

Universidad Nacional de La Plata
Facultad de Ciencias Exactas
Departamento de Ciencias Biológicas
Cátedra de Química Medicinal



2014

TRABAJO DE TESIS DOCTORAL

Búsqueda racional de nuevos fármacos antichagásicos inhibidores de la cruzipaína

Lic. Carolina L. Bellera

Director: Dr. Luis E. Bruno-Blanch

Co-Director: Dr. Alan Talevi



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral

**BÚSQUEDA RACIONAL DE NUEVOS FÁRMACOS
ANTICHAGÁSICOS INHIBIDORES DE LA
CRUZIPAÍNA**

Lic. Carolina L. Bellera

Director: Dr. Luis E. Bruno-Blanch / Co-Director: Dr. Alan Talevi

2014

“No estalla como las bombas ni suena como los tiros. Como el hambre, mata callando. Como el hambre, mata a los callados: a los que viven condenados al silencio y mueren condenados al olvido. Tragedia que no suena, enfermos que no pagan, enfermedad que no vende. El mal de Chagas no es negocio que atraiga a la industria farmacéutica, ni es tema que interese a los políticos ni a los periodistas. Elige a sus víctimas en el pobrerío. Las muerde y lentamente, poquito a poco, va acabando con ellas. Sus víctimas no tienen derechos, ni dinero para comprar los derechos que no tienen. Ni siquiera tienen el derecho de saber de qué mueren.”

Informe Clínico, Eduardo Galeano

A mis abuelas.

Agradecimientos

A mi director **Dr. Luis Bruno-Blanch**, por su confianza, sabiduría, sus consejos académicos, científicos y de la vida.

A mi co-director **Dr. Alan Talevi**, por su paciencia, por ser guía, porque sin él seguramente este logro no hubiera sido posible, por la amistad de tantos años.

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

A la **Dra. Carolina Carrillo, Dr. Carlos Labriola, Lic. Darío Balcazar y Dra. Patricia Romano**, por su inmensa ayuda en los ensayos experimentales.

Al **Dr. Jorge Gálvez**, por abrirme las puertas de su laboratorio en Valencia y por su confianza.

Al **Dr. Ramón García-Domenech, Dr. Lionello Pogliani, Farm. María Gálvez-Llompart y Farm. Ricardo Zanni**, por hacer que mi estadía en la Unidad de Diseño y Conectividad Molecular de la Facultad de Farmacia de Valencia sea inolvidable.

A mi **familia, mis padres, hermanos, abuela y cuñada**, por ser mi cable a tierra.

A mis **compañeros del LIDeB**, Andrea, Mauricio, Luciana, Pablo, Melisa, María Luisa, Laureano, Roque, Manuel, Lucas, Malala, Francisco, Esperanza, Julián, Emiliano, Rocío, Sofía, Facundo, Ruben, Yoda y a los que ya no están (Gisela, Valen, Juan, y Sung), por tantas horas compartidas, por las cenas, los after Lab, los group meetings y porque sin todos ustedes ir a trabajar todos los días con una sonrisa no sería posible.

A mis **compañeros de docencia de Biofarmacia y Farmacotecnia I**, en especial a Emilia y Francisco por bancarme estos últimos meses.

A mis **amigas y amigos**, por estar siempre.

A mis **camaradas**, Agus, Emi y Vicky, por enseñarme a mirar desde otra perspectiva a la Salud, la Educación y la Comunicación Social.

A las **comunidades de los Parajes Callejón Bajada y Yanacón Bajada** de la provincia de Santiago del Estero, por ser el puntapié inicial de mi compromiso y movilización frente al Chagas.

A mis **compañeros de ConCiencia y Pueblo (CCyP)**, por trabajar por una educación universitaria de calidad y una facultad al servicio del pueblo.

Al **Ninja**, por ser mi par.

ÍNDICE

Lista de abreviaturas	1
Objetivos	3
Capítulo 1: Introducción	5
1.1. Enfermedad de Chagas	7
1.1.1. Biología del parásito, vías de transmisión y fases clínicas de la enfermedad de Chagas	9
1.2. Farmacoterapia actual para la enfermedad de Chagas	16
1.2.1. Perspectivas sobre nuevas Quimioterapias para la enfermedad de Chagas	18
1.3. Cruzipaína como blanco terapéutico de la enfermedad de Chagas.....	22
1.3.1. Aspectos generales, estructurales y funcionales.....	22
1.3.2. Inhibidores de Cruzipaína.....	25
Referencias Capítulo 1	29
Capítulo 2: Cribado Virtual	35
2.1. Definición.....	37
2.2. Clasificación de las técnicas de Cribado virtual	40
2.2.1. Metodologías de CV basadas en el receptor.....	40
2.2.2. Técnicas de CV basadas en el ligando.....	41
2.3. Aproximación empleada en el trabajo de Tesis	47
Referencias Capítulo 2	49
Capítulo 3: Reposicionamiento de fármacos	53
3.1. Descubrimiento y desarrollo de fármacos	55
3.1.1. Estrategias generales en el descubrimiento de fármacos	55

3.1.2. Costos y fracasos en el desarrollo de fármacos	59
3.2. Cambio de paradigma en la búsqueda de nuevos fármacos	62
3.3. Reposicionamiento de Fármacos.....	64
3.3.1. Reposicionamiento de fármacos asistido por computadora.....	67
3.3.2. Reposicionamiento de fármacos asistido por computadoras para enfermedades olvidadas	74
Referencias Capítulo 3	79
Capítulo 4: Metodología	89
4.1. Generación de modelos QSAR.....	91
4.1.1. Construcción de la base de datos y partición en conjuntos de entrenamiento y de prueba.....	91
4.1.2. Características del Conjunto de Entrenamiento.....	98
4.1.3. Cálculo de descriptores moleculares	99
4.1.4. Construcción y Validación del modelo	99
4.2. Campaña Simulada de Cribado Virtual.....	107
4.3. Cribado (o Tamizado) Virtual	107
4.4. Dominio de Aplicabilidad	108
4.5. Evaluación biológica	109
4.5.1. Efecto inhibitorio sobre Cruzipaína	109
4.5.2. Efecto inhibitorio sobre la proliferación de epimastigotes de <i>T. cruzi</i>	109
4.5.3. Ensayos de invasión de tripomastigotes y proliferación a amastigotes de <i>T. cruzi</i>	110
4.5.4. Ensayo Alamar Blue.....	111
4.5.5. Ensayo in vivo. Modelo de infección aguda en ratones.	112
Referencias Capítulo 4	115
Capítulo 5: Resultados	119
5.1. Compilación de la base de datos y partición en conjunto de entrenamiento y conjunto de prueba	121

5.2. Primera Campaña de Cribado Virtual.....	122
5.2.1. Generación y Validación del modelo 1	122
5.2.2. Cribado Virtual	124
5.2.3. Evaluación Biológica	125
5.3. Segunda Campaña de Cribado Virtual	130
5.3.1. Generación y Validación del modelo 2	130
5.3.2. Cribado Virtual	132
5.3.3. Evaluación Biológica	133
5.4. Tercera Campaña de Cribado Virtual.....	135
5.4.1. Generación y Validación de modelo 3	135
5.4.2. Cribado Virtual	137
5.4.3. Evaluación Biológica	139
5.5. Ensayo <i>in vivo</i> . Modelo de infección aguda en ratones	149
5.5.1. Selección de candidatos a ser evaluados en ensayos preclínicos	149
5.5.2. Modelo de infección aguda en ratones.....	151
Referencias Capítulo 5	153
Capítulo 6: Conclusiones	157
Anexo 1	165
A.1. Índices topológicos utilizados en los modelos desarrollados.....	167
A.1.1. Topología Molecular	167
A.1.2. Matrices asociadas a grafos moleculares.....	169
A.1.3. Índices de conectividad simple	172
A.1.4. Índices de Carga	173
A.1.5. Índices derivados de la matriz de adyacencia	175
Referencias Anexo 1.....	177

Anexo 2	179
A.2.1. Estructuras de los Compuestos Activos del Conjunto de entrenamiento.....	181
A.2.2. Estructuras de los Compuestos Inactivos del Conjuntode entrenamiento. ..	185
A.2.3. Estructuras de los Compuestos Activos del Conjunto de prueba.....	188
A.2.4. Estructuras de los Compuestos Inactivos del Conjunto de prueba.	190

Lista de abreviaturas

ALD	Análisis Lineal Discriminante
AUC	Area Under the Curve (Área bajo la curva)
CMap	Connectivity Map
CV	Cribado o Tamizado Virtual
Cz	Cruzipaína
DE	Desviación Estándar
DMSO	Dimetil sulfóxido
DNDi	Drug for Neglected Disease initiative (Iniciativa Medicamentos para Enfermedades Olvidadas)
FD	Función Discriminante
FDA	Food and Drug Administration
FIV	Factor de Inflación de la Varianza
HTS	High Throughput Screening (Cribado de alta performance)
LBD	Literature based discovery (Descubrimiento basado en la literatura)
LGO	Leave Group Out (Dejar un grupo afuera)
LOO	Leave One Out (Dejar uno afuera)
MSC	Máxima Subestructura Común
OMS	Organización Mundial de la Salud
QSAR	Quantitative Structure-Activity Relationship (Relación cuantitativa estructura actividad)
ROC	Receiver Operating Characteristic
Se	Sensibilidad
Sp	Especificidad
<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>

Objetivos

Objetivo general:

-  Descubrir nuevos agentes terapéuticos aplicables en la farmacoterapia de la Enfermedad de Chagas, mediante cribado virtual (también conocido como screening o tamizado virtual) de grandes bases de datos de compuestos químicos. La diversidad química de las bases de datos utilizadas permitirá encontrar prototipos activos novedosos (nuevos líderes). 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.

Objetivos específicos:

-  Desarrollar modelos topológicos capaces de establecer qué características estructurales fundamentales debe reunir un compuesto químico para poseer actividad inhibitoria sobre la cruzipaína (Cz) del *Trypanosoma cruzi* (*T. cruzi*).
-  Aplicar el conocimiento teórico generado en la búsqueda racional, mediante Cribado Virtual (CV), de nuevos agentes terapéuticos contra la enfermedad de Chagas. Este paso implica contrastar cada estructura química de una base de datos con el modelo generado, para determinar qué compuestos de la base de datos cumplen con los requisitos estructurales definidos por el modelo. En particular, la identificación de fármacos y drogas aprobadas por la Food and Drug Administration (FDA), permitirá proponer el uso de fármacos conocidos como antichagásicos (*knowledge-based drug repurposing*).
-  Adquisición y/o síntesis de aquellas estructuras señaladas como más promisorias por los modelos teóricos desarrollados y evaluación mediante ensayos *in vitro* de su actividad inhibitoria sobre Cz, a fin de validar de manera experimental las predicciones de los modelos teóricos.

CAPÍTULO 1

INTRODUCCIÓN





1.1. ENFERMEDAD DE CHAGAS

La Enfermedad de Chagas, Chagas-Mazza o Tripanosomiasis Americana es una enfermedad parasitaria endémica de América Latina, causada por el protozoo hemoflagelado *T. cruzi*. La principal vía de transmisión es la picadura de un insecto hematófago de la subfamilia *Triatominae* conocido como vinchuca, chinche o barbeiro. Otras vías incluyen las transfusiones sanguíneas, la vía congénita y la ingestión de alimentos contaminados.

En 1909 el Dr. Carlos Ribeiro Justiniano Chagas (1879-1934) médico e investigador del Instituto Oswaldo Cruz (Brasil) descubrió la patología mientras realizaba una investigación sobre Malaria en el estado de Mina Gerais en Brasil; allí encontró enfermos que padecían una sintomatología con características propias. Chagas descubría una nueva enfermedad que más tarde llevaría su nombre. Tiempo después, descubrió su agente causal: un parásito unicelular hemoflagelado al que llamó "*Trypanosoma cruzi*" en homenaje al Dr. Oswaldo Cruz. Lo estudió a través de su ciclo evolutivo y describió el papel que desempeña la vinchuca como agente transmisor (Chagas, 1909). Años más tarde, a mediados de la década de 1920, el Dr. Salvador Mazza (1886-1947) en el contexto de una Misión de Estudio de Patologías Regionales de Argentina (MEPRA) impulsó y coordinó diferentes estudios sobre la enfermedad Chagas, identificando los primeros casos en Argentina. Sus actividades en relación a la enfermedad, incluyendo la comprobación de la existencia del parásito y sus vectores en nuestro territorio, el análisis de los reservorios naturales, la identificación de casos agudos y crónicos, la descripción de las características de cada etapa, el análisis de los métodos de diagnóstico y el ensayo (sin éxito) de tratamientos experimentales (Mazza, 1939).

La enfermedad cobraba mayor importancia a medida que aumentaba el número de casos agudos diagnosticados, hecho que se vio favorecido por la identificación del "Signo de Romaña" (inflamación del párpado como consecuencia de la picadura de la vinchuca), lo cual a su vez contribuyó a la amplia difusión de la identificación o caracterización de la enfermedad en la comunidad médica.

Hacia fines de 1930 la enfermedad era reconocida en el ámbito médico (y científico en general) como un grave problema sanitario de toda la región de América del Sur, reconocimiento sustentado por los cientos de casos certificados por Mazza y sus colaboradores y por las actividades de divulgación que ellos desarrollaron (Dias, 1939; Mazza, 1943; Sordelli, 1939).

A más de cien años de su descubrimiento la enfermedad de Chagas sigue siendo un problema sanitario tanto en Argentina como en el resto de Latinoamérica. Se han ejecutado diferentes programas de intervención orientados a la eliminación del vector, regulación de las transfusiones sanguíneas, campañas educativas, mejora de viviendas, disponibilidad de tratamientos antiparasitarios y promoción de la investigación en diferentes áreas; sin embargo, la voluntad política de dar prioridad al Chagas no es suficiente, la enfermedad continúa siendo un problema actual y global.

Durante los últimos años se consiguieron avances científicos muy importantes en relación a la biología del parásito y la comprensión de múltiples factores involucrados en la enfermedad, lo cual ha contribuido en la realización de los programas para el control vectorial y el diagnóstico de pacientes. Sin embargo, existe una brecha muy grande entre este progreso en el conocimiento y su aplicación en el tratamiento del paciente: desde hace más de 50 años existen sólo dos fármacos aprobados para el tratamiento de la enfermedad. Frente al desinterés de la industria farmacéutica a desarrollar nuevos fármacos con menor toxicidad y mayor eficacia, son el Estado y las instituciones públicas quienes deben jugar un rol protagónico. El trabajo realizado en la presente tesis propone una estrategia que a nuestro entender constituye una prometedora aproximación para la búsqueda de nuevos tratamientos: el reposicionamiento de fármacos asistido por computadora. El reposicionamiento permite una incorporación mucho más directa de los resultados obtenidos a la práctica clínica, luego de la aprobación de las autoridades correspondientes, con una inversión mucho menor que la necesaria para desarrollar un fármaco *de novo*. Se trata, por lo tanto, de una estrategia eficiente desde el punto de vista translacional.

Sin embargo, es importante aclarar que el Chagas es una enfermedad que comprende aspectos biológicos y ecológicos, médicos, políticos y sociales; por lo tanto, su solución excede las posibilidades de cada uno de estos aspectos abordados de manera separada.

Encontrar un nuevo fármaco para tratar la enfermedad no es la solución al problema; la farmacoterapia no es más que una componente de la red en la cual está inmerso el paciente o portador de Chagas.

La solución radica en realizar una tarea conjunta, donde las acciones se implementen en medidas concretas que fortalezcan el desarrollo y la participación de la comunidad involucrando a otras ciencias como la antropología, la sociología, la psicología, la ecología, la política y la economía (Storino, 2000; Storino, 2012); apuntando a un abordaje integral que esté orientado no solo a descubrir nuevos tratamientos farmacológicos, y a eliminar las vinchucas de las zonas endémicas, sino a la promoción de la salud de la población como un medio para lograr el desarrollo individual y colectivo, abarcando todos los niveles de prevención, poniendo foco en la atención médica integral del paciente con Chagas, implementando centros de estudio y control de todos los aspectos de la enfermedad (especialmente en la inserción laboral del paciente/portador), garantizando el diagnóstico temprano y la accesibilidad al tratamiento, pero sobre todo modificando la situación de marginación y olvido que padecen millones de personas. Para esto no basta la voluntad política por parte del Estado, sino que es necesaria la responsabilidad social de todos los actores que trabajamos para que el Chagas deje de ser una enfermedad olvidada.

En este primer capítulo se realiza un recorrido por los diferentes aspectos de la enfermedad, la biológica del parásito, las fases clínicas, epidemiología, tratamientos disponibles y nuevas perspectivas en la quimioterapia; en los siguientes se detallan las estrategias empleadas en el trabajo, las metodologías, los resultados obtenidos y las conclusiones a las cuales hemos arribado.

1.1.1. BIOLOGÍA DEL PARÁSITO, VÍAS DE TRANSMISIÓN Y FASES CLÍNICAS DE LA ENFERMEDAD DE CHAGAS

*1.1.1.1. CICLO DE VIDA DEL *T. CRUZI**

El *T. cruzi* presenta un ciclo de vida complejo que involucra la infección del huésped vertebrado y la transmisión por los insectos vectores (vinchucas). Básicamente, *T. cruzi* atraviesa 3 estadios morfológicos bien diferenciados: la forma no flagelada amastigote y las formas flageladas epimastigotes y tripomastigotes (Brenner, 1973; Tyler and Engman, 2001). La forma epimastigote prolifera en el intestino del insecto vector y en el recto se diferencia eventualmente transformándose en tripomastigotes metacíclicos (forma infectante). Cuando el insecto pica a otro mamífero (ocasionalmente el hombre) las heces son evacuadas en el sitio de la picadura e ingresan en el torrente sanguíneo del nuevo anfitrión a través de la piel dañada, iniciándose la segunda etapa del ciclo de vida, dentro del huésped vertebrado. Una vez que ingresa en el torrente sanguíneo del huésped en su forma tripomastigote puede penetrar diferentes células. Dentro de las mismas se transforma en amastigote, sufriendo varios ciclos de división y diferenciándose posteriormente en tripomastigotes sanguíneos (forma no infectante) que luego de la lisis celular se liberan comenzando nuevamente el ciclo (Figura 1.1) (Clayton, 2010; Murcia et al., 2013).

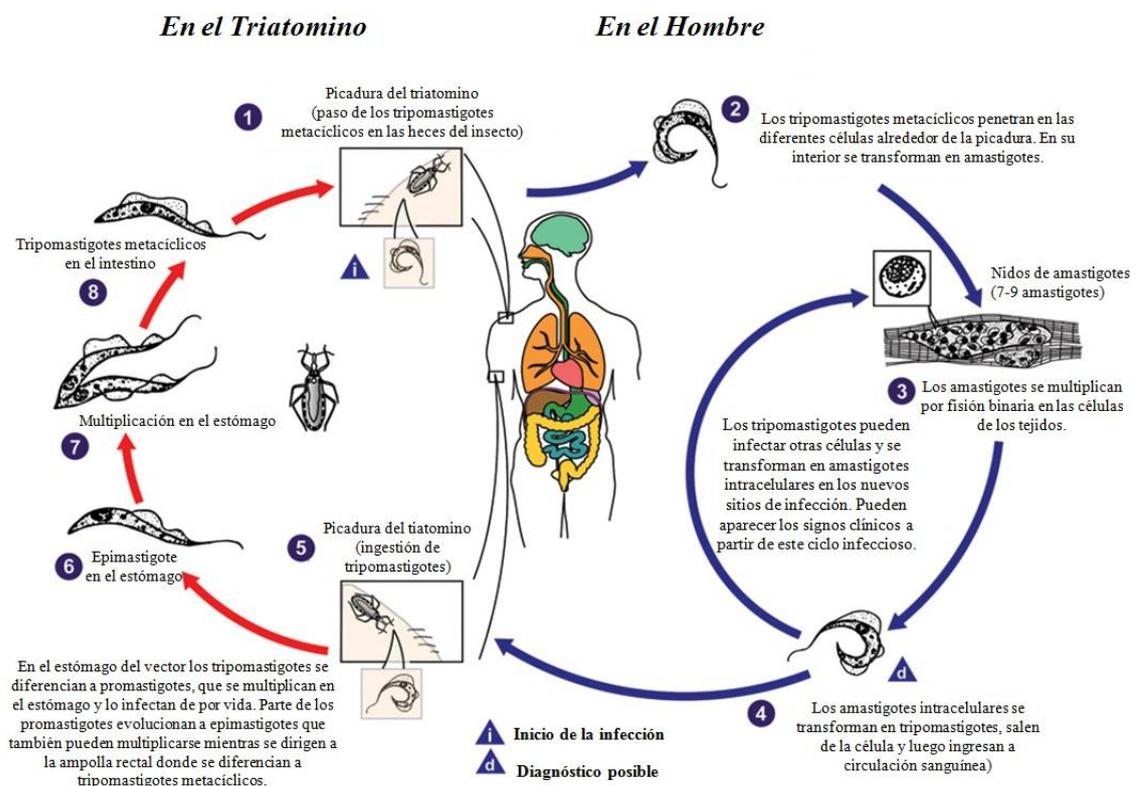


Figura 1.1. Ciclo biológico de *T. cruzi*.



1.1.1.2. VÍAS DE TRANSMISIÓN

La forma más común de transmisión de la enfermedad en zonas endémicas es a través de las heces del insecto infectado, las que contienen tripomastigotes metacíclicos. Debido a que los insectos triatominos tienen el tracto digestivo recto, inmediatamente después de alimentarse con la sangre de seres humanos y animales defecan y al rascarse la persona o el animal permite que el parásito se introduzca en la sangre. Esta transmisión es conocida como *transmisión vectorial* y comprende básicamente 3 ciclos: doméstico, peridomiciliario y selvático.

En relación al ciclo doméstico, la estructura de las casas rurales o periurbanas en zonas endémicas las hace especialmente vulnerables a la infección por vía vectorial. Por lo general los insectos hematófagos de hábitos nocturnos viven en las grietas y huecos de las paredes de adobe y los techos de paja de las viviendas precarias construidas en zonas rurales o suburbanas. Por otra parte, la estrecha asociación entre los habitantes y los animales domésticos establece una fuente de sangre abundante y de fácil acceso, por lo que el interior de las viviendas conforma un hábitat muy favorable para los insectos. *Triatoma infestans* (vinchuca) es el principal vector domiciliario en los países de América del Sur (Argentina, Brasil, Bolivia, Chile, Paraguay y Uruguay) (WHO, 2012; WHO, 2013).

El ciclo peridoméstico actúa como nexo entre los ciclos doméstico y selvático e involucra especies como marsupiales, roedores, y otros mamíferos que ocasionalmente entran y salen de las viviendas, y triatominos selváticos diferentes de *Triatoma infestans* que son atraídos a las casas por la luz y el alimento (WHO, 2012).

En el ciclo selvático intervienen triatominos selváticos que infectan a numerosas especies y subespecies de mamíferos salvajes terrestres o arbóreos. En la Figura 1.2 se observa un diagrama de interconexión simplificado entre los diferentes ciclos (Rodrigues Coura and de Castro, 2002).

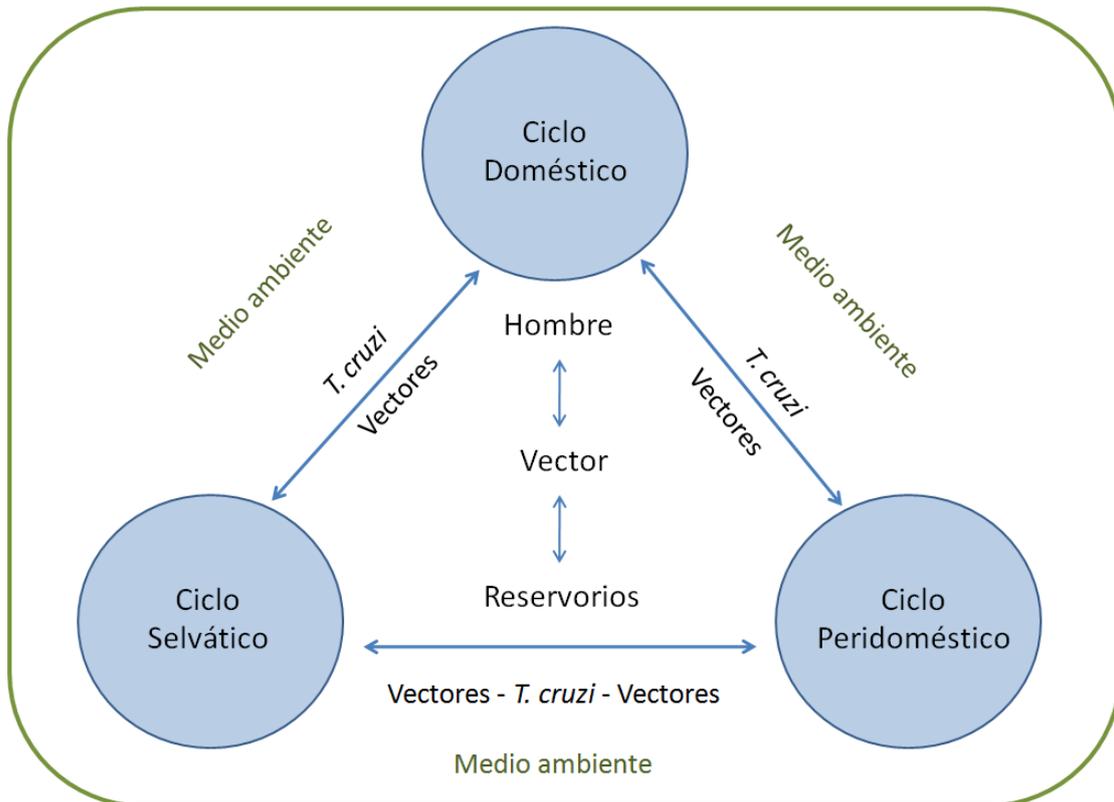


Figura 1.2. Diagrama de interconexión entre los ciclos selvático, peridoméstico y doméstico. (adaptado de Rodrigues Coura J y Pinto Dias JC, 2009)

La *transmisión congénita* cobra particular relevancia como consecuencia de los avances en el control de la transmisión vectorial, y no está limitada solo a zonas rurales sino que también se verifica cada vez con mayor frecuencia en ciudades donde no hay transmisión vectorial, debido a la migración de mujeres en edad de procrear hacia zonas urbanas (WHO, 2007).

La transmisión congénita no se puede prevenir durante el embarazo dado que no se conocen los riesgos teratogénicos de los fármacos disponibles (nifurtimox, benznidazol) además de presentar ambos principios activos alto riesgo de efectos adversos en adultos (Carlier et al., 2011). Sin embargo, los recién nacidos infectados diagnosticados y tratados durante el primer año de vida poseen casi un 100% de posibilidades de obtener curación parasitológica mediante tratamiento con benznidazol; al mismo tiempo, la frecuencia de efectos adversos en pacientes pediátricos es baja (Altchek et al., 2011;

WHO, 2012). Por lo tanto es particularmente importante la detección activa de la enfermedad en las mujeres embarazadas y el diagnóstico temprano a lactantes.

El riesgo de transmisión de madre a hijo varía según la cepa de *T. cruzi*, la parasitemia de la madre, la existencia de lesiones de la placenta y la región geográfica; en los países de América del Sur oscila entre 1% y 12% dependiendo de la región (Howard et al., 2014).

Otras vías de transmisión que ocurren con menor frecuencia son la transfusión de sangre, el trasplante de órganos y la ingesta de alimentos contaminados con *T. cruzi* (Haberland et al., 2013).

1.1.1.3. ASPECTOS CLÍNICOS DE LA ENFERMEDAD

Las manifestaciones clínicas de la enfermedad pueden clasificarse en tres etapas. La etapa aguda de corta duración, la etapa clínicamente asintomática también conocida como indeterminada y una fase crónica de larga duración.

La fase aguda aparece entre los 7 y 14 días tras la infección. Se caracteriza por presentar parasitemia circulante y sintomatología mayormente inespecífica. En ésta los parásitos invaden y se multiplican en diferentes células (macrófagos, músculo liso y estriado, neuronas). Luego sobreviene la respuesta inmune del huésped y una reducción en la respuesta inflamatoria de las células infectadas, lo que se traduce en una segunda fase indeterminada o latente, asintomática, que puede durar años (Haberland et al., 2013; Rodriques Coura and de Castro, 2002).

Cuando la penetración del parásito es a través de una laceración de la piel, aparece una zona indurada y eritematosa, denominada «chagoma», acompañada de linfadenopatía local. El signo de Romana, que es la manifestación clásica de la enfermedad de Chagas aguda, se manifiesta por edema indoloro bipalpebral y unilateral y tiene lugar cuando la vía de entrada es conjuntival. Los primeros signos se acompañan de malestar general, fiebre, anorexia, edema facial y de extremidades inferiores (Murcia et al., 2013).

La fase crónica se caracteriza principalmente por cardiopatía severa; la miocarditis crónica es lenta y progresiva, y una vez que aparece el daño miocárdico, éste tiende a empeorar. La cardiopatía chagásica crónica se denomina también miocardiopatía dilatada al manifestarse como una cardiomegalia; también pueden presentarse megaesófago, megacolon y problemas neurológicos (Haberland et al., 2013; Rodriques Coura and de Castro, 2002).

1.1.1.4. EPIDEMIOLOGIA Y TRATAMIENTO DISPONIBLE

La enfermedad de Chagas fue clasificada por la Organización Mundial de la Salud (OMS), de la misma manera que otras enfermedades tropicales infecciosas como la leishmaniasis o la enfermedad del sueño, como una de las “enfermedades olvidadas”. Estas son enfermedades que afectan principalmente a las poblaciones más pobres del planeta, y que por tanto no representan un retorno lucrativo suficiente que justifique una inversión de la industria farmacéutica en investigación y desarrollo de nuevos medicamentos para su tratamiento. El Chagas es la enfermedad tropical transmisible de mayor prevalencia en América Latina, con una carga de enfermedad que en 1990 era cinco veces mayor que la carga del paludismo y superaba a la generada en toda América por todas las otras enfermedades tropicales tomadas en conjunto. Aunque dicha carga disminuyó significativamente entre 1990 y el 2001, en este último año todavía era mayor que la producida individualmente por la malaria, la leishmaniasis, la lepra y la esquistosomiasis (WHO, 2013).

Posee una incidencia anual de 28.000 casos en todo el continente Americano. Se estima que afecta aproximadamente a 8 millones de personas y provoca unas 12.000 muertes anuales. Su costo económico ha sido estimado en unos 7 mil millones de dólares anuales (Lee et al., 2013; Nunes et al., 2013; WHO, 2013). La mortalidad ha descendido notablemente desde los años 90 a esta parte principalmente por los programas de control vectorial de los diferentes países latinoamericanos (WHO, 2007) (ver tabla 1 con reporte de la OMS).

Tabla 1.1. Datos Epidemiológicos de la OMS. Cambios en los parámetros epidemiológicos por la interrupción de la transmisión y descenso de la incidencia de la enfermedad de Chagas, 1990–2000–2006.

<i>Parámetros epidemiológicos</i>	<i>1990</i>	<i>2000</i>	<i>2006</i>
Muertes anuales	>45.000	21.000	12.500
Casos humanos de infección	30 millones	18 millones	15 millones
Nuevos casos anuales	700.000	200.000	41.200
Población en riesgo	100 millones	40 millones	28 millones

Es importante tener en cuenta que los movimientos migratorios de las zonas rurales a las zonas urbanas han modificado el mapa epidemiológico de la enfermedad, su transmisión y atención médica. Por ejemplo, tras la reducción de las densidades demográficas en áreas rurales endémicas, una menor infestación por el vector puede ser considerado como un efecto "positivo". La infección, que principalmente era rural, paso a ser urbana y en las últimas décadas el número de donantes de sangre con serología positiva ha sido elevado en países endémicos, pese a que en la mayoría de los países de Latinoamérica se ha establecido por ley la obligatoriedad de que todos los bancos de sangre dispongan de ensayos para la prevención de la transmisión de *T. cruzi* por transfusiones (Pinto Dias, 2013).

Debe subrayarse que la transmisión no se limita a los países endémicos. La migración de personas infectadas con *T. cruzi* posiciona a la Enfermedad de Chagas como un problema de salud pública global; debido a estas corrientes migratorias se han detectado numerosos casos por transmisión a través de trasplantes o transfusiones en países donde no existe transmisión vectorial del parásito.

La globalización de la enfermedad ha obligado a países no endémicos a implementar medidas de prevención como así también discutir nuevas estrategias para su control. En la Figura 1.3 se observa un mapa con las principales corrientes migratorias desde

Latinoamérica y el número estimado de personas infectadas en países y regiones no endémicos (Coura and Dias, 2009).

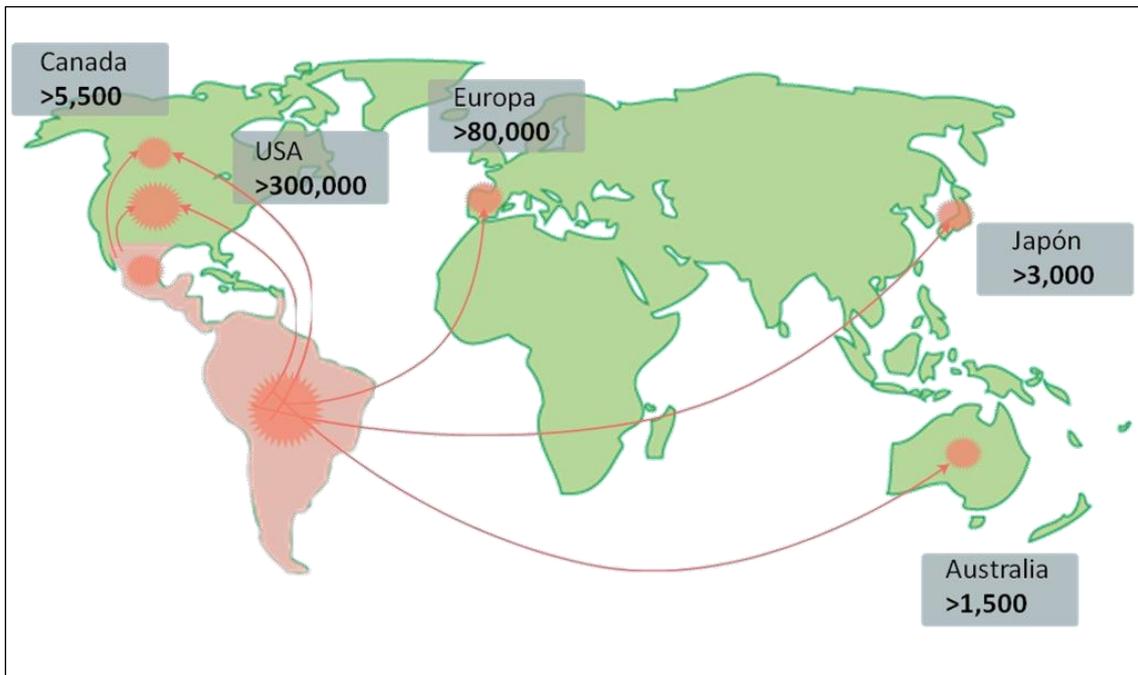


Figura 1.3. Rutas de migración desde América Latina y estimación del número total de infectados en países no endémicos.

1.2. FARMACOTERAPIA ACTUAL PARA LA ENFERMEDAD DE CHAGAS

A pesar de afectar a millones de personas en todo el mundo, al día de hoy no existen tratamientos eficaces de la enfermedad de Chagas. Desde hace más de cuatro décadas sólo existen dos drogas disponibles en el mercado para su tratamiento: nifurtimox (1972) y benznidazol (1974) (ver Figura 1.4). Ambos fármacos son nitroheterociclos.

Nifurtimox es un nitrofurano N-(3-metil-1,1-dioxo-1,4-tiazinan-4-il)-1-(5-nitro-2-furil) metanimina, que actúa por la vía de la reducción del grupo nitro a radicales aniónicos inestables, lo cual produce una reacción que genera la producción de metabolitos de oxígeno reducido altamente tóxicos; *T. cruzi* es deficiente en mecanismos de detoxificación para metabolitos de oxígeno (particularmente peróxido de hidrogeno) siendo por tanto muy sensible al estrés oxidativo (Maya et al., 2007; Urbina and

Docampo, 2003). Se han postulado también otros mecanismos de acción que sugieren la generación de metabolitos tóxicos, del tipo nitrilos insaturados de cadena abierta, por acción nitroreductasas del tipo I (Boiani et al., 2010; Hall et al., 2011).

Benznidazol es un nitroimidazol N-bencil-2-(2-nitroimidazol-1-il)acetamida cuyo mecanismo de acción involucra la modificación covalente de macromoléculas mediante intermediarios de la nitroreducción (Hardman and Limbird, 2002; Maya et al., 2007; Mecca et al., 2008).

Si bien el tratamiento de pacientes adultos y niños en fase aguda es efectivo (tasas de curación de 60-80% en adultos en fase aguda y de 80-100% en niños) las principales limitaciones de éstos fármacos son la muy alta frecuencia de efectos adversos en adultos y las reducidas tasas de curación (10-20%) en adultos en fase crónica (McKerrow et al., 2009; Murcia et al., 2012; Rodriques Coura and de Castro, 2002; Urbina and Docampo, 2003).

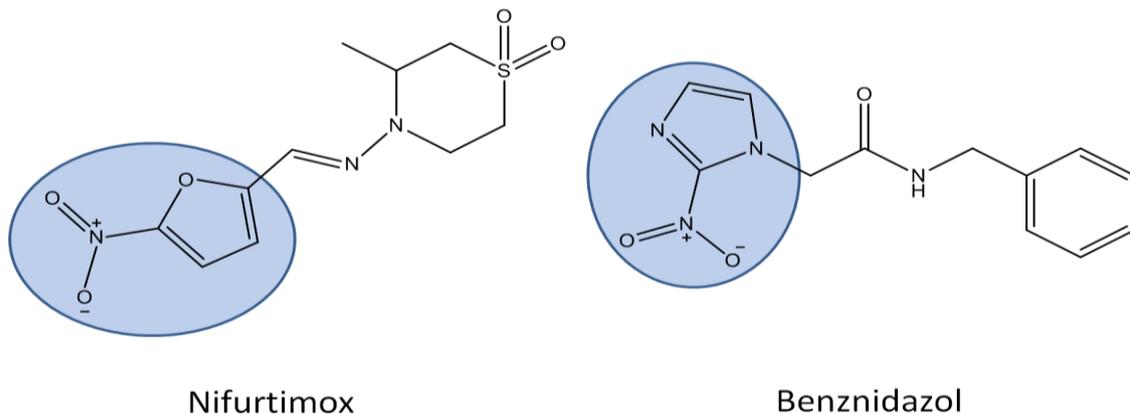


Figura 1.4. Estructuras de los fármacos disponibles en la actualidad para el tratamiento de la enfermedad de Chagas. Las regiones coloreadas corresponden a los grupos 5-nitrofurano y 2-nitroimidazol respectivamente.

Benznidazol es considerado como el tratamiento de primera elección, sin embargo nifurtimox también se utiliza en aquellos pacientes que no toleran el benznidazol (Bern et al., 2007; Rassi et al., 2010).

Los efectos adversos más frecuentes en pacientes tratados con benznidazol son las erupciones cutáneas y la neuropatía periférica; ambos se observan en hasta un 30% de los pacientes tratados. La erupción cutánea se desarrolla aproximadamente 10 días después de la iniciación del tratamiento; es una dermatitis pruriginosa fotosensible que se puede tratar con corticoides por vía tópica. Sin embargo, las reacciones severas, incluyendo los síndromes de hipersensibilidad con fiebre, linfadenopatía, y dermatitis exfoliativa deben ser reportados y requieren la interrupción inmediata de la administración del fármaco. La neuropatía inducida por benznidazol es dosis dependiente, tiene un inicio tardío y es reversible, pero su desarrollo también obliga a la interrupción inmediata del tratamiento (Bern et al., 2007). Otros efectos menos graves asociados a la terapia con dicho fármaco incluyen, anorexia, náuseas, vómitos, pérdida de peso e insomnio (Parker and Sethi, 2011; Viotti et al., 2009).

Los pacientes medicados con nifurtimox pueden presentar los siguientes efectos secundarios: anorexia, pérdida de peso, alteraciones psíquicas, excitabilidad, insomnio, manifestaciones digestivas como náuseas o vómitos, cólicos intestinales y ocasionalmente diarrea (Castro, 2006).

1.2.1. PERSPECTIVAS SOBRE NUEVAS QUIMIOTERAPIAS PARA LA ENFERMEDAD DE CHAGAS

La OMS define los requisitos que debiera reunir la medicación ideal para la enfermedad de Chagas siendo ellos (Rodrigues Coura and de Castro, 2002): i) cura parasitológica del paciente en cualquier fase de la enfermedad; ii) efectivo en un tratamiento de duración reducida; iii) accesible a los pacientes; iv) ausencia de efectos colaterales; v) posibilidad de administrar en pacientes ambulatorios; vi) ausencia de resistencia.

Aunque el desarrollo de un fármaco que reúna todas esas características es difícil, la profundización del conocimiento de la biología del *T. cruzi* permite la búsqueda de nuevos agentes farmacológicos más cercanos a la idealidad que las medicaciones disponibles actualmente, las cuales distan mucho de la misma particularmente por el

incumplimiento de los requisitos ii) y iv) de los enumerados en el párrafo anterior. Por otra parte, históricamente la accesibilidad a los tratamientos conocidos también se ha visto restringida. Actualmente, una iniciativa público-privada entre el laboratorio Elea, la Fundación Mundo Sano y el Ministerio de Salud de la Nación ha subsanado esta última problemática mediante la producción de comprimidos de benznidazol (Abarax®) para el abastecimiento local y regional de dicho medicamento. Dadas las limitaciones del restringido arsenal terapéutico disponible actualmente, se han señalado como necesidades de investigación prioritarias la investigación de nuevas dianas moleculares, **la exploración/screening de bibliotecas químicas públicas y privadas**, la explotación de la biodiversidad de las Américas en la búsqueda de nuevas terapias contra la enfermedad de Chagas y el ensayo preclínico y clínico de nuevos candidatos (WHO, 2007).

A causa de las grandes diferencias de eficacia en las distintas fases de la enfermedad y al alto porcentaje de efectos adversos de las únicas drogas disponibles ha surgido la necesidad de descubrir nuevas drogas con un perfil farmacológico más adecuado. Durante los últimos años el progresivo aumento en el conocimiento de la biología de *T. cruzi* (Croft et al., 2005; Urbina and Docampo, 2003) ha facilitado el desarrollo de nuevas aproximaciones racionales hacia una quimioterapia específica para tratar el Chagas, recurriendo a diversos blancos moleculares para identificar potenciales compuestos con propiedades anti-*T. cruzi*. En particular las investigaciones se centran en blancos moleculares diferentes a los ortólogos que posee el mamífero, o ausentes en este último.

Algunos procesos y/o enzimas que se utilizan actualmente como blancos terapéuticos son: tripanotión reductasa; biosíntesis de esteroides de membrana; metabolismo del pirofosfato; enzimas glicolíticas; trans-sialidasas y; cisteín-proteasas, entre otras (Duschak and Couto, 2007; WHO, 2007).

De todos los compuestos desarrollados en las últimas décadas quienes presentan notables posibilidades de superar los ensayos clínicos, son varios fármacos antifúngicos que demostraron actividad en modelos agudos y crónicos en ratones posaconazol, E1224 (profármaco de ravuconazol) y fexinidazol, inhibidores de la síntesis de ergosterol (Bustamante, 2014) (Figura 1.5). Posaconazol ha demostrado eficacia en varias cepas resistentes a los nitroderivados (Bustamante, 2014). En 2012, ingresó en

ensayos clínicos en Argentina, Bolivia y España, aunque estos resultados demostraron que es claramente inferior a la terapia estándar, posiblemente debido a la baja exposición sistémica al fármaco, laboratorios Merck ya ha iniciado un segundo estudio (<http://clinicaltrials.gov/show/NCT01377480>, último acceso agosto 2014). Sin embargo, se ha subrayado que, aunque estos ensayos fueron exitosos, posaconazol es muy caro debido al costoso esquema sintético y su bajo rendimiento, lo cual limita su uso generalizado en países endémicos (Lepesheva, 2013).

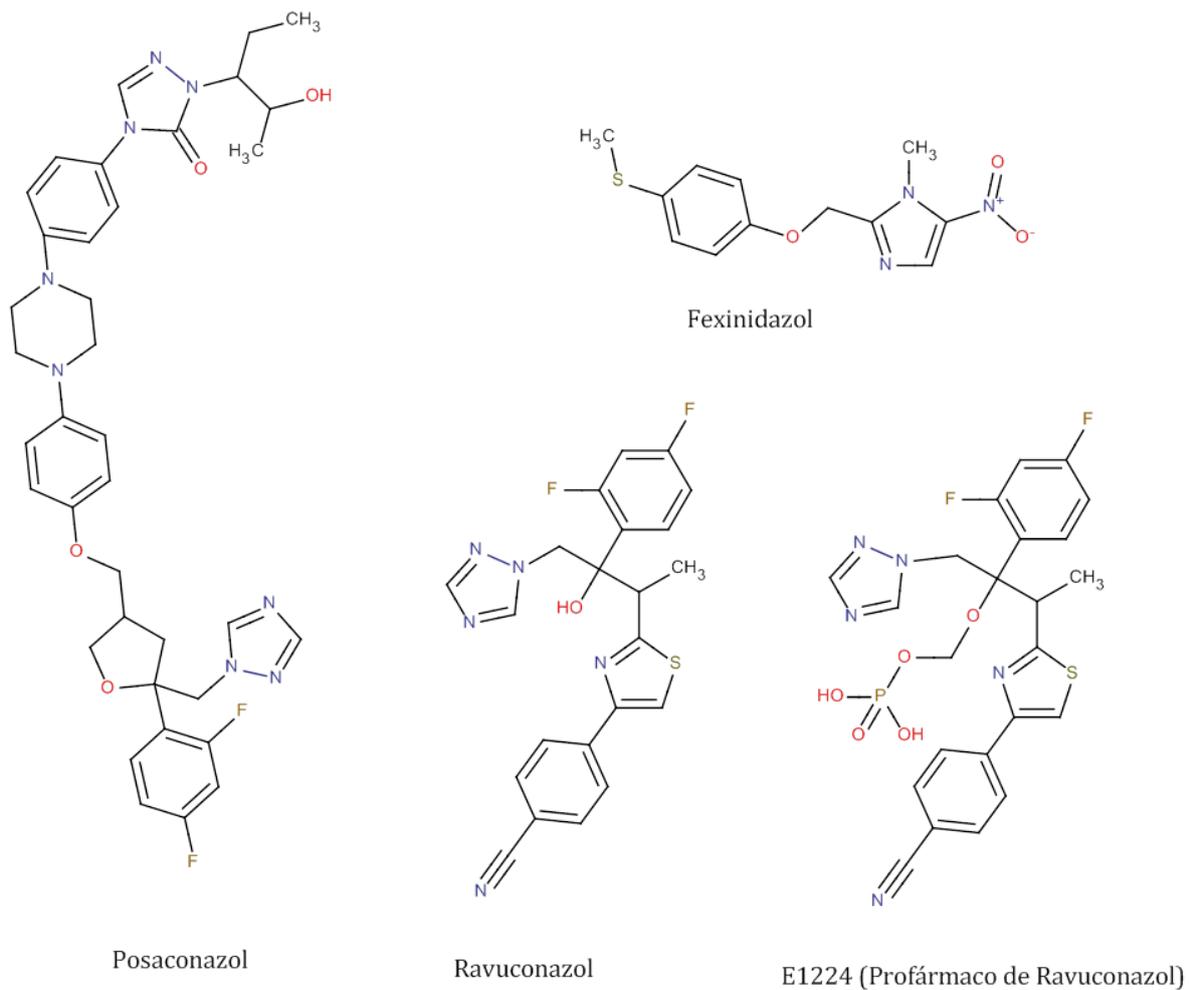


Figura 1.5. Estructuras químicas de candidatos sometidos a estudios preclínicos avanzados.

Ravuconazol es un segundo antifúngico azólico que muestra una potente actividad in vitro frente a *T. cruzi*. A pesar de su farmacocinética desfavorable en animales (vida

media muy corta) su vida media de eliminación y el volumen de distribución en humanos impulsaron un ensayo clínico de prueba de concepto de un profármaco del ravuconazol (E1224) (Buckner, 2012). E1224 fue eficaz en la eliminación del parásito al final del ciclo de tratamiento, pero no logró desarrollar una eficacia sostenida un año después del tratamiento en comparación con benznidazol y presentó algunos problemas de seguridad en dosis altas (<http://www.dndi.org/diseases-proyectos/cartera/azoles-e1224.html>, visitado por última vez agosto de 2014). Los ensayos adicionales de E1224 como una terapia combinada con benznidazol se han anunciado recientemente (Bustamante, 2014).

Fexinidazol es un nitroimidazol que atraviesa en la actualidad ensayos clínicos de Fase II/III para la Tripanosomiasis Africana (Bustamante, 2014). La droga fue evaluada frente a *T. cruzi* hace más de 30 años y más tarde fue rescatada por el Proyecto Iniciativa de Medicamentos para Enfermedades Olvidadas (en inglés *Drugs for Neglected Disease initiative*, DNDi) como un candidato a fármaco para el tratamiento de la Tripanosomiasis Africana. Después de completar la primera fase en humanos (Tarral, 2014) entró en estudios de Fase II / III a finales de 2012. También en 2012, un estudio mostró efectos comparables sobre *T. cruzi* (cepa CL-benznidazol susceptible y cepa Y parcialmente resistente) entre benznidazol y fexinidazol (este último en dosis 3 veces más alta con respecto de la terapia estándar), mientras que un efecto superior de fexinidazol en modelos de infección aguda y crónica en ratones se observó en las cepas VL-10 y colombianas resistentes a benznidazol (Bahia, 2012). En el mismo estudio fexinidazol demostró reducir la miocarditis en animales infectados con la cepa VL-10 y Colombiana. Estos descubrimientos, junto con un perfil de seguridad favorable sentaron las bases para el inicio de un ensayo de prueba de Fase II como medicación antichagásica en Bolivia hacia fines de 2013.

En el siguiente apartado se describen las características estructurales y funcionales más relevantes de la cisteína-proteasa de *T. cruzi*, Cz (blanco molecular sobre el cual se trabajó en el presente trabajo de tesis) y de los inhibidores descubiertos hasta la fecha.

1.3. CRUZIPAÍNA COMO BLANCO TERAPÉUTICO DE LA ENFERMEDAD DE CHAGAS

1.3.1. ASPECTOS GENERALES, ESTRUCTURALES Y FUNCIONALES

Cz es una cisteína-proteasa análoga a la catepsina L (también conocida como gp51/57 o cruzañina, en su forma recombinante), responsable de la mayor parte de la actividad proteolítica de *T. cruzi*, en todos los estadios de su ciclo de vida (Caffrey et al., 2000; Cazzulo, 2002; McKerrow et al., 2009).

Cz es codificada por diversos genes colocados en serie y separados por espacios intragénicos de algo más de 400 pares de bases (hasta 130 en la cepa Tulahuen 2) (Campetella et al., 1992). Esta variabilidad génica resulta en la expresión de diversas isoformas de la enzima, algunas de las cuales presentan variaciones no conservativas en la secuencia de aminoácidos que influyen en la especificidad de la misma (Duschak et al., 2003; Lima et al., 2001). Se ha demostrado que la composición de isoformas varía entre cepas y en diferentes estadios del ciclo de vida del parásito lo cual podría explicar las variaciones inter-cepa observadas en la susceptibilidad a distintos inhibidores (Gomes et al., 2009).

Diferentes estudios han demostrado que Cz se expresa de manera diferencial en los cuatro estadios principales del ciclo de vida de *T. cruzi* (McKerrow et al., 2009). Cz se encuentra en mayores niveles en la forma epimastigotes localizándose en organelas análogas a los lisosomas que almacenan proteína que es digerida por el parásito durante la etapa de diferenciación a la forma tripomastigote metacíclico (Souto-Padron et al., 1990); también se localiza en menor proporción en la membrana plasmática del parásito en todas sus formas, pero en mayores niveles en la forma amastigote (Nascimento and de Souza, 1996).

Estructuralmente, consta de una única cadena polipeptídica que posee un dominio catalítico de 215 aminoácidos y un dominio C-terminal de 130 aminoácidos (Figura 1.6) La estructura cristalográfica muestra que la cadena polipeptídica de 215 aminoácidos se pliega en dos dominios definidos. El dominio izquierdo presenta tres regiones de hélices alfa, una de las cuales contiene el residuo Cys25 en su extremo N-terminal. El dominio

derecho presenta principalmente una disposición de hojas beta antiparalelas. La tríada catalítica, formada por los residuos de Cys25, His159 y Asn175, y los distintos subsitios (S1-S4 y S1'-S3') de unión al sustrato se encuentran en una hendidura entre los dos dominios (Figura 1.6c) (Huang et al., 2003; McGrath et al., 1995).

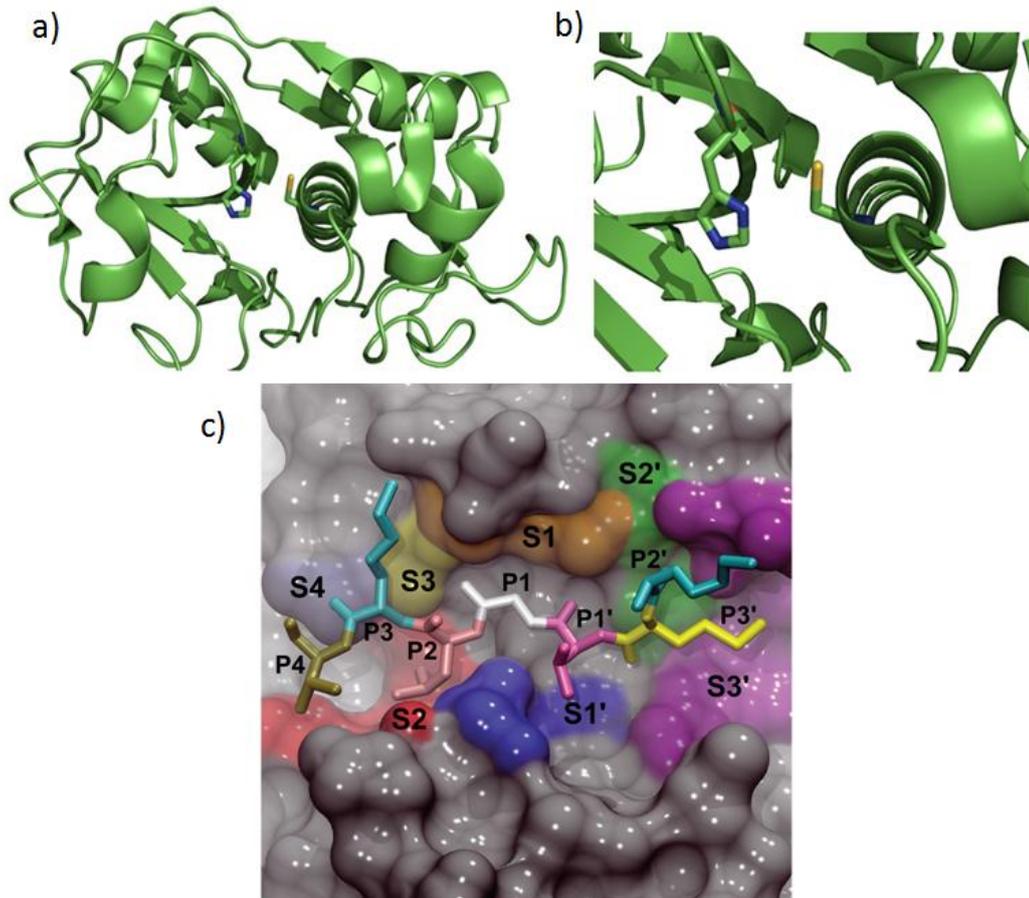


Figura 1.6. Modelo de Cz en Pymol (a), de su sitio activo (b) y subsitios (c). (Imagen c, reproducida bajo licencia Creative Commons).

En relación al mecanismo de acción propuesto a través del cual ocurre la hidrólisis, se postula que la Cys25 y la His159 forman parte del sitio activo de la enzima pero se ha cuestionado si bastan esos dos residuos, únicamente, para una actividad catalítica completa. El grupo imidazol de la histidina polariza el tiol de la cisteína permitiendo la desprotonación aún en condiciones neutras o débilmente ácidas. El par iónico S-imidazolium que se produce es altamente nucleofílico. Las cisteína proteasas hidrolizan el enlace amida de manera similar a las serina proteasas. Luego de la unión al sustrato se forma un complejo no covalente de Michaelis. El anión tiolato ataca al grupo carbonilo

del enlace carbonilo que se romperá (Figura 1.7a). Luego se forma un complejo tetraédrico el cual es estabilizado por el hueco del oxianión (Figura 1.7b). A continuación el complejo es acilado por la enzima y el primer producto es liberado. La hidrólisis del complejo aciloenzima da como resultado la formación del segundo intermediario tetraédrico (Figura 1.7c). Luego el intermediario colapsa y el ácido es liberado regenerando la enzima (Figura 1.7d) (Leung et al., 2000).

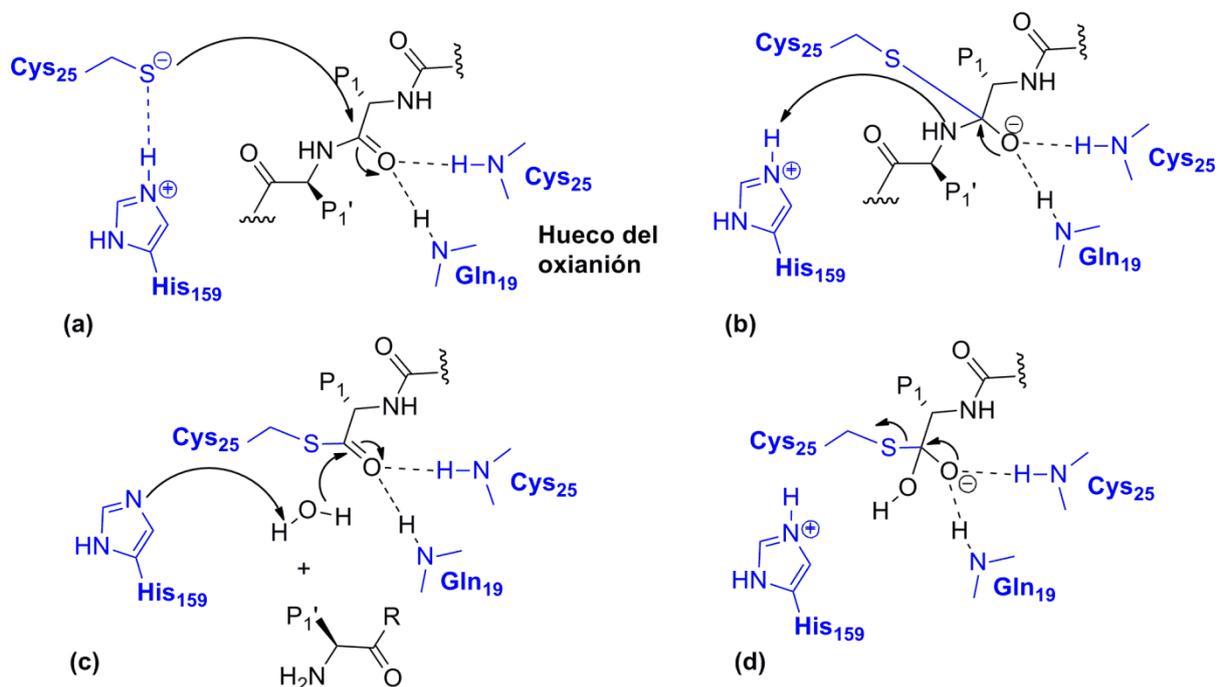


Figura 1.7. Mecanismo de acción propuesto para cisteína-proteasas.

En cuanto a la relevancia funcional de Cz, la misma juega un rol importante en diversas funciones del parásito tales como la invasión de las células huésped, degradando tejidos del huésped y facilitando la penetración del parásito en las células (Aparicio et al., 2004) y participando en la evasión de la respuesta inmune por diferentes mecanismos (Benitez-Hernandez et al., 2010). La participación de la enzima en la proliferación de epimastigotes y amastigotes y en la metaciclo génesis también ha sido demostrada (Santos et al., 2005; Tomas and Kelly, 1996); por otro lado, dada su función autocatalítica, la inhibición de Cz conduce a la acumulación de un precursor en el aparato de Golgi, provocando importantes alteraciones morfológicas, dando lugar

eventualmente a un shock osmótico y a la muerte de *T. cruzi*; estos cambios morfológicos en el aparato de Golgi se han observado en parásitos tratados con diferentes inhibidores, lo cual sugiere que independientemente del mecanismo de inhibición, todos los inhibidores de Cz producen el mismo efecto (Capaci Rodrigues et al., 2010; McKerrow et al., 2009).

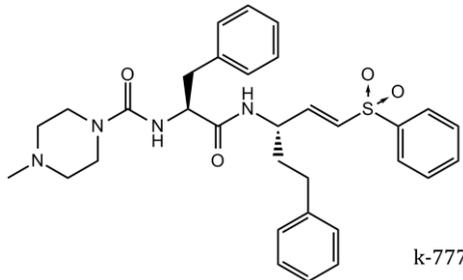
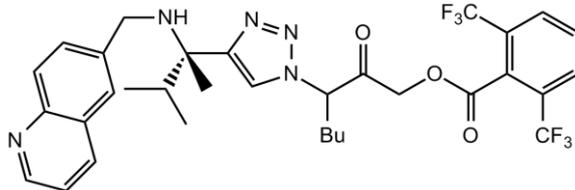
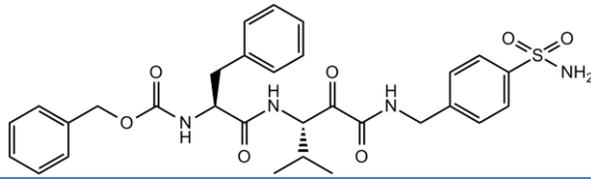
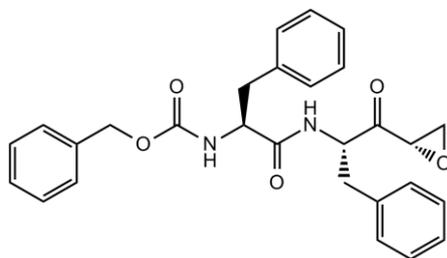
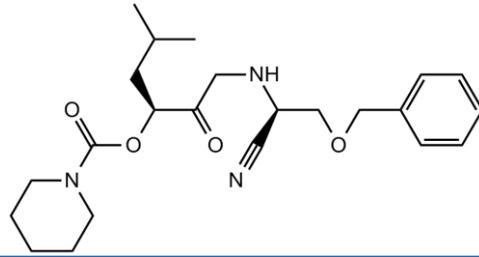
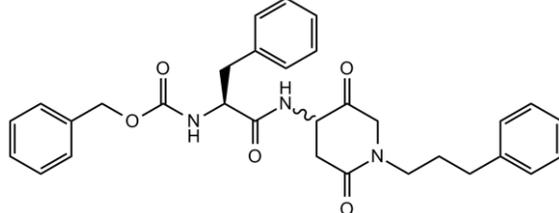
1.3.2. INHIBIDORES DE CRUZIPAÍNA

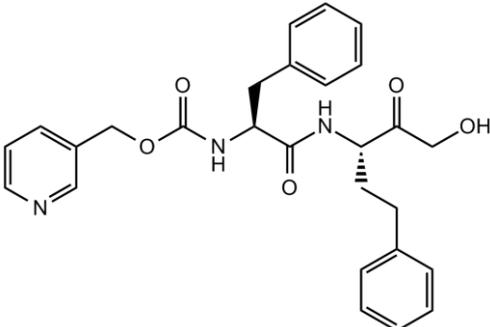
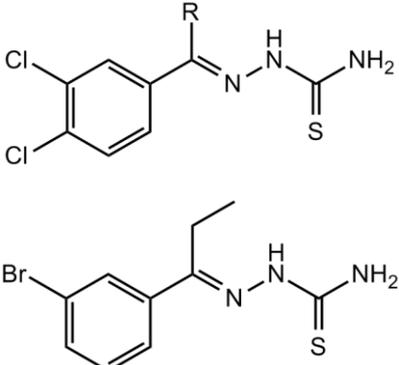
En los últimos quince años diversos grupos de investigación han enfocado sus proyectos en desarrollar inhibidores de Cz (Capaci Rodrigues et al., 2010; Duschak and Couto, 2009; McKerrow et al., 2009; WHO, 2007). La mayoría de las patentes de potenciales antichagásicos registradas entre 2000 y 2006 corresponden a inhibidores de cisteína proteasas (Duschak and Couto, 2007).

Si bien en un principio las investigaciones relacionadas a inhibidores de Cz se enfocaban en inhibidores irreversibles, generalmente peptidomiméticos, se encontraron dificultades principalmente vinculadas a la carencia de selectividad, alta reactividad, toxicidad y baja biodisponibilidad oral de los mismos (Capaci Rodrigues et al., 2010; Jacobsen et al., 2000; Mallari et al., 2008; McKerrow et al., 1999; Powers et al., 2002). Por tal motivo, muchas investigaciones actuales dirigen sus esfuerzos a inhibidores de naturaleza no peptídica, tipo fármaco, preferentemente reversibles.

Algunos ejemplos de inhibidores de Cz que han sido desarrollados y patentados hasta el momento se observan en la Tabla 2. k-777 ha sido ensayado en diversos modelos animales de Chagas agudo (Doyle et al., 2007; Engel et al., 1998).

Tabla 1.2. Ejemplos de Inhibidores reversibles e irreversibles de Cz.

<i>Familia</i>	<i>Estructura</i>	<i>Inhibición</i>
Vinilsulfonas	 <p style="text-align: right;">k-777</p>	Irreversible (Chen et al., 2008)
Aciloxi metilcetonas		Irreversible no peptidico (Brak et al., 2010)
Alpha-dicarbonilos		Análogo de estado de Transición (Choe et al., 2005)
Peptidil cetona		Irreversible (Hernandez and Roush, 2002)
Nitrilos		Reversible (Duschak and Couto, 2007; Mott et al., 2010)
Cetonas cíclicas		Reversible competitivo (Duschak and Couto, 2007; Huang et al., 2003)

Hidroximetilcetonas		Reversible competitivo (Duschak and Couto, 2007; Huang et al., 2003)
Tiosemicarbazonas		Inhibidor Irreversible (Fujii et al., 2005; Siles et al., 2006)

A pesar de la gran variedad estructural de los inhibidores de Cz que se conocen, sólo un inhibidor de Cz ha completado la fase de ensayos preclínicos: la vinilsulfona k-777.

A la fecha no hay evidencia definitiva y publicada de cura parasitológica inducida por este tipo de compuestos en cualquiera de los modelos animales utilizados. Un estudio pre-clínico en un modelo canino de la enfermedad de Chagas aguda indicó que, aunque el tratamiento con k-777 fue incapaz de curar a los animales infectados, redujo significativamente el daño cardíaco inducido por *T. cruzi* (Barr et al., 2005).

En 2002, el Instituto *One World Health* (iOWH) inició el desarrollo de k-777 como nuevo tratamiento potencial para la enfermedad de Chagas, pero en 2005 anunció que daba por finalizado este proyecto, debido a la detección de hepatotoxicidad y graves problemas con el escalado de este compuesto. La hepatotoxicidad se adjudicó a la elevación de los niveles de alanina aminotransferasa, una enzima que se utiliza como biomarcador reversible para la medida del daño en los hepatocitos, aunque a concentraciones mucho mayor que las incluidas en el rango terapéutico (McKerrow et al., 2009).

Actualmente la licencia de k-777 la posee el *National Institute of Allergy and Infectious Diseases* (NIAID) y el *National Institute of Health* (NIH). El proyecto coordinado por DNDi tiene como objetivo principal llevar a cabo los estudios preclínicos requeridos con el fin de completar la etapa de investigación del nuevo fármaco y continuar con la evaluación clínica de k777 para el tratamiento de la enfermedad. De acuerdo al último reporte de DNDi de octubre de 2013, se completaron los estudios de seguridad farmacológica, y no se observaron efectos en el electrocardiograma (ECG) o la función respiratoria, incluso a dosis alta. Los estudios de toxicidad se realizarán en monos y se espera contar con resultados en 2014, con la intención de continuar con los estudios de clínicos de fase I (DNDi, 2013).

REFERENCIAS CAPÍTULO 1

Altcheh J., Moscatelli G., Moroni S., Garcia-Bournissen F., Freilij H. Adverse events after the use of benznidazole in infants and children with Chagas disease. *Pediatrics*. **127**, 212-218. (2011).

Aparicio I.M., Scharfstein J., Lima A.P. A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infect Immun*. **72**, 5892-5902. (2004).

Bahia, M. T.; de Andrade, I. M.; Martins, T. A.; do Nascimento, A. F.; Diniz Lde, F.; Caldas, I. S.; Talvani, A.; Trunz, B. B.; Torreele, E.; Ribeiro, I., Fexinidazole: a potential new drug candidate for Chagas disease. *PLoS Negl Trop Dis*. **6** (11), e1870. (2012)

Barr S.C., Warner K.L., Kornreic B.G., Piscitelli J., Wolfe A., Benet L., McKerrow J.H. A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi*. *Antimicrob Agents Chemother*. **49**, 1560-1561. (2005).

Benitez-Hernandez I., Mendez-Enriquez E., Ostoa P., Fortoul T., Ramirez J.A., Stempin C., Cerban F., Soldevila G., Garcia-Zepeda E.A. Proteolytic cleavage of chemokines by *Trypanosoma cruzi*'s cruzipain inhibits chemokine functions by promoting the generation of antagonists. *Immunobiology*. **215**, 413-426. (2010).

Bern C., Montgomery S.P., Herwaldt B.L., Rassi A., Jr., Marin-Neto J.A., Dantas R.O., Maguire J.H., Acquatella H., Morillo C., Kirchhoff L.V., Gilman R.H., Reyes P.A., Salvatella R., Moore A.C. Evaluation and treatment of chagas disease in the United States: a systematic review. *JAMA*. **298**, 2171-2181. (2007).

Boiani M., Piacenza L., Hernandez P., Boiani L., Cerecetto H., Gonzalez M., Denicola A. Mode of action of nifurtimox and N-oxide-containing heterocycles against *Trypanosoma cruzi*: is oxidative stress involved? *Biochem Pharmacol*. **79**, 1736-1745. (2010).

Brak K., Kerr I.D., Barrett K.T., Fuchi N., Debnath M., Ang K., Engel J.C., McKerrow J.H., Doyle P.S., Brinen L.S., Ellman J.A. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for Chagas disease chemotherapy. *J Med Chem*. **53**, 1763-1773. (2010).

Brener Z. Biology of *Trypanosoma cruzi*. *Annu Rev Microbiol*. **27**, 347-382. (1973).

Buckner, F. S.; Urbina, J. A., Recent Developments in Sterol 14-demethylase Inhibitors for Chagas Disease. *Int J Parasitol Drugs Drug Resist*. **2**, 236-242. (2012).

Bustamante, J. M.; Tarleton, R. L., Potential new clinical therapies for Chagas disease. *Expert Rev Clin Pharmacol*. **7** (3), 317-325. (2014)

Caffrey C.R., Scory S., Steverding D. Cysteine proteinases of trypanosoma parasites: novel targets for chemotherapy. *Curr Drug Targets*. **1**, 155-162. (2000).

Campetella O., Henriksson J., Aslund L., Frasch A.C., Pettersson U., Cazzulo J.J. The major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* is encoded by multiple polymorphic

tandemly organized genes located on different chromosomes. *Mol Biochem Parasitol.* **50**, 225-234. (1992).

Capaci Rodrigues G., Palermo Aguiar A., da Silva Goncalves Vianez J.L.J., Macrae A., Nogueira de Melo A.C., Vermelho A.B. Peptidase inhibitors as a possible therapeutic strategy for Chagas disease. *Curr Enz Inhib.* **6**, 183-195. (2010).

Carlier Y., Torrico F., Sosa-Estani S., Russomando G., Luquetti A., Freilij H., Albajar Vinas P. Congenital Chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women. *PLoS Negl Trop Dis.* **5**, e1250. (2011).

Castro J.A. Toxic side effects of drugs used to treat Chagas disease (American Trypanosomiasis). *Hum Exp Toxicol.* **25**, 471-479. (2006).

Cazzulo J.J. Proteinases of *Trypanosoma cruzi*: potential targets for the chemotherapy of Chagas disease. *Curr Top Med Chem.* **2**, 1261-1271. (2002).

Chagas C. Nova tripanozomiose humana. *Mem Inst Oswaldo Cruz.* **1**, 159-218. (1909).

Chen Y.T., Lira R., Hansell E., McKerrow J.H., Roush W.R. Synthesis of macrocyclic trypanosomal cysteine protease inhibitors. *Bioorg Med Chem Lett.* **18**, 5860-5863. (2008).

Choe Y., Brinen L.S., Price M.S., Engel J.C., Lange M., Grisostomi C., Weston S.G., Pallai P.V., Cheng H., Hardy L.W., Hartsough D.S., McMakin M., Tilton R.F., Baldino C.M., Craik C.S. Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorg Med Chem.* **13**, 2141-2156. (2005).

Clayton J. Chagas disease 101. *Nature.* **465**, S4-5. (2010).

Coura J.R., Dias J.C. Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Mem Inst Oswaldo Cruz.* **104 Suppl 1**, 31-40. (2009).

Croft S.L., Barrett M.P., Urbina J.A. Chemotherapy of trypanosomiasis and leishmaniasis. *Trends Parasitol.* **21**, 508-512. (2005).

Dias E. O signal de Romaña na molestia de Chagas. *Acta Medica.* **3**, 60-62. (1939).

DNDi. Portfolio K777 (2013), <http://www.dndi.org/diseasesprojects/portfolio/k777.html?highlight=WyJrNzc3Il0=> (Último acceso, mayo 2014).

Doyle P.S., Zhou Y.M., Engel J.C., McKerrow J.H. A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrob Agents Chemother.* **51**, 3932-3939. (2007).

Duschak V.G., Couto A.S. An insight on targets and patented drugs for chemotherapy of Chagas disease. *Recent Pat Antiinfect Drug Discov.* **2**, 19-51. (2007).

Duschak V.G., Couto A.S. Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Curr Med Chem.* **16**, 3174-3202. (2009).

Duschak V.G., Barboza M., Couto A.S. *Trypanosoma cruzi*: partial characterization of minor cruzipain isoforms non-adsorbed to Concanavalin A-Sepharose. *Exp Parasitol.* **104**, 122-130. (2003).

Engel J.C., Doyle P.S., Hsieh I., McKerrow J.H. Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *J Exp Med.* **188**, 725-734. (1998).

Fujii N., Mallari J.P., Hansell E.J., Mackey Z., Doyle P., Zhou Y.M., Gut J., Rosenthal P.J., McKerrow J.H., Guy R.K. Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain. *Bioorg Med Chem Lett.* **15**, 121-123. (2005).

Gomes S.A., Misael D., Silva B.A., Feder D., Silva C.S., Goncalves T.C., Santos A.L., Santos-Mallet J.R. Major cysteine protease (cruzipain) in Z3 sylvatic isolates of *Trypanosoma cruzi* from Rio de Janeiro, Brazil. *Parasitol Res.* **105**, 743-749. (2009).

Haberland A., Saravia S.G., Wallukat G., Ziebig R., Schimke I. Chronic Chagas disease: from basics to laboratory medicine. *Clin Chem Lab Med.* **51**, 271-294. (2013).

Hall B.S., Bot C., Wilkinson S.R. Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites. *J Biol Chem.* **286**, 13088-13095. (2011).

Hardman J.G., Limbird L.E. Goodman & Gilman's. The pharmacological basis of therapeutics. 9 ed. McGraw-Hill, New York. (2002)

Hernandez A.A., Roush W.R. Recent advances in the synthesis, design and selection of cysteine protease inhibitors. *Curr Opin Chem Biol.* **6**, 459-465. (2002).

Howard E.J., Xiong X., Carlier Y., Sosa-Estani S., Buekens P. Frequency of the congenital transmission of *Trypanosoma cruzi*: a systematic review and meta-analysis. *BJOG.* **121**, 22-33. (2014).

Huang L., Brinen L.S., Ellman J.A. Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorg Med Chem.* **11**, 21-29. (2003).

Jacobsen W., Christians U., Benet L.Z. In vitro evaluation of the disposition of A novel cysteine protease inhibitor. *Drug Metab Dispos.* **28**, 1343-1351. (2000).

Lee B.Y., Bacon K.M., Bottazzi M.E., Hotez P.J. Global economic burden of Chagas disease: a computational simulation model. *Lancet Infect Dis.* **13**, 342-348. (2013).

Lepesheva, G. I., Design or screening of drugs for the treatment of Chagas disease: what shows the most promise? *Expert Opin Drug Discov.* **8** (12), 1479-1489. (2013).

Leung D., Abbenante G., Fairlie D.P. Protease inhibitors: current status and future prospects. *J Med Chem.* **43**, 305-341. (2000).

Lima A.P., dos Reis F.C., Serveau C., Lalmanach G., Juliano L., Menard R., Vernet T., Thomas D.Y., Storer A.C., Scharfstein J. Cysteine protease isoforms from *Trypanosoma cruzi*, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. *Mol Biochem Parasitol.* **114**, 41-52. (2001).

Mallari J.P., Shelat A.A., O'Brien T., Caffrey C.R., Kosinski A., Connelly M., Harbut M., Greenbaum D., McKerrow J.H., Guy R.K. Development of potent purine-derived nitrile inhibitors of the trypanosomal protease TbcatB. *J Med Chem.* **51**, 545-552. (2008).

Maya J.D., Cassels B.K., Iturriaga-Vasquez P., Ferreira J., Faundez M., Galanti N., Ferreira A., Morello A. Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. *Comp Biochem Physiol A Mol Integr Physiol.* **146**, 601-620. (2007).

Mazza S. La enfermedad de Chagas en la República Argentina. (1939), VI Congreso Nacional de Medicina, Ed. Pomponio, Rosario, Argentina. pp. 9-20.

Mazza S. Sobre el desacierto de adoptar un subtítulo para la Enfermedad de Chagas. *La Prensa Medica.* **30**, 433-440. (1943).

McGrath M.E., Eakin A.E., Engel J.C., McKerrow J.H., Craik C.S., Fletterick R.J. The crystal structure of cruzain: a therapeutic target for Chagas' disease. *J Mol Biol.* **247**, 251-259. (1995).

McKerrow J.H., Engel J.C., Carey C.R. Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med Chem.* **7**, 639-644. (1999).

McKerrow J.H., Doyle P.S., Engel J.C., Podust L.M., Robertson S.A., Ferreira R., Saxton T., Arkin M., Kerr I.D., Brinen L.S., Craik C.S. Two approaches to discovering and developing new drugs for Chagas disease. *Mem Inst Oswaldo Cruz.* **104 Suppl 1**, 263-269. (2009).

Mecca M.M., Bartel L.C., Castro C.R., Castro J.A. Benznidazole biotransformation in rat heart microsomal fraction without observable ultrastructural alterations: comparison to Nifurtimox-induced cardiac effects. *Mem Inst Oswaldo Cruz.* **103**, 549-553. (2008).

Morillo C.A. Study of oral Posaconazole in the treatment of asymptomatic Chronic Chagas disease. (2012), VIII taller sobre la enfermedad de Chagas importada. Avances en el tratamiento antiparasitario, Barcelona, España.

Mott B.T., Ferreira R.S., Simeonov A., Jadhav A., Ang K.K., Leister W., Shen M., Silveira J.T., Doyle P.S., Arkin M.R., McKerrow J.H., Inglese J., Austin C.P., Thomas C.J., Shoichet B.K., Maloney D.J. Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB. *J Med Chem.* **53**, 52-60. (2010).

Murcia L., Carrilero B., Segovia M. Limitations of currently available Chagas disease chemotherapy. *Rev Esp Quimioter.* **25**, 1-3. (2012).

Murcia L., Carrilero B., Saura D., Iborra M.A., Segovia M. [Diagnosis and treatment of Chagas disease]. *Enferm Infecc Microbiol Clin.* **31 Suppl 1**, 26-34. (2013).

Nascimento A.E., de Souza W. High resolution localization of cruzipain and Ssp4 in *Trypanosoma cruzi* by replica staining label fracture. *Biol Cell*. **86**, 53-58. (1996).

Nunes M.C., Dones W., Morillo C.A., Encina J.J., Ribeiro A.L. Chagas disease: an overview of clinical and epidemiological aspects. *J Am Coll Cardiol*. **62**, 767-776. (2013).

Parker E.R., Sethi A. Chagas disease: coming to a place near you. *Dermatol Clin*. **29**, 53-62. (2011).

Pinto Dias J.C. Human chagas disease and migration in the context of globalization: some particular aspects. *J Trop Med*. **2013**, 789758. (2013).

Powers J.C., Asgian J.L., Ekici O.D., James K.E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev*. **102**, 4639-4750. (2002).

Rassi A., Jr., Rassi A., Marin-Neto J.A. Chagas disease. *Lancet*. **375**, 1388-1402. (2010).

Rodrigues Coura J., de Castro S.L. A critical review on Chagas disease chemotherapy. *Mem Inst Oswaldo Cruz*. **97**, 3-24. (2002).

Santos C.C., Sant'anna C., Terres A., Cunha-e-Silva N.L., Scharfstein J., de A.L.A.P. Chagasin, the endogenous cysteine-protease inhibitor of *Trypanosoma cruzi*, modulates parasite differentiation and invasion of mammalian cells. *J Cell Sci*. **118**, 901-915. (2005).

Siles R., Chen S.E., Zhou M., Pinney K.G., Trawick M.L. Design, synthesis, and biochemical evaluation of novel cruzain inhibitors with potential application in the treatment of Chagas' disease. *Bioorg Med Chem Lett*. **16**, 4405-4409. (2006).

Sordelli A. Tripanosomiasis americana. (1939), X Conferencia Sanitaria Panamericana, Oficina Sanitaria Panamericana, Bogota. pp. 450.

Souto-Padron T., Campetella O.E., Cazzulo J.J., de Souza W. Cysteine proteinase in *Trypanosoma cruzi*: immunocytochemical localization and involvement in parasite-host cell interaction. *J Cell Sci*. **96 (Pt 3)**, 485-490. (1990).

Storino R. La cara oculta de la enfermedad de Chagas. *Rev Fed Arg Cardiol*. **29**, 31-44. (2000).

Storino R. Poder y dinero en la enfermedad de Chagas: una histórica omisión. *Salud Colectiva*. **8**, 23-25. (2012).

Tarral, A.; Blesson, S.; Mordt, O. V.; Torreele, E.; Sassella, D.; Bray, M. A.; Hovsepian, L.; Evenc, E.; Gualano, V.; Felices, M.; Strub-Wourgaft, N., Determination of an optimal dosing regimen for fexinidazole, a novel oral drug for the treatment of human African trypanosomiasis: first-in-human studies. *Clin Pharmacokinet*. **53 (6)**, 565-580. (2014)

Tomas A.M., Kelly J.M. Stage-regulated expression of cruzipain, the major cysteine protease of *Trypanosoma cruzi* is independent of the level of RNA1. *Mol Biochem Parasitol*. **76**, 91-103. (1996).

Tyler K.M., Engman D.M. The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol*. **31**, 472-481. (2001).

Urbina J.A., Docampo R. Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol.* **19**, 495-501. (2003).

Viotti R., Vigliano C., Lococo B., Alvarez M.G., Petti M., Bertocchi G., Armenti A. Side effects of benznidazole as treatment in chronic Chagas disease: fears and realities. *Expert Rev Anti Infect Ther.* **7**, 157-163. (2009).

W.H.O. World Health Organization on behalf of the Special Programme for Research and Training in Tropical Diseases. (2007).

W.H.O. Research Priorities for Chagas Disease, Human African Trypanosomiasis and Leishmaniasis. (2012).

W.H.O. Chagas disease (american tripanosomiasis). *Fact sheet nro. 340*. (2013).

CAPÍTULO 2

CRIBADO VIRTUAL



2.1. DEFINICIÓN

La cantidad de compuestos químicos conocidos ha aumentado exponencialmente en el curso de los últimos cincuenta años (Binetti et al., 2008). En 1960 el número de compuestos orgánicos e inorgánicos registrados en el *Chemical Abstract Service* (CAS) ascendía a algo más de 200000 compuestos; en 1990, alcanzaba alrededor de 10 millones de entidades químicas, mientras que en 2009 ascendía ya a 50 millones. Actualmente, el servicio ha indexado más de 87 millones de compuestos (<http://www.cas.org/>, último acceso en mayo de 2014). Aproximadamente la mitad de los mismos comprende compuestos orgánicos pequeños, tipo fármaco (la base de datos Pubchem, orientada a pequeñas moléculas orgánicas, contiene actualmente 48 millones de entradas). En este escenario, el ensayo sistemático exhaustivo de los posibles candidatos resulta inviable y surge el interrogante de cómo explorar eficientemente semejante diversidad química en busca de fármacos que interactúen selectivamente con una o más dianas moleculares. Una solución a esta problemática ha sido el desarrollo de métodos del *cribado de alto rendimiento* (en inglés, *high throughput screening*, HTS), un procedimiento en el que se combinan la robótica, la miniaturización de ensayos y el procesamiento de datos a gran escala para identificar con rapidez y mínimo gasto de muestra compuestos que modulen una diana molecular concreta (actualmente, de hecho, se utiliza el término cribado de ultra-alto rendimiento para referirse a plataformas que permiten evaluar más de 100000 compuestos químicos por día) (Mayr and Bojanic, 2009). El cribado de alto rendimiento requiere, sin embargo, acceso a tecnologías de altísimo costo que limita su aplicabilidad en la región latinoamericana. Como una alternativa más económica y racional para explorar la enorme diversidad química existente se plantea el *Cribado Virtual*.

El CV comprende un conjunto heterogéneo de técnicas computacionales o algoritmos que permiten priorizar qué compuestos de una base de datos o biblioteca química digital se ensayarán experimentalmente en modelos *in vitro* y/o *in vivo* (Reddy et al., 2007; Talevi et al., 2011). Tales modelos o algoritmos se aplican a manera de filtro o tamiz: sólo aquellos compuestos de la base de datos que superen el filtro (es decir, que superen

un valor de *score* determinado para el o los modelos o algoritmos) se ensayarán experimentalmente.

Entre las ventajas de las técnicas de CV se pueden mencionar:

 **Tiempo-Eficiencia:** el estado de la técnica permite explorar bases de datos de varios millones de estructuras químicas en períodos de tiempo del orden de semanas (Muegge and Oloff, 2006).

 **Costo-Eficiencia:** Una vez que un laboratorio o grupo de investigación ha desarrollado o adquirido una base de datos y el software necesario para aplicar una metodología de CV, el costo se reduce simplemente al tiempo que el proceso de CV demanda al personal especializado, por lo cual es un proceso relativamente poco costoso. Existen muchas bases de datos públicas o de libre acceso, como Pubchem (Baycoucheva, 2007), Zinc (Irwin and Shoichet, 2005) o DrugBank (Law et al., 2014); por otra parte, numerosas aplicaciones para desarrollar el proceso de CV son de acceso libre. Podemos mencionar por ejemplo, PowerMv (Liu et al., 2005) (software que permite realizar diversos cálculos de similitud química), PaDEL (Yap, 2011) y Open Babel (O'Boyle et al., 2011) (que permiten el cálculo de descriptores moleculares y *fingerprints*, entre otros). La mayoría de esas aplicaciones pueden ser ejecutadas adecuadamente en una computadora de escritorio de última generación. En otras palabras, los requerimientos tecnológicos para el desarrollo del CV son bastante accesibles, aún para grupos de investigación del ámbito académico de nuestra región.

 **Carácter teórico:** A diferencia del HTS, que requiere disponer de muy pequeñas cantidades de los compuestos evaluados, el CV, por su carácter teórico, no requiere *a priori* de una muestra de los compuestos de la base de datos. La adquisición o síntesis se realiza con posterioridad a la selección computacional de un número restringido de compuestos de la base de datos; solamente se ensayan compuestos promisorios que son señalados como candidatos por los modelos o algoritmos utilizados.

 **Segundos Usos:** En ciertos casos, el CV puede enfocarse en el descubrimiento de nuevas indicaciones terapéuticas para fármacos ya conocidos, estrategia que

permite responder a necesidades terapéuticas de la población en menor tiempo, debido a que es posible prescindir de ensayos preclínicos de toxicidad y ensayos clínicos orientados a establecer la seguridad y perfil farmacocinético del fármaco.

Durante los últimos años la comunidad químico-farmacéutica internacional ha enfocado su atención en la búsqueda de segundos usos de fármacos o reposicionamiento de fármacos (en inglés *drug repurposing* o *drug reprofiling*). Existe evidencia empírica de los posibles beneficios terapéuticos y económicos de la búsqueda de nuevas indicaciones de medicamentos conocidos (Ma et al., 2013; Oprea et al., 2011; Tobinick, 2009). El CV puede servir como una herramienta para ayudar a investigar sobre los segundos usos, desde una perspectiva racional y sistemática, tamizando bases de datos de compuestos con antecedentes de uso en clínica. En el siguiente capítulo se discutirá con detalle la estrategia de reposicionamiento y la aplicación del CV para ese fin.



Bioética: El CV responde a la filosofía de las pautas bioéticas vigentes a nivel internacional que proponen utilizar modelos computacionales o *in vitro* previamente a los ensayos *in vivo*, con el objeto de minimizar y optimizar los ensayos en modelos animales (Claude, 2009).

En la Figura 2.1 puede observarse el aumento del uso de las técnicas de CV en los últimos años; se denota el número de artículos de revistas científicas que llevan en su título o entre sus palabras claves el término “Virtual Screening” según el reconocido buscador Scopus ([www.scopus.com/scopus/ home.url](http://www.scopus.com/scopus/home.url) - acceso por suscripción). La búsqueda se restringió a revistas científicas (*journals*) empleando como término de búsqueda “Virtual Screening” entrecomillado.

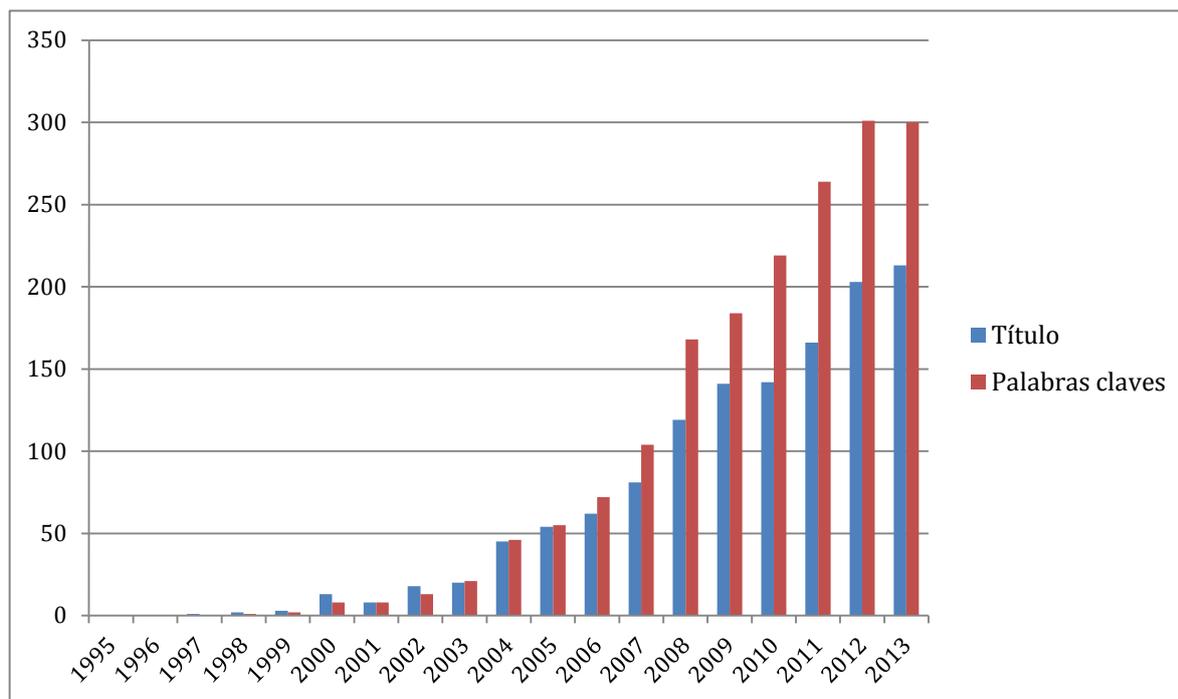


Figura 2.1. Número de artículos publicados en revistas científicas entre 1995 y 2013 que contienen el término “Virtual Screening” en el título y entre las palabras claves.

2.2. CLASIFICACIÓN DE LAS TÉCNICAS DE CRIBADO VIRTUAL

En general las técnicas de CV se clasifican en dos grandes grupos dependiendo de si se enfocan en el receptor o en el ligando.

Recurriendo al clásico paradigma de “llave y cerradura”, el CV basado en el receptor implica investigar las características estructurales del blanco terapéutico (la cerradura), pudiendo ser éste una enzima, un receptor o un blanco molecular de otra naturaleza, a fin de encontrar nuevos ligandos (llaves) que produzcan la actividad o respuesta buscada. El CV enfocado desde el ligando consiste en buscar “llaves” que “abren” la cerradura de interés, utilizando un modelo, patrón o mínimo denominador común estructural.

2.2.1. METODOLOGÍAS DE CV BASADAS EN EL RECEPTOR

Habitualmente conocidas como técnicas de acoplamiento o *docking*, comprenden tres pasos generales (Ma et al., 2013), a) generación de un modelo molecular del receptor (se

requiere el conocimiento de la estructura tridimensional del blanco molecular obtenida por rayos X, RMN o por modelado por homología), b) pre-tratamiento y muestreo conformacional de los ligandos y c) obtención de una puntuación o “score” asociado a la energía de interacción de cada potencial ligando de la base de datos al sitio activo del blanco molecular de elección.

El procedimiento de *docking* es utilizado para explorar y predecir los posibles modos de formación de complejos ligando-blanco molecular (Sousa et al., 2006). El proceso por el cual un ligando se une a su blanco molecular depende de factores entrópicos y entálpicos.

De entre todas las técnicas conocidas que pueden utilizarse en una campaña de CV, los métodos de docking proveen claramente la mejor descripción de la unión del ligando a su blanco. Sin embargo, se trata de un método muy exigente computacionalmente. Si deseamos realizar un CV sobre grandes bases de datos, las conformaciones evaluadas de los diferentes ligandos deben ser limitadas si la búsqueda tiene que ser completada en un período de tiempo razonable.

2.2.2. TÉCNICAS DE CV BASADAS EN EL LIGANDO

Podemos clasificar a las técnicas de CV basadas en el ligando en 3 grandes grupos: metodologías basadas en la similitud química, las técnicas de superposición con farmacóforos y aquellas basadas en modelos multivariable basados en descriptores.

a) Técnicas basadas en la similitud química

Consisten en medir el grado de similitud entre una estructura de referencia y cada uno de los compuestos que integran una base de datos utilizando alguna definición cuantitativa de la similitud estructural (Bajorath, 2001).

La ventaja de las técnicas basadas en la similitud es que poseen un bajo costo computacional, por lo cual la búsqueda en grandes bibliotecas de compuestos puede realizarse en un tiempo relativamente corto; además el CV por similitud puede

desarrollarse incluso con una sola molécula de referencia, de manera que dicha metodología resulta sumamente útil para aquellas situaciones en las que -ante un blanco molecular novedoso- se dispone de un número limitado de ligandos conocidos, el cual es insuficiente para aplicar técnicas como el modelado QSAR (Holliday et al., 2011). Esencialmente, se asocia a cada estructura comparada una secuencia de bits (*fingerprints*, huella digital) en la que cada bit indica la ausencia o presencia de determinado fragmento o característica molecular (u ocasionalmente, la frecuencia con la que cada característica aparece en las estructuras comparadas). Luego, se cuantifican cuántos bits tienen en común las dos estructuras, recurriendo para ello a un coeficiente de similitud.

Las variantes a tener en cuenta en una búsqueda por similitud son (Willett, 2006): a) el sistema de fragmentos, subestructuras o características moleculares utilizados para la comparación de las moléculas; b) la medida empleada para establecer la comparación cuantitativa; c) si se considera la frecuencia en la que cada fragmento o característica molecular aparece en cada una de las moléculas comparadas, o sólo la presencia o ausencia de determinada característica. Ocasionalmente, puede definirse algún criterio para ponderar la contribución de cada tipo de subestructura, y utilizar múltiples estructuras de referencia para conducir la búsqueda. Estas comparaciones permiten ordenar las moléculas de la biblioteca química según un orden decreciente de similitud con respecto a la molécula de referencia, de manera que los compuestos más similares a la estructura de referencia aparecen en los primeros lugares del ranking (Girschick et al., 2013; Hert et al., 2004).

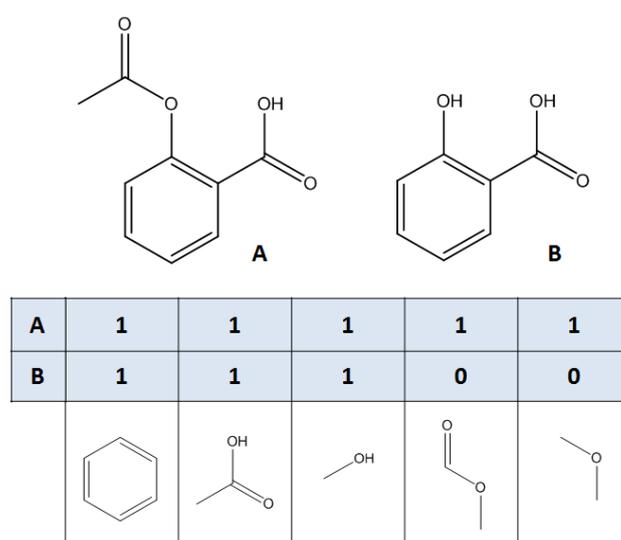
Las metodologías de similitud molecular más empleadas comprenden huellas digitales 2D binarias (2D binary fingerprints) en la cual cada bit denota la presencia o ausencia de un fragmento o una característica conformación-independiente para determinar la similitud inter-estructural (Maggiore et al., 2014; Willett, 2006).

Si bien existen diversos coeficientes de similitud, el más profusamente utilizado es el coeficiente de Tanimoto, también conocido como coeficiente de Jaccard o -en el caso de medir su complemento, la disimilitud- coeficiente de Soergel (Willett et al., 1998). El coeficiente de Tanimoto, en su versión binaria, se define como:

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

donde a representa el número de tipos de subestructuras o características presentes en el compuesto A ; b el número de tipos de subestructuras o características presentes en el compuesto B y c el número de tipos de subestructuras o características que se encuentran en ambos compuestos. Por lo tanto, el valor del coeficiente de Tanimoto oscila entre 0 y 1, donde 0 representa máxima disimilitud (es decir ningún bit o tipo de subestructura en común) y 1 representa máxima similitud (todos los bits de la huella dactilar en común). Este coeficiente constituye actualmente la medida de elección de la mayoría de los sistemas de software comerciales y también forma parte de los repositorios de acceso público más importantes, como por ejemplo PubChem (<https://pubchem.ncbi.nlm.nih.gov/>. Último acceso, junio 2014).

En la Figura 2.2, se presenta un ejemplo de cálculo del coeficiente de Tanimoto y en la tabla 2.1. puede observarse otros ejemplos de coeficientes de similitud.



a = número de bits encendidos en la estructura **A** = 5

b = número de bits encendidos en la estructura **B** = 3

c = número de bits encendidos en **A** y **B** = 3

$$\text{Coeficiente de Tanimoto} = c / (a + b - c) = 3 / (5 + 3 - 3) = 0.60$$

Figura 2.2. Ejemplo de cálculo del coeficiente de Tanimoto.

Tabla 2.1. Ejemplos de coeficientes de similitud.

Coeficiente	Expresión
Tanimoto	$\frac{c}{a + b - c}$
Cosine	$\frac{c}{\sqrt{ab}}$
Hamming	$a + b - 2c$
Russell-Rao	c/m
Forbes	cm/ab

Suponiendo que dos moléculas *A* y *B* están representadas por fingerprints 2D de naturaleza binaria (cada bit indica presencia -1- o ausencia -0- de la subestructura asociada al mismo) conteniendo un total de *m* bits (o subestructuras); asumiendo además que *a* y *b* corresponden al número de bits encendidos (iguales a 1) en las estructuras *A* y *B* respectivamente, y *c* es el número de bits encendidos en ambas estructuras.

Nota: el coeficiente de Hamming es una distancia, por lo cual los valores mayores corresponden a pares de moléculas disímiles en lugar de pares similares como ocurre con otros coeficientes.

b) Técnicas basadas en la superposición de ligandos

Los ligandos que conforman la base de datos se superponen con una molécula de referencia (o con un farmacóforo si el mismo ya fue identificado) en función de la disposición espacial de ciertas características moleculares relevantes para la interacción ligando-blanco molecular. Comúnmente las características que se superponen son difusas (es decir, pueden ser provistas por distintas funciones o grupos químicos) e involucran grupos de átomos con cargas positiva y negativa, dadores y aceptores de puente de hidrógeno y regiones o sustituyentes con carácter hidrofóbico. Los métodos de superposición, consideran la geometría 3D y habitualmente evalúan la diferencia entre dos conformaciones (la de menor energía, obtenida por análisis conformacional, y

la conformación activa definida por el farmacóforo); esto implica la necesidad de realizar una búsqueda conformacional para cada estructura de la base de datos a fin de superponer uno o varios conformeros de baja energía, lo cual se traduce en un aumento del costo computacional (Yang, 2010).

c) Técnicas basadas en descriptores

Esta metodología se basa en predecir una categoría o una propiedad biológica de interés a través de Relaciones Cuantitativas Estructura-Actividad (*Quantitative Structure-Activity Relationship* - QSAR). Los métodos *in silico* QSAR pueden ser utilizados para predecir el valor de la actividad o propiedad biológica (ej. IC₅₀) o para clasificar (ej. inhibidores y no inhibidores) compuestos antes de realizar el experimento biológico. Las predicciones de los modelos QSAR tendrán validez en el espacio químico definido por el conjunto de compuestos utilizados para generar los modelos. Ese espacio, dentro del cual las predicciones de un modelo QSAR son confiables, se denomina “dominio de aplicabilidad”. Los modelos QSAR pueden utilizarse para explicar las diferencias en cierta propiedad fisicoquímica o actividad biológica de una serie de compuestos, o para realizar búsquedas en bibliotecas de compuestos químicos (Yee and Wei, 2012).

A continuación se detallan las diferentes clasificaciones de los métodos QSAR; basadas en la dimensionalidad de los descriptores utilizados en el modelado y según la naturaleza de la variable dependiente utilizada.

c1) Clasificación de Modelos QSAR basados en la dimensionalidad de los descriptores

Un **descriptor molecular** se define como una variable numérica que codifica o refleja algún aspecto estructural de la molécula; es el resultado final de un procedimiento lógico-matemático que transforma la información química codificada en la estructura de una molécula en un número o en un resultado de varios experimentos estandarizados. (Todeschini and Consonni, 2009).

De acuerdo a la naturaleza de su definición y la complejidad de los aspectos estructurales que codifican los descriptores moleculares se clasifican en base a las

dimensiones que abarcan: 0D (descriptores adimensionales o constitucionales); 1D (unidimensionales); 2D (bidimensionales o invariantes de grafos); 3D (tridimensionales) y 4D (tetradimensionales). Teniendo en cuenta esta subclasificación podemos englobarlos en:

 *Descriptores independientes de la conformación espacial (3D) de la molécula.*

Dentro de este grupo encontramos a los descriptores moleculares 0D, 1D y 2D. Los descriptores constitucionales se obtienen a partir de su fórmula molecular, por ejemplo el peso molecular, el número de átomos o cualquier función de las propiedades atómicas. Los 1D indican frecuencia o presencia de grupos funcionales o de fragmentos estructurales de la molécula. Los descriptores 2D se basan habitualmente en la representación de la estructura molecular como un grafo, esto es, como un conjunto de vértices y de aristas que conectan los vértices; en un grafo químico los átomos y enlaces están representados por vértices y aristas respectivamente (Pogliani, 2000). Los descriptores topológicos consideran la conectividad entre átomos (vértices) en la molécula pudiendo tener en cuenta la naturaleza de los enlaces químicos y átomos.

El número de descriptores moleculares desarrollados hasta el momento es muy amplio; para una descripción detallada se puede recurrir a la última versión del Manual de Descriptores Moleculares de Todeschini y Consonni (Todeschini and Consonni, 2009).

Las metodologías que utilizan este tipo de descriptores independientes de la conformación espacial se definen como QSAR-2D (Dudek et al., 2006).

 *Descriptores dependientes de la conformación espacial (3D) de la molécula.*

En esta categoría agrupamos a los 3D y 4D; los descriptores 3D derivan de la representación tridimensional de la molécula y se basan no solo en la naturaleza y conectividad de los átomos sino también en la conformación y configuración espacial. Los descriptores 4D examinan el espacio conformacional de los objetos moleculares, son similares a los 3D, pero, con respecto a ellos, la información estructural se considera para un conjunto de conformeros (condicionalmente, la cuarta dimensión), en lugar de una conformación fija (Andrade et al., 2010).

Las metodologías 3D-QSAR son más complejas y costosas computacionalmente respecto a los enfoques 2D-QSAR, requieren una conformación de baja energía para realizar el cálculo de los descriptores, siendo la obtención de estos conformeros 3D el factor más importante para obtener modelos fiables y la principal limitación de los métodos 3D-QSAR en relación particularmente a su aplicación en el CV (Myint and Xie, 2010). Los métodos 3D-QSAR más populares, entre los que se destacan el CoMFA (Cramer et al., 1989), CoMSIA (Klebe and Abraham, 1999) y GRID/GOLPE (Nilsson et al., 1997) se basan en colocar cada molécula optimizada en una grilla y calcular el valor de la fuerza que diferentes campos ejercen sobre los vértices de la misma (colocando generalmente un átomo o molécula de prueba en cada vértice), estableciendo qué campo, y en qué zona de la grilla, es significativo para la actividad biológica. La principal diferencia de estos métodos es el tipo de descriptores que utilizan; CoMFA y CoMSIA calculan distintos campos de interacción molecular, estéricos, electrostáticos o hidrofóbicos y las contribuciones de las energías correspondientes se consideran por separado, en cambio GRID/GOLPE calcula la suma de todas las interacciones involucradas en cada punto de la grilla.

c2) Modelos QSAR basados en la naturaleza de la variable dependiente

Los modelos QSAR pueden utilizar como repuesta o *output* una variable dependiente continua, es decir, pueden emplearse para predecir el valor numérico de la propiedad modelada (por ejemplo IC₅₀, ED₅₀), o pueden generar una respuesta discreta (modelos clasificatorios o discriminantes) cuyo valor se asocia a distintas categorías de objetos (compuestos químicos).

2.3. APROXIMACIÓN EMPLEADA EN EL TRABAJO DE TESIS

En el presente trabajo empleamos, para la búsqueda de fármacos antichagásicos inhibidores de Cz, métodos 2D-QSAR basados en descriptores y enfocados desde el ligando, utilizando una variable discreta categórica. En el capítulo 4 (metodología) se detallan cada una de las etapas de modelado 2D-QSAR y las herramientas utilizadas. Es

importante destacar que la metodología elegida presenta ventajas, pero también limitaciones.

En relación a las ventajas podemos mencionar que la independencia de la conformación espacial de los compuestos en el cálculo de los descriptores 0D-2D permite acortar notablemente los tiempos empleados para el desarrollo de los modelos y fundamentalmente para su aplicación en campañas de CV, dado que las geometrías de las estructuras del conjunto de entrenamiento, del conjunto de prueba y la base de datos tamizada no requieren ser optimizadas, esto es, no necesitan un pre-procesamiento (contrariamente a lo que ocurre en las metodologías 3D-QSAR), sumado a que el costo computacional en el cálculo de los descriptores de baja dimensionalidad es relativamente bajo.

Los modelos 2D-QSAR presentan asimismo una serie de limitaciones, muchos descriptores 2D (en particular, muchos descriptores topológicos) son abstractos y difíciles de interpretar. Debido a ello en general es difícil extraer de un modelo 2D QSAR información significativa desde el punto de vista fisicoquímico acerca de los efectos o interacciones electrónicas, estérica o hidrofóbicas por los cuales un compuesto determinado interacciona con su blanco molecular. La mayoría de los modelos 2D-QSAR son eficientes pero difíciles de interpretar, actuando como una especie de caja negra. En cambio, los modelos 3D-QSAR permiten identificar las interacciones más importantes entre droga y blanco molecular para la actividad o propiedad que se estudia.

En síntesis, las aproximaciones 3D-QSAR describen con mayor claridad (en comparación con las aproximaciones 2D-QSAR) las interacciones droga-blanco molecular pero requieren un pre-procesamiento de las estructuras químicas que supone un obstáculo para el estudio de grandes bases de datos. Las metodologías 2D-QSAR, si bien por su carácter abstracto no dan gran información sobre el proceso de unión al receptor, son más eficientes dada su independencia de la conformación espacial de la molécula y la velocidad de cálculo de descriptores de baja dimensionalidad, lo que las convierte en herramientas ideales para identificar nuevos fármacos en grandes bases de datos de diversidad estructural.

REFERENCIAS CAPÍTULO 2

Andrade C.H., Pasqualoto K.F., Ferreira E.I., Hopfinger A.J. 4D-QSAR: perspectives in drug design. *Molecules*. **15**, 3281-3294. (2010).

Bajorath J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J Chem Inf Comput Sci*. **41**, 233-245. (2001).

Baycoucheva S. A new era in chemical information: PubChem, DiscoveryGate, and chemistry central. *Online (Wilton, Connecticut)*. **31**, 16-20. (2007).

Binetti R., Costamagna F.M., Marcello I. Exponential growth of new chemicals and evolution of information relevant to risk control. *Ann Ist Super Sanita*. **44**, 13-15. (2008).

Claude N. [Reducing the use of laboratory animals]. *Bull Acad Natl Med*. **193**, 1767-1772; discussion 1772. (2009).

Cramer R.D., 3rd, Patterson D.E., Bunce J.D. Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res*. **291**, 161-165. (1989).

Dudek A.Z., Arodz T., Galvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb Chem High Throughput Screen*. **9**, 213-228. (2006).

Girschick T., Puchbauer L., Kramer S. Improving structural similarity based virtual screening using background knowledge. *J Cheminform*. **5**, 50. (2013).

Hert J., Willett P., Wilton D.J., Acklin P., Azzaoui K., Jacoby E., Schuffenhauer A. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J Chem Inf Comput Sci*. **44**, 1177-1185. (2004).

Holliday J.D., Kanoulas E., Malim N., Willett P. Multiple search methods for similarity-based virtual screening: analysis of search overlap and precision. *J Cheminform*. **3**, 29. (2011).

Irwin J.J., Shoichet B.K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. **45**, 177-182. (2005).

Klebe G., Abraham U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J Comput Aided Mol Des*. **13**, 1-10. (1999).

Law V., Knox C., Djoumbou Y., Jewison T., Guo A.C., Liu Y., Maciejewski A., Arndt D., Wilson M., Neveu V., Tang A., Gabriel G., Ly C., Adamjee S., Dame Z.T., Han B., Zhou Y., Wishart D.S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. **42**, D1091-1097. (2014).

Liu K., Feng J., Young S.S. PowerMV: a software environment for molecular viewing, descriptor generation, data analysis and hit evaluation. *J Chem Inf Model*. **45**, 515-522. (2005).

Ma D.L., Chan D.S., Leung C.H. Drug repositioning by structure-based virtual screening. *Chem Soc Rev*. **42**, 2130-2141. (2013).

Maggiore G., Vogt M., Stumpfe D., Bajorath J. Molecular similarity in medicinal chemistry. *J Med Chem.* **57**, 3186-3204. (2014).

Mayr L.M., Bojanic D. Novel trends in high-throughput screening. *Curr Opin Pharmacol.* **9**, 580-588. (2009).

Muegge I., Oloff S. Advances in virtual screening. *Drug Discov Today Technol.* **3**, 405-411. (2006).

Myint K.Z., Xie X.Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *Int J Mol Sci.* **11**, 3846-3866. (2010).

Nilsson J., Wikstrom H., Smilde A., Glase S., Pugsley T., Cruciani G., Pastor M., Clementi S. GRID/GOLPE 3D quantitative structure-activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D3 receptor subtype. *J Med Chem.* **40**, 833-840. (1997).

O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R. Open Babel: An open chemical toolbox. *J Cheminform.* **3**, 33. (2011).

Oprea T.I., Bauman J.E., Bologna C.G., Buranda T., Chigaev A., Edwards B.S., Jarvik J.W., Gresham H.D., Haynes M.K., Hjelle B., Hromas R., Hudson L., Mackenzie D.A., Muller C.Y., Reed J.C., Simons P.C., Smagley Y., Strouse J., Surviladze Z., Thompson T., Ursu O., Waller A., Wandinger-Ness A., Winter S.S., Wu Y., Young S.M., Larson R.S., Willman C., Sklar L.A. Drug Repurposing from an Academic Perspective. *Drug Discov Today Ther Strateg.* **8**, 61-69. (2011).

Pogliani L. From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors. *Chem Rev.* **100**, 3827-3858. (2000).

Reddy A.S., Pati S.P., Kumar P.P., Pradeep H.N., Sastry G.N. Virtual screening in drug discovery a computational perspective. *Curr Protein Pept Sci.* **8**, 329-351. (2007).

Sousa S.F., Fernandes P.A., Ramos M.J. Protein-ligand docking: current status and future challenges. *Proteins.* **65**, 15-26. (2006).

Talevi A., Castro E.A., Bruno-Blanch L.E. Virtual Screening: An emergent, key methodology for drug development in an emergent continent – A bridge towards patentability. En: *Advanced Methods and Applications in Chemoinformatics*, E. A. Castro and A. K. Haghi (Eds.), Editorial IGI Global. pp. 229-245 (2011).

Tobinick E.L. The value of drug repositioning in the current pharmaceutical market. *Drug News Perspect.* **22**, 119-125. (2009).

Todeschini R., Consonni V. *Molecular Descriptors for Chemoinformatics* Wiley-VCH Verlag GmbH, Weinheim, Germany. (2009)

Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today.* **11**, 1046-1053. (2006).

Willett P., Barnard J.M., Downs G.M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **38**, 983-996. (1998).

Yang S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today*. **15**, 444-450. (2010).

Yap C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem*. **32**, 1466-1474. (2011).

Yee L.C., Wei Y.C. Current Modeling Methods Used in QSAR/QSPR. En: *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. M. Dehmer, et al. (Eds.), (2012)

CAPÍTULO 3

REPOSICIONAMIENTO DE FÁRMACOS



3.1. DESCUBRIMIENTO Y DESARROLLO DE FÁRMACOS

3.1.1. ESTRATEGIAS GENERALES EN EL DESCUBRIMIENTO DE FÁRMACOS

El desarrollo de un fármaco para una patología específica es un proceso largo y riguroso que incluye diferentes etapas y procedimientos. El mismo comprende desde la identificación de un compuesto activo hasta el desarrollo de los estudios que garanticen la seguridad y eficacia del mismo para el uso clínico propuesto. En la Figura 3.1 se exhibe un esquema general de las diferentes etapas implicadas en el descubrimiento y desarrollo de un fármaco.

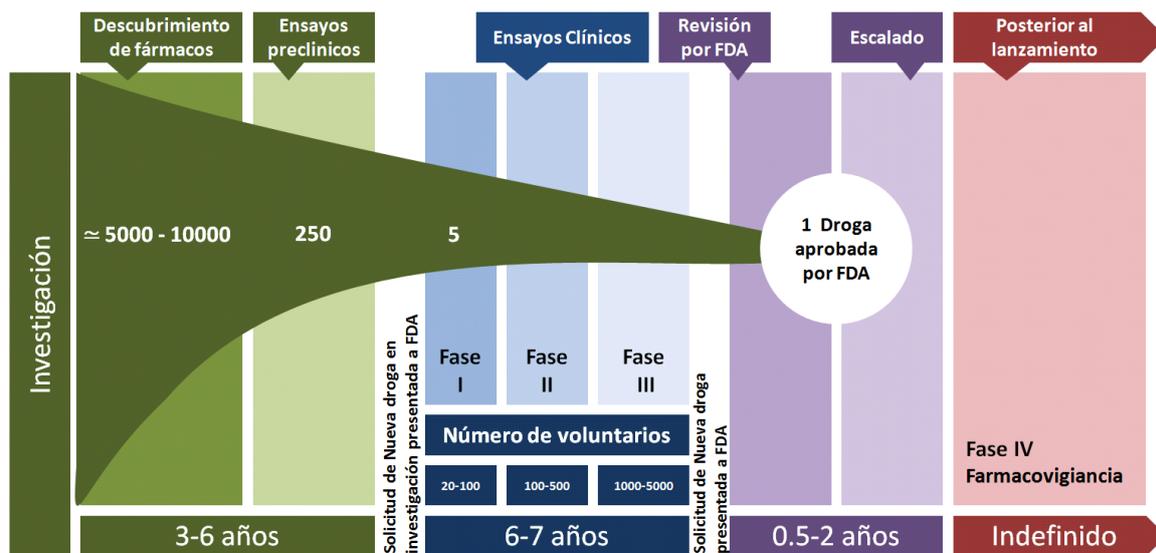


Figura 3.1. Esquema del proceso general del descubrimiento y desarrollo de un fármaco. Adaptado de Pharmaceutical Research and Manufacturers of America, Drug Discovery and Development: Understanding the R&D Process. <http://www.innovation.org> (Acceso Mayo 2014).

Podemos diferenciar dos estrategias generales de descubrimiento de fármacos (ver Figura 3.2) (Kell, 2013):



El enfoque conocido como “*cribado fenotípico*” o “*primero la función*” (en inglés, *function first*) (equivalente en genómica con el mecanismo “hacia adelante o directo”). Es la estrategia clásica en el descubrimiento de fármacos, según la cual se comienza con un organismo que muestra un determinado fenotipo asociado a una enfermedad y se ensayan distintos fármacos candidato para identificar aquellos eficaces y no tóxicos. En esta aproximación no es necesario postular (o dilucidar) un mecanismo de acción para desarrollar un fármaco exitoso; eventualmente este paso de identificación del mecanismo de acción podría devenir (mucho) más adelante.



En parte como resultado de los proyectos de secuenciación del genoma y de dilucidación del proteoma -en los cuales se identifican genes y productos génicos asociados a distintas enfermedades- surge un nuevo enfoque que se basa en la capacidad de los fármacos de modular blancos terapéuticos específicos, a bajas concentraciones, en ensayos *in vitro* que a posteriori son validados mediante ensayos *in vivo*. A diferencia del cribado fenotípico, aquí se parte de un cribado destinado a seleccionar aquellos compuestos que se unen selectivamente al blanco molecular de interés; en el caso del CV, se podrá utilizar un modelo de la estructura 3D de la proteína blanco o ligandos de referencia asociados a un mecanismo de acción específico para explorar las bases de datos. La estrategia se conoce como “*cribado basado en el blanco terapéutico*” o “*primero el blanco*” (en inglés, *target first*) (equivalente en genómica con el mecanismo “de reversa o hacia atrás”).

Ambas estrategias presentan ventajas y desventajas (Kell, 2013; Swinney and Anthony, 2011). Las fortalezas del enfoque “*primero el blanco terapéutico*” incluyen la capacidad de aplicar el conocimiento químico y molecular para investigar las hipótesis moleculares específicas, y la mayor viabilidad de aplicar el cribado a grandes bases de datos de moléculas pequeñas; adicionalmente, los avances en la quimioinformática y bioinformática contribuyen enormemente al diseño, selección u optimización de los

candidatos a fármacos para su interacción con un blanco terapéutico (Qu and Rajpal, 2012; Sliwoski et al., 2014). Una desventaja de esta aproximación es que no se conoce *a priori* la capacidad de los fármacos así identificados para modificar la progresión de una enfermedad o controlar su sintomatología, pudiendo observarse frecuentemente una ausencia de correlación entre la actividad *in vitro* de los fármacos candidatos y su actividad *in vivo*. Este hecho refleja, en parte, la selección natural de redes bioquímicas robustas, resistentes a cambios en un único nodo o parámetro de la red.

En el “cribado fenotípico”, en cambio, la actividad del candidato puede extrapolarse con mayor eficacia al impacto terapéutico en un estado determinado de enfermedad. La debilidad del enfoque es la falta de una relación clara entre el mecanismo de acción y el efecto biológico, siendo común que aquél se desconozca total o parcialmente. Un desafío explorado recientemente es el de incorporar de manera efectiva nuevas tecnologías de cribado enfocados en el fenotipo (Kell, 2013).

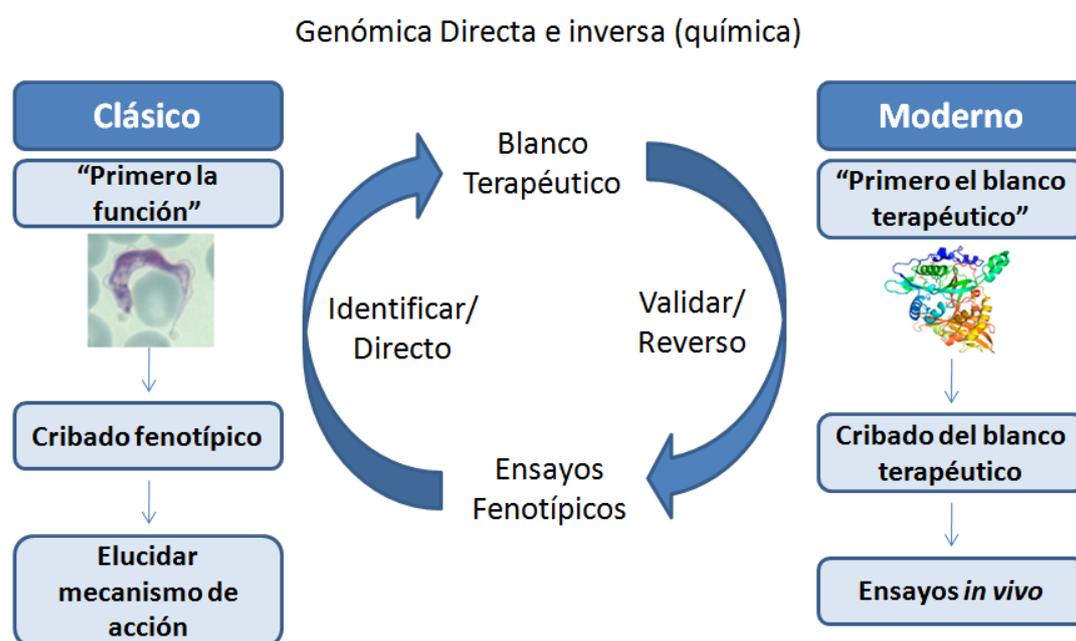


Figura 3.2. Esquema de estrategias aplicadas en el descubrimiento de fármacos.

En la era postgenómica, la alta frecuencia de fracasos del enfoque centrado en el blanco (Sams-Dodd, 2005) ha llevado al renacimiento del interés en la estrategia

centrada en la función, capaz de identificar candidatos que pueden interaccionar con más de un blanco terapéutico o diferentes rutas metabólicas. La tarea más compleja de esta estrategia es lograr un entendimiento de los complejos mecanismos de acción involucrados en las enfermedades, pero esto ha sido facilitado con el desarrollo de herramientas informáticas y nuevas técnicas de ensayo (Lee et al., 2012).

En relación a la experiencia de la industria farmacéutica según los dos enfoques mencionados, Swinney y How (Swinney and Anthony, 2011) sintetizan las estrategias aplicadas en el descubrimiento de los nuevos compuestos químicos y biológicos aprobados por la FDA para su uso en clínica en el periodo 1999-2008. 75 de esos compuestos corresponden a fármacos “primeros de su clase” (*first-in-class*), esto es, con nuevos mecanismos de acción. De éstos, 50 (67%) corresponden a pequeñas moléculas (o entidades químicas) y 25 (33%) son compuestos de origen biológico. Los resultados demuestran que la contribución del cribado fenotípico para el descubrimiento de fármacos de primera clase supera al enfocado en el blanco terapéutico (con 28 y 17 de estos fármacos novedosos procedentes respectivamente de esos dos enfoques) (ver Figura 3.3). No obstante el enfoque basado primero en el blanco terapéutico es el que predomina en el descubrimiento de “fármacos posteriores” (*follower-drugs*), fármacos que pueden aportar beneficios adicionales en la misma clase terapéutica inaugurada por los primeros de su clase.

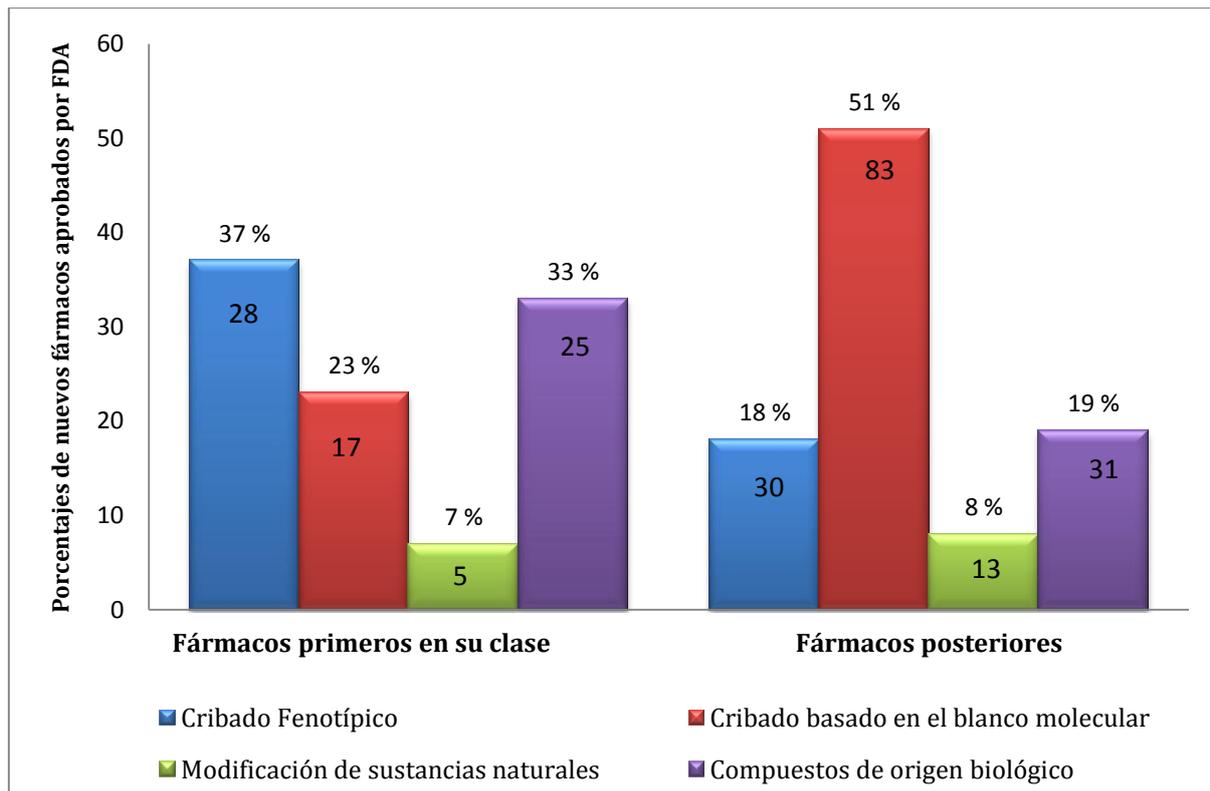


Figura 3.3. Distribución de los fármacos aprobados por la FDA en el período 1999-2008 de acuerdo a la estrategia de descubrimiento empleada. Adaptado de (Swinney and Anthony, 2011).

3.1.2. COSTOS Y FRACASOS EN EL DESARROLLO DE FÁRMACOS

Las estimaciones sobre el costo medio vinculado al desarrollo de fármacos según reportes de publicaciones analizadas por Morgan y colaboradores (Morgan et al., 2011) varían desde 92 millones de dólares de costo directo (161 millones de dólares de costo de capital) hasta 883,6 millones (1,8 billones de costo de capital). Las diferencias en las bases de datos examinadas para realizar estas estimaciones, las metodologías y la diversidad de categorías terapéuticas analizadas conducen a la amplia variación en el costo estimado. Las cifras más elevadas corresponden al análisis de información confidencial que proveen los laboratorios farmacéuticos, quienes presentan intereses respecto a la publicación de altas cifras para justificar el valor de sus productos en el mercado y para argumentar a favor de la extensión de las patentes de productos farmacéuticos.

A pesar de las diferencias en la estimación de los costos, estos números reflejan la importante inversión que insumen los estudios necesarios para que los candidatos se puedan transformar en medicamentos seguros y eficaces disponibles para su uso clínico.

El proceso de desarrollo está caracterizado por una alta tasa de fracaso. Las estimaciones de la tasa de éxito de los medicamentos que entran en las fases clínicas varían desde 11,7% a 24,0%; es decir, sólo uno o dos de cada diez proyectos que ingresan a estudios clínicos logran acceder al mercado farmacéutico bajo la forma de un medicamento innovador (Morgan et al., 2011).

Alrededor de un 27% de los fracasos de los candidatos a fármacos en las fases preclínica y clínica se deben a falta de eficacia, y más de un 30% a problemas relacionados con la seguridad/toxicidad (Ferri et al., 2013). En la Figura 3.4 se presenta la distribución de frecuencias de las causas de fracasos de aquellos candidatos que llegan a la etapa de evaluación clínica.

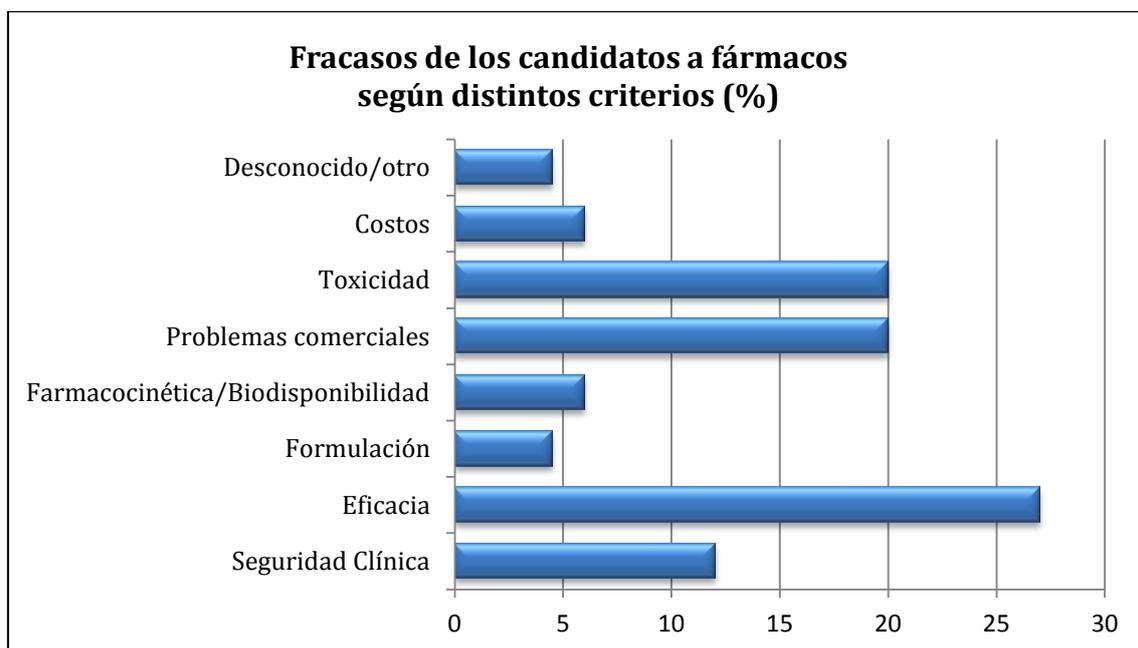


Figura 3.4. Distribución de frecuencias de causas de fracaso de los candidatos a fármacos en fase pre clínica y en fase clínica.

Un análisis reciente realizado por el laboratorio AstraZeneca (Roberts et al., 2014), ha demostrado que en el periodo 2001-2010, de entre 198 candidatos a fármacos de dicho

laboratorio, 48 (24%) fallaron por razones de seguridad entre el inicio de los estudios de toxicidad y la entrada a la fase I. De los 48 fracasos el 50% estaba relacionado a problemas de toxicidad en órganos (detectados por histopatología) o por muertes inexplicables durante las 2 semanas posteriores al inicio de la exposición. En la tabla 3.1 se detallan los problemas de toxicidad detectados para los 48 candidatos que no lograron avanzar en las fases clínicas. Los datos anteriores revelan que, si bien la falta de eficacia es una causa importante de atrición, se encuentra prácticamente a la par de la falta de seguridad de los candidatos estudiados.

Es por ello que uno de los objetivos relevantes para los laboratorios farmacéuticos es mejorar la capacidad para predecir la probabilidad de fracaso asociada a la toxicidad antes de ingresar a las fases clínicas, dado que tomar la decisión de avanzar hacia la fase clínica representa una significativa inversión financiera y de recursos de I + D (Roberts et al., 2014).

Tabla 3.1. Fracasos de fármacos debido a diferentes toxicidades.

Tipos de Toxicidad de candidatos a fármacos que han fracasado		
Predicción	Número de CFs	Toxicidad
Detectado	2	Hígado
	1	Gastrointestinal
	1	Sistema Nervioso Central
	1	Riñón
	4	Patologías cardiovasculares
	2	Pulmón
	2	Muerte por causa inespecífica
	3	Músculo
	1	Páncreas
	1	Tiroides
	1	Lenticular
	1	Toxicidad múltiple en órganos
	4	específicos
	24 (50%) ^a	Función cardiovascular
No detectado luego de 2 semanas	1	Patología Cardiovascular
	1	Sistema Nervioso Central
	1	Pulmonar
	1	Reprotoxicidad
	4 (8%)	

Retirado por razones que no involucran la toxicidad	9 (19%)	No se debe a toxicidad
Incierto	3	Riñón
	1	Hígado
	2	Pulmón
	1	Músculo
	1	Testículo
	1	Glándula Suprarrenal
	2	Toxicidad en múltiples órganos
	11 (23%)	
Total	48 (100%)	

Nota: ^a Este 50% comprende a los 20 compuestos (40%) retirados por causas atribuibles a toxicidades en órganos más los 4 (10%) compuestos debido a los riesgos cardiovasculares agudos. Adaptado de referencia (Roberts et al., 2014)

3.2. CAMBIO DE PARADIGMA EN LA BÚSQUEDA DE NUEVOS FÁRMACOS

Hace más de dos décadas, el paradigma dominante en el campo del descubrimiento de fármacos proponía el desarrollo de ligandos con máxima selectividad que actúen modulando un único blanco molecular asociado a la enfermedad objetivo (Sams-Dodd, 2005). En el marco de este paradigma, la selectividad y la potencia eran los aspectos esenciales para decidir si un candidato a fármaco progresaría a fases posteriores de desarrollo.

Tal enfoque reduccionista se basa en dos premisas: a) fármacos altamente específicos evitarían efectos secundarios debidos a interacciones del mismo con biomoléculas distintas al blanco molecular, lo que conduciría a terapias más seguras y; b) al menos algunas enfermedades podrían ser tratadas adecuadamente modulando un único blanco molecular (paradigma *"un gen, una droga, una enfermedad"*) (Kaufmann, 2008). Esta aproximación se ajusta a la metáfora de que los fármacos ideales deberían actuar como *"balas mágicas"* (en inglés *magic bullets*), expresión que emplea por primera vez Paul Ehrlich (1854-1915) a principios del siglo pasado para referirse a los compuestos que en forma específica actuaban contra los microorganismos causantes de una enfermedad con una mínima toxicidad sobre el huésped (Drews, 2000).

Sin embargo, observaciones recientes han desafiado el paradigma anterior al observarse que los fármacos utilizados actualmente en clínica a menudo se unen a más de un objetivo molecular (Marona-Lewicka and Nichols, 2007; Marona-Lewicka and Nichols, 2009; Nobeli et al., 2009; Peterson, 2008). Bajo esta nueva perspectiva es que se sugiere que una analogía de “*escopeta mágica*” (*magic shotgun*) podría ser más adecuada para describir el mecanismo de acción de muchos fármacos conocidos, en particular dentro de ciertas categorías terapéuticas (Roth et al., 2004).

Algunos autores se han referido a este tipo de ligandos multiobjetivo como “*drogas promiscuas*” (Peters, 2013). Por otro lado, si a un blanco molecular se unen diferentes ligandos químicamente heterogéneos, también se lo puede considerar como promiscuo (Perez-Nueno et al., 2011). No obstante, preferiremos el uso de los términos ligando *poliespecífico*, *multiobjetivo* o *multifuncional* sobre el de promiscuidad, bajo la perspectiva de que los mismos evocan mejor la idea de una interacción selectiva con más de un blanco molecular o, si se quiere, una promiscuidad selectiva (Medina-Franco et al., 2013; Merino et al., 2010).

Las interacciones de los fármacos con proteínas distintas al blanco molecular tradicionalmente se han visto como un aspecto indeseable asociado a efectos secundarios no deseados. Pero en muchos casos -que van desde ciertos medicamentos psiquiátricos a las modernas terapias contra el cáncer- esta poliespecificidad está intrínsecamente vinculada a la eficacia terapéutica del fármaco (Hopkins, 2009).

Varios autores proponen que el diseño de fármacos multi-objetivo que interactúan con varios blancos moleculares asociados a una misma enfermedad conducirán a medicamentos nuevos y más eficaces para una variedad de trastornos de etiología compleja, entre ellos diversos desórdenes del sistema nervioso central, tales como enfermedades neurodegenerativas, desórdenes psiquiátricos y epilepsia (Di Ianni and Talevi, 2014; Roth et al., 2004; Talevi et al., 2012 y referencias allí citadas).

Por otra parte, esta nueva visión permite también concebir y explicar el reposicionamiento de fármacos o búsqueda de segundos usos médicos (que se describe en la siguiente sección) como una alternativa viable para desarrollar nuevos medicamentos. En este caso, pensaríamos en un fármaco multi-objetivo que interactúe

selectivamente con blancos moleculares asociados a distintas enfermedades o desórdenes, o en utilizar un fármaco aprobado para una indicación para una segunda indicación en la que esté involucrado el mismo blanco molecular o una isoforma del mismo blanco molecular.

3.3. REPOSICIONAMIENTO DE FÁRMACOS

El concepto de reposicionamiento de fármacos (en inglés *drug repurposing* o *drug reprofiling*) se refiere a encontrar nuevos usos terapéuticos para medicamentos ya existentes, incluyendo aquellos que se encuentran en el mercado farmacéutico, fármacos discontinuados, abandonados y en desarrollo. Dentro de la comunidad internacional el interés en dicha estrategia ha crecido notablemente en los últimos años y en países desarrollados como Estados Unidos y Reino Unido se han generado diversos programas públicos o público-privados para promoverla (Allarakhia, 2013; Allison, 2012; Marusina et al., 2011; Murteira et al., 2013; Novac, 2013; Wadman, 2013).

El reposicionamiento de fármacos está intrínsecamente ligado al concepto del uso “fuera de lo indicado” (*off-label*) de un medicamento, que se refiere a la prescripción de un medicamento por un médico (basado en evidencia científica o empírica) para indicaciones que aún no han sido evaluadas ni aprobadas por autoridades de la salud (Meadows and Hollowell, 2008). El uso fuera de lo indicado es una práctica muy frecuente en ciertas ramas de la Medicina (ej. Psiquiatría y Pediatría).

La búsqueda de segundos usos médicos de fármacos ya conocidos posibilita introducir tratamientos innovadores en el mercado reduciendo considerablemente los tiempos (reducción de entre 3 y 12 años) y costos asociados al desarrollo de un medicamento novedoso (Arrowsmith and Harrison, 2012; Ashburn and Thor, 2004). Los candidatos reposicionados han (como mínimo) sobrevivido, previamente, a estudios toxicológicos preclínicos (y en muchos casos clínicos) y poseen un perfil farmacocinético adecuado, ya caracterizado. Cuando el fármaco reposicionado ha sido previamente aprobado para su uso clínico y comercializado, los posibles problemas de estabilidad y de manufactura ya han sido resueltos (por ejemplo, ya se han superado los estudios de

pre-formulación); finalmente, aquellos compuestos cuya patente se ha vencido pueden brindar soluciones relativamente poco costosas para nuevos problemas (Aubé, 2012).

Los reposicionamientos exitosos probablemente han contribuido al interés de expansión de las indicaciones terapéuticas de fármacos ya conocidos; por ejemplo el sildenafil originalmente fue investigado para el tratamiento de la hipertensión e isquemias cardíacas pero se transformó en un éxito de ventas por su uso en el tratamiento de la disfunción eréctil (Ghofrani et al., 2006). Por su parte la aspirina ha expandido sus indicaciones terapéuticas y hoy es ampliamente utilizada en la prevención de accidentes cerebro y cardiovasculares en pacientes con una condición cardiovascular preexistente (Patrono et al., 2005). En la tabla 3.2 se observan diferentes ejemplos de reposicionamiento de fármacos.

Tabla 3.2. Ejemplos de fármacos reposicionados.

<i>Fármaco</i>	<i>Indicación Original</i>	<i>Nueva Indicación</i>	<i>Referencias</i>
Aspirina	Inflamación	Antiplaquetario	(Patrono, 1994; Patrono et al., 2005)
Anfotericina B	Infecciones fúngicas	Leishmaniasis	(Bern et al., 2006; Coukell and Brogden, 1998; Thornton and Wasan, 2009)
Bromocriptina	Enfermedad de Parkinson, hiperprolactinemia y galactorrea	Diabetes mellitus	(Holt et al., 2010)
Bupropion	Depresión	Terapia para dejar de fumar	(Lerman et al., 2004)
Celecoxib	Osteoartritis y artritis reumatoidea en adultos	Poliposis adenomatosa familiar, cáncer de colon y de mama	(Jendrossek, 2013; Kim et al., 2010; Steinbach et al., 2000)
Duloxetina	Trastorno depresivo	Stress, Incontinencia Urinaria	(Voelker, 1998)
Finasteride	Hiperplasia prostática benigna	Pérdida de cabello	(Leyden et al., 1999; Rossi et al., 2011)

Fluoxetina	Depresión	Disforia Premenstrual	(Pearlstein et al., 1997; Steiner et al., 1995)
Galantamina	Poliomielitis, parálisis, anestesia	Enfermedad de Alzheimer	(Lilienfeld, 2002; Richarz et al., 2014)
Gemcitabina	Infecciones virales	Cáncer	(King, 1996)
Metotrexato	Cáncer	Artritis reumatoidea, Psoriasis	(Burmester et al., 2014; Emery et al., 2006)
Minoxidil	Hipertensión	Pérdida de cabello	(Melkote et al., 2009; Olsen et al., 2002)
Paclitaxel	Cáncer	Prevención de restenosis coronaria	(Byrne et al., 2013; Gershlick et al., 2004)
Raloxifeno	Cáncer de próstata y de mama	Osteoporosis	(Cranney and Adachi, 2005; Das and Crockett, 2013)
Ropinirol	Hipertensión	Enfermedad de Parkinson y síndrome idiopático de las piernas inquietas	(Adler et al., 1997; Schrag et al., 2002; Stocchi et al., 2014)
Sildenafil	Angina	Disfunción eréctil	(Ghofrani et al., 2006)
Tadalafil	Inflamación y enfermedad cardiovascular	Disfunción eréctil	(Ali, 2008)
Topiramato	Epilepsia	Obesidad	(Appolinario et al., 2002; McElroy et al., 2003)
Warfarina	Prevención de trombosis	Profilaxis secundaria seguida de infarto de miocardio	(Hurlen et al., 1994; Hurlen et al., 2002)
Zidovudina	Cáncer	HIV/SIDA	(Broder, 2010)

Los segundos usos de fármacos aprobados han sido descubiertos, mayormente, mediante serendipia (explotando de manera inteligente efectos colaterales imprevistos de un fármaco). Sin embargo, recientemente se ha explorado el reposicionamiento basado en el conocimiento, incluyendo aproximaciones quimioinformáticas, bioinformáticas (Issa et al., 2013; Liu et al., 2013; Lussier and Chen, 2011; Napolitano et

al., 2013) y análisis de literatura de alto rendimiento (Deftereos et al., 2011; Frijters et al., 2010).

A continuación se examinan las tendencias recientes en el campo de reposicionamiento de fármacos asistido por computadora y se discutirá la aplicación particular de esta estrategia en la búsqueda de nuevas soluciones terapéuticas para las enfermedades olvidadas.

3.3.1. REPOSICIONAMIENTO DE FÁRMACOS ASISTIDO POR COMPUTADORA

El reposicionamiento de fármacos asistido por computadora se basa en dos principios generales (Liu et al., 2013): a) Los medicamentos que comparten ciertas características moleculares mínimas, biológicamente relevantes, podrían interactuar con el mismo blanco molecular (enfoque centrado en los fármacos) y; b) los trastornos de salud relacionados con las mismas o similares proteínas desreguladas o disfuncionales pueden ser tratados con los mismos medicamentos (enfoque centrado en la enfermedad). Diversos métodos computacionales podrían ser útiles para revelar las relaciones proteína-fármaco o proteína-proteína ocultas. Los clasificaremos en aproximaciones bioinformáticas, quimioinformáticas y basadas en análisis de literatura de alto rendimiento, y describiremos brevemente la integración de las tres anteriores en redes multimodales fármaco-proteína o fármaco-proteína-enfermedad.

3.3.1.1. BIOINFORMÁTICA APLICADA EN REPOSICIONAMIENTO DE FÁRMACOS

La bioinformática aborda el reto de encontrar similitudes estructurales y conexiones funcionales entre los productos de los genes, y, más recientemente, las similitudes y similitudes inversa/reversa entre los patrones/perfiles de expresión de todo el genoma (firmas genéticas) relacionados con la enfermedad y el efecto de uno o varios fármacos. El perfil de expresión génica ofrece una medida global instantánea de los niveles de

transcripción en una célula determinada, tejido u organismo bajo una cierta condición experimental (Qu and Rajpal, 2012).

Un recurso público destinado a conectar enfermedades y moléculas bioactivas a través de perfiles génicos es el Connectivity Map o CMap (Instituto Board)(Lamb et al., 2006). Esta base de datos fue la primera en compilar perfiles de expresión génica derivados de la exposición/tratamiento de una serie de células humanas con un gran número de fármacos y compuestos bioactivos. En un principio se consideraron 164 perturbágenos; hoy, ese número se ha expandido a más de 1.300 moléculas aprobadas por la FDA y se ha anunciado que CMap pronto tendrá alrededor de 4.000 firmas fármaco-efecto. El recurso emplea algoritmos simples de comparación de patrones que permiten descubrir conexiones funcionales entre fármacos, genes y enfermedades, a través de cambios comunes en la expresión de genes. *¿Cómo puede ser utilizado este recurso en el reposicionamiento de fármacos?* Si se utiliza una firma genética correspondiente a un estado determinado de una enfermedad como “firma de consulta” (*query*), los fármacos que presenten una similitud inversa a la consulta serían, hipotéticamente, una terapia potencial para restaurar el estado fisiológico. De manera alternativa, si una firma fármaco-efecto es seleccionado como consulta, entonces todas las firmas fármaco-efecto almacenadas similares representarían a los fármacos con efecto similar (Figura 3.5). Ejemplos de aplicación de CMap pueden encontrarse en las siguientes referencias (Cavalla, 2013; Dudley et al., 2011; Qu and Rajpal, 2012; Sirota et al., 2011).

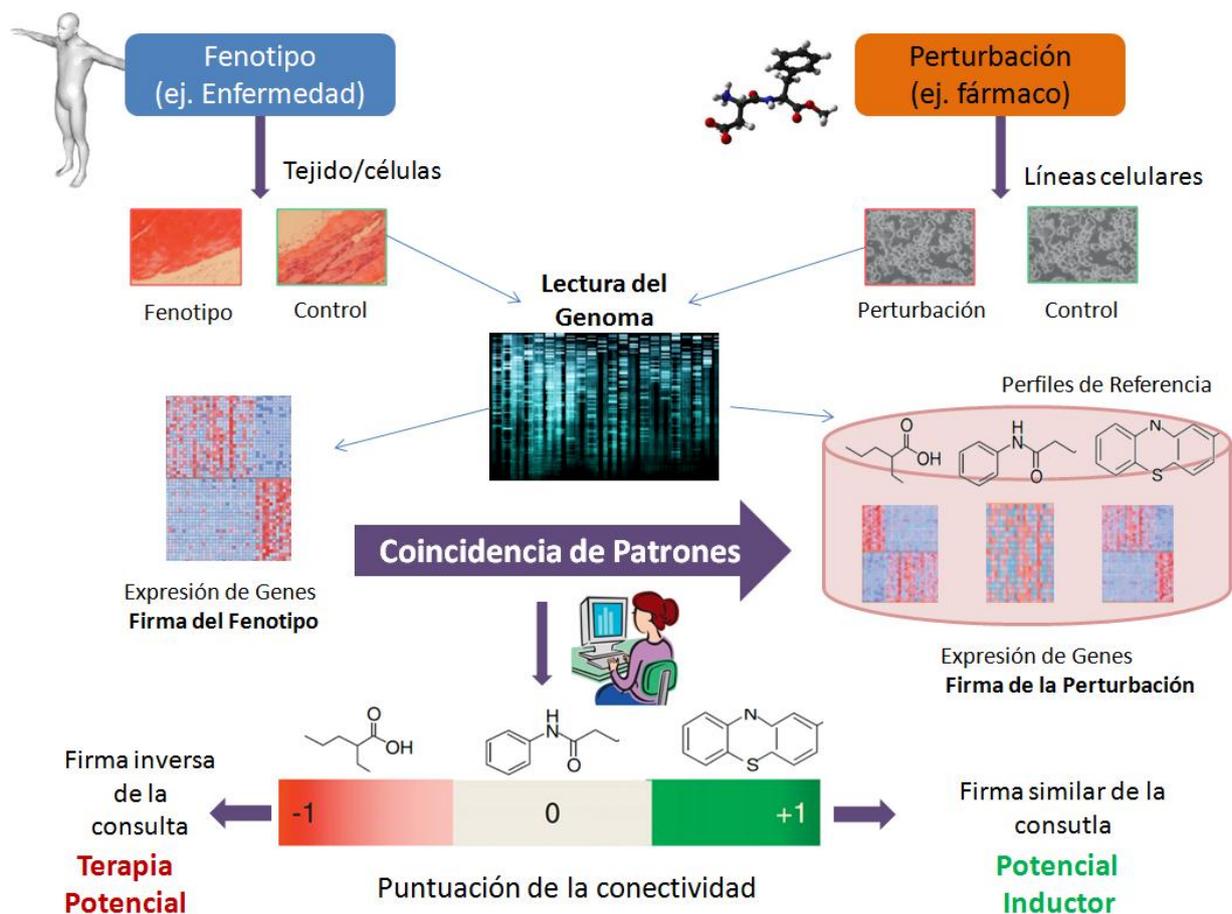


Figura 3.5. Esquema general que muestra cómo las firmas de fármaco-efecto pueden utilizarse para seleccionar potenciales agentes terapéuticos. (Adaptado de (Qu and Rajpal, 2012)).

Otros recursos bioinformáticos utilizados en el reposicionamiento de fármacos son aquellos basados en la homología. Se definen como **homólogos** dos genes, o sus correspondientes productos génicos, cuando poseen alta similitud entre sí. Son **ortólogos** cuando esta similitud deriva de una ascendencia común (relación vertical), y son **parálogos** cuando la similitud se produce dentro del mismo genoma por duplicación de un gen (relación horizontal) (ver esquema en Figura 3.6). Frecuentemente dos genes ortólogos tienen la misma función, mientras que los genes parálogos, puesto que se han originado por duplicación y posterior divergencia, deben tener diferente función o, al menos, cierto grado de especialización (Koonin, 2005).

Dos proteínas *A* y *B*, de los organismos *1* y *2*, respectivamente, son ortólogas si *A* es la proteína más similar a *B* en el organismo *1*, y viceversa. A esta definición se le suele

imponer dos restricciones adicionales: que la similitud tenga una puntuación por encima de cierto valor (para el algoritmo de alineamiento más utilizado BLAST - *Basic Local Alignment Search Tool*- suelen usarse E-valores menores de 1×10^{-5}) y que las regiones de similitud entre las dos proteínas abarquen, al menos, el 75% de sus longitudes.

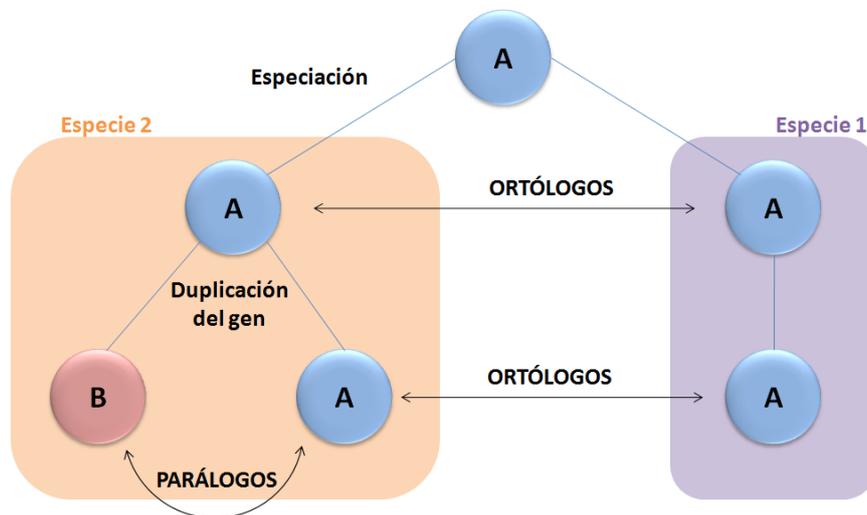


Figura 3.6. Diagrama representativo de los conceptos de ortólogos y parálogos.

Las proteínas procedentes de un ancestro común en general presentan un alto grado de similitud en la secuencia y en la estructura. Desde el punto de vista bioinformático, esta similitud ha sido ampliamente utilizada para la predicción de la estructura y la función de las proteínas (Kiefer et al., 2009; Loewenstein et al., 2009). Recientemente se ha reportado que la unión de pequeñas moléculas (ej. fármacos) también se conserva entre homólogos (Kruger and Overington, 2012; Paricharak et al., 2013). En particular, las proteínas ortólogas de especies cercanas como el humano y la rata a menudo comparten la mayoría de sus ligandos. Lo mismo ocurre para proteínas parálogas, aunque el grado de similitud entre los ligandos de proteínas de este tipo es ligeramente menor que entre las proteínas ortólogas (Kruger and Overington, 2012).

Ejemplos recientes de aplicaciones en el reposicionamiento de fármacos basados en homología pueden ampliarse en las siguientes referencias, donde se emplean recursos bioinformáticos web de libre acceso (Gfeller et al., 2014; Salazar et al., 2014).

3.3.1.2. QUIMIOINFORMÁTICA APLICADA EN REPOSICIONAMIENTO DE FÁRMACOS

La utilización de la quimioinformática en el reposicionamiento se basa principalmente en la utilización de las técnicas de CV (detalladas en el capítulo anterior) aplicadas sobre bases de datos de fármacos ya aprobados, discontinuados, abandonados y/o en desarrollo. La disponibilidad de este tipo repositorios ha allanado el camino para el desarrollo de campañas de reposicionamiento de fármacos basados en quimioinformática.

Los métodos usados se clasifican de la misma forma que las aproximaciones de CV (Talevi et al., 2011), ya que se trata de las mismas técnicas enfocadas en un subconjunto acotado y bien definido del universo de moléculas tipo-fármaco conocidas. Por ello, la aplicación del CV con fines de reposicionamiento es particularmente eficiente, punto que debe tenerse en cuenta cuando se desean realizar campañas de CV basadas en el blanco molecular. El reposicionamiento de fármacos por *docking* molecular puede ser realizado a través del enfoque en un único blanco molecular (que tiene por objetivo identificar posibles interacciones entre un único blanco de interés y diversos candidatos), o investigando la unión de un fármaco existente frente a un panel de diversos blancos terapéuticos (Ma et al., 2013).

En relación a los enfoques basados en los ligandos, comparaciones sistemáticas indican que mientras los enfoques relativamente sencillos y eficaces (por ejemplo métodos de similitud basados en huellas digitales 2D) presentan muy buenas métricas de enriquecimiento con bajo costo computacional, métodos más elaborados dependientes de la conformación como la superposición con farmacóforos en general muestran una mejor recuperación de andamiajes moleculares novedosos (*scaffold hopping*) (Good et al., 2004; Zhang and Muegge, 2006). Holliday y colaboradores han demostrado (explorando sistemáticamente una variedad de blancos moleculares mediante distintos métodos de búsqueda por similitud) que el enriquecimiento de activos aumenta cuando se utilizan diferentes moléculas de referencia y diferentes esquemas de huellas digitales en campañas de CV basado en similitud (Holliday et al., 2011). Recientemente, Wu y col. desarrollaron una idea muy atractiva donde aducen que dos indicaciones terapéuticas se correlacionarán entre sí si comparten los mismos

fármacos o fármacos similares (por lo tanto, una red que indique qué categorías terapéuticas tienen fármacos similares sería un marco valioso para el reposicionamiento sistemático de fármacos) (Wu et al., 2013a).

3.3.1.3. REPOSICIONAMIENTO BASADO EN LA LITERATURA

El proceso de generación de nuevas hipótesis generando “puentes” que una hechos científicos aparentemente no relacionados (mediante la identificación de asociaciones indirectas entre ellos) se conoce como descubrimiento basado en la literatura (*literature-based discovery - LBD*). El LBD se basa en la hipótesis de que dos conjuntos de conocimientos o conceptos A y C pueden estar relacionados entre sí si comparten un vínculo a un concepto intermedio B (Lekka et al., 2011) (de hecho, cuanto más grande es el número de conceptos compartidos entre A y C, más probable será la relación entre ellos). Este modelo luego se postuló y aplicó por primera vez durante los años ‘80 y se conoce como el modelo ABC de Swanson (Swanson, 1986; Swanson, 1988). La primera aplicación exitosa del modelo ABC descubrió el potencial terapéutico del aceite de pescado para el tratamiento del Síndrome de Raynaud, siguiendo las observaciones conocidas de que el Síndrome está relacionado con el aumento en la viscosidad de la sangre y que por otro lado el aceite de pescado reduce la viscosidad de la sangre. Esta predicción se dedujo mediante el uso del método semi-automatizado Arrowsmith; este método utiliza un marco “cerrado” en el cual el usuario proporciona la hipótesis (Figura 3.7) (Jensen et al., 2006).

El rápido incremento en el volumen de la literatura científica genera un enorme número de conexiones implícitas/indirectas (y potencialmente valiosas) entre entidades descritas en ella; la posibilidad de este tipo de conexiones para permanecer ocultas/desapercibidas aumenta sustancialmente debido a la naturaleza cada vez más divergente del conocimiento como consecuencia de la especialización científica (Cohen et al., 2012). Actualmente es difícil inclusive que un investigador pueda mantenerse al día con toda la literatura relevante de su especialidad (Jensen et al., 2006), más aún que esté actualizado en relación a descubrimientos desarrollados en otros campos del conocimiento. Por lo tanto, el desarrollo de métodos automatizados de alto rendimiento

de recuperación de la información y extracción de información relevante es cada día más necesario; en el contexto de reposicionamiento de fármacos, un enfoque de descubrimiento “abierto” (establecimiento de relaciones y posterior formulación de hipótesis) (Figura 3.7) parece ser la mejor estrategia para encontrar segundos usos médicos (Weeber et al., 2005).

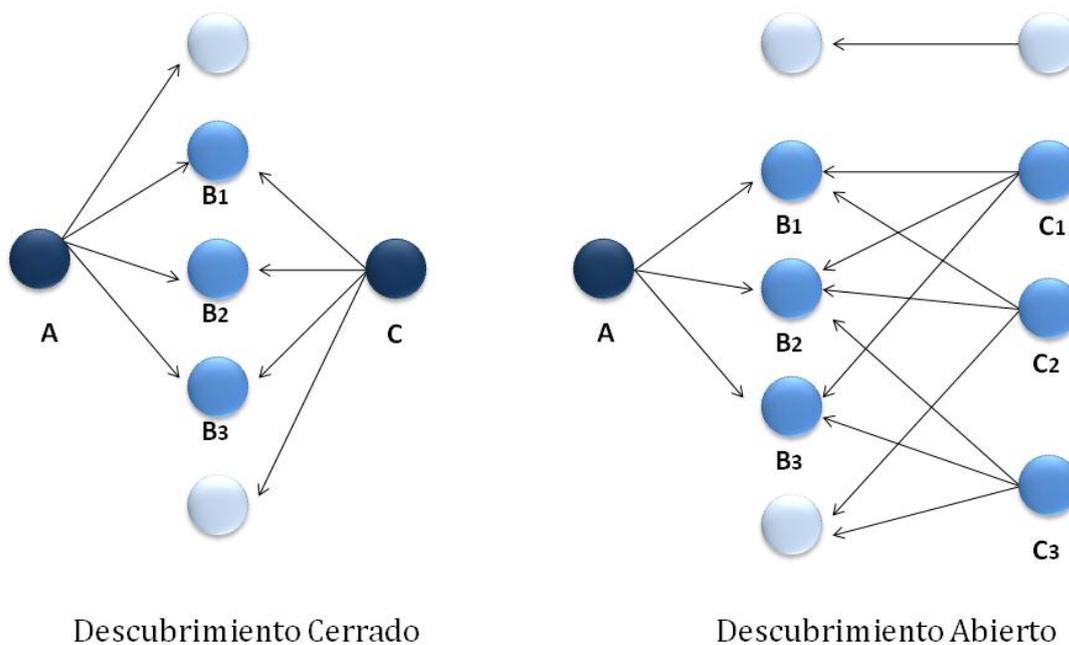


Figura 3.7. Esquema del principio ABC de las relaciones ocultas en la literatura. El descubrimiento cerrado puede ser útil para apoyar hipótesis previamente formuladas; el descubrimiento abierto es valioso para la generación de una hipótesis y por lo tanto es el enfoque más apropiado para las campañas de reposicionamiento de fármacos.

3.3.1.4. REDES APLICADAS EN EL REPOSICIONAMIENTO DE FÁRMACOS

Tal como se mencionó anteriormente, ya no es posible explorar manualmente la literatura y fuentes de información disponibles (por ejemplo, bases de datos de genes y productos génicos) para encontrar conexiones valiosas, y se necesitan enfoques computacionales para analizar esas enormes cantidades de datos (Iskar et al., 2012).

Las redes nos permiten lidiar con la complejidad mediante una simplificación de los sistemas complejos: los conceptos o entidades se representan como *nodos* mientras que las relaciones entre nodos se representan como *enlaces* (Vidal et al., 2011). En la representación las características funcionales o dinámicas de los elementos representados como nodos usualmente pierde importancia y la atención se centra en la conectividad entre los nodos, es decir en la arquitectura topológica de la red (para cuyo análisis puede recurrirse a la Teoría de Grafos matemática). La conectividad puede establecerse a través de relaciones demostradas experimentalmente o por medio de asociaciones predichas computacionalmente (por ejemplo mediante similitud química, similitud de proteínas, perfiles de expresión similares, conexiones inferidas por literatura, etc.).

El análisis de la topología de la red enfocado en el reposicionamiento de fármacos permitirá rescatar aquellos compuestos con alto grado de conectividad con múltiples blancos, o blancos moleculares vinculados a múltiples compuestos de categorías terapéuticas distintas. Mientras que los *hubs* (nodos altamente conectados) de interacciones proteína-proteína corresponderán a las proteínas esenciales cuya modulación impactará profundamente sobre la función del sistema (pudiendo preverse un perfil toxicológico desfavorable), los nodos conectados moderadamente podrían ser de interés como potenciales nuevos blancos terapéuticos (Hopkins, 2008).

Pueden encontrarse ejemplos de reposicionamiento de fármacos basado en redes en las siguientes referencias (Chen et al., 2012; Di Ianni and Talevi, 2014; Wu et al., 2013b)

3.3.2. REPOSICIONAMIENTO DE FÁRMACOS ASISTIDO POR COMPUTADORAS PARA ENFERMEDADES OLVIDADAS

Según la OMS las enfermedades olvidadas afectan a más de 750 millones de personas en todo el mundo. No obstante, sólo 21 (1,3%) de los 1.556 medicamentos registrados entre 1975 y 2004 se han desarrollado específicamente para estas condiciones, lo que refleja las fallas del mercado y el fracaso de las políticas públicas (Chirac and Torreale, 2006).

El sector farmacéutico privado es reticente a invertir en investigación y desarrollo de nuevos tratamientos, debido al limitado retorno de inversión. Por lo tanto, el sector público, las organizaciones académicas y las organizaciones sin fines de lucro juegan un papel destacado en el desarrollo de nuevas soluciones a estas enfermedades (Coles and Cloyd, 2012; Griggs et al., 2009; Moran et al., 2009). Un estudio realizado por G-Finder (Global Funding of Innovation for Neglected Diseases) revela que alrededor del 90% de los fondos para I + D en enfermedades olvidadas proviene de las organizaciones del sector público y sin fines de lucro (Moran et al., 2009). Más aún, las instituciones públicas, incluidos laboratorios públicos de investigación y universidades, han contribuido al desarrollo de casi el 90% de las nuevas indicaciones para medicamentos ya aprobados (Stevens et al., 2011).

En este escenario, la posibilidad de reposicionamiento de fármacos sin la comercialización (por incorporación directa de los resultados de la investigación a la práctica clínica, después de un examen de los datos de las autoridades reglamentarias) ha sido sugerido como una opción única para el sector académico (Coles and Cloyd, 2012). También se ha subrayado la importancia del reposicionamiento de fármacos asistido por computadoras en los países en desarrollo, donde la mayoría de la limitada y relativamente escasa inversión de I + D proviene del sector público (ver Figura 3.8), y el sector privado se muestra reacio a invertir en I + D (Talevi et al., 2011).

La discusión anterior explica entonces, por qué el reposicionamiento de fármacos constituye una estrategia clave en el campo del descubrimiento y desarrollo de fármacos para enfermedades olvidadas, habiendo una necesidad evidente de colaboración entre los sectores público y privado (Ekins et al., 2011; Pollastri and Campbell, 2011; Radisch, 2007; Sardana et al., 2011) y siendo esencial la implementación de estrategias eficientes desde el punto de vista translacional. En la tabla 3.3 se presentan algunos ejemplos de reposicionamiento para enfermedades olvidadas.

Diversas iniciativas, como el Programa Especial de Investigaciones y Enseñanzas sobre Enfermedades Tropicales de la OMS, el Proyecto Medicinas para Malaria, la Alianza Global para el desarrollo de medicamentos para Tuberculosis, DNDi y el consorcio *Open Source Drug Discovery* han reconocido al reposicionamiento de

medicamentos como una opción atractiva para proporcionar un acceso de bajo costo a los medicamentos en los países en desarrollo (Bost et al., 2010).

El potencial del reposicionamiento de fármacos asistido por computadoras centrado en enfermedades olvidadas ha sido reforzado recientemente por la transferencia al sector farmacéutico del modelo de Código Abierto (en inglés *Open Source*) proponiéndose la producción y difusión del conocimiento colectivo en el campo del descubrimiento de fármacos (Allarakhia, 2010; Allarakhia, 2013; Allarakhia and Ajuwon, 2012). Este modelo facilita el ingreso al campo de desarrollo de nuevos medicamentos de empresas/productores de países emergentes.

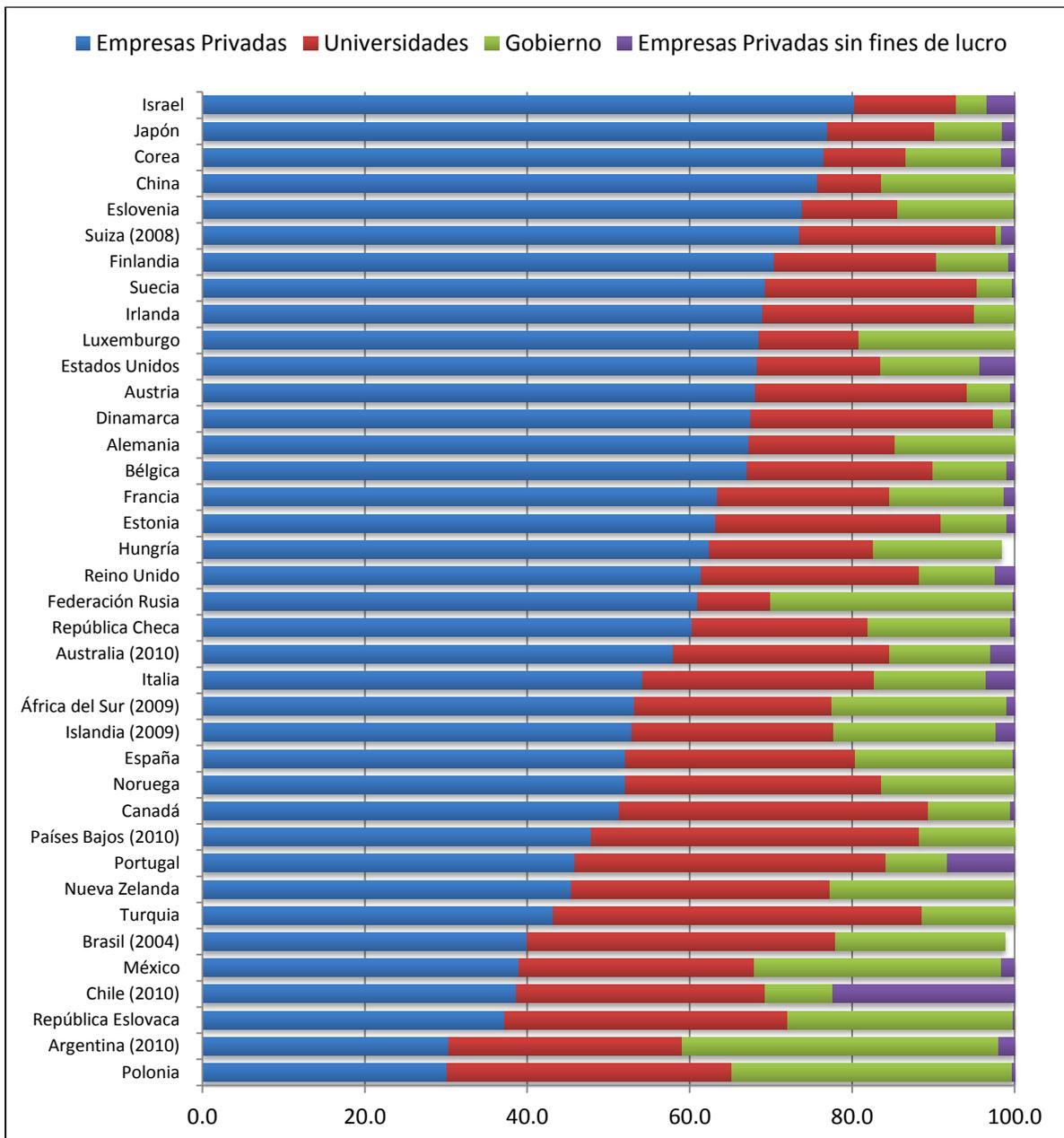


Figura 3.8. Inversiones en I+D diferenciadas por sectores (como porcentaje en gasto bruto interno). Gráfico construido con la información publicada en los reportes 2007 y 2013 de la Organización para la Cooperación y Desarrollo Económico (OECD). Se destaca la mayor inversión relativa en I+D por parte del sector público en países emergentes como Chile, Brasil y Argentina.

Fuente: OECD-Science, Technology and Industry Scoreboard. OECD, Main Science and Technology Indicators Database, www.oecd.org/sti/msti.htm, último acceso Junio, 2013.

Tabla 3.3. Ejemplos de campañas orientadas al reposicionamiento de fármacos con foco en enfermedades olvidadas.

<i>Fármaco</i>	<i>Indicación Original</i>	<i>Nueva indicación</i>	<i>Referencias</i>
Amiodarona	Antiarrítmico	Chagas	(Bellera et al., 2013; Oldfield, 2010)
Bromocriptina	Enfermedad de Parkinson	Chagas	(Bellera et al., 2013)
Tamoxifeno	Antiestrogénico	Leishmania amazonensis	(Miguel et al., 2008)
Anfotericina B	Infecciones Fúngicas	Leishmaniasis	(Freitas-Junior et al., 2012)
Ivermectina	Antiparasitario	Malaria	(Kobylnski et al., 2011)
Eflornitina	Anticáncer	Enfermedad del sueño	(No authors listed, 2005)
Astemizol	Antihistamínico	Malaria	(Chong et al., 2006)
Cicloserina	Infecciones causadas por <i>Giarda</i>	Tuberculosis	(de Carvalho et al., 2009)

REFERENCIAS CAPÍTULO 3

Adler C.H., Sethi K.D., Hauser R.A., Davis T.L., Hammerstad J.P., Bertoni J., Taylor R.L., Sanchez-Ramos J., O'Brien C.F. Ropinirole for the treatment of early Parkinson's disease. The Ropinirole Study Group. *Neurology*. **49**, 393-399. (1997).

Ali S.T. Effectiveness of sildenafil citrate (Viagra) and tadalafil (Cialis) on sexual responses in Saudi men with erectile dysfunction in routine clinical practice. *Pak J Pharm Sci*. **21**, 275-281. (2008).

Allarakhia M. Developing a framework for understanding and enabling open source drug discovery. *Expert Opin Drug Discov*. **5**, 709-714. (2010).

Allarakhia M. Open-source approaches for the repurposing of existing or failed candidate drugs: learning from and applying the lessons across diseases. *Drug Des Devel Ther*. **7**, 753-766. (2013).

Allarakhia M., Ajuwon L. Understanding and creating value from open source drug discovery for neglected tropical diseases. *Expert Opin Drug Discov*. **7**, 643-657. (2012).

Allison M. NCATS launches drug repurposing program. *Nat Biotechnol*. **30**, 571-572. (2012).

Appolinario J.C., Fontenelle L.F., Papelbaum M., Bueno J.R., Coutinho W. Topiramate use in obese patients with binge eating disorder: an open study. *Can J Psychiatry*. **47**, 271-273. (2002).

Arrowsmith J., Harrison R. Drug Repositioning: The Business Case and Current Strategies to Repurpose Shelved Candidates and Marketed Drugs. En: *Drug Repositioning*, Barratt J. and Frail, D.E (Eds). Editorial John Wiley & Sons, Inc. pp. 7-32. (2012).

Ashburn T.T., Thor K.B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. **3**, 673-683. (2004).

Aubé J. Drug Repurposing and the Medicinal Chemist. *ACS Med. Chem. Lett*. **3**, 443-444. (2012).

Bellera C.L., Balcazar D.E., Alberca L., Labriola C.A., Talevi A., Carrillo C. Application of computer-aided drug repurposing in the search of new cruzipain inhibitors: discovery of amiodarone and bromocriptine inhibitory effects. *J Chem Inf Model*. **53**, 2402-2408. (2013).

Bern C., Adler-Moore J., Berenguer J., Boelaert M., den Boer M., Davidson R.N., Figueras C., Gradoni L., Kafetzis D.A., Ritmeijer K., Rosenthal E., Royce C., Russo R., Sundar S., Alvar J. Liposomal amphotericin B for the treatment of visceral leishmaniasis. *Clin Infect Dis*. **43**, 917-924. (2006).

Bost F., Jacobs R.T., Kowalczyk P. Informatics for neglected diseases collaborations. *Curr Opin Drug Discov Devel*. **13**, 286-296. (2010).

Broder S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral Res.* **85**, 1-18. (2010).

Burmester G.R., Feist E., Dorner T. Emerging cell and cytokine targets in rheumatoid arthritis. *Nat Rev Rheumatol.* **10**, 77-88. (2014).

Byrne R.A., Neumann F.J., Mehilli J., Piniack S., Wolff B., Tiroch K., Schulz S., Fusaro M., Ott I., Ibrahim T., Hausleiter J., Valina C., Pache J., Laugwitz K.L., Massberg S., Kastrati A. Paclitaxel-eluting balloons, paclitaxel-eluting stents, and balloon angioplasty in patients with restenosis after implantation of a drug-eluting stent (ISAR-DESIRE 3): a randomised, open-label trial. *Lancet.* **381**, 461-467. (2013).

Cavalla D. Predictive methods in drug repurposing: gold mine or just a bigger haystack? *Drug Discov Today.* **18**, 523-532. (2013).

Chen B., Ding Y., Wild D.J. Assessing drug target association using semantic linked data. *PLoS Comput Biol.* **8**, e1002574. (2012).

Chirac P., Torreale E. Global framework on essential health R&D. *Lancet.* **367**, 1560-1561. (2006).

Chong C.R., Chen X., Shi L., Liu J.O., Sullivan D.J., Jr. A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat Chem Biol.* **2**, 415-416. (2006).

Cohen T., Widdows D., Schvaneveldt R.W., Davies P., Rindflesch T.C. Discovering discovery patterns with Predication-based Semantic Indexing. *J Biomed Inform.* **45**, 1049-1065. (2012).

Coles L.D., Cloyd J.C. The role of academic institutions in the development of drugs for rare and neglected diseases. *Clin Pharmacol Ther.* **92**, 193-202. (2012).

Coukell A.J., Brogden R.N. Liposomal amphotericin B. Therapeutic use in the management of fungal infections and visceral leishmaniasis. *Drugs.* **55**, 585-612. (1998).

Cranney A., Adachi J.D. Benefit-risk assessment of raloxifene in postmenopausal osteoporosis. *Drug Saf.* **28**, 721-730. (2005).

Das S., Crockett J.C. Osteoporosis - a current view of pharmacological prevention and treatment. *Drug Des Devel Ther.* **7**, 435-448. (2013).

de Carvalho L.P., Lin G., Jiang X., Nathan C. Nitazoxanide kills replicating and nonreplicating Mycobacterium tuberculosis and evades resistance. *J Med Chem.* **52**, 5789-5792. (2009).

Deftereos S.N., Andronis C., Friedla E.J., Persidis A. Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip Rev Syst Biol Med.* **3**, 323-334. (2011).

Di Ianni M., Talevi A. How Can Network-Pharmacology Contribute to Antiepileptic Drug Development? *Molecular & Cellular Epilepsy.* **1**, 19-25. (2014).

Drews J. Drug discovery: a historical perspective. *Science.* **287**, 1960-1964. (2000).

Dudley J.T., Sirota M., Shenoy M., Pai R.K., Roedder S., Chiang A.P., Morgan A.A., Sarwal M.M., Pasricha P.J., Butte A.J. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med.* **3**, 96ra76. (2011).

Ekins S., Williams A.J., Krasowski M.D., Freundlich J.S. In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discov Today.* **16**, 298-310. (2011).

Emery P., Fleischmann R., Filipowicz-Sosnowska A., Schechtman J., Szczepanski L., Kavanaugh A., Racewicz A.J., van Vollenhoven R.F., Li N.F., Agarwal S., Hesse E.W., Shaw T.M. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment: results of a phase IIB randomized, double-blind, placebo-controlled, dose-ranging trial. *Arthritis Rheum.* **54**, 1390-1400. (2006).

Ferri N., Siegl P., Corsini A., Herrmann J., Lerman A., Benghozi R. Drug attrition during pre-clinical and clinical development: understanding and managing drug-induced cardiotoxicity. *Pharmacol Ther.* **138**, 470-484. (2013).

Freitas-Junior L.H., Chatelain E., Andrade Kim H., Siqueira-Neto J.L. Visceral leishmaniasis treatment: What do we have, what do we need and how to deliver it? . *Int J Parasitol Drugs Drug Resist.* **2**, 11-19. (2012).

Frijters R., van Vugt M., Smeets R., van Schaik R., de Vlieg J., Alkema W. Literature mining for the discovery of hidden connections between drugs, genes and diseases. *PLoS Comput Biol.* **6**. (2010).

Gershlick A., De Scheerder I., Chevalier B., Stephens-Lloyd A., Camenzind E., Vrints C., Reifart N., Missault L., Goy J.J., Brinker J.A., Raizner A.E., Urban P., Heldman A.W. Inhibition of restenosis with a paclitaxel-eluting, polymer-free coronary stent: the European evaluation of paclitaxel eluting stent (ELUTES) trial. *Circulation.* **109**, 487-493. (2004).

Gfeller D., Grosdidier A., Wirth M., Daina A., Michielin O., Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* **42** (2014).

Ghofrani H.A., Osterloh I.H., Grimminger F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat Rev Drug Discov.* **5**, 689-702. (2006).

Good A.C., Hermsmeier M.A., Hindle S.A. Measuring CAMD technique performance: a virtual screening case study in the design of validation experiments. *J Comput Aided Mol Des.* **18**, 529-536. (2004).

Griggs R.C., Batshaw M., Dunkle M., Gopal-Srivastava R., Kaye E., Krischer J., Nguyen T., Paulus K., Merkel P.A. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab.* **96**, 20-26. (2009).

Holliday J.D., Kanoulas E., Malim N., Willett P. Multiple search methods for similarity-based virtual screening: analysis of search overlap and precision. *J Cheminform.* **3**, 29. (2011).

Holt R.I., Barnett A.H., Bailey C.J. Bromocriptine: old drug, new formulation and new indication. *Diabetes Obes Metab.* **12**, 1048-1057. (2010).

Hopkins A.L. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* **4**, 682-690. (2008).

Hopkins A.L. Drug discovery: Predicting promiscuity. *Nature.* **462**, 167-168. (2009).

Hurlen M., Erikssen J., Smith P., Arnesen H., Rollag A. Comparison of bleeding complications of warfarin and warfarin plus acetylsalicylic acid: a study in 3166 outpatients. *J Intern Med.* **236**, 299-304. (1994).

Hurlen M., Abdelnoor M., Smith P., Erikssen J., Arnesen H. Warfarin, aspirin, or both after myocardial infarction. *N Engl J Med.* **347**, 969-974. (2002).

Iskar M., Zeller G., Zhao X.M., van Noort V., Bork P. Drug discovery in the age of systems biology: the rise of computational approaches for data integration. *Curr Opin Biotechnol.* **23**, 609-616. (2012).

Issa N.T., Kruger J., Byers S.W., Dakshanamurthy S. Drug repurposing a reality: from computers to the clinic. *Expert Rev Clin Pharmacol.* **6**, 95-97. (2013).

Jendrossek V. Targeting apoptosis pathways by Celecoxib in cancer. *Cancer Lett.* **332**, 313-324. (2013).

Jensen L.J., Saric J., Bork P. Literature mining for the biologist: from information retrieval to biological discovery. *Nat Rev Genet.* **7**, 119-129. (2006).

Kaufmann S.H. Paul Ehrlich: founder of chemotherapy. *Nat Rev Drug Discov.* **7**, 373. (2008).

Kell D.B. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: where drug discovery went wrong and how to fix it. *FEBS J.* **280**, 5957-5980. (2013).

Kiefer F., Arnold K., Kunzli M., Bordoli L., Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**, D387-392. (2009).

Kim C.K., Joe Y.A., Lee S.K., Kim E.K., O E., Kim H.K., Oh B.J., Hong S.H., Hong Y.K. Enhancement of anti-tumor activity by low-dose combination of the recombinant urokinase kringle domain and celecoxib in a glioma model. *Cancer Lett.* **288**, 251-260. (2010).

King R.S. Gemcitabine. New first-line therapy for pancreatic cancer. *Cancer Pract.* **4**, 353-354. (1996).

Kobylinski K.C., Sylla M., Chapman P.L., Sarr M.D., Foy B.D. Ivermectin mass drug administration to humans disrupts malaria parasite transmission in Senegalese villages. *Am J Trop Med Hyg.* **85**, 3-5. (2011).

Koonin E.V. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* **39**, 309-338. (2005).

Kruger F.A., Overington J.P. Global analysis of small molecule binding to related protein targets. *PLoS Comput Biol.* **8**, e1002333. (2012).

Lamb J., Crawford E.D., Peck D., Modell J.W., Blat I.C., Wrobel M.J., Lerner J., Brunet J.P., Subramanian A., Ross K.N., Reich M., Hieronymus H., Wei G., Armstrong S.A., Haggarty S.J., Clemons P.A., Wei R., Carr S.A., Lander E.S., Golub T.R. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. **313**, 1929-1935. (2006).

Lee J.A., Uhlik M.T., Moxham C.M., Tomandl D., Sall D.J. Modern phenotypic drug discovery is a viable, neoclassic pharma strategy. *J Med Chem*. **55**, 4527-4538. (2012).

Lekka E., Deftereos S., Persidis A., Persidis A., Andronis C. Literature analysis for systematic drug repurposing: a case study from Biovista. *Drug Discovery Today*. **8**, 103-108. (2011).

Lerman C., Niaura R., Collins B.N., Wileyto P., Audrain-McGovern J., Pinto A., Hawk L., Epstein L.H. Effect of bupropion on depression symptoms in a smoking cessation clinical trial. *Psychol Addict Behav*. **18**, 362-366. (2004).

Leyden J., Dunlap F., Miller B., Winters P., Lebwohl M., Hecker D., Kraus S., Baldwin H., Shalita A., Draelos Z., Markou M., Thiboutot D., Rapaport M., Kang S., Kelly T., Pariser D., Webster G., Hordinsky M., Rietschel R., Katz H.I., Terranella L., Best S., Round E., Waldstreicher J. Finasteride in the treatment of men with frontal male pattern hair loss. *J Am Acad Dermatol*. **40**, 930-937. (1999).

Lilienfeld S. Galantamine--a novel cholinergic drug with a unique dual mode of action for the treatment of patients with Alzheimer's disease. *CNS Drug Rev*. **8**, 159-176. (2002).

Liu Z., Fang H., Reagan K., Xu X., Mendrick D.L., Slikker W., Jr., Tong W. In silico drug repositioning: what we need to know. *Drug Discov Today*. **18**, 110-115. (2013).

Loewenstein Y., Raimondo D., Redfern O.C., Watson J., Frishman D., Linial M., Orengo C., Thornton J., Tramontano A. Protein function annotation by homology-based inference. *Genome Biol*. **10**, 207. (2009).

Lussier Y.A., Chen J.L. The emergence of genome-based drug repositioning. *Sci Transl Med*. **3**, 96ps35. (2011).

Ma D.L., Chan D.S., Leung C.H. Drug repositioning by structure-based virtual screening. *Chem Soc Rev*. **42**, 2130-2141. (2013).

Marona-Lewicka D., Nichols D.E. Further evidence that the delayed temporal dopaminergic effects of LSD are mediated by a mechanism different than the first temporal phase of action. *Pharmacol Biochem Behav*. **87**, 453-461. (2007).

Marona-Lewicka D., Nichols D.E. WAY 100635 produces discriminative stimulus effects in rats mediated by dopamine D(4) receptor activation. *Behav Pharmacol*. **20**, 114-118. (2009).

Marusina K., Welsch D.J., Rose L., Brock D., Bahr N. The CTSA Pharmaceutical Assets Portal - a public-private partnership model for drug repositioning. *Drug Discov Today Ther Strateg*. **8**, 77-83. (2011).

McElroy S.L., Arnold L.M., Shapira N.A., Keck P.E., Jr., Rosenthal N.R., Karim M.R., Kamin M., Hudson J.I. Topiramate in the treatment of binge eating disorder associated with obesity: a randomized, placebo-controlled trial. *Am J Psychiatry*. **160**, 255-261. (2003).

Meadows W.A., Hollowell B.D. 'Off-label' drug use: an FDA regulatory term, not a negative implication of its medical use. *Int J Impot Res*. **20**, 135-144. (2008).

Medina-Franco J.L., Giulianotti M.A., Welmaker G.S., Houghten R.A. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov Today*. **18**, 495-501. (2013).

Melkote S., Dhurat R.S., Palav A., Jerajani H.R. Alopecia in congenital hidrotic ectodermal dysplasia responding to treatment with a combination of topical minoxidil and tretinoin. *Int J Dermatol*. **48**, 184-185. (2009).

Merino A., Bronowska A.K., Jackson D.B., Cahill D.J. Drug profiling: knowing where it hits. *Drug Discov Today*. **15**, 749-756. (2010).

Miguel D.C., Yokoyama-Yasunaka J.K., Uliana S.R. Tamoxifen is effective in the treatment of *Leishmania amazonensis* infections in mice. *PLoS Negl Trop Dis*. **2**, e249. (2008).

Moran M., Guzman J., Ropars A.L., McDonald A., Jameson N., Omune B., Ryan S., Wu L. Neglected disease research and development: how much are we really spending? *PLoS Med*. **6**, e30. (2009).

Morgan S., Grootendorst P., Lexchin J., Cunningham C., Greyson D. The cost of drug development: a systematic review. *Health Policy*. **100**, 4-17. (2011).

Murteira S., Ghezaiel Z., Karray S., Lamure M. Drug reformulations and repositioning in pharmaceutical industry and its impact on market access: reassessment of nomenclature. *J Mark Access Health Pol*. **1**, 21131. (2013).

Napolitano F., Zhao Y., Moreira V.M., Tagliaferri R., Kere J., D'Amato M., Greco D. Drug repositioning: a machine-learning approach through data integration. *J Cheminform*. **5**, 30. (2013).

No authors listed. One world problem, one world solution? *Nature Reviews Drug Discovery*. **4**, 701. (2005).

Nobeli I., Favia A.D., Thornton J.M. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. **27**, 157-167. (2009).

Novac N. Challenges and opportunities of drug repositioning. *Trends Pharmacol Sci*. **34**, 267-272. (2013).

Oldfield E. Targeting isoprenoid biosynthesis for drug discovery: bench to bedside. *Acc Chem Res*. **43**, 1216-1226. (2010).

Olsen E.A., Dunlap F.E., Funicella T., Koperski J.A., Swinehart J.M., Tschen E.H., Trancik R.J. A randomized clinical trial of 5% topical minoxidil versus 2% topical minoxidil and placebo in the treatment of androgenetic alopecia in men. *J Am Acad Dermatol*. **47**, 377-385. (2002).

Paricharak S., Klenka T., Augustin M., Patel U.A., Bender A. Are phylogenetic trees suitable for chemogenomics analyses of bioactivity data sets: the importance of shared active compounds and choosing a suitable data embedding method, as exemplified on Kinases. *J Cheminform.* **5**, 49. (2013).

Patrono C. Aspirin as an antiplatelet drug. *N Engl J Med.* **330**, 1287-1294. (1994).

Patrono C., Garcia Rodriguez L.A., Landolfi R., Baigent C. Low-dose aspirin for the prevention of atherothrombosis. *N Engl J Med.* **353**, 2373-2383. (2005).

Pearlstein T.B., Stone A.B., Lund S.A., Scheft H., Zlotnick C., Brown W.A. Comparison of fluoxetine, bupropion, and placebo in the treatment of premenstrual dysphoric disorder. *J Clin Psychopharmacol.* **17**, 261-266. (1997).

Perez-Nueno V.I., Venkatraman V., Mavridis L., Ritchie D.W. Predicting Drug Promiscuity Using Spherical Harmonic Surface Shape-Based

Similarity Comparisons. *The Open Conference Proceedings Journal.* **2**, 113-129. (2011).

Peters J.U. Polypharmacology - foe or friend? *J Med Chem.* **56**, 8955-8971. (2013).

Peterson R.T. Chemical biology and the limits of reductionism. *Nat Chem Biol.* **4**, 635-638. (2008).

Pollastri M.P., Campbell R.K. Target repurposing for neglected diseases. *Future Med Chem.* **3**, 1307-1315. (2011).

Qu X.A., Rajpal D.K. Applications of Connectivity Map in drug discovery and development. *Drug Discov Today.* **17**, 1289-1298. (2012).

Radisch J. More medicines for neglected and emerging infectious diseases. *Bull World Health Organ.* **85**, 572. (2007).

Richarz U., Gaudig M., Rettig K., Schauble B. Galantamine treatment in outpatients with mild Alzheimer's disease. *Acta Neurol Scand.* **129**, 382-392. (2014).

Roberts R.A., Kavanagh S.L., Mellor H.R., Pollard C.E., Robinson S., Platz S.J. Reducing attrition in drug development: smart loading preclinical safety assessment. *Drug Discov Today.* **19**, 341-347. (2014).

Rossi A., Cantisani C., Scarno M., Trucchia A., Fortuna M.C., Calvieri S. Finasteride, 1 mg daily administration on male androgenetic alopecia in different age groups: 10-year follow-up. *Dermatol Ther.* **24**, 455-461. (2011).

Roth B.L., Sheffler D.J., Kroeze W.K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov.* **3**, 353-359. (2004).

Salazar G.A., Meintjes A., Mazandu G.K., Rapanoel H.A., Akinola R.O., Mulder N.J. A web-based protein interaction network visualizer. *BMC Bioinformatics.* **15**, 129. (2014).

Sams-Dodd F. Target-based drug discovery: is something wrong? *Drug Discov Today*. **10**, 139-147. (2005).

Sardana D., Zhu C., Zhang M., Gudivada R.C., Yang L., Jegga A.G. Drug repositioning for orphan diseases. *Brief Bioinform*. **12**, 346-356. (2011).

Schrag A., Keens J., Warner J. Ropinirole for the treatment of tremor in early Parkinson's disease. *Eur J Neurol*. **9**, 253-257. (2002).

Sirota M., Dudley J.T., Kim J., Chiang A.P., Morgan A.A., Sweet-Cordero A., Sage J., Butte A.J. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med*. **3**, 96ra77. (2011).

Sliwoski G., Kothiwale S., Meiler J., Lowe E.W., Jr. Computational methods in drug discovery. *Pharmacol Rev*. **66**, 334-395. (2014).

Steinbach G., Lynch P.M., Phillips R.K., Wallace M.H., Hawk E., Gordon G.B., Wakabayashi N., Saunders B., Shen Y., Fujimura T., Su L.K., Levin B., Godio L., Patterson S., Rodriguez-Bigas M.A., Jester S.L., King K.L., Schumacher M., Abbruzzese J., DuBois R.N., Hittelman W.N., Zimmerman S., Sherman J.W., Kelloff G. The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis. *N Engl J Med*. **342**, 1946-1952. (2000).

Steiner M., Steinberg S., Stewart D., Carter D., Berger C., Reid R., Grover D., Streiner D. Fluoxetine in the treatment of premenstrual dysphoria. Canadian Fluoxetine/Premenstrual Dysphoria Collaborative Study Group. *N Engl J Med*. **332**, 1529-1534. (1995).

Stevens A.J., Jensen J.J., Wyller K., Kilgore P.C., Chatterjee S., Rohrbaugh M.L. The role of public-sector research in the discovery of drugs and vaccines. *N Engl J Med*. **364**, 535-541. (2011).

Stocchi F., Radicati F.G., Torti M. Drug safety evaluation of ropinirole prolonged release. *Expert Opin Drug Saf*. **13**, 383-389. (2014).

Swanson D.R. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med*. **30**, 7-18. (1986).

Swanson D.R. Migraine and magnesium: eleven neglected connections. *Perspect Biol Med*. **31**, 526-557. (1988).

Swinney D.C., Anthony J. How were new medicines discovered? *Nat Rev Drug Discov*. **10**, 507-519. (2011).

Talevi A., Castro E.A., Bruno-Blanch L.E. Virtual Screening: An emergent, key methodology for drug development in an emergent continent – A bridge towards patentability. En: *Advanced Methods and Applications in Chemoinformatics*, E. A. Castro and A. K. Haghi (Eds.), Editorial IGI Global, pp. 229-245 (2011)

Talevi A., Bellera C.L., Di Ianni M., Gantner M., Bruno-Blanch L.E., Castro E.A. CNS drug development - lost in translation? *Mini Rev Med Chem*. **12**, 959-970. (2012).

Thornton S.J., Wasan K.M. The reformulation of amphotericin B for oral administration to treat systemic fungal infections and visceral leishmaniasis. *Expert Opin Drug Deliv.* **6**, 271-284. (2009).

Vidal M., Cusick M.E., Barabasi A.L. Interactome networks and human disease. *Cell.* **144**, 986-998. (2011).

Voelker R. International group seeks to dispel incontinence "taboo". *JAMA.* **280**, 951-953. (1998).

Wadman M. NIH gambles on recycled drugs. *Nature.* **499**, 263-264. (2013).

Weeber M., Kors J.A., Mons B. Online tools to support literature-based discovery in the life sciences. *Brief Bioinform.* **6**, 277-286. (2005).

Wu L., Ai N., Liu Y., Wang Y., Fan X. Relating anatomical therapeutic indications by the ensemble similarity of drug sets. *J Chem Inf Model.* **53**, 2154-2160. (2013a).

Wu Z., Wang Y., Chen L. Network-based drug repositioning. *Mol. BioSyst.* **9**, 1268-1281. (2013b).

Zhang Q., Muegge I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem.* **49**, 1536-1548. (2006).

CAPÍTULO 4

METODOLOGÍA



4.1. GENERACIÓN DE MODELOS QSAR

4.1.1. CONSTRUCCIÓN DE LA BASE DE DATOS Y PARTICIÓN EN CONJUNTOS DE ENTRENAMIENTO Y DE PRUEBA

Las relaciones QSAR son modelos químico-matemáticos que establecen una relación cuantitativa entre la estructura química y una propiedad biológica de interés; estas relaciones pueden ser utilizadas para predecir dicha propiedad biológica (por ejemplo IC_{50}) o para discriminar compuestos químicos activos de compuestos químicos inactivos. Su desarrollo comprende fundamentalmente cuatro pasos (Figura 4.1) (Yee and Wei, 2012): 1) Compilación de la *base de datos, set de datos o dataset* de compuestos químicos cuya actividad o clase (ACTIVOS, INACTIVOS) haya sido experimentalmente determinada; 2) Cálculo de *los descriptores moleculares asociados* a la estructura molecular; 3) Partición de la base de datos en un *conjunto de entrenamiento o training set* (a partir del cual se inferirá estadísticamente el modelo) y un *conjunto de prueba o test set* que se usará para estimar la capacidad predictiva del modelo (la capacidad de generalización, la habilidad del modelo de predecir la categoría de casos o instancias no incluidos en el conjunto de entrenamiento); 4) Generación y validación del modelo propiamente dicho.

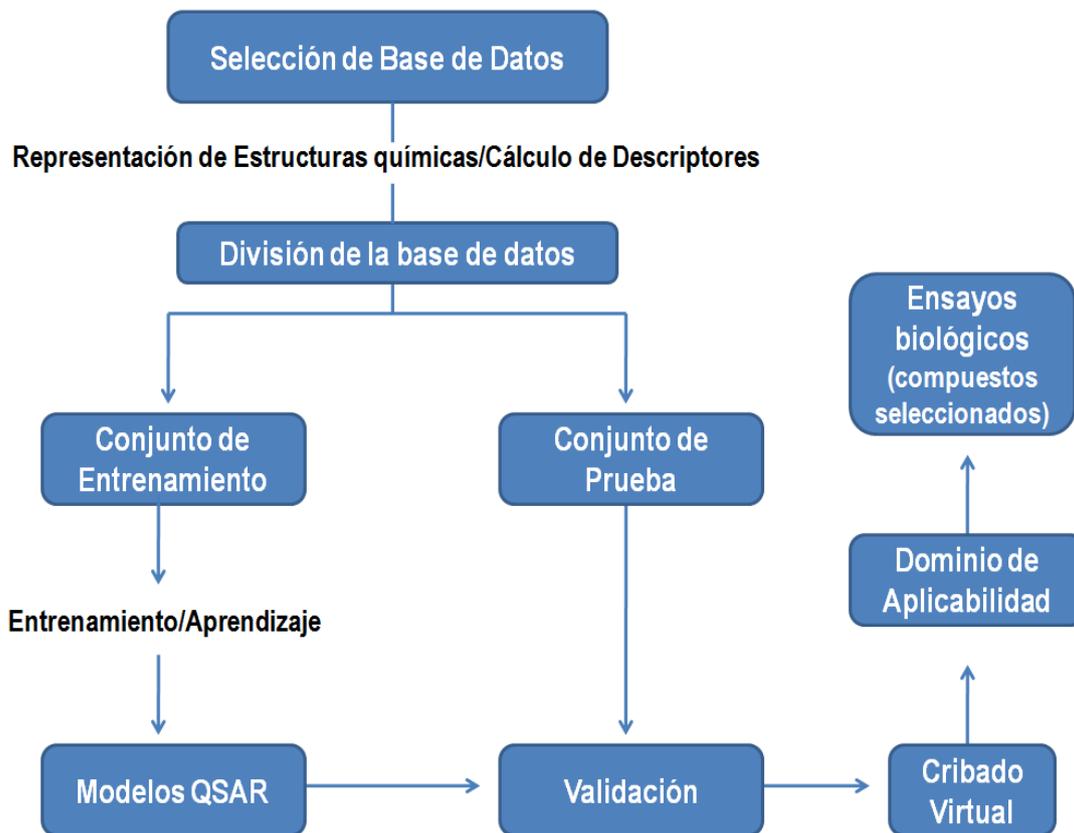


Figura 4.1. Esquema General de las etapas de desarrollo de un modelo QSAR y su aplicación en campañas de CV.

4.1.1.1. PARTICIÓN DE LA BASE DE DATOS

Un problema relevante en la generación de modelos QSAR 2D es cómo particionar la base de datos en un conjunto de entrenamiento y un conjunto de prueba representativos. Se ha demostrado que la partición al azar es una aproximación adecuada cuando se seleccionan conjuntos de entrenamiento y de prueba de tamaño similar; en cambio, cuando los conjuntos de prueba son relativamente pequeños, aproximaciones de muestreo más racionales proveen mejores resultados (Golbraikh and Tropsha, 2002a; Golbraikh et al., 2003)

De modo general existen tres grandes aproximaciones para seleccionar conjuntos de moléculas representativos de un espacio químico determinado, aunque esta clasificación

no es estricta en el sentido de que diferentes autores proponen distintas formas de clasificación (Perez, 2005; Willett, 2000):

a) Análisis de agrupamientos (*clustering*): consiste en dividir un grupo de objetos (en nuestro caso estructuras químicas) en grupos o *clusters* de manera que éstos presenten un alto grado de similitud *intra-cluster* y de disimilitud *inter-cluster*. Seleccionando objetos pertenecientes a cada *cluster*, se obtiene una excelente representatividad del espacio químico.

b) Métodos de partición: clasifican el conjunto de moléculas considerado por división recursiva de los rangos de las diferentes propiedades que describen el espacio químico. Esta aplicación es de menor costo computacional comparada con los métodos de *clustering* y suele aplicarse a base de datos muy grandes. Este enfoque requiere la identificación de un pequeño número de características o propiedades moleculares que se utilizarán como criterio de partición. El rango de valores para cada una de esas características se sub-divide en un conjunto de sub-rangos, y el producto de la combinación de todos los posibles sub-rangos luego define el conjunto de celdas (*bins*) que forman la partición. Cada molécula se asigna a la celda que coincide con el conjunto de características para esa molécula, y un subconjunto se obtiene entonces mediante la selección de un número o fracción determinada de las moléculas de cada una de las celdas resultantes.

c) Métodos basados en distancias o disimilitud: el objetivo de este método es identificar un subconjunto de las n moléculas disimiles pertenecientes a una base de datos de N compuestos. El compuesto inicial puede ser escogido aleatoriamente o puede corresponder al más disimilar de la base de datos analizada.

En el trabajo de tesis se empleó una combinación de técnicas de *clustering* para particionar el dataset en conjuntos de entrenamiento y prueba representativos.

4.1.1.2. ANÁLISIS DE AGRUPAMIENTOS (CLUSTERING)

El análisis de Agrupamientos (*Clustering*) es una técnica de análisis de datos que, cuando se aplica a un conjunto de elementos heterogéneos, identifica subgrupos homogéneos definidos por un determinado algoritmo o medida de similitud (Downs and Barnard, 2003). Entre los métodos de agrupamiento existentes, se aplican mayormente aquellos que no producen superposición, es decir, cada estructura química es asignada a un único *cluster*. Los mismos se diferencian en Jerárquicos y No Jerárquicos, a continuación se detallan los métodos empleados en el presente trabajo: LibraryMCS (Agrupamiento Jerárquico) y el método *k-means* (Algoritmo de Optimización de Agrupamiento No Jerárquico).

LibraryMCS v0.7 (ChemAxon): Se basa en la máxima subestructura común (MSC) obtenida a partir de pares de estructuras similares para agrupar conjuntos de estructuras sin recurrir a una comparación exhaustiva (y por ende computacionalmente demandante). Las estructuras moleculares se representan como grafos, en los que un átomo es representado como un vértice y un enlace covalente como un eje o arista. Una subestructura común es definida como una subestructura o subgrafo presente en dos moléculas, conteniendo los mismos tipos de átomos y de enlaces químicos. La MSC es la subestructura común que contiene el mayor número de átomos (ver como ejemplo Figura 4.2) (Kawabata, 2011). LibraryMCS construye una matriz de similitud para las estructuras consideradas y en función de los coeficientes de similitud (Talevi et al., 2012a) elige el par de estructuras más similares, determinando luego la MSC de éstas, y agrupando aquellas moléculas que presentan esta subestructura común. El proceso se repite iterativamente hasta que no se encuentran más pares de estructuras con un valor del coeficiente de similitud por encima del umbral utilizado por el algoritmo, en cuyo caso tales estructuras se presentan aisladas, es decir, no se incluyen en ninguno de los agrupamientos o *clusters* (Figura 4.3). La búsqueda de MSC basada en similitud química conduce a soluciones reproducibles pero aproximadas, ya que no hay garantía de que aquellas estructuras con mayor valor del coeficiente de similitud contengan, efectivamente, la MSC (Hariharan et al., 2011) (Figura 4.3).

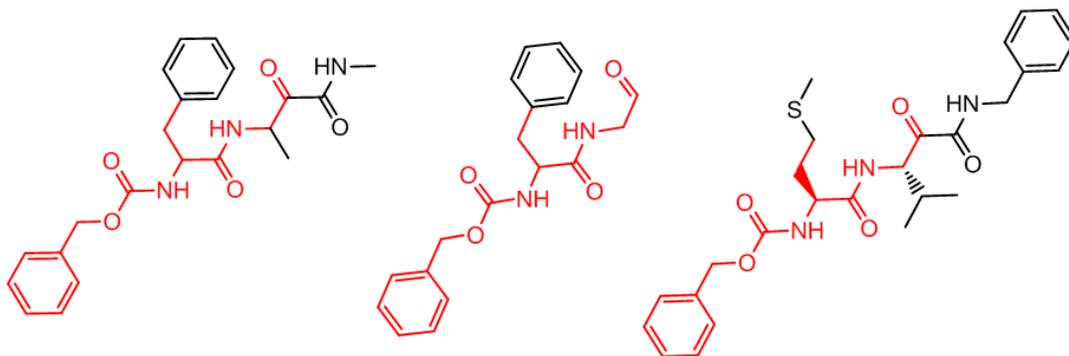


Figura 4.2. Se presenta coloreada en rojo, en las tres estructuras, un ejemplo de MSC.

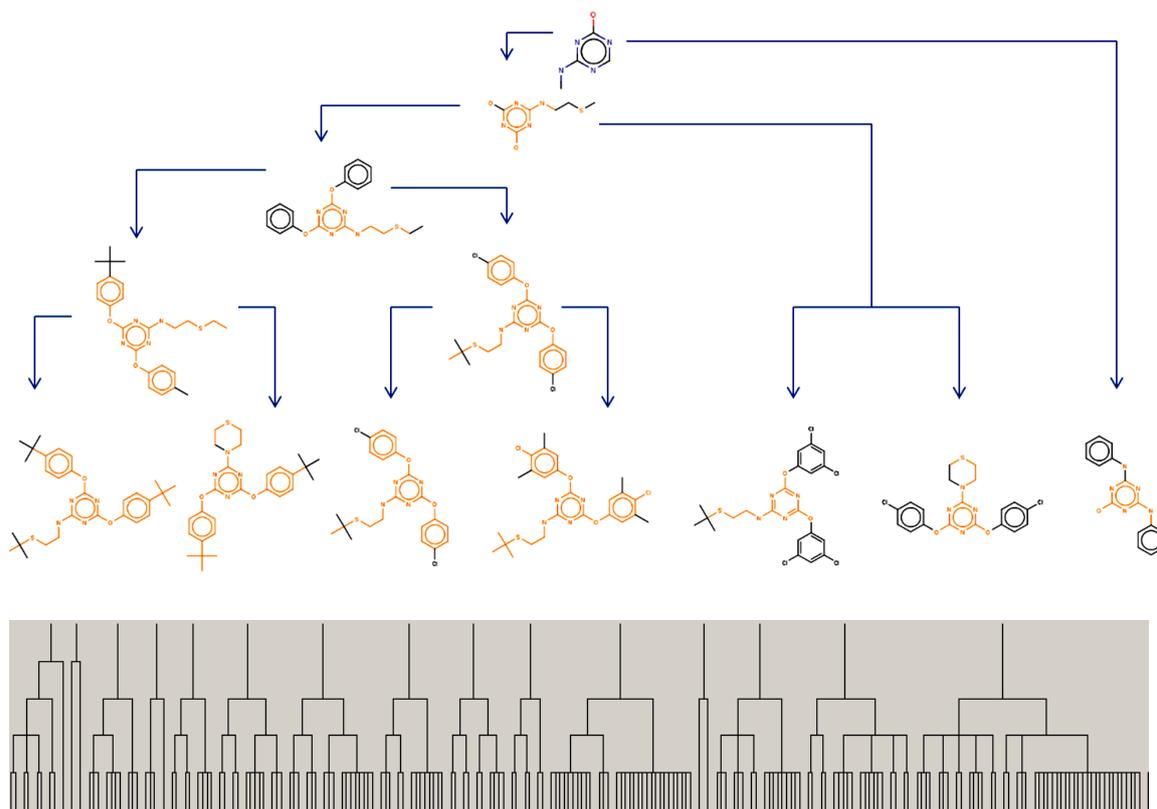


Figura 4.3. Ejemplo de agrupamiento basado en MSCs. Las estructuras de la base de datos que no poseen una subestructura común con el resto con un mínimo número de átomos especificado por el usuario, son representadas de manera aislada en el dendrograma.

Método *k-means*: involucra un algoritmo de optimización de agrupamientos que – como cualquier algoritmo de optimización de agrupamientos- obedece los siguientes pasos generales: a) encontrar o proponer una partición inicial de n objetos en k grupos; b) calcular la variación en un criterio de agrupamiento cuando un objeto de un grupo se mueve a otro; c) realizar el cambio que conduce a la mayor mejora en el valor del criterio de agrupamiento; d) el proceso anterior se repite tantas veces como lo defina el usuario o hasta alcanzar convergencia (cuando ningún desplazamiento de objetos de un grupo a otro produzca una mejoría). En el caso particular del método *k-means*, el criterio de agrupamiento es la minimización de la suma de cuadrados intragrupo (Everitt et al., 2011):

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

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 \bar{x}_m es el vector que representa la media de los elementos del grupo m en el espacio p -dimensional.

4.1.1.3. METODOLOGÍA DE PARTICIÓN DE BASE DE DATOS EMPLEADA EN LA TESIS

Tal como se mencionó anteriormente, para realizar la partición de la base de datos en conjuntos de entrenamiento y de prueba se utilizó una combinación de metodologías de agrupamiento jerárquico y de optimización. A partir de la base de datos generada con compuestos que presentan inhibición (ACTIVOS) o no inhibición (INACTIVOS) sobre la enzima Cz, se procedió en primer lugar a realizar un análisis jerárquico utilizando la herramienta LibraryMCS determinando los clusters en base a la MSC, el análisis de clustering se realizó para las categorías ACTIVOS e INACTIVOS por separado. Se fijó que

las subestructuras comunes en torno a las cuales se desarrolló el agrupamiento jerárquico debían tener al menos 9 átomos manteniendo los anillos como tales.

En base al resultado se determinó el número de grupos de la partición inicial para el *clustering k-means*, siendo el correspondiente a los *clusters* más poblados del análisis en LibraryMCS (al menos 6 compuestos para la categoría Activos y 7 para INACTIVOS).

El *clustering* por *k-means* se realizó utilizando el módulo *Cluster Analysis* de Statistica 10.0. Las “semillas” (elementos que se asignan inicialmente a los grupos y a partir de los cuales se construyen los agrupamientos iniciales) fueron elegidas al azar a partir de los agrupamientos obtenidos vía LibraryMCS para las categorías ACTIVOS e INACTIVOS. La partición inicial se realiza localizando cada objeto en el *cluster* cuya semilla se haya más próxima en el espacio p-dimensional considerado. Posteriormente, cada objeto se relocaliza hacia el grupo cuya media le resulta más próxima. En el programa Statistica se utilizaron 10 iteraciones hasta convergencia. Para caracterizar a cada elemento (compuesto) y definir el espacio p-dimensional en el que se aplicaría *k-means*, se seleccionaron distintos descriptores moleculares representativos de diferentes aspectos de la molécula, calculados con Dragon 6.0 (Milano Chemometrics, 2011): peso molecular, logaritmo del coeficiente de reparto octanol agua (log P), área de la superficie polar, número de aceptores de enlaces de H, índice de información del contenido atómico, suma de volúmenes atómicos de van der Waals. El número de dadores de puentes de H no fue incluido pues tenía poca dispersión en el dataset utilizado. Estos descriptores fueron normalizados; el valor normalizado (d_N) corresponde al cociente entre la diferencia del valor del descriptor que se desea normalizar (d_i) y la media aritmética para el dataset (μ) sobre la desviación estándar del descriptor en el dataset (σ):

$$d_N = \frac{d_i - \mu}{\sigma}$$

Luego de normalizados, los descriptores arriba mencionados fueron aplicados para el cálculo de la distancia de cada elemento a la semilla de cada grupo (en la partición inicial) y a la media de su grupo tentativo (posteriormente).

Finalmente, una vez que se identificaron de manera separada los agrupamientos en las categorías ACTIVOS e INACTIVOS, el 25% de los elementos de cada grupo se asignó aleatoriamente al conjunto de prueba, y el 75% restante se retuvo como conjunto de entrenamiento. Los elementos no agrupados (que consideraremos atípicos/*outliers*) fueron asignados aleatoriamente a los conjuntos de prueba y entrenamiento siguiendo la misma proporción previamente definida.

4.1.2. CARACTERÍSTICAS DEL CONJUNTO DE ENTRENAMIENTO

Se denomina conjunto de entrenamiento al conjunto de estructuras químicas a partir de las cuales se deriva o infiere el modelo QSAR. Se trata de una muestra acotada de compuestos químicos de la cual se infiere una relación entre la estructura molecular y una propiedad de interés; se espera que esta relación sea generalizable a la “población” de compuestos químicos de la cual el conjunto de entrenamiento es representativo. Es decir, la selección del conjunto de entrenamiento es crítica ya que define el “espacio químico” dentro del cual la aplicación del modelo tendrá validez. El valor de la actividad/propiedad biológica predicha para un compuesto externo al conjunto de entrenamiento será confiable si y sólo si el compuesto posee cierta similitud con alguna o varias de las estructuras incluidas en el conjunto de entrenamiento. El espacio químico dentro del cual la aplicación del modelo QSAR es confiable se denomina “dominio de aplicabilidad” o “dominio de aplicación” del modelo (como hemos determinado el mismo se explica en la sección correspondiente del presente capítulo). En nuestro caso, deseamos aplicar los modelos generados en la selección de compuestos a partir de repositorios químicos que comprenden compuestos de gran heterogeneidad estructural. Siendo la población que deseamos analizar tan diversa, es conveniente que el conjunto de datos generado sea lo más inclusivo posible.

Otra característica que debe reunir el conjunto de entrenamiento es que los valores observados de la propiedad estudiada deberían presentar una buena distribución en un rango de entre tres y cuatro órdenes logarítmicos. Idealmente, la distribución debería ser uniforme; este requisito es difícil de cumplir y se considera adecuado que la distribución sea aproximadamente normal. Este requisito tiene un fundamento lógico:

pensemos por ejemplo que si tenemos un conjunto de entrenamiento formado solamente por compuestos que poseen alta y mediana actividad, el modelo QSAR generado sólo va a predecir con exactitud valores de compuestos muy activos y moderadamente activos, pero no nos permitirá diferenciar compuestos activos de inactivos (es decir, no identificará qué características estructurales son desfavorables a la actividad por no estar los compuestos poco activos e inactivos representados dentro del conjunto de entrenamiento del que se infiere el modelo). En palabras de Jaworska y colaboradores, el dominio de aplicabilidad no sólo se refiere a las variables independientes del modelo, sino también a la respuesta o variable dependiente del mismo (Jaworska et al., 2005).

Un último requisito esencial es que los compuestos del conjunto de entrenamiento posean un mecanismo de acción común, esto es, interaccionen con el mismo receptor o blanco molecular y presenten el mismo sitio de unión.

4.1.3. CÁLCULO DE DESCRIPTORES MOLECULARES

Las estructuras moleculares de toda la base de datos fueron representadas en Hyperchem 8.0 (HyperCube INC, 2007); luego para cada una de las estructuras que conforman el conjunto de entrenamiento se calculó un pool de descriptores (0D-2D). Se utilizaron para las diferentes campañas de CV (detalladas en el capítulo 5) los programas de cálculo de descriptores: Dragon v4.0 (Milano Chemometrics, 2003); Dragon v6.0 (Milano Chemometrics, 2011) y Desmol11 (Unidad de Topología Molecular y Diseño de Fármacos, Universidad de Valencia, España).

4.1.4. CONSTRUCCIÓN Y VALIDACIÓN DEL MODELO

Desarrollar un buen modelo QSAR con una variable cuantitativa para ser aplicado en campañas de CV exige la existencia de un grupo de compuestos que presenten diversidad estructural para el cual la propiedad modelada haya sido determinada experimentalmente bajo las mismas condiciones; adicionalmente, se requiere que los

valores de dicha variable se encuentren bien distribuidos a lo largo de tres o cuatro órdenes de magnitud. Como se mencionó ya, el requisito de la diversidad estructural del conjunto de entrenamiento es particularmente importante cuando se desea aplicar el modelo en campañas de cribado, ya que el mismo se utilizará para predecir la propiedad para un conjunto de estructuras grande y generalmente diverso. Por otro lado, se puede asumir que los errores aleatorios que el modelo producirá al predecir la propiedad serán mayores o iguales a los errores aleatorios presentes en los datos experimentales; por lo tanto, se desea disponer de datos experimentales sin errores sistemáticos o aleatorios significativos.

En nuestro caso, el conjunto de entrenamiento fue compilado a partir de literatura, lo que supone una limitación para desarrollar un modelo QSAR con una variable dependiente continua (por ejemplo, IC_{50} o $\log IC_{50}$), dado que los datos reportados no corresponden a las mismas condiciones experimentales (proviene de diferentes ensayos en distintos laboratorios). Para superar esta limitación se recurrió a la técnica conocida como **Análisis Lineal Discriminante (ALD)** (Yee and Wei, 2012) a fin de obtener un modelo clasificador capaz de distinguir inhibidores de Cz de compuestos sin capacidad inhibitoria.

ALD es un método de aprendizaje supervisado destinado a encontrar una combinación lineal de variables independientes (en nuestro caso, descriptores moleculares) capaces de diferenciar entre objetos de dos o más categorías (en nuestro caso, dos: ACTIVOS e INACTIVOS). La función discriminante (FD) obtenida corresponde a un plano en el espacio g -dimensional (siendo g el número de descriptores incluidos en el modelo) que, idealmente, deja a uno y otro lado los compuestos activos e inactivos. La ecuación 4.1 presenta la forma general de una FD :

$$FD = \sum_{i=1}^g w_i d_i$$

siendo w_i el coeficiente de regresión asociado al descriptor molecular d_i (al normalizarlos, los coeficientes de regresión informan el peso de cada variable independiente en la diferenciación de las categorías consideradas).

Cada clase de objeto se asocia a un valor de una variable dependiente binaria arbitraria que funciona como etiqueta de las clases. En nuestro caso, el valor 1 se asoció a los compuestos ACTIVOS y el -1 a los INACTIVOS. Dado que la función que se busca no predecirá una variable continua sino la clase a la que pertenece cada elemento de un conjunto de objetos, se ha propuesto que el ALD y otras técnicas destinadas a encontrar modelos clasificadores pueden ser útiles para manejar datos ruidosos (por ejemplo, un punto final asociado a gran variabilidad o en aquellos casos, como el nuestro, en los que se compilan datos experimentales de distintos laboratorios) (Talevi et al., 2012c).

4.1.4.1. SELECCIÓN DE DESCRIPTORES INCLUIDOS EN EL MODELO

La técnica de *Stepwise forward* (selección hacia delante, paso a paso) se utilizó para la selección de los descriptores incluidos en los modelos (Van Waterbeemd, 1995). Dicha estrategia consiste en realizar una regresión por cuadrados mínimos por pasos hasta encontrar el mejor modelo. En cada etapa la variable independiente (en nuestro caso un descriptor) del pool de variables mejor correlacionada con la variable dependiente es incluida en el modelo de regresión; en el segundo paso se busca introducir en el modelo, de entre las variables independientes que quedan en el pool de descriptores, aquella que consigue explicar mejor la varianza que no ha sido explicada por la primera variable independiente; el proceso continúa hasta que la adición de una nueva variable no contribuye a una reducción significativa de la varianza sin explicar o hasta que el usuario lo defina por algún otro criterio.

Adicionalmente se utilizó un valor de *tolerancia* β de 0,5 para evitar incluir pares de descriptores altamente correlacionados. La tolerancia se define por:

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

donde R_j es el coeficiente de regresión entre la variable independiente j y todas las demás variables independientes/covariables incluidas en el modelo. La tolerancia es la inversa del Factor de Inflación de la Varianza (FIV); en general, se acepta que si el FIV es menor a 5 no existe alta multicolinealidad (Rawlings et al., 1998). En nuestro caso, un valor de tolerancia de 0,5 determinaría un FIV de 2.

4.1.4.2. CRITERIOS DE ELECCIÓN DEL MEJOR MODELO GENERADO POR ALD

Para la elección de la mejor FD se emplearon los siguientes criterios: a) Sólo se consideraron modelos cuyos coeficientes de regresión fueran significativos a un nivel de 0,05; b) evaluación del parámetro λ de Wilks; c) observación del principio de parsimonia; d) baja probabilidad de sobreajuste (en inglés, *overfitting*); e) evaluación de la capacidad predictiva de los modelos sobre el conjunto de prueba.

 *Estadístico λ de Wilks*: es un parámetro ampliamente utilizado para evaluar la habilidad clasificadora de una FD. Es un parámetro de distancia y se calcula a partir de la transformación escalar de las matrices de covarianza de las varianzas entre- e intra-grupo. En un análisis de clasificación en el que se contemplan dos categorías, el λ de Wilks se determina cómo la relación de la suma de cuadrados dentro de una categoría y la suma de cuadrados totales, es decir la dispersión total dentro de la categoría (Roy and Kar, 2014). El estadístico se define como:

$$\lambda \text{ Wilks} = \frac{\text{Suma de cuadrados intragrupo}}{\text{Suma de cuadrados totales}} = \frac{|S|}{|T|}$$

donde S es la matriz de varianzas-covarianzas combinada, calculada a partir de las matrices de varianzas-covarianzas de cada grupo y T es la matriz de varianzas-covarianzas total calculada sobre todos los casos como si pertenecieran a un único grupo. Cuando los grupos se encuentren superpuestos en el espacio multidimensional considerado, los valores del numerador y denominador serán aproximadamente iguales y su cociente será 1; a medida que los grupos se van separando, la variabilidad inter-grupos irá aumentando y la variabilidad intra-grupos se hará menor con respecto a la total, disminuyendo de esta forma el valor del cociente. Por lo tanto un λ de Wilks igual 0 corresponderá a una discriminación perfecta de las 2 categorías y λ igual a 1 indica una capacidad de clasificación nula (gran parecido entre los grupos).

-  *Principio de Parsimonia*: también conocido como Navaja de Occam, es un principio metodológico y filosófico, según el cual “No ha de presumirse la existencia de más cosas que las absolutamente necesarias” es decir, cuando dos explicaciones logran explicar una misma observación, la explicación más sencilla es la que se prefiere *a priori*. En este sentido frente a la situación de encontrar varios modelos con buenos desempeños estadísticos se optará por aquel que incluya menor número de descriptores.
-  *Sobreajuste (overfitting)*: se refiere al efecto de sobreentrenar un algoritmo de aprendizaje, es decir, que el modelo “memorice” las características del conjunto de entrenamiento (aumentando su poder explicativo o descriptivo) a expensas de sacrificar poder predictivo/capacidad de generalización (capacidad predictiva sobre compuestos no utilizados para el entrenamiento). Se exigió una relación de al menos 10 entre el número de compuestos del conjunto de entrenamiento y el número de descriptores incorporados al modelo, de modo de reducir la probabilidad de sobreajuste (Peduzzi et al., 1995; Peduzzi et al., 1996).
-  *Validación Externa*: comprende evaluar la capacidad del modelo de predecir la clase de un conjunto de compuestos diferente al utilizado en la generación del modelo. En la siguiente sección se discute en detalle.

4.1.4.3. VALIDACIÓN DEL MODELO

El proceso de validación es la etapa crítica del modelado y su objetivo es evaluar la robustez y la capacidad predictiva de los modelos generados (Roy and Kar, 2014; Talevi et al., 2012b). Los métodos de validación pueden clasificarse en métodos internos y externos; en los primeros es el propio conjunto de entrenamiento el que se utiliza para validar el modelo generado; en la validación externa, en cambio, un conjunto de prueba conformado por compuestos que no forman parte del conjunto de entrenamiento se utiliza para predecir la propiedad o categoría de cada compuesto.

 Técnicas de Validación Interna:

Validación cruzada (cross-validation): Las estrategias por excelencia son las técnicas denominadas *Leave One Out* (LOO -“Dejar uno afuera”) y *Leave Group Out* (LGO -“Dejar algunos a fuera”), esta última también conocida como *Leave K Out*, *Leave Some Out*, *Leave Many Out*. Consisten en, una vez generado el modelo, retirar del conjunto de entrenamiento n compuestos (n igual a 1 corresponde a la técnica LOO), recalcular el modelo con los compuestos remanentes y utilizar éste último para predecir la clasificación de los compuestos removidos; el proceso se repite hasta que todos los compuestos del conjunto de entrenamiento son removidos al menos una vez. En nuestro caso (modelos clasificadores) hemos utilizado la técnica LGO expresando los resultados de la validación como la media aritmética del % de buenas clasificaciones de todos los nuevos modelos generados.

Test de Randomización o Aleatorización de Fischer (también conocido en inglés como *Scrambling model*): se realiza con el objetivo de minimizar la probabilidad de correlación fortuita entre los descriptores y la variable dependiente. Implica justamente aleatorizar los valores de la variable dependiente en el conjunto de entrenamiento, cancelando por tanto cualquier relación que pudiera existir entre estructura y propiedad, y generar nuevos modelos para evaluar si existe probabilidad de correlación azarosa entre las variables independientes y la variable dependiente. Idealmente se espera que los modelos generados por aleatorización sean estadísticamente inferiores al modelo elegido. Se evalúan los parámetros λ de Wilk, el estadístico F y la media del % de buenas clasificaciones.

 Validación Externa:

Consiste en predecir el valor de la variable dependiente para un conjunto de compuestos que no han sido utilizados para generar el modelo. La gran mayoría de los especialistas coinciden en reconocer a la validación externa como una etapa imprescindible (Golbraikh and Tropsha, 2002b; Roy and Kar, 2014; Tropsha et al., 2003). Sin embargo existen opiniones encontradas respecto a la conveniencia de resignar casos del conjunto de entrenamiento para incluirlos en

un conjunto de prueba cuando no se dispone de un conjunto de entrenamiento con gran cantidad de compuestos. Por ejemplo Hawkins (Hawkins, 2004) afirma que utilizar un conjunto de prueba externo para evaluar la predicción del modelo es adecuado cuando el mismo está formado por más de 20 de estructuras (y por lo menos 50, idealmente) y sólo si el dataset es realmente grande (varios cientos de compuestos). Otros autores sugieren utilizar un 10 o 20 % del total de compuestos de la base de datos (Yasri and Hartsough, 2001). En nuestro caso utilizamos un 25 % de la totalidad de la base de datos generada, y cumplimos con la condición de Hawkins respecto al número mínimo de compuestos en el conjunto de entrenamiento para utilizar un conjunto de prueba.

4.1.4.4. EVALUACIÓN DE LA PERFORMANCE DEL MODELO CLASIFICADOR

4.1.4.4.1. CURVAS ROC

Dada la aplicación que se le dará al modelo (definir, en una base de datos de miles de compuestos cuáles son los que presentan mayores probabilidades de poseer la actividad buscada) resulta relevante establecer un valor de corte en la respuesta del modelo (*score* FD) asociado a un adecuado balance entre *sensibilidad* (*Se*, es decir la tasa de verdaderos positivos, *VP*) y *especificidad* (*Sp*, es decir tasa de verdaderos negativos, *VN*).

$$Se = \frac{VP}{VP + FN}$$

$$Sp = \frac{VN}{VN + FP}$$

donde *FN* denota los falsos negativos (en nuestro caso, inhibidores que son predichos como no inhibidores por el modelo) y *FP* representa los falsos positivos (no inhibidores que son predichos como inhibidores). A tal fin resultan valiosas las *Curvas ROC* (*Receiver Operating Characteristic*) en las que, para distintos valores de corte de la respuesta o *score* del modelo, se grafica el balance de *Se* versus $1-Sp$.

El área bajo la curva (AUC, area under the curve) ROC constituye una medida valiosa para evaluar, además, si el modelo se comporta significativamente mejor que una

clasificación al azar o que otro modelo (Triballeau et al., 2005). Tal como puede observarse en la Figura 4.4 los ejes del gráfico de la curva ROC adoptan valores entre 0 y 1 (0% y 100%), delimitando un cuadrado de área = 1,00. Un modelo clasificador se considera no-discriminativo si su curva ROC no es estadísticamente distinta de la línea de no-discriminación (línea azul de la Figura 4.4) la cual está asociada a un AUC = 0,50 (la línea de no-discriminación divide en dos mitades iguales el cuadrado de área = 1,00). A medida que el AUC del modelo se acerca al valor 1,00 (clasificación perfecta), mayor será su capacidad discriminativa.

Como afirman Triballeau y colaboradores en el primer reporte del uso de estas curvas en el marco de una campaña de CV, la selección de un determinado balance entre Se y Sp no es una cuestión estadística sino que depende del contexto. En nuestro caso, debido al presupuesto limitado para adquirir candidatos, decidimos priorizar Sp sobre Se para reducir el número de falsos positivos (compuestos que ensayaríamos en un ensayo biológico pero no tendrían la actividad predicha sobre la enzima). Esto significa, no obstante, que para reducir el número de falsos positivos nos arriesgamos a perder andamios/motivos estructurales potencialmente valiosos (Triballeau et al., 2005). En la Figura 4. 4 se observa un esquema de construcción de la curva ROC.

Para evaluar la performance de los modelos elegidos se tuvieron en cuenta los comportamientos de las curvas ROC (y el parámetro AUC) tanto para el conjunto de entrenamiento como para el conjunto de prueba.

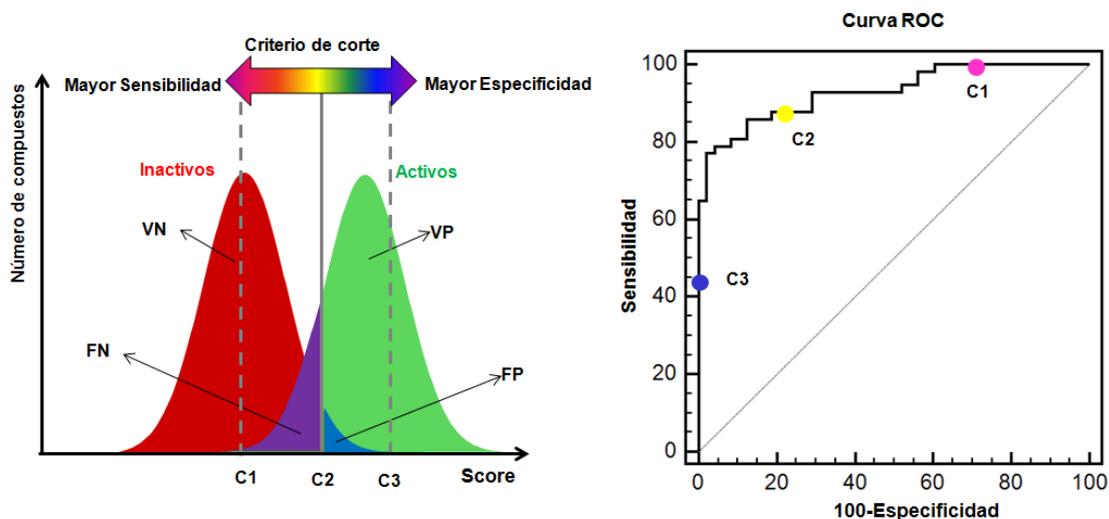


Figura 4.4. Esquema de construcción de una curva ROC.

4.2. CAMPAÑA SIMULADA DE CRIBADO VIRTUAL

Los experimentos conducidos por Truchon y Bayly sugieren que, cuando se usa para evaluar el modelo un conjunto de compuestos/base de datos demasiado pequeño, el valor de área bajo la curva ROC calculado presenta un error considerable que se reduce y converge a un valor constante para bases de datos de mayor tamaño (Truchon and Bayly, 2007). Por otro lado, debe tenerse en cuenta que en una aplicación real de CV nos encontraremos, en general, con un reducido número de activos inmersos en un número de inactivos muy grande. Por tal motivo, el conjunto de prueba resulta insuficiente para evaluar la utilidad del modelo en una aplicación real de CV y es conveniente estimar aquella de manera más realista, estudiando el comportamiento del modelo al ser aplicado a una base de datos en la que un número pequeño de positivos se encuentra inmerso en un alto número de negativos, o presuntos negativos. A tal fin, los 21 compuestos activos del conjunto de prueba se dispersaron entre los 21 inactivos y 444 presuntos inactivos que, junto a los demostrados inactivos del conjunto de prueba, actúan como “señuelos” (en inglés *decoys*). Los presuntos inactivos se obtuvieron realizando una búsqueda por similitud química en la base de datos Pubchem, eligiendo compuestos con una similitud química (calculada en base al coeficiente de Tanimoto) mayor a 0,95 con respecto a los inactivos del conjunto de prueba. La base de datos simulada obtenida de esta forma contiene entonces 21 inhibidores de Cz entre 465 no inhibidores conocidos o presuntos, de tal manera que los inhibidores representan menos del 5% de la base de datos total, lo cual constituye un test de desafío más estricto a la capacidad de enriquecimiento de activos del modelo

4.3. CRIBADO (O TAMIZADO) VIRTUAL

Una vez seleccionado el mejor modelo clasificador se procede al cribado de la base de datos elegida. Para eso se realiza el cálculo de los descriptores que forman parte del algoritmo clasificador para cada uno de los compuestos de la base de datos, y luego de incluir dichos valores en el modelo se obtiene la clasificación predicha de todos los compuestos. En este trabajo, se utilizaron 2 bases de datos que contienen fármacos y

compuestos químicos que son utilizados actualmente para diferentes usos: a) Merck Index 12 ed., presenta 5.570 compuestos químicos tipo fármaco; b) Drug Bank 3.0 (Knox et al., 2011) es una base de datos que compila fármacos aprobados por la FDA y candidatos atravesando ensayos clínicos en Estados Unidos , además de aditivos alimentarios aprobados por la FDA. Para las campañas en las cuales se utilizó esta base de datos se consideraron un total de 6.684 compuestos, excluyendo *a priori* aquellos de origen biotecnológico.

4.4. DOMINIO DE APLICABILIDAD

Obtener un modelo robusto y correctamente validado, puede ser insuficiente para predecir de forma confiable la propiedad para la totalidad del universo de compuestos químicos que conforman una base de datos sobre la cual se realiza el cribado. De hecho, sólo las predicciones (clasificaciones, en nuestro caso) de los nuevos compuestos que se encuentren dentro del dominio de aplicabilidad del modelo se pueden considerar fiables y no como extrapolaciones del modelo (P. Gramatica, 2007). El dominio de aplicabilidad se define por la naturaleza de los compuestos químicos que conforman el conjunto de entrenamiento. Existen varios enfoques para calcular el dominio de aplicabilidad de los modelos QSAR (Dimitrov et al., 2005; P. Gramatica, 2007), una de las formas más sencillas es empleando la *Regla de la Palanca* (en inglés, *leverage*): se determinan los valores de palanca (h_i) de cada compuesto 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).

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

Un compuesto seleccionado por CV de una base de datos, estará dentro del dominio de aplicabilidad del modelo cuando presente un valor de h inferior al umbral (h^*), definido generalmente como $h^*=3f/n$, donde f es el número de descriptores incluidos en el modelo y n el número de compuestos que conforman el conjunto de entrenamiento.

4.5. EVALUACIÓN BIOLÓGICA

La verificación experimental de los candidatos seleccionados por el modelo QSAR 2D, constituye la etapa definitiva para demostrar la validez del modelo. A continuación se detallan los ensayos *in vitro* e *in vivo* realizados en el trabajo de tesis.

4.5.1. EFECTO INHIBITORIO SOBRE CRUZIPAÍNA

Para estudiar el efecto de los candidatos seleccionados sobre Cz, se purificó parcialmente la enzima mediante precipitación con sulfato de amonio seguida de cromatografía de afinidad en columna sobre concavalina A-sefarosa (Sigma), de acuerdo a procedimientos descritos previamente en literatura (Labriola et al., 1993). La actividad de la Cz parcialmente purificada fue ensayada con Bz-Pro-Phe-Arg-pNA 250 μM (Sigma) como sustrato, incubando en buffer DTT (ditiotretiol) 6,5 μM y Tris-HCl pH 7 (Cazzulo et al., 1990), en presencia y ausencia de los compuestos ensayados. El producto de reacción fue medido espectrofotométricamente a temperatura ambiente y a 410 nm durante 5 minutos (espectrofotómetro Beckman Coulter TM DU530 Life Science UV-visible). Los valores obtenidos fueron convertidos en pmol de sustrato hidrolizado por minuto utilizando un coeficiente de extinción de 8.800 $\text{M}^{-1} \text{cm}^{-1}$ (p-nitroanilina). El efecto inhibitor del compuesto testeado fue expresado como porcentaje de actividad residual respecto al control sin inhibidor.

4.5.2. EFECTO INHIBITORIO SOBRE LA PROLIFERACIÓN DE EPIMASTIGOTES DE *T. CRUZI*

Epimastigotes de la cepa Y de *T. cruzi* fueron cultivados a 28°C en medio BHT con 20 mg/l de hemina, 20% de suero fetal bovino inactivado y antibióticos (100 $\mu\text{g}/\text{ml}$ de estreptomicina y 100 $\mu\text{g}/\text{ml}$ de penicilina) (Cazzulo et al., 1985), agregando la concentración del candidato a ensayar correspondiente (0-200 μM). Los cultivos fueron iniciados a una concentración de 10^7 células/ml y se realizó el seguimiento de la

proliferación diariamente por conteo de células en cámara hemocitométrica (cámara de Neubauer). A fin de observar el efecto del fármaco sobre la proliferación durante períodos prolongados, una vez que los cultivos alcanzan la fase estacionaria (cada 7 días) se repican mediante dilución a 10^7 células/ml adicionando medio fresco con el compuesto a ensayar.

4.5.3. ENSAYOS DE INVASIÓN DE TRIPOMASTIGOTES Y PROLIFERACIÓN A AMASTIGOTES DE T. CRUZI

La cepa Y-GFP de *T. cruzi* (provista por el Dr. Schenkman, S. Departamento de Microbiología, Inmunología y Parasitología de la Universidad Federal de Sao Paulo, Brasil) se mantiene en un cultivo de células Vero según técnicas descriptas previamente (Barclay et al., 2011); los mioblastos de rata (línea celular H9c2) se mantienen en el medio de cultivo DMEM (Gibco Laboratorio, Argentina) suplementado con 10% de SFB (Suero Fetal Bovino) y antibióticos a 37°C en una atmósfera de 95% de aire y 5% de CO₂.

Para realizar los experimentos, las células se siembran en placas de 24 pocillos. Los tripomastigotes (cepa Y-GFP) se tratan con los fármacos seleccionados durante los 30 minutos previos a la infección y luego se transfieren a monocapas de células H9c2 durante un período de 12 horas en DMEM suplementado con 3% de SFB y antibióticos (medio de infección) en ausencia y presencia de diferentes concentraciones de inhibidores de Cz. Los tripomastigotes no unidos se eliminan por lavado con PBS (solución de buffer fosfato salino).

Las monocapas infectadas se incuban por un período adicional de 48 horas en medio de infección suplementado con los compuestos de prueba para estudiar el efecto de los inhibidores de Cz en las primeras etapas de la infección.

Los ensayos de proliferación de amastigotes se llevan a cabo de una manera similar pero utilizando diferentes tiempos. Las células H9c2 son infectadas durante un período de 24 hs sin inhibidores y se tratan durante 96 hs con los inhibidores de Cz antes de la fijación. Las células se lavan 3 veces con PBS, se fijan con solución de paraformaldehído

al 3% en PBS durante 15 min a temperatura ambiente, y se inactivan con NH_4Cl 50 mM (reactivo de lisis).

Posteriormente las células son permeabilizadas con 1% de saponina en PBS con 1% de BSA (albúmina de suero bovino) y se incuban con faloidina-TRITC (tiocianato de tiorodamina) (Sigma) para teñir los límites celulares. Las células se montan con Mowiol (medio de montaje) que contiene un marcador de ADN Hoechst (Life Technologies) y se examinan por microscopía confocal (Barclay et al., 2011).

El porcentaje de células infectadas y el número de amastigotes por célula se cuantifican y analizan para cada condición con la prueba de comparaciones múltiples de Tukey-Kramer utilizando el software estadístico KyPlot.

4.5.4. ENSAYO ALAMAR BLUE

Los estudios de viabilidad celular se realizan mediante el ensayo Alamar Blue que permite cuantificar la proliferación celular evaluando la potencial toxicidad de los inhibidores de Cz en células huésped (Rassmusen, 1999). Se procede según las instrucciones del fabricante (DAL-1025, Biosource-Life Technologies): las células H9c2 cultivadas en placas de 96 pocillos se lavan tres veces con PBS y se incuban en medio de infección en ausencia o presencia de diferentes concentraciones de los inhibidores de la Cz durante 48 horas. Posteriormente, las células se lavan, se agrega el reactivo Alamar Blue al medio y se incuban durante 6 horas a 37 °C antes de realizar las mediciones. El nivel de crecimiento de las células se basa en la detección de la actividad metabólica celular la cual será proporcional a la absorbancia a 540 nm.

4.5.5. ENSAYO IN VIVO. MODELO DE INFECCIÓN AGUDA EN RATONES

4.5.5.1. CRITERIO EMPLEADO PARA PASAR AL MODELO DE INFECCIÓN AGUDA

Se emplearon tres criterios básicos para dar prioridad a las pruebas de los candidatos seleccionados en un modelo de infección aguda en ratones:

a) Si las concentraciones efectivas en epimastigotes de *T. cruzi* eran similares a las concentraciones plasmáticas alcanzadas en el estado estacionario durante un régimen de dosis múltiples para la indicación aprobada previamente del candidato reposicionado. Idealmente, debe ser empleada para la nueva indicación (en nuestro caso un efecto antichagásico) dosis iguales o inferiores a las utilizadas en clínica para la indicación terapéutica ya aprobada.

b) Si las indicaciones terapéuticas originales proveen un beneficio adicional para el paciente que padece la enfermedad de Chagas (por ejemplo un fármaco antichagásico que presenta un efecto adicional cardioprotector puede ayudar a controlar la infección y también los síntomas de la enfermedad).

c) La gravedad de los efectos adversos conocidos asociados a las dosis utilizadas para la indicación original del fármaco seleccionado. Lo ideal sería que los candidatos reposicionados presentes efectos adversos menos frecuentes y/o leves en comparación con los fármacos disponibles para el tratamiento de la enfermedad (nifurtimox y benznidazol). Por otra parte no serán viables como potenciales antichagásicos en la fase crónica de la enfermedad aquellos candidatos que presentes efectos adversos incompatibles con la patología (por ejemplo, fármacos contraindicados para pacientes cardíacos).

4.5.5.2. MODELO DE INFECCIÓN AGUDA EN ANIMALES

Para el modelo de infección aguda se utilizan ratones de la cepa outbred C57BL/6J de ambos sexos con un peso inicial de 7-9 gramos (Andrade and Andrews, 2005). Los animales se mantienen en condiciones libre de patógenos y alojados en habitaciones con

temperatura controlada (22-25°C) recibiendo agua y comida *ad libitum*. Todos los procedimientos se realizan en conformidad con las normas bioéticas del Comité Institucional para el Cuidado y Uso de Animales de Laboratorio (CICUAL) de la Facultad de Ciencias Médicas de la Universidad Nacional de Cuyo. Se utilizan grupos experimentales de 3 animales cada uno: siendo uno de ellos el control negativo, y los demás grupos tratados con los fármacos seleccionados para evaluar o benznidazol (control positivo) según corresponda.

Los animales son infectados por vía intraperitoneal con 10^5 células de la forma tripomastigote de *T. cruzi* (cepa Y-GFP). Luego de 5 días de la infección (5dpi) se determina la parasitemia (determinación de formas circulares de *T. cruzi*) y se inicia el tratamiento que continúa hasta el día de la muerte de los animales. El tratamiento consiste en administrar diariamente un volumen de 200 μ l del candidato seleccionado (en DMSO 1%) por vía intraperitoneal.

La parasitemia se determina contando el número de parásitos presentes en la sangre del ratón (por punción directa de la cola). A los 5 dpi se analiza y cuantifica el número de parásitos en una gota de sangre (5 μ l) mediante un examen microscópico en 50 campos (400X) (Andrade and Andrews, 2005). A los 10, 14 y 20 dpi se evalúa la parasitemia contando el número de parásitos en una cámara de Neubauer y los resultados se expresan como número de parásitos/ml. Las tasas de supervivencia se determinan mediante inspección diaria de las jaulas.

REFERENCIAS CAPÍTULO 4

Andrade L.O., Andrews N.W. The Trypanosoma cruzi-host-cell interplay: location, invasion, retention. *Nat Rev Microbiol.* **3**, 819-823. (2005).

Barclay J.J., Morosi L.G., Vanrell M.C., Trejo E.C., Romano P.S., Carrillo C. Trypanosoma cruzi Coexpressing Ornithine Decarboxylase and Green Fluorescence Proteins as a Tool to Study the Role of Polyamines in Chagas Disease Pathology. *Enzyme Res.* **2011**, 657460. (2011).

Cazzulo J.J., C. C.F.M., Martinez J. Some kinetic properties of a cysteine proteinase (Cruzipain) from Trypanosoma cruzi. *Biochim. Biophys. Acta.* **1037**, 186-191. (1990).

Cazzulo J.J., Franke de Cazzulo B.M., Engel J.C., Cannata J.J. End products and enzyme levels of aerobic glucose fermentation in trypanosomatids. *Mol Biochem Parasitol.* **16**, 329-343. (1985).

Dimitrov S., Dimitrova G., Pavlov T., Dimitrova N., Patelwicz G., Niemela J., Mekenyan O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model.* **45**, 839-849. (2005).

Downs G.M., Barnard J.M. Clustering Methods and Their Uses in Computational Chemistry, in: K. B. Lipkowitz and D. B. Boyd (Eds.), Reviews in Computational Chemistry, Wiley. pp. 1-40. (2003)

Everitt B.S., Landau S., Leese M., Stahl D. Optimization Clustering Techniques, in: Everitt (Ed.), Cluster Analysis, Wiley. pp. 111-142. (2011)

Golbraikh A., Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J Comput Aided Mol Des.* **16**, 357-369. (2002a).

Golbraikh A., Tropsha A. Beware of q²! *J Mol Graph Model.* **20**, 269-276. (2002b).

Golbraikh A., Shen M., Xiao Z., Xiao Y.D., Lee K.H., Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des.* **17**, 241-253. (2003).

Hariharan R., Janakiraman A., Nilakantan R., Singh B., Varghese S., Landrum G., Schuffenhauer A. MultiMCS: a fast algorithm for the maximum common substructure problem on multiple molecules. *J Chem Inf Model.* **51**, 788-806. (2011).

Hawkins D.M. The problem of overfitting. *J Chem Inf Comput Sci.* **44**, 1-12. (2004).

Jaworska J., Nikolova-Jeliazkova N., Aldenberg T. QSAR applicability domain estimation by projection of the training set descriptor space: a review. *Altern Lab Anim.* **33**, 445-459. (2005).

Kawabata T. Build-up algorithm for atomic correspondence between chemical structures. *J Chem Inf Model.* **51**, 1775-1787. (2011).

Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A.C., Wishart D.S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035-1041. (2011).

Labriola C., Sousa M., Cazzulo J.J. Purification of the major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* by affinity chromatography. *Biol Res.* **26**, 101-107. (1993).

P. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **26**, 694-701. (2007).

Peduzzi P., Concato J., Feinstein A.R., Holford T.R. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol.* **48**, 1503-1510. (1995).

Peduzzi P., Concato J., Kemper E., Holford T.R., Feinstein A.R. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* **49**, 1373-1379. (1996).

Perez J.J. Managing molecular diversity. *Chem Soc Rev.* **34**, 143-152. (2005).

Rassmusen E.S. Use of fluorescent redox indicator to evaluate cell proliferation and viability. *In vitro Mol Toxicol.* **12**, 47-58. (1999).

Rawlings J.O., Pantula S.G., Dickey D.A. Regression Diagnostics, in: J. O. Rawlings, et al. (Eds.), *Applied Regression Analysis: A Research Tool*, Editorial Springer. pp. 369-377 (1998)

Roy K., Kar S. How to judge predictive quality of classification and regression based QSAR models, Betham science (En prensa). (2014)

Talevi A., Castro E.A., Bruno-Blanch L.E. Recent studies on similarity measures and its applications to chemoinformatics and drug design. En: *Recent Trends on QSAR in the Pharmaceutical Perceptions*, M. T. Hassan Khan (Ed.), Editorial Betham Science. pp. 272-297 (2012a)

Talevi A., Castro E.A., Bruno-Blanch L.E. Towards Accurate Virtual Screening Technologies: Sampling of Chemical Space and Applicability Domain Estimation. En: *Chemical Information and Computational Challenges*, M. V. Putz (Ed.). Nova Science Publishers. pp. 277-297, (2012b)

Talevi A., Bellera C.L., Di Ianni M., Duchowicz P.R., Bruno-Blanch L.E., Castro E.A. An integrated drug development approach applying topological descriptors. *Curr Comput Aided Drug Des.* **8**, 172-181. (2012c).

Triballeau N., Acher F., Brabet I., Pin J.P., Bertrand H.O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem.* **48**, 2534-2547. (2005).

Tropsha A., Gramatica P., Gombar V.K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science.* **22**, 69-77. (2003).

Truchon J.F., Bayly C.I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J Chem Inf Model.* **47**, 488-508. (2007).

Van Waterbeemd H. Discriminant Analysis for Activity Prediction. , in: V. W. H (Ed.), *Chemometric methods in molecular design* Wiley. pp. 265-282. (1995)

Willett P. Chemoinformatics - similarity and diversity in chemical libraries. *Curr Opin Biotechnol.* **11**, 85-88. (2000).

Yasri A., Hartsough D. Toward an optimal procedure for variable selection and QSAR model building. *J Chem Inf Comput Sci.* **41**, 1218-1227. (2001).

Yee L.C., Wei Y.C. Current Modeling Methods Used in QSAR/QSPR, in: M. Dehmer, et al. (Eds.), *Statistical Modelling of Molecular Descriptors in QSAR/QSPR.* (2012)

CAPÍTULO 5

RESULTADOS



5.1. COMPILACIÓN DE LA BASE DE DATOS Y PARTICIÓN EN CONJUNTO DE ENTRENAMIENTO Y CONJUNTO DE PRUEBA

Se compiló a partir de literatura una base de datos balanceada formada por 163 compuestos, 82 de ellos pertenecientes a la categoría *ACTIVOS* y los 81 restantes correspondientes a la categoría *INACTIVOS* (Ang et al., 2011; Beaulieu et al., 2010; Brak et al., 2008; Choe et al., 2005; dos Santos Filho et al., 2009; Du et al., 2000; Ferreira et al., 2010; Hernandez et al., 2010; Huang et al., 2003; Mott et al., 2010; Rodrigues et al., 2002; Scheidt et al., 1998). Las estructuras químicas de los compuestos de ambas categorías se presentan en el *Anexo 2*.

El procedimiento de *clustering* jerárquico reveló 5 grupos de por lo menos 6 compuestos en la categoría activa y 7 grupos de por lo menos 7 compuestos en la categoría inactiva. De acuerdo al *clustering* MCS, existen 4 compuestos en la categoría activa que pueden ser considerados atípicos ya que no presentan una mínima subestructura común de 9 átomos con ninguna de las estructuras restantes de esa categoría o son agrupados a lo sumo de a pares (es decir, presentan una mínima subestructura común de nueve átomos con sólo un compuesto de su misma categoría) (Figura 5.1). La categoría inactiva presentó 14 compuestos atípicos.

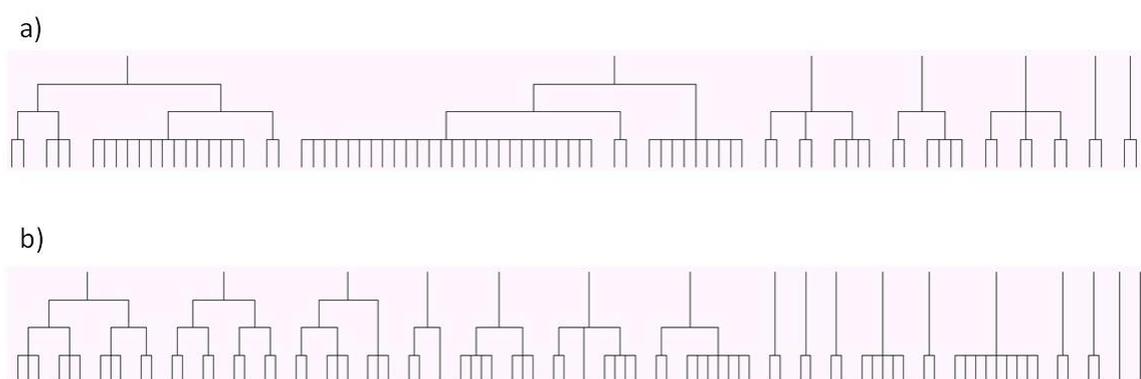


Figura 5.1. Dendrogramas obtenidos por LibraryMCS correspondientes a los compuestos Activos (a) y a los compuestos inactivos (b).

A partir de los resultados obtenidos del *clustering* jerárquico se seleccionó al azar un compuesto correspondiente a cada *cluster* de cada categoría, a fin de utilizar dichos compuestos como “semillas” en la técnica de *clustering k-means*. Una vez optimizado el *clustering* jerárquico por medio del algoritmo *k-means*, se sustrajo el 25% de cada grupo reservándolo para el conjunto de prueba, mientras que el 75% remanente se utilizó para calibrar el modelo. De esta manera el conjunto de entrenamiento utilizado para la generación y validación de los modelos empleados en las campañas de CV, quedó conformado por 121 compuestos (61 activos y 60 inactivos) y el conjunto de prueba por 42 compuestos (21 activos y 21 inactivos).

A continuación se presentan los resultados obtenidos con los 3 modelos generados separados por Campañas de CV realizadas. Los descriptores incluidos en los modelos se describen con detalle en el *Anexo 1*.

5.2. PRIMERA CAMPAÑA DE CRIBADO VIRTUAL

5.2.1. GENERACIÓN Y VALIDACIÓN DEL MODELO 1

Se calcularon un total de 877 descriptores moleculares (0D-2D) con el programa Dragon 4.0 en su versión académica (Milano Chemometrics 2006). A partir de 30 subconjuntos aleatorios de no más de 254 descriptores cada uno, se procedió a generar modelos mediante ALD, de todos los modelos generados el siguiente fue el que presentó mejor desempeño considerando todos los criterios de selección descritos en el capítulo anterior:

$$FD = -0,959 + 1,764 * \mathbf{GGI7} - 0,287 * \mathbf{nS} + 0,280 * \mathbf{nCN}$$

$$N=121 \quad \lambda = 0,56 \quad F(3,117) = 31,24 \quad p < 0,00001$$

Donde **GGI7** representa el índice de carga topológica de Gálvez de orden 7, **nS** el número de átomos de azufre y **nCN** el número de nitrilos alifáticos.

Las magnitudes de los coeficientes de regresión normalizados del modelo son, respectivamente: 0,583; 0,183; 0,136; esto demuestra que, **GGI7** es la variable más relevante del modelo (es decir con mayor poder clasificatorio).

El modelo presenta una excelente relación de casos por cada descriptor (más de 40) que indica una muy baja probabilidad de sobreajuste, confirmada posteriormente en los resultados de la validación externa. Además, la correlación entre los descriptores incluidos en el modelo es insignificante, siendo 0,147 el coeficiente de correlación entre los dos descriptores más correlacionados.

Utilizando 0 como punto de corte, el porcentaje de buenas clasificaciones fue el siguiente: para el conjunto de entrenamiento, 74 % de buenas clasificaciones para compuestos inactivos, 80 % para activos y un 77% de buenas clasificaciones generales. En relación al conjunto de prueba 81% de buenas clasificaciones de activos, 90,5 % inactivos, con un 86% de buenas clasificaciones generales. Estos resultados parecen confirmar que no ha ocurrido sobreajuste, ya que el rendimiento en el conjunto de prueba es similar (y de hecho, mejor) que en el conjunto de entrenamiento.

El rendimiento promedio de los modelos aleatorizados fue de 63,6% (DE = 3,5), demostrándose que como era de esperar los modelos aleatorios son estadísticamente inferiores al modelo real.

Los resultados de la validación cruzada (LGO) resultaron en un porcentaje promedio de buenas clasificaciones de 77,6% (DE=17) (promedio de las 24 particiones, con 5 compuestos removidos en cada una de ellas), que es muy similar al funcionamiento del modelo original en el conjunto de entrenamiento.

Siguiendo las sugerencias de Truchon y Bayly (Truchon and Bayly, 2007); estudiamos el comportamiento del modelo al ser aplicado a una base de datos en la que un número pequeño de positivos se encuentra inmerso en un alto número de inactivos, o presuntos inactivos.

La Figura 5.2 muestra las curvas ROC construidas utilizando el programa MedCalc (v12.4.0) para los conjuntos de entrenamiento, conjunto de prueba y para la base de datos simulada. Como se describió oportunamente, tales gráficas se construyen para

seleccionar racionalmente el valor de corte del score del modelo que se utilizará en el CV. Sobre la base del análisis de las curvas ROC fue definido un punto de corte de 0,29 para diferenciar compuestos activos de inactivos en la campaña de CV. De acuerdo a la información de la curva ROC este valor corresponde a una Se de 67% y una Sp de 95% en el conjunto de entrenamiento, y para el conjunto de prueba 75% de Se y 100 % de Sp. El AUC para ambos conjuntos fue de 0,893 y 0,921 respectivamente y de 0,846 para la base de datos simulada.

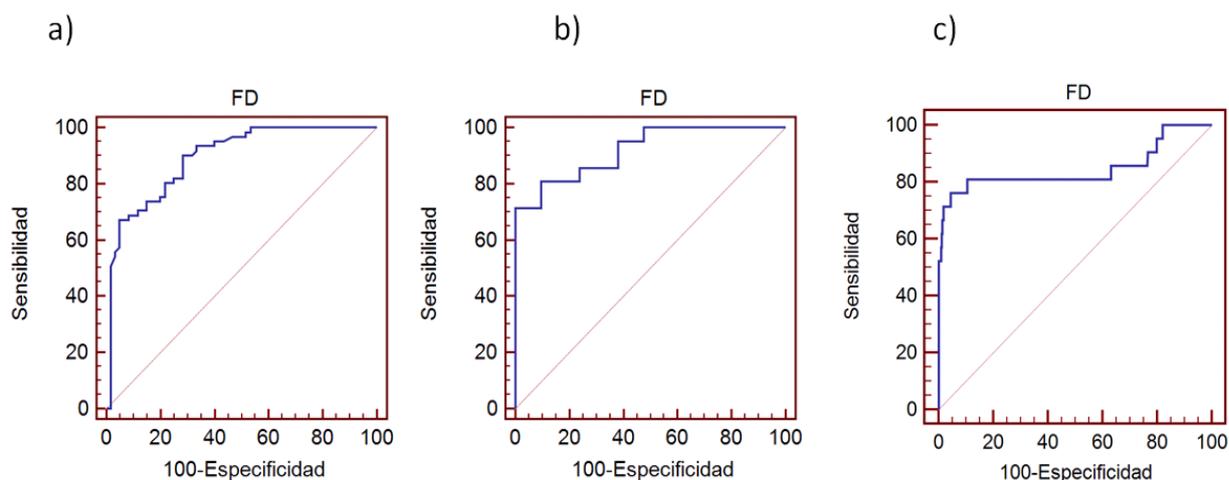
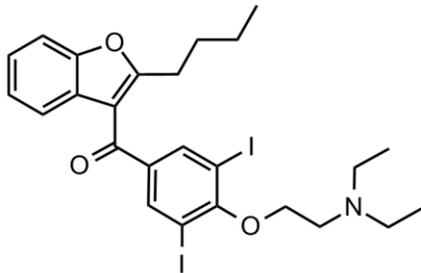


Figura 5.2. Curva ROC correspondiente al conjunto de entrenamiento (a), conjunto de prueba (b) y CV simulado (c) correspondientes al modelo aplicado en la primera campaña de CV.

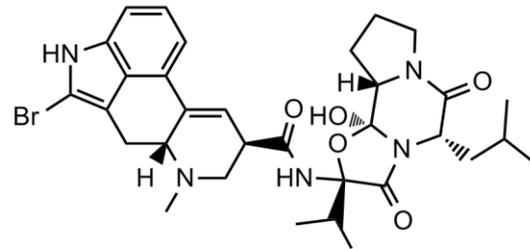
5.2.2. CRIBADO VIRTUAL

Se tamizó la base de datos DrugBank 3.0 (Knox et al., 2011); sólo fueron considerados fármacos aprobados, pequeñas moléculas experimentales y nutracéuticos (*a priori* fueron excluidos los compuestos biotecnológicos). De 6684 compuestos, 256 fueron seleccionados por el modelo clasificador encontrándose dentro del dominio de aplicabilidad del modelo determinado por la regla de la palanca (Tropsha et al., 2003); 54 de estos son fármacos aprobados, que constituyen los candidatos más inmediatos para el potencial reposicionamiento. En base a la accesibilidad fueron adquiridos 4 fármacos para ser evaluados experimentalmente: clorhidrato de amiodarona (antiarrítmico), oxalato de escitalopram (antidepresivo), colchicina (fármaco utilizado

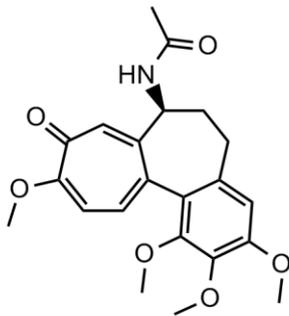
para el tratamiento de la gota) y bromocriptina (aprobado como antiparkinsoniano y antidiabético). Los dos primeros fueron donados por los laboratorios Vannier y Bagó respectivamente; los dos últimos fueron adquiridos de Droguería Saporitti. En la Figura 5.3 se presentan las estructuras correspondientes a cada uno.



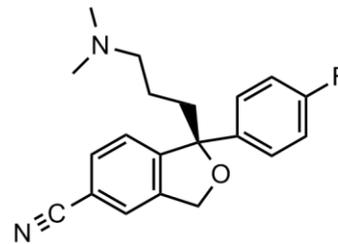
Amiodarona



Bromocriptina



Colchicina



Escitalopram

Figura 5.3. Estructuras químicas de fármacos seleccionados y adquiridos para ser evaluados experimentalmente.

5.2.3. EVALUACIÓN BIOLÓGICA

Los 4 candidatos seleccionados por el CV fueron evaluados experimentalmente en el ensayo enzimático de extracto crudo de Cz. La Figura 5.4 muestra el efecto de los 4 candidatos (soluciones de 100 μ M) sobre la actividad de extracto crudo de Cz. Puede observarse que amiodarona y bromocriptina mostraron un efecto inhibitorio significativo sobre la actividad de Cz en extracto crudo de células.

A continuación se realizó un ensayo en extracto purificado de Cz con estos dos candidatos a fin de verificar la existencia de una inhibición dependiente de la

concentración. En ambos casos se observó que la inhibición es concentración-dependiente. La concentración inhibitoria media (IC₅₀) para amiodarona fue de 219,8 μ M y para bromocriptina 84,2 μ M (resultados en Figura 5.5).

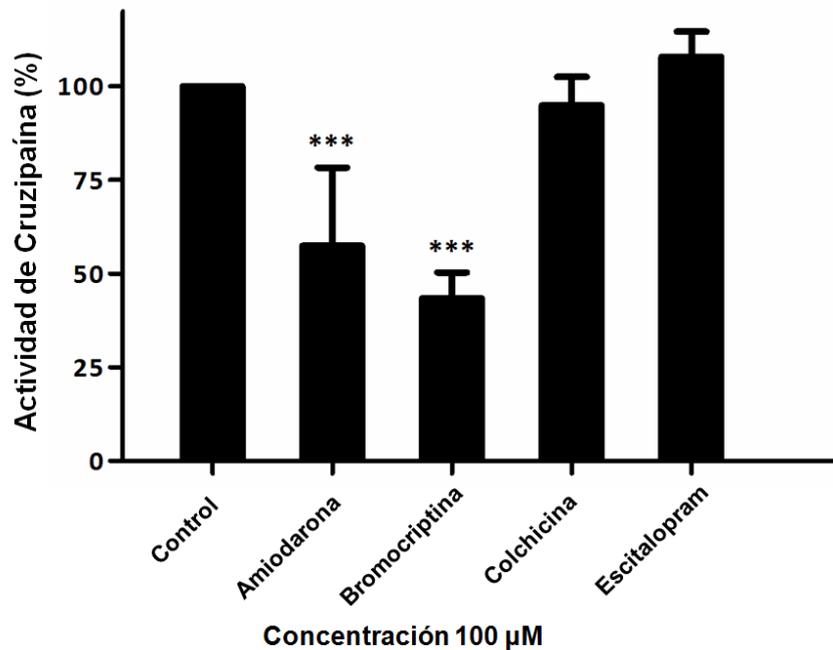


Figura 5.4. Efecto inhibitorio de los cuatro candidatos seleccionados sobre la actividad de CZ en extracto crudo de células de *T. cruzi*. La actividad de la enzima se expresa como porcentaje de la condición control (2% DMSO). Los resultados representan la media de tres experimentos independientes. Las barras indican DE y los asteriscos indican diferencias significativas respecto del control (**p<0,005).

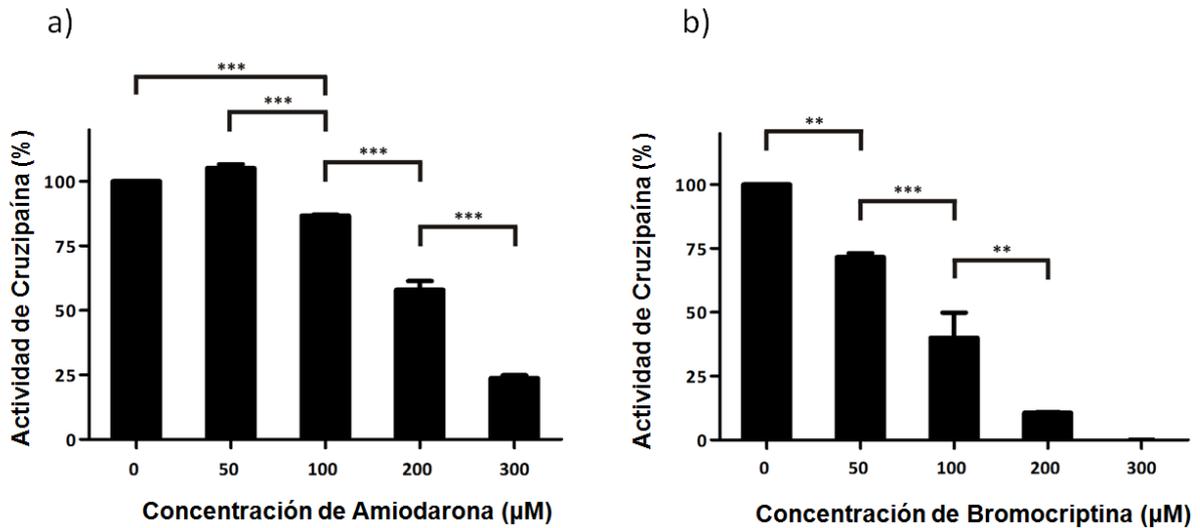


Figura 5.5. Efecto inhibitorio dependiente de la concentración de amiodarona (a) y bromocriptina (b) sobre la actividad de Cz purificada. Ambos candidatos se ensayaron en un intervalo de concentraciones de 0 a 300 μM . La actividad residual de Cz se expresó como porcentaje del control (0 μM , 2% DMSO). Los resultados representan la media de dos experimentos independientes. Los asteriscos indican diferencias significativas (** $p < 0,01$; *** $p < 0,005$).

Ambos candidatos también mostraron un efecto notorio sobre la proliferación de epimastigotes de *T. cruzi* (Figura 5.6) con valores de concentración inhibitoria media de 22 μM para amiodarona y de 15 μM para bromocriptina en la fase logarítmica media de los controles (cuarto día). El hecho de que ambos principios activos presentan mayor efecto sobre el cultivo de epimastigotes que en la actividad de Cz podría explicarse por un efecto de los compuestos sobre más de un blanco molecular del parásito. Por ejemplo, ya ha sido reportada la participación de amiodarona en la homeostasis de Ca^{+2} y en la biosíntesis de ergosterol en *T. cruzi* (Benaim et al., 2006). Sin embargo es importante destacar que esta es la primera vez que se observa que el efecto de amiodarona sobre tripanosomas, podría estar también relacionado con la inhibición de Cz.

En la Figura 5.7 se observa el cambio en la morfología del parásito frente a Bromocriptina.

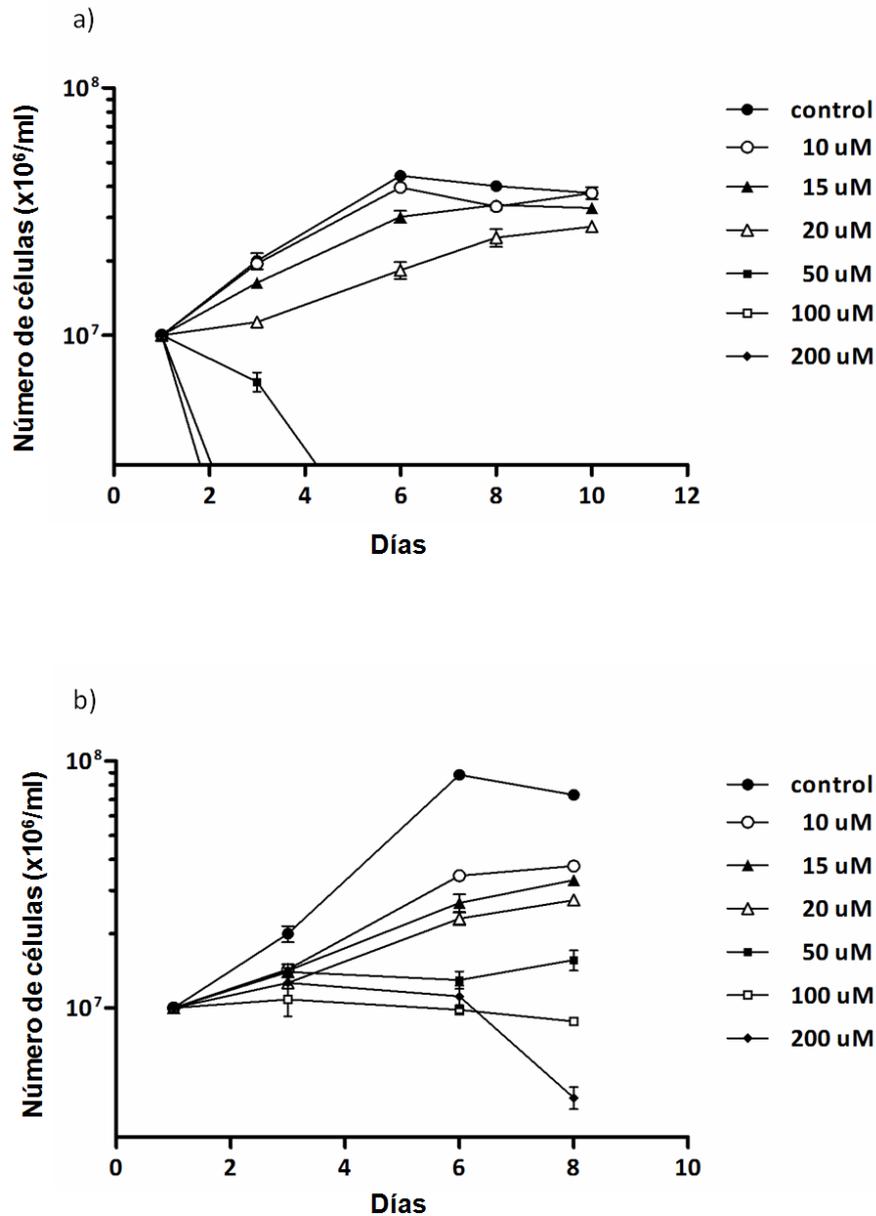
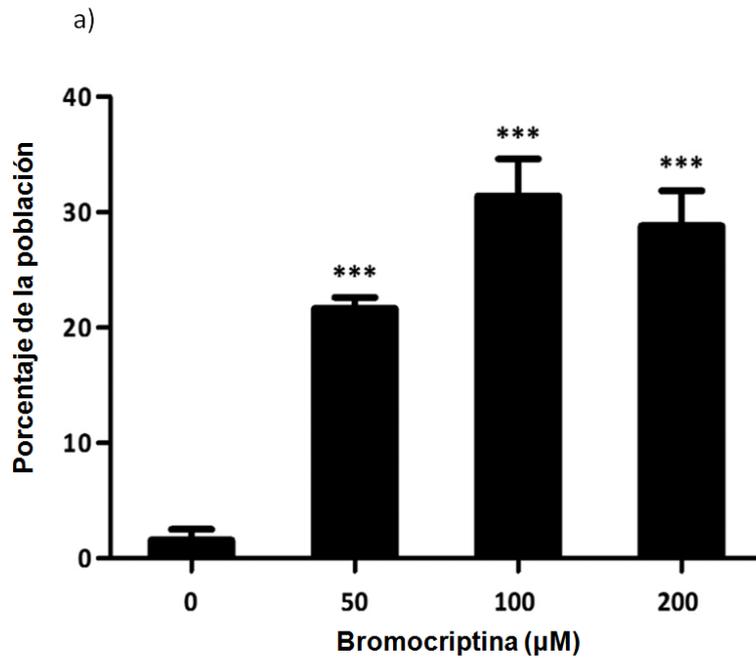


Figura 5.6. Efectos sobre la proliferación de epimastigotes de (a) amiodarona y (b) bromocriptina. Para determinar la tasa de crecimiento, se sembraron 10^7 células/ml en medio de BHT y se mantuvo a 28°C durante 10 u 8 días, respectivamente. La condición de control se realizó con 2% DMSO. Los parásitos se contaron utilizando una cámara hemocitométrica (cámara de Neubauer). Los resultados representan la media \pm DE de un experimento representativo.



b)

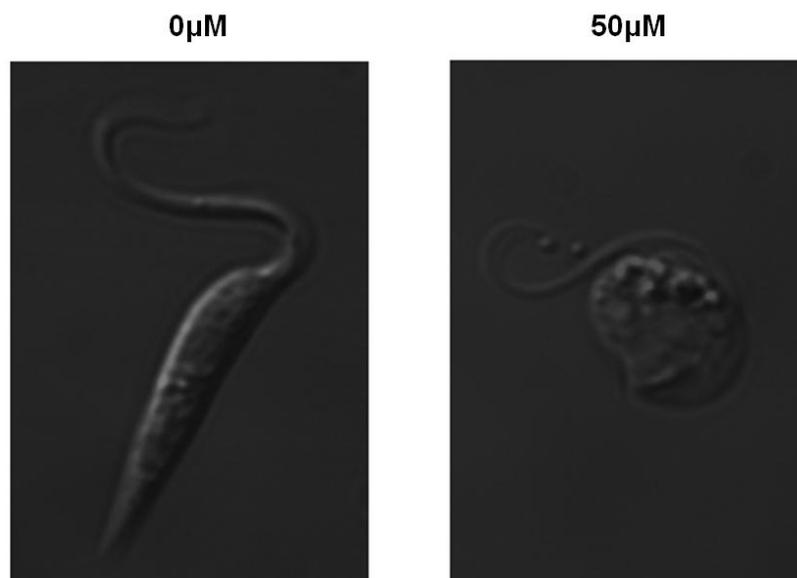


Figura 5.7. Efecto de bromocriptina sobre la morfología de *T. cruzi*. (A) Cuantificación de cambios morfológicos en las células que han sido cultivadas en ausencia o presencia de bromocriptina (50-200 μM) durante cuatro días. Al menos 100 células fueron contadas para cada condición. Los resultados representan las medias de dos experimentos independientes. (***) $P < 0,005$, respecto a $0\mu\text{M}$). (B) Microscopía de contraste diferencial de interferencia (DIC), imágenes de las células incubadas durante cuatro días en ausencia (morfología normal) o presencia (morfología redondeada) de $50\mu\text{M}$ bromocriptina. Los resultados representan las imágenes de un experimento representativo.

5.3. SEGUNDA CAMPAÑA DE CRIBADO VIRTUAL

5.3.1. GENERACIÓN Y VALIDACIÓN DEL MODELO 2

Para el desarrollo del segundo modelo clasificador, calculamos –para cada compuesto de los conjuntos de entrenamiento y prueba- 3.764 descriptores independientes de la conformación mediante el uso del programa Dragon 6.0 (Milano Chemometrics, 2011). En esta oportunidad se particionó aleatoriamente el pool total de 3.764 descriptores en 25 grupos de no más de 254 descriptores cada uno; de cada subconjunto de descriptores se obtuvo un modelo mediante ALD, resultando seleccionado por su buena performance el siguiente:

$$FD = 0,819 + 0,536 * VE1_X - 1,187 * C-018 + 0,495 * F-084 - 0,688 * NsssN - 0,187 * H-051 - 0,427 * O-056 - 0,09 * Sds - 5,311 * SpPosA_A$$

$$N=121 \quad \lambda = 0,445 \quad F(8,112) = 17,476 \quad p < 0,00001$$

donde **VE1_X** representa la suma de los coeficientes del autovector asociado con el último autovalor (el más negativo) de la matriz chi (en el Anexo 1 se detalla la explicación del descriptor); **C-018** corresponde al número de elementos de la forma =CHX (= representa un doble enlace, C un átomo de carbono y X un heteroátomo); **F-084** representa átomos de flúor unidos a carbonos con hibridación Sp²; **NsssN** denota el número de nitrógenos terciarios; **H-051** es el número de átomos de hidrógeno unidos a un carbono alfa; **O-056** se refiere al número de OHs; **Sds** simboliza la suma de los E-estados del tipo de átomo =S (proveyendo información de la frecuencia de tal tipo de átomo y la accesibilidad de sus electrones) y; **SpPosA_A** es la suma espectral positiva y normalizada de la matriz adyacencia.

Las magnitudes de los coeficientes de regresión normalizados del modelo son, respectivamente: 0,564; 0,185; 0,168; 0,160; 0,137; 0,135; 0,142 y 0,108, mostrando que **VE1_X** es la variable independiente más relevante (con mayor poder clasificador) del modelo. Debe destacarse que el modelo presenta una muy buena relación de casos por descriptor (alrededor de 15) lo que señala una baja probabilidad de sobreajuste, como lo

han demostrado luego los resultados de la validación externa. Considerando un score de 0 como valor de corte para separar activos e inactivos, el modelo presenta un 87% de buenas clasificaciones sobre los compuestos inactivos del conjunto de entrenamiento, 90% de buenas clasificaciones entre los compuestos activos del conjunto de entrenamiento, y un 88% de buenas clasificaciones globales. Con respecto al conjunto de prueba, estos porcentajes, en ese mismo orden, son 81; 90,5 y 86%. En otras palabras, el desempeño del modelo al clasificar a los compuestos del conjunto de entrenamiento es muy similar a su desempeño al clasificar a los compuestos del conjunto de prueba.

El desempeño promedio de los modelos obtenidos por aleatorización fue 68,4 (DE = 4,1) demostrándose que los modelos aleatorizados se comportan significativamente peor que el modelo real, indicando una muy baja probabilidad de correlación azarosa entre las variables independientes y dependiente. La validación cruzada resultó en un porcentaje promedio de buenas clasificaciones de 79% (DE=16) (resultado promedio de las 20 particiones con 6 compuestos removidos en cada una). Es notable que las particiones peor clasificadas en la validación cruzada fueron aquellas que presentaban compuestos atípicos según se definieran en la sección 5.1.

La Figura 5.8 muestra las curvas ROC para los conjuntos de entrenamiento, de prueba y la base de datos simulada. Las AUC asociadas a las curvas ROC para los conjuntos de entrenamiento y prueba fueron de 0,930 y 0,923, en ese orden. Debe destacarse que el área bajo la curva ROC para la base de datos simulada fue de 0,953 (mayor a las obtenidas para los conjuntos de entrenamiento y de prueba) mientras que 16 de los 21 inhibidores de Cz incluidos en tal base fueron rankeados entre el 5% de los compuestos mejor rankeados. En este caso se eligió como valor de corte el score 0,06, que corresponde a una Se de 87% y una Sp de 88% en el conjunto de entrenamiento, y una Se de 81% y una Sp de 95% en el conjunto de prueba.

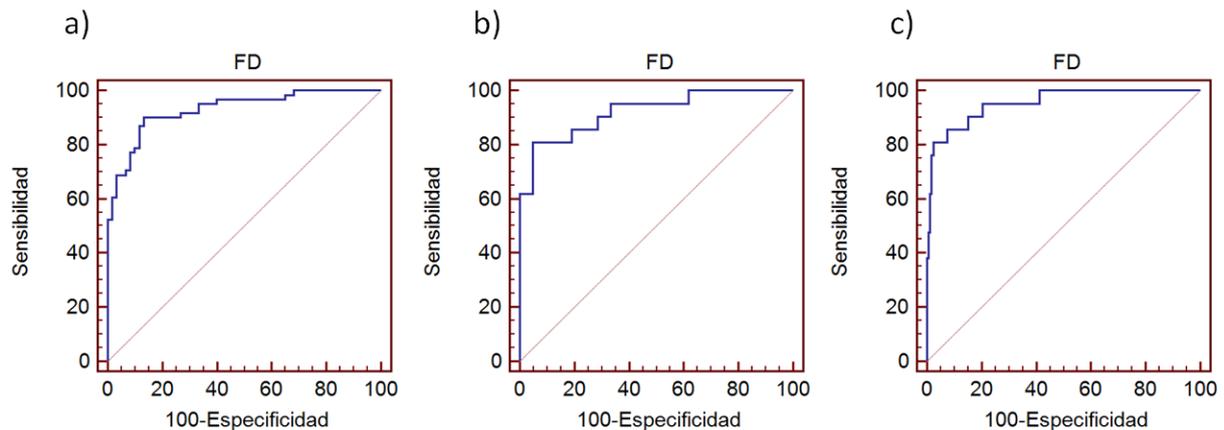


Figura 5.8. Curvas ROC correspondientes al conjunto de entrenamiento (a), conjunto de prueba (b) y CV simulado (c), correspondientes al modelo aplicado en la segunda campaña de CV.

5.3.2. CRIBADO VIRTUAL

De las 6684 estructuras de DrugBank 3.0 fueron seleccionadas 64 por el modelo como inhibidores de Cz, perteneciendo estas al dominio de aplicabilidad. 54 de ellas corresponden a fármacos de uso clínico, los candidatos más directos para reposicionamiento de fármacos. En base a su accesibilidad, tres de ellos (Figura 5.9) fueron adquiridos y evaluados sobre Cz. Los mismos fueron cisaprida (agente procinético), paroxetina (antidepresivo) y levotiroxina (reemplazo hormonal en pacientes con hipotiroidismo). Entre ellos, sólo paroxetina había sido anteriormente evaluada sobre Cz, con resultados negativos, mediante una campaña de cribado de alta performance en la que se evaluó el efecto de 197000 compuestos sobre la hidrólisis del sustrato Z-Phe-ARg-AMC por Cz (Jadhav et al., 2010). Los 3 fármacos fueron adquiridos en Drogueria Saporitti.

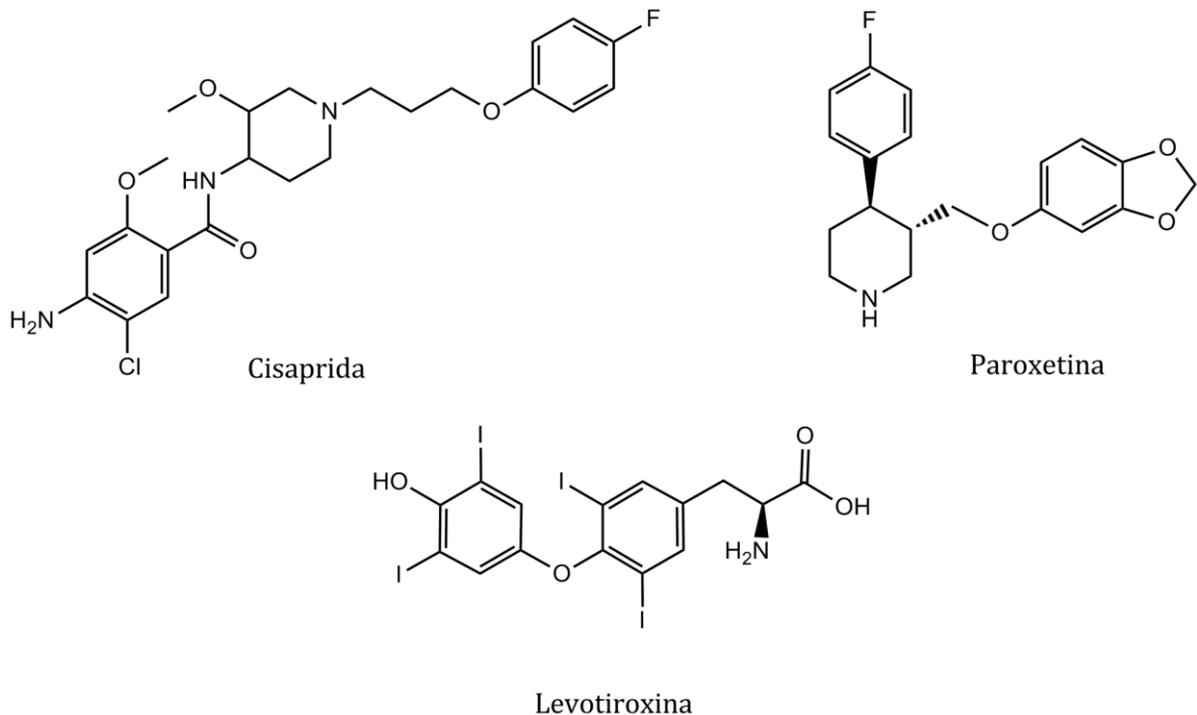


Figura 5.9. Estructuras de los compuestos adquiridos luego de la segunda campaña de CV.

5.3.3. EVALUACIÓN BIOLÓGICA

Usando Bz-Pro-Phe-Arg-pNa como sustrato cromogénico, la levotiroxina demostró un efecto inhibitorio significativo sobre la Cz en extracto puro (no así en el caso de cisaprida y paroxetina) (Figura 5.10). El efecto de levotiroxina sobre extracto purificado de Cz demostró ser dependiente de la concentración, con una IC_{50} de $38,4 \pm 6,8 \mu M$ (Figura 5.11).

La proliferación de los epimastigotes de *T. cruzi* se vio afectada por levotiroxina progresivamente en el tiempo y de forma concentración-dependiente (Figura 5.12). El efecto fue claramente notorio durante la tercera semana del ensayo mostrando una IC_{50} de $121,8 \pm 11,4 \mu M$ en la fase exponencial de crecimiento (día 17).

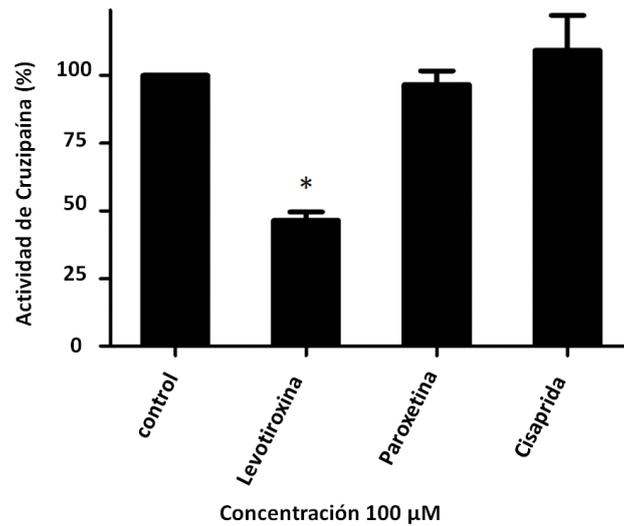


Figura 5.10. Efecto inhibitorio de los tres candidatos seleccionados sobre la actividad de Cz en extracto crudo de células de *T. cruzi*. La actividad de la enzima se expresa como porcentaje de la condición control (2% DMSO). La concentración final de cada compuesto fue de 100 μM. Los resultados representan la media de tres experimentos independientes. Las barras indican DE y los asteriscos indican diferencias significativas respecto del control (* $p < 0,05$).

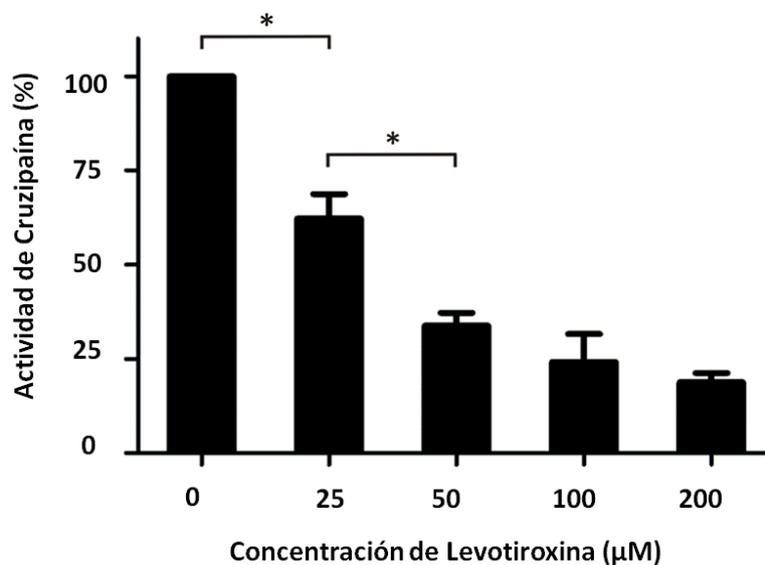


Figura 5.11. Efecto inhibitorio dependiente de la dosis de levotiroxina sobre la actividad de Cz purificada. La levotiroxina se ensayó en un intervalo de concentración de 0-200 μM. La actividad remanente de Cz se expresó como un porcentaje respecto al control (0 μM levotiroxina, 2% de DMSO). Los resultados representan la media de tres experimentos independientes. Los asteriscos indican diferencias significativas (* $p < 0,05$).

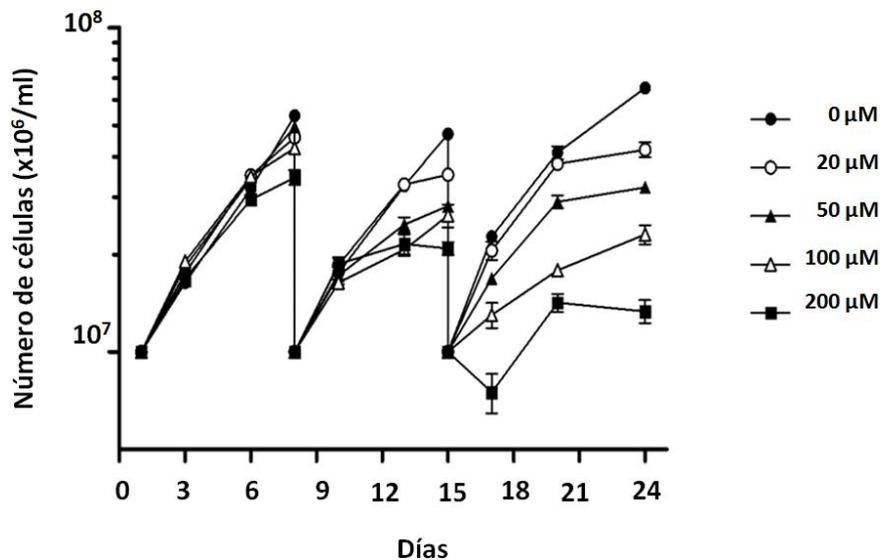


Figura 5.12. Efecto de levotiroxina sobre la proliferación de epimastigotes de *T. cruzi*. Para determinar la tasa de crecimiento, 10^7 células/ml se sembraron en medio de BHT y se mantuvieron a 28°C durante una semana. La condición de control fue realizada con $0\ \mu\text{M}$ levotiroxina, 2% de DMSO. Los parásitos se contaron usando una cámara hemocitométrica. Los resultados representan la media \pm DE de un experimento representativo.

5.4. TERCERA CAMPAÑA DE CRIBADO VIRTUAL

5.4.1. GENERACIÓN Y VALIDACIÓN DE MODELO 3

Para el desarrollo del tercer modelo utilizamos un set de 62 descriptores topológicos ampliamente conocidos como, índices de conectividad simples o de grafo de Randić-Kier-Hall (mX_t , ${}^mX_{tv}$) (Kier and Hall, 1983; Kier et al., 1976), índices de carga hasta de quinto orden (J_m , G_m , J_{mv} , G_{mv}) (Galvez et al., 1994), cocientes y diferencias en índices de conectividad de valencia o no valencia (mC_t y mD_t), número de pares de ramificaciones a distancia topológica n con n de 0 a 4 (PR_n) y número de vértices con valencia n siendo n 3 o 4 (V_n). Caracterizamos a cada compuesto con el set de 62 descriptores mediante el programa DESMOL11 (Unidad de Topología Molecular y Diseño de Fármacos, Universidad de Valencia, España).

El siguiente modelo capaz de clasificar inhibidores de no inhibidores de Cz se obtuvo por ALD:

$$FD = -4,762 + 1,325 * G4V - 1,248 * X4V + 0,421 * G1 + 0,171 * VV3$$

N=121

$\lambda=0,428$

F(4,116)=33,44

p<0,00001

donde **G4V** representa el índice de carga topológica de valencia de orden 4; **X4V** es el índice de conectividad tipo *path* de orden 4; **G1** denota al índice de carga topológica de orden 1y **VV3** al número de vértices con valencia 3.

El modelo presenta una excelente relación observaciones/descriptor (alrededor de 25), que indica una baja probabilidad de sobreajuste, confirmada posteriormente en los resultados de la validación externa. Cuando se utiliza 0 como punto de corte para diferenciar compuestos activos de inactivos, el modelo presenta un 86% de buenas clasificaciones para la categoría INACTIVA del conjunto de entrenamiento, el 91,7% de las buenas clasificaciones para la categoría ACTIVA, y un total de 88,5% de buenas clasificaciones generales.

En cuanto al conjunto de prueba independiente, el modelo clasifica correctamente el 81% de la población activa y el 85,7% de los compuestos inactivos, con una buena clasificación general del 83%.

El rendimiento promedio de los modelos obtenidos por aleatorización fue 65,7% (DE = 4,6) que muestra que los modelos aleatorizados fueron superados significativamente por el modelo real, como se esperaba. El resultado de la validación cruzada se tradujo en un porcentaje medio de las buenas clasificaciones de 85,8% (DE=12) (resultado promedio de 12 particiones removidas, de 5 compuestos cada una), que es muy similar a los resultados de la validación externa.

Recurrimos a las curvas ROC (Figura 5.13) para seleccionar racionalmente el valor de corte del score del modelo que se utilizará en el CV. El AUC para el conjunto de entrenamiento y conjunto de prueba fueron, 0,931 y 0,862 respectivamente. En base a los resultados obtenidos definimos 0,5 como punto de corte para diferenciar compuestos activos de inactivos en la campaña de CV. De acuerdo con los datos de las curvas ROC, esto corresponde a una Se del 77% y una Sp del 98% en el conjunto de entrenamiento y una Se de 71% y una Sp de 95% en el conjunto de prueba. Tal como en

las campañas anteriores en la selección del punto de corte, hemos priorizado Sp por encima de Se para evitar la selección de FP en la campaña de CV.

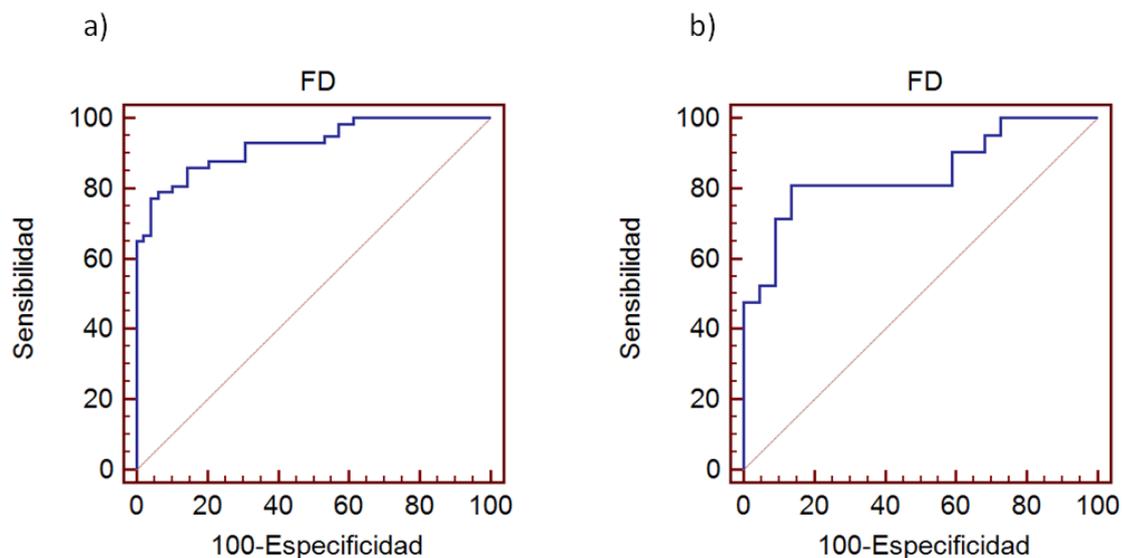


Figura 5.13. Curva ROC correspondiente al conjunto de entrenamiento (a) y conjunto de prueba (b) correspondientes al modelo aplicado en la tercera campaña de CV.

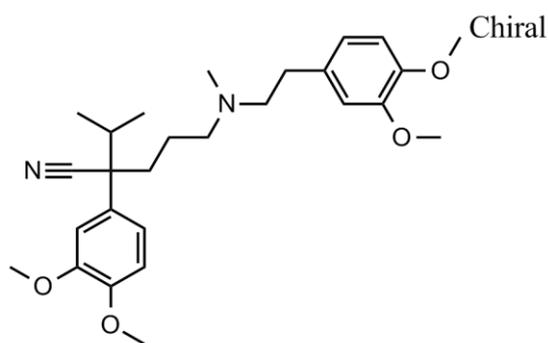
5.4.2. CRIBADO VIRTUAL

El modelo se aplicó en el CV de 5.570 compuestos químicos de la base de datos Merck Index 12^a edición. 154 candidatos pertenecientes al dominio de aplicabilidad del modelo presentan una puntuación por encima del *score* seleccionado; 34 de ellos corresponden a fármacos aprobados, que son los candidatos más favorables, con fines directos de reposicionamiento, tal como hemos señalado anteriormente.

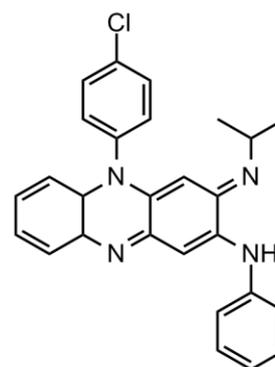
Sobre la base de su accesibilidad, cuatro de ellos (Figura 5.14) fueron adquiridos y probados experimentalmente en el ensayo enzimático en extracto crudo de Cz. Los candidatos seleccionados fueron verapamilo (utilizado en el tratamiento de la hipertensión y la arritmia cardíaca), la clofazimina (presenta actividad antiinflamatoria y es utilizado en el tratamiento frente a la lepra), benidipina (utilizado en el tratamiento de la hipertensión y la enfermedad isquémica cardíaca) y saquinavir (un fármaco

antirretroviral utilizado en el tratamiento del VIH). Todos los fármacos a evaluar se adquirieron en Sigma Aldrich.

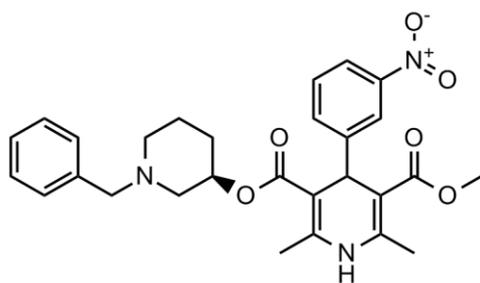
Esta campaña fue realizada durante mi estadía en el año 2012, en la Unidad de Diseño Molecular y Conectividad Molecular, Facultad de Farmacia, Universidad de Valencia, España (subsidió de viaje UNLP-2012). Dado que el uso de campañas de CV simulado para validar el modelo fue implementado con posterioridad, y que no se dispone del programa Desmol11 utilizado para generar el modelo de la tercera campaña, la etapa de CV simulado no fue realizada para este tercer modelo.



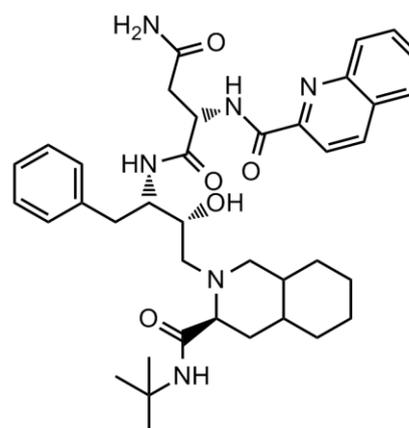
Verapamilo



Clofazimina



Benidipina



Saquinavir

Figura 5.14. Estructuras de los fármacos adquiridos luego de la tercera campaña de CV.

5.4.3. EVALUACIÓN BIOLÓGICA

Como primera aproximación bioquímica se evaluó el efecto de los cuatro candidatos seleccionados (a 100 μM) sobre la actividad de Cz de *T. cruzi* en extracto purificado, utilizando Bz-Pro-Phe-Arg-pNA como sustrato cromogénico.

Mediante una búsqueda bibliográfica y el análisis de la base de datos de ensayos biológicos de PubChem (NCBI PubChem Bioassay Database) encontramos que clofazimina ha sido probada en Cz previamente con resultados ambiguos. Doak y colaboradores reportan la actividad inhibitoria en presencia y ausencia de Triton 0.01% (IC_{50} de 40 y 6 μM) (AIDs: 484274, 484275) (Doak et al., 2010). Jadhav y colaboradores realizaron una campaña de cribado de alta performance en la que se evaluó el efecto de 197.000 compuestos sobre la hidrólisis del sustrato Z-Phe-Arg-AMC por Cz en presencia y ausencia de detergente reportando que no presenta actividad inhibitoria sobre Cz (Jadhav et al., 2010) (AIDs: 1476, 1478). Por lo tanto decidimos realizar el ensayo de clofazimina para comprobar la actividad inhibitoria sobre Cz y en caso de obtener resultados positivos continuar con los ensayos de proliferacion de *T. cruzi*. En relación a saquinavir los resultados de Jadhav no fueron concluyentes sobre el efecto sobre Cz, por lo que también se decidió que valía la pena evaluarlo experimentalmente (AIDs:1476, 1478). No se encontraron reportes anteriores de efecto inhibitorio de benidipina ni verapamilo sobre Cz .

Los resultados del ensayo de inhibición de Cz demostraron que verapamilo no afecta a la actividad de la enzima; los otros tres candidatos mostraron efectos inhibitorios significativos (Figura 5.15). Tales inhibiciones son dependientes de la concentración, con una IC_{50} de $106,29 \pm 20,5 \mu\text{M}$ para clofazimina (Figura 5.16a), $72,7 \pm 8,1 \mu\text{M}$ para benidipina (Figura 5.16b) y $54,12 \pm 8,25 \mu\text{M}$ para saquinavir.

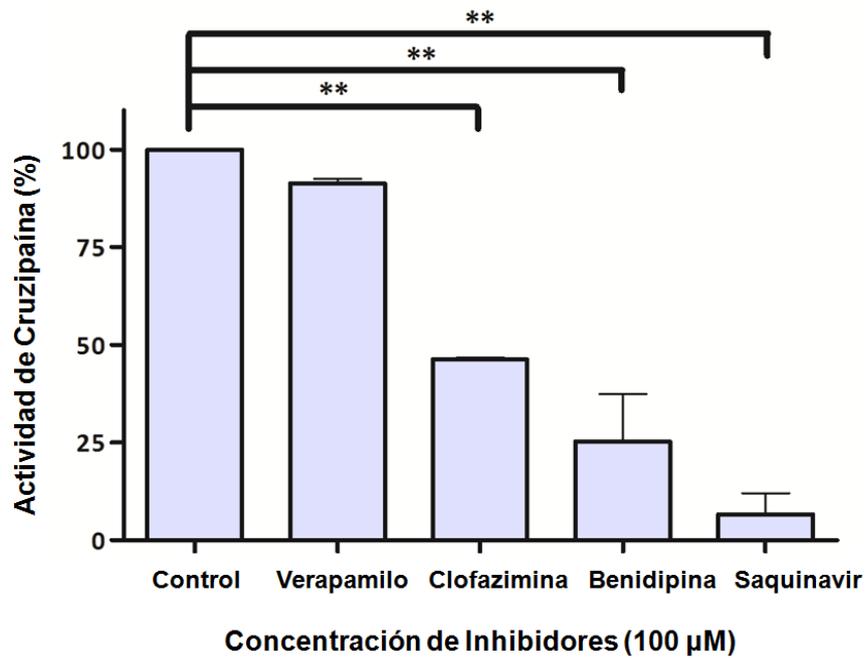


Figura 5.15. Efecto inhibitorio de los cuatro candidatos seleccionados en la tercer campaña de CV sobre la actividad de Cz de *T. cruzi* purificada. La concentración final de cada compuesto fue de 100 µM. La actividad de Cz se expresa como porcentaje de la condición control (2% DMSO). Los resultados representan la media de dos experimentos independientes. Los asteriscos indican diferencias significativas con el control (** p < 0,01).

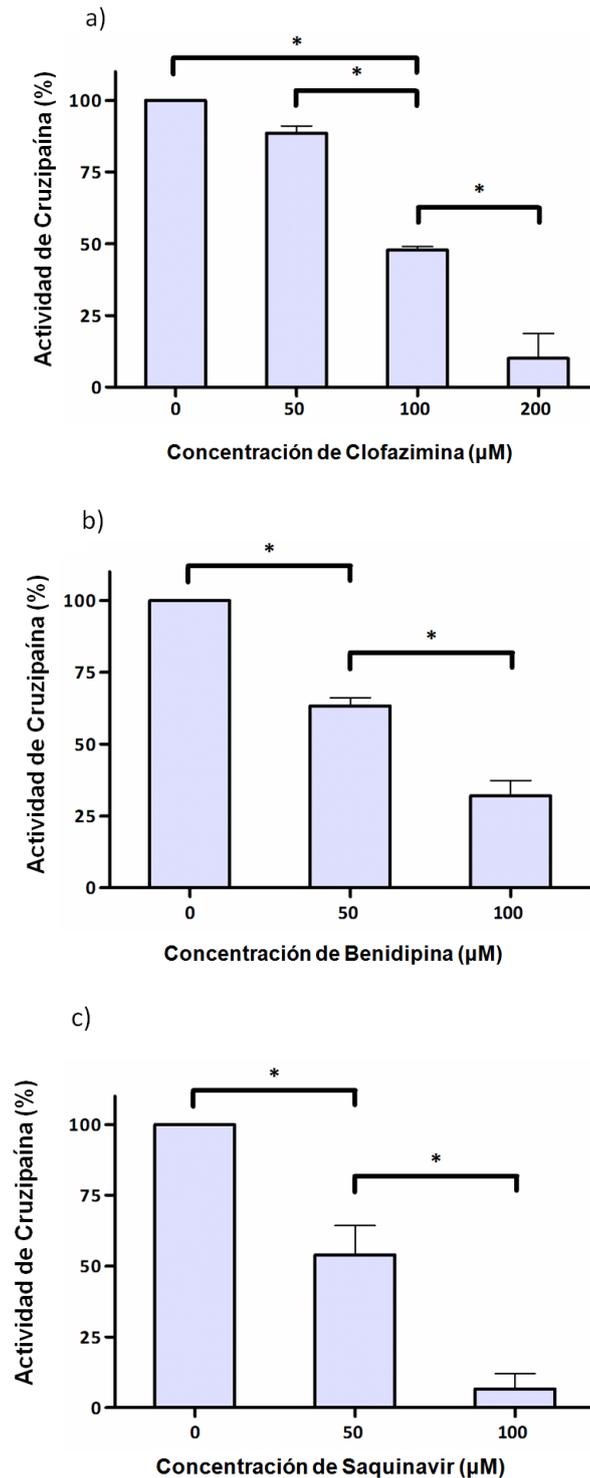


Figura 5.16. Efecto inhibitorio dependiente de la concentración sobre Cz purificada de clofazimina (a), benidipina (b) y saquinavir (c). La actividad remanente de la enzima se expresa como porcentaje del control (2% DMSO). El rango de concentraciones ensayadas para cada fármaco depende de su solubilidad. Los resultados representan la media de dos experimentos independientes. Los asteriscos indican diferencias significativas (* $p < 0,05$).

En el siguiente ensayo se evaluó la capacidad de clofazimina, benidipina y saquinavir de afectar la proliferación de diferentes formas parasitarias de *T. cruzi*. Inicialmente, se ha demostrado que afectan claramente la proliferación de epimastigotes de *T. cruzi* (Figura 5.17) presentando valores de concentración inhibitoria media (IC_{50}) de $10,6 \mu\text{M}$ para clofazimina (Figura 5.17a) y $19,5 \mu\text{M}$ para benidipina (Figura 5.17b) en la fase logaritmica media de los controles (cuarto día). Las concentraciones mayores a $20 \mu\text{M}$ de clofazimina resultaron letales en pocas horas. Saquinavir afectó la proliferación de epimastigotes (Figura 5.17c), sin embargo su efecto se observó tardíamente, en la fase estacionaria; no fueron evaluadas concentraciones mayores de saquinavir debido a su baja solubilidad en DMSO.

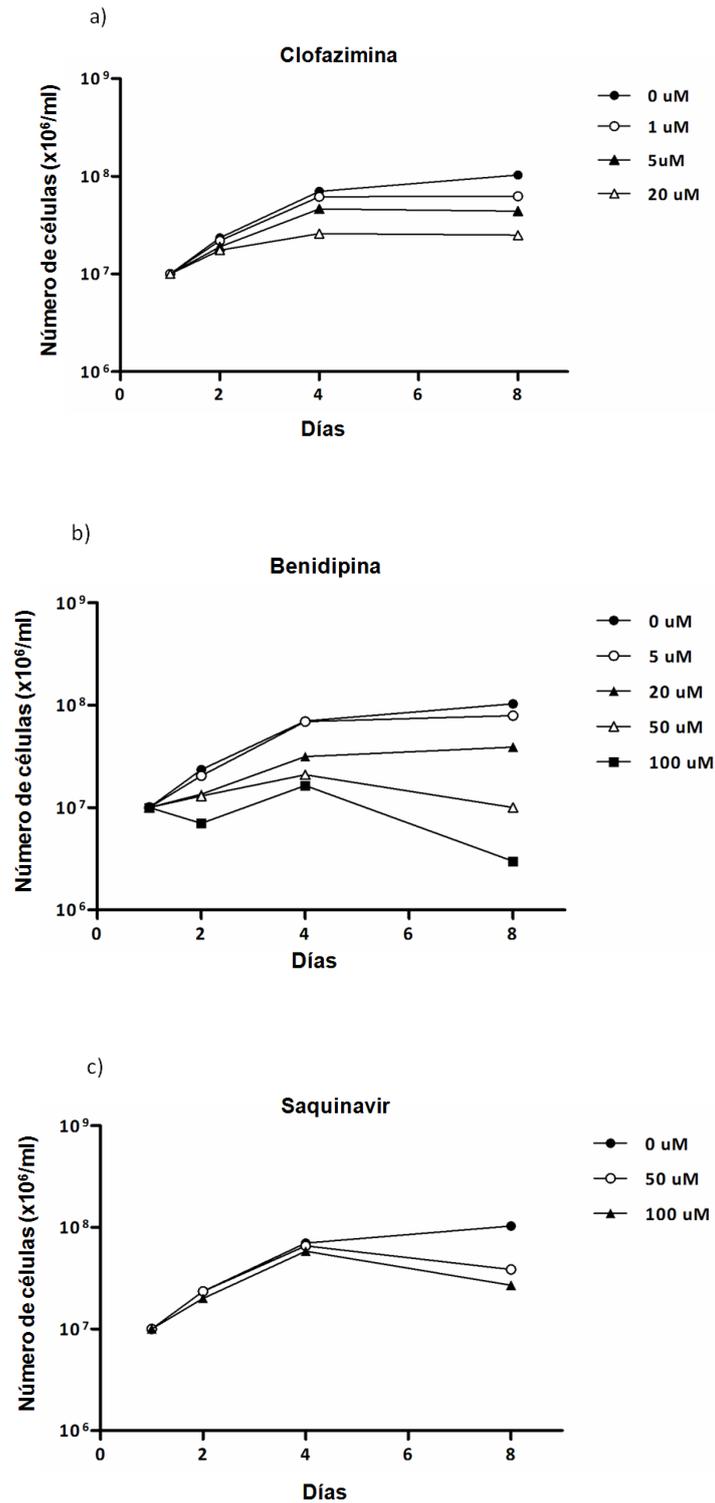


Figura 5.17. Efectos sobre la proliferación de epimastigotes de clofazimina (a), benidipine (b) y saquinavir (c). Para determinar la tasa de crecimiento, se sembraron 10⁷ células / ml en medio de BHT y se mantuvo a 28 °C durante 8 días, respectivamente. El rango de concentraciones ensayadas para cada droga se debe a la solubilidad de las mismas. La condición de control se realizó con 2% DMSO. Los parásitos se contaron utilizando una cámara hemocitométrica (cámara de Neubauer). Los resultados representan la media ± DE.

Los inhibidores de Cz seleccionados en esta campaña se estudiaron a continuación en distintas formas intracelulares de *T. cruzi* (método descrito en el capítulo 4 sección 4.4.3).

Los compuestos se ensayaron sobre los cultivos celulares de tripomastigotes en medio axénico durante los 30 minutos previos a la infección y luego se transfirieron a monocapas de células H9c2 durante un período de 12 horas en DMEM suplementado con 3% de SFB y antibióticos (medio de infección) en ausencia y presencia de diferentes concentraciones de inhibidores de Cz (procedimiento detallado en la sección 4.4.3 del capítulo 4) (ver esquema en la Figura 5.18a). En este contexto, los parásitos fueron expuestos a los 3 fármacos seleccionados durante el proceso de invasión, transporte a través de vacuolas y replicación en amastigotes. Cada inhibidor se ensayó a una concentración no citotóxica, definida por otros investigadores (2,5-10 μM) (Kim et al., 1998; Ono et al., 2002; Van Rensburg et al., 1993). Dichas concentraciones se encuentran dentro del rango de IC_{50} determinado en los ensayos de epimastigotes (ver Figura 5.17).

La infección se cuantificó por microscopia confocal de las muestras fijadas. Los parásitos se visualizaron directamente debido a la expresión estable de TcH2b-GFP (histona 2B marcada con GFP, proteína verde fluorescente) (Ramirez et al., 2000) y las células huésped se observaron mediante tinción de la actina del citoesqueleto con faloindina-TRITC (Figura 5.18b). Los resultados muestran que los tres candidatos redujeron más de tres veces el número de células infectadas (Figura 5.18c). El número de parásitos en cada célula también se redujo significativamente lo que indica que la división de amastigotes recientemente formados se altera en presencia de inhibidores de Cz (Figura 5.18d).

Luego se analizó el efecto de los inhibidores de Cz sobre la proliferación de amastigotes. En este caso las células fueron infectadas durante 24 hs en un medio de control y después de los lavados fueron expuestas a los fármacos y se incubaron por un tiempo adicional de 96 hs (Figura 5.19a). Los datos cuantitativos muestran que los inhibidores de Cz también redujeron el número de amastigotes por célula (Figura 5.19b).

La posible toxicidad de los inhibidores de Cz sobre las células huésped también fue evaluada mediante el ensayo AlamarBlue (ver detalles en el capítulo 4). Como se muestra en la Figura 5.20, la actividad metabólica de las células huésped bajo cada tratamiento fue similar a la concición control confirmando que los inhibidores de Cz tienen acción selectiva sobre *T. cruzi*.

En la tabla 5.1 se presenta un resumen de los modelos generados, parámetros de las curvas ROC del conjunto de prueba, resultados de cada campaña de CV realizada y fármacos que resultaron activos en ensayos *in vitro* frente a Cz y proliferación de epimastigotes. A continuación se detallan los motivos por los cuales se seleccionaron 3 de los 5 fármacos que resultaron activos para evaluarlos en un modelo *in vivo* de infección aguda.

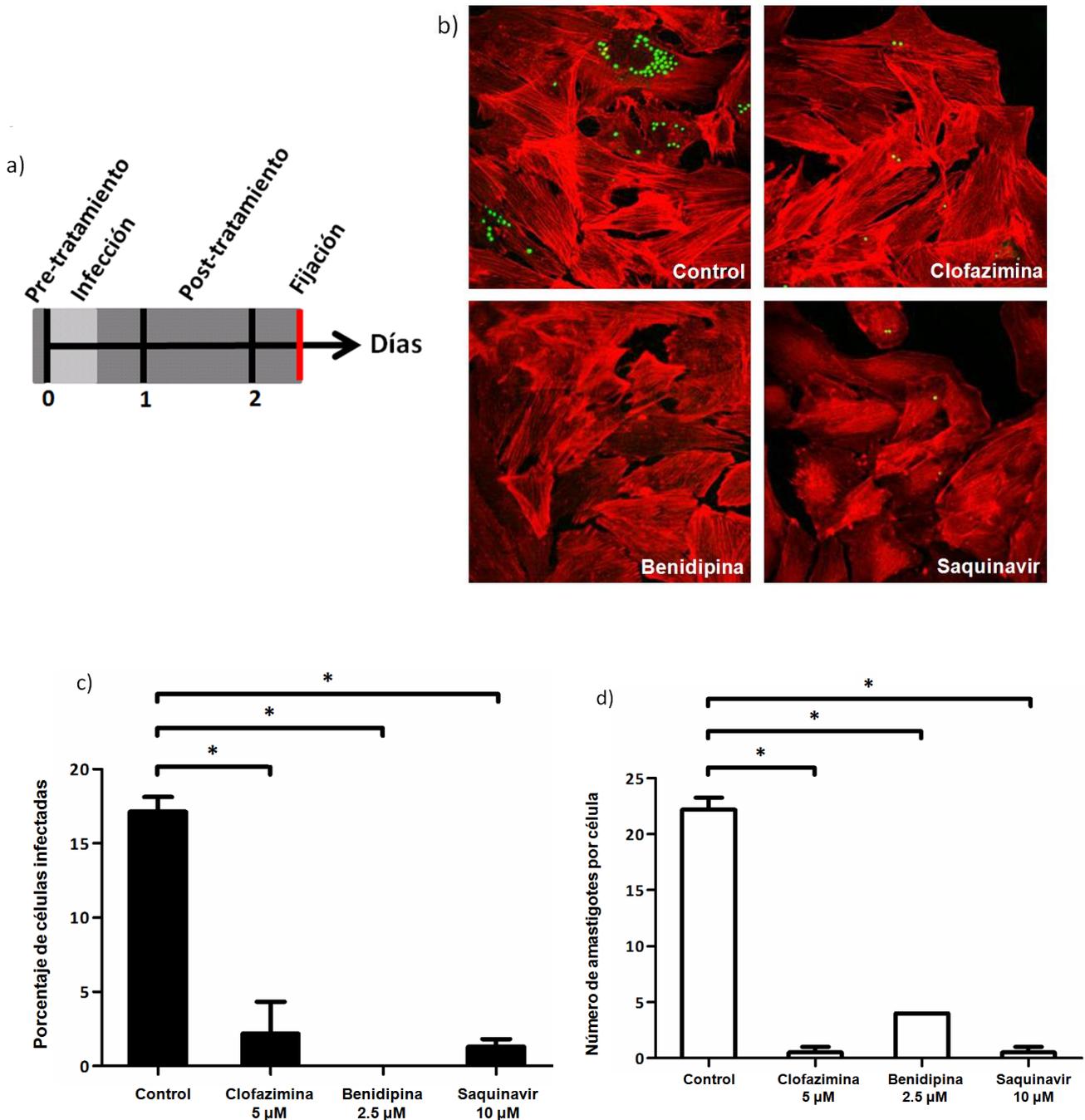


Figura 5.18. Efecto de clofazimina, benidipina y saquinavir sobre la infectividad de *T. cruzi*. a) Representación esquemática del protocolo realizado; b) Imágenes obtenidas en microscopio confocal Olympus FV1000, los amastigotes se observan en color verde (tinción de la actina) y las células en color rojo; c) Porcentaje de células infectadas; d) Número de parásitos por célula en presencia de los inhibidores indicados. Para el análisis estadístico se utilizó la prueba de Turkey-Kramer. Los datos representan la media \pm DE de al menos tres experimentos independientes (número de células contadas \approx 100). Los asteriscos indican diferencias significativas (* $p < 0,05$).

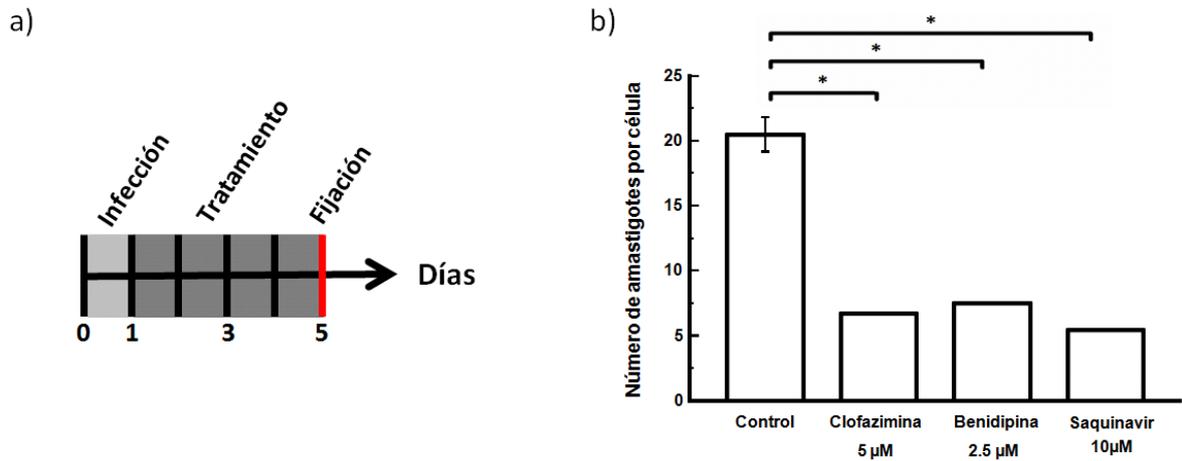


Figura 5.19. Efectos de clofazimina, benidipina y saquinavir sobre la proliferación de amastigotes de *T. cruzi*: Se utilizaron amastigotes TCT de la cepa Y-GFP de *T. cruzi* para infectar monocapas de células H9c2 (MOI = 10) durante 24 hs. Las células se lavaron, se añadió medio fresco con inhibidores y se incubaron por un período adicional de 96 hs. Posteriormente fueron fijadas, teñidas y se prepararon para el análisis microscópico como se describe en el capítulo 4. a) Representación esquemática del protocolo realizado. b) Número de amastigotes por célula en los inhibidores indicados.

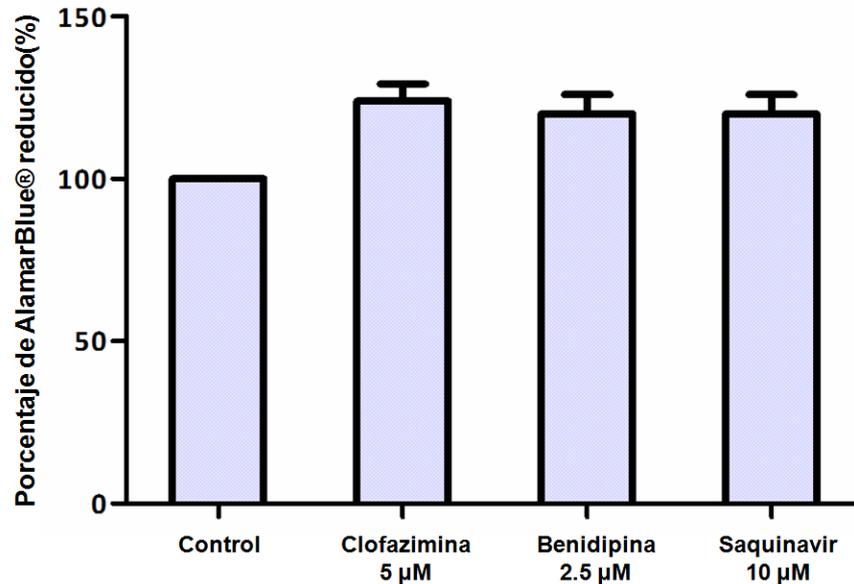


Figura 5.20. Ensayo AlamarBlue. Células H9c2 fueron tratadas con inhibidores de Cz durante 48 horas en las concentraciones indicadas y la actividad metabólica se evaluó con el ensayo AlamarBlue. Los resultados se expresan como porcentaje de reducción de reactivo por la actividad mitocondrial de las células. Los datos representan la media \pm DE de al menos tres experimentos independientes.



Tabla 5.1. Resumen de los modelos generados y utilizados en las 3 campañas de CV.

Modelo	Descriptorios incluidos	Parámetros de la curva ROC (correspondientes al conjunto de prueba)		Resultado del CV (compuestos pertenecientes al dominio de aplicabilidad)	Fármacos seleccionados Aprobados por la FDA (candidatos al reposicionamiento)	Fármacos Adquiridos	IC ₅₀ de Fármacos activos en ensayos de proliferación de epimastigotes
		AUC	Se Sp				
Modelo 1 $\lambda = 0,56$	GGI7 nS nCN	0,921	75% 100%	256	54	Amiodarona Bromocriptina Colchicina Escitalopram	Amiodarona (IC ₅₀ =22 μ M) Bromocriptina (IC ₅₀ =15 μ M)
Modelo 2 $\lambda = 0,445$	VE1_X C-018 F-084 NsssN H-051 O-056 Sds SpPosA_A	0,923	81% 95%	64	54	Cisaprida Paroxetina Levotiroxina	Levotiroxina (IC ₅₀ =121,8 μ M)
Modelo 3 $\lambda = 0,428$	G4V X4P G1 VV3	0,862	71% 95%	154	34	Saquinavir Verapamilo Benidipina Clofazimina	Clofazimina (IC ₅₀ =10,6 μ M) Benidipina (IC ₅₀ =19,5 μ M)

5.5. ENSAYO *IN VIVO*. MODELO DE INFECCIÓN AGUDA EN RATONES

5.5.1. SELECCIÓN DE CANDIDATOS A SER EVALUADOS EN ENSAYOS PRECLÍNICOS

Se emplearon los criterios descriptos en el capítulo anterior para definir cuáles de todos los candidatos seleccionados que presentaron resultados positivos en los ensayos enzimáticos y biológicos de las tres campañas de CV realizadas avanzarían hacia fase pre-clínica. En este sentido, encontramos que los fármacos amiodarona y bromocriptina (seleccionados en la primer campaña de CV) presentan la limitación de que con las dosis empleadas habitualmente en clínica, alcanzan concentraciones plasmáticas en el estado estacionario del orden de 4 μM para amiodarona (Rotmensch et al., 1984) y 0,001 μM para bromocriptina (Flogstad et al., 1994); estas concentraciones distan mucho de los valores que se obtuvieron en el ensayo de proliferación de epimastigotes (IC_{50} de 22 μM para amiodarona y de 15 μM para bromocriptina).

Levotiroxina (seleccionado en la segunda campaña de CV) presenta 2 inconvenientes importantes, está contraindicada en pacientes cardíacos y las concentraciones plasmáticas en el estado estacionario que se consiguen con las dosis de este fármaco empleadas en clínica (3×10^{-5} μM) (Cerutti et al., 1999) están muy por debajo de las concentraciones que resultaron efectivas sobre la proliferación de epimastigotes (121,8 μM), lo que limita el potencial de este candidato como fármaco antichagásico, además de las posibles alteraciones metabólicas que podría producir esta terapia hormonal al ser administrada a una persona con función tiroidea normal.

Clofazimina, se indica principalmente en el tratamiento de la lepra (Cholo et al., 2012). Se administran dosis de 200-300 mg al día bajo supervisión médica, por no más de 3 meses, para el tratamiento del eritema nudoso leproso. Cabe destacar que los tratamientos actuales para tratar la enfermedad de Chagas tienen una duración de 2 a 3 meses. La absorción de clofazimina administrada por vía oral es variable, la administración diaria de 100, 300, 400 y 600 mg de este fármaco a pacientes que presentan lepra resulta en niveles medios plasmáticos de 1,5; 2,1; 3,0 y 8,5 μM respectivamente (Cholo et al., 2012); es decir los niveles alcanzados en el estado

estacionario con clofazimina en las dosis utilizadas para la indicación original son similares a las concentraciones eficaces obtenidas en los ensayos sobre *T. cruzi*. Los efectos adversos de clofazimina están relacionados generalmente con la dosis. Principalmente afecta a la piel, los ojos y el tracto gastrointestinal; estos efectos son tolerables y reversibles al cesar la terapia. En relación a los efectos secundarios del tracto gastrointestinal son leves a moderados (dolor abdominal, epigástrico, náuseas, diarreas, vómitos); con menor frecuencia pueden presentarse efectos graves a dosis altas que requieren supervisión médica.

La coincidencia entre los niveles de clofazimina que muestran efectos antichagásicos y los utilizados para la indicación clínica aprobada anteriormente sumado a un perfil razonable de efectos adversos, hacen que clofazimina sea un buen candidato para pasar a los ensayos pre-clínicos.

Benidipina es un bloqueante de canales de calcio que actualmente se utiliza para el tratamiento de la hipertensión y la angina de pecho (Yao et al., 2006). Existe abundante evidencia científica sobre las propiedades cardioprotectoras de este fármaco (Asanuma et al., 2001; Gao et al., 2001; Liu et al., 2004; Wang et al., 2004) lo cual podría representar un beneficio adicional para los pacientes con Chagas. Sus efectos adversos más frecuentes son leves y se encuentran relacionados con su actividad vasodilatadora, incluyendo palpitaciones, sofocos y jaquecas. Además, otras dihidropiridinas de uso clínico han demostrado previamente efecto antitripanosomal (Reimao et al., 2010).

Por último saquinavir no fue candidato para los ensayos pre-clínicos. Aunque generalmente es un fármaco bien tolerado por los pacientes, la concentración máxima en el estado estacionario es de alrededor de 0,15 μM (Noble and Faulds, 1996), muy inferior a las concentraciones que afectaron la proliferación de epimastigotes y amastigotes. Por otro lado al ser un fármaco antiviral no existe evidencia o sugerencias de que podría presentar beneficios adicionales para pacientes que padecen Chagas.

Por estas razones, fueron clofazimina y benidipina los fármacos evaluados en el modelo de infección aguda en ratones.

5.5.2. MODELO DE INFECCIÓN AGUDA EN RATONES

Los fármacos seleccionados para pasar al modelo animal, clofazimina y benidipina, fueron ensayados individualmente en un modelo murino de infección aguda experimental en una dosis de 20 mg/kg/día y 10 mg/kg/día respectivamente (Figura 5.20). En otros dos grupos de animales se administraron 1% de DMSO (control) y 100 mg/kg/día de benznidazol) para incluir grupos no tratados y tratados con el fármaco aprobado de elección respectivamente. Los valores más elevados de parasitemia se observaron en los animales no tratados luego de 14 días de la infección (14 dpi) mientras que el grupo tratado con benznidazol presenta el menor valor en las mismas condiciones.

Benidipina y clofazimina redujeron significativamente la parasitemia en un 40-60% en comparación con los valores del control negativo (Figura 5.20a). Todos los animales no tratados murieron entre los días 12 y 14 luego de la infección. Se observó una respuesta similar en el grupo tratado con benidipina (control: media de $13,6 \pm 0,5$ dpi; benidipina: media de $13,6 \pm 1,5$ dpi). Por el contrario, la mortalidad en el grupo tratado con clofazimina fue más lenta (media: $14,6 \pm 0,5$ dpi), mientras que con benznidazol se observó un 50% de mortalidad a los 25 dpi).

Por último hemos realizado un análisis *post-mortem* para verificar la presencia de parásitos en el corazón mediante microscopia convencional de secciones de tejido cardíaco. Realizando una tinción con hematoxilina y eosina (HE) se pueden visualizar claramente los nidos de amastigotes de *T. Cruzi* entre las fibras musculares (Figura 5.20b). Los datos muestran que el número de nidos de amastigotes cardíacos se redujo significativamente en los corazones de los animales tratados con clofazimina; tal como se esperaba benznidazol también produjo un bajo parasitismo cardíaco (Figura 5.20c).

Analizando en conjunto los ensayos, estos resultados ponen en evidencia la acción beneficiosa del tratamiento con clofazimina en el proceso de infección de los animales y justifican nuevos experimentos (en un futuro) aplicando modelos crónicos de infección y posteriores ensayos clínicos.

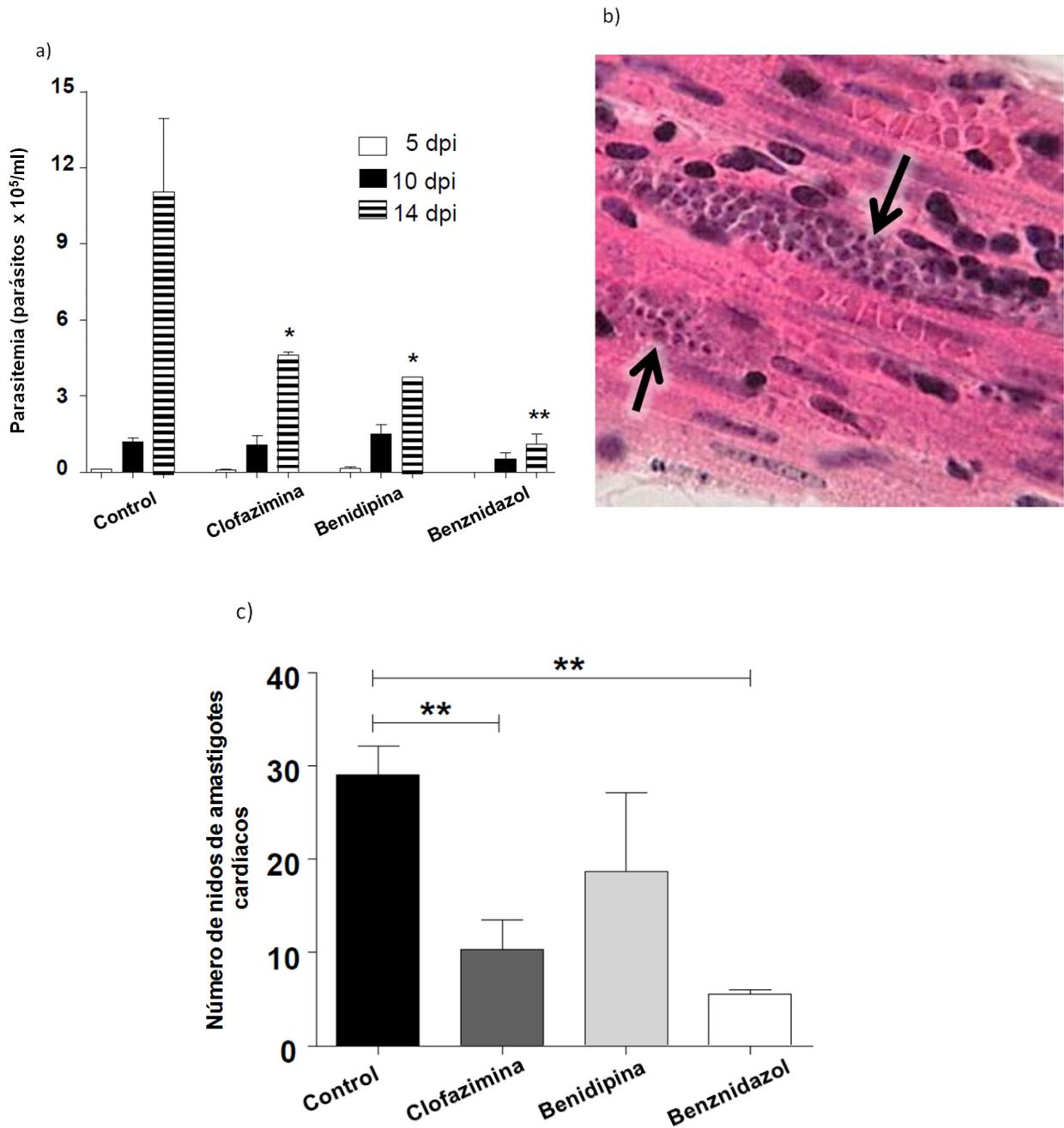


Figura 5.20. Efectos de clofazimina y benidipina sobre modelo murino de infección aguda. Los ratones (3 animales por grupo) se infectaron con 1×10^5 tripomogastigotes provenientes del cultivo celular) vía intra peritoneal y la parasitemia fue evaluada a 5dpi. Los animales con parásitos circulantes fueron tratados diariamente con 1% de DMSO en solución salina (control) o con clofazimina, benidipina y benznidazol a la dosis indicada. La parasitemia se cuantificó a los 5, 10, 14 y 20 dpi y los resultados se expresaron en número de parásitos/ml de sangre (a). Se analizaron los cortes de tejido cardíaco de animales infectados luego de realizar la tinción con HE para visualizar los nidos de amastigotes (b). El número de nidos de *T. cruzi* fueron cuantificados por microscopía convencional de los tejidos cardíacos de cada animal para cada tratamiento (c). Los asteriscos indican diferencias significativas respecto a los controles (* $p < 0,05$; ** $p < 0,01$).

REFERENCIAS CAPÍTULO 5

Ang K.K., Ratnam J., Gut J., Legac J., Hansell E., Mackey Z.B., Skrzypczynska K.M., Debnath A., Engel J.C., Rosenthal P.J., McKerrow J.H., Arkin M.R., Renslo A.R. Mining a cathepsin inhibitor library for new antiparasitic drug leads. *PLoS Negl Trop Dis.* **5**, e1023. (2011).

Asanuma H., Kitakaze M., Node K., Takashima S., Sakata Y., Asakura M., Sanada S., Shinozaki Y., Mori H., Tada M., Kuzuya T., Hori M. Benidipine, a long-acting Ca channel blocker, limits infarct size via bradykinin- and NO-dependent mechanisms in canine hearts. *Cardiovasc Drugs Ther.* **15**, 225-231. (2001).

Beaulieu C., Isabel E., Fortier A., Masse F., Mellon C., Methot N., Ndao M., Nicoll-Griffith D., Lee D., Park H., Black W.C. Identification of potent and reversible cruzipain inhibitors for the treatment of Chagas disease. *Bioorg Med Chem Lett.* **20**, 7444-7449. (2010).

Benaïm G., Sanders J.M., Garcia-Marchan Y., Colina C., Lira R., Caldera A.R., Payares G., Sanoja C., Burgos J.M., Leon-Rossell A., Concepcion J.L., Schijman A.G., Levin M., Oldfield E., Urbina J.A. Amiodarone has intrinsic anti-*Trypanosoma cruzi* activity and acts synergistically with posaconazole. *J Med Chem.* **49**, 892-899. (2006).

Brak K., Doyle P.S., McKerrow J.H., Ellman J.A. Identification of a new class of nonpeptidic inhibitors of cruzain. *J Am Chem Soc.* **130**, 6404-6410. (2008).

Cerutti R., Rivolta G., Cavalieri L., Di Giulio C., Grossi E., Vago T., Baldi G., Righini V., Marzo A. Bioequivalence of levothyroxine tablets administered to a target population in steady state. *Pharmacol Res.* **39**, 193-201. (1999).

Choe Y., Brinen L.S., Price M.S., Engel J.C., Lange M., Grisostomi C., Weston S.G., Pallai P.V., Cheng H., Hardy L.W., Hartsough D.S., McMakin M., Tilton R.F., Baldino C.M., Craik C.S. Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorg Med Chem.* **13**, 2141-2156. (2005).

Cholo M.C., Steel H.C., Fourie P.B., Germishuizen W.A., Anderson R. Clofazimine: current status and future prospects. *J Antimicrob Chemother.* **67**, 290-298. (2012).

Doak A.K., Wille H., Prusiner S.B., Shoichet B.K. Colloid formation by drugs in simulated intestinal fluid. *J Med Chem.* **53**, 4259-4265. (2010).

dos Santos Filho J.M., Leite A.C., de Oliveira B.G., Moreira D.R., Lima M.S., Soares M.B., Leite L.F. Design, synthesis and cruzain docking of 3-(4-substituted-aryl)-1,2,4-oxadiazole-N-acylhydrazones as anti-*Trypanosoma cruzi* agents. *Bioorg Med Chem.* **17**, 6682-6691. (2009).

Du X., Hansell E., Engel J.C., Caffrey C.R., Cohen F.E., McKerrow J.H. Aryl ureas represent a new class of anti-trypanosomal agents. *Chem Biol.* **7**, 733-742. (2000).

Ferreira R.S., Simeonov A., Jadhav A., Eidam O., Mott B.T., Keiser M.J., McKerrow J.H., Maloney D.J., Irwin J.J., Shoichet B.K. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J Med Chem.* **53**, 4891-4905. (2010).

Flogstad A.K., Halse J., Grass P., Abisch E., Djoseand O., Kutz K., Bodd E., Jervell J. A comparison of octreotide, bromocriptine, or a combination of both drugs in acromegaly. *J Clin Endocrinol Metab.* **79**, 461-465. (1994).

Galvez J., Garcia-Domenech R., Salabert M.T., Soler R. Charge Indexes. New Topological Descriptors. *J Chem Inf Comput Sci.* **34**, 520-525. (1994).

Gao F., Gong B., Christopher T.A., Lopez B.L., Karasawa A., Ma X.L. Anti-apoptotic effect of benidipine, a long-lasting vasodilating calcium antagonist, in ischaemic/reperfused myocardial cells. *Br J Pharmacol.* **132**, 869-878. (2001).

Hernandes M.Z., Rabello M.M., Leite A.C., Cardoso M.V., Moreira D.R., Brondani D.J., Simone C.A., Reis L.C., Souza M.A., Pereira V.R., Ferreira R.S., McKerrow J.H. Studies toward the structural optimization of novel thiazolyldrazone-based potent antitrypanosomal agents. *Bioorg Med Chem.* **18**, 7826-7835. (2010).

Huang L., Brinen L.S., Ellman J.A. Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorg Med Chem.* **11**, 21-29. (2003).

Jadhav A., Ferreira R.S., Klumpp C., Mott B.T., Austin C.P., Inglese J., Thomas C.J., Maloney D.J., Shoichet B.K., Simeonov A. Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *J Med Chem.* **53**, 37-51. (2010).

Kier L.B., Hall L.H. General definition of valence delta-values for molecular connectivity. *J Pharm Sci.* **72**, 1170-1173. (1983).

Kier L.B., Murray W.J., Randic M., Hall L.H. Molecular connectivity V: connectivity series concept applied to density. *J Pharm Sci.* **65**, 1226-1230. (1976).

Kim R.B., Fromm M.F., Wandel C., Leake B., Wood A.J., Roden D.M., Wilkinson G.R. The drug transporter P-glycoprotein limits oral absorption and brain entry of HIV-1 protease inhibitors. *J Clin Invest.* **101**, 289-294. (1998).

Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A.C., Wishart D.S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035-1041. (2011).

Liu H.R., Gao F., Tao L., Yan W.L., Gao E., Christopher T.A., Lopez B.L., Hu A., Ma X.L. Antiapoptotic mechanisms of benidipine in the ischemic/reperfused heart. *Br J Pharmacol.* **142**, 627-634. (2004).

Mott B.T., Ferreira R.S., Simeonov A., Jadhav A., Ang K.K., Leister W., Shen M., Silveira J.T., Doyle P.S., Arkin M.R., McKerrow J.H., Inglese J., Austin C.P., Thomas C.J., Shoichet B.K., Maloney D.J. Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB. *J Med Chem.* **53**, 52-60. (2010).

- Noble S., Faulds D. Saquinavir. A review of its pharmacology and clinical potential in the management of HIV infection. *Drugs*. **52**, 93-112. (1996).
- Ono T., Liu N., Kusano H., Nogaki F., Makino T., Muso E., Sasayama S. Broad antiproliferative effects of benidipine on cultured human mesangial cells in cell cycle phases. *Am J Nephrol*. **22**, 581-586. (2002).
- Ramirez M.I., Yamauchi L.M., de Freitas L.H., Jr., Uemura H., Schenkman S. The use of the green fluorescent protein to monitor and improve transfection in *Trypanosoma cruzi*. *Mol Biochem Parasitol*. **111**, 235-240. (2000).
- Reimao J.Q., Scotti M.T., Tempone A.G. Anti-leishmanial and anti-trypanosomal activities of 1,4-dihydropyridines: In vitro evaluation and structure-activity relationship study. *Bioorg Med Chem*. **18**, 8044-8053. (2010).
- Rodrigues C.R., Flaherty T.M., Springer C., McKerrow J.H., Cohen F.E. CoMFA and HQSAR of acylhydrazide cruzain inhibitors. *Bioorg Med Chem Lett*. **12**, 1537-1541. (2002).
- Rotmensch H.H., Belhassen B., Swanson B.N., Shoshani D., Spielman S.R., Greenspon A.J., Greenspan A.M., Vlasses P.H., Horowitz L.N. Steady-state serum amiodarone concentrations: relationships with antiarrhythmic efficacy and toxicity. *Ann Intern Med*. **101**, 462-469. (1984).
- Scheidt K.A., Roush W.R., McKerrow J.H., Selzer P.M., Hansell E., Rosenthal P.J. Structure-based design, synthesis and evaluation of conformationally constrained cysteine protease inhibitors. *Bioorg Med Chem*. **6**, 2477-2494. (1998).
- Tropsha A., Gramatica P., Gombar V. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb Sci*. **22**, 69-77. (2003).
- Truchon J.F., Bayly C.I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J Chem Inf Model*. **47**, 488-508. (2007).
- Van Rensburg C.E., Van Staden A.M., Anderson R. The riminophenazine agents clofazimine and B669 inhibit the proliferation of cancer cell lines in vitro by phospholipase A2-mediated oxidative and nonoxidative mechanisms. *Cancer Res*. **53**, 318-323. (1993).
- Wang N., Minatoguchi S., Chen X.H., Arai M., Uno Y., Lu C., Misao Y., Nagai H., Takemura G., Fujiwara H. Benidipine reduces myocardial infarct size involving reduction of hydroxyl radicals and production of protein kinase C-dependent nitric oxide in rabbits. *J Cardiovasc Pharmacol*. **43**, 747-757. (2004).
- Yao K., Nagashima K., Miki H. Pharmacological, pharmacokinetic, and clinical properties of benidipine hydrochloride, a novel, long-acting calcium channel blocker. *J Pharmacol Sci*. **100**, 243-261. (2006).

CAPÍTULO 6

CONCLUSIONES



Durante el presente trabajo de tesis se obtuvieron 3 modelos clasificadores con la capacidad de diferenciar inhibidores de no inhibidores de Cz, un blanco molecular validado para la búsqueda de nuevos tratamientos para la enfermedad de Chagas.

Los modelos generados presentaron una muy buena relación casos/descriptores y similar poder clasificador sobre el conjunto de entrenamiento y el conjunto de prueba, lo que sugiere buena capacidad predictiva y ausencia de sobreajuste. Dado que sólo incluyen descriptores independientes de la conformación, los mismos son particularmente aptos para la exploración eficiente de bibliotecas virtuales de fármacos mediante campañas de CV, ya que no se requiere análisis conformacional previo de las estructuras de tales bibliotecas.

Teniendo en cuenta el potencial del reposicionamiento de fármacos asistido por computadora, el mismo fue aplicado en la selección de candidatos antichagásicos a partir de la colección de pequeñas moléculas de uso clínico o en fase clínica de evaluación de la base DrugBank (modelos 1 y 2) y de la base de datos Merck Index 12ed (modelo 3). El análisis de curvas ROC permitió seleccionar racionalmente un valor de corte apropiado para diferenciar activos de inactivos. Se ha priorizado obtener una alta especificidad en el cribado para minimizar el número de FP (compuestos predichos como inhibidores de la Cz, pero que no poseen actividad inhibitoria).

De todos los candidatos seleccionados y adquiridos que fueron evaluados experimentalmente en ensayos enzimáticos y de inhibición de la proliferación de epimastigotes de *T. cruzi*. amiodarona (antiarrítmico), bromocriptina (antiparkinsoniano), levotiroxina (utilizado en el tratamiento de desórdenes de la tiroides), benidipina (antihipertensivo), clofazimina (antibiótico) y saquinavir (antirretroviral), demostraron un efecto inhibitorio sobre Cz dependiente de la concentración y efectos antiproliferativos en *T. cruzi*.

Tales resultados permitieron validar experimentalmente los modelos, por un lado, e ilustrar el potencial de la estrategia propuesta para el desarrollo de nuevos medicamentos para enfermedades olvidadas. No obstante, debe señalarse que al analizar los valores de inhibición de los fármacos seleccionados sobre la actividad de la

enzima y frente a *T. cruzi*, podemos inferir que la inhibición de la enzima Cz no es el único mecanismo involucrado, existiendo probablemente otros mecanismos que participan en la inhibición del crecimiento del parásito.

En relación a los ensayos pre-clínicos (modelo murino de infección aguda de la enfermedad de Chagas) se decidió avanzar únicamente con benidipina y clofazimina por los motivos detallados en la sección 5.3.1 del capítulo 5; estos fármacos presentaron una dosis efectiva sobre *T. cruzi* similar a los niveles plasmáticos en estado estacionario obtenidos en clínica con las dosis que habitualmente se administran para las indicaciones ya aprobadas; por otra parte, el perfil de efectos adversos es tolerable y compatible con la principal manifestación de la enfermedad (cardiomiopatía grave); en el caso de benidipina presenta adicionalmente efecto cardioprotector, que podría significar un beneficio adicional para el paciente con Chagas.

Cabe destacar que uno de los candidatos ensayados, clofazimina, no sólo ha demostrado ser eficaz en la reducción de los niveles de parásitos en el modelo de infección aguda en animales, sino también en la reducción del número de nidos de amastigotes de *T. cruzi* en el tejido cardíaco, lo que sugiere que podría ser útil para prevenir el desarrollo de daño cardíaco en personas que sufren de Chagas. Por este motivo se justifica realizar, a futuro, nuevos experimentos utilizando modelos crónicos de infección y posteriores ensayos clínicos.

El presente trabajo integra con éxito la búsqueda racional de fármacos asistida por computadora con la biología molecular, celular y ensayos pre-clínicos, lo cual confirma la utilidad de CV para desarrollar el reposicionamiento de fármacos basados en el conocimiento orientado a enfermedades olvidadas. Se visualiza una optimización de los recursos aplicados en el proceso de selección de potenciales nuevas terapias con una relativamente pequeña inversión de costos y tiempo.

Como hemos mencionado en el capítulo 3, la falta de innovación en el campo de los tratamientos antichagásicos responde a que el Chagas se encuadra en las llamadas “enfermedades olvidadas” en las cuales el sector farmacéutico privado se ha mostrado – históricamente- reticente a invertir debido al limitado retorno de inversión percibido. Debido a ello, son el sector público y las organizaciones sin fines de lucro los actores

salientes en la investigación y desarrollo de soluciones terapéuticas para estas enfermedades.

Por esto, la implementación de estrategias eficientes desde el punto de vista traslacional es fundamental para que los avances en el campo de la investigación alcancen a los potenciales beneficiarios de los mismos; en nuestro caso, que los compuestos con actividad antichagásica se transformen en un medicamento innovador disponible para los pacientes con Chagas.

Durante este trabajo de tesis se han realizado 5 presentaciones a congresos (2 nacionales y 3 internacionales), 4 publicaciones en revistas científicas con referato, 1 trabajo enviado, además de 1 capítulo de libro y 1 revisión. Se han obtenido por concurso 2 subsidios de jóvenes investigadores de la Universidad Nacional de La Plata, uno de ellos para realizar una estadía de investigación en el exterior en el año 2012 (estadía de 2 meses en la Unidad de Diseño y Conectividad Molecular de la Universidad de Farmacia de Valencia España bajo la tutoría del Dr. Jorge Gálvez). En el marco de dicha estadía se generó la tercera campaña de CV (con excepción de los ensayos biológicos, desarrollados en Argentina). El otro subsidio fue otorgado en el año 2013 para financiar parte de los ensayos experimentales del trabajo de de tesis.

En el año 2012 se ha obtenido una beca de la Bibliotheca Alexandrina (BA) y de The Academy of Sciences for the Developing Countries (TWAS) para asistir a TWAS/BioVisionAlexandria.NXT 2012 y BioVision Alexandria Conference (Alejandría, Egipto); el mismo año el Ministerio de Ciencia e Innovación Productiva (MINCyT) y el Centro Argentino Brasileño de Nanotecnología (CABN) otorgó una beca para asistir a la Escuela de Nanotecnología y Terapéutica en la Facultad de Farmacia de la Universidad de Porto Alegre, Brasil. En 2013 la Sociedad Iberoamericana de Bioinformática otorgó una beca para asistir al 4to. Congreso Argentino de Bioinformática y Biología Computacional (4CAB2C) y 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio).

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

Artículos:

 Pizzo C., Saiz C., Talevi A., Gavernet L., Palestro P., Bellera C., Blanch L., Benítez D., Cazzulo J. J., Chidichimo A., Wipf P., Mahler S. G. Synthesis of 2-hydrazolyl-4-thiazolidinones based on multicomponent reactions and biological evaluation against *Trypanosoma Cruzi*. *Chem Biol Drug Des.* **77**(3), 166-172 (2011).

 Talevi A., Bellera C. L., Di Ianni M., Duchowicz P. R. , Bruno-Blanch L. E., Castro E.A. An integrated drug development approach applying topological descriptors. *Curr Comput Aided Drug Des.* **8**(3), 172-181. (2012).

 Bellera C.L., Balcazar D., Alberca L., Labriola C., Talevi A., Carrillo C. Application of computer-aided drug repurposing in the search of new cruzipain inhibitors: discovery of amiodarone and bromocriptine inhibitory effects. *J Chem Inf Model.* **53**(9), 2402-2408. (2013).

 Bellera C.L., Balcazar D., Alberca L., Labriola C., Talevi A., Carrillo C. Identification of levothyroxine antichagasic activity through computer-aided drug repurposing. *Scientific World Journal.* **2014**, article279618. (2014).

 Bellera C. L., Balcazar D., Vanrell M. C., Casassa F., Labriola C., Gálvez J., Bruno-Blanch L. E., Romano P. S., Carrillo C., Talevi A. Knowledge-based drug repurposing: identification of antichagasic activity of clofazimine, benidipine and saquinavir. *Eur J Med Chem.* Enviado (Septiembre, 2014).

Revisiones:

 Talevi A., Bellera C. L., Di Ianni M., Gantner M., Bruno-Blanch L. E., Castro E. A. CNS drug development – lost in translation?. *Mini Rev Med Chem.* **12**(10), 959-970. (2012).

 Bellera C.L., Sbaraglini M.L., Balcazar D., Fraccaroli L., Varnell M.C., Casassa A.F., Labriola C., Romano P.S., Carrillo C., Talevi A. High-throughput drug repositioning for the discovery of new antichagasic medications. *Mini Rev Med Chem.* Aceptado (Octubre, 2014).

Capítulos de Libros:

-  Bellera C. L., Di Ianni M., Sbaraglini M. L. , Bruno-Blanch L. E., Castro E.A. Talevi A. Knowledge-based drug repurposing: a rational approach towards the identification of novel medical applications of known drugs. En: *Frontiers in Computational Chemistry* (2014). Ed. Bentham Science.

ANEXO 1



A.1. ÍNDICES TOPOLÓGICOS UTILIZADOS EN LOS MODELOS DESARROLLADOS

A.1.1. TOPOLOGÍA MOLECULAR

La topología se define como la parte del álgebra que estudia las posiciones y conexiones entre los diferentes elementos dentro de un conjunto, existen diversas disciplinas donde es empleada (redes de comunicación, circuitos eléctricos, optimización de vías de comunicación, biología, ciencias económicas, sociales, químicas, etc.). La **Topología molecular**, se basa en la aplicación de la teoría de grafos para caracterizar las estructuras químicas, analizando las posiciones e interconexiones de los átomos en una molécula. Se define un *grafo* como un conjunto de *nodos* o *vértices* donde algunos de ellos se encuentran conectados mediante *aristas* o *ejes*. Un grafo G de una estructura química está definido por los siguientes elementos (Figura A1):

-  **Vértices o nodos:** representan a los átomos de la molécula; los denotaremos $V(G)=\{v_1,v_2,\dots,v_n\}$, siendo n el número de átomos que componen la estructura.
-  **Ejes o aristas:** representan los enlaces químicos entre los átomos de la molécula, $E(G)=\{e_{12},e_{23},e_{34},\dots\}$.
-  **Circuito:** cuando el grafo contiene una estructura cíclica.
-  **Grado o valencia del vértice:** se define como el número de aristas del que v_i es extremo y se designa como δ_i . La suma de las valencias de cada uno de los vértices de que está compuesto el grafo es igual al doble del número de aristas del grafo.
-  **Longitud** de una secuencia es el número de aristas que tiene.
-  **Subgrafo:** es una porción de un grafo; entendiendo por tal, cualquier parte de un grafo constituido por una o varias de sus aristas interconectadas. Los subgrafos se clasifican según su orden m y su tipo t . El orden de un subgrafo no es más que

el número de aristas que contiene. A su vez, se clasifican en: *path*, *cluster*, *path-cluster*, *chain*.

- Tipo *path* ($t=p$): subgrafos en los cuales los grados de sus vértices son menores o iguales a 2.
- Tipo *Cluster* ($t=c$): formado por aquellos subgrafos que tienen al menos algún vértice con grado 3 o 4, pero ninguno de grado 2.
- Tipo *Path-cluster* ($t=pc$): corresponde a los subgrafos que incluyen vértices de grado 2, además de alguno con valores 3 o 4.
- Tipo *Chain* ($t=ch$): constituido por secuencias de enlaces conteniendo al menos un ciclo.

La representación del grafo de la molécula, es el primer paso para definir los índices o descriptores topológicos, que llevan información implícita acerca de los átomos, enlaces entre ellos y ensamblaje topológico. A continuación se construye la matriz topológica y a partir de esta matriz, por medio de los algoritmos adecuados, se obtienen los índices topológicos más importantes. La característica fundamental de un índice topológico es que sea invariante de grafo, es decir, su valor es independiente de la numeración de los vértices del grafo (Stankevich et al., 1988).

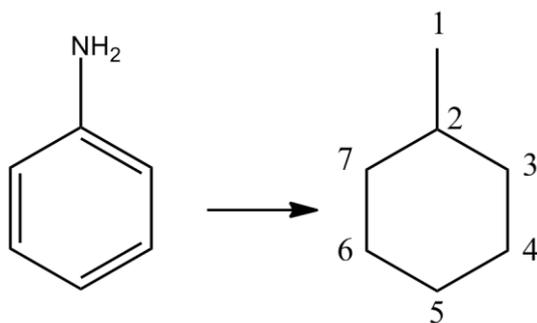


Figura A1. Grafo asociado a la estructura química de la anilina. Los índices topológicos son invariantes del grafo y, por lo tanto, es indistinto dónde se inicie la numeración del grafo.

A.1.2. MATRICES ASOCIADAS A GRAFOS MOLECULARES

Con el objetivo de transformar la información brindada del grafo molecular a una entidad numérica (índice o descriptor) es necesario realizar una representación algebraica del grafo. Existen varias operaciones como matrices, polinomios, operadores y espectros del grafo; en esta oportunidad detallaremos dos de las matrices más utilizadas, la matriz de adyacencia y de distancia.

La **matriz de adyacencia o matriz de conectividad (A)** da información sobre que pares de vértices están unidos por aristas $E(G)$. Dado una grafo G con n vértices, la matriz de adyacencia $A=A(G)$ es la matriz cuadrada simétrica de n filas por n columnas, cuyos elementos $[A]_{ij}$ se definen como:

$$[A]_{ij} = \begin{cases} 1 & \text{si } i \neq j \text{ y } (i,j) \in E(G) \\ 0 & \text{si } i = j \text{ ó } (i,j) \notin E(G) \end{cases}$$

En la Figura A2 se ilustra a modo de ejemplo el grafo y la matriz de adyacencia para el isopentano y el ciclohexano.

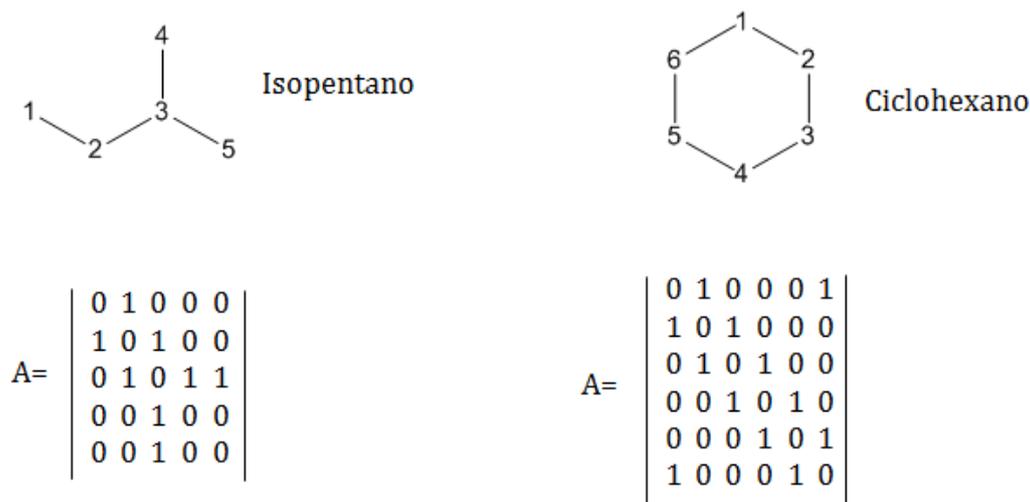


Figura A2. Representación de isopentano y ciclohexano como grafos y matrices de adyacencia.

La **matriz de distancia** $D(G)$ está formada por los elementos d_{ij} (distancias), que representan al número de aristas que unen los vértices i y j por el camino más corto, los elementos de la matriz $D(G)$ quedan definidos por:

$$[D]_{ij} = \begin{cases} d_{ij} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

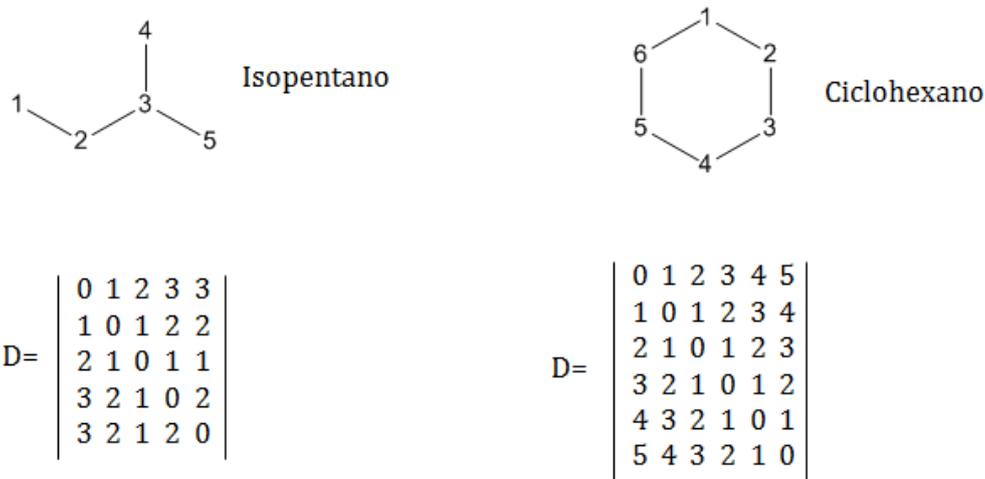


Figura A3. Matrices de distancia asociadas a los grafos de isopentano y ciclohexano.

Estas dos representaciones algebraicas del grafo sólo tienen en cuenta la conectividad entre los átomos, es decir no se considera la naturaleza de los mismos o de los enlaces covalentes que los unen. Significa, que ninguno de los índices derivados de estas simples matrices podría diferenciar por ejemplo a la anilina del fenol, ya que ambos compuestos poseen grafos isomorfos.

De manera que es necesario ponderar de distinta manera los vértices y aristas del grafo según las propiedades químicas de los átomos y enlaces involucrados (Ivanciuc et al., 1999).

Un grafo de vértices y aristas pesadas entonces consistirá en un conjunto de símbolos químicos que permitan diferenciar cada tipo de átomo, un set de órdenes topológicos de las aristas Bo_{ij} , un conjunto de vértices pesados $Vw(w)$ y un conjunto de ejes pesados

$Ew(w)$. Generalmente B_0 es 1 si el enlace entre i y j es simple; 1.5 cuando el enlace es aromático; 2 cuando es un doble enlace y 3 cuando se trata de un triple enlace (Ivanciuc, 2000).

A partir de grafos de vértices y aristas pesadas; quedan definidos los elementos de la matriz de adyacencia $A(w)$ y de la matriz de distancia $D(w)$:

$$[A(w)]_{ij} = \begin{cases} Vw(w)_i & \text{si } i = j \\ Ew(w)_{ij} & \text{si } i \neq j \text{ y } e_i \in E(G) \\ 0 & \text{si } e_i \notin E(G) \end{cases}$$

$$[D(w)]_{ij} = \begin{cases} Vw(w)_i & \text{si } i = j \\ d(w)_{ij} & \text{si } i \neq j \end{cases}$$

donde $V(w)_i$ representa el peso de los vértices v_i , $Ew(w)_{ij}$ es el peso de los enlaces e_{ij} , y w es el esquema pesada.

Los vértices también pueden ser ponderados según diferentes propiedades atómicas P (por ejemplo, el número atómico Z , la masa atómica A , la electronegatividad relativa X , el radio atómico r (Å), entre otros. El siguiente algoritmo general se aplica para obtener los $Vw(P)_i$:

$$Vw(P)_i = 1 - P_c/P_i$$

donde P_c es la propiedad atómica del átomo de C y P_i es la propiedad atómica del átomo i . Los ejes o aristas por su parte son ponderados de la siguiente manera:

$$Ew(P)_i = P_c P_i / B_0 P_i P_j$$

Como se menciono anteriormente a partir de las matrices pueden obtenerse polinomios y espectros del grafo, que serán de utilidad para el cálculo de diferentes índices o descriptores.

El polinomio característico de la matriz de adyacencia derivada del grafo $Ch(G,x)$ es:

$$Ch(G, x) = \det(xI - M) = \sum_{n=0}^N c_n x^{N-n}$$

donde I representa a la matriz unidad y M a una matriz simétrica, por ejemplo la de adyacencia o distancia. El determinante (det) de una matriz se calcula como:

$$\det(M) = \sum_{i=1}^{N!} (-1)^{p_i} P_i M_{11} M_{22} \dots M_{NN}$$

donde P_i es un operador que permuta los índices de las columnas de los elementos de la matriz; la sumatoria comprende $N!$ permutaciones de los índices de columna y p_i indica el número de trasposiciones necesarias para restaurar el orden original.

Los autovalores de matriz serán las raíces del polinomio característico de un grafo y el conjunto de los autovalores definen el espectro del grafo, los autovalores de las diferentes matrices son utilizados como índices topológicos. (Consonni and Todeschini, 2012; Szabo and Ostlund, 1996).

A continuación se detallan las definiciones de los índices topológicos incluidos en los modelos desarrollados en el trabajo de tesis.

A.1.3. ÍNDICES DE CONECTIVIDAD SIMPLE

Los índices de conectividad simple fueron introducidos por Randic (Randic, 1975) y luego desarrollados por Kier y Hall (Kier, 1985; Kier and Hall, 1981), son los descriptores más empleados en estudios QSAR y cuantifican la estructura molecular al incorporar fragmentos de los subgrafos, se representan por ${}^m x_t$ donde t representa el tipo y m el orden del subgrafo, están definidos definidos por:

$${}^m x_t = \sum_{j=1}^{n_t^m} S_j^m$$

n_t^m es el número de subgrafos de tipo t y de orden m . Los términos S_j^m se definen como el inverso de la raíz cuadrada del producto de las valencias de los vértices que conforman el subgrafo, donde j define a un subgrafo en particular y el número de valencias δ_i depende del tipo de subgrafo:

$$S_j^m = \left[\prod_{i=1}^{m+1} \delta_i \right]^{-1/2}$$

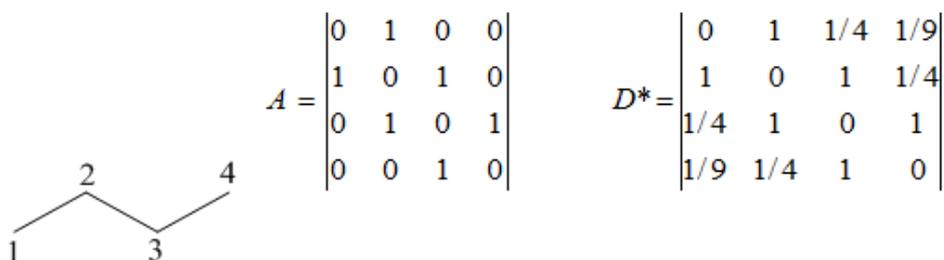
En el modelo 3 se incorporó un índice de conectividad simple de orden 4 y tipo path (X4P, nomenclatura de DRAGON), es decir se tuvieron en cuenta solo aquellos subgrafos en los que las valencias de sus vértices son menores o iguales a 2 sin formar un ciclo.

A.1.4. ÍNDICES DE CARGA

Los **índices de carga simple** evalúan la transferencia de carga entre pares de átomos así como la transferencia global de carga intramolecular (Galvez et al., 1994). Puesto que la influencia de la carga decrece con el cuadrado de la distancia se utiliza como el primer descriptor de carga la “matriz de la inversa de la distancia al cuadrado” (*inverse square distance matrix*), denotada como D^* . Para evitar dividir por cero se conserva la diagonal de la matriz distancia.

M es la matriz obtenida a partir del producto entre la matriz adyacencia A y la matriz D^* . El término de carga CT se define como $m_{ij} - m_{ji}$, donde las m simbolizan los elementos de la matriz M (para $i \neq j$); para $i=j$ CT es igual a la valencia del vértice para $i = j$.

Para ilustrar el cálculo de CT consideremos el caso del n-butano.



$$M = A \times D^* = \begin{vmatrix} 1 & 0 & 1 & 1/4 \\ 1/4 & 2 & 1/4 & 10/9 \\ 10/9 & 1/4 & 2 & 1/4 \\ 1/4 & 1 & 0 & 1 \end{vmatrix}$$

Entonces,

$$CT_{14} = CT_{23} = 1/4 - 1/4 = 0$$

$$CT_{12} = -CT_{34} = 0 - 1/4 = 1/4$$

$$CT_{13} = -CT_{24} = 1 - 10/9 = -1/9$$

En este punto puede subrayarse:

- La diagonal de M representa la valencia topológica del vértice correspondiente.
- Para i distinto de j el término CT_{ij} es una medida de la carga neta transferida **desde el átomo j al i**. Si el término es negativo es i el que está transfiriéndole carga a j.

Luego, se define el índice de carga **G** de orden m , como la suma en valor absoluto de los términos c_{ij} definidos para los vértices ij situados a una distancia d_{ij} igual a m .

$$G^m = \sum_{i=1}^{N-1} \sum_{j=i+1}^N |CT_{ij}| \delta(m, d_{ij})$$

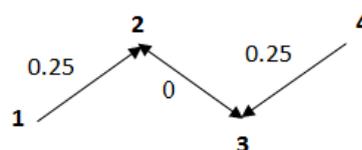
Donde: m es el orden del índice, es decir un valor prefijado de distancia entre los vértices i y j ; N es el número de vértices en el grafo; d_{ij} es la distancia entre los vértices i y j ; δ es el delta de Kronecker, $\delta(\alpha, \beta) = 0$ si $\alpha \neq \beta$ y $\delta(\alpha, \beta) = 1$ si $\alpha = \beta$.

En el caso del n-butano los pares de vértices situados a una distancia $m = 1$ son tres: (v_1, v_2) , (v_2, v_3) y (v_3, v_4) ; los pares situados a distancia $m = 2$ son dos: (v_1, v_3) y (v_2, v_4) . Por último, a distancia $m = 3$ sólo hay un par de vértices: (v_1, v_4) , por lo tanto:

$$G^1 = |CT_{12} + CT_{23} + CT_{34}| = 0,25 + 0 + 0,25 = 0,5$$

$$G^2 = |CT_{13} + CT_{24}| = 0,11 + 0,11 = 0,22$$

$$G^3 = |CT_{14}| = 0$$



Los **índices de carga de valencia** G_k^V se calculan siguiendo el procedimiento anterior usando una matriz modificada A' en cuya diagonal se introduce la electronegatividad de los heteroátomos.

En el modelo 1 se utilizó un índice de carga simple de orden 7 (GGI7, nomenclatura del programa Dragon) y en el modelo 3 un índice de carga de valencia de orden 4 (G4V, nomenclatura del programa Desmol11).

A.1.5. ÍNDICES DERIVADOS DE LA MATRIZ DE ADYACENCIA

$VE1_X$ representa la suma de los coeficientes del autovector asociado con el último autovalor (el más negativo) de la matriz chi. Tal matriz es una matriz de adyacencia modificada, obtenida al ponderar cada enlace entre dos vértices por la conectividad de los vértices (Randic, 1992). Los elementos de la matriz chi se definen como sigue:

$$X_{i,j} = \begin{cases} \frac{1}{\sqrt{m,n}} & \text{si } i, j \text{ son adyacentes} \\ 0 & \text{caso contrario} \end{cases}$$

siendo m y n las valencias de los vértices involucrados en el enlace químico considerado.

$VE1_X$ está definido por:

$$VE1_X = \sum_{i=1}^{nSK} |l_i|$$

donde nSK indica el número de átomos distinto de hidrogeno y l_i indica el coeficiente de orden i del último vector propio de la matriz Chi (X) y la suma se extiende sobre todos los coeficientes de vector propio que será igual al número de vértices del grafo (Balaban et al., 1991; Consonni and Todeschini, 2012).

$SpPosA_A$ es el índice de suma espectral positiva y normalizada de la matriz de adyacencia, se deriva a partir de la función de los valores propios positivos de la matriz de adyacencia y se aplica un factor de normalización para reducir la influencia de tamaño de la molécula (Balaban et al., 1991; Consonni and Todeschini, 2012):

$$SpPosA_A = \frac{1}{nSK} \sum_{i=1}^{n^+} (\lambda_i^+)$$

donde nSK es el número total de vértices del grafo; λ_i^+ representa los valores propios positivos de la matriz de adyacencia y n^+ indica que la sumatoria solo es de dicho valores positivos.

REFERENCIAS ANEXO 1

Balaban A.T., Ciubotariu D., Medeleanu M. Topological indices and real number vertex invariants based on graph eigenvalues or eigenvectors. *J. Chem. Inf. Comput. Sci.* **31**, 517-523. (1991).

Consonni V., Todeschini R. Multivariate Analysis of Molecular Descriptors in: M. Dehmer, et al. (Eds.), *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. pp. 111-147. (2012)

Galvez J., Garcia-Domenech R., Salabert M.T., Soler R. Charge Indexes. New Topological Descriptors. *J Chem Inf Comput Sci.* **34**, 520-525. (1994).

Ivanciuc O. QSAR Comparative study of Wiener descriptors for weighted molecular graphs. *J. Chem. Inf. Comput. Sci.* **40**, 1412-1422. (2000).

Ivanciuc O., Ivanciuc T., Balaban T.A. Vertex- and edge-weighted molecular graphs and derived structural descriptors. , *Topological indices and related descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, Amsterdam. pp. 169-220. (1999)

Kier L.B. A shape index from molecular graphs. *Quant Struct-Act Relat.* **4**, 109-116. (1985).

Kier L.B., Hall L.H. Derivation and significance of valence molecular connectivity *J Pharm Sci.* **70**, 583-589. (1981).

Randic M. On characterization of molecular branching *J Am Chem Soc.* **97**, 6609-6615. (1975).

Randic M. Similarity based on extended basis descriptors. *J Chem Inf Comput Sci.* **32**, 686-692. (1992).

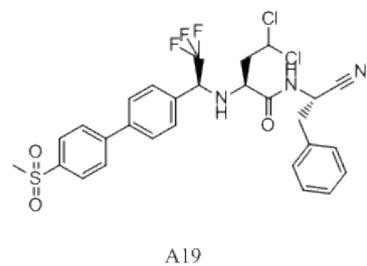
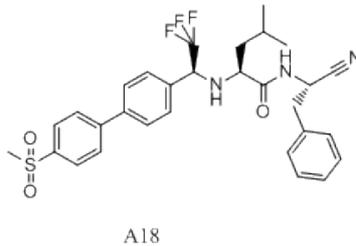
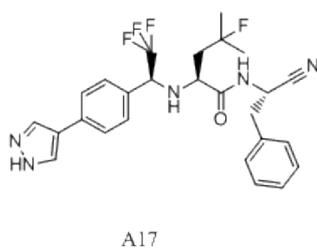
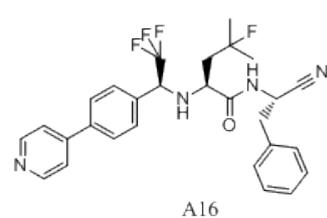
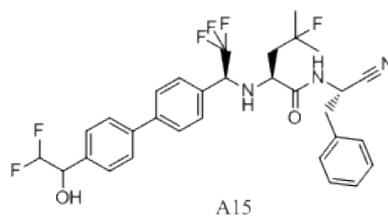
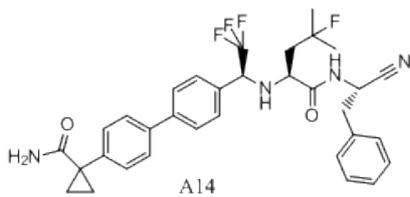
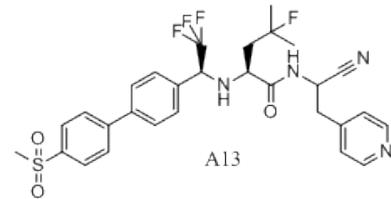
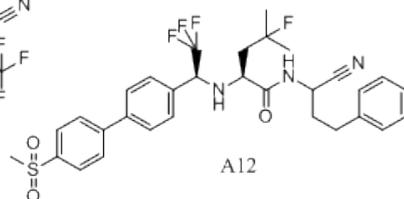
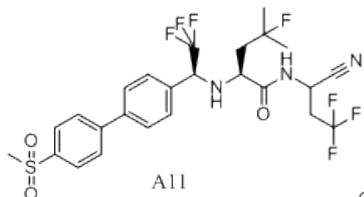
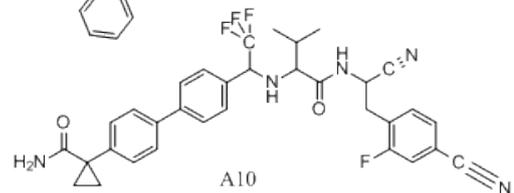
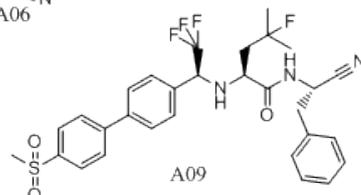
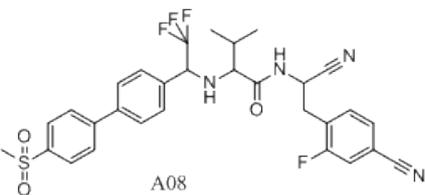
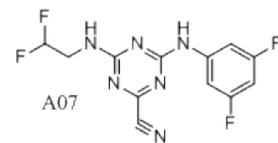
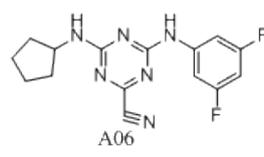
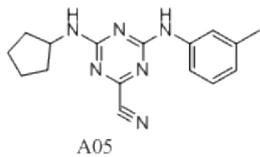
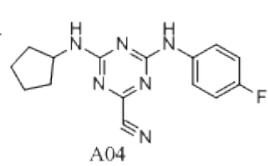
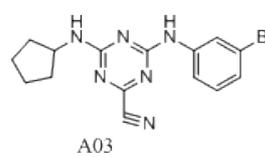
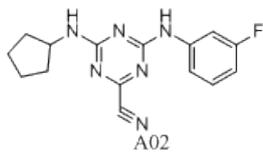
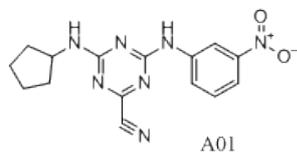
Stankevich M.I., Stankevich I.V., Zefirov N.S. Topological Indices in Organic Chemistry. *Russ. Chem. Rev.* **57**, 191-208. (1988).

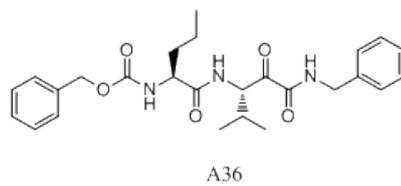
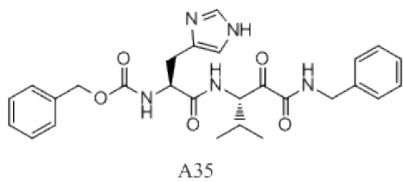
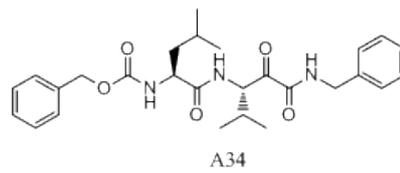
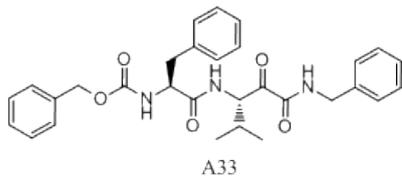
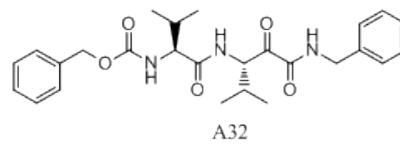
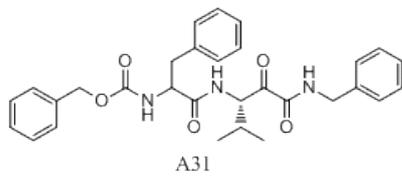
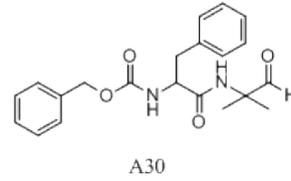
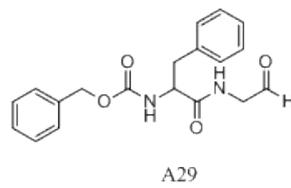
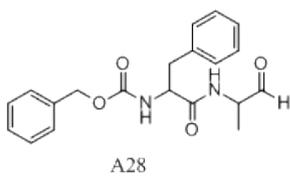
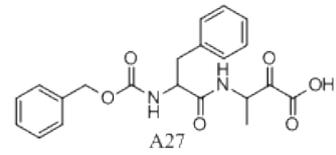
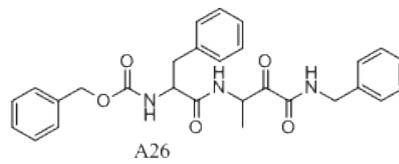
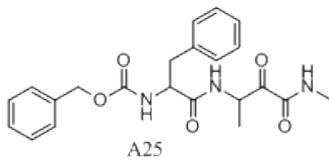
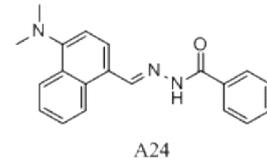
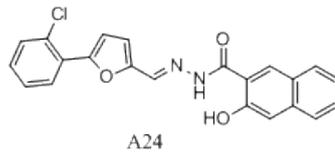
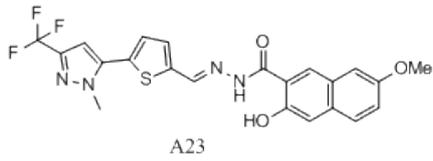
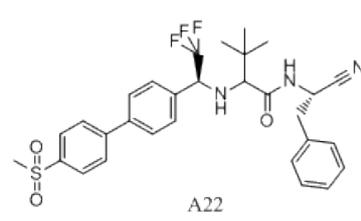
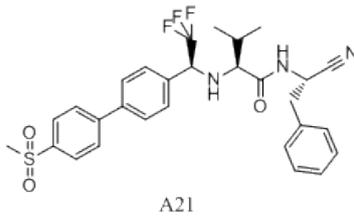
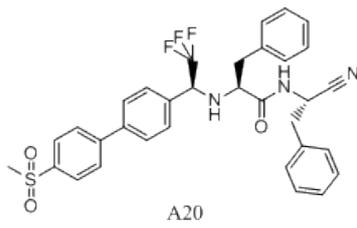
Szabo A., Ostlund N.S. Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory, Mathematical Review, Dover Publications, New York. pp. 2-38. (1996).

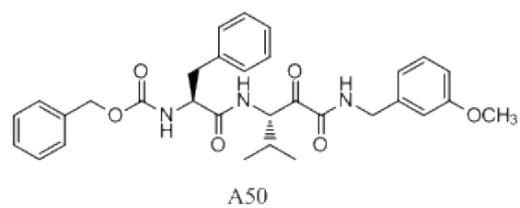
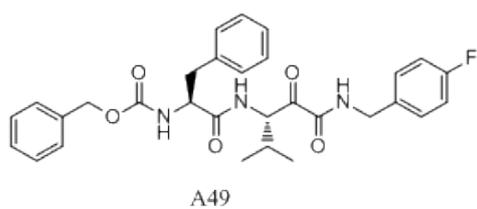
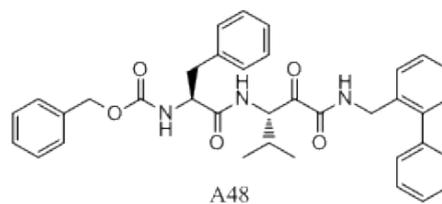
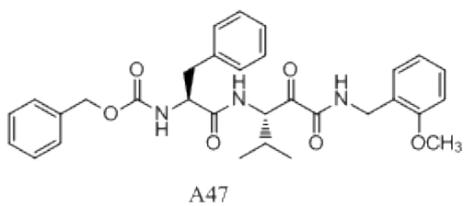
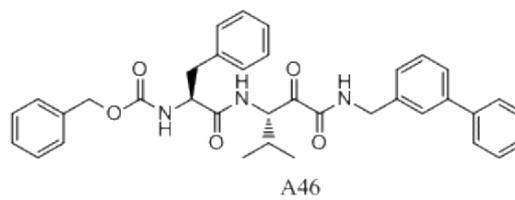
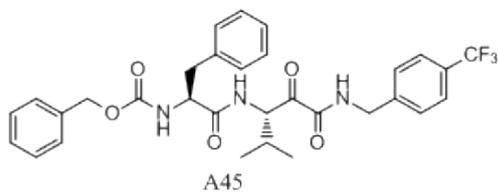
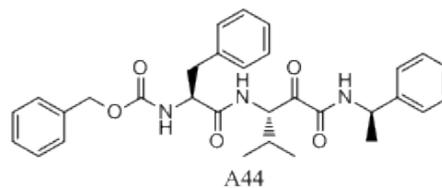
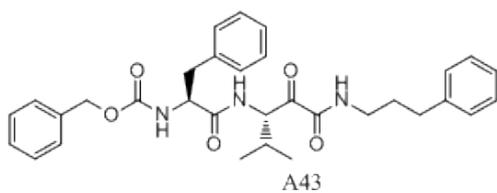
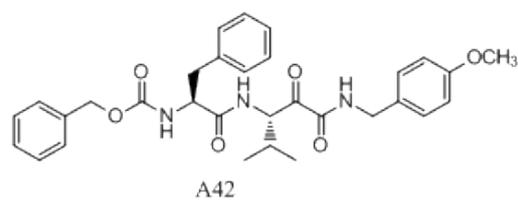
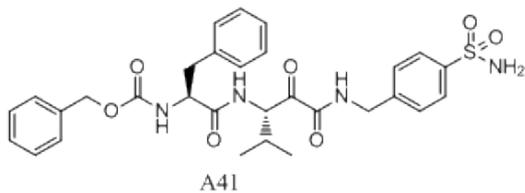
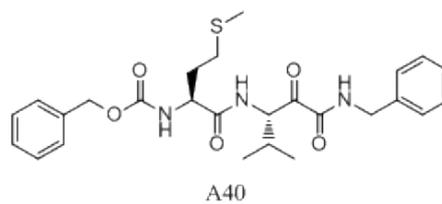
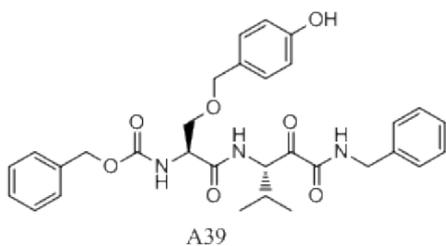
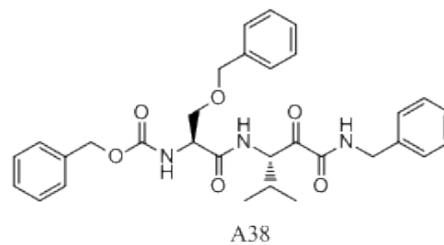
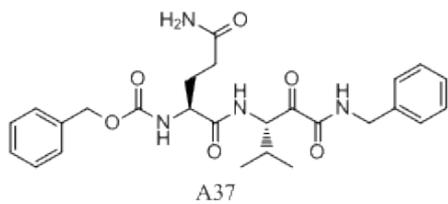
ANEXO 2

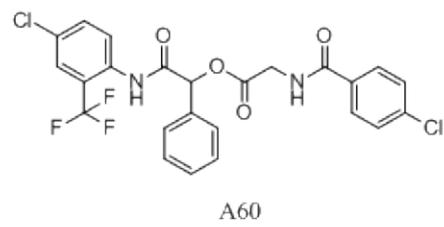
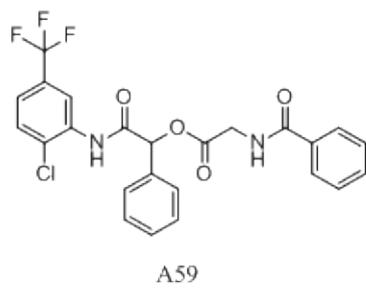
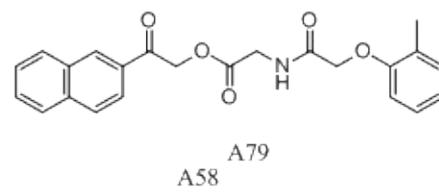
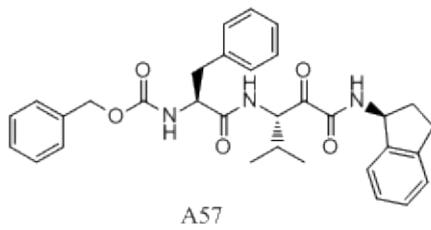
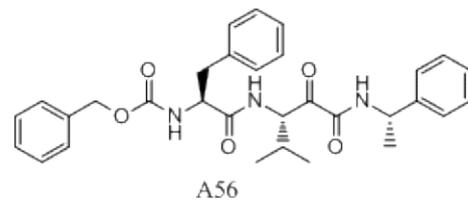
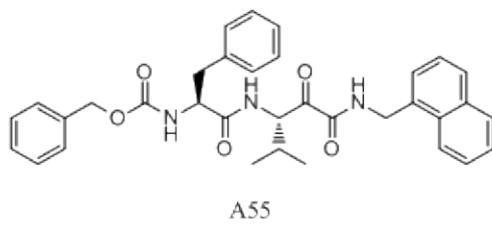
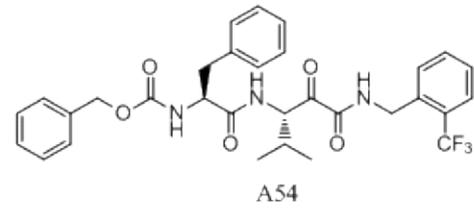
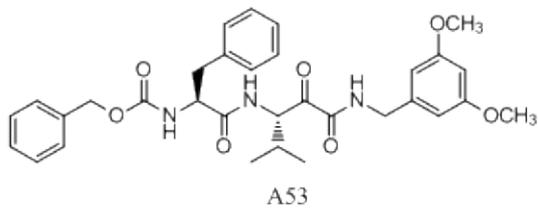
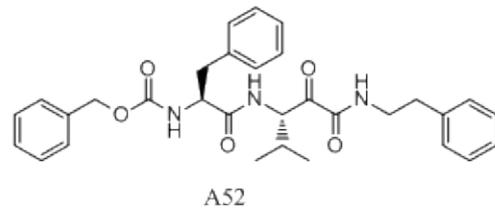
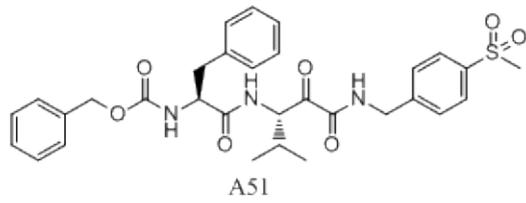


A.2.1. ESTRUCTURAS DE LOS COMPUESTOS ACTIVOS DEL CONJUNTO DE ENTRENAMIENTO

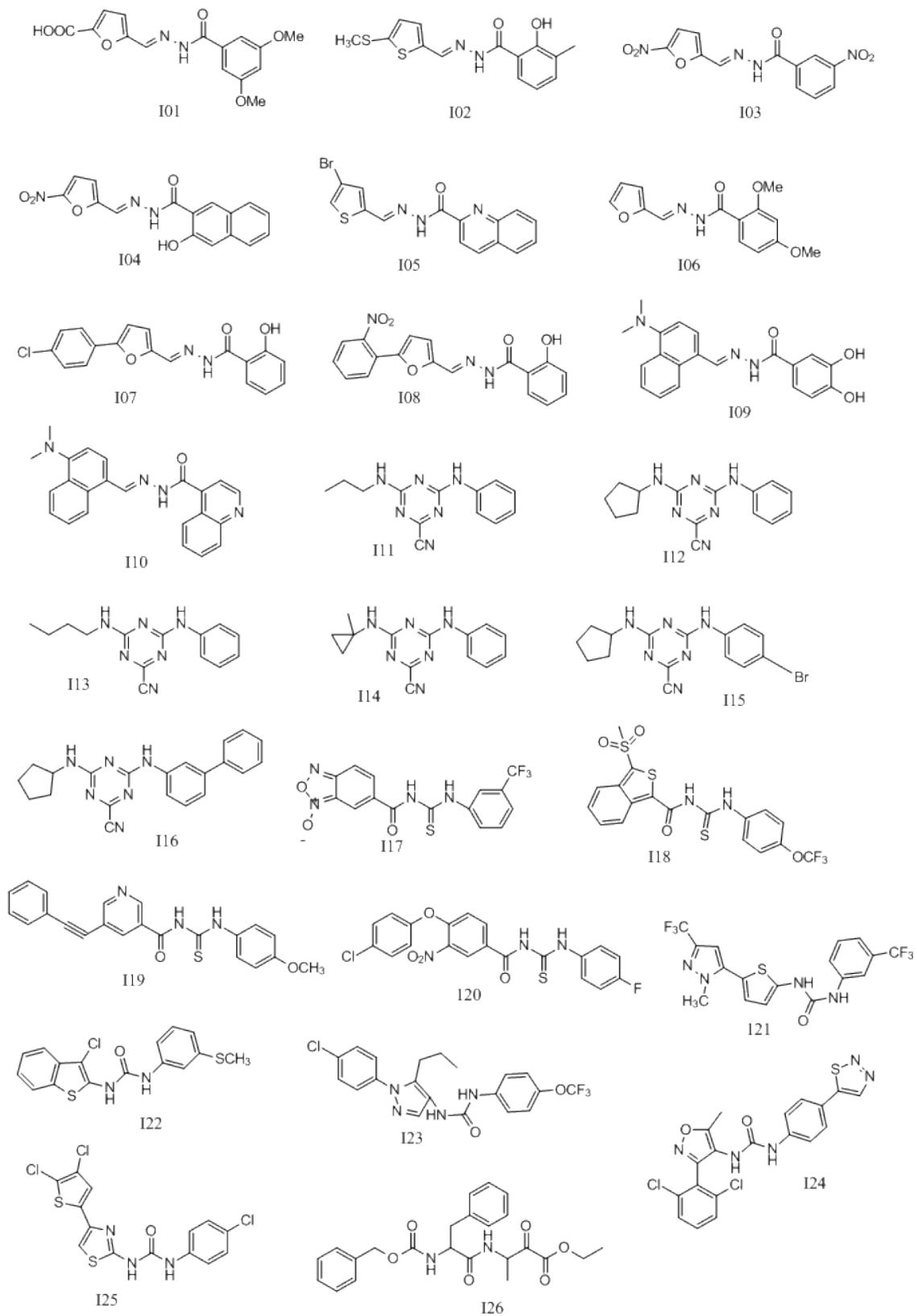


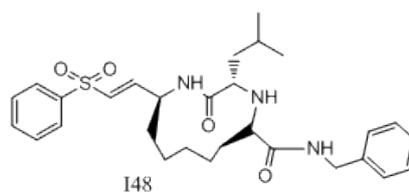
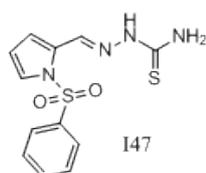
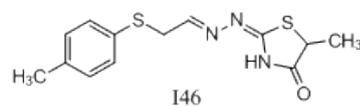
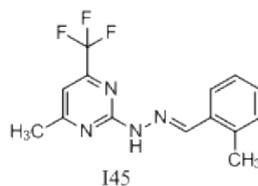
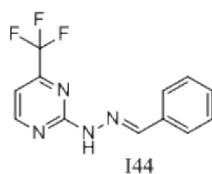
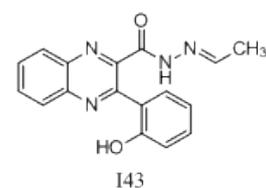
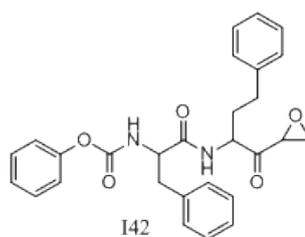
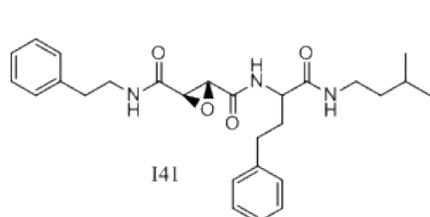
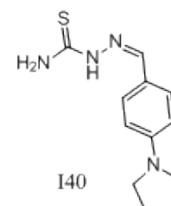
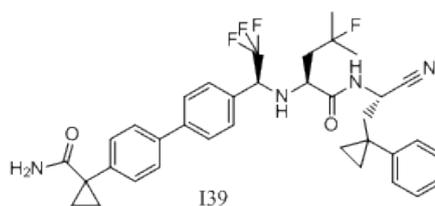
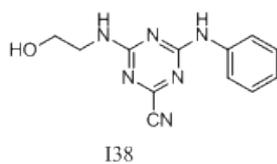
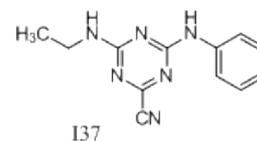
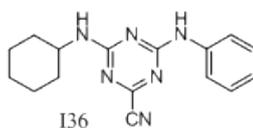
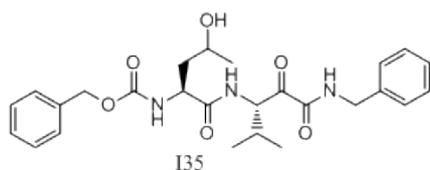
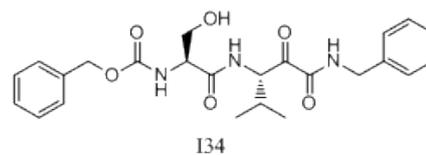
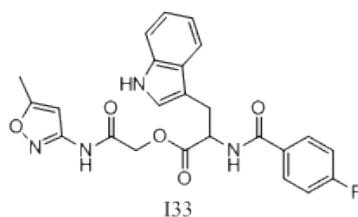
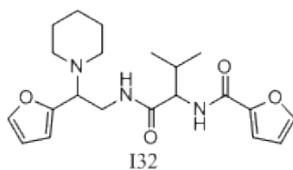
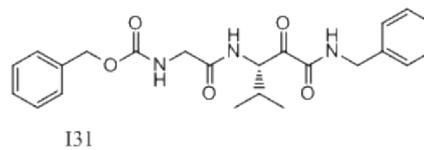
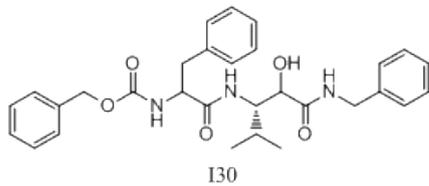
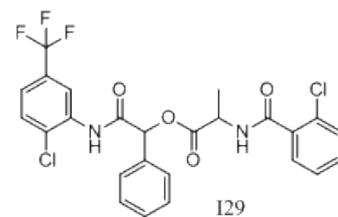
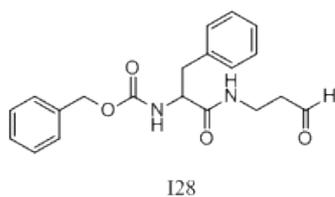
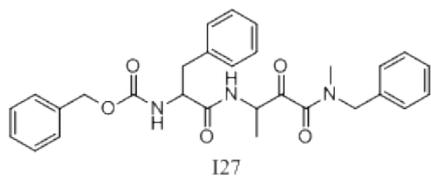


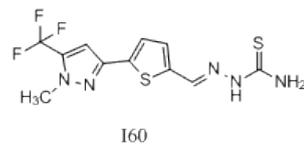
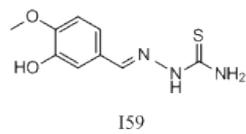
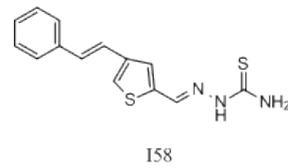
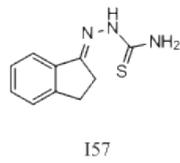
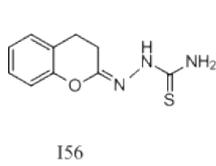
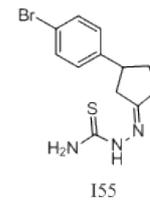
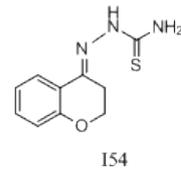
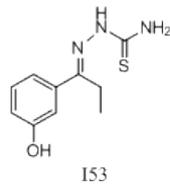
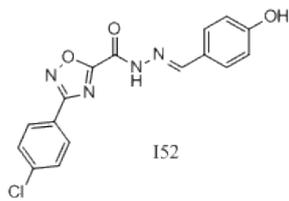
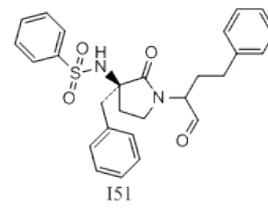
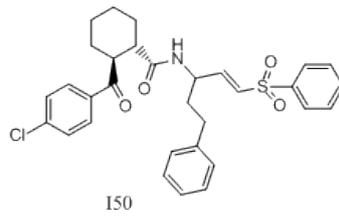
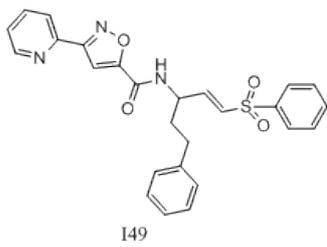




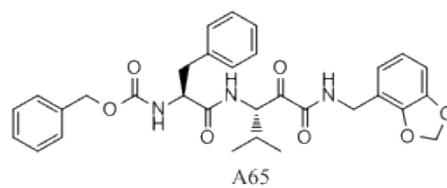
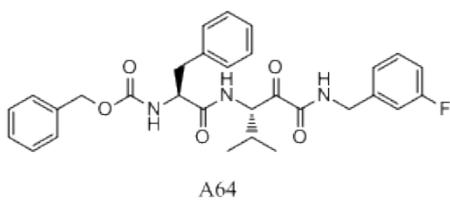
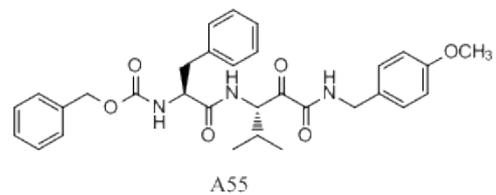
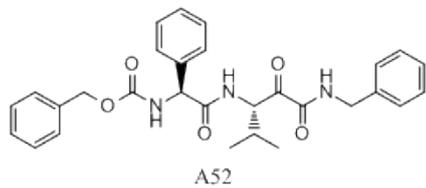
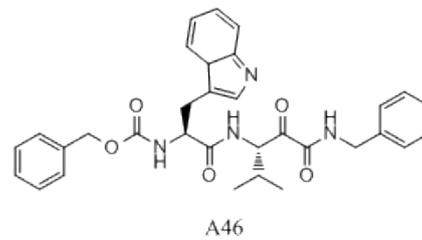
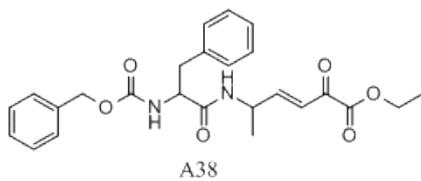
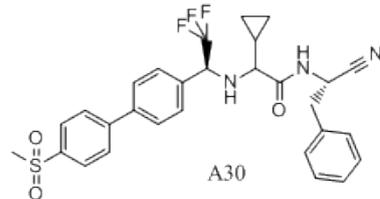
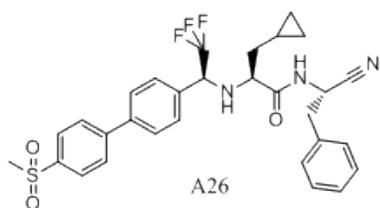
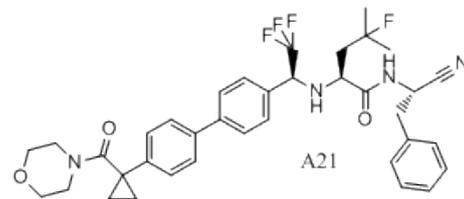
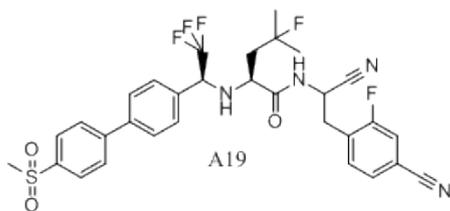
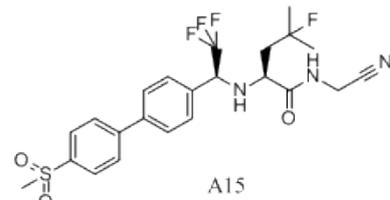
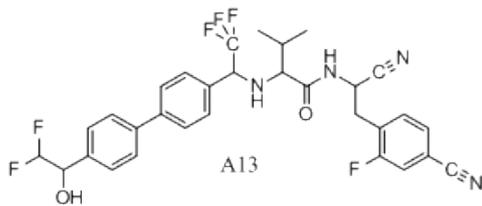
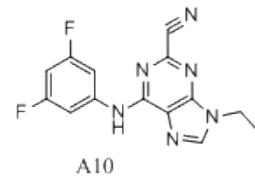
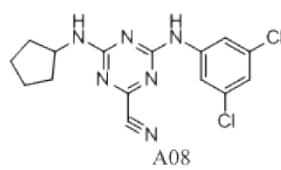
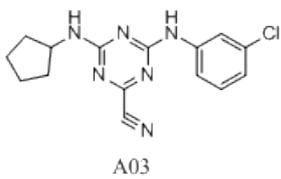
A.2.2. ESTRUCTURAS DE LOS COMPUESTOS INACTIVOS DEL CONJUNTO DE ENTRENAMIENTO

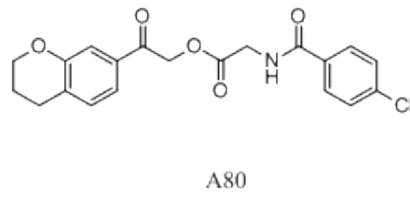
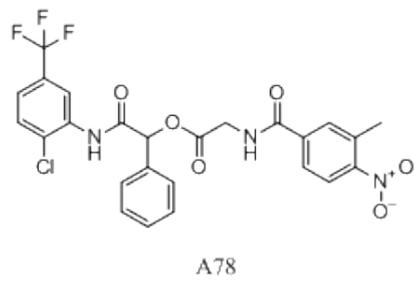
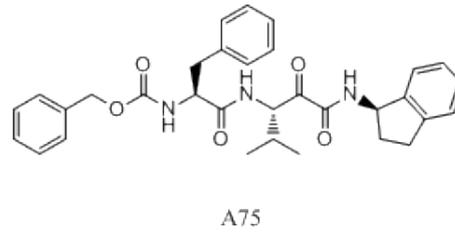
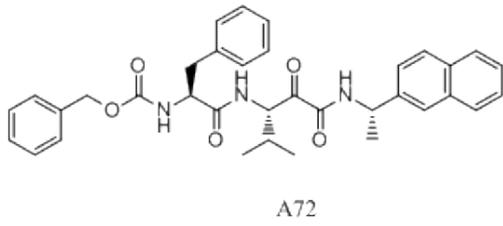
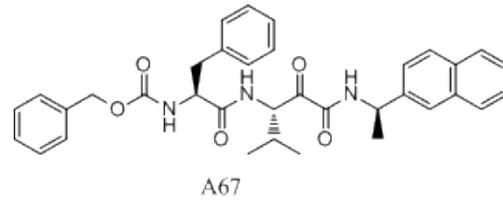
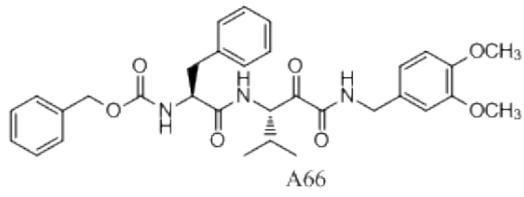




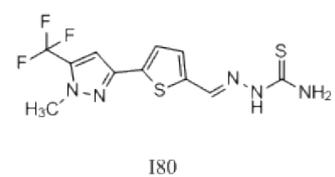
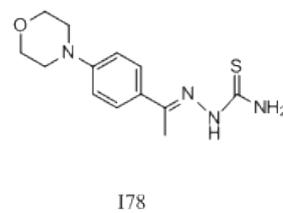
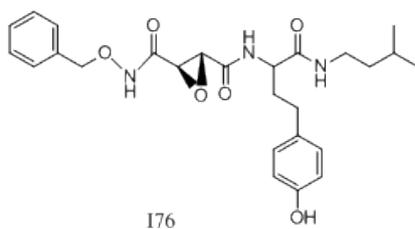
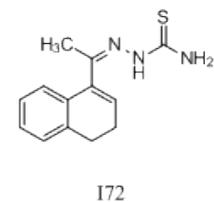
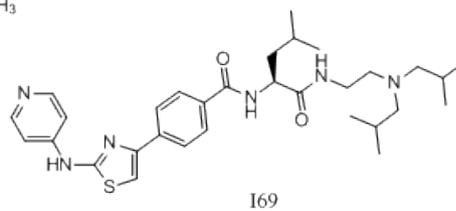
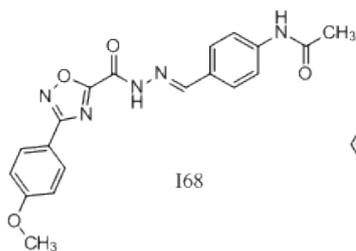
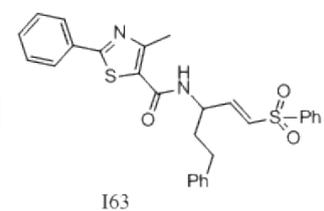
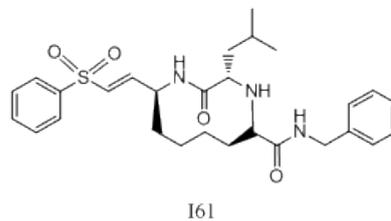
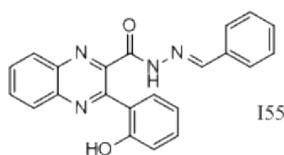
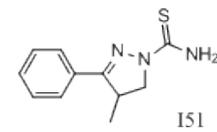
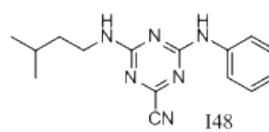
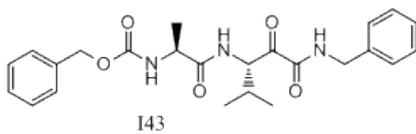
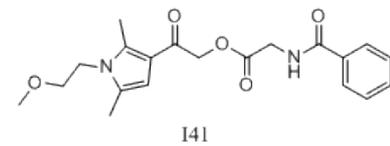
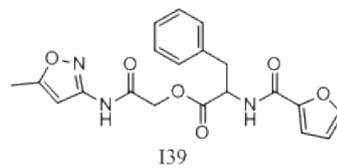
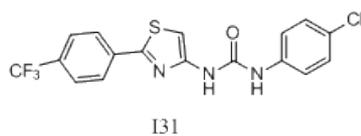
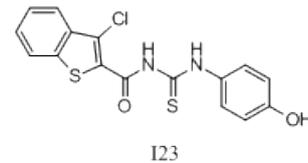
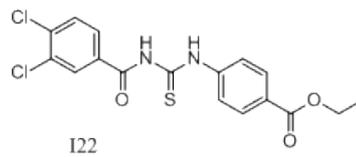
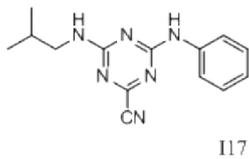
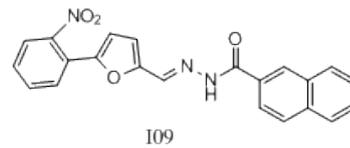
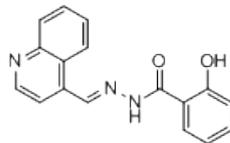
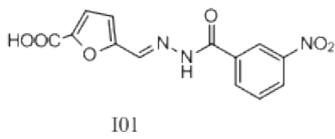


A.2.3. ESTRUCTURAS DE LOS COMPUESTOS ACTIVOS DEL CONJUNTO DE PRUEBA





A.2.4. ESTRUCTURAS DE LOS COMPUESTOS INACTIVOS DEL CONJUNTO DE PRUEBA



Acta 1584

En el día de la fecha se reunió el jurado designado por el HCD de la Facultad de Ciencias Exactas de la UNLP en el expediente 0700-004404/09-008 constituido por los Dres Juan Bautista Rodríguez, Guillermo Lebedie y Augusto Pich Otero para evaluar el trabajo de tesis doctoral realizado por la Lic. Caroline Bellera titulado "Búsqueda racional de nuevos fármacos antichepáticos inhibidores de cruzipaina" dirigido por el Dr. Luis E. Bruno-Blenda y co-dirigido por el Dr. Alan Televi para optar al grado académico de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata.

El trabajo de tesis abordó un enfoque de diseño racional de fármacos basado en el cribado virtual y el reposicionamiento de fármacos para el tratamiento de la enfermedad de Chagas.

Ambos temas son de mucha relevancia en la actualidad permitiendo ahorrar tiempos y costos para la selección y evaluación de potenciales candidatos terapéuticos.

Los resultados obtenidos han permitido encontrar potenciales candidatos para el tratamiento de la enfermedad de Chagas.

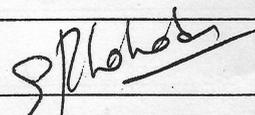
El manuscrito presentado ha logrado describir de manera clara y concisa el trabajo

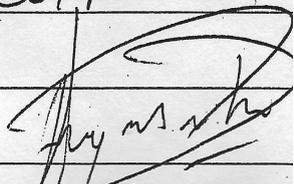
y pertinente. La presentación oral fue bien estructurada logrando resumir los principales logros alcanzados en el desarrollo de esta tesis doctoral y dejar bien planteadas las perspectivas futuras.

En base a lo anteriormente expuesto, el jurado resuelve por unanimidad aprobar la tesis con una calificación de sobresaliente 10 (diez).

La Plata, 30 de Octubre de 2014


Juan B. Rodriguez


Guillermo R. L. Lopez


Augusto P. de O.

COPIA FIEL

