



**UNIVERSIDAD MICHOACANA  
DE SAN NICOLÁS DE  
HIDALGO**



**INSTITUTO DE INVESTIGACIONES  
QUÍMICO BIOLÓGICAS**

**INSTITUTO DE INVESTIGACIONES  
QUÍMICO- BIOLÓGICAS**

**“DISEÑO POR QSAR DE COMPUESTOS POTENCIALMENTE  
ACTIVOS CONTRA *Leishmania chagasi*”.**

Tesis que para obtener el grado de

Maestra en Ciencias Químicas

Presenta:

***QFB. ROSALÍA MANRIQUE SOTO.***

Director de tesis:

**DC. LUIS CHACÓN GARCÍA.**

Morelia Mich., Diciembre, 2012.

El presente trabajo de investigación se llevó a cabo en el laboratorio de Diseño Molecular de Fármacos del Instituto de Investigaciones Químico Biológicas, de la Universidad Michoacana de San Nicolás de Hidalgo, bajo la asesoría del D.C. Luis Chacón García.

*“La mente que se abre a una nueva idea,  
jamás volverá a su tamaño original”.*

*Albert Einstein.*

*Para mi familia con amor,  
Mis Padres y Hermanos.*

## A G R A D E C I M I E N T O S .

Agradezco de corazón a todas las personas que de una u otra forma han contribuido en la realización de este proyecto.

Con especial cariño y respeto para el *D.C Luis Chacón García*, por la amistad, la confianza, el apoyo y dedicación que me ha otorgado como profesor y amigo.

Al equipo de QSAR del laboratorio de Diseño Molecular de Fármacos, sin ellos realmente este trabajo no se hubiera realizado:

*QFB. Ana Lilia Ocampo Néstor, QFB. Anai Zavala Franco, QFB. Jonahtan Lira Chávez.*

A tod@s l@s compañer@s del Instituto de Investigaciones, por su apoyo y amistad, especialmente para las químicas: *QFB. Melissa Tapia Juárez, QFB. Juanita Suárez García y MC. Fany Elisa Álvarez Esquivel*, por darme la oportunidad de conocerlas, crecer y aprender, pero sobre todo por la gratificación de tenerlas como amigas toda la vida.

A mis compañeros de la primera generación de la Maestría en Ciencias Químicas, *QFB. Gerardo Morán López y QFB. Ramón Guzmán Mejía.*

A los investigadores: *DC. Claudia Contreras Celedón, DC. Erck Cuevas Yáñez, DC. Janett Betzabé González Campos y DC. Mario Armando Gómez Hurado.* Por el apoyo brindado en la revisión y análisis de este proyecto.

Parte de este trabajo fue presentado en:

**“QSAR de tetrahydro-1-benzoazepinas activas contra *Leishmania chagasi*”.**

Manrique Soto Rosalía<sup>1</sup>, Palma Rodríguez Arilio<sup>2</sup>, Gómez Ayala Sandra L.<sup>1, 2</sup>,  
Chacón García Luis. <sup>\*1</sup>

\*Laboratorio de Diseño Molecular. EdificioB-1. Instituto de Investigaciones Químico  
Biológicas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán,  
México.

Laboratorio de Síntesis Orgánica, Universidad Industrial de Santander,  
Bucaramanga, Colombia.

En la 8va. Reunión de la Academia de Química Orgánica. Benemérita Universidad  
Autónoma de Puebla, México. Facultad de Ciencias Químicas. Marzo 2012.

## ÍNDICE GENERAL.

ÍNDICE DE FIGURAS.....	II
ÍNDICE DE TABLAS.....	III
ÍNDICE DE GRÁFICAS.....	V
ABREVIATURAS.....	VI
RESUMEN.....	VIII
SUMMARY.....	X
<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
<b>2. ANTECEDENTES.....</b>	<b>4</b>
2.1 LEISHMANIASIS.....	4
2.2 TRATAMIENTO.....	7
2.3 TÉCNICA QSAR EN EL DESARROLLO DE NUEVOS FARMACOS.....	12
2.4 SÍNTESIS DE DERIVADOS ANTIPARASITARIOS CONTRA <i>LEISHMANIA CHAGASI</i> .....	16
<b>3. OBJETIVOS.....</b>	<b>20</b>
3.1 OBJETIVO GENERAL.....	20
3.2 OBJETIVOS PARTICULARES.....	20
<b>4. ESTRATEGIA EXPERIMENTAL GENERAL.....</b>	<b>21</b>
<b>5. METODOLOGÍA.....</b>	<b>22</b>
5.1 DESCRIPCIÓN METODOLOGICA.....	22
<b>6. RESULTADOS Y DISCUSIÓN.....</b>	<b>32</b>
6.1 CARACTERIZACIÓN DE LOS MODELOS SELECCIONADOS POR ESTUDIO QSAR, PARA ACTIVIDAD BIOLÓGICA CONTRA PROMASTIGOTES DE <i>L. CHAGASI</i> .....	36
I. Análisis del Modelo 2:.....	36
I.I Descriptores correspondientes al Modelo 2.....	40
II. Análisis del Modelo 3:.....	43
II.I Descriptores correspondientes al Modelo 3.....	46
III. Modelo 10:.....	47
III.I Descriptores correspondientes al Modelo 10.....	50
6.2 PREDICCIÓN DE LAS PROPUESTAS DE LOS DERIVADOS DE LA BENZOAZEPINA.....	58
6.3 ESTUDIO QSAR DE LA CITOTOXICIDAD.....	60
IV. Modelo 8:.....	64
IV.I Descriptores correspondientes al modelo 8.....	66
V. Modelo 12:.....	70
V.I Descriptores correspondientes al Modelo 12.....	73
<b>7. CONCLUSIONES.....</b>	<b>78</b>
<b>8. PERSPECTIVAS.....</b>	<b>79</b>
<b>9. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>80</b>

## ÍNDICE DE FIGURAS.

<b>Figura 1.</b> Distribución genográfica en México de los principales agentes causales de <i>Leishmania</i> .....	4
<b>Figura 2.</b> Ciclo biológico de <i>Leishmania sp.</i> .....	5
<b>Figura 3.</b> Anitmoniato de meglumina (Glicantime®).....	8
<b>Figura 4.</b> Estibogluconato de sodio (Penstostam ®).....	9
<b>Figura 5.</b> Enfotericina B.....	10
<b>Figura 6.</b> Isoetionato de pentamidina.....	10
<b>Figura 7.</b> Metodología de investigación QSAR.....	14
<b>Figura 8.</b> Estructura general de las tetrahidronafto (A, B) azepinas, (C, D) tetrahidro-1-benzoazepinas, derivados con actividad biológica contra <i>Leishmania chagasi</i> .....	17
<b>Figura 9.</b> Síntesis del <i>cis</i> -2-aril-4-hidróxitetrahydro-1-benzoazepina.....	18
<b>Figura 10.</b> Estructura de la cabeza de serie para QSAR.....	21
<b>Figura 11.</b> Diagrama general de la Metodología QSAR.....	22
<b>Figura 12.</b> Propuesta de compuestos potencialmente activos contra <i>Leishmania chagasi</i> .....	76

## ÍNDICE DE TABLAS.

<b>Tabla 1.</b> Características generales de los descriptores Dragon® 5.3.	16
<b>Tabla 2.</b> Actividad antiparasitaria <i>in vitro</i> y citotoxicidad del tetrahydro-1-benzoazepina-ol, agosnista para <i>Leishmania chagasi</i> .	23
<b>Tabla 3.</b> Serie de exploración de los derivados de la <i>cis</i> -2-aril-hidroxitetrahydro-1-benzoazepina, con actividad biológica ( ).	25
<b>Tabla 4.</b> Familia de descriptores calculados con el software Dragon 5.3 ®	26
<b>Tabla 5.</b> Parámetros usados por Moby Digs 1.0 ®, por QSAR con AGs.	27
<b>Tabla 6.</b> Los 50 modelos con sus descriptores y parámetros estadísticos.	32
<b>Tabla 7.</b> Matriz de correlación total de descriptores.	34
<b>Tabla 8.</b> Comparación entre la IC <sub>50</sub> experimental y la IC <sub>50</sub> calculada en el Modelo 2.	36
<b>Tabla 9.</b> Coeficiente de correlación Modelo 2.	37
<b>Tabla 10.</b> Fragmentación del Modelo 2 para promastigotes <i>L. chagasi</i> .	38
<b>Tabla 11.</b> Comparación de la Act. Exp (IC <sub>50</sub> ) y Act. Calc. (IC <sub>50</sub> ), para el Modelo 3.	43
<b>Tabla 12.</b> Coeficiente de correlación Modelo 3.	44
<b>Tabla 13.</b> Fragmentación del Modelo 3.	44
<b>Tabla 14.</b> Comparación de la Act. Exp (IC <sub>50</sub> ) y Act. Calc. (IC <sub>50</sub> ), para el Modelo 10.	47
<b>Tabla 15.</b> Coeficiente de correlación Modelo 10.	48
<b>Tabla 16.</b> Fragmentación del Modelo 10.	48
<b>Tabla 17.</b> Esquema general de los descriptores para actividad biológica por modelo.	52
<b>Tabla 18.</b> Derivados benzoazepínicos para citotoxicidad.	55
<b>Tabla 19.</b> Propuesta de análogos potencialmente activos contra <i>Leishmania chagasi</i> .	58

<b>Tabla 20.</b> Serie de exploración de los derivados benzoazepínicos para citotoxicidad.	60
<b>Tabla 21.</b> Los 50 modelos con sus descriptores y principales parámetros estadísticos para citotoxicidad.	61
<b>Tabla 22.</b> Comparación entre $CC_{50}$ experimental y $CC_{50}$ calculada para el Modelo 8.	63
<b>Tabla 23.</b> Coeficiente de correlación Modelo 8.	64
<b>Tabla 24.</b> Fragmentación del Modelo 8 para citotoxicidad.	65
<b>Tabla 25.</b> Comparación entre la $CC_{50}$ experimental y $CC_{50}$ calculada para el Modelo 12.	70
<b>Tabla 26.</b> Coeficiente de correlación Modelo 12.	70
<b>Tabla 27.</b> Fragmentación del Modelo 12 para citotoxicidad.	71
<b>Tabla 28.</b> Propuesta de análogos benzoazepínicos.	73

## ÍNDICE DE GRÁFICAS.

<b>Gráfica 1.</b> Correlación de Log-(1/IC <sub>50</sub> ) Exp. vs. Log-(IC <sub>50</sub> ) Cal. Modelo 2.....	37
<b>Gráfica 2.</b> Correlación de Log-(1/IC <sub>50</sub> ) Exp. vs. Log-(IC <sub>50</sub> ) Cal. Modelo 3.....	43
<b>Gráfica 3.</b> Correlación de Log-(1/IC <sub>50</sub> ) Exp. vs. Log-(IC <sub>50</sub> ) Cal. Modelo 10.....	47
<b>Gráfica 4.</b> Dispresión general de IC <sub>50</sub> Exp. vs. IC <sub>50</sub> Calc. Para promastigotes de <i>Leishmania chagasi</i> .....	53
<b>Gráfica 5.</b> Correlación de de Log-(1/CC <sub>50</sub> ) Exp. vs. Log-(CC <sub>50</sub> ) Cal. Modelo 8.....	65
<b>Gráfica 6.</b> Correlación de de Log-(1/CC <sub>50</sub> ) Exp. vs. Log-(CC <sub>50</sub> ) Cal. Modelo 12.....	71
<b>Gráfica 7.</b> Comparación de la IC <sub>50</sub> del Modelo 2 y la CC <sub>50</sub> del Modelo 8, para los derivados benzoazepínicos.....	75
<b>Gráfica 8.</b> Comparación de la IC <sub>50</sub> del Modelo 2 y la CC <sub>50</sub> del Modelo 12, para los derivados benzoazepínicos.....	76

## ABREVIATURAS.

<i>L.</i>	<i>Leishmania.</i>
<i>SP.</i>	<i>Especies.</i>
QSAR	Estudio de relación cuantitativa estructura-actividad.
ATP	Adenosín trifosfato.
mm	Milímetros.
®	Registrado.
<i>Et. al.</i>	Colaboradores.
THP-1	Línea celular monocítica humana.
IC <sub>50</sub>	Concentración Inhibitoria 50.
CC <sub>50</sub>	Concentración Citotóxica 50.
MM	Mecánica Molecular.
Kcal	Kilocalorias.
Å	Angstroms.
mol	Molar.
AM1	Método semiempírico Austin Model 1.
Log	Logaritmo
AGs	Algoritmos genéticos
Q <sup>2</sup>	Capacidad predictiva del modelo.
Q <sup>2</sup> boot	Capacidad predictiva optimizada.
R <sup>2</sup>	Indicador de la precisión del modelo (Coeficiente de determinación 100%).
r <sup>2</sup>	Coeficiente de determinación de validez cruzada.
F	Parámetro de Fisher.
s	Desviación estándar.
Kx	Índices multivariados de correlación.
n	Número de datos.
µM	Micromolar.

SIC4	Información relacionada con la simetría vecinal de orden 4.
CIC4	Cotenido de información complementaria, simetría 4.
PJI3	Descriptor geométrico 3D.
BELv3	Descriptor propio de la carga.
VdW	Van der Waals.
TSA	Superficie total molecular.
X0A	Índice promedio de conectividad 0.
EEig11r	Borde del orden de adyascencia, descriptor molecular de autovalor
RDF95p	11.
3D-MorSE	Factor de distribución radial de polarizabilidad 9.5. Representación molecular 3D de las estructuras basadas en la
Mor06m	difracción de electrones.
Mor24e	Señal 06 de 3D-Morse, para masas atómicas. Señal 24 de 3D-Morse, para electronegatividades atómicas de
Mor10p	Sanderson.
Mor20u	Señal 10 de 3D-Morse, para polarizabilidad atómica.
R <sub>1,2,3</sub>	Señal 20 de 3D-Morse, no ponderado. Sustituyentes en el compuesto benzoazepinico cabeza de serie.

## RESUMEN.

La *Leishmaniasis* es una enfermedad producida por un parásito protozoario perteneciente al género *Leishmania sp.* Se manifiesta en tres formas clínicas: cutánea, mucocutánea y visceral. El tratamiento de primera elección para las diferentes formas clínicas de *Leishmaniasis* son las sales de antimonio pentavalente, como primera elección, y anfotericina B como fármaco de segunda elección; en general estos fármacos no son eficaces y resultan altamente tóxicos.

El desarrollo alcanzado por la química computacional ha favorecido la generación de sistemas que permitan calcular las características fisicoquímicas de compuestos químicos, estos sistemas son capaces de generar datos con aplicación en la investigación experimental, tanto para la interpretación de los resultados obtenidos y la planificación, así como para ajustar la información experimentalmente. Uno de los métodos que relacionan la estructura química con la actividad biológica asistida por computadoras es QSAR, análisis cuantitativos de relación estructura actividad, en la cual es necesario partir de la actividad biológica de una serie de compuestos y que ésta, esté en función de las características estructurales del compuesto regidas por descriptores moleculares como: la geometría, la energía molecular, la electronegatividad, radio atómico, entre otras.

En el presente trabajo se realizó el estudio QSAR asistido por inteligencia artificial usando algoritmos genéticos, para 15 derivados de la *cis*-2-aril-4-hidroxitetrahydro-1-benzazepina, para actividad biológica y citotoxicidad. Se obtuvieron 50 modelos matemáticos de cada serie, junto con las validaciones estadísticas interna y externa de los modelos. Los mejores modelos, que explican y predicen la actividad biológica y citotoxicidad para cada estudio realizado mostraron que los descriptores, CIC4, BELv3 y PJ13, relacionan la simetría estructural, el valor de la carga y la geometría de compuestos químicos en el caso

de la actividad biológica y los descriptores 3D-MorSE, que representan el tamaño y la carga nuclear para citotoxicidad, son los determinantes para la selección de buenos compuestos que tengan, buena actividad biológica con el mínimo de toxicidad.

**Palabras clave:** *Leishmaniasis*, *Leishmania sp.*, protozooario, QSAR, algoritmos genéticos, descriptores.

## S U M M A R Y.

Leishmaniasis is a disease caused by a protozoan parasite belonging to the genus *Leishmania* sp. It occurs in three clinical forms: cutaneous, mucocutaneous and visceral. The treatment of choice for different clinical forms of leishmaniasis are pentavalent antimony salts, as first choice, and amphotericin B as second choice, in general, these drugs are not effective and are highly toxic.

The development achieved by computational chemistry has led to the generation of systems to calculate the physicochemical properties of chemical compounds, these systems are able to generate data with applications in experimental research, both for interpreting the results and planning as well information as to adjust the experimentally. One method of linking the chemical structure with biological activity is QSAR computer-assisted analysis, quantitative structure activity relationship, which is necessary from the biological activity of a series of compounds and it may be in function of the composite structural characteristics as governed by molecular descriptors: geometry, molecular energy, electronegativity, atomic radius and others.

In the present work, the study artificial intelligence assisted QSAR genetic algorithms for 15 derivatives of cis-2-aryl-4-1-benzazepine hydroxytetrahydro for biological activity and cytotoxicity. 50 were obtained mathematical models of each series, along with the internal and external validation of statistical models. The best models that explain and predict the biological activity and cytotoxicity for each study showed that the descriptors, CIC4, and PJI3 BELv3, related structural symmetry, the value of the load and the geometry of chemical compounds for activity biological and 3D-Morse descriptors representing the nuclear size and charge to cytotoxicity, are the determinants for the selection of compounds having good, good biological activity with minimal toxicity.

**Keywords:** *Leishmaniasis*, *Leishmania* sp., protozoan, QSAR, genetic algorithms, descriptors.

## 1. INTRODUCCIÓN.

Las enfermedades infecciosas incluyendo las causadas por parásitos, son de las primeras causas de muerte a escala mundial; infecciones por helmintos intestinales y protozoos están entre las más comunes. La epidemiología de estas parasitosis se favorece por los factores ecológicos, socioeconómicos y culturales. En México se han desarrollado una gran variedad de investigaciones para observar la prevalencia de los parásitos, los cuales se relacionan directamente con las condiciones sanitarias ambientales como son: agua potable, letrinas, drenajes, alcantarillado, disposición de desechos sólidos influyendo también las condiciones de la vivienda como son: pisos de tierra, falta de drenaje y falta de agua potable entre otros.<sup>1</sup>

Una de las enfermedades parasitarias que en los últimos años ha causado inquietud en el área médica es la “*Leishmaniasis*”, que se caracteriza por ser un conjunto de enfermedades, producidas por distintas especies de un protozoario hemoflagelado perteneciente al género *Leishmania*. Esta enfermedad que es de evolución crónica, se caracteriza por implicar a piel, mucosas y vísceras dependientes de la especie causante y de la respuesta inmune del huésped.<sup>2</sup> Tiene en común el agente causal *Leishmania*, el vector (insectos dípteros hematófagos), el reservorio (vertebrados) y el parasitismo de las células del sistema fagocítico mononuclear (sobre todo macrófagos). Generalmente afecta zonas rurales tropicales donde prevalecen escasos recursos socioeconómicos y culturales.<sup>3,4</sup>

A pesar de tener información sobre la enfermedad, el tratamiento no siempre es del todo accesible en cuanto a costos y lo más importante, los medicamentos no son eficaces y resultan altamente tóxicos. Es por esto que el desarrollo de nuevos fármacos, cada vez más seguros, adecuados y efectivos en el tratamiento de enfermedades, es una tarea que requiere la capacidad de deducción, la institución y en muchos casos, la suerte. Si se conoce la base biológica de la enfermedad o de un desarreglo metabólico, es posible diseñar un medicamento,

utilizando un mecanismo de aproximación al proceso patológico, que interactúe con la molécula responsable.

El desarrollo alcanzado por la computación y la química computacional ha favorecido la generación de sistemas que permitan calcular la geometría y la energía molecular. Estos sistemas son capaces de generar datos con una amplia aplicación en la investigación experimental, tanto para la interpretación de los resultados obtenidos y la planificación de los resultados, así como para deducir información experimentalmente. Uno de los métodos que relacionan la estructura química con la actividad biológica asistida por computadoras es QSAR.<sup>5</sup>

Las siglas QSAR, provienen de la expresión en inglés “*Quantitative Structure-Activity Relationship*” que significa en español “Estudios cuantitativos de relación estructura actividad”. Para realizar un estudio QSAR es necesario partir de que la actividad biológica de un compuesto activo y el producto cuantificable de su interacción con el receptor, es una función de las características estructurales de la molécula.<sup>6, 7</sup>

La base de QSAR es la correlación entre los valores experimental de la actividad y descriptores moleculares teóricos que reflejan las propiedades moleculares de los compuestos. La relación QSAR es la correlación cuantitativa de las propiedades estructurales de un compuesto con sus efectos biológicos. De esto se parte que un estudio QSAR tiene los siguientes objetivos: 1) La predicción de la actividad de las moléculas probadas usando una serie de moléculas y 2) La construcción de ideas sobre el mecanismo de acción de un grupo de compuestos o explicar el mecanismo de acción, útiles para el diseño de nuevos compuestos con mejor actividad biológica y menor toxicidad.<sup>8, 9</sup>

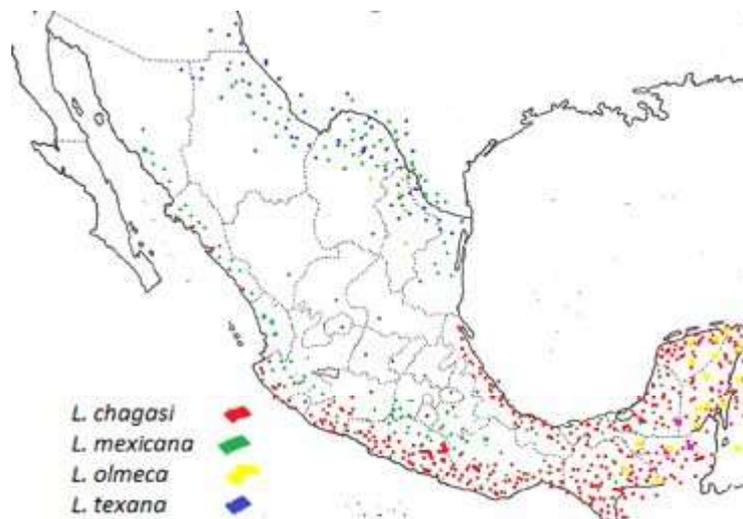
Con esta información se hace hincapié en que la *Leishmaniasis* es una enfermedad del pasado que prevalece en la actualidad y que requiere medidas urgentes para su tratamiento. Por tal motivo en este proyecto se diseñaron por QSAR compuestos del tipo *cis*-2-aril-4-hidroxtetrahidro-1-benzoazepinas tomando como base una serie de benzoazepinas con actividad biológica probada contra *Leishmania chagasi*, con ello explicar sus actividades y como consecuencia de estos compuestos ampliar el panel de antileishmaniacicos con posibilidades de menos efectos colaterales y secundarios indeseados y sobre todo con mayores posibilidades de éxito frente a los disponibles hasta el momento.<sup>10</sup>

## 2. ANTECEDENTES.

### 2.1 LEISHMANIASIS.

La *Leishmaniasis* es una enfermedad endémica en 88 países de los cinco continentes, con un total de 350 millones de personas en riesgo y más de 12 millones afectadas. Entre 1.5 y 2 millones de casos nuevos ocurren cada año. El 90% de los casos de *Leishmaniasis* cutánea ocurren en Brasil, Perú, y México.<sup>11</sup>

Los Estados del sureste de México son los más afectados por *Leishmaniasis*; en 1999 el Estado de Quintana Roo ocupó la tasa más alta de afección y para el año 2002 estuvo en el segundo lugar después de Tabasco y las poblaciones que se encuentran en la rívera del Río Hondo (frontera con Belice) y el área limítrofe con Campeche y Guatemala que comprenden la zona selvática de la península (Figura 1).<sup>12, 13</sup>



**Figura 1.** Distribución geográfica en México de los principales agentes causales de *Leishmaniasis*.

De acuerdo con el Instituto de la salud (Indesalud), en Campeche México, en 2004, se registraron 115 casos de *Leishmaniasis*, de los cuales aproximadamente el 30% de la población pertenece a comunidades alejadas de los principales centros de salud.<sup>14</sup>

La *Leishmaniasis* es transmitida por la picadura de hembras flebótomo, especies de *Phlebotomus* en Europa, Asia y África, y por especies endémicas de *Lutzomyia (Nyssomyia)*, *Olmecca (Psychodiadae)*, al que se conoce como “mosca chiclera” en América, un pequeño mosquito de 1.5 – 2.0 mm de tamaño, se ha referido al *Psychodopygus* como una segunda especie en las Américas, sin embargo el consenso entre los taxonomistas es que el *Psychodopygus* es simplemente un subgénero dentro del género *Lutzomyia*. Todas las *Leishmaniasis* poseen un ciclo de vida similar, (Figura 2): Uno principalmente silvestre donde la *Leishmania sp.*, protozoo, un parásito intracelular obligado, circula entre los reservorios naturales, manteniendo el ciclo con la participación de vectores propios de la zona. En un segundo ciclo, estos vectores infectados pueden atacar al hombre y a los animales domésticos.<sup>15, 16</sup>

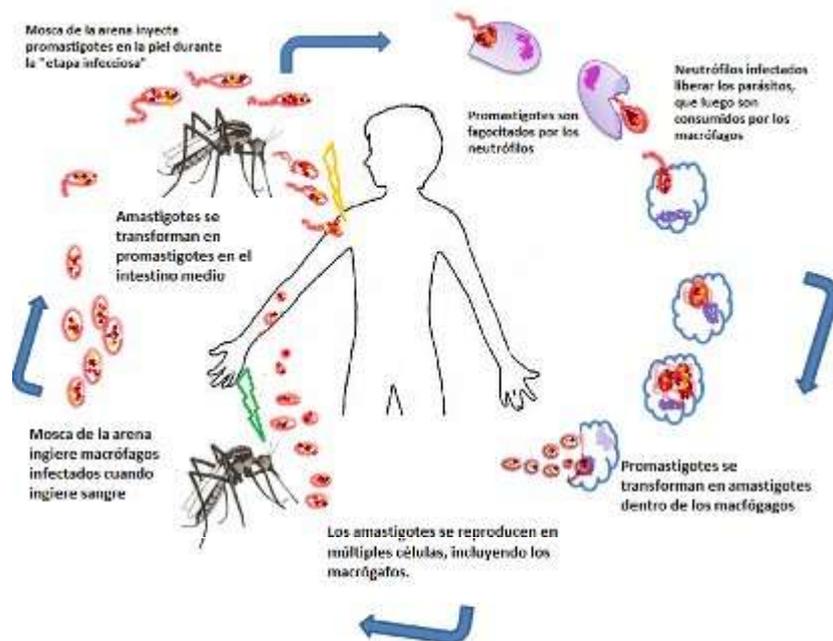


Figura 2. Ciclo biológico de la *Leishmania sp.* (Modificado de Al-Jawbreh., et al.2008).

Cuando una hembra de flebótomo ingurgita sangre de un animal parasitado, ingiere macrófagos con amastigotes, estos se liberan para multiplicarse en el tubo digestivo, adquiriendo la forma de promastigotes infectivos. Ante una nueva ingesta, el flebótomo inocula saliva para vasodilatar y con ella, pasan los promastigotes al animal y penetran en los macrófagos cerrándose así el ciclo.<sup>17, 18</sup>

Hay dos tipos epidemiológicos de *Leishmaniasis*: la forma antroponótica, entre una persona enferma, flebótomo que actúa de vector, y una persona sana; y la forma zoonótica, entre un animal enfermo que actúa como reservorio, el flebótomo como vector, y la persona sana. Las manifestaciones clínicas de la *Leishmania* están relacionadas en parte a la cepa del agente infectante, al medio ambiente y la respuesta inmune del hospedero. Las formas clínicas caracterizadas son: *Leishmaniasis cutánea*, *L. mucocutánea*, *L. visceral*.<sup>4, 19</sup>

La enfermedad es endémica en áreas geográficas circunscritas en el noreste de África, sur de Europa, Medio Oriente, sureste de México, América central y del sur. Las manifestaciones clínicas dependen de la especie de *Leishmania* involucrada y de la respuesta inmunológica del huésped. En México, la forma más característica es la mucocutánea ulceral localizada llamada “úlceras de los chicleros”, el agente causal es *L. mexicana* y *L. chagasi*.<sup>4, 9</sup>

## 2.2 TRATAMIENTO.

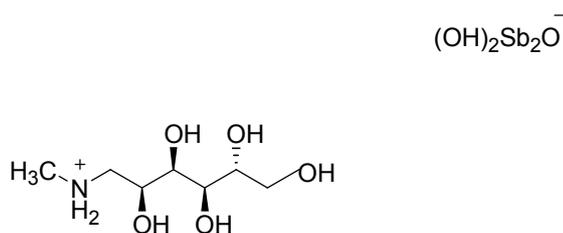
El tratamiento de esta parasitosis varía de acuerdo con el tipo de *Leishmaniasis*, la severidad de la enfermedad, el lugar donde se infecta el individuo y el estado nutricional e inmunológico del paciente. Debido a que la respuesta ante la terapia no es la misma y varía entre individuos, existen varios tratamientos médicos que incluyen a fármacos de primera y segunda línea.

El tratamiento de primera elección para todas las formas de *Leishmaniasis* son los antimoniales pentavalentes en forma de antimonio de meglumina (Glucantime®, Laboratorios Aventis) o estibogluconato de sodio (Pentostam®, Laboratorios GSK). El tratamiento puede durar hasta 40 días.<sup>20</sup>

Los mecanismos de acción de estos fármacos se asocian al bloqueo de la glicólisis, metabolismo de ácidos grasos y la formación de ATP, debiendo convertirse en trivalentes para ejercer su acción farmacológica. Son más eficaces mientras más temprano sea el tratamiento, sin embargo no hay estudios *in vivo* que comparen su eficacia y su citotoxicidad controlando variables en el hospedero.  
21, 22

**a) ANTIMONIATO DE MEGLUMINA.**

Las reacciones adversas más comunes del antimonitato de meglumina (Figura 3) incluyen: Dolor en el sitio de aplicación, cefalea, vómito, náuseas, nefrotoxicidad, pancreatitis, mialgias, artralgias, fatiga, reacciones de hipersensibilidad, arritmias ventriculares.<sup>23</sup>



**Figura 3.** Antimoniato de Meglumina (Glucontime®).

**b) ESTIBOGLUCONATO DE SODIO.**

En el estibogluconato de sodio (Figura 4), las reacciones adversas más comunes incluyen dolor en el sitio de aplicación, vómito, náuseas, elevación de transaminasas, mialgias, artralgias, fatiga, nefrotoxicidad y cefalea cardiotoxicidad. Está contraindicado en el embarazo y no deben ser usados en pacientes con problemas cardíacos, hepáticos, renales o pancreáticos. En general los antimoniales tienen reacciones adversas músculo-esqueléticas y toxicidad sobre el bloqueo del metabolismo energético.<sup>24</sup>

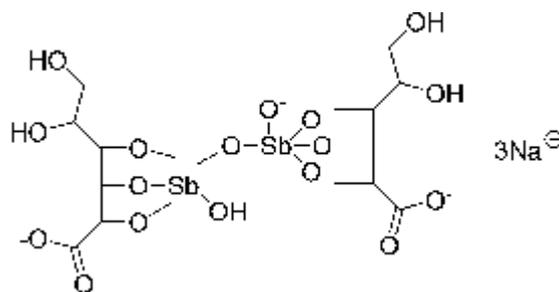


Figura 4. Estibogluconato de sodio (Pentostam®)

Los fármacos de segunda elección son la Anfotericina B y el Isotionato de pentamidina. Ambas son más tóxicas que los antimoniales y requieren que su uso sea intrahospitalario, pueden representar cura clínica a pesar de su toxicidad.<sup>25</sup>

#### c) ANFOTERICINA B.

El tratamiento de segunda elección en la *Leishmaniasis* visceral, es la anfotericina B (Figura 5), esta se fija a los esteroides de la membrana de las células eucariotas, con mayor afinidad al ergosterol del parásito, alterando su permeabilidad con el desequilibrio de los iones sodio, potasio e hidrogeniones, provocando la muerte del parásito. Es usado en regiones en las que se presenta alta resistencia al tratamiento con pentavalentes. Sin embargo como el fármaco también se une al colesterol de las células humanas, la anfotericina B, presenta efectos tóxicos, en particular a nivel renal, hepático y cardiaco, entre sus efectos colaterales sobresalen: escalofríos, fiebre, náuseas y vómitos, las cuales aparecen durante los dos primeros días de tratamiento, anemia, desórdenes de la coagulación, aumento o disminución de la protrombina, dolor abdominal, anorexia, calambres, diarrea, dispepsia, dolor epigástrico, náuseas, estomatitis, pérdida de peso. Sin embargo, el uso de este fármaco en regiones endémicas tiene como limitaciones el costo del tratamiento, las dificultades de administración y su toxicidad.<sup>26, 27</sup>



y eficacia de este fármaco. La miltefosina es un fosfolípido sintético activo tópicamente y por vía oral, químicamente similar a los fosfolípidos naturales, presenta una actividad antiparasitaria directa independiente del sistema inmune. El costo casi inaccesible para las zonas de prevalencia parasitaria.<sup>28, 29</sup>

Por los datos relacionados con el tratamiento, resulta evidente la importancia de buscar un fármaco que sea eficaz contra la enfermedad, que no produzca resistencia ante los parásitos y que los efectos secundarios sean menores. En este sentido, la metodología QSAR es una herramienta que permite proponer estructuras que representen ser buenos candidatos para experimentarse contra *Leishmania in vitro*.

### 2.3 TECNICA QSAR EN EL DESARROLLO DE NUEVOS FARMACOS.

Diseñar fármacos asistido por computadoras engloba métodos teóricos y técnicas computacionales para modelar o imitar el comportamiento de las moléculas, en este sentido se opta por la metodología QSAR (Quantitative Structure-Activity Relationship), por sus siglas en inglés se traduce como Estudios cuantitativos de relación estructura-actividad, siendo una de las herramientas más importantes para el diseño y descubrimiento de fármacos.<sup>30, 31</sup>

Para realizar un estudio QSAR es necesario que la actividad biológica del fármaco, producto de su interacción con el receptor, sea función de las características estructurales de la molécula. El modelo extratermodinámico de Hansch brinda una explicación matemática del fenómeno recogido en la siguiente ecuación:<sup>32, 33</sup>

$$\ln A = f_h(X_h) + f_e(X_e) + f_s(X_s) + cte$$

Donde: A, es la actividad y  $f_h(X_h)$ ,  $f_e(X_e)$ ,  $f_s(X_s)$  son funciones de índices o parámetros hidrofóbicos, electrónicos o estéricos respectivamente.

Aunque no fue precisamente Hansh a quien se le atribuye el QSAR, sí fue quien sistematizó el estudio basándose en su paradigma conocido como “El paradigma de Hansch”<sup>34</sup>, que establece que:

1. La actividad biológica es función de la estructura del fármaco.
2. La estructura del fármaco implica ciertas propiedades globales como la carga neta, solubilidad y ciertas propiedades locales como la distribución de la hidrofobicidad, carga y volumen en determinadas posiciones de la molécula.

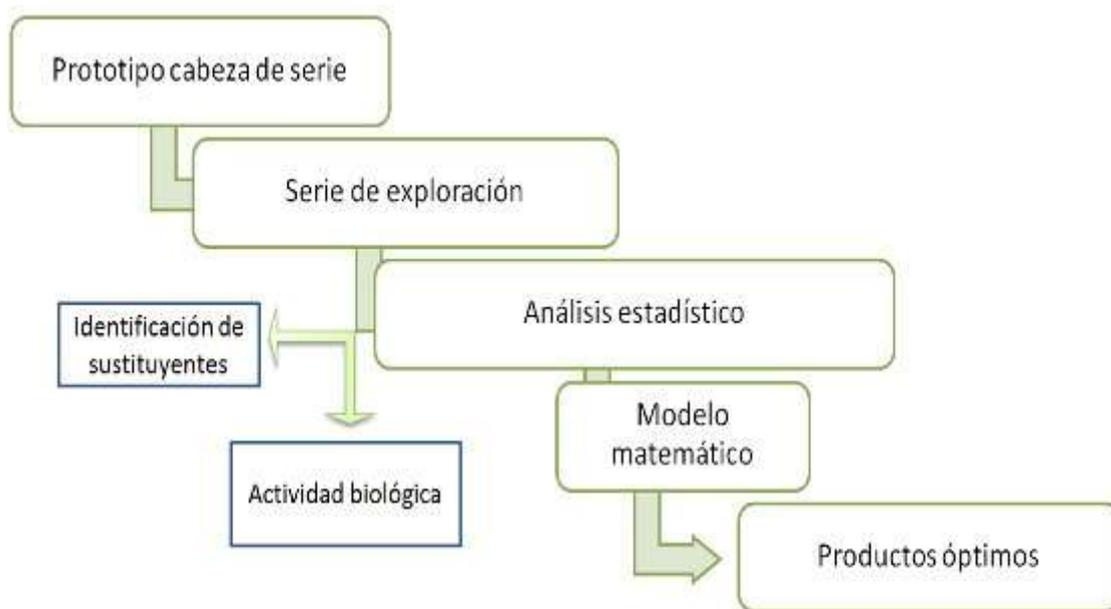
3. Estas propiedades globales y locales pueden ser cuantificadas mediante ciertos parámetros.

4. Siempre existe una función que relaciona los cambios de actividad biológica con los cambios en las propiedades globales y locales, aunque puede ser que no sea sencilla ni evidente.<sup>35</sup>

Por lo tanto la Relación Cuantitativa Estructura-Actividad (QSAR) es la correlación cuantitativa de las propiedades estructurales de un compuesto con su actividad biológica cuantificable. Donde se correlacionan diferentes propiedades fisicoquímicas de un conjunto de moléculas con su actividad biológica conocida experimentalmente, teniendo como objetivos: 1) La predicción de la actividad de las moléculas probadas, en modelos desarrollados usando una serie de moléculas y 2) La explicación molecular de la actividad que puede conducir a un diseño de nuevos compuestos de mejor actividad y menor toxicidad.<sup>36, 37</sup>

Las funciones estructura-actividad que propone el método QSAR (Figura 7), pueden buscar los productos más activos entre un conjunto de candidatos. El punto de partida implica en todos los casos la existencia de un cabeza de serie, donde se tienen propiedades susceptibles a ser mejoradas para obtener siempre un producto lo más biodisponible posible.<sup>38</sup>

Cuando se dispone de un prototipo, es preciso diseñar una serie de exploración, constituida por un conjunto de análogos del prototipo y que permiten el establecimiento de las primeras relaciones estructura-actividad.



**Figura 7.** Metodología de investigación en el QSAR.

Los miembros de la serie de exploración están constituidos por un núcleo común, y los sustituyentes o fragmentos variables, que son característicos de cada producto de la serie. Dichos sustituyentes variables han de ser identificados por descriptores, que serán utilizados como variables independientes ( $X_i$ ) en el modelo. Los valores de actividad biológica ( $A$ ) expresados regularmente en forma logarítmica se utilizan como variable dependiente.<sup>39, 40</sup>

Para la optimización del modelo QSAR se opta por métodos adaptativos, como los Algoritmos Genéticos (AGs), que en un sentido amplio es un modelo de técnicas usadas por un sistema biológico, donde usa la reproducción, mutación y adaptación, para producir los descendientes que puedan sobrevivir en el ambiente donde se desarrollan emulando la hipótesis de Darwin en “La supervivencia de los más aptos”.<sup>41, 42</sup> De manera análoga, en QSAR son usados para resolver problemas de búsqueda y optimización de datos, como operaciones de estrategia que mejoran la calidad de soluciones a un problema en particular, seleccionando descriptores de relación mínima y con varianza significativa. En AGs, cada individuo de la población está delimitado por un cromosoma de valores. En

términos genéticos, cada variable se denomina gen y un conjunto de variables se llama un cromosoma, donde la población de la primera generación se seleccionó al azar. En esta fase en un QSAR, un “descriptor molecular” es el resultado final de un procedimiento matemático que transforma la información química codificada dentro de una representación simbólica de una molécula, en un número útil o el resultado de un cierto experimento estandarizado.<sup>43, 44</sup>

Los descriptores moleculares, son números que proveen datos teóricos y experimentales de propiedades fisicoquímicas de un determinado compuesto. Estas propiedades moleculares no dependen solo de la composición química sino también de su actividad, y solamente una parte de esta información se puede extraer experimentalmente y por lo que el resto debe ser de manera teórica.<sup>45, 46</sup>

Esto significa que el número arrojado por el descriptor puede representar un enfoque de la interpretación de las propiedades moleculares o puede formar parte en un modelo para la predicción de alguna propiedad sobresaliente de otras moléculas. Un aspecto fundamental de los descriptores es predecir y comprender hechos experimentales, como propiedades fisicoquímicas, biológicas o actividades diversas. El estudio sobre los descriptores es complejo, su interpretación puede ser débil, provisional, o carecer por completo de interpretación, de estos últimos debe tenerse en cuenta su capacidad predictiva o la utilidad en su aplicación a problemas reales, los cuales pueden ser un fuerte motivo para su uso. En cuanto a los descriptores con pobre capacidad de predicción, a veces son útiles cuando se codifica información química estructural: como peso molecular o la geometría de la molécula, descriptores geométricos, etc.<sup>47,48</sup> En la Tabla 1, se muestran las características más sobresalientes de los descriptores involucrados en un estudio QSAR utilizados por el programa Dragon ®.

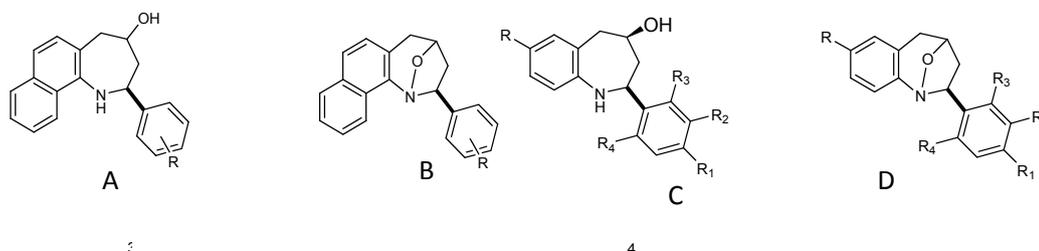
Tabla 1. Características generales de los descriptores, por Dragon 5.3®.<sup>49</sup>

PROPIEDAD DEL DESCRIPTOR	CARACTERÍSTICAS GENERALES
<i>Datos Geométricos</i>	Longitud de enlace, Ángulo de enlace, Ángulo de torsión, Estructura tridimensional, Distancia interatómica, Intermediarios de reacción, Estados de transición.
<i>Termodinámica y energía</i>	Energía molecular, Calor de formación, Población conformacional, Entropía, Energía de activación, Superficie de energía potencial, Camino de reacción, Energía de solvatación.
<i>Propiedades electrónicas</i>	Distribución de carga, Momento dipolar, Potencial de ionización, Afinidad electrónica, Afinidad protónica, Polarización, Campo de potencial electrostática.
<i>Propiedades espectroscópicas</i>	Frecuencia vibracional, Energía de excitación ultravioleta, Coeficiente de extinción.
<i>Interacción Molecular</i>	Regla de Woodward-Hoffmann, Energía de asociación, Sitios de unión de macromoléculas, pKa's.
<i>Propiedades de Transporte</i>	Volumen molecular, Área de superficie molecular.

## 2.4 SÍNTESIS DE DERIVADOS ANTIPARASITARIOS CONTRA *Leishmania chagasi*.

Como se mencionó, la búsqueda de nuevos compuestos activos contra *Leishmania* es un área de investigación que urgentemente requiere ser atendida. En este sentido se han desarrollado una gran cantidad de trabajos por laboratorios de investigación de los países más afectados por la enfermedad. Uno de estos trabajos es el que se llevó a cabo en la Universidad Industrial de Santander por el grupo del Doctor Alirio Palma en el Laboratorio de Síntesis Orgánica, en donde se sintetizó una serie de *cis*-2-aril-4-hidroxitetrahydro-1-bezoazepinas (Figura 8) activas contra *Leishmania chagasi*.<sup>7</sup> Aunque los compuestos no fueron probados contra *L. Mexicana* que es de mayor importancia en México, es de esperar un mecanismo de acción molecular similar al resto de las especies de *Leishmania*.

Así mismo, *L. chagasi*, también es un parásito que cohabita en centro y sur América incluyendo nuestro país.

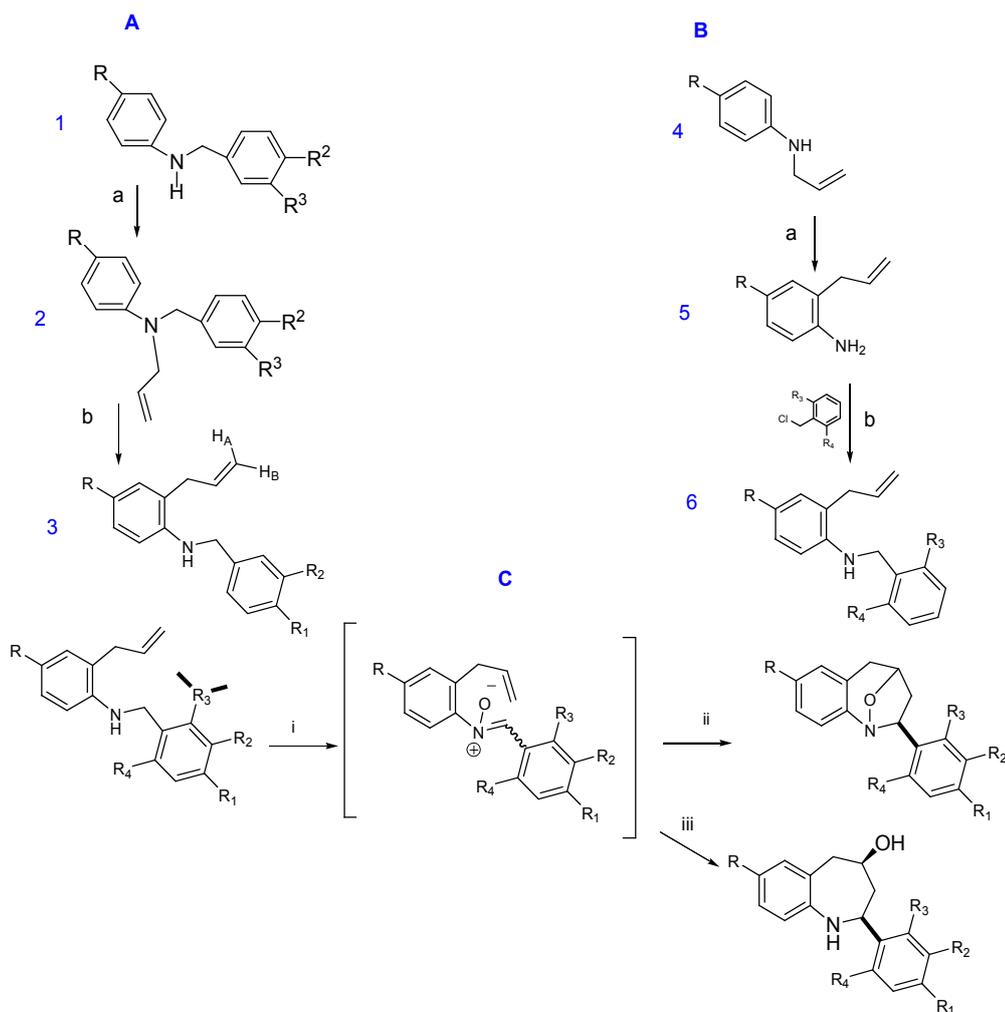


**Figura 8.** Estructura general de las tetrahidronafto (A, B) azepina, (C, D) tetrahidro-1-benzoazepinas, derivados con actividad biológica contra *Leishmania sp.*

Los derivados de tetrahidro-1-benzoazepina han sido ampliamente investigados en la química farmacéutica. Sintéticamente contienen una unidad de heterocíclicos que poseen una amplia escala de actividades biológicas, contra los amastigotes y promastigotes de especies de *Leishmania*. Con base a esta información se tomó como referencia la síntesis del sistema de anillos tetrahidro-1-benzoazepina, expuestos en la Figura 9.

Palma, *et al.*, en el 2008, reportaron la actividad antiparasitaria de *cis*-2-aril-4-hidroxitetrahidronafto [1,2-b] benzodiazepinas, contra *L. chagasi*. Los compuestos que evaluaron fueron sintetizados a partir de N-bencil-sustituido- $\alpha$ -naftilaminas, que en la reacción con exceso de bromuro de alilo en presencia de carbonato de potasio en acetona seca a reflujo, se convirtieron en N-alil-N-bencil-sustituido- $\alpha$ -naftilaminas, con buenos rendimientos.<sup>50</sup> Posteriormente Acosta L., *et al.*, 2010., describieron la síntesis estereoselectiva de las 1,4-epoxi-2-exo-vinilo (isopropenil) tetrahidro-1-benzoazepinas y *cis*-2-vinil (isopropenil)-4-hidroxitetrahidro-1-benzoazepinas, compuestos que permiten la síntesis de análogos variando el sustituyente del nitrógeno, en el último paso de la ruta sintética, los autores llevaron a cabo la apertura del 1,4-epoxi-cicloadductos

obteniendo la respectiva tetrahidro-1-benzoazepina-4-ol, los cuales presentan una notable actividad contra promastigotes de *L. chagasi*.<sup>51</sup>



**Figura 9.** Síntesis del *cis*-2-aryl-4-hidroxitetrahydro-1-benzoazepina. (Modificado de, Gómez Ayala, *et al.*, 2010).

De tal forma, se describió la síntesis y la actividad antiparasitaria de dos series de compuestos, 1,4-epoxi-2-exo-aril y *cis*-2-aril-4-1-hidroxitetrahydrobenzoazepinas, así como la determinación de su citotoxicidad en THP-1<sup>7</sup> (las células huésped de *L. chagasi*), se refiere a una línea celular monocítica humana que expresa naturalmente receptores de reconocimiento de patrones de tipo Toll, para demostrar cómo estos receptores son capaces de interactuar con diferentes patógenos para poder estimular alguna respuesta inmune más específica como la inflamación, son expresados en diferentes tipos celulares, principalmente células del sistema inmune tales como células dendríticas, macrófagos, neutrófilos y linfocitos, además de células endoteliales y epiteliales.<sup>52</sup> La actividad antiparasitaria y citotoxicidad de los compuestos evaluados se expresa como la concentración requerida para inhibir el 50% y el 90% de los parásitos (IC<sub>50</sub> e IC<sub>90</sub>), y la concentración requerida para matar el 50% y el 90% de las células de mamíferos (CC<sub>50</sub> y CC<sub>90</sub>).<sup>53</sup>

La importancia de la síntesis descrita hasta aquí, es que se han propuesto compuestos con actividad biológica contra los promastigotes del parásito y baja toxicidad frente a líneas celulares humanas. Desafortunadamente, no se conoce hasta el momento el mecanismo de acción de dichos compuestos.

Como parte de la colaboración de nuestro grupo de trabajo con el equipo del Dr. Palma, en el Laboratorio de Síntesis Orgánica de la Escuela de Química, de la Universidad Industrial de Santander, Bucaramanga en Colombia, se llevó a cabo el estudio QSAR de los compuestos derivados del tetrahydro-1-benzoazepina-4ol, a partir de los mismos se plantearon estructuras novedosas que presentaron actividad superior a las sintetizadas, pero sobretodo explicar en función de qué propiedades moleculares depende la actividad y la citotoxicidad, de dicha serie de compuestos.

### 3. OBJETIVOS.

#### 3.1 OBJETIVO GENERAL.

Diseñar compuestos que estructuralmente sean candidatos activos contra *Leishmania sp.*, utilizando estudios QSAR a partir de una base de datos.

#### 3.2 OBJETIVOS PARTICULARES.

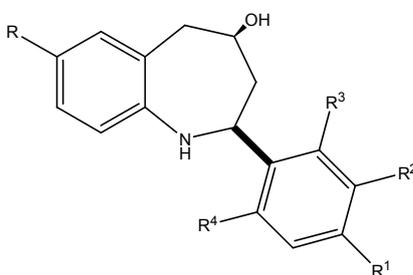
A) Obtener modelos QSAR que expliquen la actividad biológica experimental y predigan la actividad calculada de los compuestos derivados del *cis*-2-aril-4-hidroxitetrahydro-1-benzoazepina.

B) Obtener modelos QSAR, de la misma serie de compuestos activos contra *Leishmania chagasi* que permitan explicar la citotoxicidad en líneas celulares de mamíferos (THP-1).

C) Diseñar compuestos novedosos análogos al *cis*-2-aril-4-hidroxitetrahydro-1-benzoazepina, a partir del estudio QSAR realizado y predecir su actividad biológica contra *Leishmania chagasi*, así como su baja citotoxicidad.

#### 4. ESTRATEGIA EXPERIMENTAL GENERAL.

Para el estudio QSAR que permitiese el diseño de estructuras análogas potencialmente activas contra *Leishmania chagasi*, se tomaron como referencia 15 compuestos y se utilizó como cabeza de serie el compuesto *cis*-2-aril-4-hidroxitetrahydro-1-benzoazepina (Figura 10).<sup>7</sup>



**Figura 10.** Estructura de la cabeza de serie, para QSAR.

## 5. METODOLOGÍA.

Para llevar a cabo el diseño molecular por QSAR se tomaron en cuenta las siguientes fases (Figura 11):

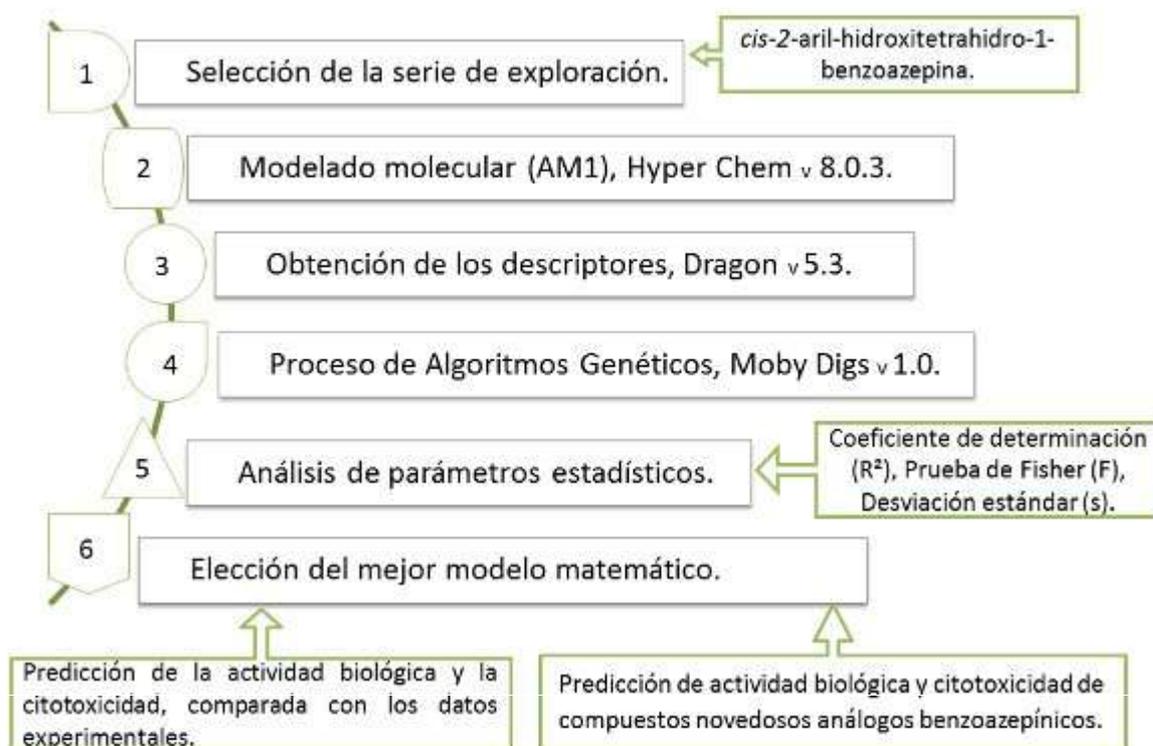
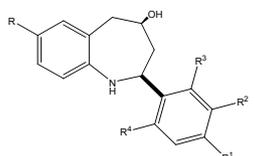


Figura 11. Diagrama general de la metodología QSAR.

### 5.1 DESCRIPCIÓN METODOLÓGICA.

(1) Selección de la serie de exploración a partir de la base de datos de los compuestos *cis*-2-aril-hidroxitetrahydro-1-benzoazepina, considerando para el estudio aquellos cuya actividad partiera de un intervalo de actividad cuantificable, es decir, se descartaron los compuestos con valor de  $IC_{50} > 100$  , (Tabla 2).<sup>7</sup>

**Tabla 2.** Actividad antiparasitaria *in vitro* y citotoxicidad del tetrahidro-1-benzoazepina-4ol, agonista para *Leishmania chagasi*.

 Tetrahidro-1-benzoazepina-4ol							
Compuesto	Sustituyentes					Promastigotes	*THP-1 cells
	R	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	*IC <sub>50</sub>	*CC <sub>50</sub>
B	OCF <sub>3</sub>	H	H	H	H	35.5	103.59
C	Br	H	H	H	H	45.05	184.52
D	Cl	H	H	H	H	22.12	188.07
F	H	Cl	H	H	H	67.89	113.85
H	H	H	OCH <sub>3</sub>	H	H	97.09	-----
I'	H	H	CH <sub>3</sub>	H	H	-----	176.38
J	H	H	Cl	H	H	41.23	166.58
K	CH <sub>3</sub>	CH <sub>3</sub>	H	H	H	-----	171.66
L	Cl	Cl	H	H	H	14.1	49.21
M	Cl	Br	H	H	H	9.54	153.11
N	F	Cl	H	H	H	17.11	100.32
O'	H	H	H	Cl	H	84.44	-----
Q	CH <sub>3</sub>	H	H	Cl	H	41.34	85.19
R	Cl	H	H	Cl	H	50.22	75.53
S	Cl	H	H	F	H	54.45	110.21
V	Cl	H	H	Cl	Cl	14.66	59.59
W	Cl	H	H	Cl	F	39.75	73.22

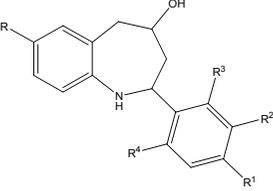
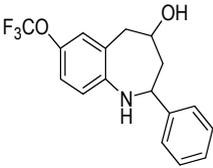
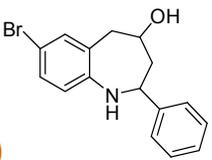
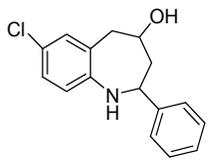
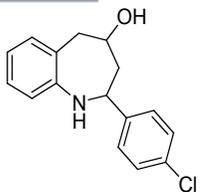
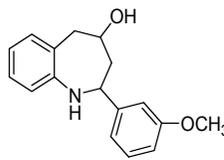
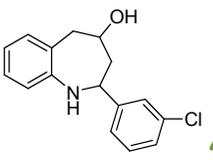
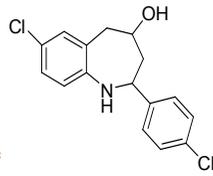
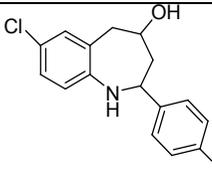
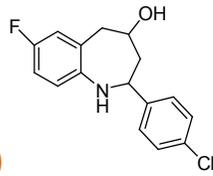
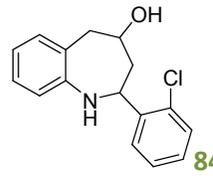
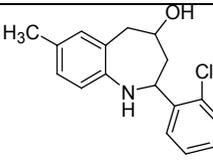
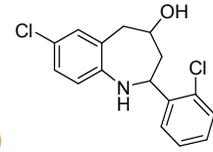
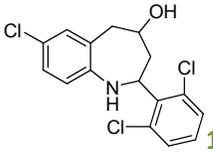
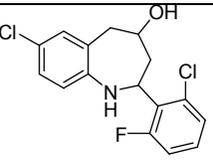
\* THP-1 cels (Líneas celulares de mamíferos), IC<sub>50</sub> (Concentración Inhibitoria 50), CC<sub>50</sub> (Concentración Citotóxica 50). Compuestos señalados con (') sólo incluyen la serie de citotoxicidad.

(2) Modelado Molecular: Optimización geométrica de las estructuras seleccionadas mediante mecánica molecular (MM) y métodos semiepíricos (AM1). El método semiempírico “Austin 1” (AM1) considera únicamente los electrones de valencia, aminorando así significativamente la complejidad y por lo tanto el tiempo de uno de los pasos computacionales más complicados.<sup>54 55</sup>

La elección de estos métodos se consideró por estar parametrizados para las estructuras orgánicas de baja complejidad. Cabe resaltar que por las características de los descriptores a utilizar en el siguiente paso del procedimiento no se requiere, ni sería útil, el uso de métodos *Ab initio* o funcionales de la densidad.<sup>56</sup> Para el modelado se utilizó el software Hyper Chem versión 8.0.3®,<sup>57</sup> en donde la optimización los compuestos se llevó a cabo hasta obtener convergencia en una variación de ciclos, se usó como base el algoritmo Polak-Ribiere,<sup>58</sup> con parámetros de condiciones de terminación de 0.1 kcal/ (Å mol) y 525 ciclos como máximo. Las estructuras con la geometría optimizada mostradas en la Tabla 3, fueron guardadas como archivos con extensión “Hin”.

En la Tabla 3, se muestra la serie de exploración de los compuestos análogos del *cis*-2-aryl-hidroxitetrahydro-1-benzoazepina, con actividad biológica contra *Leishmania chagasi*, elegidos para el estudio QSAR.<sup>7</sup>

**Tabla 3.** Serie de exploración de derivados del *cis*-2-*aryl*-hidroxitetrahydro-1-benzazepina, con actividad biológica ( ).

<b>Cabeza de serie</b>		<b>Serie de exploración</b>				
		 35.5(B)	 45.04(C)	 22.12(D)*	 67.89 (F)	 97.09(H)
 41.23(J)*	 14.1(L)*	 9.54(M)	 17.11(N)	 84.44(O)		
 41.34(Q)	 50.22(R)*	 54.45(S)	 14.66(V)	 2.76(W)		

\* Moléculas seleccionadas para la validación interna de actividad contra promastigotes agonistas de *L. chagasi*.

(3) Las estructuras modeladas con geometría optimizada se importaron al software Dragon versión 5.3®, para obtener los datos numéricos de los descriptores que se encuentran organizados en 20 familias (Tabla 4). Se calcularon 1664 descriptores moleculares para cada uno de los compuestos dando un total de 3 descriptores moleculares teóricos por modelo, que fueron usados para valorar la relación de las propiedades moleculares y la actividad de cada compuesto.<sup>37</sup>

**Tabla 4.** Familias de descriptores calculados en el software Dragon 5.3.

<b>Categoría del descriptor</b>	<b>Cantidad de descriptores</b>	<b>Familia</b>
Constitucional	48	1
Topológico	119	2
“Walk and pathcounts”	47	3
Indicadores de conectividad	33	4
Contenido de información estructural	47	5
Autocorrelaciones 2D	96	6
Borde del índice de adyacencia	107	7
Descriptores propios del valor de carga	64	8
Indicadores topológicos de carga	21	9
Indicadores basados en valores propios	44	10
Perfiles rándico-moleculares	41	11
Geométricos	74	12
Función de Distribución Radial (RDF)	150	13
3D-MorSE	160	14
WHIM	99	15
GETAWAY	197	16
Conteo de grupos funcionales	154	17
Fragmentos átomo centrado	120	18
Descriptores de carga	14	19
Propiedades moleculares	29	20

Algunas familias 

--	--	--	--

 de descriptores no tienen traducción al español y se usan su nombre original.

(4) Se seleccionaron cuatro compuestos al azar con la finalidad de ser utilizados en la validación externa, que estriba en predecir la actividad sin haber participado en la selección de modelos; es decir, se asume que no se cuenta con sus datos de actividad experimental al hacer el cálculo y que al predecirse por los modelos finales, la actividad calculada debe coincidir o estar cercana a sus datos experimentales. Una vez obtenidos los descriptores moleculares se seleccionan las ecuaciones matemáticas mediante un análisis de regresión lineal múltiple, tomando en cuenta las características estadísticas: el número de compuestos en regresión (n), el coeficiente de regresión o determinación ( $R^2$ ), la F-test (valor de Fischer) para la significación estadística (F), la desviación estándar de la muestra (s), la validación cruzada y el coeficiente de correlación. Dado que se trata de miles de variables y que es imposible realizar millones de análisis de regresión, se utilizó inteligencia artificial mediante la técnica de Algoritmos Genéticos (AGs) con el uso del software Moby Digs® versión 1.0-2004<sup>lix</sup>, en el que se utilizaron parámetros estandarizados que se muestran en la Tabla 5.

Para los criterios en el uso de los AGs, llevan a cabo cuatro evoluciones en todo el proceso donde finalmente se obtuvieron 3 descriptores respresentativos para cada modelo.

**Tabla 5.** Parámetros usados por Moby Digs 1.0®, por QSAR con AGs.<sup>45</sup>

Parámetros Evoluciones	Mínimo de descriptores por modelo	Máximo de descriptores por modelo	Número de ciclos por generación	Número de ciclos por variables	Reproducción/ Mutación	Tamaño de la muestra	Obtención de los Mejores modelos
------------------------	-----------------------------------	-----------------------------------	---------------------------------	--------------------------------	------------------------	----------------------	----------------------------------

**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

1, 2	2	7	2000	20000	0.8	100	10=100
3	2	5	2000	20000	0.8	70	50=100
4	2	3	2000	20000	0.8	50	50=50

Sin embargo de este proceso en Moby Digs, también se obtienen los siguientes parámetros estadísticos, tomados en cuenta para estimar la calidad del o los modelos seleccionados: <sup>lx, lxi</sup>

a) **Coefficiente de determinación, ( $r^2$ ):** Es la razón de la variación explicada entre la variación total se denomina coeficiente de determinación. Si existen cero variaciones explicadas, la razón es cero. Si hay cero variaciones inexplicadas, la razón es uno. En otros casos, la razón está entre cero y uno. Dado que la razón nunca es negativa, entonces se denota como ( $r^2$ ).

b) **Coefficiente de correlación múltiple al cuadrado ( $R^2$ ):** Es la varianza total de la respuesta se explica por un modelo de regresión. Se puede calcular a partir de la suma de los cuadrados residual RSS:

$$= 1 - \frac{\sum (e_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (e_i)^2}{\sum (y_i - \bar{y})^2}$$

Donde: SST es la suma total de cuadrados alrededor de la media. Un valor de uno indica un ajuste perfecto, es decir, un modelo con término de error cero.

En Moby Digs, ( $R^2$ ) se expresa como una cantidad por ciento por lo tanto en la estadística expresada en Excel se procedió a calcular la ( $r^2$ ) por 100%, para que los valores comparativos fueran homogéneos.<sup>45</sup>

c) **F de Fisher:** La distribución (F), determina si 2 conjuntos tienen diferentes grados de diversidad; por lo tanto representa los niveles de significancia; es decir determina si la relación observada entre las variables dependientes e independientes se produce por el azar.

d) **Desviación estándar (s):** Es la medida de dispersión más importante y se representa con la letra (s), con ella se puede ubicar la posición de una observación con relación a una serie de datos, es decir mide la dispersión de los valores respecto a la media.

e) **Q<sup>2</sup>:** Es la capacidad predictoria, deriva de la validación cruzada, es la destreza del modelo, para realizar la predicción, es la varianza explicada en la predicción:

$$Q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Donde PRESS es la suma de los cuadrados de error de predicción y TSS es la suma total de los cuadrados.

f) **Kx:** Es el coeficiente total de correlación, donde se muestra la evaluación de la relación que existe entre las variables del modelo.

$$Kx = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Donde  $\lambda$ , son los valores propios obtenidos a partir de la matriz de correlación del conjunto de datos X (n, p), siendo n el numero de objetos y p, el numero de variables.

g) **Q<sup>2</sup>boot**: También conocido como *bootstrap*, son técnicas de simulación que reutilizan los datos observados para constituir un universo del cual extraer repetidas muestras.

Todas las diferencias al cuadrado entre la respuesta verdadera y la respuesta calculada de los objetos del conjunto de evaluación son recopiladas en PRESS (suma de los cuadrados de los residuos predictivos), es la suma de cuadrados de las diferencias entre la respuesta observada y la estimada por técnica de validación.

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Donde:

$\hat{y}_i$  : La respuesta del objeto *i-th* estimado, mediante el uso de un modelo obtenido sin utilizar el objeto *i-th*.

Este procedimiento de construcción del conjunto de ensayo y evaluación del conjunto se repite miles de veces, PRESS se suma y el poder de predicción promedio es calculado y la capacidad de predicción promedio es calculada por (Q<sup>2</sup>boot). Finalmente esta técnica se propone como un procedimiento para estimar el error estadístico en cualquiera de sus formas.<sup>lxii</sup>

(5) De los 50 modelos obtenidos, se validaron manualmente la correlación entre descriptores y los parámetros estadísticos más importantes:  $R^2$ , F y s mediante el uso de Microsoft Excel®,<sup>lxiii</sup> donde el coeficiente de regresión o determinación ( $R^2$ ) con valores más cercanos al 100 representa el mejor ajuste de la regresión, es calculado para evaluar la capacidad predictiva del modelo, siendo una medida relativa de ajuste por la ecuación de regresión, que representa la parte de la variación en los datos observados, indica la buena capacidad predictiva del modelo QSAR. La prueba de Fisher (F), refleja la proporción de la varianza explicada por el modelo y la (s) es la varianza debido al error en la regresión, los valores altos de la prueba indican que el modelo es estadísticamente significativo. Por su parte la correlación entre descriptores representa la repetición de las variables.

(6) Se seleccionaron como mejores modelos aquellos que cumplieron con los parámetros de restricción utilizados en estudios QSAR<sup>lxiv</sup> estadísticamente significativos y que son:

$$R^2: 0-100; Q^2: 0-100; Q^2 \text{ Boot: } 0-100; Kx \leq 50; s: 0-1; F = *(\alpha 0.05)$$

\* Los datos de la (F) de Fischer crítica se toman de las tablas de referencia y dependen de los grados de libertad  $\alpha$  0.05, como rango considerando un 95% de confianza.

(7) La validación externa se llevó a cabo sustituyendo los valores de los compuestos seleccionados de la serie de exploración (D, J, L y R) en los modelos obtenidos y comparando su actividad calculada con la actividad experimental.

(8) La predicción de la actividad de cada compuesto propuesto potencialmente activo contra promastigotes de *Leishmania chagasi*, se llevó a cabo modelando sus estructuras, obteniendo los descriptores moleculares y sustituyendo sus valores en las ecuaciones obtenidas.<sup>lxv</sup>

## 6. RESULTADOS Y DISCUSIÓN.

Por estrategia metodológica es necesario que el diseño QSAR se realice con modelos matemáticos que contengan el menor número de descriptores posibles. Esto es importante sobre todo en términos de explicación, idealmente se considera adecuado un descriptor por cada cinco compuestos sometidos al estudio, sin embargo no es requisito fundamental siempre y cuando los modelos estén validados adecuadamente; así, se puso como restricción inicial en el software Moby Digs considerar modelos que contuviesen un máximo de tres descriptores. Los 50 modelos obtenidos se muestran en la Tabla 6, en donde además de sus principales descriptores estadísticos utilizados como criterio de “mejores modelos”, se aprecian: la  $R^2$  (Correlación entre el descriptor y la actividad), la  $Q^2$  (Capacidad predictiva del modelo), la  $Q^2$  boot (Capacidad predictiva optimizada) y la Kx (los índices multivariados de correlación entre los descriptores seleccionados).<sup>lxvi</sup>

**Tabla 6.** Los 50 modelos con sus descriptores y principales parámetros estadísticos.

No.	Tamaño	Modelos	R <sup>2</sup>	Q <sup>2</sup>	Q <sup>2</sup> boot	Kx	F	s
1	3	JGI2 Mor13u MLOGP	91.81	85.45	82.08	73.72	41.12	0.097
<b>2</b>	<b>3</b>	<b>SIC4 BELv3 PJI3</b>	<b>90.75</b>	<b>85.23</b>	<b>81.05</b>	<b>28.58</b>	<b>35.97</b>	<b>0.103</b>
<b>3</b>	<b>3</b>	<b>CIC4 BELv3 PJI3</b>	<b>90.68</b>	<b>85.18</b>	<b>81.5</b>	<b>28.69</b>	<b>35.67</b>	<b>0.104</b>
4	3	JGI2 Mor13u BLTD48	91.61	85.1	81.54	73.66	40.03	0.098
5	3	RDF100u RDF095p Mor21v	86.46	82.51	57.18	60.57	23.42	0.125
6	3	JGI8 RDF095p H-047	87.41	80.47	71.24	49.31	25.46	0.12
7	3	GGI4 RDF100u RDF095p	86.24	80.36	46.87	39.75	22.98	0.126
8	3	GATS4e RDF095p O-060	89.09	80.31	74.58	58.75	29.93	0.112
9	3	BEHe6 BELe6 RDF095p	85.94	79.98	0	35.8	22.41	0.127
<b>10</b>	<b>3</b>	<b>XOA EEig11r RDF095p</b>	<b>88.64</b>	<b>79.97</b>	<b>73.14</b>	<b>28.69</b>	<b>28.6</b>	<b>0.114</b>
11	3	AMW RDF100u RDF095p	85.23	79.92	54.35	45.7	21.16	0.13
12	3	RDF100u RDF060p RDF095p	85.76	79.87	55.47	31.88	22.08	0.128
13	3	BELv3 RDF100u RDF095p	90	79.42	68.81	63.15	32.98	0.107
14	3	X4sol MATS4p GATS4e	88.08	78.85	72.78	25.12	27.11	0.117
15	3	PCR X4sol MATS4p	85.63	78.65	71.75	20.84	21.85	0.129
16	3	PCR Mor13u BLTD48	87.47	78.25	71.34	43.43	25.6	0.12
17	3	X2A PJI3 Qpos	83.97	78.25	44.44	37.21	19.21	0.136
18	3	X2A PJI3 Qneg	83.97	78.25	45.3	37.21	19.21	0.136
19	3	PCR Mor13u MLOGP	87.48	78.21	74.53	43.47	25.63	0.12
20	3	RDF100u RDF095p R3p+	83.3	78.05	36.16	53.58	18.29	0.139
21	3	MATS4p GATS4m GATS4e	86.9	78.05	70.01	44.72	24.32	0.123
22	3	GATS8e RDF100u RDF095p	84.25	78.03	48.56	37.87	19.61	0.135
23	3	X4sol RDF100u RDF095p	84.8	78.01	42.12	34.54	20.45	0.132
24	3	MATS4p RDF100u RDF095p	84.66	77.87	43.8	43.96	20.24	0.133
25	3	RDF100u RDF095p BLTD48	84.59	77.79	36.04	58.68	20.12	0.133
26	3	RDF100u RDF095p O-060	82.9	77.62	30.34	68.49	17.78	0.14
27	3	X1sol X4sol JGI9	87.83	77.61	71.43	48.26	26.45	0.118
28	2	RDF100u RDF095p	82.38	77.32	63.74	60.1	28.06	0.136
29	3	BELe6 RDF100u RDF095p	83.36	77.24	31.92	55.7	18.36	0.138
30	3	GATS4m RDF100u RDF095p	84.23	77.13	48.38	39.38	19.59	0.135
31	3	X1sol RDF100u RDF095p	84.83	76.98	50.66	35.84	20.51	0.132
32	3	PJI3 RDF100u RDF095p	82.87	76.58	26.87	57.41	17.74	0.14
33	3	X4sol JGI9 HGM	84.82	76.51	0	12.23	20.49	0.132

**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

34	3	SEigv RDF100u RDF095p	83.06	76.38	31.09	38.19	17.98	0.14
35	3	RDF100u RDF095p H-047	83.16	75.72	34.58	63.13	18.11	0.139
36	3	EEig11r RDF100u RDF095p	84.2	75.56	50.07	45.76	19.54	0.135
37	3	RDF100u RDF095p Mor22p	82.57	75.4	0	34.69	17.37	0.142
38	3	JGI8 RDF100u RDF095p	83.97	74.78	49.82	45.6	19.2	0.136
39	3	BELv3 RDF095p H-047	86.54	74.71	66.73	54.1	23.57	0.124
40	3	PCR BELv3 PJI3	82.13	74.58	58.12	30.98	16.85	0.143
41	3	LP1 RDF100u RDF095p	82.38	74.22	17.03	35.73	17.15	0.142
42	3	Pol RDF100u RDF095p	82.39	74.2	11.11	31.62	17.15	0.142
43	3	VEA1 RDF100u RDF095p	82.4	73.96	18.75	32.39	17.16	0.142
44	3	RDF100u RDF095p Mor13u	82.42	73.95	30.49	59.4	17.2	0.142
45	3	XOA RDF100u RDF095p	82.46	73.91	25.71	35.9	17.23	0.142
46	3	X4sol MATS4p JGI9	85.74	73.63	69.6	31.44	22.05	0.128
47	2	BELv3 PJI3	79.16	73.17	68.77	39.82	22.79	0.148
48	3	BEHe6 RDF100u RDF095p	82.39	72.83	13.69	44.08	17.15	0.142
49	3	Pol RDF095p H-047	81.57	72.76	0	36.03	16.23	0.146
50	1	JGI9	46.52	31.29	32.9	0	11.31	0.228

En color café se representan los modelos seleccionados con mejor estadística.

Los 50 modelos obtenidos por Moby Digs<sup>45</sup>, son considerados como “mejores modelos” de acuerdo a los parámetros establecidos en el programa. La mayoría de ellos contienen valores estadísticamente aceptables. Sin embargo, al tratarse de un método estadístico resulta conveniente llevar a cabo estrategias adicionales para la selección de los modelos útiles. Uno de ellos es la baja correlación entre descriptores que está dado por el valor (Kx), el cual debe ser menor de 50, ésta baja correlación se puede apreciar en la matriz de correlación de la Tabla 7, donde se hace un estudio comparativo con todos los descriptores que participan en la validación, marcados con un color distintivo los Modelos 2, 3 y 10 respectivamente, mostrando su baja correlación entre los descriptores de cada modelo seleccionado.

**Tabla 7.** Matriz de correlación total de descriptores.

DESCRIPTORES	SIC4	BELv3	PJ13	CIC4	BEHe6	BELe6	RDF095p	X0A	EEig11r	X4sol	MATS4p	GATS4e	PCR	JGI9	HGM	PoI	H-047
<b>SIC4</b>	1.000																
<b>BELv3</b>	<b>-0.226</b>	1.000															
<b>PJ13</b>	<b>0.219</b>	<b>0.398</b>	1.000														
<b>CIC4</b>	-0.999	<b>0.246</b>	<b>-0.203</b>	1.000													
<b>BEHe6</b>	0.320	0.452	0.468	-0.303	1.000												
<b>BELe6</b>	-0.463	0.528	0.065	0.473	-0.365	1.000											
<b>RDF095p</b>	-0.217	0.441	0.727	0.234	0.441	0.441	1.000										
<b>X0A</b>	0.011	-0.206	0.001	0.001	-0.641	-0.641	<b>0.160</b>	1.000									
<b>EEig11r</b>	-0.176	0.521	0.064	0.200	0.705	0.705	<b>0.219</b>	<b>0.448</b>	1.000								
<b>X4sol</b>	0.054	-0.342	0.204	-0.063	0.277	-0.641	0.033	0.551	0.033	1.000							
<b>MATS4p</b>	0.305	-0.631	0.202	-0.326	-0.363	-0.327	-0.040	0.002	-0.793	0.351	1.000						
<b>GATS4e</b>	0.333	-0.064	-0.302	-0.355	-0.121	-0.208	-0.479	-0.513	-0.348	0.113	0.270	1.000					
<b>PCR</b>	0.290	0.242	-0.281	-0.301	0.156	-0.211	-0.456	-0.368	0.032	0.094	-0.139	0.863	1.000				
<b>JGI9</b>	-0.036	-0.373	0.331	0.040	-0.358	0.197	0.437	0.260	-0.338	-0.113	0.449	-0.585	-0.836	1.000			
<b>HGM</b>	0.346	-0.780	0.331	-0.375	-0.413	-0.502	-0.674	-0.145	-0.663	0.213	0.660	0.464	0.198	0.008	1.000		
<b>PoI</b>	0.005	0.033	0.315	0.008	0.682	-0.604	0.153	0.923	0.643	0.615	-0.194	-0.359	-0.118	-0.072	-0.257	1.000	
<b>H-047</b>	-0.217	0.743	0.163	0.232	0.078	0.845	0.416	-0.638	0.209	-0.594	-0.500	-0.083	0.025	-0.157	-0.587	-0.487	1.000

Modelo2  
Modelo3  
Modelo10

Así mismo, se llevó cabo un análisis estadístico independiente de Moby Digs, que dedujo mediante algún programa de cálcula básico como Excel, la correlación entre los descriptores por cada modelo, sometiéndose a la comparación entre la actividad experimental y la actividad calculada de cada modelo seleccionando los mejores. Los resultados se desglosan en el siguiente apartado.

### **6.1 Caracterización de los modelos seleccionados por estudio QSAR, para actividad biológica contra promastigotes de *L. chagasi*.**

Una vez elegidos los modelos que mostraron independencia de los descriptores participantes, se corroboraron, mediante una regresión lineal, los parámetros de  $R^2$ , F de Fisher, buscando los valores de 95% de confianza, constante de la ecuación, así como la desviación estándar. Otro parámetro de elección de modelos fue la determinación de la actividad calculada por cada modelo contra la observada experimentalmente debiendo existir congruencia y confirmando esta característica en cada gráfica. De esta forma, se eliminaron modelos resultando útiles únicamente los Modelos 2, 3, 10; los cuales se analizan y detallan a continuación:

#### **I. Análisis del Modelo 2:**

##### **Ec.1**

$$\text{Log} - (1/CI_{50}) = 3.775 \text{ SIC } (\pm 1.610) + 63.471 \text{ BELv3 } (\pm 7.110) - 3.750 \text{ PJI3 } (\pm 0.545) - 106.124$$

$$n = 15; R^2 = 92.67; F = 29.538; s = 0.103; Q^2 = 85.23; Kx = 28.58$$

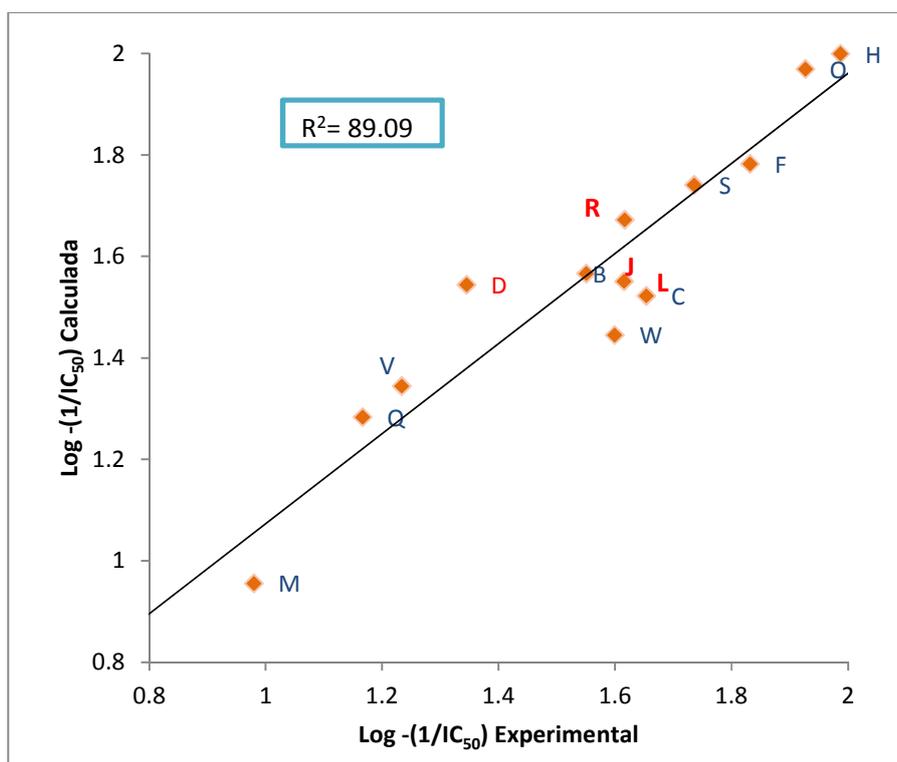
Los resultados estadísticos para el Modelo 2, en la actividad experimental y la calculada se muestran en la Tabla 8, donde se representan marcados de color verde los valores de los compuestos, el (M) con IC<sub>50</sub> experimental de 9.54 y la observada de 9.02 y el compuesto (L) IC<sub>50</sub> experimental de 14.1 y observada de 14.47, la diferencia en estos compuestos es poco significativa, indicando buena correlación entre sus actividades.

**Tabla 8.** Comparación entre la IC<sub>50</sub> experimental y la IC<sub>50</sub> calculada en el Modelo 2.

Descriptores			Compuestos	Actividad	
SIC4	BELv3	PJ13		IC <sub>50</sub> ( ) Experimental	IC <sub>50</sub> ( ) Calculada
0.919	1.689	0.795	B	35.5	36.815
0.933	1.687	0.787	C	45.04	33.261
0.933	1.681	0.616	F	67.89	60.590
0.958	1.699	0.888	H	97.09	99.816
0.978	1.679	0.848	M	9.54	9.021
0.978	1.682	0.795	N	17.11	22.103
0.967	1.686	0.685	O	84.44	93.184
0.956	1.688	0.787	Q	41.34	47.015
0.978	1.687	0.774	S	54.45	55.0266
0.933	1.682	0.766	V	14.66	19.201
0.978	1.684	0.802	W	39.75	27.871
0.933	1.687	0.781	D	22.12	35.030
0.978	1.682	0.74	J	41.23	35.540
0.967	1.681	0.816	L	14.1	14.478
0.978	1.685	0.776	R	50.22	40.376

Los valores obtenidos para los compuestos D, J, L, R, (marcados de color naranja), que fueron los seleccionados para la validación ciega, en el Modelo 2 (Tabla 8), se observa que la actividad calculada se encuentra poco alejada de la actividad experimental, mostrando diferencias poco significativas, partiendo de que para estos compuestos la Ecuación 1 (Ec.1), es representativa. Así mismo, los

datos experimentales y los observados tienen poca dispersión respecto a los valores con el resto de los compuestos (Gráfica 1). Los datos tienden a la línea recta, lo cual aumenta el valor estadístico de este Modelo, probando que se puede considerar como útil para predecir actividad de compuestos diseñados no incluidos en el estudio.



**Gráfica 1.** Relación de  $\text{Log}-(1/IC_{50})$  Exp. vs  $\text{Log}-(1/IC_{50})$  Calc. Modelo 2.

**Tabla 9.** Coeficiente de correlación Modelo 2.

Descriptor	SIC4	BELv3	PJI3
SIC4	1		
BELv3	0.0186	1	
PJI3	0.0829	0.1962	1

Adicionalmente, en la Tabla 9 se muestra la baja correlación entre los descriptores que constituyen el Modelo 2. Es importante que sus características fisicoquímicas correspondan a distintas familias, de esta forma reduciendo la posibilidad de ambigüedad.

Con la finalidad de ponderar los descriptores en el Modelo 2, se fragmentó e hizo un análisis de regresión para cada uno de ellos o bien la combinación de dos de ellos. En la Tabla 10 se muestran los valores obtenidos de cada fragmentación. En el caso del Modelo 2.2, se identificó que la combinación de los descriptores BELv3 y PJI3, con una  $R^2 = 89.09$ ,  $F = 26.615$  y la desviación estándar ( $s$ )= 0.129, es la fragmentación que estadísticamente tiene los mejores valores, por lo que este análisis llevó a pensar que estos descriptores son los determinantes en la actividad para la serie de compuestos benzoazepínicos contra promastigotes *L.chagasi*.

**Tabla 10.** Fragmentación del Modelo 2 para Promastigotes de *L.chagasi*.

<b>Modelo</b>	<b>Número de descriptores</b>	<b>Descriptores</b>	<b>R<sup>2</sup></b>	<b>F</b>	<b>s</b>
<b>2</b>	3	SIC4, BELv3, PJI3	92.667	29.538	0.103
<b>2.1</b>	2	SIC4, BELv3,	43.77	3.052	0.270
<b>2.2</b>	2	BELv3, PJI3	86.93	26.615	0.129
<b>2.3</b>	2	SIC4, PJI3	09.33	0.412	0.342
<b>2.4</b>	1	SIC4	01.65	0.151	0.335
<b>2.5</b>	1	BELv3	43.12	6.824	0.255
<b>2.6</b>	1	PJI3	09.15	0.906	0.322

Aunque los modelos obtenidos tienen sentido matemático, resulta importante dar interpretación a los descriptores. En el caso de estudios QSAR basados en descriptores teóricos, dichos descriptores suelen resultar del producto de una compleja ecuación matemática difícil de interpretar químicamente. Sin embargo, es conveniente, en la medida de lo posible, dar una interpretación práctica de la presencia de los descriptor en la ecuación.

## I.I Descriptores correspondientes al Modelo 2.

En el Modelo 2, sus descriptores **SIC4**, **BELv3** y **PJI3**, pertenecen a las familias: Descriptores del contenido de información estructural, Descriptores propios del valor de carga y Descriptores Geométricos, respectivamente.<sup>67</sup>

La categoría del descriptor **SIC4**, es en la Familia 5: “Contenido de Información Estructural”. Se refiere a los índices de contenido de información, se calculan sobre la base de los átomos de pares de equivalencia en una molécula llena de hidrógenos. Un par de átomos se dice que es equivalente a un determinado nivel-*r*, si son del mismo elemento y su vecindad es equivalente al nivel-*r*. La equivalencia de las vecindades de un par de átomos *a* y *b* es equivalente al nivel-*r*, donde se define como sigue: Sea conjunto (*a*) el conjunto de caminos de longitud hasta (*r*) procedente del átomo (*a*). De manera similar, el conjunto (*b*) correspondiente a un átomo de (*b*) se define. Entonces los átomos *A* y *B* se dice que son equivalentes al nivel-*r*, si existe un uno-a-uno sobre asignación entre los conjuntos establecidos (*a*) y el conjunto (*b*). Decimos que existe una asignación entre un par de rutas de si los átomos correspondientes y los bonos son los mismos en ambas de las trayectorias. Por ejemplo, si la ruta de CC = NO existe en cada uno del conjunto conjuntos (*a*) y el conjunto (*b*), entonces estos dos caminos pueden ser asignados entre sí.<sup>68</sup>

El descriptor **SIC4**, corresponde a los índices de “Información relacionada con la simetría vecinal de orden 4” (cuatro átomos de carbono vecinos). Considera a los valores numéricos moleculares obtenidos por la cuantificación de diversas características estructurales y fisicoquímicas del compuesto, por lo que se prevee que pueda cuantificar estos atributos con el fin de determinar el comportamiento de la molécula y la forma en que ésta interactúa unida a un sistema fisiológico.<sup>69</sup>

Por su parte, el descriptor **BELv3**, pertenece a la Familia 8, su categoría es “Descriptores propios del valor de carga”. Dirige los valores propios de la carga y

se basa en las electronegatividades de las fuerzas electrostáticas de Van der Waals (VdW). Diversas características de los descriptores que pertenecen a esta familia son seleccionados para los QSAR 2D, son: VdW superficie de la zona, Van der Waals superficie de la molécula, (s), total de Van der Waals superficie con potencial electrostático negativo de la molécula, (-p), total de Van der Waals superficie con potencial electrostático positivo de la molécula, (+p), momento dipolar, el recuento de elementos, recuento de ruta, conglomerado, distancia basada en índices topológicos, índice de conectividad en áreas hidrófobas e hidrófilas, es decir, el valor más alto en la superficie VdW, distancia entre el punto más hidrofóbico e hidrofílico en la superficie de contacto de VdW, donde las interacciones polares pueden tener lugar y se caracterizan por una distribución específica electrónica obtenida mediante la asignación atómica, y las cargas parciales en la superficie del disolvente accesible.<sup>70</sup>

Para el descriptor BELv3, es la superficie la que envuelve esferas fusionadas y rígidas centradas en las coordenadas del átomo (núcleos atómicos), teniendo radios iguales a algunos de los valores recomendados de los radios de Van der Waals. Las esferas interaccionan unas con otras de tal manera que la distancia entre los centros de dos esferas es igual a la longitud del enlace formal. La superficie de Van der Waals (también conocida como superficie total molecular, TSA) se define entonces como la superficie exterior de la unión de todas las esferas de este tipo en la molécula, es decir, el área de la superficie de Van der Waals molecular (m). De manera análoga, la de Van der Waals volumen (V), también llamada molecular, intrínseco.<sup>71</sup>

El descriptor **PJ13**, pertenece a la Familia 12 y corresponden los “Descriptores geométricos 3D”. En un descriptor topológico, también llamado descriptor molecular 3D, se deriva de una representación geométrica de la molécula, más específicamente de las coordenadas x, y, z, donde se posicionan los átomos de la molécula.<sup>56</sup>

Éste es un descriptor molecular definido de varias maneras, pero siempre se deriva de la estructura tridimensional de la molécula. En general, los descriptores geométricos se calculan ya sea en una geometría optimizada molecular obtenida por el método de la química computacional o coordenadas por índices topográficos que lo constituyen como subconjunto especial de descriptores geométricos, se calcula sobre la representación gráfica de las moléculas, pero usando las distancias geométricas entre átomos en lugar de las distancias topológicas.

Una representación geométrica, implica el conocimiento de la relatividad de las posiciones de los átomos en el espacio 3D, los descriptores geométricos suelen proporcionar más información y poder de discriminación para estructuras moleculares similares y conformaciones de moléculas. A pesar de su alto contenido de información, los descriptores geométricos suelen presentar algunos inconvenientes; requieren geometría optimización, donde se intenta localizar un mínimo de energía potencial en la superficie y así conseguir la estructura de equilibrio de un sistema molecular. Además, para moléculas flexibles, varias conformaciones de la molécula pueden estar disponibles.<sup>56, 72</sup>

Por estas razones, los descriptores topológicos, “huellas dactilares” basados en el recuento de fragmentos y otros descriptores simples, se prefieren generalmente para la detección de grandes bases de datos de moléculas y la búsqueda de relaciones entre las estructuras moleculares y propiedades complejas, tales como actividades biológicas, a menudo pueden realizarse de manera eficiente mediante el uso de descriptores geométricos, explotando su contenido de información general.

La mayoría de los descriptores geométricos se calculan directamente a partir la representación geométrica de la molécula y otras magnitudes derivadas de las coordenadas, como distancias interatómicas o distancias de un origen específico, muchos de ellos son derivados de la matriz de geometría molecular definida por todas las distancias geométricas entre pares de átomos.

$$\begin{matrix} & 0 & \dots & 12 & \dots & 1 \\ \equiv & 21 & \dots & 0 & \dots & 2 \\ & & & 1 & \dots & 2 & \dots & 0 \end{matrix}$$

La matriz de geometría, dada por (G), es una representación molecular en donde los átomos se consideran como puntos individuales en el espacio 3D molecular. Es una matriz cuadrada simétrica ( $A \times A$ ), siendo (A) el número de átomos de la molécula, donde cada entrada (r) es la distancia euclídea (espacios geométricos) entre los átomos.<sup>56,57</sup>

## II. Análisis del Modelo 3:

### Ec. 2

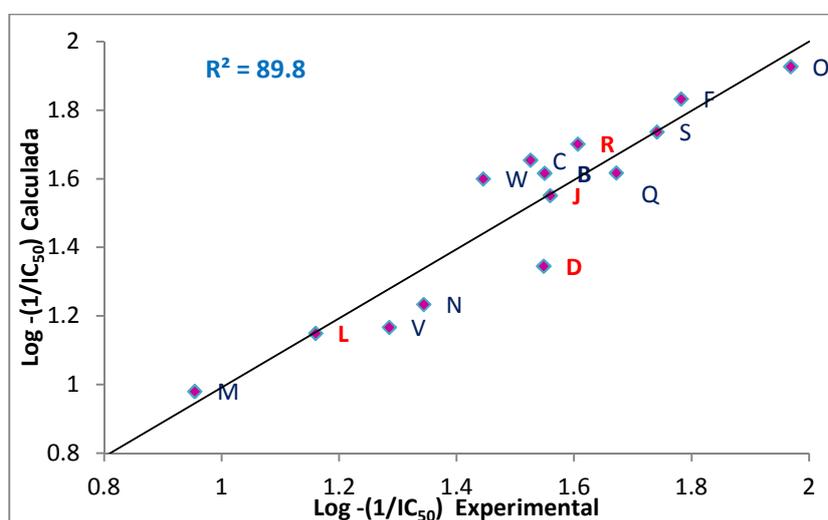
$\text{Log-}(1/CI_{50}) = -0.730 \text{ CIC4 } (\pm 0.307) + 63.794 \text{ BELv3 } (\pm 7.108) - 3.740 \text{ PJI3 } (\pm 0.540) - 102.902$
---

$n = 15; R^2 = 92.76; F = 29.897; Q^2 = 90.68; Kx = 28.69; s = 0.103$
---

En la Tabla 11, se observan los valores de actividad experimental y actividad calculada, donde el compuesto (L) con actividad experimental de 14.1 y actividad observada de 14.45 y el compuesto (M) con actividad experimental de 9.54 y 9.00 de actividad calculada, demuestran la capacidad predictiva del modelo. Así mismo, los valores de los compuestos seleccionados para su validación, se muestran en la Gráfica 2, donde tienden a la línea recta, lo cual aumenta el valor predictivo de este Modelo 3, ya que ambas actividades están correlacionadas ( $R^2 = 89.8$ ).

**Tabla 11.** Comparación entre la Actividad experimental ( $IC_{50}$ ) y la Actividad calculada ( $IC_{50}$ ) para el Modelo 3.

Descriptores			Actividad		
CIC4	BELv3	PJI3	Compuestos	$IC_{50}$ ( ) Experimental	$IC_{50}$ ( ) Calculada
0.43	1.689	0.795	B	35.5	36.287
0.343	1.687	0.787	C	45.04	33.546
0.343	1.681	0.616	F	67.89	60.593
0.224	1.699	0.888	H	97.09	100.089
0.114	1.679	0.848	M	9.54	9.004
0.114	1.682	0.795	N	17.11	22.083
0.171	1.686	0.685	O	84.44	93.109
0.23	1.688	0.787	Q	41.34	46.989
0.114	1.687	0.774	S	54.45	55.153
0.343	1.682	0.766	V	14.66	19.284
0.114	1.684	0.802	W	39.75	27.891
0.343	1.687	0.781	D	22.12	35.325
0.114	1.682	0.74	J	41.23	35.461
0.171	1.681	0.816	L	14.1	14.457
0.114	1.685	0.776	R	50.22	40.411



**Gráfica 2.** Correlación de  $\text{Log}-(1/IC_{50})$  Exp. vs  $\text{Log}-(1/IC_{50})$  Calc. Modelo 3.

**Tabla12.** Coeficiente de correlación Modelo 3.

<b>Descriptor</b>	<b>CIC4</b>	<b>BELv3</b>	<b>PJI3</b>
<b>CIC4</b>	1		
<b>BELv3</b>	0.0257	1	
<b>PJI3</b>	0.1962	0.1962	1

Adicionalmente en la Tabla 12, se muestra el coeficiente de correlación de los descriptores para el Modelo 3, donde efectivamente no existe correlación entre ellos, aunado a esta información se llevó a cabo la fragmentación del Modelo 3 (Tabla 13), que mostró en el Modelo 3.2 la mejor estadística, con una  $R^2 = 86.93$ ,  $F = 26.615$  y  $s = 0.1$ , al igual que el Modelo 2, el descriptor BELv3 es el más relevante.

**Tabla13.** Fragmentaciones para el Modelo 3.

<b>Modelo</b>	<b>Número de descriptores</b>	<b>Descriptor</b>	<b><math>R^2</math></b>	<b>F</b>	<b>s</b>
<b>3</b>	3	CIC4, BELv3, PJI3	92.76	29.897	0.103
<b>3.1</b>	2	CIC4, BELv3,	43.22	3.044	0.270
<b>3.2</b>	2	BELv3, PJI3	86.93	26.615	0.129
<b>3.3</b>	2	CIC4, PJI3	09.47	0.418	0.341
<b>3.4</b>	1	CIC4	01.84	0.168	0.335
<b>3.5</b>	1	BELv3	43.12	6.824	0.255
<b>3.6</b>	1	PJI3	09.15	0.906	0.073

## II.I Descriptores correspondientes al Modelo 3.

La similitud entre los Modelos 2 y 3 es evidente y se espera un comportamiento similar entre ambos. La diferencia se encuentra en los descriptores CIC4 y SIC4 que aunque pertenecen a la misma Familia 5, que agrupa descriptores de “contenido de información estructural”, existen algunas pequeñas diferencias que se pueden encontrar en los índices de simetría de vecindad (es decir, IC, TIC, SIC, BIC, CIC), cuando se calcula para moléculas pequeñas. En particular, para TIC, **SIC**, BIC, y **CIC** de pedido grande (4 y 5), cero en lugar del código de valor faltante se harán si el trayecto de orden correspondiente no existe en la molécula.<sup>57</sup>

La representación de la molécula es utilizada para derivar descriptores, congéneres. De manera particular el descriptor **CIC4**: Se refiere al “contenido de información complementaria (simetría barrio de la 4-fin)”.

Dado que el mecanismo exacto de la actividad del fármaco en muchos casos es desconocido, son deseables descriptores que tengan atributos sobre los compuestos, de esta forma evaluar su capacidad de predecir la actividad. En particular este descriptor se basa en los índices de información respecto al sector de la simetría del átomo, de “orden vecinal 4”, combina la forma y la información electrónica para caracterizar las moléculas y por lo tanto codifica características responsables de las interacciones entre las moléculas polares.<sup>53,57</sup>

### III. Modelo 10:

#### Ec. 3

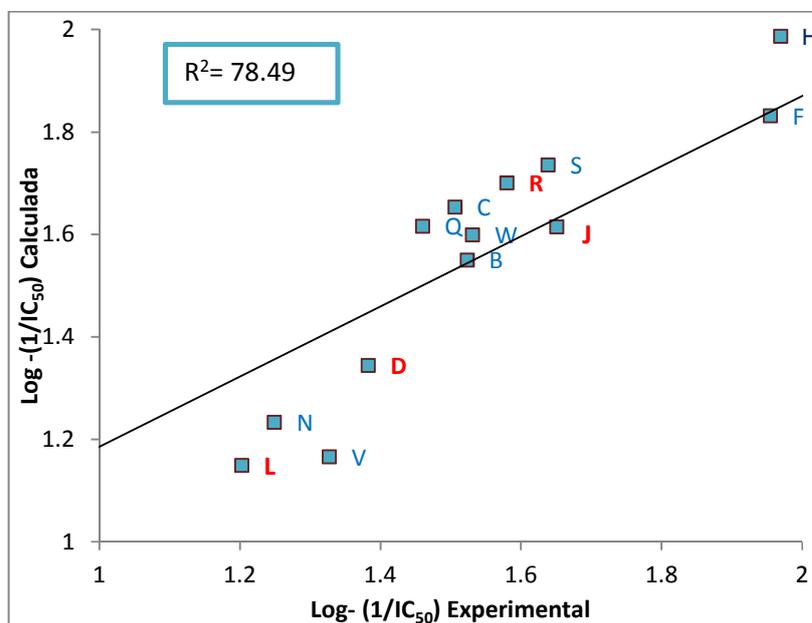
$$\text{Log}(1/CI_{50}) = -36.659 X0A (\pm 6.533) + 1.686 \text{EEig11r} (\pm 0.030) - 0.192 \text{RDF095p} (\pm 0.074) + 26.597$$

$$n = 15; R^2 = 87.73; F = 16.683; Q^2 = 79.97; Kx = 28.69; s = 0.134$$

La validación externa con los compuestos D, J, L, R, en el Modelo 10, resultó, aproximada a los valores experimentales como se aprecian en la Tabla 14, aunque más alejada de los valores promedio obtenidos en los Modelos 2 y 3. La Gráfica 3, muestra la correlación entre las actividades experimental y calculada, donde se aprecia que, aunque los valores tienden a la recta existe cierta dispersión se consideran como “outlier” es decir un puntos que distan del resto de los datos en un análisis estadístico. De acuerdo con lo anterior el Modelo 10 no es el mejor modelo para predecir actividad biológica, sin embargo para corroborarlo se deben tomar en cuenta todos los aspectos estadísticos para validación del Modelo en cuestión.

**Tabla 14.** Comparación entre la Actividad experimental ( $IC_{50}$ ) y Actividad calculada ( $IC_{50}$ ) para el Modelo 10.

X0A	Descriptores		Compuestos	Actividad	
	EEig11r	RDF095p		$IC_{50}$ ( ) Experimental	$IC_{50}$ ( ) Calculada
0.715	0.857	1.597	B	35.5	33.355
0.697	0.354	0.709	C	45.04	32.062
0.697	0.61	0.622	F	67.89	90.032
0.697	0.803	2.235	H	97.09	93.127
0.706	0.424	1.682	M	9.54	12.783
0.706	0.44	1.083	N	17.11	17.742
0.697	0.591	0.142	O	84.44	103.475
0.706	0.591	1.309	Q	41.34	28.845
0.706	0.656	0.952	S	54.45	43.495
0.713	0.657	1.243	V	14.66	21.254
0.713	0.693	0.502	W	39.75	33.955
0.697	0.358	1.384	D	22.12	24.137
0.697	0.438	0.692	J	41.23	44.761
0.706	0.439	1.315	L	14.1	15.945
0.706	0.591	0.687	R	50.22	38.010



**Gráfica 3.** Relación de  $\text{Log}-(1/IC_{50})$  Exp vs  $\text{Log}-(1/IC_{50})$  Calc. Modelo 10.

Adicionalmente en la Tabla 15, se muestra la relación de los descriptores que constituyen el Modelo matemático, infiriendo que no hay correlación significativa entre ellos.

**Tabla15.** Coeficiente de correlación Modelo 10.

<b>Descriptores</b>	<b>X0A</b>	<b>EEig11r</b>	<b>RDF095p</b>
<b>X0A</b>	1		
<b>EEig11r</b>	0.1258	1	
<b>RDF95p</b>	0.0317	0.1047	1

La fragmentación del Modelo 10, no arroja ningún dato significativo que indique la ponderación de algún descriptor aunque se podría especular que el descriptor X0A es el que más peso tiene en la ecuación. (Tabla 16)

**Tabla 16.** Fragmentación Modelo 10.

<b>Modelo</b>	<b>Número de descriptores</b>	<b>Descriptores</b>	<b>R<sup>2</sup></b>	<b>F</b>	<b>s</b>
<b>10</b>	3	X0A, EEig11r, RDF095p	87.73	16.683	0.134
<b>10.1</b>	2	X0A, EEig11r	76.15	12.773	0.175
<b>10.2</b>	2	EEig11r, RDF095p	32.54	1.929	0.295
<b>10.3</b>	2	X0A, RDF095p	34.79	2.134	0.290
<b>10.4</b>	1	X0A	32.83	4.400	0.277
<b>10.5</b>	1	EEig11r	16.97	1.840	0.308
<b>10.6</b>	1	RDF095p	05.75	0.549	0.328

### III.I Descriptores correspondientes al Modelo 10.

El descriptor **X0A** es un “Indicador Topológico” que corresponde a los “Índices promedio de conectividad” de la Familia 4, específicamente a los índices de conectividad Kier-Hall que es obtenido a partir de gráficas moleculares que excluyen todos los átomos de hidrógeno. Sus siglas provienen de “Average connectivity index chi-0” (X0A).<sup>56,57</sup>

La conectividad índice Chi-0 a Chi-5, es el promedio del índice de conectividad a través de Chi Chi-0-5 3. Se refiere a un balance en el índice de conectividad a través de Chi-0-5. Promedio de balance de conectividad índice Chi-0 a Chi-5, conectividad índice Chi-0 se define como:

$$D_{ij}^* = \sum_{k \in \text{path}(i,j)} \frac{1}{\text{length}(k)}$$

Donde: la suma se extiende a través de la secuencia de bonos (bordes) a lo largo de la ruta más corta entre los átomos i y j.

Grado multigrafo, la distancia del átomo i-ésimo se define como la suma de la fila correspondiente de la matriz de distancia multigrafo:

$$D_i^* = \sum_{j=1}^n D_{ij}^*$$

El descriptor **EEig11r**: Pertenece a la Familia 7, se refiere los “Índices del borde de adyacencia”, con autovalor de 11 desde el borde adyacente hasta la matriz que pondera la resonancia del compuesto. Es un descriptor topológico molecular de autovalor 11, desde el borde adyacente derivado de la matriz ponderada por integrales de resonancia, que codifica la conectividad entre los bordes de un gráfico.<sup>55</sup>

Los descriptores **RDF095p**, forman parte de la Familia 13, sus siglas significan “Factor de Distribución Radial” que corresponden a las integrales valoradas en la polarizabilidad atómica 9.5 p. Estas descripciones se basan en la distribución de distancia geométrica de la representación de una molécula y que constituyen un código de función de distribución radial (RDF) que muestra ciertas características en común con el 3D-MorSE. Formalmente, la función de distribución radial de un conjunto de átomos puede ser interpretado como la distribución de probabilidad de encontrar un átomo en un volumen esférico de radio. Esta característica es la única respecto a la disposición tridimensional de los átomos, y es invariante contra traslación y rotación de la molécula completa. Además, el código RDF puede ser restringido a los tipos de átomos específicos o rangos de distancia para un espacio con cierta estructura tridimensional, como por ejemplo las distancias entre los enlaces y tipos de átomos, así como para describir un impedimento estérico o estructura y/o las propiedades de actividad de un compuesto.

La función de distribución radial muy a menudo se utiliza para la interpretación de los patrones de difracción obtenidos en los experimentos de difracción de polvo de rayos-X. Al incluir características de las propiedades atómicas, los descriptores permiten la discriminación de los átomos de una molécula para casi todas las propiedades que se pueden atribuir a un átomo.

La función de distribución radial en esta forma cumple con todos los requisitos para un descriptor de estructura 3D: es independiente del número de átomos, es decir, el tamaño de una molécula, es único en cuanto a la disposición tridimensional de los átomos, e invariante contra traslación y la rotación de la molécula entera. Por otra parte, el descriptor RDF vectorial es interpretable mediante el uso de reglas simples y, por lo tanto, proporciona una posibilidad de decodificación reversible. Además de la información sobre distribución de las distancias interatómicas en la molécula completa, el vector de RDF proporciona una valiosa información adicional, como las distancias de enlace, los sistemas planos y no planos, así como los tipos de átomos. En este hecho es un descriptor

que toma en consideración la elucidación códigos asistida por un ordenador. <sup>34, 56,57</sup>

La función de distribución radial de un conjunto de (n) átomos puede ser interpretada como la distribución de la probabilidad de encontrar un átomo en un volumen esférico de radio (R). Un típico descriptor RDF se denota por  $RDF_{sw}$  donde, (s) y (w) toman los valores  $10 \leq s \leq 155$  en 5 unidades de (y), se define de la siguiente manera:

$$f(r) = \frac{1}{n} \sum_{i,j} w_i w_j \delta(r - r_{ij})$$

Donde:

(f) es un factor de escala,

(r) es la distancia euclídea entre los átomos (i) y (j),  $w_i$  y  $w_j$  son los pesos de los átomos respectivamente,

(n) es el número total de átomos,

( $\beta$ ) es el parámetro de suavizado que define la distribución de probabilidad de que el individuo inter-atómico distancia, ( $\beta$ ) se puede interpretar como el factor de temperatura que define el movimiento de los átomos.

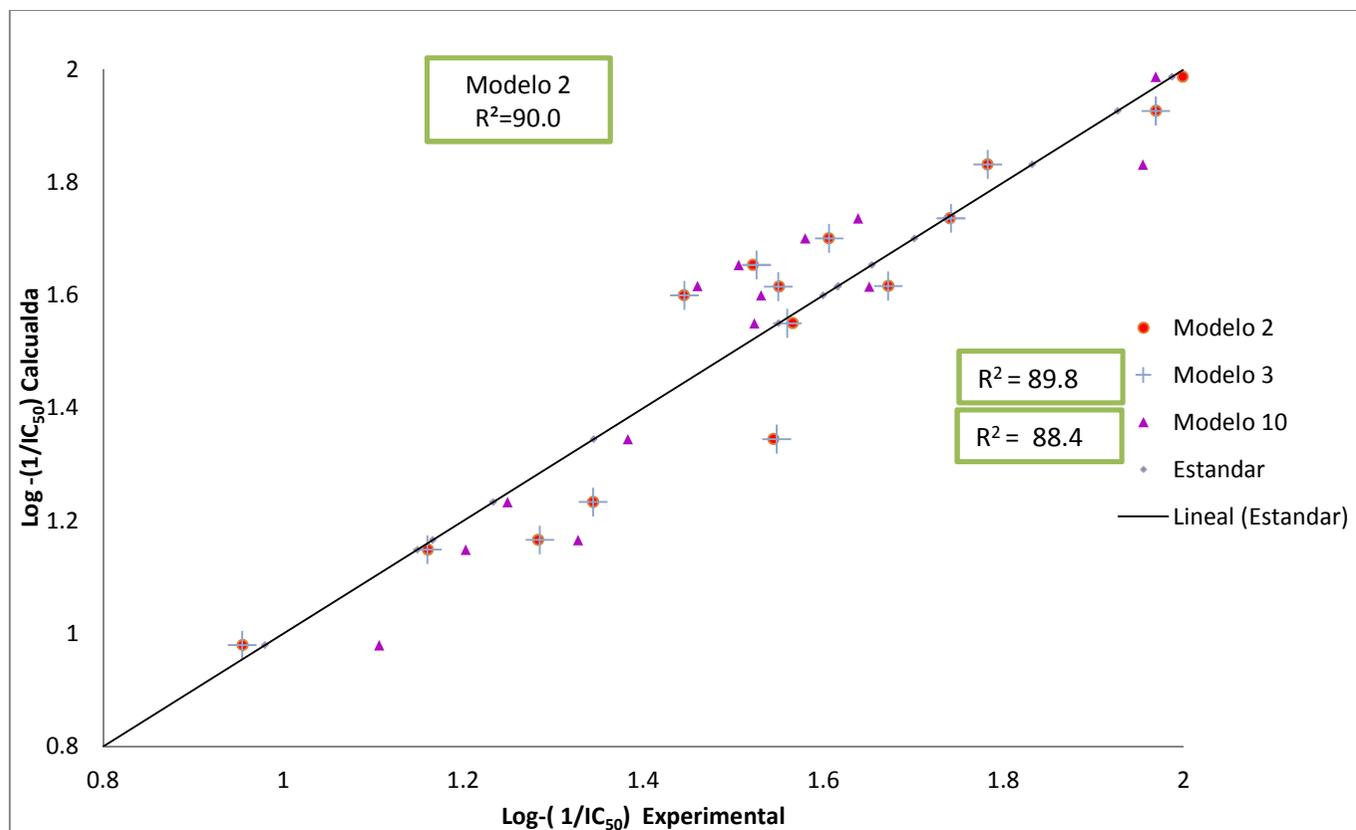
Como ya se mencionó los descriptores moleculares representan información química, resultado de lógica y procedimientos matemáticos, por lo tanto deben ser derivados de procedimientos bien establecidos que les confieran la posibilidad de calcular propiedades para cualquier conjunto de moléculas. Es evidente que un descriptor o un pequeño número de descriptores no pueden representar totalmente la complejidad molecular o el modelo de todas las respuestas físicoquímicas y las interacciones biológicas, por lo que podemos decir que se trata de modelos aproximados usados por su funcionalidad, dependiendo de las condiciones que se necesiten para algún compuesto seleccionado, en la Tabla 17

se presenta de manera general un resumen de cada uno de los Modelos (2, 3 y 10) analizados con sus respectivos descriptores, se infiere que la capacidad predictiva permite tamizar aun más el procedimiento, obteniendo dos mejores modelos. Los modelos 2 y 3 pueden ser confiables en predecir actividad de compuestos desconocidos y mantienen un claro parentesco en sus (descriptores SIC4 y CIC4) que pertenecen a la Familia 5, aportando características respecto a la simetría, y el valor de carga relacionadas con las electronegatividades de Van der Waals. El Modelo 10 a diferencia, aporta características topológicas sobre la conectividad y adyacencia de los átomos dependiendo de los radios atómicos.

**Tabla 17.** Esquema general de los descriptores por modelo.

<b>MODELOS</b>			
	<b>2</b>	<b>3</b>	<b>10</b>
	<b>SIC4</b>	<b>CIC4</b>	<b>X0A</b>
<b>Características Específicas de los descriptores</b>	Estructural de simetría	Estructural de simetría	Topológico de conectividad
	<b>BELv3</b>	<b>BELv3</b>	<b>EEig11r</b>
	Valor de carga	Valor de carga	Topológico de adyacencia
	<b>PJI3</b>	<b>PJI3</b>	<b>RDF095p</b>
	Geométrico	Geométrico	Distribución radial

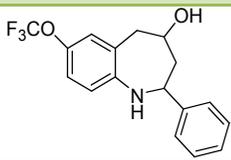
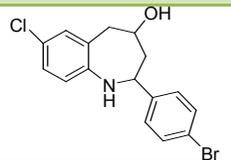
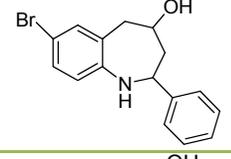
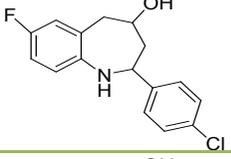
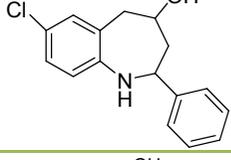
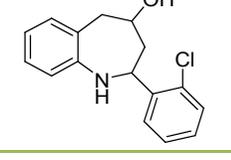
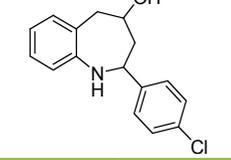
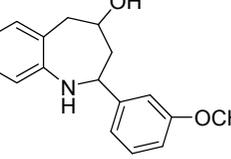
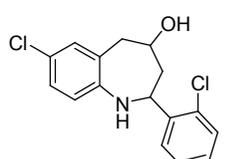
Con la información obtenida hasta ahora, incluyendo la estadística y las características de los descriptores de los Modelos matemáticamente buenos, en la Gráfica 4 se muestra el comportamiento de los Modelos seleccionados respecto a la funcionalidad por el QSAR, donde se observa que efectivamente la actividad calculada de cada uno de los compuestos está relacionada con su actividad experimental, observando que los valores respectivamente tienden a la recta; sin embargo, el Modelo 2 con una  $R^2$  de 90.0, es el Modelo que estadísticamente fue estimado como el “mejor”, el mismo que se usará para los siguientes análisis de esta investigación, sin embargo para darle mayor confiabilidad al trabajo se tomarán en cuenta los Modelos, 3 y 10 para actividad biológica.



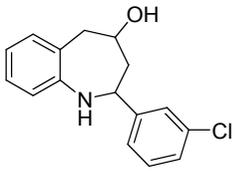
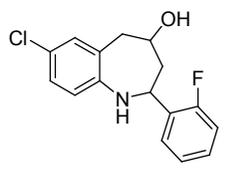
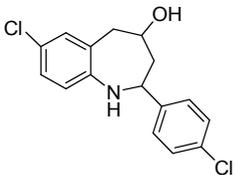
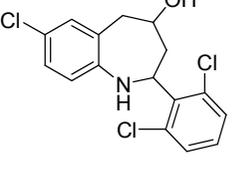
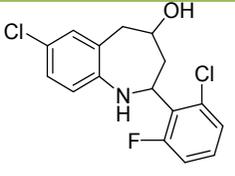
**Gráfica 4.** Dispersión general de  $\text{IC}_{50}$  experimental vs.  $\text{IC}_{50}$  calculada para promastigotes de *L. chagasi*.

De acuerdo con las características estructurales y fisicoquímicas de los descriptores teóricos analizados en los tres Modelos propuestos para predecir actividad biológica, el parecido de los Modelos 2 y 3 en su estadística y en su composición de variables independientes es similar, aunque en lo que refiere a su capacidad de predicción, de acuerdo a los resultados mostrados en la Tabla 18, el Modelos 2 es el de mejor capacidad predictiva y dista demasiado del Modelo 10, en la tabla también se muestra la serie de compuestos analizados en el estudio QSAR, con su actividad experimental y la actividad calculada por los tres Modelos, considerando esto los compuestos más activos seleccionados fueron “L” con actividad experimental de  $14.1 \mu$  y actividades calculadas de:  $14.47 \mu$  (Ec.1),  $14.45 \mu$  (Ec.2) y  $15.94 \mu$  (Ec.3), otro compuesto es el “M” con actividad experimental de  $9.54 \mu$  y de actividades calculadas:  $9.02 \mu$  (Ec.1),  $9.00 \mu$  (Ec.2) y  $12.78 \mu$  (Ec.3), aunque los Modelos son de carácter explicativo, la capacidad predictiva permite pronosticar que la actividad de este tipo de compuestos está relacionada con la Familia de los descriptores.

**Tabla 18.** Derivados benzoazepínicos para actividad biológica contra promastigotes de *L. chagasi*.

Compuestos Benzoazepínicos	Act. Exp. IC <sub>50</sub> (μ )	Actividad Calculada IC <sub>50</sub> ( μ )			Compuestos Benzoazepínicos	Act. Exp. IC <sub>50</sub> ( μ )	Actividad Calculada IC <sub>50</sub> ( μ )		
		Modelo 2	Modelo 3	Modelo 10			Modelo 2	Modelo 3	Modelo 10
	35.5 B	36.8	36.28	33.35		9.54 M	9.02	9.00	12.78
	45.04 C	33.26	33.54	32.06		17.11 N	22.10	22.08	17.74
	22.12 D	35.03	35.32	24.13		84.44 O	93.18	93.10	103.47
	67.89 F	60.59	60.59	90.03		41.34 Q	47.01	46.98	28.84
	97.09 H	99.81	100.08	93.12		50.22 R	40.37	40.41	38.01

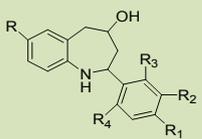
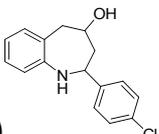
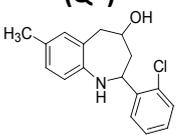
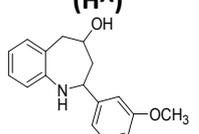
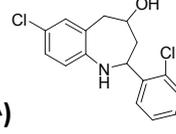
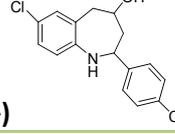
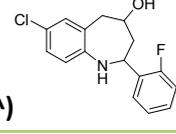
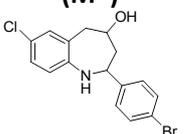
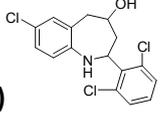
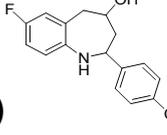
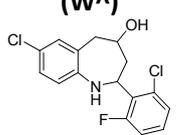
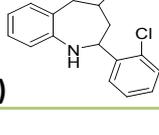
**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

	41.23 J	35.54	35.46	44.76		54.45 S	55.02	55.15	43.49
	14.1 L	14.47	14.45	15.94		14.16 V	19.20	19.28	21.25
						39.75 W	27.87	27.89	33.95

## 6.2 PREDICCIÓN DE LAS PROPUESTAS DE LOS DERIVADOS DE LA BENZOAZEPINA.

Como uno de los objetivos de este estudio es la propuesta de estructuras novedosas de tipo *cis*-2-aril-4-hidroxitetrahydro-1-benzazepinas, que de acuerdo a sus características estructurales y fisicoquímicas obtenidas a partir de descriptores teóricos, puedan presentar actividad, se llevó a cabo una propuesta inicial de 20 compuestos con variaciones en los sustituyentes en el grupo fenilo, conocidos como (R<sub>1</sub>, R<sub>2</sub> y R<sub>3</sub>) dentro del compuesto cabeza de serie. La predicción de la actividad se llevó a cabo con el Modelo 2 y los resultados se resumen en la Tabla 19. Los compuestos 14-R<sup>^</sup> sustituyente en (R<sub>1</sub>) con actividad calculada de 3.96 μ , el 15-S<sup>^</sup> sustituyente en (R<sub>3</sub>) con actividad calculada de 3.09 μ y el 18-V<sup>^</sup> con sustituyente en (R<sub>1</sub>) y actividad calculada de 4.11 μ , tienen actividades bajas respecto a los demás compuestos, presuntivamente podrían ser buenos compuestos, sin embargo se necesita una validación experimental de la actividad biológica para comprobarlo.

**Tabla 19.** Propuesta de análogos potencialmente activos contra *Leishmania chagasi*.

<b>Compuesto</b> <b>Cabeza de serie</b> 	<b>Modelo 2</b> <b>IC<sub>50</sub> μ</b>			<b>Compuesto</b>	<b>Modelo 2</b> <b>IC<sub>50</sub> μ</b>		
	<b>R<sub>3</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>1</sub></b>		<b>R<sub>3</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>1</sub></b>
<b>(F<sup>Λ</sup>)</b> 	--	---	14.55 1-(Cl)	<b>(Q<sup>Λ</sup>)</b> 	---	5.95 11-(Cl)	25.75 12-(Cl)
<b>(H<sup>Λ</sup>)</b> 	21.86 2-(OCH <sub>3</sub> )	----	155.10 3-(OCH <sub>3</sub> )	<b>(R<sup>Λ</sup>)</b> 	---	11.11 13-(Cl)	<b>3.97</b> 14-(Cl)
<b>(L<sup>Λ</sup>)</b> 	31.91 4-(Cl)	262.96 5-(Cl)	----	<b>(S<sup>Λ</sup>)</b> 	<b>3.09</b> 15-(F)	65.05 16-(F)	---
<b>(M<sup>Λ</sup>)</b> 	15.22 6-(Br)	4.64 7-(Cl)	---	<b>(V<sup>Λ</sup>)</b> 	---	65.05 17-(Cl)	<b>4.11</b> 18-(Cl)
<b>(N<sup>Λ</sup>)</b> 	14.44 8-(Cl)	5.64 9-(Cl)	---	<b>(W<sup>Λ</sup>)</b> 	---	6.5 19-(Cl)	39.57 20-(Cl)
<b>(O<sup>Λ</sup>)</b> 	9.13 10-(Cl)	---	---				

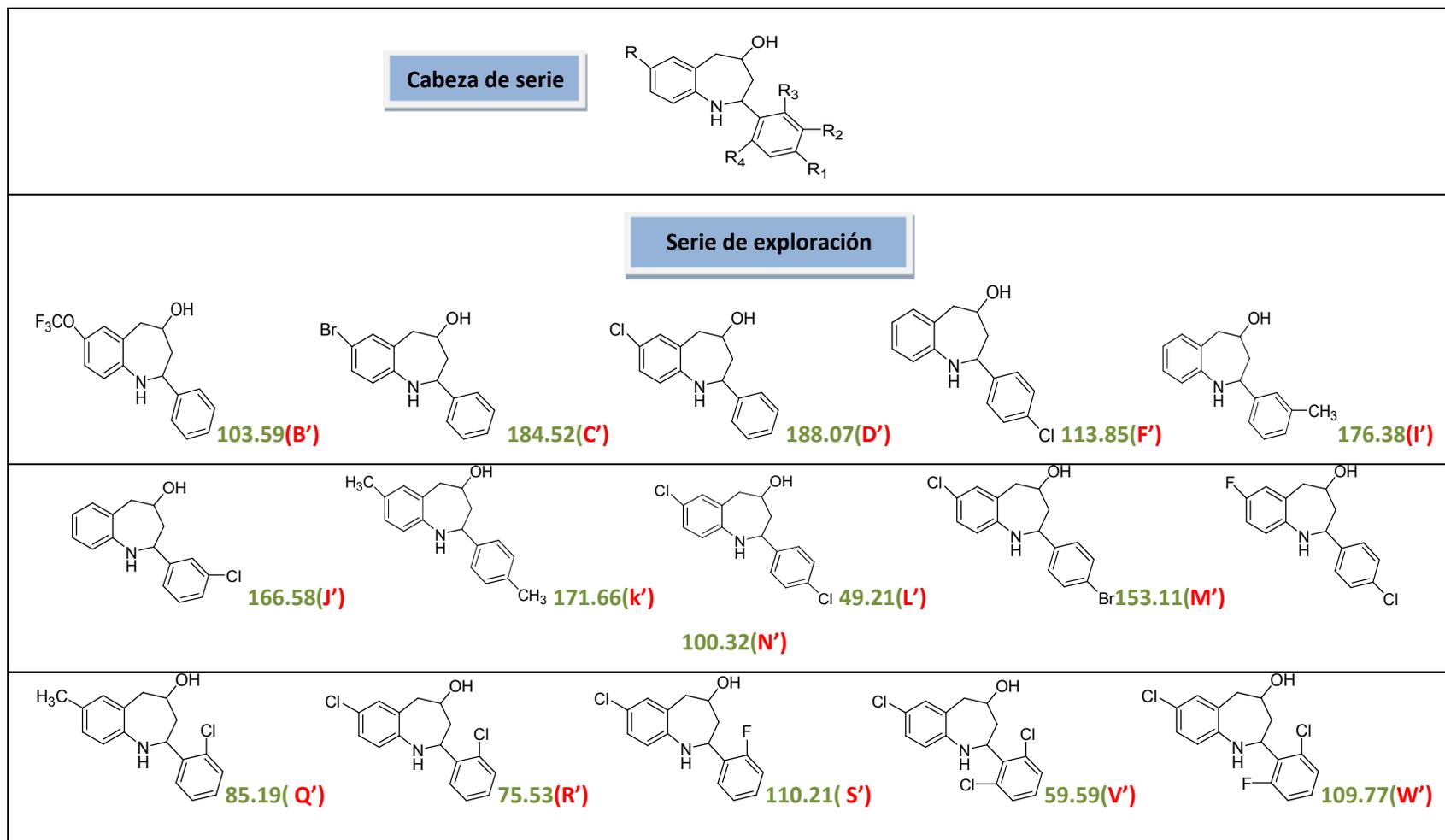
### 6.3 ESTUDIO QSAR DE LA CITOTOXICIDAD.

Como bien se sabe, el método de QSAR relaciona la estructura con la actividad de los compuestos químicos, por lo que además de elucidar su actividad biológica contra *Leishmania chagasi*, la investigación se abordó con los estudios de citotoxicidad de la misma serie de exploración, ya que al ser parásitos intracelulares obligados, semejan rutas metabólicas alterando el buen funcionamiento celular. Por esta razón, es posible, al menos en teoría, proponer estructuras que cumplan con los requisitos deseables de un fármaco, es decir, selectividad y lo menos posible de efectos secundarios indeseables.

Uno de los objetivos es obtener modelos QSAR derivados de la serie *cis*-2-aril-4-hidroxitetrahydro-1-benzoazepinas,<sup>7</sup> que permitan explicar la citotoxicidad de compuestos contra *Leishmania chagasi*, que sean selectivos y que lo menos tóxicos, esta es la razón de proponer nuevas estructuras benzoacepínicas con el afán de analizar las características fisicoquímicas de citotoxicidad, así como obtener una mejor optimización en los modelos matemáticos. Como metodología se siguieron los parámetros antes mencionados en el análisis de la actividad biológica con los software: Hyper Chem®, Dragon® y Moby Digs®, tomando como referencia la concentración citotóxica ( $CC_{50}$ ) en líneas celulares de mamíferos THP-1. El uso de los valores de citotoxicidad depende de si causan daño celular a bajas concentraciones, e indica, abaja dosis para presentar toxicidad y una alta concentración refleja daño celular, significa que para que cause toxicidad cierto compuesto tendrían que manejarse dosis altas.

En la Tabla 20, muestra la serie de exploración de los 15 derivados benzoacepínicos usados para el análisis de citotoxicidad, para confirmar dicho estudio se realizó una validación externa donde se seleccionaron los compuestos: K', L', M' y N', sin integrarlos en el análisis interno y posteriormente obtener sus valores citotóxicos.

**Tabla 20.** Serie de exploración de los derivados benzoazepínicos para citotoxicidad<sup>4</sup>.



**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

Los 50 modelos obtenidos en la Tabla 21, evaluados con los valores de citotoxicidad, son considerados como “mejores modelos” de acuerdo a los parámetros establecidos en el programa, la mayoría de ellos contienen valores estadísticamente aceptables, sin embargo se eligieron los Modelos 8 y 12 por que cumplen con los requerimientos estadísticos del Moby Digs, al tener una  $R^2 = 87.67$  y  $89.78$  respectivamente indicando que son buenos modelos, pues el valor promedio de la  $R^2$  tiende al 100, y en el caso de la Kx, para el Modelo 8 es de 46.82 y para el Modelo 10 una  $Kx = 41.49$ , en ambos casos menores de 50, indicando que entre sus descriptores no existe correlación y al no existir proporcionan mejor información sobre las características de los compuestos respecto a la citotoxicidad.

**Tabla 21.** Los 50 modelos con sus descriptores y principales parámetros estadísticos, para citotoxicidad.

No.	Tamaño	Modelos	R2	Q2	Q2boot	Kx	F	s
1	3	Mor24u Mor10p R5m+	84.82	76.16	62.33	20.5	20.49	0.081
2	3	Me Mor24e Mor10p	84.57	75.74	65.2	31.03	20.1	0.081
3	3	ZM1V RDF110m Mor10p	84.24	75.4	0	15.21	19.6	0.082
4	3	SP01 Mor10e G1m	84.64	75.2	42.87	28.24	20.2	0.081
5	3	G(O..F) RDF110m Mor10p	84.41	74.84	0	19.16	19.85	0.082
6	3	Mor10p HATS4u R4m+	83.29	73.97	62.29	41.27	18.28	0.085
7	3	Mor24u Mor10p HATS4u	84.52	73.52	65.93	28.68	20.02	0.081
8	3	Mor06m Mor24e Mor10p	87.67	73.11	70.03	46.82	26.07	0.073
9	3	DP03 Mor10p R4m+	82.51	72.89	55.42	40.48	17.3	0.087
10	3	Mor24e Mor10p R5m+	82.51	72.84	61.7	23.45	17.3	0.086
11	3	Mor24e Mor10p HATS4u	83.74	72.83	63.24	34.07	18.88	0.083
12	3	Mor24u Mor06m Mor10p	89.78	72.2	72.28	41.49	32.21	0.066
13	3	Mor24u Mor10p R4p	86.85	71.59	63.92	53.6	24.22	0.075
14	3	SEigp RDF110m Mor10p	85.3	71.22	0	22.47	21.28	0.079
15	3	Mor24u Mor10p Vu	84.14	71.11	62.72	19.9	19.46	0.082
16	3	EEig06r Mor24e Mor10p	80.1	70.69	0	31.45	14.76	0.092
17	3	Dz RDF110m Mor10p	84.57	70.38	0	8.76	20.1	0.081
18	3	Se RDF110m Mor10p	83	70.1	0	23.06	17.91	0.085
19	3	ESpm02d Mor24e Mor10p	82.9	69.95	55.67	34.25	17.78	0.086
20	3	Mor24e Mor10p R4p	86.02	69.78	60.48	59.06	22.57	0.077
21	3	ESpm04x Mor24e Mor10p	82.15	69.2	52.93	36.38	16.87	0.087

**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

22	3	Me Mor24u Mor10p	81.91	69.14	56.28	24.8	16.6	0.088
23	3	EEig10d Mor24e Mor10p	86.77	69.09	60.48	24.41	24.04	0.075
24	3	VEp1 Mor24e Mor10p	82.57	68.35	50.44	28.03	17.37	0.086
25	3	Mor24e Mor10p Vu	82.43	68.09	58.23	26.18	17.2	0.087
26	3	HNar Mor24e Mor10p	81.31	67.62	54.51	44.94	15.95	0.089
27	3	DP03 Mor24e Mor10p	80	67.01	42.21	37.96	14.67	0.093
28	3	ESpm04x RDF110m Mor10p	81.23	66.89	0	24.8	15.87	0.09
29	3	ESpm02d Mor24u Mor10p	81.38	66.56	51.2	28.78	16.02	0.089
30	3	SP01 Mor24e Mor10p	81.15	66.15	29.29	41.68	15.78	0.09
31	3	EEig10d Mor10p R4m+	77.76	65.66	39.95	44.14	12.82	0.098
32	3	VEp1 Mor24u Mor10p	81.6	64.91	48.98	23.15	16.26	0.089
33	3	ESpm04x Mor24u Mor10p	80.87	64.85	48.25	31.14	15.5	0.09
34	3	EEig06r Mor24u Mor10p	78.76	64.68	42.77	26.26	13.6	0.095
35	3	DP03 Mor24u Mor10p	79.88	64.68	47.68	34.29	14.56	0.093
36	3	Mor24e Mor10p G1m	77.12	64.49	48.66	28.9	12.36	0.099
37	3	Mor11v Mor24e Mor10p	76.78	64.33	52.9	40.13	12.13	0.1
38	3	ESpm02d RDF110m Mor10p	81.56	64.27	0	23.46	16.21	0.089
39	3	Mor24e Mor10p C-024	78	63.93	53.12	61.68	13	0.097
40	3	Mor24e Mor10p nCbH	78	63.93	53.58	61.68	13	0.097
41	3	Mor24e Mor10p nCb-	78	63.93	54.11	61.68	13	0.097
42	3	X0 Mor24u Mor10p	82.28	63.9	45.53	21.56	17.03	0.087
43	3	Me RDF110m Mor10p	80.38	63.84	0	15.85	15.03	0.092
44	3	ESpm02r RDF110m Mor10p	83.24	63.73	0	16.28	18.21	0.085
45	3	ZM1 Mor24e Mor10p	83.02	63.7	46.31	27.45	17.93	0.085
46	3	ESpm02r Mor24e Mor10p	83.35	63.67	50.46	26.97	18.36	0.084
47	2	Mor24e Mor10p	73.01	61.07	54.51	43.99	16.23	0.103
48	2	DP03 Mor10p	73.22	59.74	47.03	20.08	16.4	0.102
49	1	Mor10p	55.16	41.45	40.09	0	15.99	0.127
50	1	HNar	43.26	26.43	31.67	0	9.91	0.143

Dado que los modelos obtenidos son semejantes, en el sentido estadístico, se decidió llevar a cabo la validación externa. Como bien se sabe un modelo QSAR es un modelo matemático sustentado estadísticamente y que su objetivo final es que sea funcional, para lo cual quien tiene la última palabra es la validación externa.

Esto es que si el modelo pretende ser predictivo debe entonces tener la capacidad de, predecir actividad de compuestos no incluidos en el sistema.

Idealmente, no se debe conocer la actividad del o los compuestos externos hasta su comparación una vez obtenida su actividad teórica de acuerdo al modelo elegido. La actividad teórica debe aproximarse a la experimental aunque no necesariamente ser la misma. A continuación se corrobora tal validación en los Modelos 8 y 12.

**IV. Modelo 8:**

**Ec.4**

$\text{Log } -(1/CC_{50}) = 0.122 \text{ Mor06m } (\pm 0.033) - 0.825 \text{ Mor24e } (\pm 0.174) + 1.4377 \text{ Mor10p } (\pm 0.177) + 2.043$
$n = 15; \quad R^2 = 87.66; \quad F = 26.065; \quad s = 0.07; \quad Q^2 = 73.11; \quad Kx = 46.82$

Con base a lo anterior se procedió a la comparación de las  $CC_{50}$  experimental y calculada, donde en la Tabla 22, marcado con azul el compuesto (F'), tiene una  $CC_{50}$  calculada  $113.9 \mu$  muy parecida a la experimental  $113.5 \mu$ , y con respecto a los compuestos seleccionados para la validación externa, los valores comparativos de citotoxicidad de los compuestos K' y M', tienen poca discrepancia, sin embargo para decir que es un buen modelo predictivo se opta por hacer el estudio estadístico completo.

**Tabla 22.** Comparación entre la  $CC_{50}$  experimental y calculada para el Modelo 8.

Mor06m	Mor24e	Mor10p	Compuesto	$CC_{50}$ ( ) Exp.	$CC_{50}$ ( ) Calc.
1.237	0.437	0.167	B'	103.59	118.623
3.282	0.311	-0.015	C'	184.52	146.697
2.452	0.202	0.018	D'	188.07	159.400
2.36	0.245	-0.051	F'	113.85	113.902

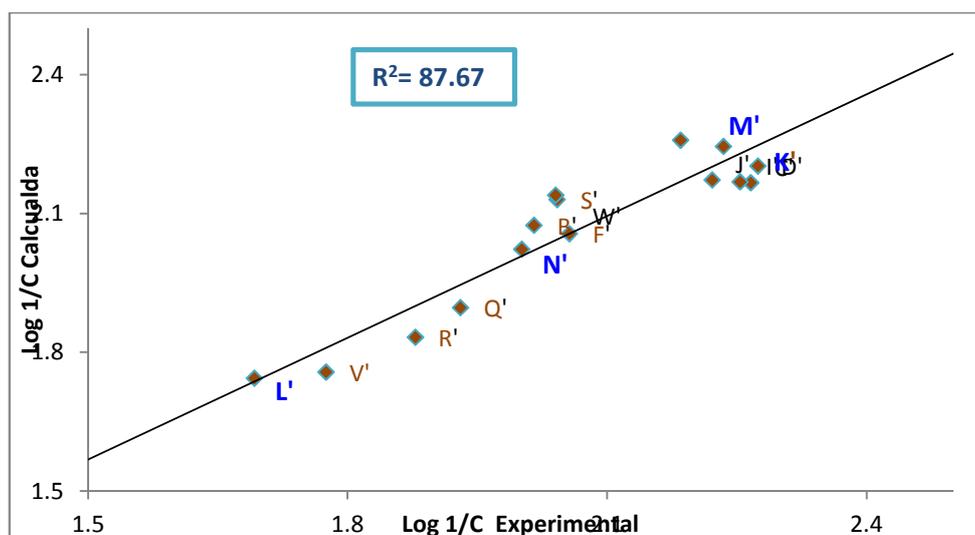
**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

2.36	0.301	0.059	I'	179.38	147.376
2.282	0.295	0.065	J'	166.58	148.758
2.67	0.269	-0.175	Q'	85.19	78.760
3.364	0.334	-0.241	R'	75.53	68.001
2.323	0.26	0.012	S'	110.21	134.958
3.849	0.077	-0.482	V'	59.59	57.220
3.093	0.063	-0.16	W'	109.77	137.942
1.319	0.411	0.264	K'	171.66	175.841
3.62	0.535	-0.209	I'	49.21	55.435
4.129	0.295	-0.032	M'	153.11	181.416
1.849	0.134	-0.095	M'	100.32	105.322

En la Tabla 23 se puede apreciar que los descriptores no están correlacionados, apuntando ser buen Modelo predictivo y confirmando con la Gráfica 5, los puntos de ambas actividades tienden a la recta, lo que quiere decir que éstas tienen relación puesto que su  $R^2$  es de = 87.67. Se procedió a la fragmentación del modelo (Tabla 24) donde se puede apreciar de manera significativa que el Modelo (8.2), que contiene los descriptores Mor06m, Mor10p, tiene una  $R^2$  de 84.3, y que los Modelos, (8.3) y (8.6) tienen  $R^2$  altas con respecto a los demás modelos de la fragmentación, indicando estadísticamente que las características físicoquímicas de los compuestos que dependen del descriptor Mor10p, ya que integra en los tres modelos.

**Tabla 23.** Coeficiente de correlación Modelo 8.

Descriptores	Mor06m	Mor24e	Mor10p
<b>Mor06m</b>	1		
<b>Mor24e</b>	0.3913	1	
<b>Mor10p</b>	0.7347	0.4298	1



Gráfica 5. Relación de Log  $-(1/CC_{50})$  Exp vs. Log  $-(1/CC_{50})$  Calc. Modelo 8.

Tabla 24. Fragmentación del Modelo 8 para citotoxicidad.

Modelo	Número de descriptores	Descriptores	R <sup>2</sup>	F	S
8	3	Mor06m, Mor24e, Mor10p	87.6	26.065	0.072
8.1	2	Mor06m, Mor24e	13.7	0.638	0.175
8.2	2	Mor06m, Mor10p	84.3	21.558	0.074
8.3	2	Mor24e, Mor10p	66.5	7.973	0.109
8.4	1	Mor06m	13.7	1.430	0.165
8.5	1	Mor24e	06.2	0.597	0.172
8.6	1	Mor10p	56.2	11.589	0.117

#### IV.I Descriptores correspondientes al modelo 8.

Los descriptores clase **MorSE**, (Representación molecular 3D de las estructuras basadas en la difracción de electrones), son índices moleculares 3D que representan diferentes fuentes de información química en términos de tamaño, forma, simetría y distribución atómica y son obtenidos a partir de sus coordenadas de la estructura tridimensional con diferentes esquemas de

numeración, en las electronegatividades de Sanderson (que se fundamenta en el recíproco del volumen atómico).<sup>55, 56, 57</sup>

Para el cálculo de la representación 3D-MorSE de los compuestos de estructuras basadas en electrones en cuanto al peso atómico de propiedades físico-químicas tales como la masa atómica, cargas atómicas parciales, polarizabilidad atómica fueron considerados, con el fin de obtener de manera uniforme la longitud de los descriptores. Estos descriptores, se derivan de la simulación de espectros de infrarrojo usando una función de dispersión generalizada.

Los descriptores 3D-MorSE se calculan para 5 diferentes propiedades atómicas: caso no ponderado (u), masa atómica (m), volumen de Van der Waals (v), la electronegatividad atómica de Sanderson (e) y la polarizabilidad atómica (p).

Un típico descriptor MorSE se denota por:  $1 \leq s \leq 32$  y  $w$  (u, m, v, e y p)

Morse (x) donde:

s y w, toman los valores  $1 \leq s \leq 32$

u, es no ponderado

m, es ponderada por la masa

v, se pondera por van der Waals volumen

E, se pondera por electronegatividad

p, es ponderada por polarizabilidad

Los descriptores MoRSE están definidos por lo siguiente:

$$(i, j) = (i, j) = \frac{\sin(\theta_{ij})}{(r_{ij})^2}$$

Donde:

( $r_{ij}$ ): es la distancia euclídea entre los átomos

(i y j) y ( $w_i$  y  $w_j$ ): son los pesos de los átomos i y j, respectivamente.

Estos descriptores pertenecen a la Familia 14, integrada por 160 descriptores, convierten la estructura tridimensional a un número de variables fijas, el método usa una versión modificada de la transformación molecular usando estudios en difracción de electrón. Varias propiedades atómicas tales como masa atómica, cargas atómicas parciales, residual de electronegatividades atómicas y polarizabilidades atómicas que pueden ser codificado a un set de valores 3D-MorSE que son independientes.

La siguiente expresión se usa para calcular los descriptores MorSE.

$$= \frac{\sum_{i,j} \frac{w_i w_j}{r_{ij}^S}}{\sum_{i,j} \frac{w_i w_j}{r_{ij}^S}}$$

Donde:

Morsw: es la intensidad del electrón dispersado

(w): propiedad atómica

( $r_{ij}$ ): son las distancias interatómicas entre los átomos i,j

(NAT): es el número de átomos

(S): es una medida del ángulo de dispersión dado por  $s = 4\pi \sin(\theta/2)/\lambda$ ,  $\theta$  es el ángulo de dispersión y  $\lambda$  es la longitud de onda.

**Mor06m:** Señal 06 de 3D-MorSE, ponderado por las masas atómicas.

**Mor24e:** Señal 24 de 3D-MorSE, ponderados por electronegatividades atómicas Sanderson. Estas escalas se basan en las energías de ionización y los radios covalentes, sabiendo que la electronegatividad es una característica del estado del sistema; los electrones tienden a fluir desde una región de electronegatividad baja a una región de alta electronegatividad, con la formación de una molécula, las electronegatividades atómicas pasan de la electronegatividad del grupo orbital y funcional a la electronegatividad molecular, además también puede estimarse mediante un enfoque topológico.<sup>55</sup>

La electronegatividad es una propiedad donde los electrones tienden al flujo de una región de baja electronegatividad a una región de alta electronegatividad. Para el caso de los descriptores MorSE la electronegatividad atómica de Sanderson: propone que la electronegatividad de un átomo es una medida de la “densidad” de su nube electrónica, comparada con la de un átomo inerte hipotético con el mismo número de electrones, ya que si un átomo es muy electronegativo, es decir, atrae fuertemente a otros electrones, los propios electrones los mantendrá muy próximos entre sí. Sanderson define la densidad electrónica media (DE) mediante la relación:

$$= \frac{Z}{r^3}$$

Donde (*r*) es el radio covalente en Angstroms.

Mor24e, (3D-MorSE-s24) ponderado por electronegatividad atómica de Sanderson, se refiere al tamaño y carga nuclear, indicando que si un átomo atrae intensamente su propia atmósfera electrónica de la misma forma atraerá un electrón de otro átomo por la cercanía de sus orbitales. Si por otra parte las repulsiones interelectrónicas son suficientes como para que disminuya la atracción de un átomo sobre sus propios electrones, tal átomo no atraerá fuertemente a un electrón externo aunque disponga de orbitales vacíos. En cuanto a la polarizabilidad, nos hablara de la respuesta de la molécula a un campo externo.<sup>56,</sup>

<sup>57</sup>

**Mor10p:** Señal 10 de 3D-MorSE, ponderado por la polarizabilidad atómica. Los descriptores MorSE usan las coordenadas espaciales de la estructura en 3D, estas coordenadas son transformadas en un curva de difracción electrónica y esta curva se basada en propiedades moleculares y en este caso el descriptor Mor10p, mide la polarizabilidad del átomo.<sup>56,57</sup>

## V. Modelo 12:

### Ec. 5

$\text{Log} - (1/\text{CC}_{50}) = -0.936 \text{ Mor24u} (\pm 0.172) + 0.134 \text{ Mor06m} (\pm 0.031) + 1.428 \text{ Mor10p} (\pm 0.159) + 2.049$
$n = 15; R^2 = 89.77; F = 32.209; Q^2 = 72.2; K_x = 41.49; s = 0.06$

Se siguió el mismo procedimiento para el Modelo 12, en la Tabla 25 muestra la comparación entre la concentración citotóxica ( $\text{CC}_{50}$ ,  $\mu$ ) experimental y la calculada, donde, se obtuvo que el compuesto (V') su  $\text{CC}_{50}$  experimental fue de  $59.59 \mu$  y la calculada de  $58.07 \mu$ ; como se observa existe una diferencia poco significativa, para confirmar el análisis del Modelo 12, se procedió al cálculo del coeficiente de correlación mostrado en la Tabla 26, se observa, la baja correlación entre los descriptores y para comparar los valores significativos del compuesto (V') en la Gráfica 6, los valores de todos los compuestos tienden a la recta es decir las  $\text{CC}_{50}$  experimental y calculada tienen cierta relación, sin embargo los modelos K', L', M', N' que son los compuestos usados para la validación externa tienen diferencia significativa respecto a la (Ec. 4) del Modelo 8 por lo que se infiere que la (Ec. 5) del Modelo 12, aunque estadísticamente sea bueno, no lo es para predecir este tipo de compuestos.

Es importante partir de la molécula de la cual se tiene una referencia, es decir la molécula V'; donde se nota que la actividad dada es cuando los descriptores toman los valores de;  $\text{Mor24u} = 0.124$   $\text{Mor06m} = 3.849$  y  $\text{Mor10p} = -0.482$ , observando una cifra negativa para el caso de  $\text{Mor10p}$  (baja), en cuanto a  $\text{Mor24u}$  y  $\text{Mor06m}$  cifras positivas (altas), haciendo una comparación con los

**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

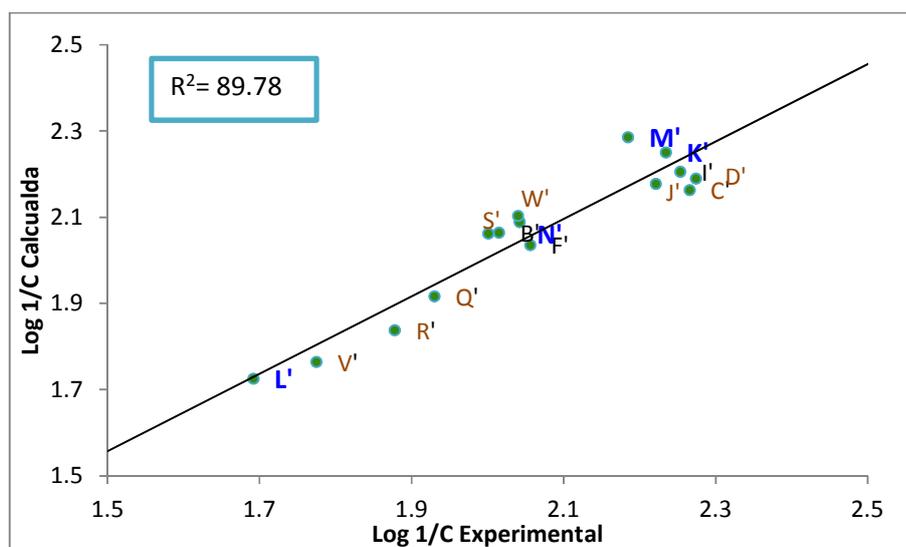
descriptores Mor10p se refieren a la electronegatividad de Sanderson, el valor de  $\sigma$ , es representado por 10 e indica la dispersión.

**Tabla 25.** Comparación entre la  $CC_{50}$  experimental y calculada para el Modelo 12.

Descriptores			Compuesto	$CC_{50}$ ( )	$CC_{50}$ ( )
Mor24u	Mor06m	Mor10p		Exp.	Calc.
0.417	1.237	0.167	B'	103.59	115.919
0.329	3.282	-0.015	C'	184.52	145.413
0.231	2.452	0.018	D'	188.07	154.695
0.277	2.36	-0.051	F'	113.85	108.495
0.263	2.36	0.059	I'	179.38	160.570
0.291	2.282	0.065	J'	166.58	150.483
0.26	2.67	-0.175	Q'	85.19	82.419
0.343	3.364	-0.241	R'	75.53	68.817
0.311	2.323	0.012	S'	110.21	122.625
0.124	3.849	-0.482	V'	59.59	58.070
0.144	3.093	-0.16	W'	109.77	126.819
0.378	1.319	0.264	K'	171.66	177.958
0.549	3.62	-0.209	I'	49.21	53.097
0.294	4.129	-0.032	M'	153.11	192.941
0.108	1.849	-0.095	M'	100.32	115.300

**Tabla 26.** Coeficiente de correlación Modelo 12.

Descriptores	Mor24u	Mor06m	Mor10p
Mor24u	1		
Mor06m	0.3698	1	
Mor10p	0.4300	0.7347	1



**Gráfica 6.** Relación de Log - (1/CC<sub>50</sub>) Exp vs. Log - (1/CC<sub>50</sub>) Calc. Modelo 12.

Se procedió a la fragmentación del Modelo 12, y dada la similitud entre los dos Modelos (8 y 12) seleccionados para citotoxicidad, se tiene como resultado que al igual que el Modelo 8 el descriptor que despunta es el Mor10p, esto se comprueba en el Modelo (12.2) con  $R^2 = 84.3$  y el Modelo (12.3) con  $R^2 = 66.9$ , ambos tienen la combinación de descriptores donde el descriptor que se repite es el Mor10p.

**Tabla 27.** Fragmentación del Modelo 12 para citotoxicidad.

Modelo	Número de descriptores	Descriptores	R <sup>2</sup>	F	s
12	3	Mor24u, Mor06m, Mor10p	89.7	32.209	0.066
12.1	2	Mor24u, Mor06m	13.7	0.636	0.175
12.2	2	Mor06m, Mor10p	84.3	21.558	0.074
12.3	2	Mor24u, Mor10p	69.9	9.303	0.103
12.4	1	Mor24u	13.7	1.430	0.165
12.5	1	Mor06m	04.5	0.428	0.173
12.6	1	Mor10p	56.28	11.589	0.117

## V.I Descriptores correspondientes al Modelo 12.

Como se analizó en el apartado anterior, los Modelos 8 y 12 comparten los descriptores de la Familia 14 “3D-MorSE”, Mor06m, Mor10p, pero la característica del Modelo 12 es que tiene el descriptor **Mor24u**, que corresponde a una señal 24, no cuantificada o ponderada.<sup>55</sup>

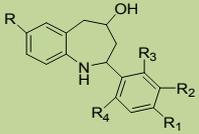
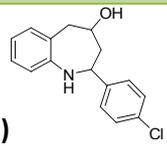
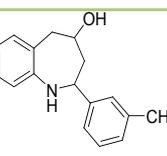
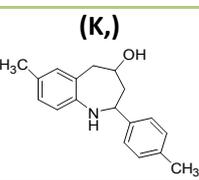
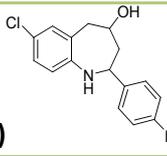
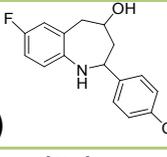
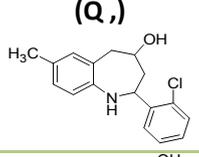
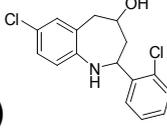
En este caso, las características fisicoquímicas de los compuestos propuestos que explican la citotoxicidad, están dadas por los descriptores de la Familia 3D-MorSE, en ambos Modelos (8 y 12), en general las características de este descriptor, están basadas en los índices moleculares 3D que se representan en términos de tamaño, forma, simetría, distribución atómica y son obtenidos a partir de sus coordenadas en la estructura tridimensional con diferentes esquemas de numeración, en las electronegatividades, por mencionar algunas, en la descripción del Modelo 8, se dió una amplia explicación al respecto.

En la Tabla 28 se muestran las actividades de los compuestos novedosos para el análisis de citotoxicidad en líneas celulares THP-1, con los sustituyentes en *Orto*, *Meta* y *Para* ( $R_3$ ,  $R_2$  y  $R_1$ ), donde se observan los valores citotóxicos de los mismos a partir de las ecuaciones (Ec. 4 y Ec. 5) de los Modelos 8 y 12 respectivamente. Se llevó a cabo la comparación con sus respectivas ecuaciones, donde se puede apreciar que la correlación citotóxica en el Modelo 8 con el compuesto (K,- $R_2$ ) tuvo una  $CC_{50}$  calculada de  $664.57 \mu$  y los compuestos del Modelo 12: (K,- $R_2$ )  $CC_{50}$  calculada de  $729.96 \mu$  y (W,-  $R_2$ ) calculada de  $633.90 \mu$  .

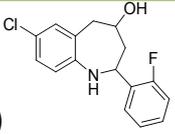
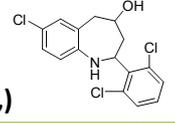
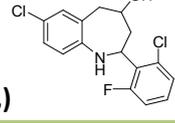
Lo anterior se resume en lo siguiente, los valores de citotoxicidad se formulan mediante el daño celular demostrando que, a bajas concentraciones la baja dosis presenta toxicidad, en cambio una alta concentración que refleje daño celular significará que, para que cause toxicidad tendrían que manejarse dosis

altas, por lo tanto si se hace una comparación con el valor de referencia de la anfotericina B de 8.79  $\mu$  , los compuestos diseñados en esta investigación presentan concentraciones altas, por tanto, se necesitarían dosis altas para ser compuestos citotóxicos.

**Tabla 28.** Propuesta de análogos benzoazepínicos para valorar citotoxicidad.

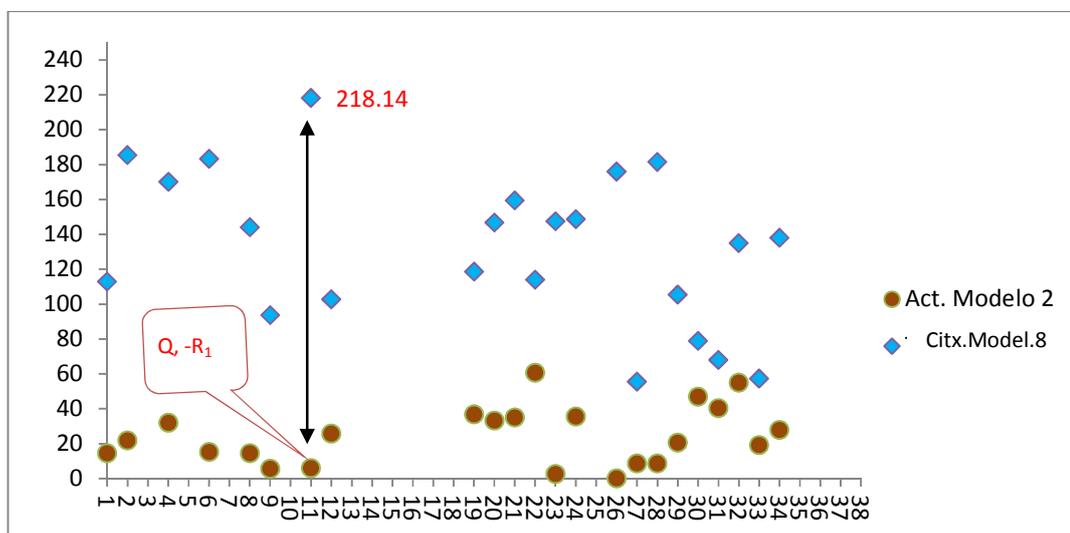
<b>Compuesto</b> <b>Cabeza de serie</b> 	<b>Modelo 8</b> <b>CC<sub>50</sub> <math>\mu</math></b>			<b>Modelo 12</b> <b>CC<sub>50</sub> <math>\mu</math></b>		
	<b>R<sub>3</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>1</sub></b>	<b>R<sub>3</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>1</sub></b>
<b>(F,)</b> 	112.79 (Cl)	---	---	114.64 (Cl)	---	---
<b>(I,)</b> 	185.39 (CH <sub>3</sub> )	----	271.90 (CH <sub>3</sub> )	170.81 (CH <sub>3</sub> )	---	270.73 (Cl)
<b>(K,)</b> 	170.13 (CH <sub>3</sub> )	<b>644.57</b> (CH <sub>3</sub> )	----	178.47 (CH <sub>3</sub> )	<b>729.96</b> (CH <sub>3</sub> )	---
<b>(M,)</b> 	183.21 (Br)	338.64 (Br)	---	191.13 (Br)	359.46 (Br)	<b>192.94</b> (Br)
<b>(N,)</b> 	144.01 (Cl)	93.57 (Cl)	---	128.63 (Cl)	77.70 (Cl)	---
<b>(Q,)</b> 	---	410.71 (Cl)	<b>218.14</b> (Cl)	---	449.52 (Cl)	225.33 (Cl)
<b>(R,)</b> 	----	102.65 (Cl)	---	---	98.11 (Cl)	---

**“Diseño por QSAR de compuestos potencialmente activos *Leishmania chagasi*”**

 <b>(S,)</b>	---	426.65 (F)	564.27 (F)	---	450.01 (F)	660.39 (F)
 <b>(V,)</b>	---	442.01 (Cl)	313.20 (Cl)	---	550.62 (Cl)	304.37 (Cl)
 <b>(W,)</b>	---	597.16 (Cl)	330.45 (Cl)	---	<b>633.90</b> (Cl)	355.95 (Cl)

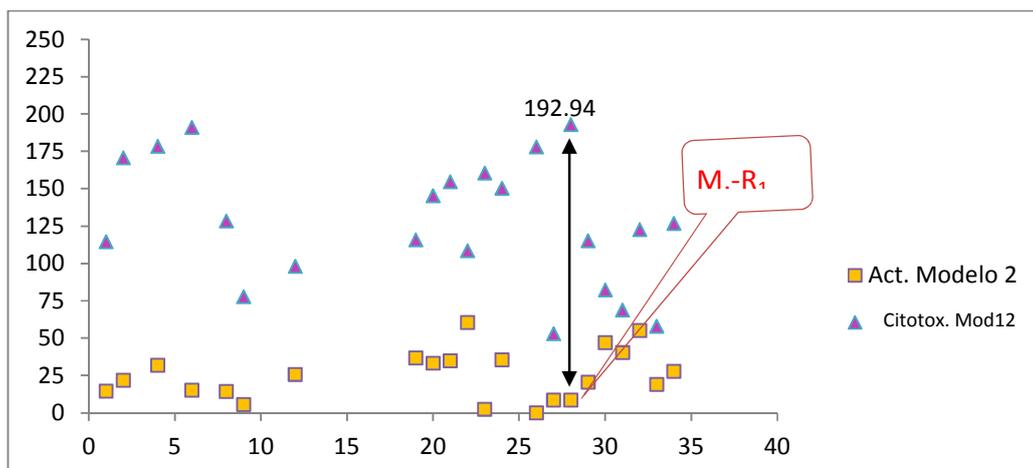
Lo anterior se corrobora al hacer una comparación entre la actividad biológica calculada del Modelo 2 y la citotoxicidad de los Modelos 8 y 12 de los compuestos benzodiazepínicos potencialmente activos, donde se despreciaron los valores de  $IC_{50}$  mayores de 50 y valores de  $CC_{50}$  mayores de 300 .

En la Gráfica 7, se compara la actividad calculada de los compuestos propuestos usando el Modelo 2 y el Modelo 8 de citotoxicidad, se analizó el compuesto Q,-R<sub>1</sub>, marcado con rojo en la Tabla 28, donde el punto más alto de citotoxicidad tiene un valor de 218.14 y su valor en actividad de 5.92, se considera que al tener estos valores un compuesto es potencialmente activo, puesto que se necesitaría gran cantidad del mismo para ser citotóxico estimando buena actividad contra el parásito.



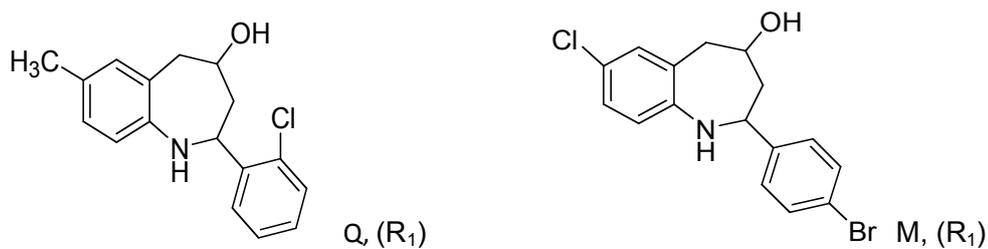
**Gráfica 7.** Comparación de la IC<sub>50</sub> Modelo 2 vs. CC<sub>50</sub> Modelo 8, de los derivados benzoazepínicos.

Finalmente en la Gráfica 8, se muestra la correlación entre los valores de la actividad del Modelo 2 contra los de citotoxicidad del Modelo 12, donde se aprecia que el compuesto, M,-R<sub>1</sub> con una IC<sub>50</sub> 8.85 baja y su CC<sub>50</sub> 192.94 alta (Tabla 20), es buen candidato para ser potencialmente activo contra *Leishmania chagasi*, al presentar estos valores se considera que tiene buena actividad biológica y baja citotoxicidad.



**Gráfica 8.** Comparación de la  $IC_{50}$  Modelo 2 vs.  $CC_{50}$  Modelo 12, de los derivados benzoazepínicos.

El diseño molecular por QSAR de nuevos compuestos potencialmente activos contra *Leishmania chagasi*, ha sido un trabajo laborioso donde finalmente se proponen los siguientes compuestos Q, $(R_1)$  con el Modelo 8 de citotoxicidad y M, $(R_1)$ , con el Modelo 12 de la misma serie y mostrados en la Figura 12. Dado lo anterior es posible proponer los siguientes compuestos activos contra *Leishmania chagasi*.



**Figura 12.** Propuesta de compuestos potencialmente activos contra *Leishmania chagasi*.

## 7. CONCLUSIONES.

La metodología QSAR, predice y explica la actividad biológica y la citotoxicidad de compuestos del tipo benzoazepínicos potencialmente activos contra *Leishmania chagasi*, de los cuales se desconoce su mecanismo de acción.

La ventaja de que el método QSAR proporcione modelos matemáticos entre un conjunto de datos, es que se puede utilizar para diseñar nuevos compuestos en función de los descriptores eficaces como: Indicadores de conectividad, Contenido de información estructural, Descriptores del valor de carga y los 3D-MorSe, características físicoquímicas como: los datos geométricos, las propiedades electrónicas, las interacciones moleculares, entre otras, que afectan la actividad parasitaria, por lo que se pueden utilizar para proponer compuestos novedosos del tipo benzoazepínicos con un óptimo de actividad biológica y un mínimo de citotoxicidad.

## 8. PERSPECTIVAS.

Se recomienda la síntesis de los derivados benzoazepínicos, Q, (R<sub>1</sub>) y M, (R<sub>1</sub>), que presentaron buena actividad biológica contra el *Leishmania chagasi* y baja citotoxicidad en líneas celulares THP-1.

Hacer un estudio QSAR donde se compare la actividad biológica y citotoxicidad experimentales de estos compuestos propuestos, con sus actividades y valores de citotoxicidad calculados en este proyecto de investigación.

## 9. REFERENCIAS BIBLIOGRÁFICAS.

- <sup>1</sup> Al-Jawbreh A., et al., (2008). Identification of geographically distributed subpopulations of *Leishmania major* by microsatellite analysis. 8: 183
- <sup>2</sup> Gamboa A. R. (2009). *Leishmaniasis cutánea*. Revista médica de Costa Rica y Centroamérica. 588: 169-172.
- <sup>3</sup> Castaño Gerra E. (2004). *Leishmaniasis*. Revista del Hospital del niño, Panamá, 20:152-159.
- <sup>4</sup> Arias J., (1996). Epidemiología y control de la *Leishmaniasis* en las Américas por país o territorio. Division de Control y Prevention de Enfermedades, Organization Panamericana de la Salud Brasilia, D.F., Brasil Cuademo Tecnico No. 44. Organización Panamericana de la Salud.
- <sup>5</sup> Puzyn T., Mark T. D. G., Challenges and advances in computational chemistry and physics, (8), Recent Advances in QSAR Studies Methods and Applications. Ed. Springer.
- <sup>6</sup> Lee, K. W.; Kwon, S. Y.; Hwang, S.; Lee, J.; and Kim, H. (1996). Quantitative Structure-Activity Relationships (QSAR) Study on C-7 Substituted; Bull. Korean Chem. Soc., Vol. 17, pp.147-152.
- <sup>7</sup> Cuevas G., Cortés F. (2003). Introducción a la química computacional. 1a. ed. Fondo de Cultura Económica. México, D.F. pp. 23-24.
- <sup>8</sup> López J., Olmo E. y San Feliciano A. F.A. (2005). Interacción Fármaco-Diana, bases para el diseño de nuevos fármacos. Departamento de Química, universidad de Salamanca, Esp., Enfermedades emergentes. 7(1): 52-56.

<sup>9</sup> Selassie, C. D.; (2003) History of Quantitative Structure-Activity Relationships; Burger's Medicinal Chemistry and Drug Discovery: 15.

<sup>10</sup> Gómez-Ayala S., Castrillón J. A., Palma A., et al., (2010). Synthesis, structural elucidation and in vitro antiparasitic activity against *Trypanosoma cruzi* and *Leishmania chagasi* parasites of novel tetrahydro-1-benzazepine derivatives. Bioorganic and Medicinal Chemistry. 18: 4721-4739.

<sup>11</sup> Vargas M.F., Torres Gro. E. (2011), *Leishmaniasis* en México, Departamento de Dermatología, Mex., DF. Educación Médica Continua, Iberoamericana Latin Am., (39) 4: 163-183.

<sup>12</sup> Jheman Zetina J. (2008). *Leishmaniasis cutánea* en el estado de Quintana Roo, México. Revista Mexicana. 52(1): 3-9.

<sup>13</sup> Treviño-Garza N., (2011). Prevención del *Leishmaniasis* en México. Secretaría de Salud, Subsecretaría de prevención y promoción de la salud. San Chistobal de las casas, Chiapas, Mex.

<sup>14</sup> Instituto Nacional de Salud., (2011). Boletín Médico del Hospital Infantil de México, Federico Gómez. Publicación bimestral, México. 68: 2.

<sup>15</sup> Santos-Preciado y Franco-Paredes., (2011). Neglected tropical diseases: 72 years after the establishment of the Instituto de Salubridad y Enfermedades Tropicales (Institute of Health and Tropical Diseases) in Mexico. Bol Med Hosp Infant Mex; 8(2):83-85.

<sup>16</sup> Stalin, Moreno-Villacis; Anita, Buestan-Cedill; Eduardo, Garzon-Aldas; (2007). *Leishmaniasis* Cutánea: Presentación en Placa Verrucosa a Propósito de un Caso; Revista Científica. sociedad ecuatoriana de dermatología; 4:1

<sup>17</sup> Hernández-Flores J.J., Morales-Aguirre J.J., et. al., (2007). Visceral *Leishmaniasis* treated with amphotericin B, Bol. Med. Hosp. Infant. Mex. v.64 n.1

<sup>18</sup> Ingerborg Becker. 2010. “Combaten *Leishmaniasis*”, Facultad de Medicina de la UNAM, El Universal.mx, (Nota periodística), <http://www.eluniversal.com.mx/sociedad/6306.html>

<sup>19</sup> L. Sanchez-Saldana; E. Saenz-Anduaga, et.al., *Leishmaniasis*; Dermatología Peruana 2004; vol 14, No 2. 4. [http://www.who.int/leishmaniasis/disease\\_epidemiology/en/index.html](http://www.who.int/leishmaniasis/disease_epidemiology/en/index.html)

<sup>20</sup> Gil E., Cunha L., et. al., (2007). Importancia de los Compuestos Inorgánicos en el Tratamiento de la *Leishmaniasis*, Latin American Journal of Pharmacy. 26 (3): 454-61

<sup>21</sup> Henao H., Osorio Y., et, et al., (2004). Eficacia y toxicidad de los antimoniales pentavalentes (Glucantime ®) y (Pentostame ®) en un modelo animal de *Leishmaniasis* cutánea americana, aplicación de la luminometría. Biomédica. 24(4): 393-402.

<sup>22</sup> Bruton, J., and Laurence L., et. al., (2006). Las bases farmacológicas de la terapéutica de Goodman y Gilman. Vol. II. Edición 11, México-Bogotá, Ed. Mc Graw Hill. 40: 1052-1053, 1314-1316

<sup>23</sup> Alzamora L., Solís H., et al., (2007). Leishmanicidal activity of methanolic extract from four ecotypes of *Lepidium peruvianum*, Chacón (Brassicaceae). Facultad de Ciencias Biológicas UNMSM, Perú-Lima. Rev. peru. biol. 13(3): 211 - 214

<sup>24</sup> Gil E., Cunha L., et al., (2007) Importancia de los Compuestos Inorgánicos en el Tratamiento de la *Leishmaniasis*. Faculdade de Farmacia; Instituto de Química, Universidade Federal de Goiás, Perú-Lima. Latin American Journal of Pharmacy. 26 (3): 454-61.

<sup>25</sup> Bruton J., and Laurence, L., (2006). Manual de Farmacología y Terapeutica de Goodman y Gilman. Edición 11., México-Bogotá, Ed. Mc. Graw Hill, pag. 680-685.

<sup>26</sup> Oficina General de Epidemiología (OGE), Instituto Nacional de Salud (INS). (2000). Medica, Magíster en enfermedades infecciosas y parasitarias por la Universidad de Brasília, Brasil.

<sup>27</sup> Mollinedo Pérez S. (1999). Manual práctico de tratamiento de la Leishmaniasis tegmentaria americana. Unidad de parasitología y entomología, INLASA. Publicación técnica N°10.

<sup>28</sup> Soto J. y Soto P. (2006). Miltefosina oral para el tratamiento de la *Leishmaniasis*. CIBIC, Centro de Investigaciones Bioclínicas de la Fundación FADER, Bogotá, Colombia. Biomédica; 26(Supl.1):207-17

<sup>29</sup> Goodarzi M., et al., (2010). QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides. European Journal of Medicinal Chemistry. 30: 1-11.

<sup>30</sup> Goudarzi N., Goodarzi M., y Chen T. (2012). QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm coupled with multiple linear. Medicinal Chemistry Research; 21:437-443.

<sup>31</sup> Cruz-Montegudo, M., Borges, F., Perez-Gonzalez, M., Cordeiro, M.N.D.S. (2007). Computational modeling tools for the design of potent antimalarial

bisbenzamidines: Overcoming the antimalarial potential of pentamidine. *Bioorganic and Medicinal Chemistry*, 15(15), 5322-5339.

<sup>32</sup> Deeb O., Jawabreh M., (2012). Exploring QSARs for Inhibitory Activity of Cyclic Urea and Nonpeptide-Cyclic Cyanoguanidine Derivatives HIV-1 Protease. Inhibitors by Artificial Neural Network, *avances in Chemical Engineering and Science*. (2) 82-100.

<sup>33</sup> Puzyn T., Leszczynski J., Cronin T. D. *Recent Advances in QSAR Studies: Methods and Applications*. Editorial Springer. New York, U.S.A. pp. 1 - 45.

<sup>34</sup> Pastor M. y Álvarez-Builla. (1999). Técnicas QSAR en diseño de fármacos. Departamento de Química Orgánica. Universidad de Alcalá, Esp., 69: 71-98.

<sup>35</sup> Kubinyi Hugo. Germany E. Manual QSAR Parameters, , [www.kubinyi.de](http://www.kubinyi.de)

<sup>36</sup> Escalona J.C., Carrasco R., Padrón J. A. (2008). Introducción al diseño de fármacos. Folleto para la docencia de la asignatura de Farmacia. Universidad de Oriente. La Habana, Cuba.

<sup>37</sup> Malleshappa N.N., Haron M. (2011). Acomparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinasa (EGFR) inhibitors: A rational approach of anticancer drug designer. King Saud University. *Journal of salud Chemical Society*.

<sup>38</sup> Gago Badenas F. (1994). Métodos computacionales de modelado molecular y diseño de fármacos. Monografía I. Diseño de medicamentos. Real Academia Nacional de Farmacia. Universidad de Alcalá de Henares, Madrid. pp. 253 – 311.

<sup>39</sup> Pastor M. y Álvarez-Builla. (1999). Técnicas QSAR en diseño de fármacos. Departamento de Química Orgánica. Universidad de Alcalá, Esp., 69: 71-98.

<sup>40</sup> Topliss, J. G.; (1993). Some observations on classical QSAR; Perspectives in Drug Discovery and Design, 1: 253-268

<sup>41</sup> González J. E., Poltev V. I. (2002). La simulación computacional de procesos genéticos a nivel molecular. *Elementos* 47. pp. 31 - 35.

<sup>42</sup> Beasley, D., R-Bull, D., Ralph-R, M. (1993). An Overview of Genetic Algorithms: Part 2, Research Topics. *University Computing*, 15(4), 170-181.

<sup>43</sup> Devillers James. (1996). Genetic Algorithms in Molecular Modeling. Principles of QSAR and Drug Desidn. Academic press.

<sup>44</sup> Melián B. B., Moreno P. J. A., Moreno V. J. M. (2009). Algoritmos genéticos: una visión práctica. *Revista Didáctica de las Matemáticas* 71: 29 - 47.

<sup>45</sup> Prado-Prado F. J., Martínez de la Vega O., Uriarte E., Ubeira FM., Chou K.C., González-Díaz H. (2009). Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modelling and comparative multi-distance study of the giant components of antiviral drug–drug complex networks. *Bioorganic & Medicinal Chemistry* 17 (2): 569–575

<sup>46</sup> Molecular Descriptor Guide, Versión 1.0.2. (2008). U.S. Enviromental Protection Agency.

<sup>47</sup> Melián B. B., Moreno P. J. A., Moreno V. J. M. (2009). Algoritmos genéticos: una visión práctica. *Revista Didáctica de las Matemáticas* 71: 29 - 47.

<sup>48</sup> Sarchitect Designer. Versión 2.5, User guide, Algorithms for life. Stand Life Sciences, Copyright 2005-2008.

<sup>49</sup> Todeschini, R, Consonni, V., *et. al.* Chemometrics and QSAR Research Group. Milano Italy. Software Dragon® Versión 5.3 profesional para windows 2005.

<sup>50</sup> Palma A., *et al.*, (2008). Synthesis and in vitro activity of new tetrahydronaphtho [1, 2-*b*] azepine derivates against *Trypanosoma cruzi* and *Leishmania chagasi* parasites. *Bioorganic and Medicinal Chemistry Letters*. 10: 1016.

<sup>51</sup> Acosta M.L., Palma A., Bahsas A., (2010). Rational use of substituted N-allyl and N, N, N-diallylanilines in the stereoselective synthesis of novel 2-alkenyltetrahydro-1-benzazapines. *Tetrahedron* 66: 8392-8401.

<sup>52</sup> Scott Ph., y Ph. D. (2011) *Leishmania — A Parasitized Parasite*. School of Veterinary Medicine, University of Pennsylvania, Philadelphia. *N Engl J Med*; 364:1773-1774

<sup>53</sup> Cheng Y, Prusoff WH (December 1973). "Relationship between the inhibition constant (K<sub>1</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction". *Biochem Pharmacol* 22 (23): 3099–108.

<sup>54</sup> Chemical Structure Drawing Standard 1985-2001. Cambridge Soft Corporation. Software CHemDraw Ultra, version 7.0

<sup>55</sup> Nicolás-Vázquez M. I., Marín C. E., Castro M. F. M., Miranda R. R. 2006. Algunos aspectos básicos de la química computacional. 1a. ed. Universidad Nacional Autónoma de México. Cuatitlán Izcalli, Estado de México. pp. 37 - 38.

<sup>56</sup> Press W, H., *et. al.*, (1968). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press. Chapter 10.

<sup>57</sup> HyperCube Inc., 2000. Software Hyper Chem ® Versión 6.03 para Windows.

<sup>58</sup> Malleshappa N.N., Haron M. (2011). Acomparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinasa (EGFR) inhibitors: A rational approach of anticancer drug designer. King Saud University. Journal of salud Chemical Society.

<sup>lix</sup> Todeshini, R, Consonni V., et al., (2004). Moby Digs User Manual. Milano Italia: Milano Chemometrics and QSAR Research Group.

<sup>lx</sup> Spiegel., Murray R., Larry J. y Stephens; Estadística; 3ra. Edición; (2002); Editorial McGrall-Hill. 317 p.

<sup>lxi</sup> Box, G. E. P.; Huter, W. G.; Hunter, J. S.; (1999). Estadística para investigadores; Editorial Reverté S.A.; 1° Edición México; pp.- 463-524.

<sup>lxii</sup> Solanas A., y Sierra V., (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. Anuario de psicología; Universidad de Barcelona. 55:143-154.

<sup>lxiii</sup> Microsoft Corporation. (2010). Microsoft® Excel® Ayuda de Excel: coef.de.correl (función COEF.DE.CORREL).

<sup>lxiv</sup> Nicolás-Vázquez M. I., Marín C. E., Castro M. F. M., Miranda R. R. (2006). Algunos aspectos básicos de la química computacional. 1a. ed. *Universidad Nacional Autónoma de México*. Cuatitlán Izcalli, Estado de México. pp. 37 - 38.

<sup>lxv</sup> Stanton, D. T.; (2003) On the Physical Interpretation of QSAR Models; J. Chem. Inf. Comput. Sci. 43:1423-1433.

<sup>lxvi</sup> Casoni D., Sârbu C., (2011). Modeling of food preservatives chromatographic lipophilicity applying genetic algorithm and multiple linear regression. Babes-Bolyai

University, Faculty of Chemistry and Chemical Engineering, ACADEMIA ROMÂNĂ.  
56(4), 381-389.

<sup>67</sup> Inanciu Ovidiu., (2000). QSAR and QSPR molecular descriptors computed from the resistance distance and electrical codetencematies. Texas, USA. Departament of marine. Models in Chemistry, 137 :( 5-6), (607-631).

<sup>68</sup> Sarchitect. Theory – Descriptors  
[http://www.strandls.com/sarchitect/documents/manual\\_html/descctheory.html](http://www.strandls.com/sarchitect/documents/manual_html/descctheory.html)  
(2012)

<sup>69</sup> Sun H., Gong H., Liu H., *et al.*, (2010) Theoretical study on the structures, isomerization and stability of SiC<sub>4</sub> isomers. Theor Chem Acc 126:15–25.

<sup>70</sup> Inanciu Ovidiu., (2000). QSAR and QSPR molecular descriptors computed from the resistance distance and electrical codetencematies. Texas, USA. Departament of marine. Models in Chemistry, 137 :( 5-6), (607-631).

<sup>71</sup> Todeschini R., Consonni V., *et.al.* (2009). Molecular Descriptors for Chemoinformatics. Vol. I y II. Wiley-Vch., Verlag Gmbh & Co. kGaA Weinleim.

<sup>72</sup> Todeschini R., Consonni V., (2000). Handbook of Molecular Descriptors. WILEY-VCH, Editorial Board.