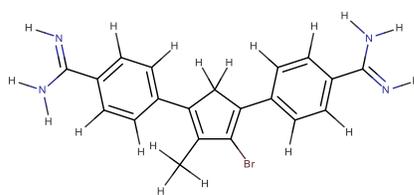
A83



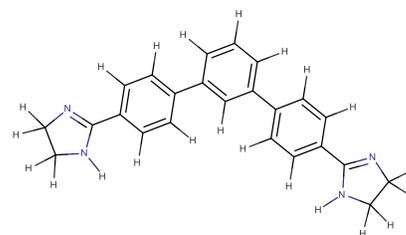
A84



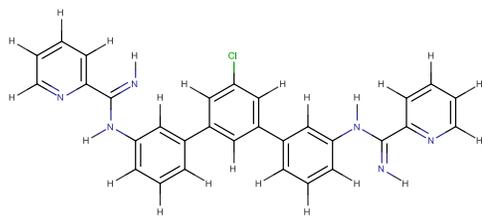
A85



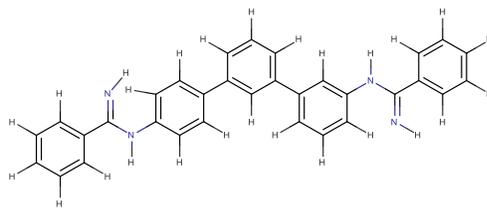
A86



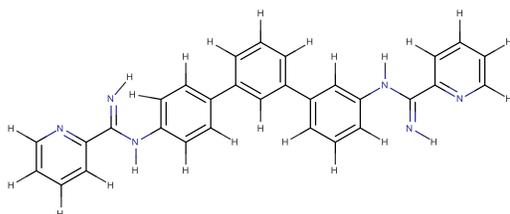
A89



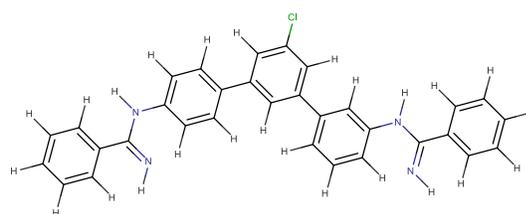
A117



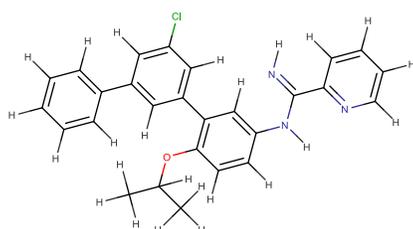
A118



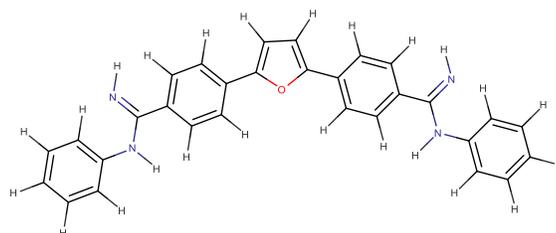
A119



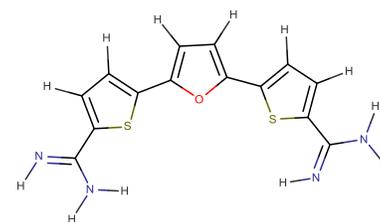
A120



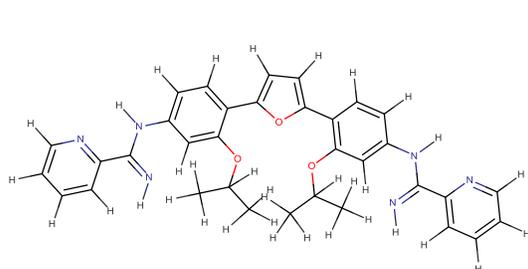
A123



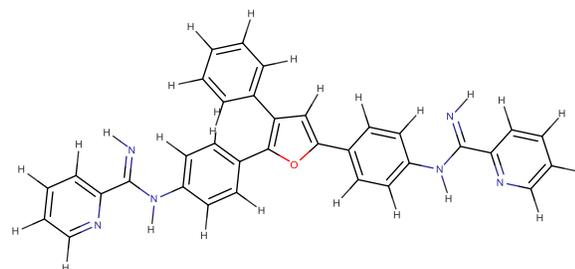
A126



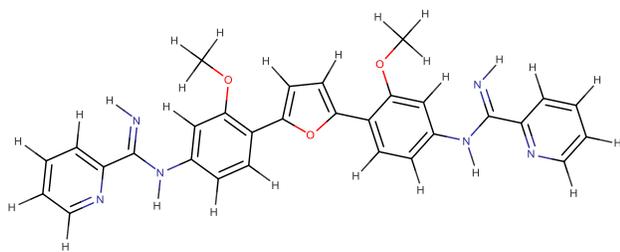
A127



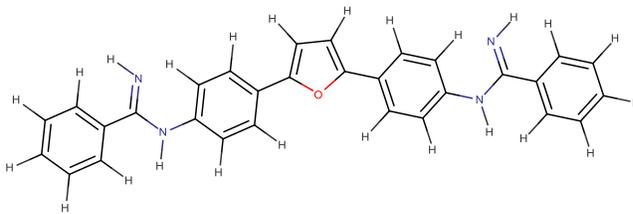
A129



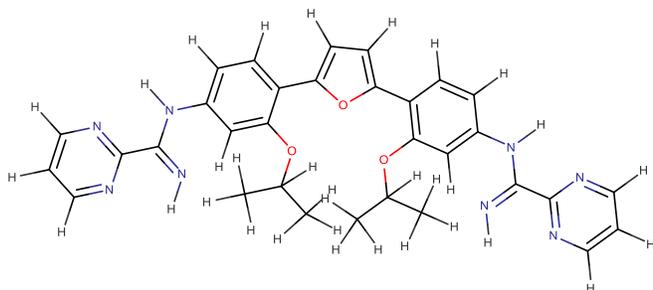
A131



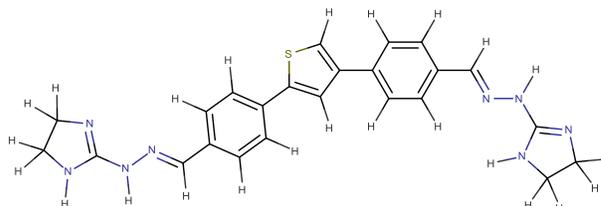
A132



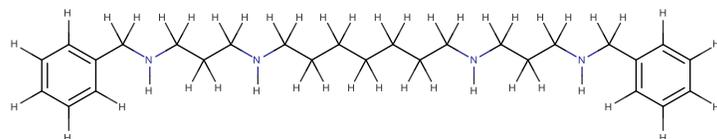
A133



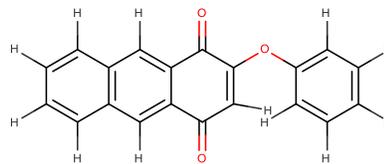
A134



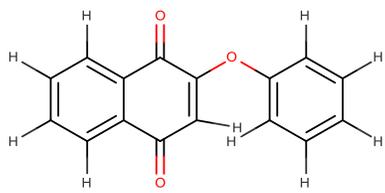
A137



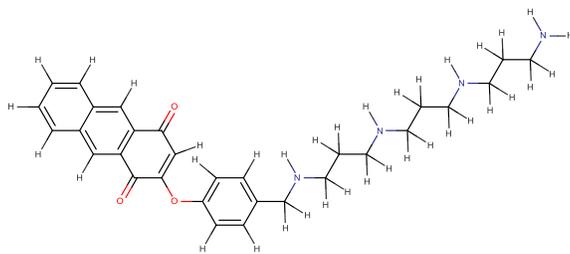
A141



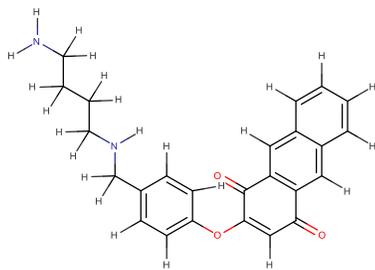
A142



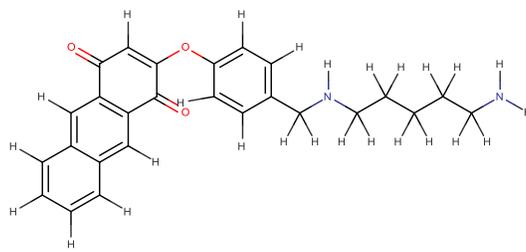
A143



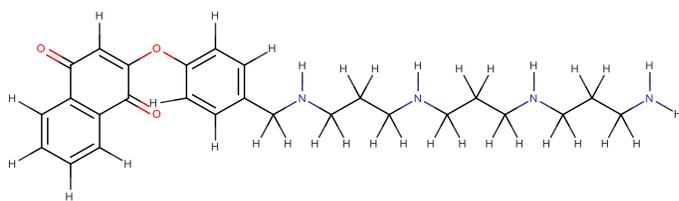
A145



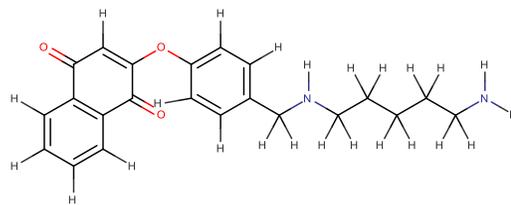
A147



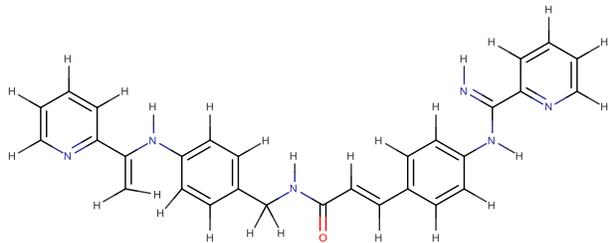
A148



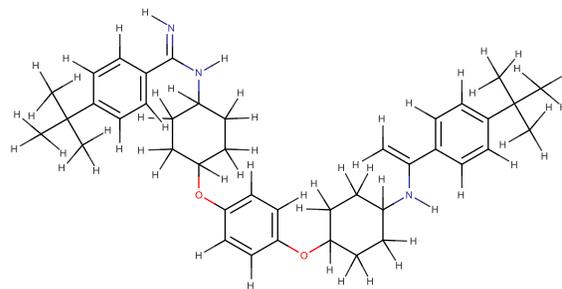
A150



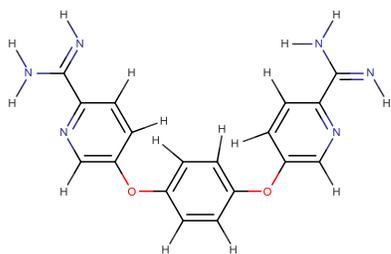
A153



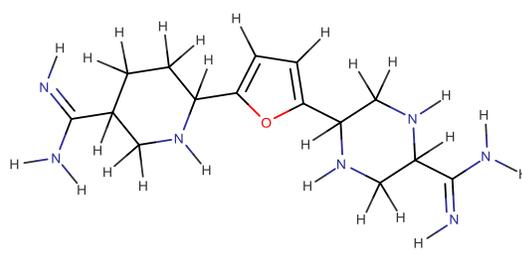
A154



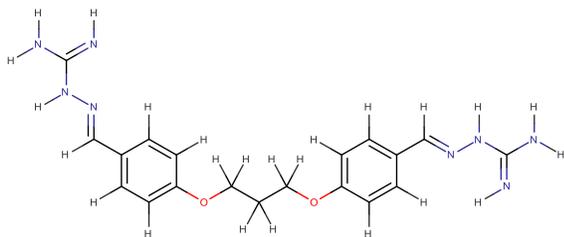
A155



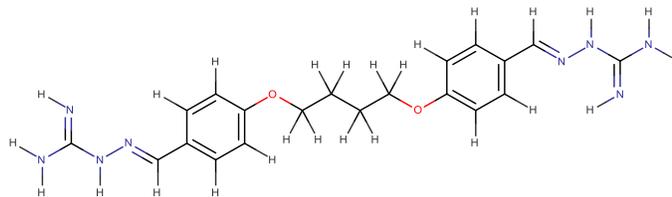
A157



A159



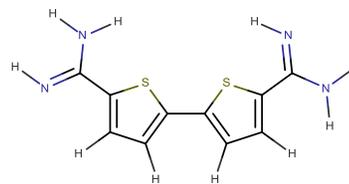
A161



A162

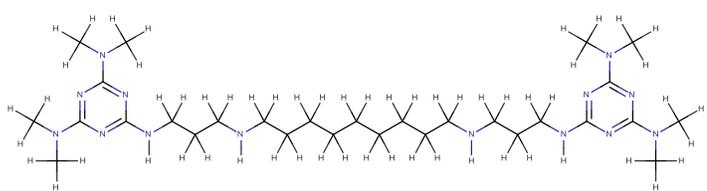


A163

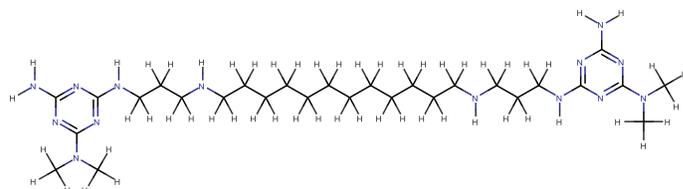


A164

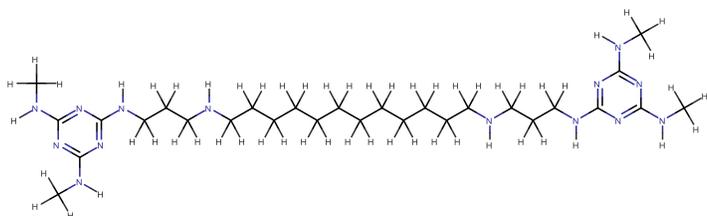
MATERIAL SUPLEMENTARIO



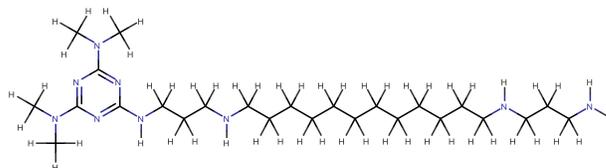
A179



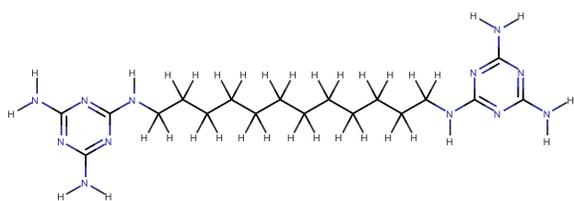
A181



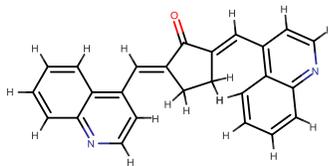
A182



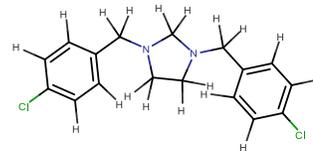
A188



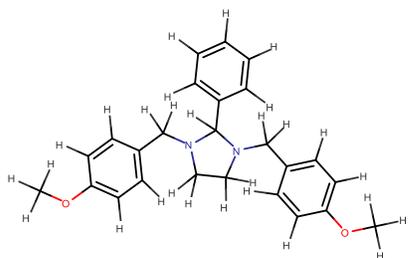
A195



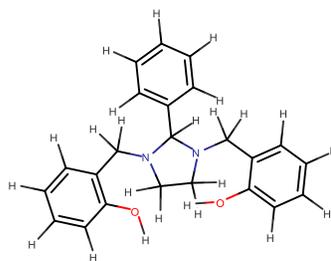
A206



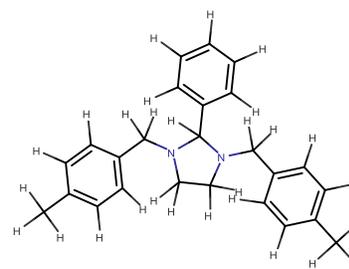
A227



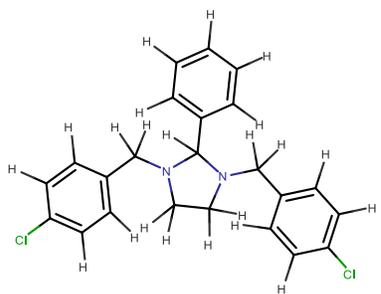
A239



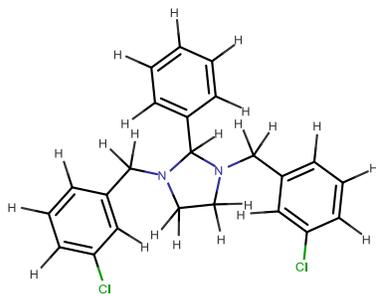
A240



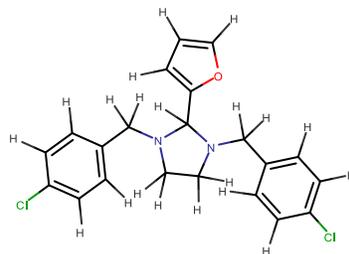
A241



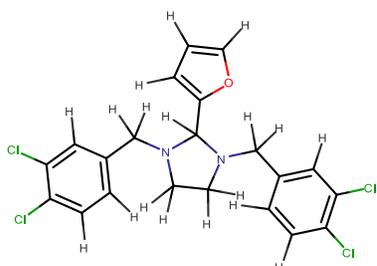
A242



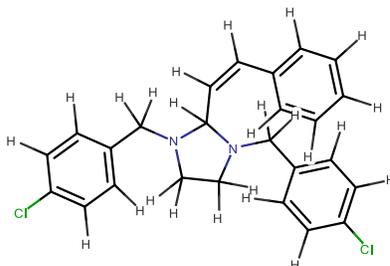
A243



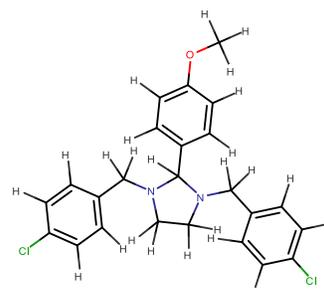
A245



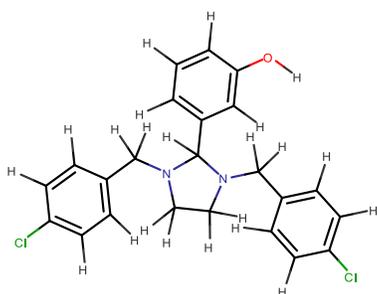
A246



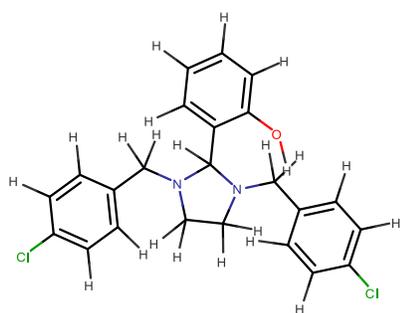
A248



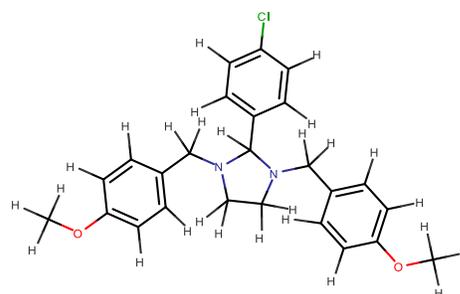
A249



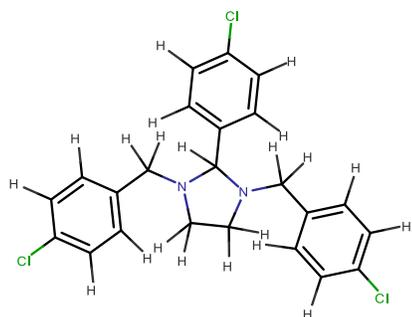
A250



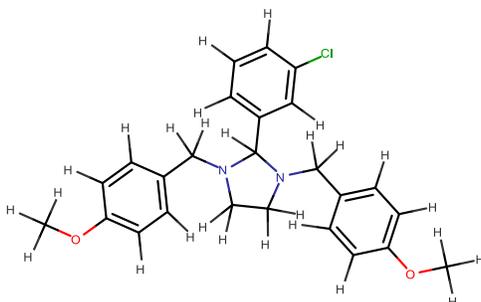
A251



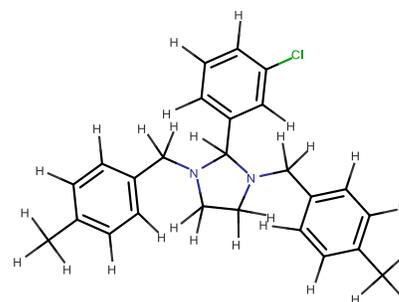
A252



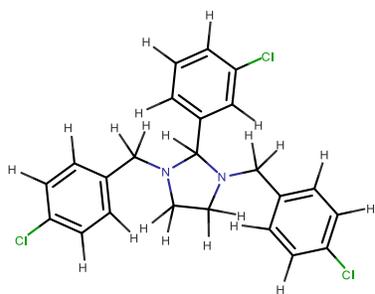
A255



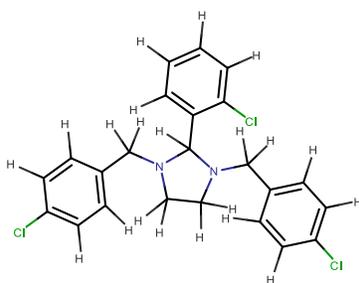
A257



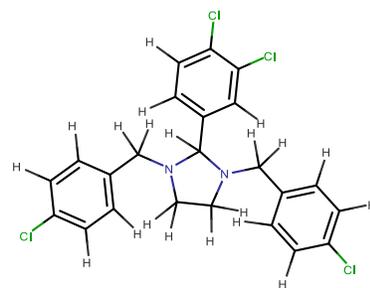
A258



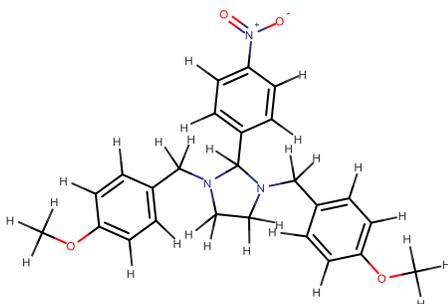
A259



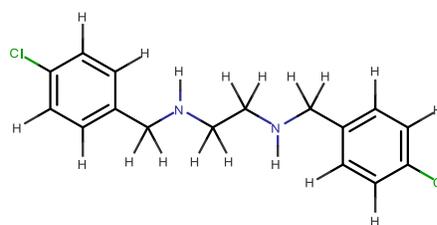
A261



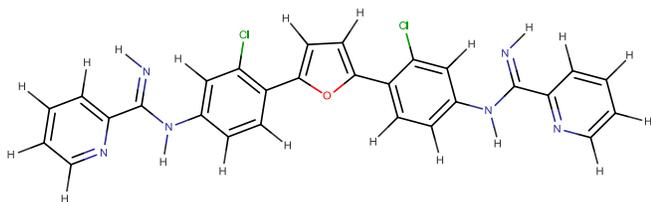
A263



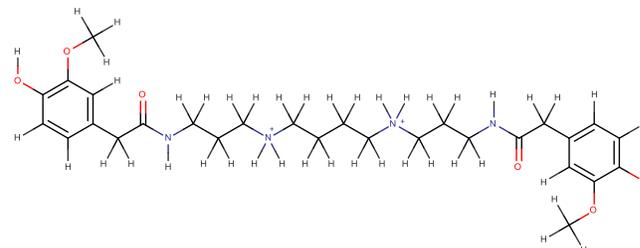
A265



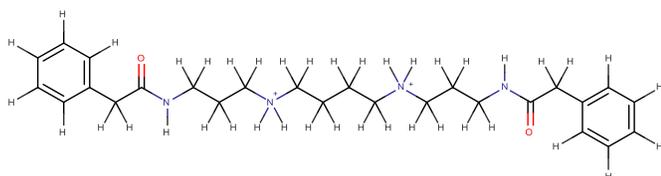
A269



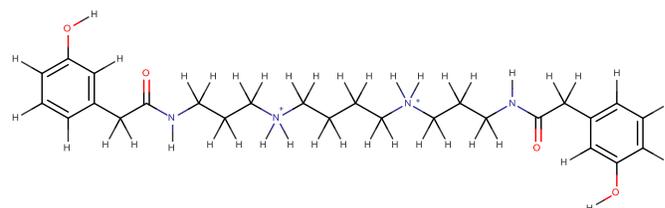
I05



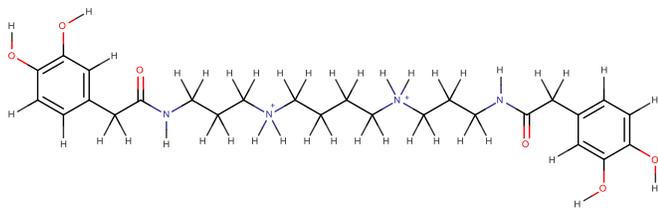
I07



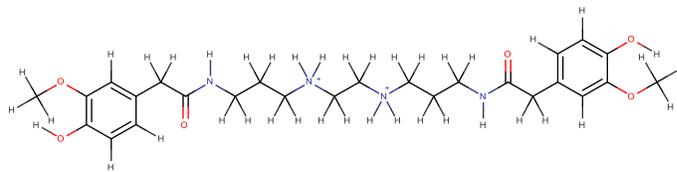
I08



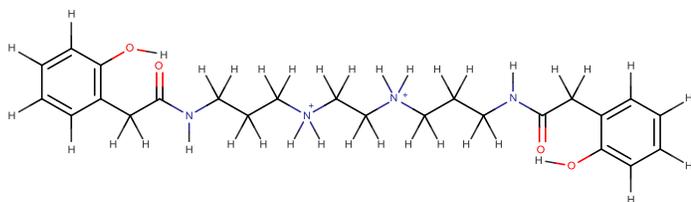
I10



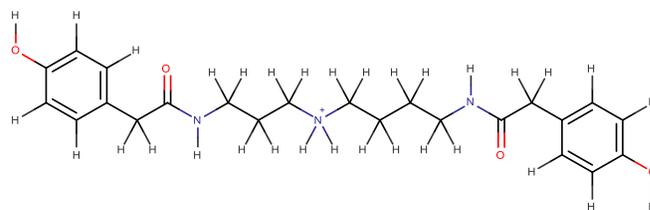
I12



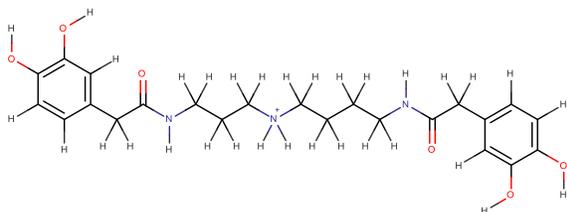
I14



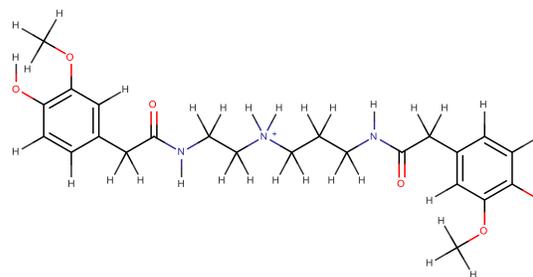
I15



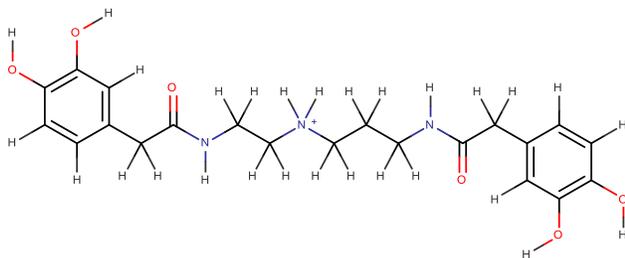
I17



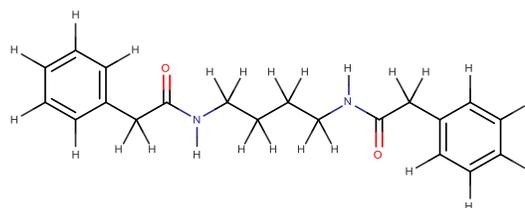
I18



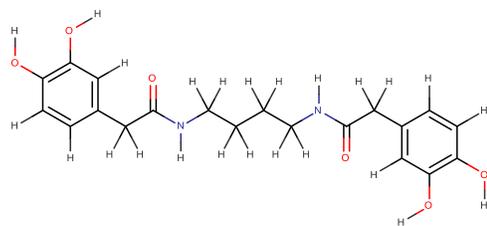
I19



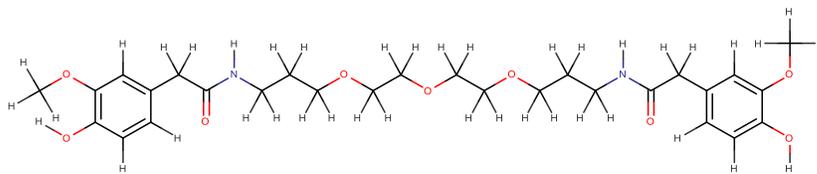
I21



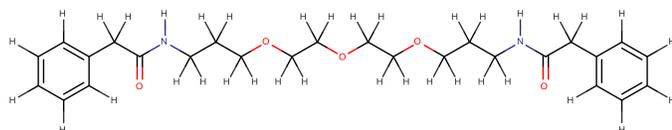
I23



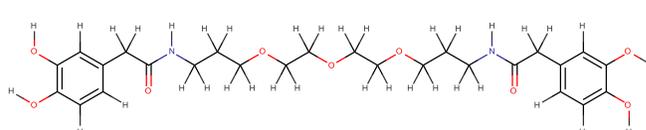
124



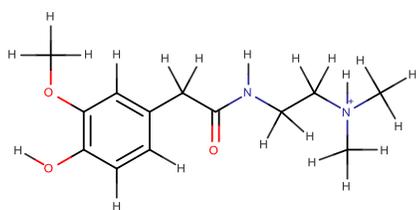
125



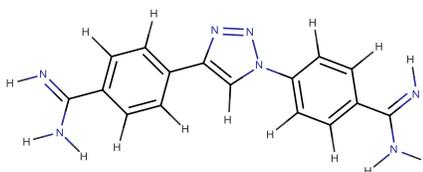
126



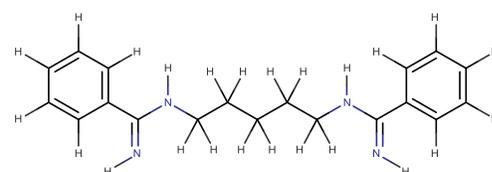
127



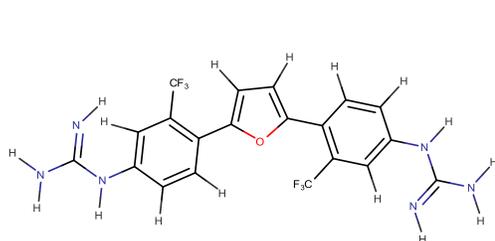
129



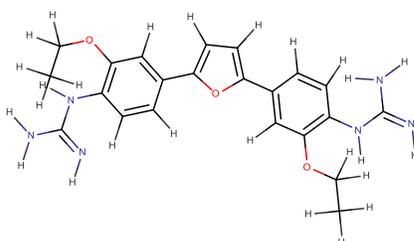
142



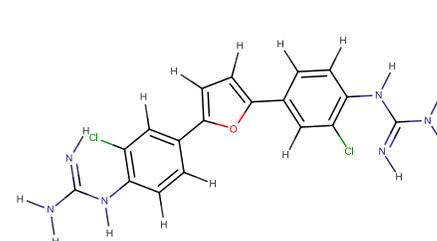
149



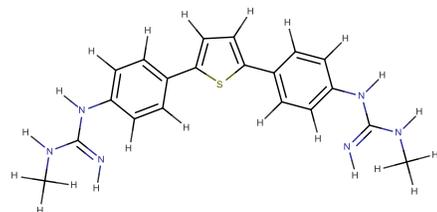
158



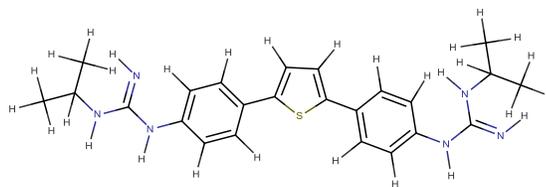
161



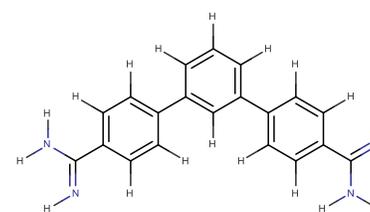
163



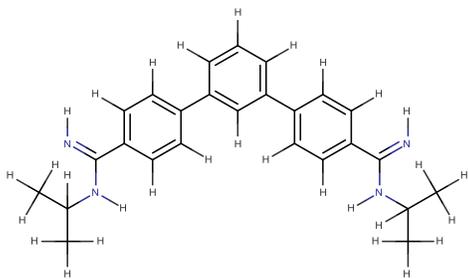
174



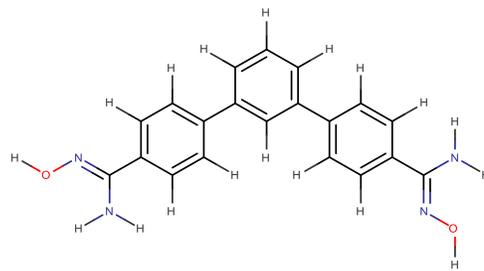
175



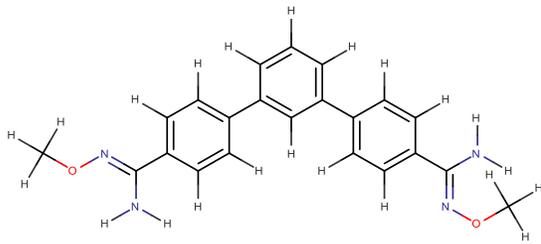
187



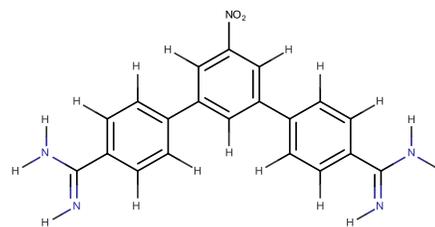
188



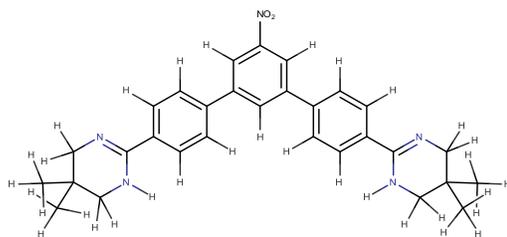
191



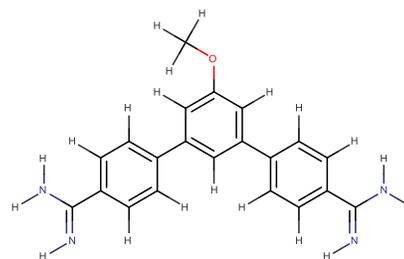
192



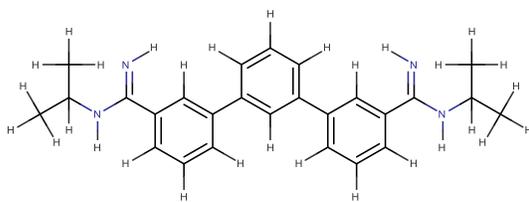
194



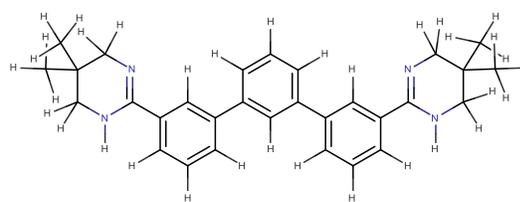
195



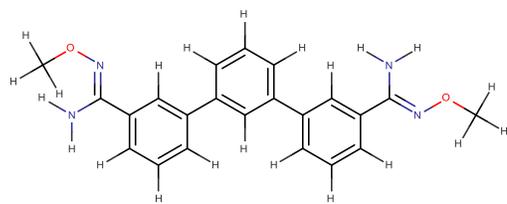
198



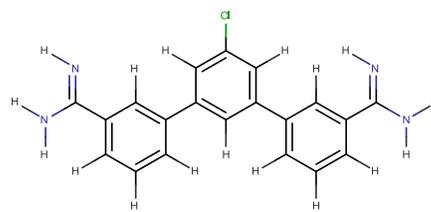
1100



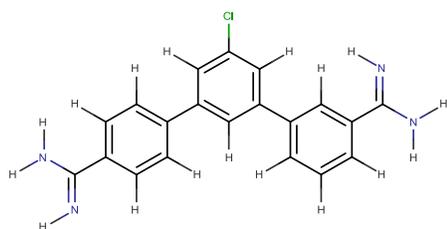
1102



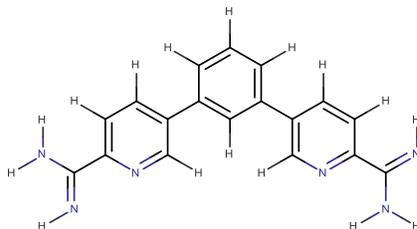
I104



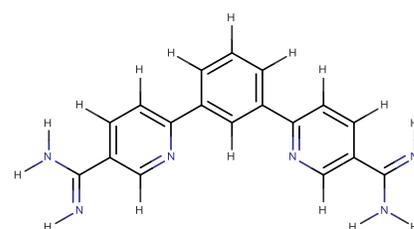
I105



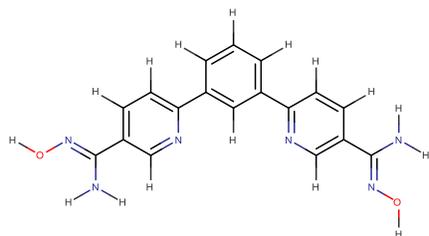
I108



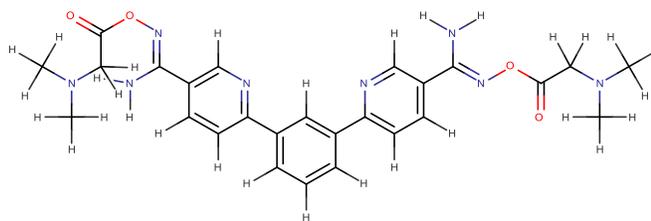
I109



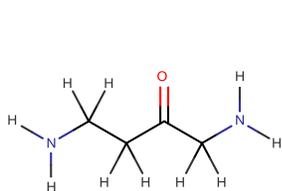
I111



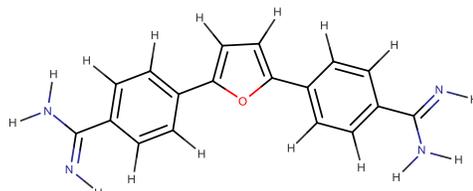
I112



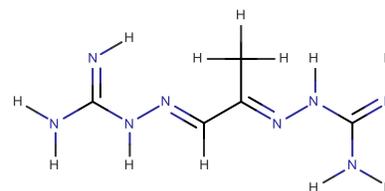
I114



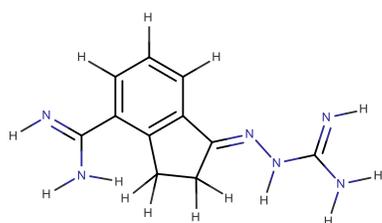
I124



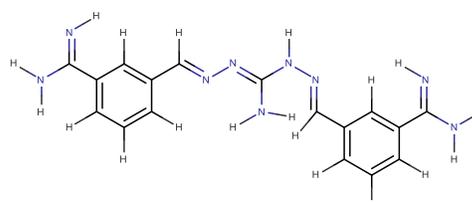
I125



I138

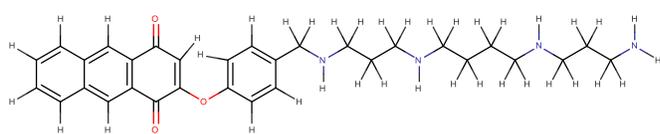


I139

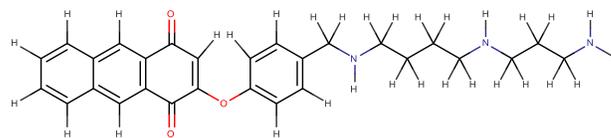


I140

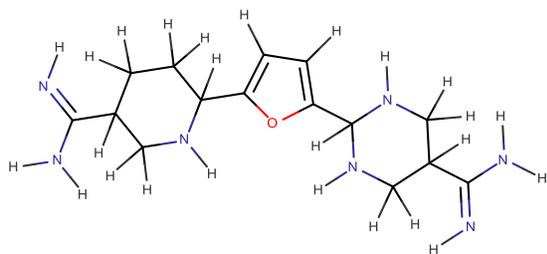
MATERIAL SUPLEMENTARIO



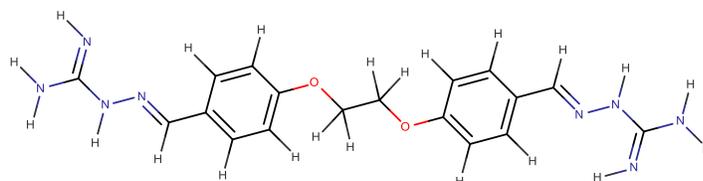
I144



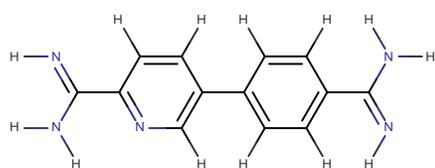
I146



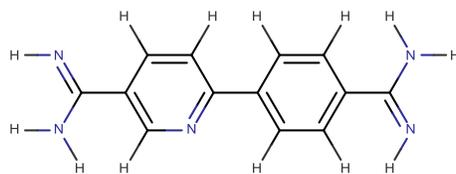
I156



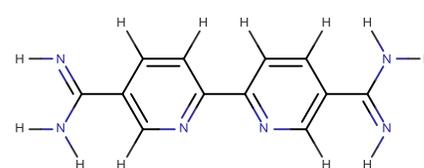
I160



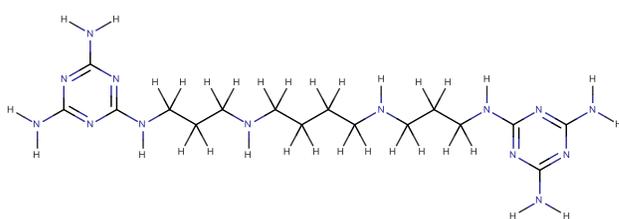
I165



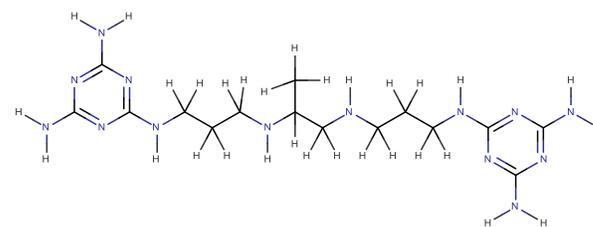
I167



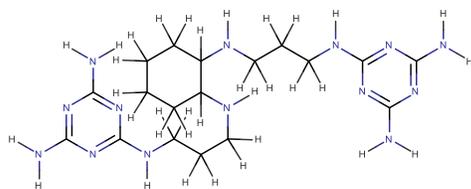
I169



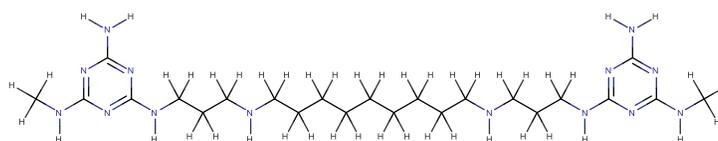
I170



I173

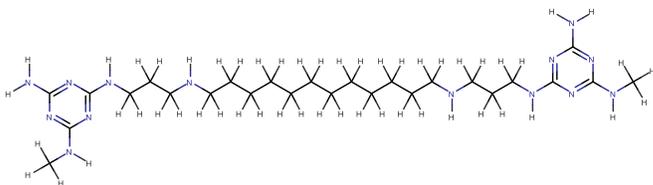


I174

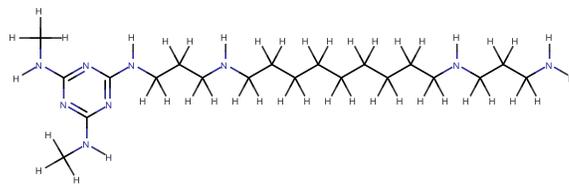


I176

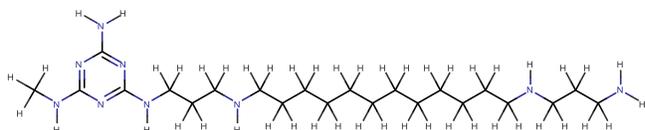
MATERIAL SUPLEMENTARIO



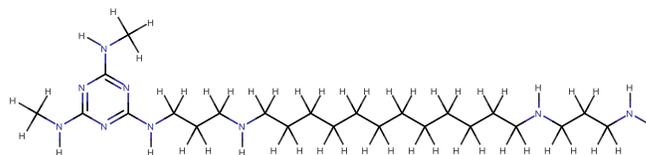
I180



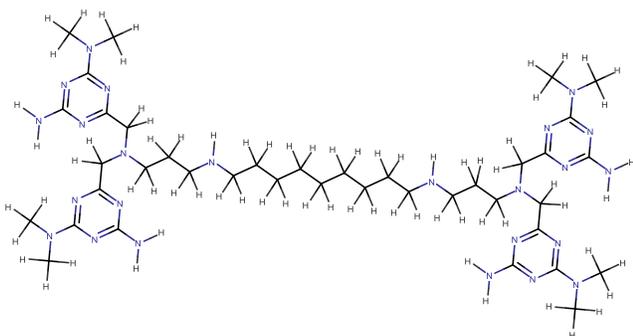
I184



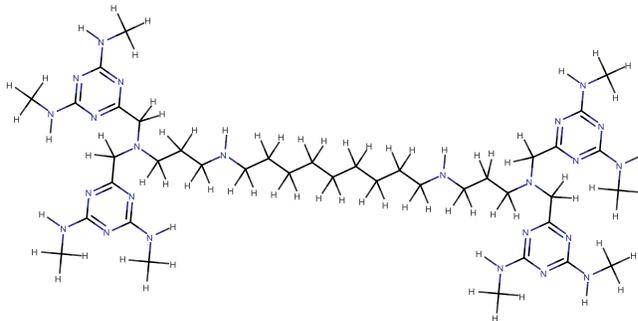
I185



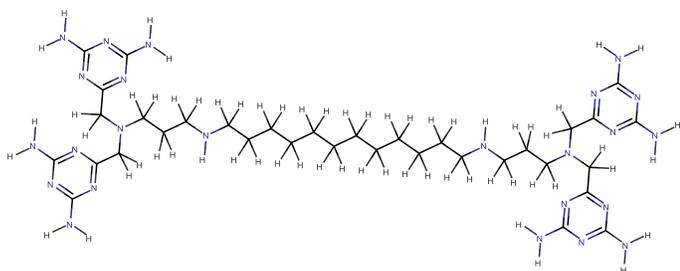
I187



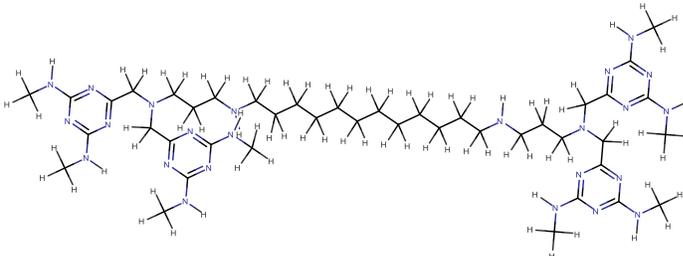
I190



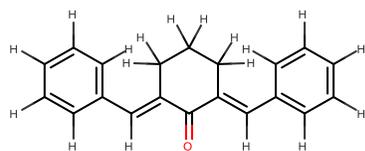
I191



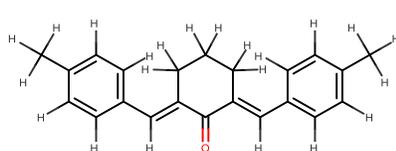
I192



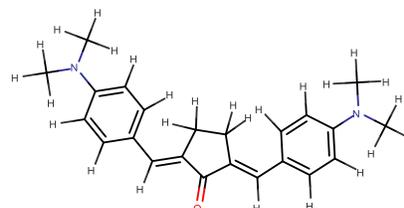
I194



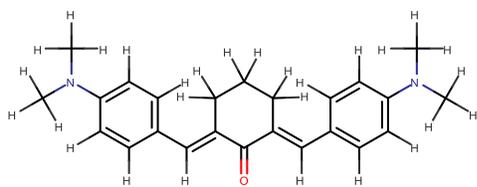
I198



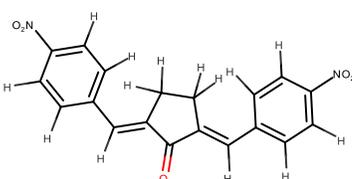
I200



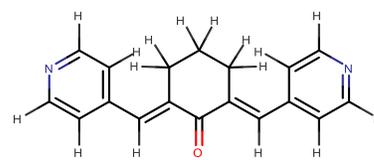
I201



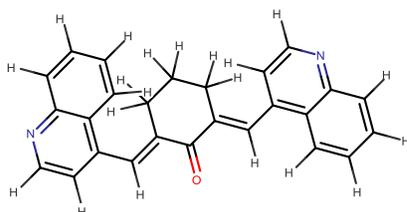
I202



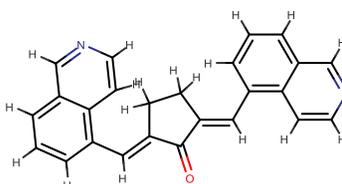
I203



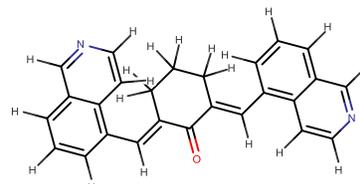
I205



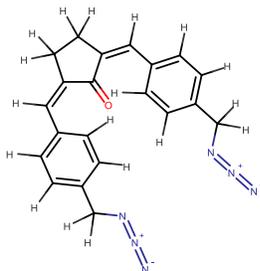
I207



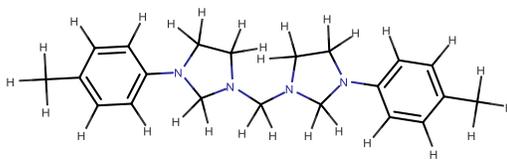
I208



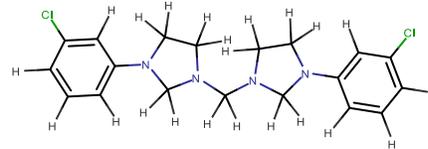
I209



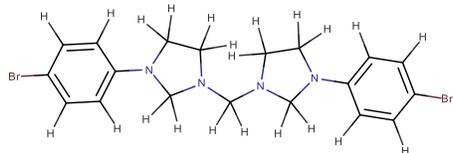
I214



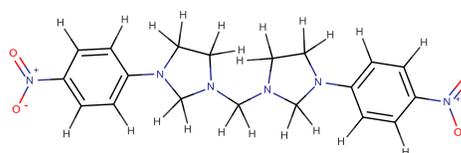
I216



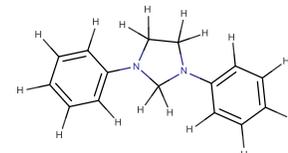
I219



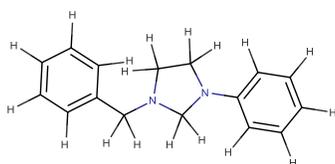
I220



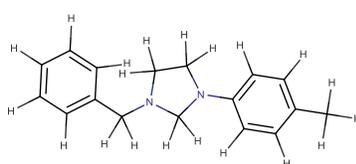
I222



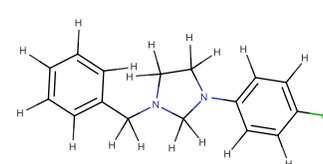
I223



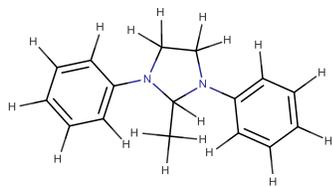
I224



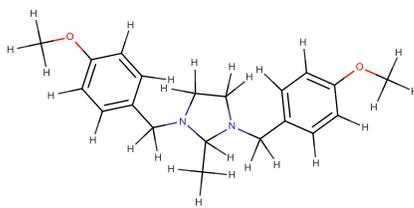
I225



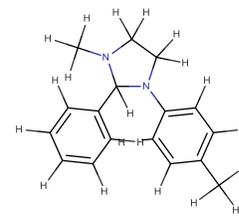
I226



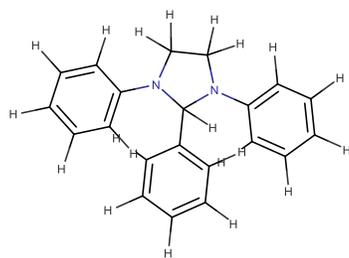
I230



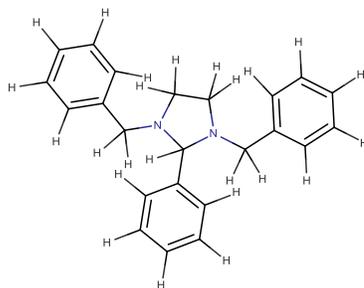
I231



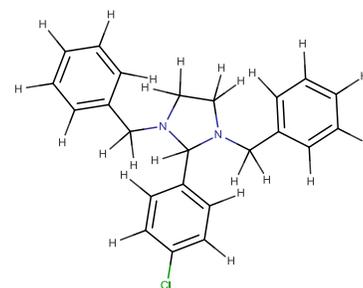
I233



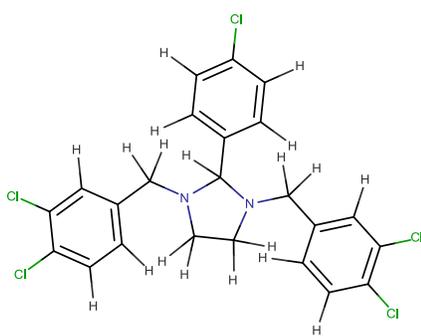
I234



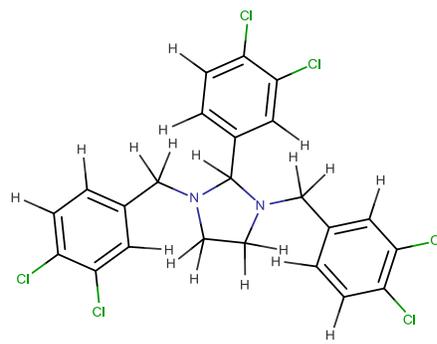
I238



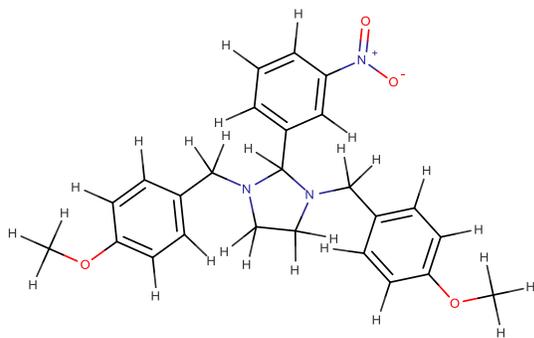
I254



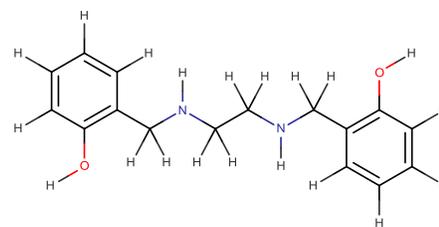
I256



I264

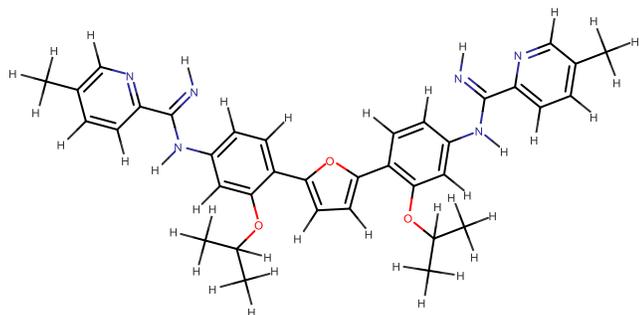


I267

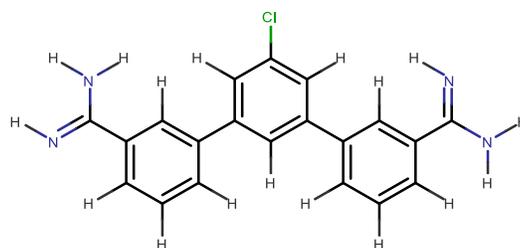


I270

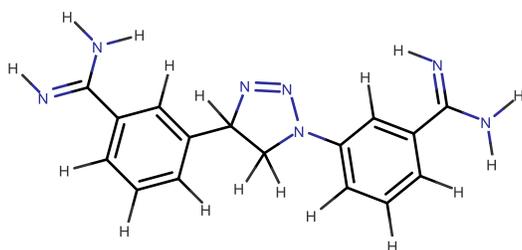
Material Suplementario 3b: Estructuras de los compuestos del conjunto de prueba.



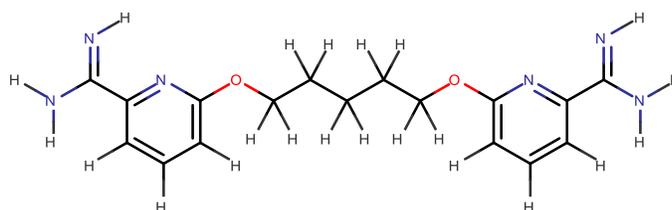
A04



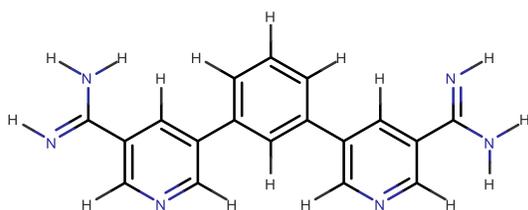
A33



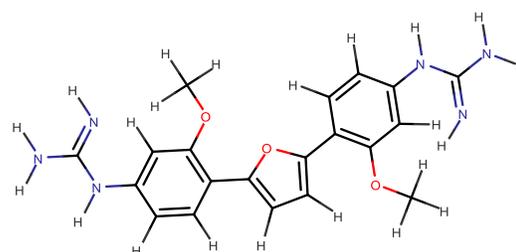
A36



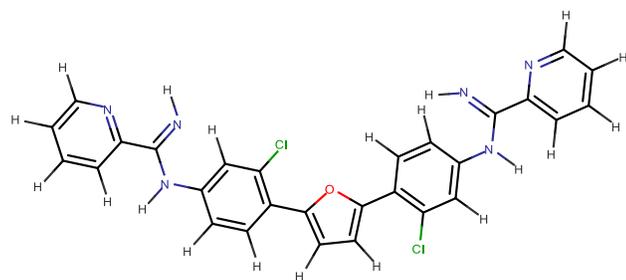
A43



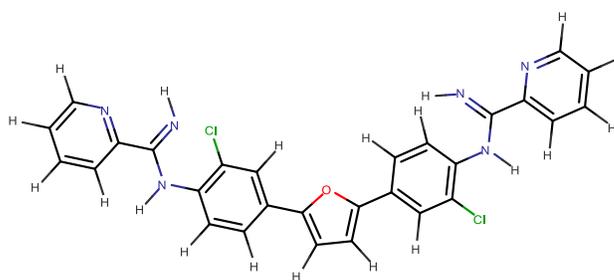
A52



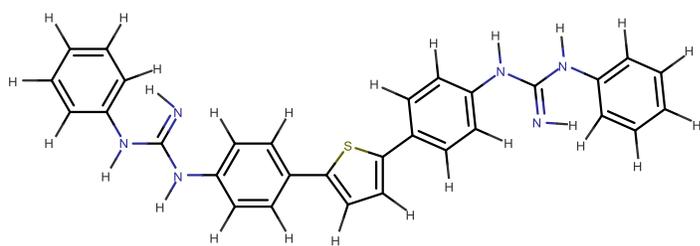
A55



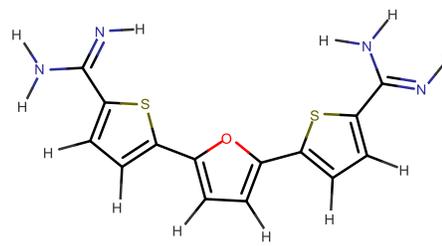
A67



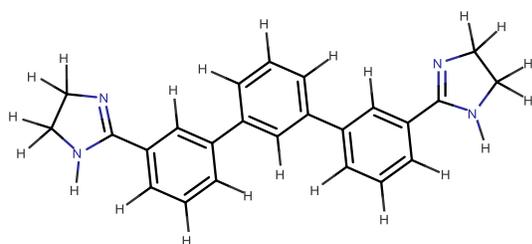
A71



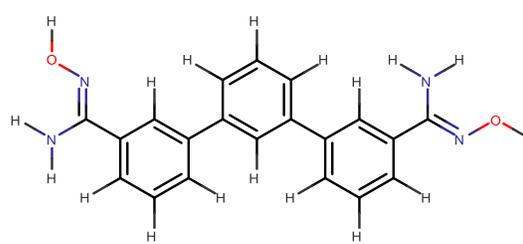
A78



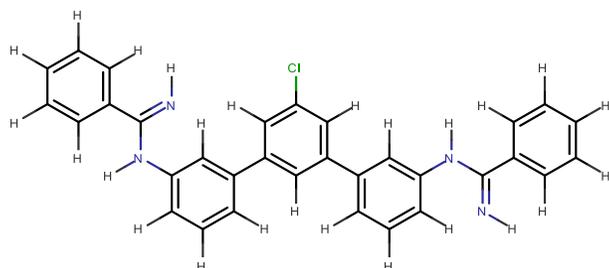
A82



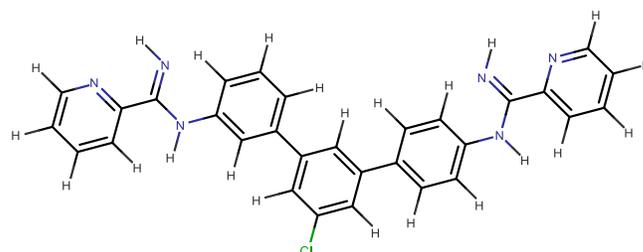
A101



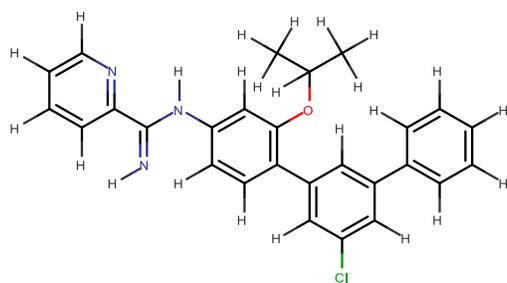
A103



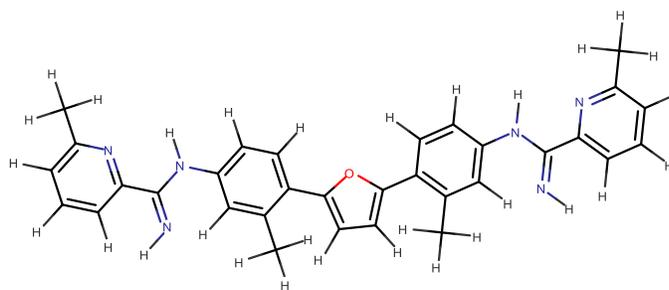
A116



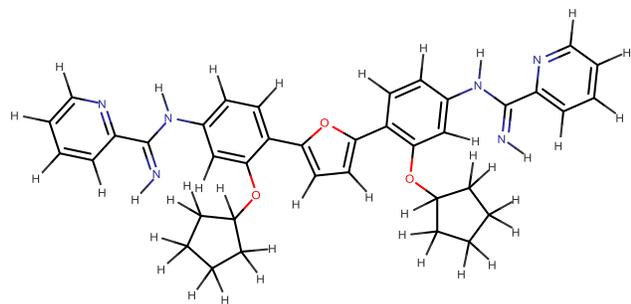
A121



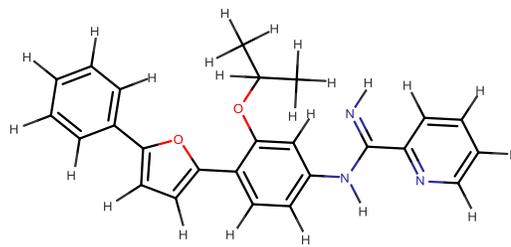
A122



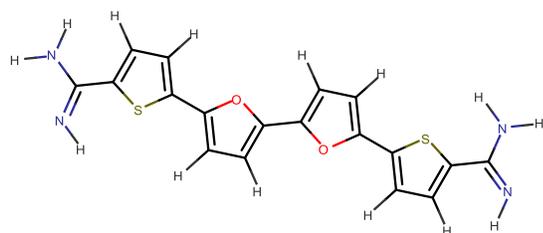
A130



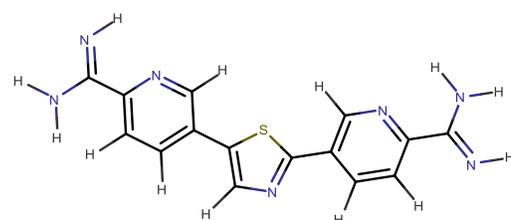
A135



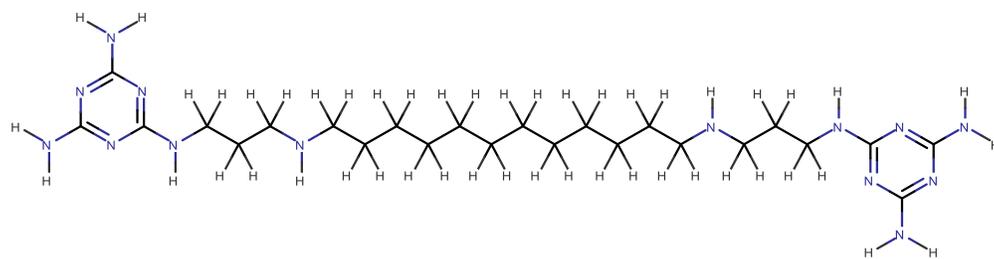
A136



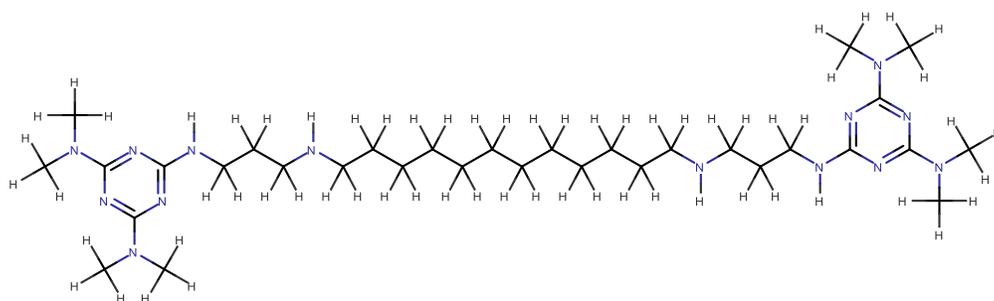
A166



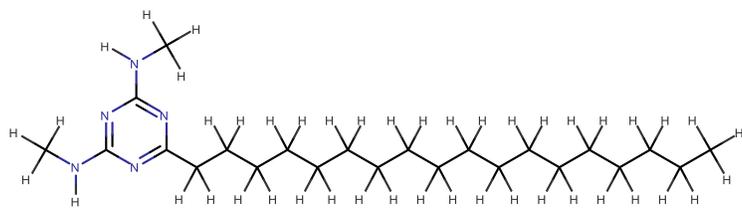
A168



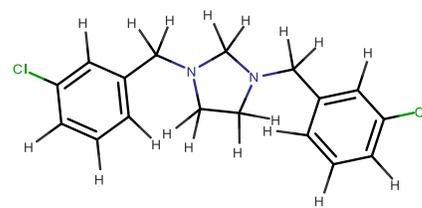
A172



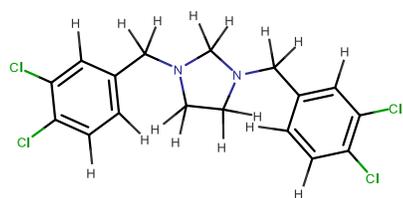
A183



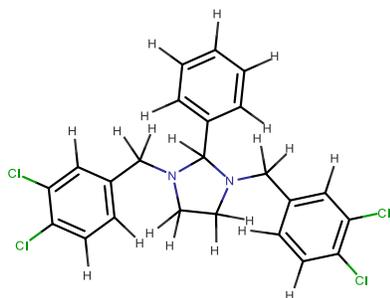
A196



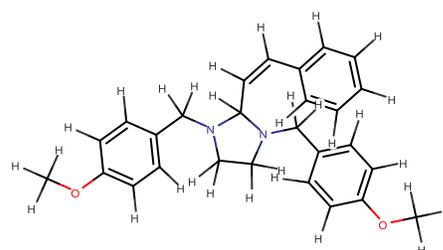
A228



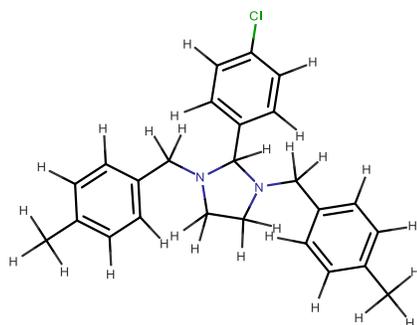
A229



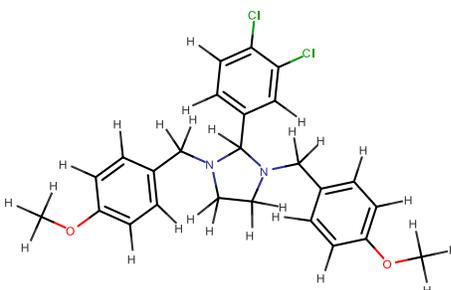
A244



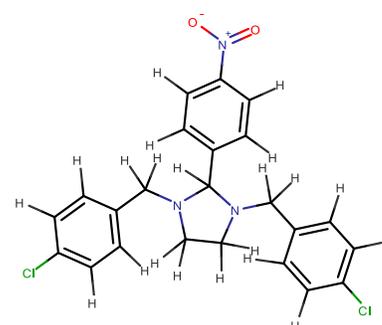
A247



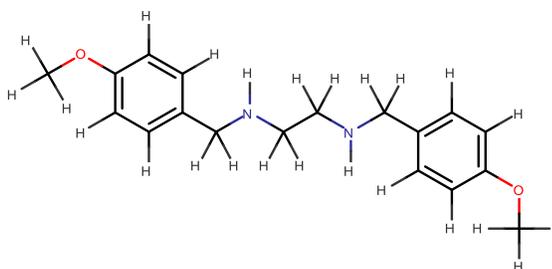
A253



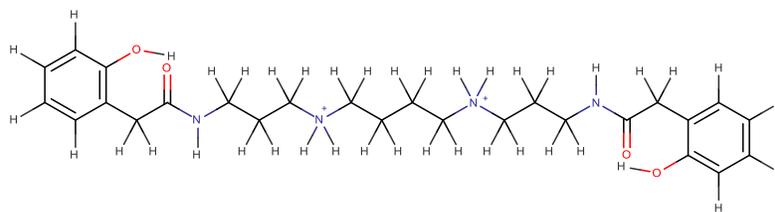
A262



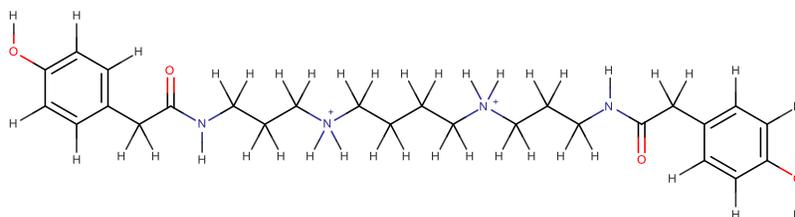
A266



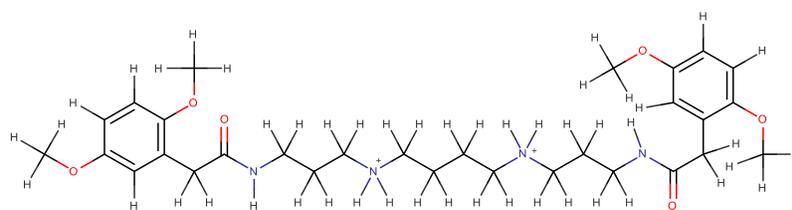
A68



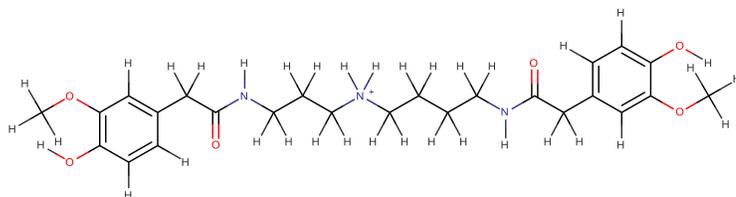
109



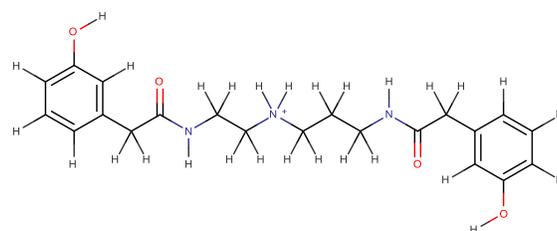
111



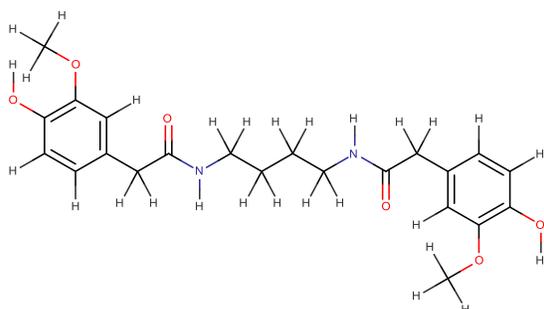
113



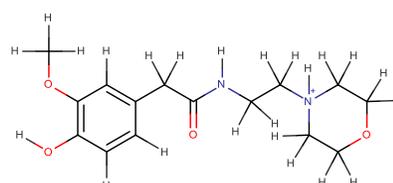
116



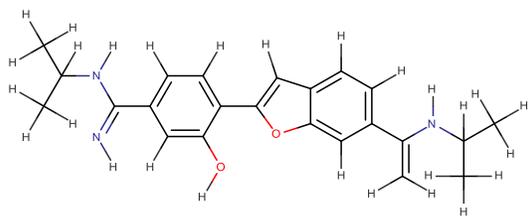
120



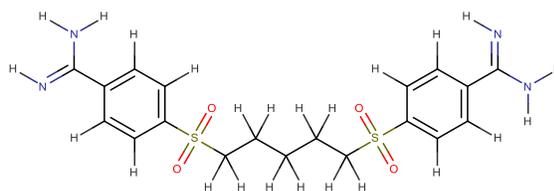
122



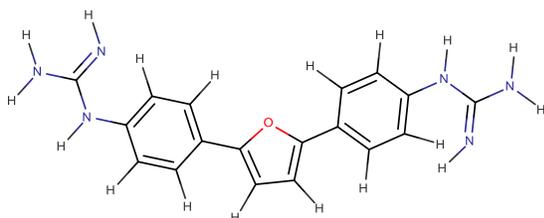
128



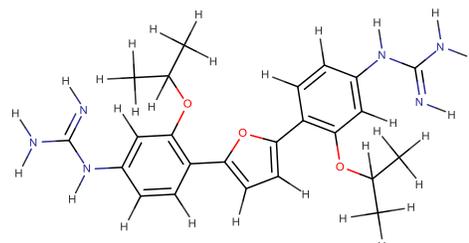
141



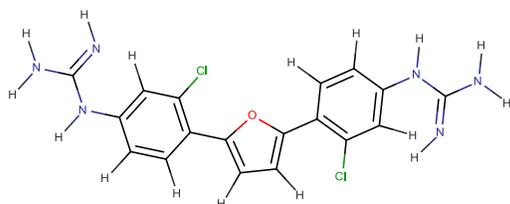
147



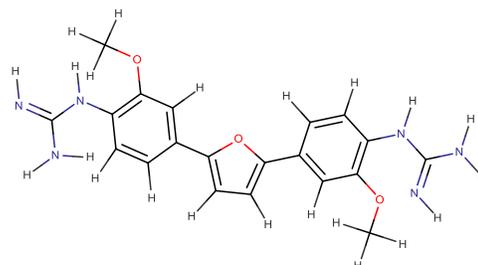
153



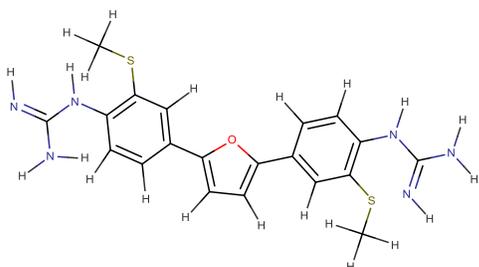
156



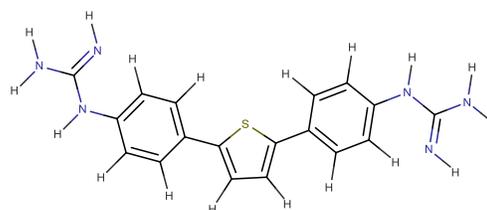
157



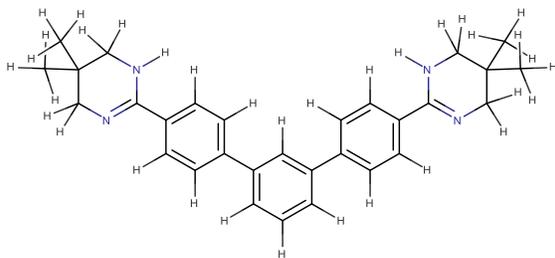
160



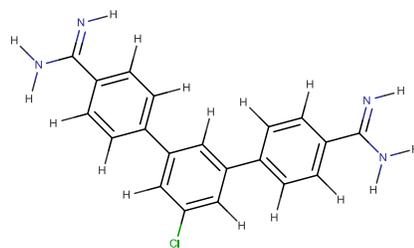
162



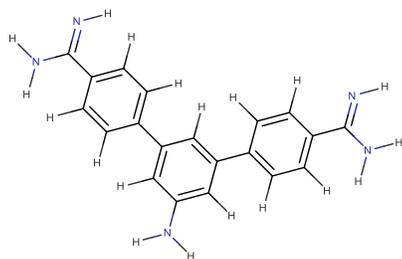
173



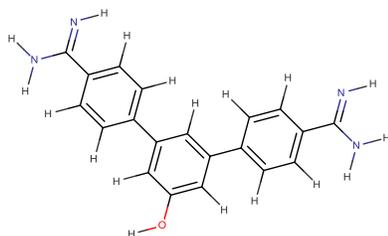
I90



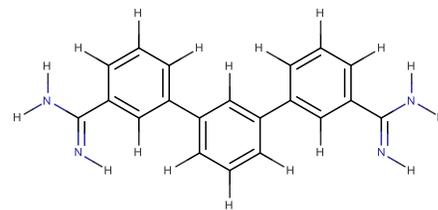
I93



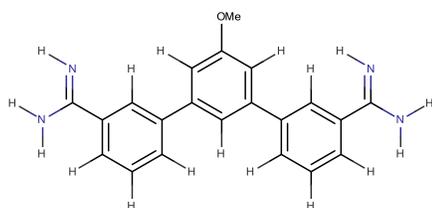
I96



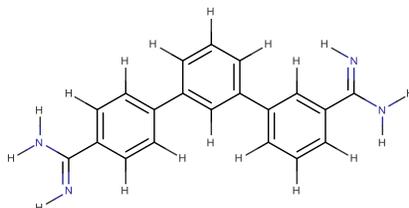
I97



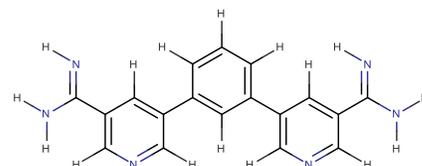
I99



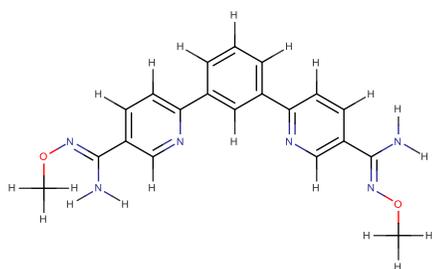
I106



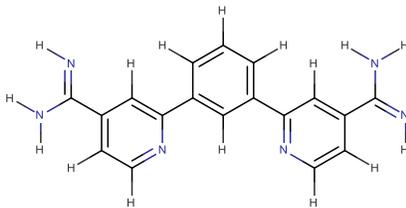
I107



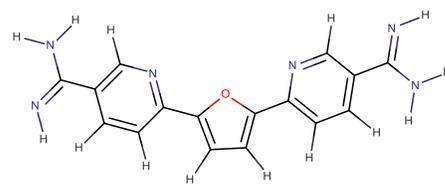
I110



I113

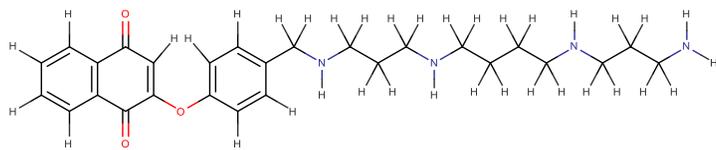


I115

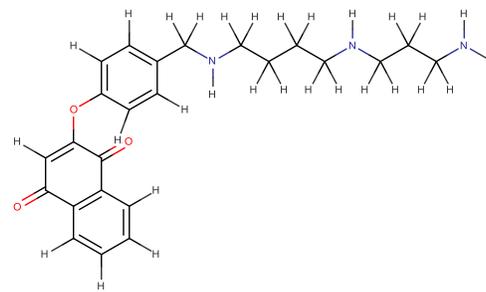


I128

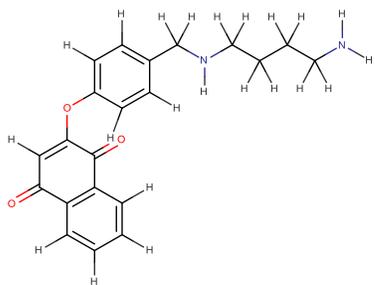
MATERIAL SUPLEMENTARIO



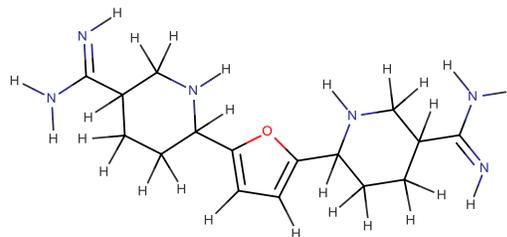
I149



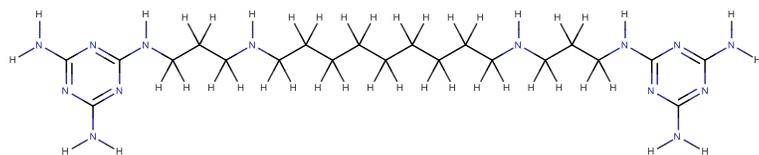
I151



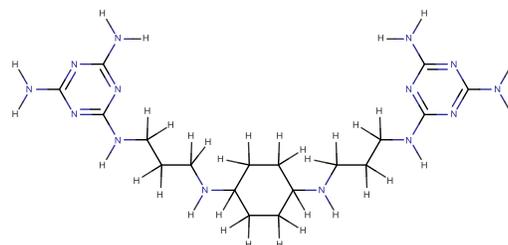
I152



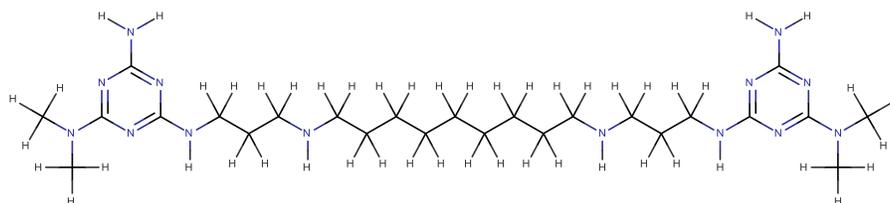
I158



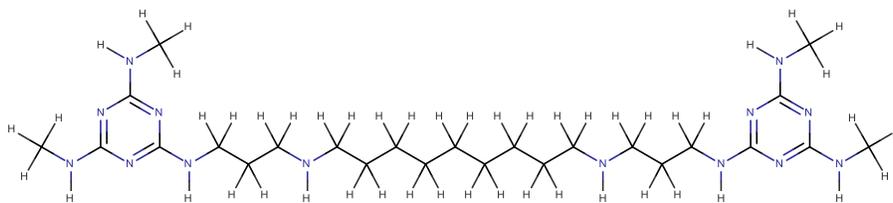
I170



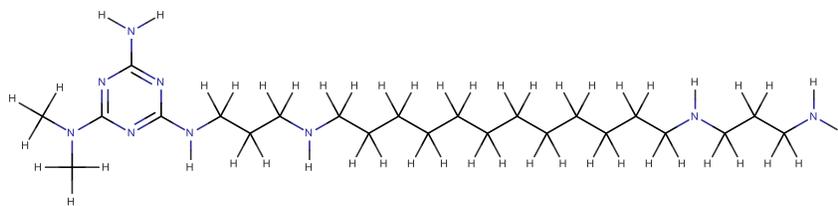
I175



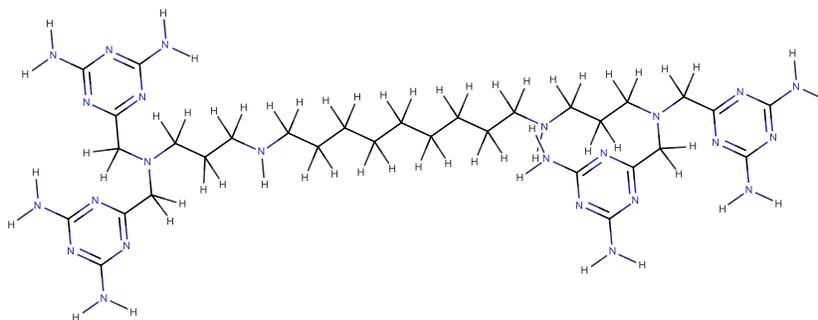
I177



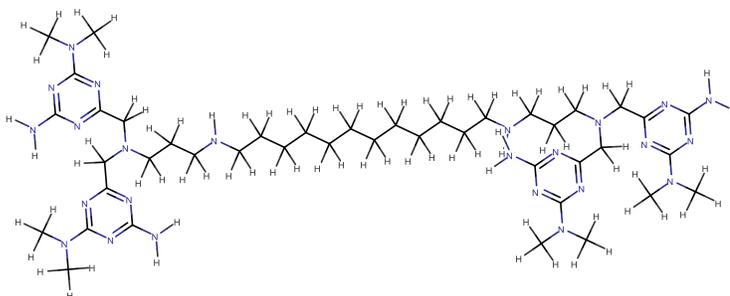
I178



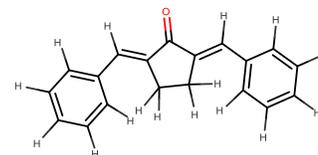
I186



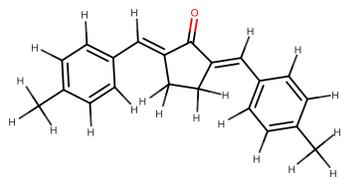
I189



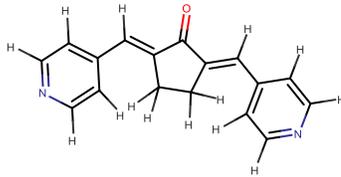
I193



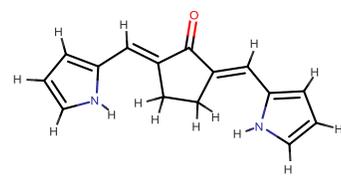
I197



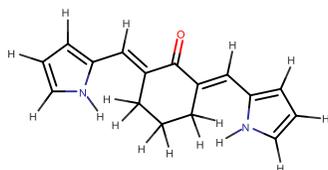
I199



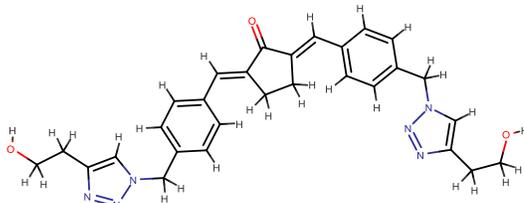
I204



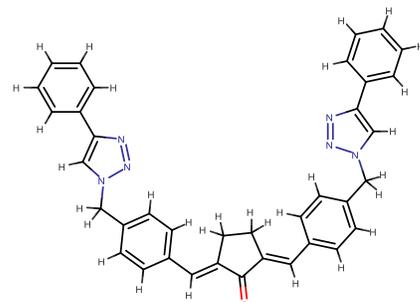
I210



I211



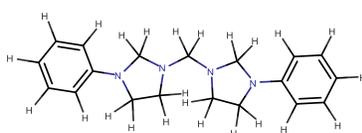
I212



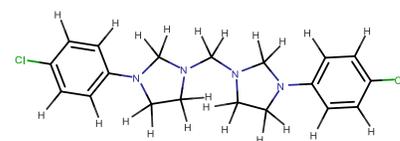
I213



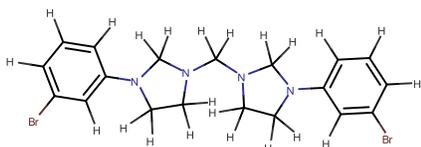
I215



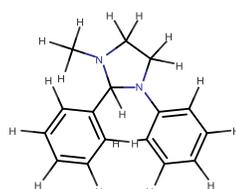
I217



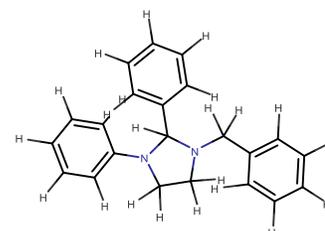
I218



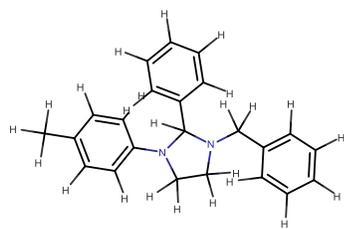
I221



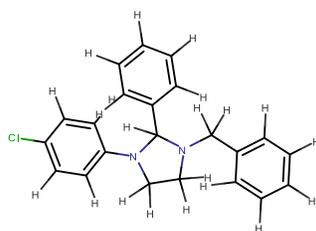
I232



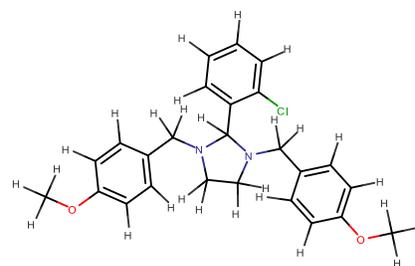
I235



I236



I237



I260

Material Suplementario 4: Lista de descriptores moleculares seleccionados por los 6 mejores modelos utilizados en la primera campaña de SV.

NOMBRE	DESCRIPCIÓN	BLOQUE
ATSC5e	Autocorrelación centrada de Broto-Moreau de quinto grado ponderado por electronegatividades de Sanderson	Autocorrelaciones 2D
B04[C-O]	Presencia/ausencia del par C - O a una distancia topológica 4	Pares de átomos 2D
B05[N-N]	Presencia/ausencia del par C - O a una distancia topológica 5	Pares de átomos 2D
B05[O-O]	Presencia/ausencia del par C - O a una distancia topológica 5	Pares de átomos 2D
B06[C-C]	Presencia/ausencia del par C - O a una distancia topológica 6	Pares de átomos 2D
B06[C-N]	Presencia/ausencia del par C - O a una distancia topológica 6	Pares de átomos 2D
B08[C-Cl]	Presencia/ausencia del par C - O a una distancia topológica 8	Pares de átomos 2D
B09[C-Cl]	Presencia/ausencia del par C - O a una distancia topológica 9	Pares de átomos 2D
B09[N-N]	Presencia/ausencia del par C - O a una distancia topológica 9	Pares de átomos 2D
B10[C-Cl]	Presencia/ausencia del par C - O a una distancia topológica 10	Pares de átomos 2D
B10[C-S]	Presencia/ausencia del par C - O a una distancia topológica 10	Pares de átomos 2D
C-007	Número de fragmentos CH ₂ X ₂ (X es un átomo electronegativo)	Fragmentos centrados en el átomo
C-019	Número de fragmentos =CRX (X es un átomo electronegativo)	Fragmentos centrados en el átomo
C-034	Número de fragmentos R--CR..X (X es un átomo electronegativo)	Fragmentos centrados en el átomo
C-038	Número de fragmentos Al-C(=X)-Al (X es un átomo electronegativo)	Fragmentos centrados en el átomo
CATS2D_03_DL	Par de átomos Donor-Lipofílico CATS2D a una distancia de 3	CATS 2D
CATS2D_04_LL	Par de átomos Lipofílico-Lipofílico CATS2D a una distancia de 04	CATS 2D
CATS2D_09_DD	Par de átomos Donor-Donor CATS2D a una distancia de 9	CATS 2D
CATS2D_09_DL	Par de átomos Donor-Lipofílico CATS2D a una distancia de 9	CATS 2D
D/Dtr05	Índice de anillo de distancia/desvío de orden 5	Descriptores de anillo
D/Dtr09	Índice de anillo de distancia/desvío de orden 9	Descriptores de anillo
Eig14_EA(ri)	Autovalor número 14 de la matriz de adyacencia ponderado por resonancia integral	Índices de adyacencia de bordes

F01[C-S]	Frecuencia del par C - S a la distancia topológica 1	Pares de átomos 2D
F04[O-O]	Frecuencia del par C - S a la distancia topológica 4	Pares de átomos 2D
F05[C-S]	Frecuencia del par C - S a la distancia topológica 5	Pares de átomos 2D
F06[N-O]	Frecuencia del par C - S a la distancia topológica 6	Pares de átomos 2D
F08[C-S]	Frecuencia del par C - S a la distancia topológica 8	Pares de átomos 2D
GATS1m	Autocorrelación de Geary de primer grado ponderado por la masa	Autocorrelaciones 2D
GATS5s	Autocorrelación de Geary de quinto grado ponderado por el estado I	Autocorrelaciones 2D
H-048	H unidos a C2(sp3)/C1(sp2)/C0(sp)	Fragmentos centrados en el átomo
J_Dz(p)	Índice de Balaban de la matriz de Barysz ponderado por polarizabilidad	Descriptorios basados en matriz 2D
JGI6	Índice topológico de carga de orden 6	Autocorrelaciones 2D
MATS2p	Autocorrelación de Moran de grado 3 ponderada por la polarizabilidad atómica	Autocorrelaciones 2D
N-072	Número de fragmentos RCO-N</>N-X=X (X= átomos electronegativos)	Fragmentos centrados en el átomo
nCar	Número de C(sp2) aromáticos	Recuentos de grupos funcionales
nCIC	Número de anillos (número ciclomático)	Descriptorios de anillo
nR=Ct	Número de C(sp2) terciarios alifáticos	Recuentos de grupos funcionales
NssNH2+	Número de átomos de tipo ssNH2+	Índices de estado E de tipo de átomo
O-060	Número de fragmentos Al-O-Ar/ Ar-O-Ar / R..O..R / R-O-C=X (X= átomos electronegativos)	Fragmentos centrados en el átomo
P_VSA_p_3	Tipo P_VSA en la polarizabilidad, compartimiento 3	Descriptorios tipo P_VSA
SaasC	Suma de estados electrónicos de aasC	Índices de estado E de tipo de átomo
SM03_EA(dm)	Momento espectral de orden 3 desde la matriz de adyacencia de borde ponderado por momento dipolar	Índices de adyacencia de bordes
SM14_EA(dm)	Momento espectral de orden 14 desde la matriz de adyacencia de borde ponderado por momento dipolar	Índices de adyacencia de bordes
SM15_EA(dm)	Momento espectral de orden 15 desde la matriz de adyacencia de borde ponderado por momento dipolar	Índices de adyacencia de bordes

SM4_Dz(Z)	Momento espectral de orden 4 de la matriz de Barysz ponderado por número atómico	Descriptores basados en matriz 2D
X5Av	Índice de conectividad de valencia medio de orden 5	Índices de conectividad

Material Suplementario 5: Breve explicación teórica de los descriptores involucrados en el mejor modelo individual de la primera campaña de SV.

Modelo 45:

$$FD = -2.96972 + 0.01213*nCar + 0.28881*C-034 + 0.79792*B10[C-Cl] + 0.66370*B05[O-O] - 1.50430*ATSC5e + 0.15263*SM4_Dz(Z) - 0.40497*NssNH2+ - 0.03304*CATS2D_04_LL + 0.06255*CATS2D_09_DL + 0.33169*F04[O-O]$$

* **nCar** pertenece al grupo “**descriptores basados en el conteo de grupos funcionales**” (*count descriptors*):

Estos son los descriptores moleculares más simples y se basan en el conteo de elementos definidos de la molécula. Estos descriptores pueden ser definidos considerando grupos químicos funcionales conocidos, los cuales presentan una reactividad específica y característica, tales como NOH; NCOOH; NNH₂; NC=O; NOCH₃; NSH; NH₂C=; NBENZ; ..., en los cuales se representan el número de grupos funcionales oxidrilos, carboxilos, aminos, carbonilos, metóxidos, tioles, metilenos y fenilos, respectivamente.

Dentro de este grupo de descriptores se encuentran los “**descriptores de subestructuras**” (*substructure descriptors*), los cuales son descriptores vectoriales que cuentan la frecuencia con la que aparecen en una molécula ciertas características estructurales definidas (grupos funcionales, fragmentos moleculares, pares de puntos de farmacóforos, pares y triángulos de átomos, superficies de triángulos, etc.) en moléculas o variables binarias que especifican su presencia/ausencia.

El descriptor **nCar** pertenece a este subgrupo e indica el número de carbonos sp² aromáticos (simplificado C(sp²)) en la molécula. En nuestro caso se observa que los compuestos activos del training set tienen una mayor cantidad de C(sp²) aromáticos con respecto a los inactivos. Por ejemplo, 62 de los 92 compuestos activos tienen más de 15 C(sp²) aromáticos, mientras que solo 34 de los compuestos inactivos presentaron esta característica. Además, sólo uno de los inactivos tiene más de 20 C aromáticos, mientras que 24 de los activos superan este valor (Todeschini & Consonni 2009).

* **C-034** pertenece al grupo “**descriptores basados en fragmentos centrados en el átomo**” (*atom-centred fragments*):

Estos son descriptores de subestructuras de rango corto que cuentan el número de ocurrencias de ciertos fragmentos en la molécula. El fragmento es definido en términos del elemento químico en los que se centra el mismo, la naturaleza de los átomos diferentes de hidrógeno a los que se enlaza y el tipo de enlace químico involucrado. Se consideran los átomos enlazados con hasta cuatro átomos distintos de hidrógeno. Por lo tanto, los fragmentos más grandes contienen cinco átomos y cuatro enlaces químicos.

El descriptor **C-034** indica el número de fragmentos R--CR..X, donde R es cualquier átomo unido a través del carbono, "--" indica un enlace aromático como en el benceno o un enlace deslocalizado como el enlace N-O en un grupo nitro, ".." indica un enlace aromático y X indica un átomo electronegativo (que en este caso podrán ser O, N, S). En definitiva, el descriptor indica la presencia de un carbono que forma parte de un heterociclo aromático. El 40% de los activos

presentaron al menos un fragmento R--CR..X, mientras que sólo un 9% de los inactivos presentaron al menos uno de estos fragmentos.

* **B10[C-Cl]**, **B05[O-O]** y **F04[O-O]** pertenecen al grupo “**Pares de átomos 2D**” (**2D atom pairs**):

Estos son descriptores de subestructura definidos en términos de cualquier par de átomos y tipo de enlace que los conecta. Un par de átomos está compuesto por dos átomos diferentes al hidrógeno y una separación interatómica:



La separación se refiere a la distancia topológica entre ellos (número de enlaces entre los átomos por el camino más corto). Estos descriptores son también llamados pares de átomos topológicos.

Un tipo particular de este grupo son los “**descriptores de recuento de distancia**” (**SE vectors**), los cuales son una implementación particular de los pares de átomos topológicos propuestos por Clerc y Terkovic en 1990 (Clerc & Terkovic 1990). Estos codifican información sobre la frecuencia de ocurrencia de cualquier combinación de dos tipos de átomos separados por la distancia topológica aludida.

El descriptor **B10[C-Cl]** indica la presencia/ausencia de un par de átomos C - Cl a una distancia topológica de 10. Encontramos que en 21 de los 92 activos hay presencia de un par de átomos C-Cl a distancia topológica de 10, y sólo 6 de los inactivos presentaban esta característica.

El descriptor **B05[O-O]** indica la presencia/ausencia de un par de átomos de O - O a una distancia topológica 5. En este caso, 11 activos y sólo 4 inactivos presentaron el par de átomos O-O a la distancia considerada.

El descriptor **F04[O-O]** pertenece al subgrupo de descriptores que miden la frecuencia con la que, en una dada molécula, aparece un par de átomos definidos (en este caso, O - O) a una distancia topológica de 4. Aproximadamente un 15% de los compuestos activos tienen al menos un par de átomos O-O separados por una distancia topológica de 4, mientras que sólo el 2% de los compuestos inactivos presentan esta característica.

* **ATSC5e** pertenece a los “**descriptores basados en la autocorrelación 2D de Broto-Moreau**”:

La autocorrelación de Broto-Moreau es una autocorrelación espacial definida en un grafo molecular G como (Moreau and Broto 1980a; Moreau and Broto 1980b), que se describe con la **Ecuación MS.1**:

$$ATS_k = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A w_i \cdot w_j \cdot \delta(d_{ij}; k)$$

Donde w es una propiedad atómica, A es el número de átomos en una molécula, k es el llamado *lag* (podría traducirse, tal vez, como desfasaje) y d_{ij} es la distancia topológica entre los átomos i -ésimo y j -ésimo; $\delta(d_{ij}; k)$ es una función delta de Kronecker igual a 1 si $d_{ij} = k$, y 0 de otra forma. La autocorrelación ATS_0 es un caso especial y se calcula con la **Ecuación MS.2**:

$$ATS_0 = \sum_{i=1}^A w_i^2$$

esto es, la suma de los cuadrados de las propiedades atómicas. Las propiedades atómicas típicamente consideradas son las masas moleculares, polarizabilidades, cargas y electronegatividades. Puede observarse, para el caso general, que la autocorrelación tenderá a valores altos: a) para moléculas grandes, donde pueden encontrarse mayor cantidad de átomos separados por distancias k y; b) en el caso de moléculas orgánicas, cuando aparezcan heteroátomos cargados o caracterizados por alta polarizabilidad, electronegatividad o peso atómico, según corresponda, a la distancia topológica definida por el *lag*.

Los descriptores de autocorrelación espacial promedio son obtenidos dividiendo cada término por el correspondiente número de contribuciones, evitando así cualquier dependencia con el tamaño molecular, **Ecuación MS.3:**

$$\overline{ATS}_k = \frac{1}{2\Delta_k} \cdot \sum_{i=1}^A \sum_{j=1}^A w_i \cdot w_j \cdot \delta(d_{ij}; k)$$

Donde Δ_k es la suma de los deltas de Kronecker, esto es, el número total de pares de vértices a la distancia igual a k (Wagener et al. 1995).

ATSC5e indica la autocorrelación centrada de Broto-Moreau de quinto grado ponderado por electronegatividades de Sanderson. Sólo 9 de los activos obtuvieron un valor mayor a 0,5, en tanto ése número ascendió a 28 para los inactivos, por lo que mayores valores en esta autocorrelación afectan negativamente a la actividad y el coeficiente de regresión es negativo en la semicorrelación correspondiente.

* **SM4_Dz(Z)** pertenece al grupo “**descriptores basados en la matriz 2D (2D matrix-based descriptors)**”:

Los momentos espectrales de la matriz M , denotados $u^k(M; w)$ y calculados con un esquema ponderado w , son definidos por la **Ecuación MS.4:**

$$u^k(M; w) = \sum_{i=1}^n \lambda_i^k = \sum_{i=1}^n [M^k]_{ii}$$

Donde $k=1, \dots, n$ es el orden del momento espectral, λ_i^k el autovalor de la matriz M , y la última suma pasa por los elementos diagonales del k -ésimo poder de la matriz M (Todeschini & Consonni 2009).

SM4_Dz(Z) indica el momento espectral de orden 4 de la matriz de Barysz ponderado por número atómico.

La matriz de distancia de Barysz 2D es una matriz de distancia ponderada simétrica que cuenta contemporáneamente la presencia de heteroátomos y enlaces múltiples en la molécula M ; es definida por la **Ecuación MS.5** (Barysz et al. 1983):

$$[{}^Z D]_{ij} = \begin{cases} d_{ij}(Z, \pi^*) & \text{si } i \neq j \\ 1 - \frac{Z_C}{Z_i} & \text{si } i = j \end{cases} \quad d_{ij}(Z, \pi^*) = \sum_{b=1}^{d_{ij}} \left(\frac{1}{\pi_b^*} \cdot \frac{Z_C^2}{Z_{b(1)} \cdot Z_{b(2)}} \right)$$

Donde Z_c es el número atómico del átomo de carbono, Z_i es el número atómico del i -ésimo átomo, y π_b^* es el orden de enlace; $d_{ij}(Z, \pi^*)$ es una distancia topológica ponderada calculada por la suma los pesos de las aristas sobre todos los enlaces d_{ij} involucrados en el camino más corto entre los vértices v_i y v_j , siendo d_{ij} la distancia topológica; los subíndices $b(1)$ y $b(2)$ representan los dos vértices incidentes al enlace b considerado.

* **NssNH2+** pertenece al grupo “índices estado-E basados en los tipos de átomos” (*atom-type E-state indices*):

NssNH2+ indica el número de subestructuras del tipo $\text{-NH}_2^- [+1]$. Ningún compuesto activo del training set presenta átomos de este tipo, mientras que 8 compuestos inactivos tienen este tipo de átomos.

* **CATS2D_04_LL** y **CATS2D_09_DL** pertenecen al grupo “CATS 2D” (*Chemically Advanced Template Search*):

Estos son descriptores que remiten al concepto de farmacóforo, sólo que en lugar de contemplar la distancia geométrica entre dos tipos de puntos farmacofóricos de interés se considera la distancia topológica (independiente de la conformación). El descriptor señala el número de veces que se encuentra en la molécula el par de puntos farmacofóricos de interés a la distancia considerada (Schneider et al. 1999; Fechner et al. 2003).

Los cinco posibles puntos farmacofóricos definidos son los más habitualmente incluidos en modelos farmacofóricos: donador de puentes de hidrogeno (D), aceptor de puentes de hidrogeno (A), ionizables o cargados positivamente (P), ionizable o cargado negativamente (N), y lipofílico (L). Si un átomo no pertenece a alguno de los 5 tipos, no es considerado. Más aún, un átomo puede ser asignado a uno o más tipos (en particular, átomos que pueden comportarse como donores o aceptores de puentes de H).

Para cada molécula, el número de ocurrencias de los 15 pares PP posibles (DD, DA, DP, DN, DL, AA, AP, AN, AL, PP, PN, PL, NN, NL, LL) es determinado y luego asociado con el número de enlaces intervinientes entre los dos puntos considerados, en donde se utiliza la ruta más corta.

CATS2D_04_LL indica el número de pares de átomos Lipofílico-Lipofílico CATS2D a una distancia topológica de 4. Con respecto a los activos, sólo 8 de estos tienen al menos 18 pares de átomos de este tipo, mientras que 14 inactivos presentan esta característica.

CATS2D_09_DL indica el número de pares de átomos Donador-Lipofílico CATS2D a una distancia topológica de 9. 37% de los compuestos activos poseen al menos 6 pares de átomos de este tipo, mientras que solo el 16% de los inactivos presentan esta característica.

Material Suplementario 6: Scripts utilizados en la segunda campaña de SV.

Script utilizado para la aleatorización y partición del conjunto de descriptores en 1000 subconjuntos:

```
conjuntos.descriptores <- function (archivo="trainingset.csv", conjuntos=1000,
descrip.conjunto=200, semilla=125)
{is.installed <- function(mypkg) { is.element(mypkg, installed.packages()[,1]) }
if (is.installed("data.table") == FALSE) {install.packages("data.table")}
if (is.installed("caret") == FALSE) {install.packages("caret")}
library(data.table)
library(caret)
set.seed(semilla)
df <- as.data.frame(fread(input = archivo, check.names = TRUE))
df <- df[, apply(df, 2, function(x) !any(is.na(x)))]
df <- df[, -nearZeroVar(df)]
n <- conjuntos
j <- c(1:ncol(df))
k <- j[-c(1,2)]
dflist <- list()
for(i in 1:n) {dflist[[i]] <- df[, c(2,sample(k,descrip.conjunto))]}
dflist}
```

Script utilizado para la obtención de los 1000 modelos lineales

```
forward.stepwise.testF.lm <- function(tabla.conjunto , punto.corte=0.05 , steps=6 ,
punto.corte.vif = 2)
{is.installed <- function(mypkg) { is.element(mypkg, installed.packages()[,1])}
if ( is.installed("car") == FALSE) {install.packages("car")}
library(car)
df <- as.data.frame(tabla.conjunto)
null <- lm(clase ~1,data = df, na.action = na.omit)
full <- lm(clase~.,data = df, na.action = na.omit)
repeat {paso <- add1(null,scope=formula(full),test="F", na.action = na.omit)
menor.probabilidad <- which.min(paso$`Pr(>F)` )
elegido <- noquote(row.names(paso)[menor.probabilidad])
formulacion <- paste(" .~. +", elegido)
if ( paso$`Pr(>F)`[menor.probabilidad] >= punto.corte ) {break}
null <- update(null, formula. = formulacion)
if (length(null$coefficients) == 3){test.vif <- data.frame(vif(null))
if (test.vif[nrow(test.vif),1] >= punto.corte.vif) { formulacion2 <- paste(" .~.
",paste(noquote(rownames(test.vif)[nrow(test.vif)]),collapse = " - "))
null<-update(null, formula. = formulacion2)
full <- update(full, formula. = formulacion2) ### elimino las variables del modelo full que no
cumplen con el VIF para que no intenten ingresar otra vez al modelo} }
if (length(null$coefficients) > 3) {
test.vif <- data.frame(vif(null))
if (any(test.vif[,1] >= punto.corte.vif)) {
```

```
formulacion2 <- paste(".~. -",paste(noquote(rownames(test.vif)[nrow(test.vif)]),collapse = " -
"))
null<-update(null, formula. = formulacion2)
full <- update(full, formula. = formulacion2) }}
test <- drop1(null, test = "F")
if (any(test$`Pr(>F)`>= punto.corte, na.rm = TRUE)) {
formulacion3<- paste(".~. -",paste(noquote(rownames(test)[which(test$`Pr(>F)`>=
punto.corte)]),collapse = " - "))
null<-update(null, formula. = formulacion3)
full<-update(full, formula. = formulacion3)}
if (length(null$coefficients) >= steps + 1) {break}
if (length(full$coefficients) == length(null$coefficients)) {break}}
result <-null
result}
```

Material Suplementario 7: Breve explicación teórica de los descriptores involucrados en el mejor modelo individual de la segunda campaña de SV.

Modelo 348:

$$FD = 3.12777 + 0.03474 * F06[C-CI] + 0.20805 * S-107 - 0.04291 * F05[N-N] + 0.39611 * C-039 - 0.34582 * SM5_B(s) + 6.25705 * Eta_epsi_A + 0.53013 * nSO2OH - 1.28338 * SpMax_H2 + 0.44827 * Eig04_AEA(ri) - 1.73390 * ATSC1e + 0.05975 * CATS2D_09_PL + 0.02805 * SaaO$$

* **F06[C-C]** y **F05[N-N]** corresponden al subgrupo “**pares de átomos 2D**” descrito en el Material Suplementario 4. El primero indica la frecuencia del par de átomos C - C a una distancia topológica de 6. Aproximadamente un 50% de los compuestos activos tienen al menos un par de átomos O-O separados por una distancia topológica de 6, mientras que sólo el 23% de los compuestos inactivos presentan esta característica.

El segundo representa la frecuencia de un par de átomos N - N a distancia topológica 5. Sólo uno de los compuestos activos presenta este par de átomos a la distancia topológica, mientras que 18 de los compuestos inactivos presentan al menos uno de uno de estos pares de átomos a la distancia seleccionada.

* **S-107** y **C-039** pertenecen al grupo de “**descriptores basados en fragmentos centrados en el átomo**” descrito en el Material Suplementario 4. El primero representa la cantidad de grupos R₂S/RS-SR, donde R es cualquier átomo unido a través del carbono y S representa al átomo de azufre. 9 de los compuestos activos del conjunto de entrenamiento presentan al menos uno de estos grupos, mientras que sólo 2 de los inactivos poseen este grupo.

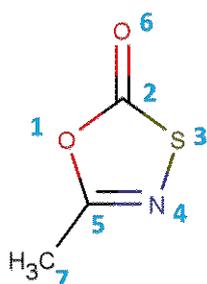
El segundo indica la frecuencia de grupos Ar-C(=X)-R, donde Ar representa un grupo aromático, R es cualquier átomo unido al carbono C y “=X” indica un átomo electronegativo unido por un doble enlace. 6 de los compuestos activos presentan al menos uno de estos grupos, mientras que sólo 2 inactivos presentan la característica.

* **SM5_B(s)** pertenece al subgrupo de “**descriptores basados en la matriz 2D**” e indica el momento espectral de orden 5 de la matriz Burden ponderado por el estado-I. De los compuestos activos solo hay 8 que presentan valores mayores a 10,7 mientras que hay 18 compuestos inactivos que superan ese valor.

La matriz de Burden es una matriz de adyacencia ponderada que se define por la **Ecuación MS.6** (Burden 1989):

$$[B]_{ij} = \begin{cases} \pi_{ij}^* \cdot 10^{-1} & \text{si } (i, j) \in E(G) \\ Z_i & \text{si } i = j \\ 0.001 & \text{si } (i, j) \notin E(G) \end{cases}$$

Donde los elementos de la diagonal son los números atómicos Z_i , los elementos fuera de la diagonal que representan dos átomos enlazados v_i y v_j , son la función de orden de enlace convencional π_{ij}^* , esto es, 0.1, 0.2, 0.3, y 0.15 para un enlace simple, doble, triple y aromático respectivamente. Todos los otros elementos de la matriz son seteados a 0.001. Además, los elementos de la matriz que corresponden a los enlaces terminales son ponderados por 0.01. En la **Figura MS.1** se encuentra un ejemplo de la obtención de la matriz de Burden.



$$B =$$

Átomo	1	2	3	4	5	6	7
1	8	0.1	0.001	0.001	0.1	0.001	0.001
2	0.1	6	0.1	0.001	0.001	0.21	0.001
3	0.001	0.1	16	0.1	0.001	0.001	0.001
4	0.001	0.001	0.1	7	0.2	0.001	0.001
5	0.1	0.001	0.001	0.2	6	0.001	0.011
6	0.001	0.21	0.001	0.001	0.001	8	0.001
7	0.001	0.001	0.001	0.001	0.11	0.001	6

Figura MS.1: Matriz de Burden matrix para el compuesto 5-metil-1,3,4-oxatiazol-2-ona.

El estado intrínseco (estado-I) del átomo i -ésimo es un invariante de vértice local calculado a partir del grafo molecular mediante la **Ecuación MS.7** (Kier & Hall 1990; Kier & Hall 1999):

$$I_i = \frac{(2/L_i)^2 \cdot \delta_i^v + 1}{\delta_i}$$

Donde L_i es el número cuántico principal asociado al átomo del vértice i , δ_i^v es el número de electrones de valencia y δ_i es el número de electrones sigma del i -ésimo átomo en la estructura molecular (sin considerar hidrógenos).

* **Eta_epsi_A** pertenece a este grupo “**Índices ETA**” (**Extended Topochemical Atom índices**):

Son invariantes de grafos y vértices locales (Roy & Ghosh 2003), definidos en el marco de la teoría del entorno Valence Electron Mobile (entorno VEM), de acuerdo a la cual se considera que un vértice en el grafo molecular sin hidrógenos consiste en un núcleo y un ambiente de valencia electrónica.

El recuento de núcleos α_i es un invariante de vértice local calculado como, **Ecuación MS.8**:

$$\alpha_i = \frac{Z_i - Z_i^v}{Z_i^v} \cdot \frac{1}{L_i - 1}$$

Donde Z_i , Z_i^v , y L son los números atómicos, el número de electrones de valencia, y el número cuántico principal, respectivamente. La suma de los valores α de los átomos distintos al hidrogeno es un descriptor molecular simple relacionado con el volumen molecular.

Al combinar el recuento de núcleos α de un átomo con su número de electrones de valencia Z_i^v , la medida de electronegatividad ETA para el i -ésimo átomo se define por la **Ecuación MS.9**:

$$\varepsilon = -\alpha_i + 0,3 \cdot Z_i^v$$

El descriptor **Eta_epsi_A** corresponde a la medida de la electronegatividad ETA promedio. El 87% de los compuestos activos tiene electronegatividad ETA promedio por encima de 0.55 mientras que sólo 67% de los inactivos superan ese valor.

* El descriptor **nSO2OH** pertenece al subgrupo de **descriptores basados en el conteo de grupos funcionales** que fue previamente descrito y corresponde al número de ácidos (tio-/ditio-) sulfónicos en la molécula.

* **SpMax_H2** pertenece al subgrupo de **descriptores basados en la matriz 2D** e indica el primer autovalor de la matriz de distancia cuadrada recíproca. 59% de los compuestos activos presentan autovalores por encima de 0,4 mientras que sólo el 28% de los inactivos superan ese

valor. Como el valor en la ecuación es negativo, indica que son deseables para la actividad menores valores de este descriptor.

Eig04_AEA(ri) pertenece al grupo “**índices de adyacencia de borde**” (**Edge adjacency indices**) y denota el cuarto autovalor de la matriz de adyacencia de borde aumentada ponderada por la integral de resonancia. 96% de los compuestos activos presentan valores de este descriptor por encima de 3,04, mientras que el 73% de los inactivos superan ese valor.

La **matriz de adyacencia**, denotada E , también llamada matriz de enlace, codifica información acerca de la conectividad mediante aristas del grafo molecular, **Ecuación MS.10** (Todeschini & Consonni 2009):

$$[E]_{ij} = \begin{cases} 1 & \text{si } (i,j) \text{ son enlaces adyacentes} \\ 0 & \text{de otra forma} \end{cases}$$

Es una matriz simétrica de dimensiones $B \times B$, donde B es el número de enlaces, y es usualmente derivada de un grafo molecular sin hidrógenos.

El **grado de la arista** ε_i es calculado a partir de la matriz de adyacencia mediante la **Ecuación MS.11**:

$$\varepsilon_i = \sum_{j=1}^B [E]_{ij}$$

La **matriz ponderada de adyacencia de ejes** wE es derivada de un grafo molecular cuyas aristas se ponderan por la aplicación de cualquier esquema de ponderación w_i , el cual codifica la información acerca de cada enlace en la molécula. Este esquema de ponderación puede estar basado en características de los enlaces, tales como distancias de enlace o dipolos de enlace, o en propiedades de los átomos involucrados en cada enlace, como masa atómica, electronegatividad atómica, cargas atómicas, etc.

En lugar de numerar los vértices (como en el caso de la matriz adyacencia), en la **matriz de ponderada de adyacencia de ejes** wE se numeran los ejes y se calculan los elementos de la matriz de acuerdo al siguiente esquema:

$${}^w[E]_{ij} = \begin{cases} w_j & \text{si } (i,j) \text{ son enlaces adyacentes} \\ 0 & \text{de otra forma} \end{cases}$$

donde los elementos de la diagonal de la matriz para un grafo ponderado son 0 para aristas no adyacentes, mientras que, si dos aristas e_i y e_j son adyacentes, la entrada $i - j$ es definida por el peso w_j de la arista j y la entrada simétrica $j - i$ es definida por la ponderación w_i de la arista i , dando como resultado una matriz asimétrica.

La **matriz de adyacencia de ejes ponderada por el orden de enlace** ${}^\pi E$ es obtenida mediante grafos ponderados cuyos ejes son ponderados por el orden de enlace π (Estrada & Montero 1993):

$${}^\pi[E]_{ij} = \begin{cases} \pi_j & \text{si } (i,j) \text{ son enlaces adyacentes} \\ 0 & \text{de otra forma} \end{cases}$$

Por analogía con la matriz adyacencia de ejes ponderada por el orden de enlace, se ha propuesto una matriz de adyacencia de ejes ponderada por la resonancia kE , reemplazando los órdenes de enlace con parámetros k_{C-X} usados en la matriz de Huckel y relacionados a la matriz integral de resonancia β_{C-X} de los enlaces entre el heteroátomo X y el átomo de carbono por la **Ecuación MS.12**:

$$\beta_{C-X} = k_{C-X} \cdot \beta_{C-C}$$

donde β_{C-C} es la resonancia de los enlaces carbono-carbono. Pueden encontrarse en literatura varios valores de k_{C-X} .

* **ATSC1e** pertenece al grupo de **“descriptores basados en la autocorrelación 2D de Broto-Moreau”** y corresponde a la autocorrelación centrada de Broto-Moreau de primer grado ponderada por la electronegatividad de Sanderson. Solamente 16 de los compuestos activos presentan valores de este descriptor por encima de 0,1 mientras que 31 de los inactivos superan ese valor.

* **CATS2D_09_PL** pertenece al subgrupo **CATS 2D** e indica el CATS2D Positivo-Lipofílico a una distancia topológica de 9. Hay 11 compuestos activos y 7 compuestos inactivos que tienen más de 2 pares PL a la distancia seleccionada.

* **SaaO** pertenece al subgrupo **índices estado-E basados en los tipos de átomos” (Atom-type E-state índices)** e indica la suma de los estados electrotopológicos del grupo aOa, donde “a” representa un anillo aromático. 24 de los compuestos activos presentaron valores mayores a 0 para este descriptor, mientras que sólo 9 de los inactivos presentaron esta característica.

REFERENCIAS

- Barysz, M. et al, 1983. On the distance matrix of molecules containing heteroatoms. *Chemical Applications of Topology and Graph Theory*, Elsevier, pp. 222–230.
- Burden, F.R., 1989. Molecular identification number for substructure searches. *Journal of Chemical Information and Computer Sciences*, 29, pp.225–227.
- Clerc, J.T. & Terkovic, A.L., 1990. Versatile topological structure descriptor for quantitative structure/property studies. *Analytica Chimica Acta*, 235, pp.93–102.
- Estrada, E. & Montero, L.A., 1993. Bond order weighted graphs in molecules as structure property indices. *Molecular Engineering*, 2, pp.363–373
- Fechner, U., et al., 2003. Comparison of correlation vector methods for ligand-based similarity searching. *Journal of Computer-Aided Molecular Design*, 17, pp.687–698.
- Kier, L.B. & Hall, L.H., 1990. An electrotopological-state index for atoms in molecules. *Pharmaceutical Research*, 7, pp.801–807.
- Kier, L.B. & Hall, L.H., 1999. Molecular Structure Description. *The Electrotopological State*, Academic Press, p. 246.
- Moreau, G. & Broto, P., 1980a. Autocorrelation of molecular structures. Application to SAR studies. *Nouveau journal de chimie*, 4, pp.757–764.
- Moreau, G. & Broto, P., 1980b. The autocorrelation of a topological structure: a new molecular descriptor. *Nouveau journal de chimie*, 4, pp.359–360.
- Naylor, S. & Schonfeld, J.M., 2014. Therapeutic drug repurposing, repositioning and rescue - Part I: Overview. *Drug Discovery World*, 16(1), pp.49–62.
- Roy, K. & Ghosh, G., 2003- Introduction of extended topochemical atom (ETA) indices in the valence electron mobile (VEM) environment as tools for QSAR/QSPR studies. *Internet Electron. Journal of Molecular Design*, 2, pp.599–620.
- Schneider, G., et al., 1999. “Scaffold-hopping” by topological pharmacophore search: a contribution to virtual screening. *Angewandte Chemie International Edition*, 38, pp.2894–2895.
- Todeschini, R. & Consonni, V., 2009. *Molecular Descriptors for Chemoinformatics*. Wiley-VCH.
- Wagener, M. et al, 1995. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic Ah receptor activity by neural networks. *Journal of the American Chemical Society*, 117, pp.7769–7775.



SOLO COMPLETAR SI LA NOTA ES "APROBADO CON MENCIÓN ESPECIAL"

FECHA: 21/12/2018

TESISTA: ALBERCA LUCAS NICOLAS

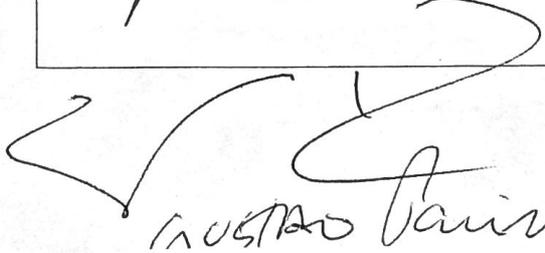
TESIS:

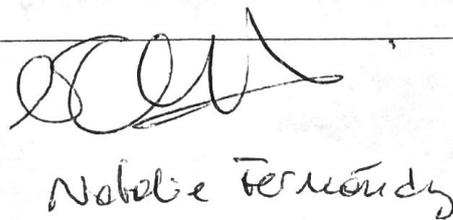
ACTA: 1884

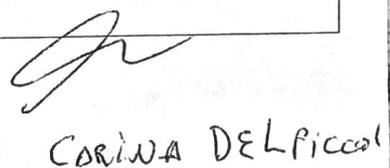
LIBRO: 9

DICTAMEN

El trabajo doctoral del Dr Alberca se basó en la identificación y caracterización de nuevos agentes teropentales Anti-chagásicos. Para tal propósito se realizaron distintos protocolos de cubos virtuales utilizando modelos predictivos. El trabajo muestra una gran rigurosidad metodológica y analítica, de tal modo el número de cubos y alternancias evaluadas para lograr sus objetivos. La presentación oral fue clara y el presentador demostró gran conocimiento del tema, contestando con soltura las preguntas del jurado. Por tales motivos, este jurado dictamina unánimemente que este trabajo doctoral es aprobado con Mención Especial.


GUSTAVO LARINI


Noélie Fernández


CORINA DELIPPICA