

**UCLV**  
Universidad Central  
"Marta Abreu" de Las Villas



**FQF**  
Facultad de  
Química y Farmacia

## **TRABAJO DE DIPLOMA**

**Empleo de técnicas de inteligencia artificial en la identificación de nuevos compuestos con potencial actividad contra *Leishmania major*.**

**Autor: Oliany Salgado Vernier**

**Tutores: MSc. Naiví Flores Balmaseda**

**Dr. Juan Alberto Castillo Garit**

Este documento es Propiedad Patrimonial de la Universidad Central “Marta Abreu” de Las Villas, y se encuentra depositado en los fondos de la Biblioteca Universitaria “Chiqui Gómez Lubian” subordinada a la Dirección de Información Científico Técnica de la mencionada casa de altos estudios.

Se autoriza su utilización bajo la licencia siguiente:

**Atribución- No Comercial- Compartir Igual**

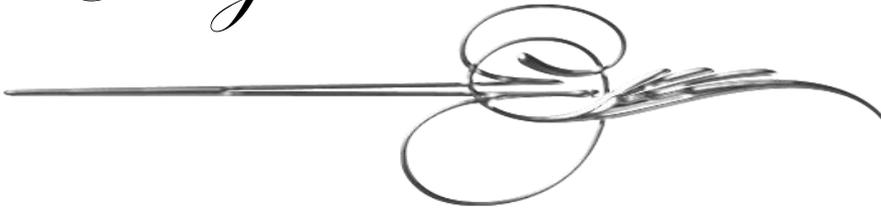


Para cualquier información contacte con:

Dirección de Información Científico Técnica. Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní. Km 5½. Santa Clara. Villa Clara. Cuba. CP. 54 830

Teléfonos.: +53 01 42281503-1419

*Exergo*



*“El conocimiento no es una vasija que se llena, sino un fuego que se enciende.”*

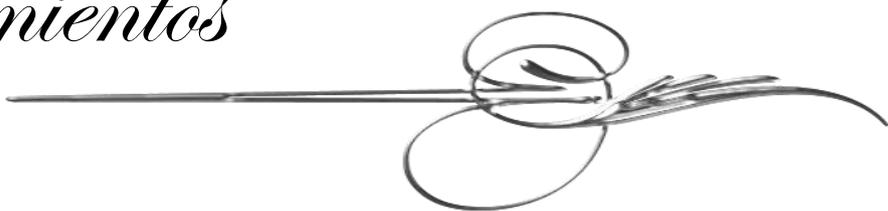
*Plutarco*

*Dedicatoria*



*“A mi mamá por todo el apoyo  
durante estos cinco años.”*

# *Agradecimientos*



*A mi familia por estar presente en todo momento.*

*A todas las personas que fueron partícipe de alguna forma y lo hicieron posible.*

*A mis niñas todas por tener paciencia conmigo y hacer de este tiempo algo inolvidable, de beti sus clases y enojos, de jenni sus confesiones y terapias colectivas, de iliex sus despertares matutinos con algo para contar, de yuliet sus locuras y chistes, de mi sol su risa y sentido del humor, de naylen su cariño, cosas que nos distinguen y hacen de nosotras un grupo divertido, melancólico, que no quiere terminar aquí esta historia.*

## Resumen

La leishmaniasis es una enfermedad zoonótica producida por parásitos, ya sea en su forma amastigote o promastigote, del género *Leishmania*, endémica en, al menos, 88 países ubicados en zonas tropicales y subtropicales. En la actualidad se conocen más de 20 especies del parásito las cuales se agrupan por las diferencias que presentan en cuanto a características ecoepidemiológicas y clínicas en dos subgéneros. Las manifestaciones clínicas de la enfermedad dependen de las propiedades del hospedero y de las propiedades del parásito, afectando así principalmente a menores de cinco años y personas con baja inmunidad. A partir de una base de datos, se realiza el cálculo de descriptores moleculares con el Dragón, además de la utilización de otros programas de estudios QSAR como el weka y otras técnicas de inteligencia artificial. Como resultado se obtuvieron cuatro modelos IBk, J48, MLP, SMO con la capacidad de identificar actividad antileishmaniasica, se realizó el cribado virtual de un total de 5174 compuestos, de los cuales 4662 pertenecen al DrugBank, 492 son compuestos de síntesis y 20 compuestos aislados a partir de plantas, de ellos 165, 127, 24 y 14 respectivamente resultaron activos frente a promastigotes *L. major*.

## Summary

Leishmaniasis is a zoonotic disease caused by parasites, whether in the amastigote or promastigote form, of the genus *Leishmania*, endemic in at least 88 countries located in tropical and subtropical zones. At present more than 20 species of the parasite are known which are grouped by the differences they present in terms of ecoepidemiological and clinical characteristics in two subgenres. The clinical manifestations of the disease depend on the properties of the host and the properties of the parasite, affecting mainly children under five years of age and people with low immunity. From a database, the calculation of molecular descriptors with the Dragon is carried out, in addition to the use of other QSAR study programs such as weka and other artificial intelligence techniques. As a result, four IBk, J48, MLP, SMO models with the ability to identify antileishmaniasis activity were obtained, a total of 5174 compounds were screened, of which 4662 belong to the DrugBank, 492 are synthesis compounds and 20 isolated compounds from plants, of them 165 127, 24 and 14 respectively resulted active against *L. major* promastigotes.



# *Introducción*



La leishmaniasis es una enfermedad zoonótica producida por parásitos del género *Leishmania*, endémica en, al menos, 88 países ubicados en zonas tropicales y subtropicales incluyendo América Central, América del Sur, India, el Oriente Medio, el Norte de África y el Sur de Europa. Se estima que aproximadamente 12 millones de personas se encuentran infectadas por *Leishmania* y cada año se reportan alrededor de 2 millones de nuevos infectados. Aproximadamente, 350 millones viven en zonas de riesgo.(1)3

En la actualidad se conocen más de 20 especies del parásito las cuales se agrupan por las diferencias que presentan en cuanto a características ecoepidemiológicas y clínicas en dos subgéneros: el subgénero *Leishmania* (*L*) cuyas especies se encuentran distribuidas tanto en el Viejo Mundo (Europa, Asia y África) como en el Nuevo Mundo (América) y el subgénero *Viannia* (*V*) cuyas especies están restringidas al Nuevo Mundo. William Leishman y Charles Donovan fueron los primeros en demostrar la presencia del parásito en el bazo de un grupo de pacientes que se pensaba sufrían una enfermedad similar a la malaria y que posteriormente fue conocida como leishmaniasis visceral. Ellos hicieron este descubrimiento simultáneamente pero independientemente en 1903, por esto, la especie descubierta fue denominada *Leishmania donovani* en honor a ellos.(1)

Existen varias formas clínicas de la enfermedad, entre ellas la leishmaniasis cutánea la cual presenta el 90 % de los casos afectados en 8 países: Afganistán, Pakistán, Siria, Arabia Saudita, Argelia, Irán, Brasil y Perú. Por otra parte, el mayor número de casos de leishmaniasis visceral se encuentra en India, Bangladesh, Nepal, Sudán y Brasil.2 En América, los parásitos se transmiten a partir de un reservorio animal o humano por la picadura de insectos hembra hematófagas del género *Lutzomyia*(1)(1-3) y producen las formas clínicas conocidas como leishmaniasis cutánea (LC) cuando afecta la piel, leishmaniasis mucosa (LM) cuando afecta los tejidos mucosos y leishmaniasis visceral (LV) cuando afecta la médula ósea, el hígado y el bazo. A su vez, la LC se puede presentar en forma localizada (LCL) o difusa (LCD) y puede, además, afectar en forma simultánea piel y mucosas, enfermedad que se conoce como leishmaniasis mucocutánea (LMC) 2

El tratamiento de la leishmaniasis no es sencillo, está limitado a unos cuantos medicamentos que no siempre son eficaces y se asocian a importantes efectos adversos. Se cree que en algunos casos que muestran curación clínica, el parásito no ha sido eliminado totalmente del organismo. Pero, ¿por qué la *Leishmania* es tan difícil de tratar? La respuesta probablemente está en las complejas interacciones que se dan entre el parásito y el sistema inmune del ser humano.<sup>1</sup> La prevalencia de la enfermedad viene en aumento (en el período de 2001-2016 se reportaron 55.530 casos humanos de LV en las Américas con un promedio anual de 3.457 casos y en el 2016 la tasa de Incidencia de LC fue 21,71 casos por 100.000 habitantes, presentando un aumento al compararla con el 2015 (18,35/100.000 hab.) (9). Por otro lado, la leishmaniasis ya no se considera una enfermedad rural que afecta principalmente a hombres en edad laboral activa sino que se ha convertido desde hace algunos años en una enfermedad de transmisión doméstica rural y semiurbana, donde toda la familia, especialmente los niños, representan la población de mayor riesgo (10), demostrando el proceso de urbanización de la enfermedad (11,12).<sup>2</sup>

La lesión que se produce en el hospedero vertebrado por la picadura del insecto vector infectado con el parásito, induce una respuesta inflamatoria que involucra la migración de diferentes células, principalmente macrófagos y linfocitos, hacia dicho sitio a fin de reparar el tejido dañado e iniciar el proceso de cicatrización (13). Dentro del hospedero mamífero, los promastigotes inoculados por el vector invaden las células del sistema de células fagocíticas mononucleares (monocitos/macrófagos) presentes en el sitio de la infección al igual que las células reclutadas durante la respuesta inflamatoria o que son fagocitados por ellas (14). Una vez intracelulares, los promastigotes se transforman en amastigotes para así sobrevivir y multiplicarse dentro del fagolisosoma de su célula hospedera (1)(15-17).<sup>2</sup>

En 2017, los delegados de los países endémicos participantes de la Reunión Regional de leishmaniasis, aprobaron el Plan de Acción de la leishmaniasis en las Américas 2017-2022, que detalla las metas, indicadores y acciones para cumplir con los compromisos de la Resolución CD 55 R09 del 2016, lo cual incluye entre otros aspectos la necesidad de descubrir y nuevos fármacos para el tratamiento de la enfermedad.<sup>(1)</sup>

El enfoque de diseño/descubrimiento de fármacos asistido por computadoras ofrece una alternativa más económica y rápida en el campo de la síntesis y evaluación de compuestos con determinada actividad farmacológica(2). Estos métodos llamados “*in silico*” engloban todas las técnicas asistidas por ordenadores usadas en el diseño/descubrimiento y optimización de compuestos con propiedades deseadas y ha jugado un rol fundamental en el desarrollo de un número de fármacos que actualmente se encuentran en el mercado. Los procedimientos “*in silico*” evitan los altos costos de los procesos de síntesis y bioensayos y se realizan solamente después de la exploración de los conceptos iniciales con modelos de Relaciones Cuantitativas Estructura-Actividad/Propiedad [conocidos internacionalmente por sus siglas en inglés QSAR/QSPR (*QuantitativeStructureActivity/PropertyRelationships*)] (3). En la actualidad la Facultad de Química y Farmacia de la Universidad Central “Marta Abreu” de las Villas realiza investigaciones dirigidas al desarrollo de nuevos principios activos para la industria farmacéutica en las ramas de la medicina veterinaria, humana y la sanidad vegetal. En la misma se encuentra, el Grupo de Diseño Molecular Asistido por Computadora e Investigaciones Bioinformáticas (CAMD-BIR Unit, por sus siglas en inglés) ha presentado resultados relevantes en este campo empleando métodos computacionales cuya aplicación en estudios QSAR ha permitido el desarrollo de nuevos métodos grafo-teóricos para el diseño molecular de compuestos orgánicos potencialmente bioactivo(2).

Los compuestos con actividad antileishmaniásica son objeto de estudio a nivel mundial, pero dada la diversidad de especies del género *Leishmania*, el ciclo de vida del parásito, y la variedad de manifestaciones clínicas que presenta la enfermedad se hace muy engorroso encontrar nuevos compuestos efectivos(4).

Por todo lo anterior podemos identificar el siguiente Problema científico: En la actualidad no existen fármacos totalmente efectivos para el tratamiento de las formas de leishmaniasis causadas por *Leishmania major*.

Como vía para darle solución se formula la siguiente hipótesis: Es posible la identificación de nuevos compuestos con potencial actividad contra *Leishmania major* mediante el empleo de métodos QSAR de inteligencia artificial.

Teniendo en cuenta lo anterior, se propone como objetivo general:

Identificar nuevos compuestos antileishmaniásicos mediante el empleo de métodos *in silico*.

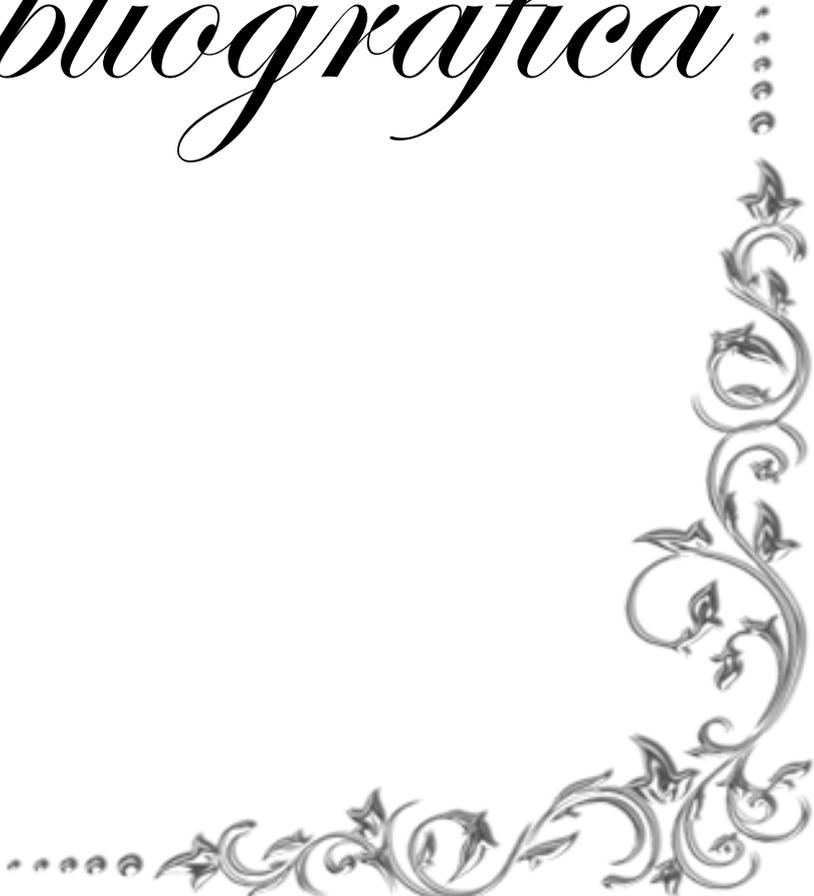
Siendo nuestros objetivos específicos:

- Confeccionar una base de datos de compuestos a los que se les ha evaluado experimentalmente su actividad contra promastigotes de *Leishmania major*.
- Obtener modelos QSAR que permitan, tras su validación, predecir la actividad de compuestos frente a promastigotes de *L. major* utilizando descriptores moleculares.
- Identificar nuevos compuestos con potencial actividad contra este parásito mediante el cribado virtual de bases de datos.



*Revisión*

*Bibliográfica*



## 1.1 Leishmaniasis. Generalidades

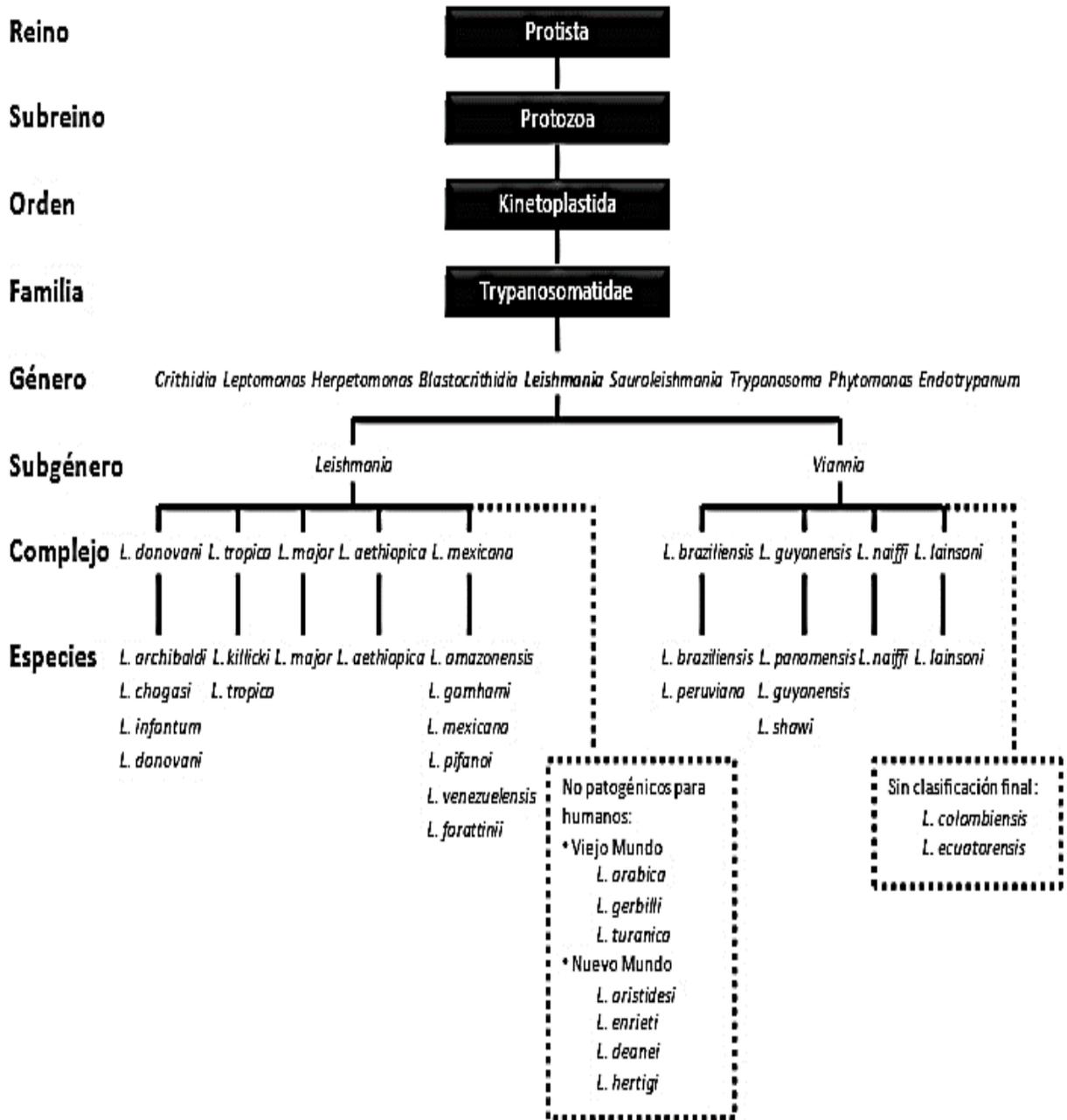
La leishmaniasis es una enfermedad zoonótica causada por protozoos del género *Leishmania* (*L.*). Sus manifestaciones clínicas van desde úlceras cutáneas, que cicatrizan espontáneamente, hasta formas fatales en las cuales se presenta inflamación severa del hígado, el bazo, trastornos renales y afectaciones de la médula ósea(5, 6). El agente se transmite a través de la picadura de hembras de los mosquitos chupadores de sangre pertenecientes a los géneros *flebotomos* del viejo mundo y *Lutzomyia* del nuevo mundo, de la familia *Psychodidae*.(1, 5) Afecta tanto a perros como a humanos. Sin embargo, animales silvestres como zarigüeyas, coatíes, osos hormigueros, ratas, zorros, lobos entre otros, son portadores asintomáticos del parásito, por lo que son considerados como animales reservorios(5, 7, 8).

Existen dos tipos epidemiológicos de leishmaniasis: la forma antroponótica, entre persona enferma, flebótomo que actúa de vector y persona sana; y la forma zoonótica, entre un animal enfermo (reservorio), el flebótomo (vector) y la persona sana. Algunas de las especies tienen como reservorio a roedores del entorno doméstico, que pueden dar proporciones epidémicas a la enfermedad en el ser humano si la infección se propaga entre ellos, así como también, la forma antroponótica de la leishmaniasis visceral en la India(9). Es un riesgo real para los viajeros, en especial si se visitan zonas rurales áridas del norte de África y el suroeste de Asia, o endémicas de América Latina.(1, 8, 9).

## 1.2 Taxonomía

La leishmaniasis es causada por parásitos intracelulares del orden *Kinetoplastida*, familia *Trypanosomatidae* y del género *Leishmania*. Su taxonomía es compleja y ha evolucionado a lo largo de los años. Figura 1. En el pasado, eran clasificadas según la localización geográfica y el aspecto clínico que producía la enfermedad. En la actualidad, las agrupaciones taxonómicas se basan en características bioquímicas y moleculares, por lo que ya no coinciden con los patrones clínicos específicos de la enfermedad. De esta manera, el género *Leishmania* se divide en dos subgéneros: *Leishmania* y *Viannia*. Todos los miembros del género *Leishmania* son parásitos de mamíferos. Los dos subgéneros están separados en base a su localización en el intestino del vector cada subgénero se divide en complejos. En el subgénero *Leishmania* se encuentran cinco complejos: *L. major*, *L. tropica*, *L. aethiopica*, *L.*

*donovani* y *L. mexicana*. Los complejos de *L. major*, *L. tropica* y *L. aethiopica* están formados por una sola especie que le da nombre a su respectivo complejo.4

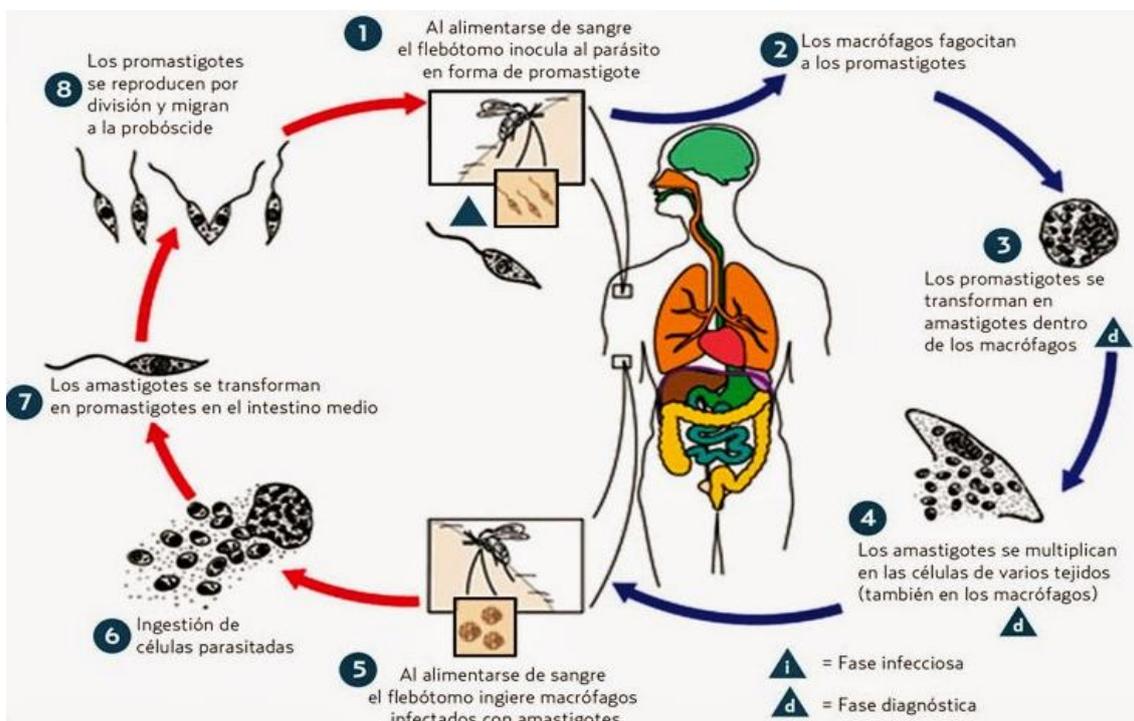


**Figura1.** Especies de *Leishmania*.

### 1.3 Ciclo de Vida

El ciclo comienza cuando el vector ingiere sangre humana con los macrófagos cargados de *Leishmania*. Figura 2. En el intestino medio del vector, el parásito sale de

los macrófagos y se transforma en el promastigote procíclico que es la forma extracelular que se divide activamente y que se encuentra adherida a la pared intestinal. Posteriormente, se transforma en el promastigote metacíclico que es la forma que no se multiplica ni se adhiere a la pared del intestino. De esta manera, es arrastrado hacia las piezas bucales del insecto. El insecto entonces se vuelve infectante para la próxima persona que sea picada. Este proceso demora entre 4 a 7 días, período que coincide con el tiempo aproximado en el cual el insecto necesita volver a alimentarse. Al picar al ser humano, le inyecta los promastigotes metacíclicos que rápidamente invaden a los macrófagos residentes y se transforman en amastigotes intracelulares. De esta manera, el ciclo continúa.(10, 11)



**Figura 2.**Ciclo de Vida de *Leishmania* ssp.

#### 1.4 Formas de *Leishmania*.

El ciclo biológico de la *Leishmania* comprende esencialmente el paso alternativo de un hospedador vertebrado a otro invertebrado, y viceversa, con dos formas morfológicas principales, la intracelular o amastigote en las células del sistema fagocítico mononuclear del hospedador vertebrado y la forma extracelular o promastigote en el tracto intestinal de los flebótomos.

La forma amastigote, inmóvil, se presenta al microscopio óptico tras tinción con colorantes habituales, como un cuerpo oval, de una longitud y anchura que oscilan, respectivamente, entre 3-5  $\mu\text{m}$  y 1,5-2,5  $\mu\text{m}$ . En su citoplasma, se observa un núcleo voluminoso y esferoidal, generalmente excéntrico, y un kinetonúcleo próximo al núcleo de aspecto bacilar o bastoniforme(12). Son exclusivamente intracelulares (macrófagos) pero pueden encontrarse en el intersticio en los casos en los que el parásito se replica hasta producir la ruptura de la célula huésped(5).

La forma promastigote, fusiforme, extracelular y móvil, presenta un tamaño superior. Su talla varía entre los 10-30  $\mu\text{m}$  de largo y los 1,5-3  $\mu\text{m}$  de ancho. Si bien en el examen en fresco sólo puede observarse un largo flagelo libre en su región anterior, su tinción permite observar además la presencia de un núcleo oval central y un kinetonúcleo bastoniforme claramente pre nuclear. En 1979 se demostró que el parásito presenta en el tubo digestivo del vector otro estadio flagelado, distinto al promastigote, al que se denominó paramastigote debido a la situación lateral del kinetoplasto respecto al núcleo(13).

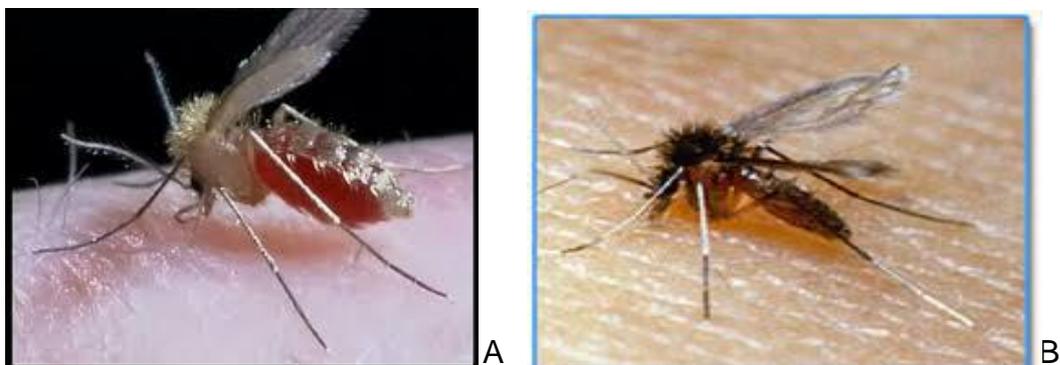
## 1.5 Vectores

Los flebotomos son pequeñas moscas de 2 a 3 mm, con el cuerpo y las alas cubiertos de pelos. Abundan todo el año en la zona tropical y en verano en zona templada. Existen más de 600 especies o subespecies, de las cuales menos de 50 han sido implicadas en la transmisión de las leishmaniasis. Se reconocen cinco géneros principales: *Phlebotomus*, *Warileya*, *Sergentomyia*, *Lutzomyia* y *Brumptomyia*. Solamente *Phlebotomus* en el Viejo Mundo y *Lutzomyia* en el Nuevo Mundo son reconocidas como vectores de las leishmaniasis (14, 15).Figura 3.

Descansan durante el día en los rincones y vuelan al atardecer. Las hembras, que son las únicas hematófagas, son más activas a la caída del día, estas ingieren sangre para desarrollar sus ooquistes, lo cual dura más o menos una semana, durante la cual reposa. Ponen entre 50 y 100 huevos en suelos ricos en humus y humedad, y eclosionan en 8 a 10 días. Las larvas se desarrollan hasta el estado adulto en 1 a 2 meses, según la temperatura. Los flebotomos infectados tienen dificultad en absorber su alimento sanguíneo lo cual puede ser un factor de multiplicación de las picaduras (picaduras exploratorias) y por ende un aumento de la transmisión. Esta es limitada

por la capacidad intrínseca de cada especie, por su antropofilia, y por su esperanza de vida (14, 16).

Varias especies pican también en el día cuando son molestadas por el ingreso del hombre a su medio ambiente (efecto de intrusión). Esta actividad juega un papel muy importante en la contaminación de los que trabajan en el desmonte de la selva. Las hembras son predominantemente exófilas (reposan fuera de las casas durante la maduración de los huevos) y endófilas (que reposan dentro de las casas durante la maduración de los huevos), estas últimas se pueden controlar mediante la fumigación con insecticidas de las paredes internas de las habitaciones(14, 15).



**Figura 3.** Vectores de los géneros *Phlebotomus* (A) y *Lutzomyia* (B)

### 1.6 Reservorios

Los reservorios son animales domésticos, como el perro, y otros animales salvajes (figura 4). En una región dada, una especie de *Leishmania* es fácilmente mantenida por un sólo huésped reservorio, sin importar si otros mamíferos pueden ocasionalmente encontrarse infestados.

Un animal es reconocido como reservorio si representa, en el medio natural, una fuente de infección para los flebótomos y si la tasa de animales que se encuentran contagiados es elevada, además debe presentar una infección prolongada que no le produzca la muerte rápidamente, para que represente una fuente real de infección. A diferencia de los animales salvajes, el perro presenta los síntomas clínicos y muere de leishmaniasis, por invasión de sus vísceras en un plazo de seis meses a dos años (11, 14, 17).



**Figura 4.** Reservorios animales entre ellos perro, perezoso y rata.

### 1.7 Principales moléculas de superficie

Los promastigotes procíclicos están cubiertos por un glicocálix grueso de 7 nm de espesor. El glicocálix de los promastigotes metacíclicos es aún más grueso con aproximadamente 17 nm de espesor; sin embargo, los amastigotes prácticamente no tienen glicocálix. El glicocálix está compuesto por glicoproteínas y otras sustancias glicosiladas que se encuentran ancladas a la membrana celular por una unión de glicosilfosfatidilinositol (GPI). Una de las moléculas más importantes de la superficie de los promastigotes es el lipofosfoglicano (LPG). La estructura de LPG varía según las especies de *Leishmania* pero básicamente está compuesto por unidades repetitivas de un disacárido y un fosfato unidos a la membrana por el ancla de GPI. Las variaciones en las cadenas laterales que se unen a la estructura central de LPG son importantes entre las distintas especies de *Leishmania*. La *L. major* tiene una estructura altamente ramificada; mientras que el LPG de *L. donovani* prácticamente no tiene ramificaciones. El LPG juega un papel importante en la supervivencia del parásito y en la modulación de la respuesta inmunológica del hospedero.

La glicoproteína 63 (gp63) también llamada leishmanolisina o proteasa mayor de superficie es una proteinasa de superficie de 63 kDa. Se expresa en grandes cantidades (más de 500 mil copias) y está distribuida en todo el cuerpo del promastigote incluyendo el flagelo.<sup>7</sup> A pesar de esto, es 10 veces menos frecuente que el LPG. La gp63 es una metaloproteínasa de zinc con un amplio rango de sustratos tales como caseína, gelatina, albúmina, hemoglobina y fibrinógeno.<sup>(18)</sup> La

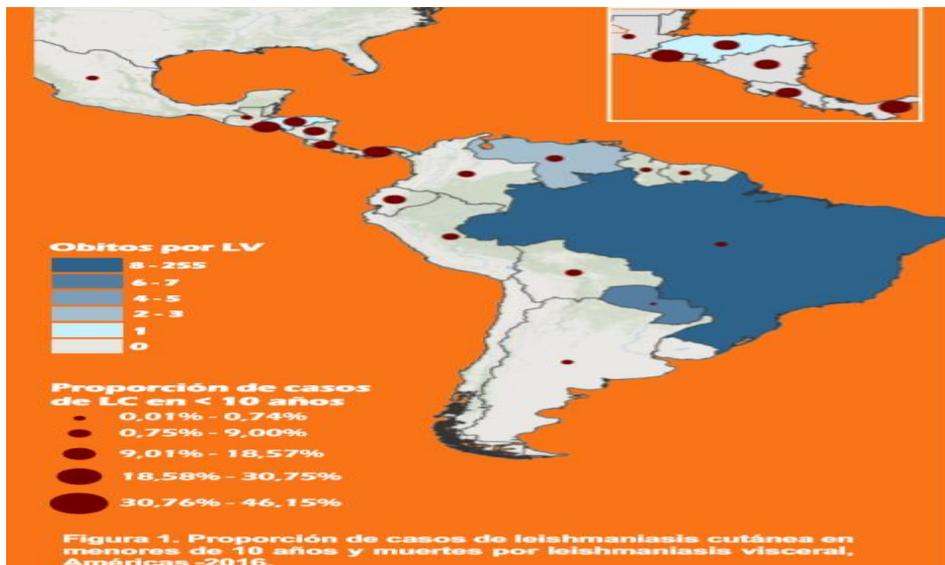
gp63 se relaciona con la entrada del parásito al macrófago y favorece la fagocitosis y la sobrevivencia del amastigote dentro del macrófago. También, afecta la función del complemento aumentando la resistencia del parásito a la lisis mediada por complemento.<sup>7</sup> La molécula más abundante en la superficie de los promastigotes es el fosfolípido de glicosilinositol (GIPL), un tipo de glicolípido con un ancla de GPI. El GIPL es 10 veces más abundante que el LPG; sin embargo, su tamaño pequeño lo mantiene cerca de la membrana celular y cubierto por las moléculas de LPG. Se cree que tiene un rol protector en la superficie del promastigote.<sup>(19)</sup>

### **1.8 Distribución geográfica**

La distribución de los casos de leishmaniosis abarca tanto el Nuevo como el Viejo mundo, en regiones de clima tropical y subtropical, siendo los países más afectados a nivel mundial: Más del 90% de los casos de leishmaniasis visceral se presenta en cinco países: Bangladés, India, Nepal, Sudán y Brasil; cerca del 90% de los casos de leishmaniasis mucocutánea se producen en tres países: Bolivia, Brasil y Perú; y el 90% de los casos de leishmaniasis cutánea se presentan en siete países: Afganistán, Arabia Saudí, Brasil, Argelia, Irán, Perú, Paraguay y Siria. La incidencia anual se estima entre uno y 1,5 millones de casos de leishmaniasis en el mundo. (8, 20) La *L. major* provoca la forma cutánea de leishmaniasis y predomina en la región de África al sur del Sahara, en la zona suroeste de Asia y Medio Oriente.

En la región americana, el 96% de los casos de leishmaniasis visceral se reportan en Brasil y desde el 2012, las muertes por esta enfermedad han aumentado, alcanzando en el 2016 una tasa de letalidad de 7,9% en las Américas, que es considerada como la más elevada cuando es comparada con otros continentes. Figura 5. La proporción de casos de leishmaniasis cutánea en niños menores de 10 años en la región también alcanzó en el 2016, su mayor valor (15,5%) y algunos países, registraron más del 40% de la población afectada en ese grupo etario. Cerca del 96% de los casos en esta región, están concentrados en Brasil donde se destaca las regiones Noreste, Sudeste y Centro-Oeste, no obstante, se observa una expansión geográfica en Argentina, Colombia, Paraguay y Venezuela. En 2016, también se puede observar esa dispersión en Roraima, Norte del Brasil, donde fueron registrados casos de leishmaniasis en las áreas de frontera con Venezuela, lo que requiere una mayor

atención y fortalecimiento de la vigilancia en los municipios de esos dos países.(20, 21)



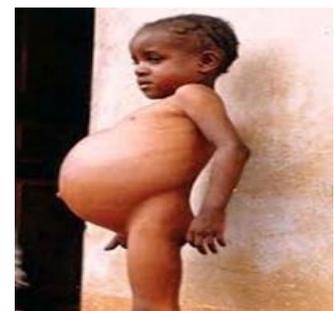
**Figura 5.** Repartición territorial de LC y LV en Las Américas.

### 1.9 Formas clínicas de la leishmaniasis

Las manifestaciones clínicas de la enfermedad dependen de las propiedades del hospedero y de las propiedades del parásito. En el caso del parásito es importante la especie de *Leishmania* y su perfil isoenzimático. Existen dos formas clínicas principales de leishmaniasis:

#### Leishmaniasis visceral (LV)

Se caracteriza por la aparición de fiebre, pérdida de peso, hepatoesplenomegalia, linfadenopatía, pancitopenia e hipergammaglobulinemia entre 3 a 8 meses después de la picada del vector infectado.(21) Puede existir hiperpigmentación de la piel por lo cual se le conoce como enfermedad negra o “kala azar”. Figura 6



**Figura 6.** Paciente que padece LV

Usualmente tiene un curso crónico y puede llevar a la muerte, generalmente por infecciones secundarias. Es causada principalmente por *L. donovani*, *L. infantum* y *L. chagasi*; sin embargo, *L. amazonensis* y *L. tropica* también la pueden producir. Una variante es la leishmaniasis post-kala azar que es causada por *L. donovani* en la India y África. Aparece un tiempo después de un cuadro resuelto de leishmaniasis visceral

y se caracteriza por máculas, pápulas o nódulos que inician en una distribución peri-oral y luego se diseminan a otras áreas del cuerpo.(21)

### Leishmaniasis cutánea (LC) y mucocutánea (LMC)

Existen tres formas clínicas principales de LC: LC localizada (LCL), LC difusa (LCD) y leishmaniasis mucocutánea (LMC). La LCL se caracteriza por la aparición de una o múltiples pápulas que aumentan de tamaño en áreas expuestas. Estas lesiones típicamente se ulceran (Ver figura 7). Las lesiones se localizan principalmente en las extremidades inferiores. Típicamente cursan con linfadenopatía y lesiones satélites.(17) La lesión puede demorar entre 3 y 18 meses para curar en más del 90 % de los casos.(22) El período de incubación oscila entre 2 a 6 semanas. Las principales especies causales son miembros de los complejos *L. mexicana*, *L. braziliensis*, *L. guyanensis*, *L. tropica*, *L. aethiopica* y *L. major*. (21)Una variante de la LCL es la leishmaniasis recidivans, que se caracteriza por la aparición de lesiones tuberculoides alrededor de las cicatrices de úlceras cutáneas curadas. Las lesiones tienen una cuenta baja de parásitos en la biopsia y tienden a ser rebeldes al tratamiento.(22)



**Figura 7.** Formas clínicas de leishmaniasis **A** leishmaniasis cutánea, **B** LMC y **C** LCD

La LCD es una forma anérgica de la enfermedad en la que ocurre diseminación de lesiones nodulares cargadas en parásitos en toda la piel. Raras veces afecta la cara, manos y pies. Ocurre con mayor frecuencia con especies del complejo *L. mexicana* y con *L. aethiopica*.<sup>2</sup> La LMC también se conoce como espundia y se caracteriza por inflamación granulomatosa de la mucosa de la nariz, cavidad oral y faringe. Puede existir destrucción del tabique nasal produciendo nariz de tapir y además la inflamación del paladar genera el signo conocido como la “cruz palatina de Escomel”.

Las lesiones mucosas pueden aparecer simultáneamente con un cuadro de LC; sin embargo, lo más frecuente es que aparezcan después. Inclusive pueden transcurrir años desde que la lesión cutánea ha cicatrizado.(17) El período de incubación más frecuente es de 1 a 3 meses. Las especies más frecuentemente asociadas son: *L. braziliensis*, *L. panamensis*, *L. guyanensis*, y *L. amazonensis*.(22)

#### Leishmaniasis dérmica postkala-azar (PKDL)

Es la forma más severa y normalmente mortal si no se trata. El periodo de incubación puede ser de meses o años. Afecta a los órganos internos, produce fiebre irregular, pérdida de peso, hinchazón de bazo e hígado y anemia; también puede afectar a la mucosa bucal, genital y conjuntival. Tras el tratamiento y recuperación, los pacientes pueden desarrollar LC crónica, la cual requiere un tratamiento largo y caro (23). La PKDL suele aparecer entre 6 meses y 1 año o más, después de la aparente curación de la leishmaniasis visceral. Es posible que no haya antecedentes de leishmaniasis. Las máculas hipopigmentadas o eritematosas presentes en cualquier parte del cuerpo pueden volverse papulares o nodulares e infiltrativas, especialmente en la cara (figura 8). Las personas que desarrollan esta manifestación crónica constituyen verdaderos reservorios, por la presencia de nódulos cargados de parásitos en algunas zonas del cuerpo (15, 24).



**Figura 8.** Leishmaniasis dérmica postkala-azar (PKDL)

#### **1.10 Situación Epidemiológica**

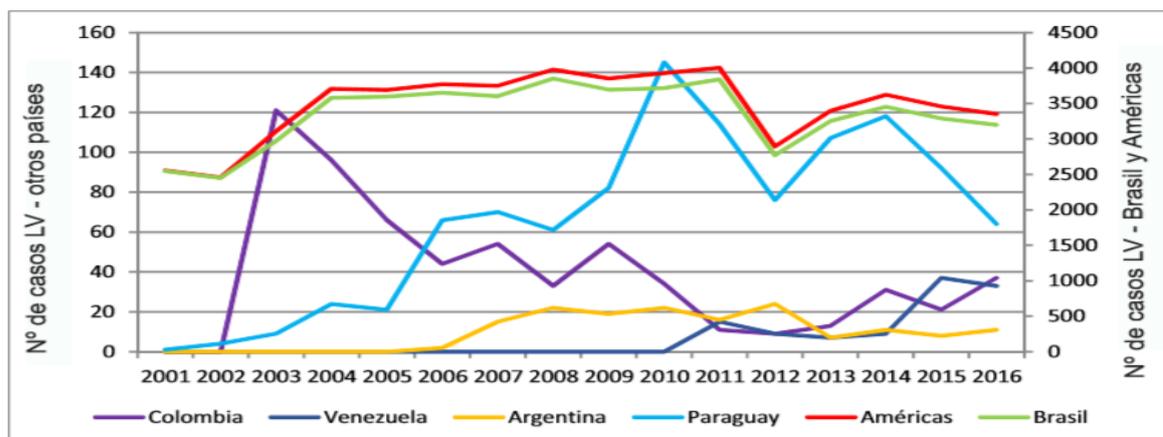
La leishmaniasis cutánea (LC) y mucosa (LM) es endémica en 18 países de las Américas, sin embargo, los casos reportados a la Organización Panamericana de la Salud (OPS/OMS), no incluyen los datos de Guyana Francesa, dado que son reportados directamente a Francia. En el período 2001-2016 se notificaron 892.846 nuevos casos de LC, distribuidos en 17 de los 18 países endémicos. A partir del 2009 y hasta el 2015, hubo una reducción en el registro de casos de la LC en la región. En el 2016 se registró un aumento del 6.15% con relación al año anterior, siendo

observado ese mismo aumento en los registros de casos de las sub regiones Centroamericana (66%) y Andina (27%). La LV es una enfermedad potencialmente fatal con distribución mundial en 76 países, siendo endémica en 12 países de las Américas. (figura 9)(20)

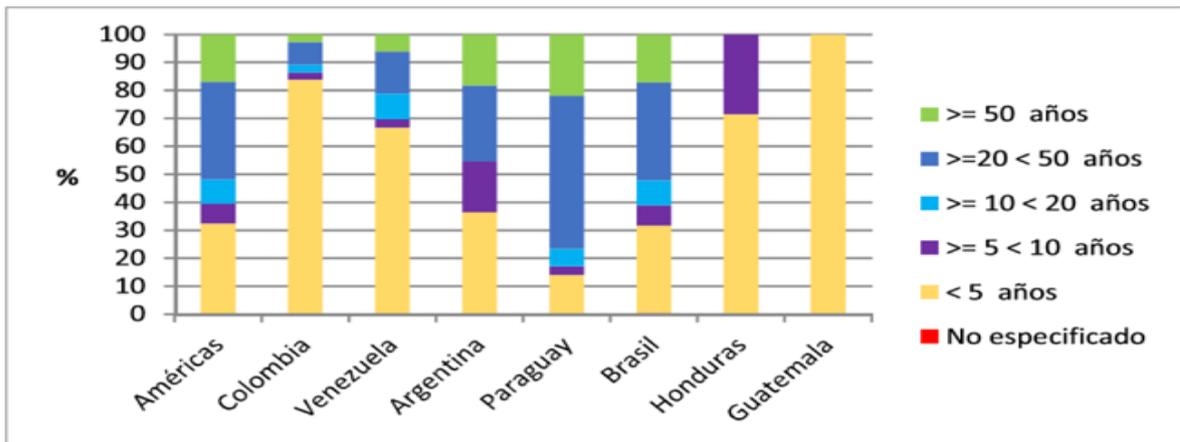


**Figura 9.** Casos de LV, América, 2016.

En el período de 2001-2016 se reportaron 55.530 casos humanos de LV en las Américas con un promedio anual de 3.457 casos. En el 2016, se observa una reducción del 67% en el número de casos de LV en Paraguay cuando es comparado con el 2013. En ese mismo período, los casos en Colombia y Venezuela tuvieron un incremento, de 13 a 37 y de 7 a 33 casos/año, respectivamente (Figura 10). El mayor número de casos de LV fue reportado en el sexo masculino con 66,3% (2.223) y los grupos de edad más afectados fueron los < de 10 años con un 39,1% (1.313), seguido por el grupo etario de  $\geq 20 < 50$  años (34,7%) y de mayores de 50 años (16,96%). En Colombia, Honduras y Venezuela los niños menores de cinco años son los más afectados con el 83,78%, 71,43 % y 66,67% de los casos, respectivamente. Además, en Guatemala, los dos casos ocurridos fueron en ese mismo grupo etario (Figura 11).(20)

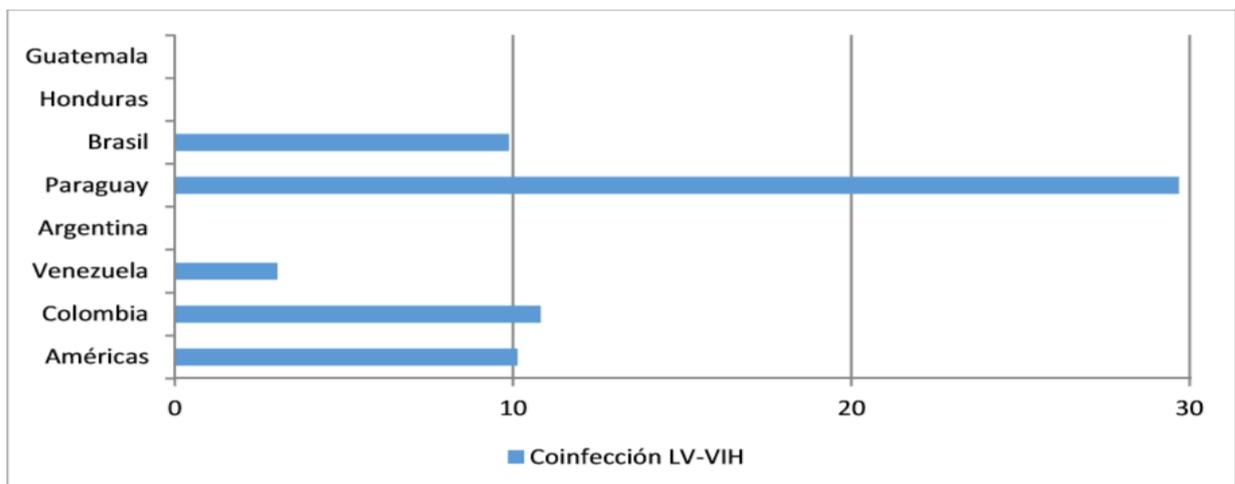


**Figura 10.** Casos de LV, en los países con mayor número de casos, América, 2001-2016.



**Figura 11.** Proporción de casos de LV por grupos de edad y países, América, 2016.

En 2016, hubo un incremento de 32% en los casos de coinfección LV- VIH, con un total de 340 casos (10,14%), siendo que 316 fueron registrados en Brasil, 19 en Paraguay, 4 en Colombia y un caso en Venezuela (Figura 12).(20)



**Figura 12.** Proporción de casos de coinfección LV-VIH por países, América, 2016.

### 1.11 Inmunología contra la leishmaniasis: el modelo de *Leishmania major*

Los estudios realizados tanto *in vitro* como *in vivo* en el modelo murino de infección por *Leishmania (L) major* (una especie de *Leishmania* presente solamente en el Viejo Mundo) han demostrado que la resolución de la infección o enfermedad depende de los mecanismos de respuesta inmune mediada por células Th (CD4+) y sus productos; las células efectoras Th1 son las que se asocian con el control de la infección o con una respuesta cicatrizante a través de la producción de interferón gamma (IFN- $\gamma$ ) y la posterior producción de factor de necrosis tumoral alfa (FNT- $\alpha$ ) por parte de los macrófagos activados debido a la acción del IFN-  $\gamma$ ; por su parte, las

células efectoras Th2 se asocian con el desarrollo de formas graves de la enfermedad por medio de la producción de interleucina (IL) 4 e IL-10 (18).

La interacción de la *Leishmania* y el sistema inmune es muy compleja y varía según la especie de *Leishmania* y según las características del hospedero. La infección por *Leishmania major* es una de las más estudiadas y esta puede llegar eventualmente a generar una respuesta inmune protectora contra subsecuentes infecciones por la misma especie; sin embargo, esto no ocurre con todas las especies de *Leishmania*.(21) La interacción con el sistema inmunitario comienza desde el momento de la picadura del insecto. La introducción de las piezas bucales lacera los vasos sanguíneos de la unión dermo-epidérmica. Además, el insecto inyecta sustancias vasodilatadoras y la enzima hialuronidasa provocando la formación de una pequeña “piscina de sangre” con la que se alimenta. Inmediatamente, el sistema inmune innato interviene para detener este proceso mediante la activación del complemento, la liberación de cininas que producen vasoconstricción, la activación de la coagulación que produce trombosis de los vasos lacerados y la llegada de macrófagos residentes y de neutrófilos sanguíneos al sitio de la lesión. Esto en teoría podría limitar la capacidad del insecto para alimentarse y para transmitir la leishmaniasis. Sin embargo, la saliva del insecto contiene potentes sustancias vasodilatadoras: maxadilán (*Lutzomyia*) o adenosina (*Phlebotomus*), antiagregantes plaquetarios, apirasa y sustancias estimuladoras de la producción de prostaglandina E2 (PGE2). Estos mecanismos utilizados por el insecto para favorecer su alimentación reducen la inflamación y facilitan la transmisión de la *Leishmania*.(21) De hecho, en un estudio se demostró que las lesiones de ratones inoculados con promastigotes de *L. major* y saliva de flebótomo crecieron más grandes y más rápido que las lesiones de ratones inoculados exclusivamente con promastigotes de *L. major*.(25)

### **1.12 Tratamiento general de la leishmaniasis**

El tratamiento solo se debe administrar tras la confirmación de la enfermedad, ya que varía de acuerdo al tipo de leishmaniasis, la severidad de la misma, el lugar donde se infecta el individuo y el estado nutricional e inmunológico del paciente(15, 26).La política farmacéutica de los países endémicos y las decisiones terapéuticas deberían basarse en la relación entre los riesgos y los beneficios de cada medicamento, el tipo

de servicio de salud, la disponibilidad de medicamentos antileishmaniásicos y consideraciones de salud pública, como la prevención de la fármaco-resistencia (15). Dado que no existen vacunas para prevenir la leishmaniasis humana, la mejor forma de protección de los individuos frente a dicha enfermedad, recae en el tratamiento terapéutico de enfermos y reservorios(27).

#### *Antimoniales pentavalentes (SbV):*

Los antimoniales pentavalentes han sido el tratamiento convencional de primera línea contra la leishmaniasis visceral. El desoxicolato de amfotericina B y la pentamidina se han utilizado como fármacos de segunda línea (15, 28). El antimoniato de meglumina y el estibogluconato de sodio (figura 11) se consideran fármacos de primera línea de tratamiento (26, 29), poseen similitud química, su toxicidad y eficacia están relacionadas con el contenido de antimonio. La inyección debe administrarse por vía I.M o I.V en infusión con aguja fina para evitar todo riesgo de trombosis. Para tratar la LC estos fármacos pueden administrarse directamente en las lesiones, a pesar de ello el tratamiento sistémico se debe elegir siempre como primera opción (27, 30).

Los efectos colaterales comunes son: anorexia, vómitos, náuseas, dolor abdominal, malestar, mialgias, artralgias, cefaleas, sabor metálico, letargo, elevación de las enzimas pancreáticas, leucopenia, anemia y trombocitopenia. En dependencia de la dosis y la duración del tratamiento ocurren alteraciones electrocardiográficas, prolongación del intervalo QT y arritmias. La hepatotoxicidad, cardiotoxicidad y la muerte súbita son efectos colaterales graves, pero infrecuentes, en caso de que aparezcan se cambiará de fármaco. Los SbV siguen siendo eficaces y en África, América del Sur, Bangladesh, Nepal e India (excepto North Bihar), prefieren el tratamiento a la dosis de 20 mg/kg/día por vía parenteral durante 28-30 días(15).

#### *Desoxicolato de amfotericina B*

La amfotericina B es un antibiótico poliénico. Se recomienda administrar una dosis de prueba de 1 mg, en infusión, seguida de una dosis plena 4 a 6 horas más tarde. Son frecuentes las reacciones a la infusión, como la fiebre alta y los escalofríos, al igual que la tromboflebitis de la vena inyectada. La nefrotoxicidad también es común, y lleva a la interrupción frecuente del tratamiento en algunos pacientes. Otros efectos tóxicos infrecuentes, pero graves, son la hipopotasemia y la miocarditis. El tratamiento

debería administrarse siempre en un hospital para que el paciente pueda ser sometido a monitorización continua. La anfotericina B se había usado raramente en la leishmaniasis, si bien actualmente se emplea como alternativa a la pentamidina o a los antimoniales en el tratamiento de la leishmaniasis visceral; los inconvenientes son: su costo, (entre 25.000 y 312.000 pesetas según el preparado comercial y el esquema terapéutico empleado) casi el doble que el de los antimoniales, y sus efectos adversos(31, 32).

Se han utilizado varias formulaciones, como la anfotericina B liposómica, el complejo lipídico de anfotericina B o la dispersión coloidal de anfotericina B, cuya eficacia es similar a la del desoxicolato de anfotericina B, pero tienen una toxicidad significativamente menor. Se administran en infusión intravenosa durante 2 h. En algunos pacientes se producen reacciones leves a la infusión y dolor de espalda. Ocasionalmente se observan nefrotoxicidad o trombocitopenia transitorias. En la cuenca mediterránea, el tratamiento de elección para los pacientes inmunocompetentes es la anfotericina B liposómica (L-AmB) en una dosis total de 40 mg/kg. La infusión periódica de L-AmB ha demostrado prevenir la recaída en este subconjunto de los pacientes (15).

### *Paromomicina*

La paromomicina (aminosidina) es un antibiótico aminoglucósido que se suele administrar por vía intramuscular. Las dosis de 15 mg/kg y 20 mg/kg en forma de sulfato son equivalentes a 11 mg/kg y 15 mg/kg en forma de base, respectivamente. El acontecimiento adverso más frecuente es el dolor leve en el lugar de la inyección, en un 2% de los casos se produce ototoxicidad reversible. La nefrotoxicidad es rara. Algunos pacientes pueden presentar hepatotoxicidad, por la elevación de las enzimas hepáticas; también se han descrito casos de tetania. Existe una formulación tópica para la leishmaniasis cutánea (15). La principal ventaja de la paromomicina es su bajo coste, en comparación con el resto de terapias (33).

### *Pentamidina*

La pentamidina es una diamidina con acción frente a ciertas especies de *Leishmania*(31, 34). Se une a los tejidos y se excreta de forma prolongada en el tiempo, lo que permite su administración en días alternos. Está disponible sólo por vía

parenteral(15).Con frecuencia produce taquicardia, hipotensión, náuseas, vómitos, mareos, exantema, enrojecimiento facial y sabor metálico. En el 6% a 9% de los pacientes produce hipoglucemia como resultado de la destrucción de las células del páncreas, y excepcionalmente hiperglucemia y diabetes mellitus. Hasta en un 25% de los pacientes tratados aparece deterioro de la función renal. Otros efectos adversos, como leucocitopenia, trombocitopenia, fiebre, hipocalcemia, confusión, alucinaciones y arritmias ventriculares, son menos frecuentes (15). La administración intramuscular con frecuencia produce dolor, inflamación y abscesos estériles (31). Se reserva para casos refractarios o cuando no hay otros tratamientos disponibles debido a que su elevada toxicidad, y la aparición de resistencias han provocado que sea relegada a un segundo plano, principalmente para tratamientos combinados. (27, 35).

### *Miltefosina*

La miltefosina es el único tratamiento de administración oral frente a *Leishmania*(36)se desarrolló inicialmente como antineoplásico oral, pero después se demostró que posee actividad antileishmaniásica(15). Produce frecuentemente efectos gastrointestinales como anorexia, náuseas, vómitos y diarrea. La mayoría de los episodios son breves y se resuelven a pesar de la continuación del tratamiento. Ocasionalmente los efectos colaterales pueden ser graves.

Debe administrarse después de las comidas, y en caso de necesitar dosis múltiples deben dividirse. Es potencialmente teratógeno por lo que no debe utilizarse en embarazadas ni en mujeres en edad de procrear en las que no se pueda garantizar una anticoncepción adecuada durante el tratamiento y los 3 meses siguientes(15).

### *Derivados azólicos: ketoconazol, fluconazol, itraconazol*

Son antifúngicos orales que poseen una eficacia variable en el tratamiento de la leishmaniasis(15).

### **1.13 Acceso a los medicamentos y a los productos diagnósticos:**

El acceso a los medicamentos para el tratamiento de las leishmaniasis visceral, cutánea y mucocutánea es un problema en los países pobres con el mayor número de casos; a pesar de los muchos esfuerzos que la OMS, las organizaciones médicas no gubernamentales y los fabricantes han hecho por mejorar el acceso a los

antileishmaniásicos. En el contexto de los sistemas de salud débil, existente en muchos países endémicos, hay varios factores que contribuyen a la falta de acceso a los medicamentos:

- Son relativamente caros y a menudo inasequibles, pese a las reducciones de precios negociadas por la OMS.
- Los protocolos terapéuticos nacionales no suelen reflejar los últimos avances, y son pocos o ninguno los medicamentos para la leishmaniasis que están incluidos en las listas nacionales de medicamentos esenciales.
- Los medicamentos para los cuales se han negociado precios preferenciales en los países de ingresos bajos y medios no siempre están registrados por las empresas allí donde resultan necesarios, dado que no hay un mercado lucrativo, y cuando no están registrados se necesita un permiso especial para importarlos. Dichos medicamentos tampoco suelen estar registrados en países con muy pocos casos, en los que los médicos tienen grandes dificultades para obtener las pequeñas cantidades de medicamentos necesarias.
- El suministro de medicamentos no es continuo. La mayoría de los antileishmaniásicos son producidos por un solo fabricante. Los problemas relacionados con la calidad, la baja capacidad de producción y la ausencia de provisiones adecuadas de las necesidades (todo lo cual redundará en largos plazos de entrega de los encargos) producen regularmente agotamiento de las reservas en los países endémicos.
- No hay reservas de amortiguación centrales a las que acceder en esos casos, y no hay ninguna plataforma que indique las necesidades de medicamentos; por consiguiente, no se pueden calcular las cantidades necesarias a nivel mundial, y la producción de medicamentos no se puede planificar adecuadamente.

En general, se debería reforzarla reglamentación de las políticas farmacéuticas, la garantía de la calidad y la monitorización del acceso a los medicamentos para la leishmaniasis, teniendo en cuenta sus precios, registro y necesidades mundiales. El acceso no controlado a los medicamentos (por ejemplo, la disponibilidad de la miltefosina sin receta médica) puede ser causa de uso indebido, tratamiento subóptimo y, a largo plazo, fármaco-resistencia. El uso de medicamentos falsificados (lotes tóxicos de antimoniales, miltefosina falsificada) ha producido en el pasado

varias muertes evitables .El suministro de productos diagnósticos para la leishmaniasis visceral presenta problemas similares, dado que el mercado es relativamente pequeño y no lucrativo. Además, en las regiones endémicas circulan pruebas diagnósticas de calidad no certificada (15).

#### **1.14 Estudios *in silico***

El enfoque de diseño/ descubrimiento de fármacos asistido por computadoras ofrece una alternativa más económica y rápida en el campo de la síntesis y evaluación de compuestos con determinada actividad farmacológica(37).Estos métodos llamados “*in silico*” engloban todas las técnicas asistidas por ordenadores usadas en el diseño/ descubrimiento y optimización de compuestos con propiedades deseadas y ha jugado un rol fundamental en el desarrollo de un número de fármacos que actualmente se encuentran en el mercado. Los procedimientos “*in silico*” se realizan solamente después de la exploración de los conceptos iniciales con modelos QSAR/ QSPR (38).

#### **Estudios QSAR**

Los estudios QSPR/QSAR, de sus siglas en inglés, *QuantitativeStructure-PropertyRelationships/QuantitativeStructure-ActivityRelationships* o en español, Relaciones Cuantitativas Estructura-Propiedad/Relaciones Cuantitativas Estructura-Actividad, constituyen enfoques cuantitativos orientados a encontrar relaciones entre la estructura molecular y las propiedades/actividades moleculares medidas o calculadas (39). Los estudios QSPR/QSAR datan desde el 1868 y desde entonces se han aplicado a la modelación de disímiles propiedades moleculares de naturaleza física, química y biológica (40-42).

#### *Estudios de Relación Cuantitativa Estructura-Actividad (QSAR)*

En el área de estudio de nuevos fármacos de la industria farmacéutica, institutos de investigación privados y públicos se han desarrollado herramientas computacionales (*in silico drug design*) y se han establecido nuevas disciplinas entorno de la Química Farmacéutica. Una de ellas es la metodología que relaciona numéricamente estructuras químicas con sus actividades biológicas, denominada QSAR (acrónimo de *QuantitativeStructure-ActivityRelationships*).

Los métodos para correlacionar cuantitativamente características estructurales con actividad biológica fueron introducidos por Hansch y Fujita en la década de 1960, con el transcurso del tiempo los métodos QSAR han emergido no sólo como una vía para el descubrimiento de compuestos con una actividad deseada, sino además como un método con aplicaciones al estudio del mecanismo de acción de los mismos.

Un estudio QSAR reúne un conjunto de técnicas computacionales relacionadas con diseño y visualización espacial virtual de moléculas, cálculo de propiedades fisicoquímicas moleculares (descriptores), bioinformática y estadística. Todo esto con el fin de hacer una predicción teórica de la actividad biológica que permita el diseño teórico de posibles nuevos fármacos, evitando pasar por el proceso a prueba y error de síntesis orgánica. Al ser una ciencia que existe sólo en un entorno virtual, desmaterializado de necesidades de infraestructura (tales como equipo, instrumentos, materiales y personal de laboratorio), con enfoque en las relaciones estructura (química) - actividad (biológica), el diseño de candidatos a nuevos fármacos es mucho más económico y rápido. En estos métodos, se utilizan múltiples descriptores moleculares (DMs) de la estructura química combinados con la aplicación de técnicas quimiométricas para derivar modelos, por ello se requiere disponer de datos biológicos de calidad, definir descriptores químicos relevantes y elegir un modelo adecuado.

Dado que los bioensayos son más costosos, es común que los experimentos no se repitan de forma debida para asegurar un valor estadísticamente validado, por su parte al ser un método estadístico, el estudio QSAR incluye este componente de incertidumbre de la variable dependiente por lo que la predicción puede fallar. En contraste, los valores de las variables independientes o descriptores, no tienen dicha naturaleza, ya que cada valor es reproducible e invariable pues descansa en un algoritmo de cálculo definido en el software. Eso no quiere decir que los descriptores no fallan, al contrario, hay que discriminar entre descriptores inadecuados (mala selección en ciertos casos), redundantes (covarianza o colinealidad) u otros mal programados (mal diseño del algoritmo).

Los pasos comunes de un estudio QSAR constan de:

1. Formulación del problema, en la cual se determina el objeto de análisis y el nivel de información requerido.
2. Parametrización cuantitativa de la estructura molecular de los compuestos químicos orgánicos/secuencia de biopolímeros
3. Medición de la propiedad de interés (efectos biológicos u otros)
4. Elección del tipo de modelo QSAR que se va a desarrollar
5. Selección de los compuestos (diseño estadístico de las series)
6. Análisis matemático de los datos y validación (interna y externa) de los modelos obtenidos
7. Interpretación de los resultados y aplicación de los modelos desarrollados al diseño/descubrimiento de un nuevo compuesto líder, desarrollando procedimientos de cribado virtual.

Sin embargo, el desarrollo de cualquier estudio QSAR, es un ciclo iterativo.

Las ventajas de QSAR son el bajo costo (programas gratuitos, información relativamente accesible, no se utiliza instrumental ni reactivos químicos, etc.), el uso de interfaces que facilitan el manejo y diseño, además de que la construcción de las moléculas y el cálculo de descriptores pueden ser sumamente rápidos. Sus desventajas son la familiarización con metodologías computacionales (diferentes sistemas operativos e interfaces gráficas, manejo de bases de datos, desarrollo del software) y en este sentido, la resolución de diferentes problemas de computo (compatibilidad, actualizaciones, registros, formatos de datos) así como el hecho de tener que disponer de datos de actividad biológica de las moléculas en estudio (43).

### Descriptores o Índices Moleculares

Los Descriptores o Índices Moleculares (DMs o IMs) definidos como el resultado final de un procedimiento lógico-matemático el cual transforma la información química, codificada dentro de una representación simbólica de una molécula, en un número (o conjunto de estos) útil (útiles) o en el resultado de alguno de los experimentos

ya estandarizados. Juegan un papel fundamental en los estudios *in silico* pues pueden brindar una visión más amplia en la interpretación de las propiedades moleculares y/o son capaces de tomar parte en un modelo para la predicción de notables propiedades moleculares.

### Algunos softwares empleados en los estudios *in silico*

#### WEKA

WEKA (*Waikato Environment for Knowledge Analysis*, o en español «Entorno para análisis del conocimiento, de la Universidad de Waikato») (figura 17) es una plataforma de software libre para el aprendizaje automático y la minería de datos escrito en Java. El paquete WEKA contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. WEKA soporta varias tareas estándar de minería de datos, especialmente, pre-procesamiento de datos, clasificación, regresión, visualización, y selección. Todas las técnicas de WEKA se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos). No puede realizar minería de datos multirelacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con WEKA(26).Figura 13



**Figura 13.** Software WEKA

## ChemAxon

ChemAxon es una plataforma integrada de programas y herramientas para tareas quimioinformática fundamentalmente. Incluye un conjunto de componentes especializados en la búsqueda en bases de datos químicos, la edición y visualización de las estructuras químicas, la predicción de las propiedades físico-químicas de moléculas y otras herramientas para la manipulación de estructuras químicas.

Varios de estos componentes pueden ser accedidos a través de la Interfaz gráfica de usuarios, la Interfaz para la programación de aplicaciones (API, del inglés *Programming Interface*) o la línea de comandos. Específicamente, contiene el módulo 'cxcalc', utilizado en este trabajo para el cálculo de descriptores como la carga de la molécula, la constante de ionización, clogP, entre otros (44).

## JChem

JChem para Office integra el manejo de estructuras y capacidades de visualización de dentro de un entorno de Microsoft Office. Incluye JChem para Excel que es una integración con el Excel. Con JChem para Office, la química se realiza en vivo en su aplicación de Office y las estructuras son totalmente compatibles dentro de hojas de cálculo, documentos, diapositivas y correos electrónicos. Su implementación es robusta con carga rápida / desplazamiento de miles de filas de estructura y copiar y pegar en toda la suite Microsoft Office. Su funcionalidad más avanzada incluye: cálculos basados en estructuras, la descomposición de grupos R y enumeración de bibliotecas (44, 45).

Las principales características de JChem para Excel son:

- Estructuras de importación y los datos de las bases de datos y archivos.
- Exportar datos a archivos.
- Añadir, estructuras de edición.
- Ordena, estructuras de filtro y los datos asociados.
- Descomposición R- grupos.
- Funciones de Excel personalizada químicos para el cálculo de las propiedades de la estructura (44, 45).

## Técnicas de inteligencia artificial

### KNN

Significa *K-Nearest-Neighbor* por sus siglas en inglés, también llamado algoritmo del vecino más cercano. Este método sirve para clasificar nuestra información y obtener predicciones. Básicamente se basa en tomar la información que deseamos clasificar, calcular las distancias de todos los miembros de la base de datos (comúnmente un conjunto de entrenamiento) y se toma un número particular de vecinos (K) que son los que tuvieron distancias menores, luego se visualiza la clase de los mismos, de allí se determina que tipo es el dato que estamos buscando. Con este tipo de algoritmos es necesario conocer bien la información y como parte medular decidir el tipo de cálculo para las distancias y el número de vecinos a considerar. Otro punto importante es el factor de búsqueda, también es común establecer un rango de búsqueda.

Entre sus características principales se encuentran:

- No requiere de la construcción de un modelo.
- Trabaja con los datos originales (no requiere representaciones especiales de los mismos).
- Realiza su predicción en base a información local (esto cuando se aplican heurísticas de búsqueda).
- Puede producir resultados equivocados si el cálculo de distancia no es el adecuado.
- En general, mientras K sea mayor (sin ser esto una regla), es decir, mientras más vecinos sean tomados en cuenta, la probabilidad de una predicción correcta aumenta (46).

### Redes neuronales (RN)

Estructura inspirada en un modelo simplificado de las neuronas biológicas. Se forma de un conjunto de elementos sencillos (neuronas) que tiene varias entradas y una salida. Estos elementos se interconectan entre sí para formar redes (red neuronal). Las RN se entrenan para aprender relaciones de entrada-salida mediante la presentación de ejemplos, modificando los pesos.

Una vez entrenada, la red neuronal se puede utilizar para diversas tareas:

- Clasificación
- Clasificación no-supervisada
- Asociación
- Complementar patrones (47)

"Una red neuronal artificial (RNA) es un conjunto de algoritmos matemáticos que procesan información y encuentran relaciones no lineales entre el conjunto de datos, y cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona" (48). Las RNA tratan de emular ciertas características propias de los cerebros humanos, como la capacidad de memorizar y de asociar hechos por tanto son modelos artificiales y simplificados del cerebro humano, un sistema que es capaz de adquirir conocimiento a través de la experiencia (48)

Perceptrón multicapa: posibilita la solución de problemas complejos. A diferencia de las anteriores arquitecturas de RNA, la estructura de red multicapa (*MultiLayerPerceptron-MLP*), posee al menos tres niveles de neuronas: una capa de entrada, una capa oculta y una capa de salida.

#### Máquinas de soporte vectorial (SVM)

Las máquinas de soporte vectorial (SVM, por sus siglas en inglés) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vapnik y Cortés (1995) y su equipo AT&T, que han surgido como métodos relacionados con problemas de clasificación y regresión. Su buen desempeño ha llevado a su uso en una gran variedad de problemas, algunos investigadores (Fletcher, Hussain y Shawe-taylor, 2010; Huang, Nakamori y Wang, 2005; Kara et al., 2011a; Kim, 2003b; Masoud, 2014; Moein Aldin, Dehghan Dehnavi y Entezari, 2012; Net- y Pino, 2012; Velásquez, Olaya y Franco, 2010; Yuan, 2011), han utilizado SVM para solucionar problemas de clasificación y regresión relacionados a la predicción de series de tiempo, mostrando ser mejores en comparación a otras metodologías tradicionales como modelos econométricos y en algunos casos a modelos de aprendizaje de máquina (44).

Las SVM se empezaron a emplear para resolver problemas de clasificación y reconocimiento de patrones para luego extenderse en el estudio de predicción de series de tiempo. Los problemas de clasificación se emplean para obtener resultados

de tipo cualitativo, por ejemplo, determinar la clase de un dato de entrada o características, mientras que las de tipo regresión son más útiles en problemas cuantitativos, cuando se trata de obtener una salida numérica al dato de entrada (Camps y Bruzzone, 2009; Fernando y Gutiérrez, 2011; Lember, 2012). Un punto a favor de utilizar este tipo de modelos es que el desempeño de las SVM no depende del tamaño de la muestra que se va utilizar para el problema, por lo que puede ser utilizado para una cantidad limitada de datos en contraste con otras metodologías que presentan mejor desempeño cuando el tamaño de la muestra es grande. Asimismo, el algoritmo detrás de las SVM se puede ajustar a problemas no lineales y la solución se realiza bajo programación cuadrática, lo cual hace que su solución sea única y generalizable (Abe, 2005; Cristianini, 2000; Velásquez et al., 2010).

La idea detrás de las SVM es que a partir de unos inputs de entrada al modelo, se etiquetan las clases y se entrena una SVM construyendo un modelo que sea capaz de predecir la clase de los nuevos datos que se introduzcan al modelo. La SVM representa en un eje de coordenadas los vectores de entrenamiento, separando las clases por un espacio lo más grande posible. Cuando nuevos datos son introducidos al modelo, estos se colocan sobre el mismo eje y en función de la cercanía de los grupos antes separados, los cuáles serán clasificados en una u otra clase (44).

### Árboles de clasificación

Los métodos basados en árboles (o árboles de decisión) son bastante populares en *data mining*, pudiéndose usar para clasificación y regresión. Estos métodos se derivan de una metodología previa denominada *automatic interaction detection*. Son útiles para la exploración inicial de datos y apropiados cuando hay un número elevado de datos, y existe incertidumbre sobre la manera en que las variables explicativas deberían introducirse en el modelo. Sin embargo, no constituyen una herramienta demasiado precisa de análisis.

En conjuntos pequeños de datos es poco probable que revelen la estructura de ellos, de modo que su mejor aplicación se encuentra en grandes masas de datos donde pueden revelar formas complejas en la estructura que no se pueden detectar con los métodos convencionales de regresión.

Problemas donde los árboles de clasificación se pueden usar:

1. Regresión con una variable dependiente continua.
2. Regresión binaria.
3. Problemas de clasificación con categorías múltiples ordinales.
4. Problemas de clasificación con categorías múltiples nominales.

#### Ventajas de los árboles de clasificación

1. Los resultados son invariantes por una transformación de monótona de las variables explicativas.
2. La metodología se adapta fácilmente en situaciones donde aparecen datos *missing*, sin necesidad de eliminar la observación completa.
3. Están adaptados para recoger el comportamiento no aditivo, de manera que las interacciones se incluyen de manera automática.
4. Incluye modelos de regresión así como modelos de clasificación generales que se pueden aplicar de manera inmediata para diagnosis (44).

Dentro del aprendizaje automático el tema a tratar es árboles de clasificación o también conocidos como árboles de decisión. Los factores que han influido en su difusión son: accesibilidad a diferentes implementaciones, la explicación que aporta a la clasificación, la posibilidad de ser representados gráficamente, y la rapidez de clasificar nuevos patrones. Los árboles de clasificación están dentro de los métodos de clasificación supervisada, teniendo una variable dependiente o clase. La construcción del árbol de clasificación se realiza mediante un proceso de inducción. Los árboles de clasificación se basan en una estructura en forma de árbol, donde las ramas representan conjuntos de decisiones, las cuales generan reglas para la clasificación de un conjunto de datos en subgrupos de datos. Las ramificaciones se generan de forma recursiva hasta que se cumplan ciertos criterios de parada (49).



*Materiales*

*y Métodos*



## MATERIALES Y MÉTODOS

En el presente capítulo se describe todo el proceso realizado como parte de la investigación hasta identificar los nuevos compuestos con posible actividad contra promastigotes de *Leishmania major*. Se parte de la construcción minuciosa de una base de datos, la representación química de los compuestos y el cálculo de descriptores moleculares para llegar a la conformación de modelos capaces de identificar estos compuestos a través del cribado virtual de una base de datos internacional y compuestos de síntesis(50).

### 2.1 Gestión de la data.

Todos los compuestos incluidos en la investigación, al igual que su perfil biológico, fueron recopilados de *BioAssay* publicados en el PubChem(51), los cuales consisten en estudios realizados en específico para promastigotes de *Leishmania major* y reportados por varias revistas de alto impacto en la *Web of Sciences*(52-58), todas publicadas por la editorial ELSEVIER en el período (1941-2018). Fueron recopilados inicialmente un total de 341 compuestos, probados contra la forma promastigote del parásito de *Leishmania major*. Además, se extrajeron los datos relacionados con la forma del parásito en que se realizó cada experimento, el valor reportado de  $IC_{50}$ , las características del procedimiento llevado a cabo en el ensayo, así como la masa molar y la estructura química a partir de los SMILE canónicos. El parámetro empleado para clasificar a los compuestos en activos e inactivos fue la  $IC_{50}$ ; considerándose como activos aquellos que tras su evaluación presentaban un valor de  $IC_{50}$  menor a 1,5  $\mu$ M, e inactivos los que no cumplían dicho requisito.

### 2.2 Estrategias Computacionales.

Como regla general un modelo QSAR intenta establecer una correlación matemática entre la estructura de un grupo de compuestos químicos y su actividad biológica, generalmente por medio del uso de un análisis de regresión o discriminante con el empleo de DMs. La figura 14 muestra la estrategia general para el desarrollo de estudios QSAR.



**Figura 14.** Esquema general para la obtención de un modelo QSAR.

Todos los cálculos se efectuaron en un ordenador (PC) Intel(R) Core™ i5-2500 CPU @ 3.30GHz. El Programa DRAGON 5.5 Professional (59) se usó para el cálculo de los DMs 0D-2D (un conjunto de más de 2000 DMs)(Figura 15). Los datos se almacenaron en matrices, utilizando el programa Microsoft Excel; donde las filas representaron a los compuestos químicos (casos) y las columnas a los DMs (variables). El análisis estadístico se efectuó con el paquete de procesamiento WEKA 3.6 del 2009 (60) para Windows.



**Figura 15.** Software DRAGON versión 5.5

### 2.3 Parametrización de la estructura química.

La mayoría de las estructuras incluidas en la base de datos (341 en total) fueron descargadas del PubChem en formato SDF 2D Structure y el resto fueron representadas usando el módulo ChemDraw Ultra 8.0 del software ChemOffice2005 ver. 9.0 (61), estas últimas fueron seguidamente exportadas como fichero con extensión sdf del tipo MDL. Antes de proceder al cálculo de las variables las estructuras fueron curadas y estandarizadas utilizando el OpenBabel ver. 2.3 para añadir hidrógenos explícitos, neutralizar cargas y remover sales que pudiesen afectar el cálculo. La eliminación de compuestos repetidos se llevó a cabo con el programa

ISIDA (62) y la forma Edi-SDF versión 5.03 para Windows (Figura 16). Los ficheros obtenidos en modo SDF se llevaron al programa DRAGON (63) donde se calcularon las diferentes familias de descriptores moleculares de 0D-2D que están implementadas en el mismo, obteniendo un conjunto de 2489 DMs, sin eliminar las variables constantes, cercanas a constantes y con una correlación mayor al 90%. Los datos se almacenaron en matrices del Excel.

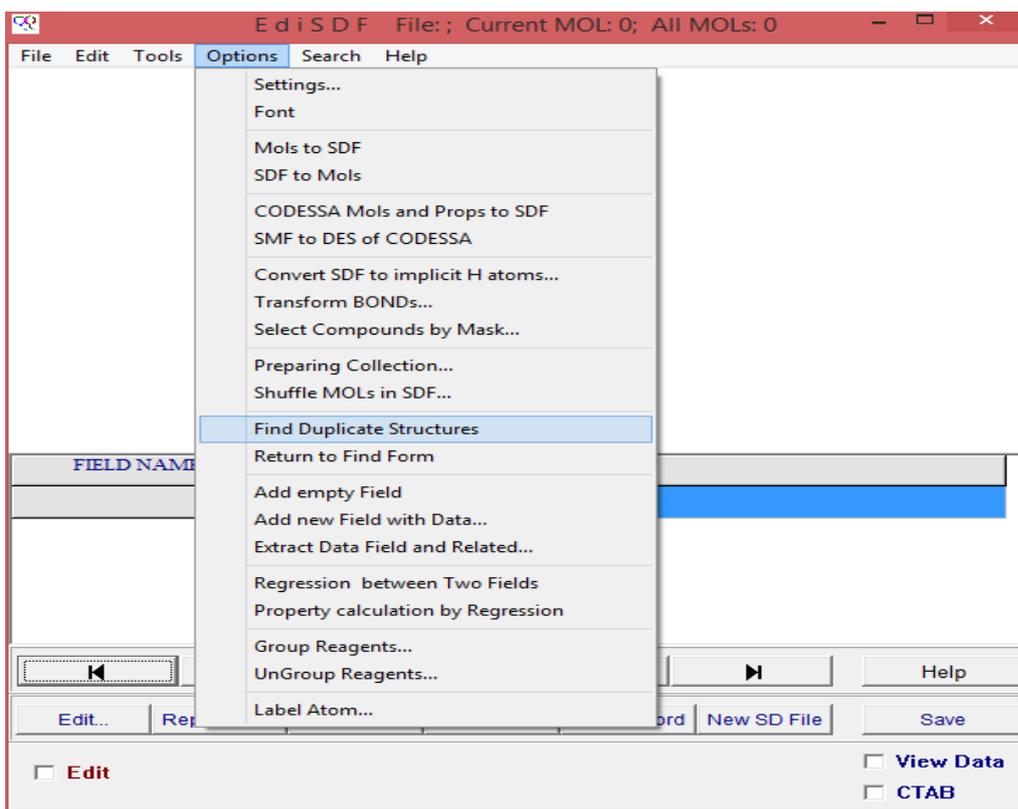


Figura 16. ISIDA forma Edi-SDF versión 5.03 para Windows

## 2.4 Análisis estadístico de los datos.

Para asegurar la diversidad estructural de los compuestos incluidos en la base de datos se realizó el Análisis de Conglomerados (AC) como vía de evaluación de la distribución y diversidad estructural existente dentro de los grupos de observaciones activas e inactivas respectivamente implementados en el paquete de procesamiento *STATISTICA 8.0*. En dicho estudio se procedió a estandarizar todas las matrices de DMs anteriormente calculadas. Con posterioridad, los conjuntos de compuestos activos e inactivos, respectivamente, se dividieron a su vez en diferentes subconjuntos mediante dos análisis de conglomerados del tipo *k*-MCA (siglas en inglés de Análisis de Conglomerados de *k*-Medias, también llamado análisis de

agrupaciones) (64-66), los que se efectuaron para cada grupo. De este modo quedó asegurada la representatividad de elementos del mismo dominio en los subconjuntos obtenidos en cada caso. De cada conglomerado se seleccionaron, de forma aleatoria, los compuestos para la conformación de la serie de entrenamiento, de predicción y la validación externa las cuales constituyen un 65 %, 20 % y 15 % de la data inicial, respectivamente. Este procedimiento asegura que las clases químicas (determinadas por los conglomerados derivados de *k*-MCA) estén representadas en ambas series de compuestos (64-66). Los compuestos incluidos en la SP nunca fueron usados en el desarrollo de los modelos cuantitativos.

Para la selección de atributos se utilizó el WEKA 3.6 (60) del 2009, empleando varios evaluadores de atributos y métodos de búsqueda específicos para cada técnica de clasificación. Los mejores atributos seleccionados por cada método se introdujeron en los ficheros finales que se emplearían en el WEKA para la clasificación. En esta se emplearon las técnicas: *k vecino más cercano* (IBK), árbol de clasificación (J48), red neuronal artificial (MLP por sus siglas en inglés *Multi Layer Perceptron*) y máquina de soporte vectorial (SMO por sus siglas en inglés *Sequential Minimal Optimization*) (67) para obtener los mejores modelos capaces de predecir la actividad de los compuestos incluidos en la serie de entrenamiento y predicción.

## 2.5 Obtención y validación de los modelos

La obtención de los modelos finales se llevó a cabo empleando los atributos seleccionados teniendo en cuenta la ley de parsimonia (*lex parsimoniae*) ó Rasurador de Ockham: “*Numquam ponenda est pluralitas sine necessitate*” o “La pluralidad nunca se debe postular sin necesidad (de ello)”, a la hora de seleccionar el mejor modelo por cada técnica, este debía poseer la significación estadística más alta con la menor cantidad de variables posible pues cuando las explicaciones múltiples están disponibles para un fenómeno, se prefiere la versión más simple) (68, 69).

Se calcularon los parámetros estadísticos más utilizados en la literatura médico-estadística: la exactitud(Q), el coeficiente de correlación de Matthews(C), la sensibilidad(Sens), la especificidad(Espec) y la razón de falsos positivos (*FPr*)

$$Q = 100 * (VP + VN) / (VP + FP + VN + FN)$$

$$FP_r = 100 * FP / (FP + VN)$$

$$\text{Sensibilidad (Sens)} = 100 * VP / (VP + FN)$$

$$\text{Especificidad (Espec)} = 100 * VP / (VP + FP)$$

$$C = 100 * (VP * VN - FP * FN) / \sqrt{(VN + FN) * (VN + FP) * (VP + FP) * (VP + FN)}$$

Donde VP y VN son los verdaderos positivos y negativos, respectivamente y FP y FN son los falsos positivos y negativos, correspondientemente.

Para demostrar el poder predictivo de los modelos se desarrollaron procedimientos de validación externa, empleándose la serie de predicción obtenida mediante el análisis de conglomerados descrito en el epígrafe anterior (para compuestos evaluados en promastigotes). En cada caso se determinaron los parámetros estadísticos antes mencionados los cuales nos permitieron promover la aceptación de estos modelos QSAR a favor de los órganos regulatorios (70). También fue empleada una serie de validación externa que incluyó un grupo de 48 compuestos (20 activos y 28 inactivos) seleccionados al azar a partir de nuestra base de datos primaria de la cual constituyen el 15%, los cuales nunca fueron empleados en el desarrollo de los modelos y cumplían los requerimientos de nuestro trabajo, por lo que permitirían validar el desempeño de los modelos obtenidos para promastigotes de *L. major*.

Otra forma de evaluar el desempeño de un clasificador se basa en las curvas ROC (acrónimo de *Receiver Operating Characteristic*). Una curva ROC es un gráfico con la Razón de Falsos Positivos ( $1 - \text{Spec}$ ) en el eje X y la Razón de Verdaderos Positivos ( $T_{\text{razón}} = \text{Sens}$ ) en el eje Y mediante la variación del umbral de decisión. Se utilizó el área bajo la curva ROC (AURC, por sus siglas en inglés) como otro indicador de la calidad de los modelos. Mientras más cercano a uno es el valor del área, el comportamiento del clasificador estaría más cercano al clasificador perfecto.

## 2.6 Cribado virtual

DrugBank es una base de datos disponible en la Universidad de Alberta que engloba en una fuente única información bioinformática y quimioinformática además de proporcionar datos sobre fármacos (71). Se ha utilizado ampliamente con el objetivo de facilitar el descubrimiento *in silico* y el diseño de compuestos, predicción del

metabolismo de los fármacos y sus interacciones, así como para la educación farmacéutica en general. En su versión de 2008 contiene cerca de 4800 entradas, incluyendo más de 1480 pequeñas moléculas y 128 fármacos bióticos, todos aprobados por la FDA (*Food and Drug Administration* por sus siglas en inglés), más de 3.200 fármacos experimentales y adicionalmente cerca de 2.500 secuencias de proteínas (50). Cada tarjeta de fármaco (*DrugCard*) contiene más de 100 campos de datos, la mitad de ellos en relación con la estructura química y la otra con la diana terapéutica. Los usuarios pueden realizar consultas de distintas formas. Está amparada por David Wishart y Craig Knox (72). Esta base de datos internacional fue cribada por los modelos obtenidos para identificar compuestos potencialmente activos contra promastigotes de *L. major*.

En este trabajo fueron empleados además para el cribado 512 compuestos, la mayoría de ellos de nueva síntesis obtenidos en laboratorios de Química Orgánica de las Universidades de Rostock [415], Alemania, el Conservatorio Nacional de Artes y Oficios (CNAM) de París, Francia [69], el Departamento de Bioquímica de la Facultad de Medicina en la Universidad de la República, Montevideo, Uruguay [8] y compuestos aislados a partir de la planta *Boldoa purpurascens Cav* [20]. en colaboración de la Universidad Central de Las Villas y el Departamento de Química de la Universidad Católica de Leuven, Bélgica.

#### *Sistema multclasificador ensamblado basado en modelos QSAR.*

El principio de parsimonia (*Occam's Razor*) fue tomado siempre en cuenta para la selección del modelo, teniendo presente que cuando se disponen de múltiples soluciones para un fenómeno, se prefiere la versión más simple (68). Aunque cada modelo capta por separado una gran fracción de la información química contenida en la base de datos a través de los DMs que usa, esta fracción de información necesita complementariedad del resto de los modelos para su completitud. Por ello se utilizó un clasificador ensamblado a partir de todos los modelos QSAR considerados. En el presente trabajo además de las técnicas individuales, se utilizó un sistema ensamblado o multclasificador con el fin de mejorar los resultados de los modelos obtenidos (73, 74) puesto que a partir de la combinación adecuada de varios clasificadores tienen una tendencia general a mejorar los resultados (75, 76). Generalmente, se empleó el voto mayoritario simple. En esta técnica cada clasificador

tiene un voto que tiene igual valor. En esencia, la clasificación más popular es la que resulta elegida por el conjunto como decisión final. La variante más básica es el voto de la pluralidad, donde la clase con la mayoría de los votos gana (77, 78).

Tras el cribado se procedió a la identificación, análisis y reporte de todas las estructuras identificadas como potencialmente activas contra promastigotes de *Leishmania major*, los cuales podrán ser empleados en futuras investigaciones relacionadas con el descubrimiento de nuevos compuestos para el tratamiento de las leishmaniasis causadas por esta especie.



*Resultados y  
discusión*



## RESULTADOS Y DISCUSIÓN

En este capítulo se presentan los resultados de la generación de los modelos a partir de la construcción de una nueva base de datos, el cálculo de DMs y la identificación de nuevos compuestos potencialmente activos contra promastigotes de *L. major* mediante el cribado virtual de una base de datos internacional y compuestos de síntesis.

### 3.1 Gestión de la data.

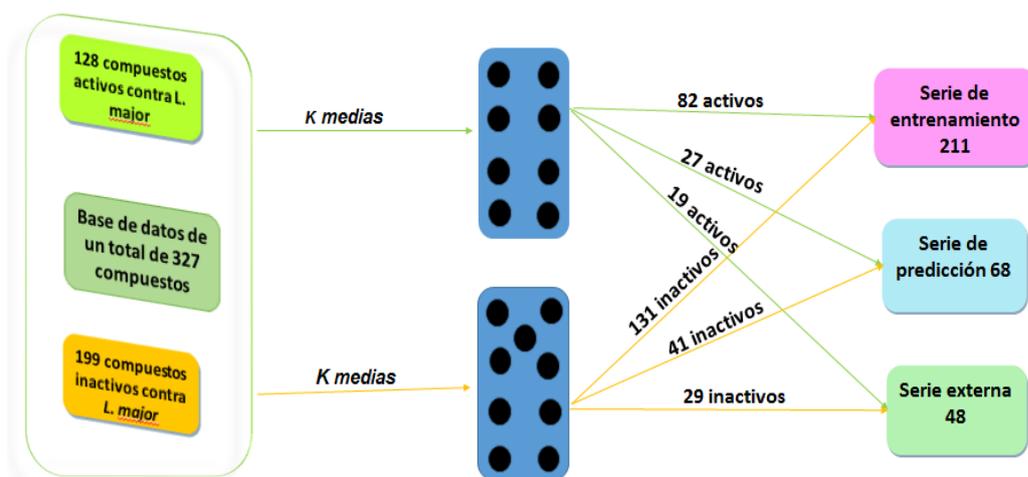
De un total de 341 compuestos obtenidos a partir de bioensayos del PubChem probados contra las dos formas del parásito, se seleccionaron 336 para construir la base de datos final de compuestos contra promastigotes de *L. major*. El hecho de que todos hayan sido probados experimentalmente a través de ensayos en los que se emplearon procedimientos muy similares posibilita la obtención de una mayor veracidad en los resultados evitando la futura clasificación “errónea” de compuestos. Atendiendo al parámetro ( $IC_{50}$ ) empleado para clasificar a los compuestos, la base de datos final incluyó 132 compuestos activos y 204 inactivos, teniendo en cuenta los diferentes modos de acción y diversos patrones estructurales. La data posee un alto grado de variabilidad estructural (Anexo 2) compuestas por familias estructurales fusionadas que incluye a derivados del naftaleno, furanos, piridina fosfatos, fenoles, cicloalcanos, posee grupos esteroidales, derivados del ácido salicílico, diamidinas aromáticas, arilquinuclidinas, pirimidinas sustituidas, sistemas heterocíclicos y cíclicos tanto saturados como insaturados entre otros.

*Parametrización de la estructura química. Estrategias computacionales. Análisis estadístico de los datos.*

Las estructuras incluidas en la base de datos (336 en total) luego de curadas y estandarizadas se llevaron al software DRAGON donde se calcularon las diferentes familias de descriptores moleculares de 0-2D obteniéndose 2489 variables, quedando al final 1151 luego de proceder a eliminar aquellas con un coeficiente de correlación mayor a 0.9. Algunos de los parámetros implementados en el software e incluidos para el cálculo fueron las propiedades moleculares, descriptores de búsqueda y análisis subestructural, conteo de fragmentos, huellas digitales de frecuencia e índices de conectividad, entre otros.

La calidad de los métodos QSAR dependen de forma crítica de la selección de los compuestos que son utilizados como serie de entrenamiento (SE) para construir los modelos de clasificación para lo cual el aspecto más importante es garantizar la adecuada diversidad estructural de la data; lo cual permite abarcar, en la medida de las posibilidades, la mayor amplitud del espacio experimental. Se realizó análisis de conglomerados (AC) implementado en el paquete de procesamiento STATISTICA 8.0, como vía de evaluación de la distribución y diversidad estructural existente dentro de los grupos de observaciones activas e inactivas respectivamente. Los compuestos activos e inactivos se dividieron a su vez en diferentes subconjuntos mediante dos análisis de conglomerados del tipo k-MCA (siglas en inglés de Análisis de Conglomerados de *k-Medias*, también llamado análisis de agrupaciones o análisis de conglomerados). El procedimiento empleado se refleja en la figura 9. Fueron eliminados de la base de datos nueve estructuras químicas muy diferentes a las demás y que quedaban aisladas de manera independiente formando clústeres de un solo compuesto, siendo identificados como *outliers*. De este modo quedó asegurada la representatividad de los 327 elementos restantes del mismo dominio en los subconjuntos obtenidos en cada caso. De cada conglomerado se seleccionaron, de forma aleatoria, los compuestos para la conformación de la serie de entrenamiento, de predicción y la validación externa quedando 211, 68 y 48 respectivamente, figura 9.

Tras la selección de atributos en WEKA con el empleo de varios evaluadores, filtros y métodos de búsqueda específicos para cada técnica de clasificación, a partir de 1151 descriptores del Dragón, se realizó una reducción de variables hasta obtener aquellas más adecuadas logrando alcanzar seis para IBK y J48, siete para MLP y ocho para SVM, estas se emplearon en la elaboración de los ficheros utilizados en el WEKA para la conformación de los cuatro modelos de clasificación finales.



**Figura 17.** Algoritmo general utilizado para diseñar serie entrenamiento, de predicción y la serie externa empleando k -MCA.

Los algoritmos y parámetros estadísticos correspondientes a los modelos obtenidos para la base de datos de promastigotes se describen a continuación:

### **Técnica de k vecinos más cercanos (IBk)**

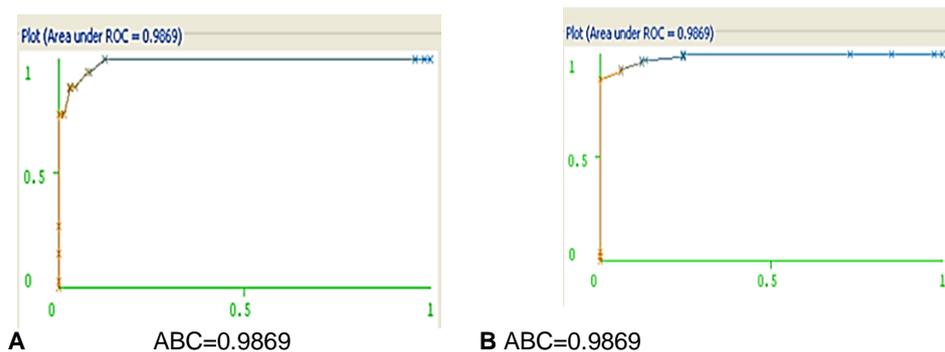
El modelo IBk obtenido utilizó el siguiente algoritmo de clasificación para WEKA: `[WEKA.classifiers.lazy.IBk-K3-W0-AWEKA.core.neighboursearch.LinearNNSearch-A"WEKA.core.EuclideanDistance-R first-last" -t IBk_training.arff -T IBk_Test.arff -o]` para 6 variables, que incluyeron: dos descriptores topológicos derivados de la matriz del eje de adyacencia que codifican la conectividad entre los ejes de grafos (*EEig01x* y *EEig01d*), dos descriptores que se basan en el conteo de grupos funcionales químicos e incluyen *anCconj* que describe el número de carbonos sp<sup>2</sup> conjugados no aromáticos y *nPyrazines* el cual corresponde al número de pirazinas, así como un descriptor de átomos basados en fragmentosO-057que describe el número de grupos hidroxilosénólicos, fenólicos y carboxílicos y B04[O-O] descriptor de huellas binarias 2D que mide la presencia o ausencia de enlaces O-O a la distancia de cuatro carbonos. Donde k (clasificador k-vecinos más cercanos) selecciona el valor apropiado de vecinos basado en la validación cruzada y W (*window Size* o en español tamaño de ventana) obtiene el número máximo de instancias permitidas en la serie de entrenamiento, un valor de 0 significa que no hay límite para el número de instancias de formación, la adición de nuevos casos por encima de este valor se traducirá en casos viejos siendo removidos. El modelo clasificó correctamente un 92.42%de las instancias para la serie de entrenamiento y un 88.24% para la serie de predicción. En la tabla I se observan los valores obtenidos con respecto a la sensibilidad,

especificidad, coeficiente de correlación de Matthews, relación de falsos positivos y exactitud para las SE y SP, resultan significativos los valores de exactitud y sensibilidad obtenidos para la SE (92.42% y 93.90%). En la figura 18 y 19 se presentan los valores de Área Bajo la Curva (ABC) para las SE y SP, puede observarse que en todos los casos su valor es muy próximo a uno, lo cual es indicativo del alto valor predictivo de este modelo.

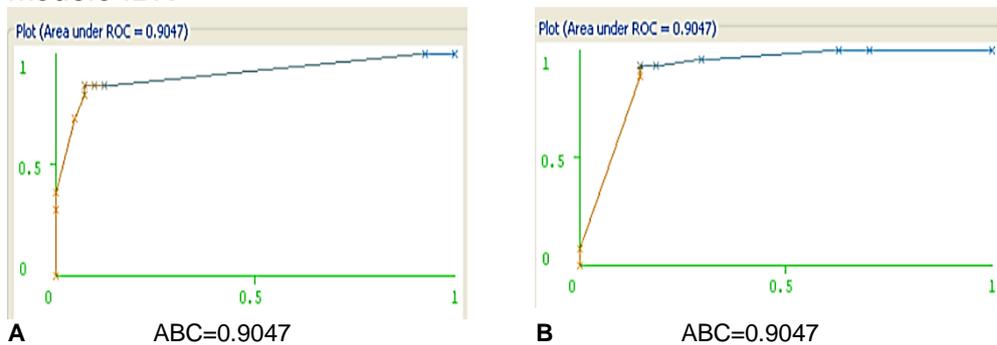
**Tabla I:** Parámetros estadísticos obtenidos en el modelo IBk (Clasificador k- vecino más cercano).

Serie	C	Q (%)	Sensibilidad(%)	Especificidad(%)	FPr (%)
Entrenamiento	0.84	92.42	93.90	87.50	8.53
Predicción	0.75	88.24	85.19	85.19	9.76

**Legenda:** Q exactitud, C coeficiente de correlación de Matthews, FPr Razón de falsos positivos.



**Figura 18.** Curvas ROC para compuestos activos(a) e inactivos(b) de la SE para el modelo IBK



**Figura 19.** Curvas ROC para compuestos activos(A) e inactivos(B) de la SP para el modelo IBK

En 2015 y 2017 Flores y colaboradores (79, 80) obtuvieron modelos con la técnica IBk para promastigotes de *Leishmania amazonensis* y amastigotes de *L. infantum*

respectivamente en los que emplearon el algoritmo de clasificación para Weka: [`weka.classifiers.lazy.IBk-K3-W0-weka.core.neighboursearch. LinearNNSearch-weka.core.ManhattanDistance-Rfirst-last`] -t `IBk_training.arff` -T `IBk_Test.arff` -o] para 7 y 6 variables, similar al empleado por nosotros a diferencia de la distancia que en nuestro caso fue la Euclideana. Los resultados de exactitud para SE (92.42%) y SP (88.24%) fueron superiores en nuestro modelo (*L. amazonensis* SE (89.53%), SP (83.33%) y *L. infantum* SE (88.01%), SP (85.05%). Los valores de falsos positivos de este trabajo SE (8.53%), SP (9.76%) son inferiores a los obtenidos por estos investigadores.

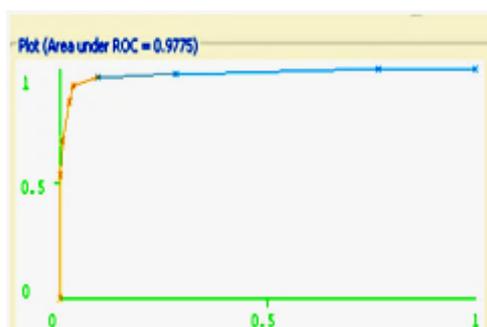
### **Técnica de árboles de clasificación (J48)**

El modelo de clasificación de árbol (J48) para promastigotes presentó 6 variables: las cuales coinciden con las empleadas para la obtención del modelo IBk comentado anteriormente e incluyen dos descriptores topológicos derivados de la matriz del eje de adyacencia (*EEig01x* y *EEig01d*), dos descriptores que se basan en el conteo de grupos funcionales químicos (*nCconj* y *nPyrazines*), un descriptor de átomos basados en fragmentos O-057 y un descriptor de huellas binarias 2D B04[O-O], así como el siguiente esquema algorítmico: [`WEKA.classifiers.trees.J48 -C 0.18 -M 2 -t J48_training.arff -T J48_Test.arff -o`] teniendo que C es 0.18 (*confidenceFactor*: factor de confianza utilizado para la poda) donde valores menores incurren en un incremento de la poda, M es dos (*minNumObj*: número mínimo de instancias por hoja) (Figura 22). Esta técnica mostró un 94.79% (SE) y 88.24% SP de compuestos correctamente clasificados. Resultan significativos los valores de sensibilidad (92.68%), especificidad (93.83%) y falsos positivos (3.88%) alcanzados en la SE así como de sensibilidad para la SP (88.00%), los cuales son muestra de la robustez de nuestro modelo. Tabla II. En la figura 20 y 21 se muestran los valores de Área Bajo la Curva (ABC) ROC, (ABC=0.9775).

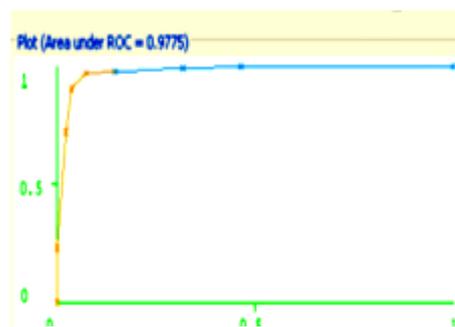
**Tabla II:** Parámetros estadísticos para el modelo J48 (Árbol de clasificación)

Serie	C	Q(%)	Sensibilidad(%)	Especificidad(%)	$FP_r$ (%)
Entrenamiento	0.89	94.79	92.68	93.83	3.88
Predicción	0.75	88.24	81.48	88.00	7.32

**Legenda:** Q exactitud, C coeficiente de correlación de Matthews; FPr Razón de falsos positivos.

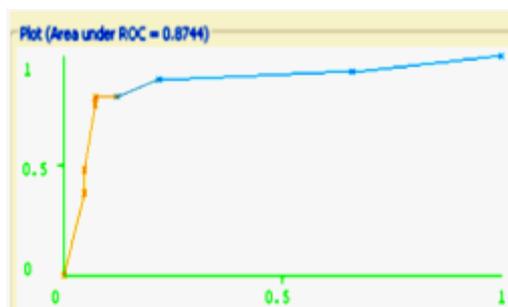


A ABC=0.9775

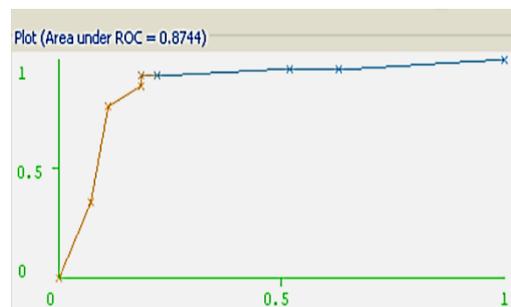


B ABC=0.9775

**Figura 20.** Curvas ROC para compuestos activos (A) e inactivos (B) de la SE para el modelo IBK.

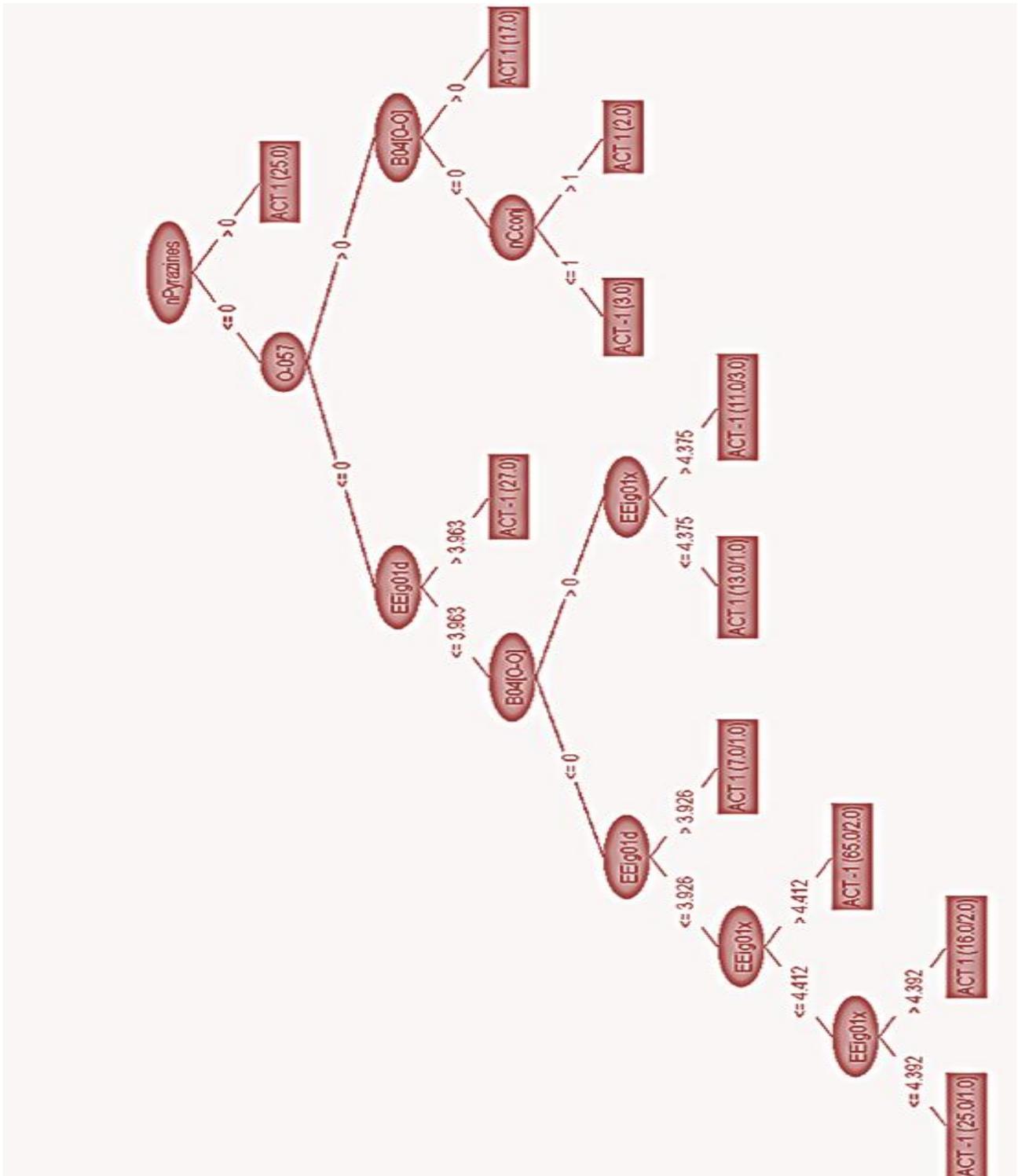


A ABC=0.8744



B ABC=0.8744

Figura 21. Curvas ROC para compuestos activos (A) e inactivos (B) de la SP para el modelo IBK



En reportes anteriores consultados en la literatura (79, 80) los algoritmos utilizados en esta técnica para la obtención de modelos en amastigotes de *L. infantum* y promastigotes de *L. amazonensis* emplean como factor de confianza C 0.09 y 0.24 y

como número mínimo de instancias M 5 y 3; la técnica utilizada para *L. major* tuvo valores de C=0.18 y M=2, incluyendo 7 y 8 variables respectivamente en nuestro caso se emplearon solo 6 variables. La exactitud obtenida en nuestro modelo es superior así como son inferiores los valores de falsos positivos.

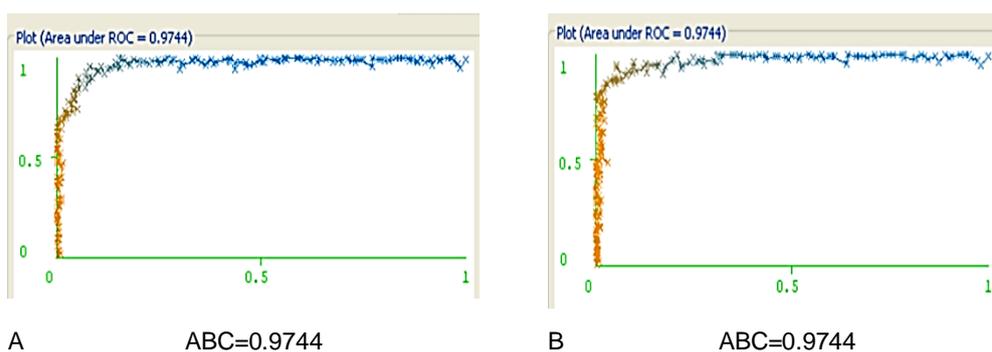
### **Técnica de red neuronal MLP**

La red neuronal artificial MLP (*Multi-Layer Perceptron* por sus siglas en inglés o perceptrón multicapas) utilizó 7 variables: *MAXDN* un descriptor topológico que mide la máxima variación negativa electrotopológica, *X4Av* el índice de conectividad de valencia promedio *chi4*, *SIC2* que calcula el contenido de información estructural de acuerdo a la simetría de vecindario de orden 2, *BELm3* que evalúa el valor propio más bajo n. 3 de matriz de Barden / ponderado por masas atómicas, un descriptor nR=Cs que realiza el conteo de grupos funcionales como carbonos secundarios alifáticos(sp<sup>2</sup>), *nPyrazines* el cual corresponde al número de pirazinas, B04[N-N] descriptor de huellas binarias 2D que mide la presencia o ausencia de enlaces N-N a la distancia de cuatro carbonos y su esquema algorítmico fue: `[WEKA.classifiers.functions.MultilayerPerceptron-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 7 -t MLP_training.arff -T MLP_Test.arff -o]` siendo L (*earning Rateo intervalo de incremento*) que constituye la suma actualizada de los pesos, H (*hidden Layers* o capas ocultas), define las capas ocultas de la red neuronal. Los valores de ABC ROC para los compuestos activos e inactivos de la SE y SP se muestran en las figuras 23 y 24 resultando superiores para la SE. Los parámetros estadísticos que más resaltan son los altos valores de exactitud y especificidad para la SE y SP así como los bajos valores de falsos positivos. Tabla III.

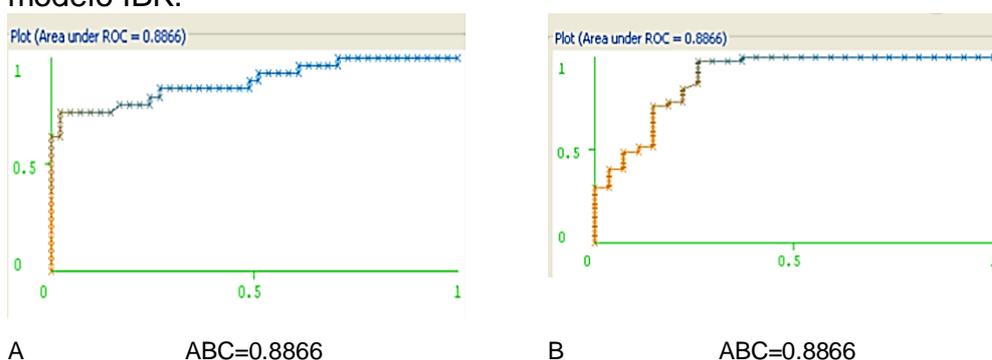
**Tabla III.** Parámetros estadísticos obtenidos para modelo MLP (Perceptrón Multicapas).

Serie	C	Q(%)	Sensibilidad (%)	Especificidad(%)	FPr(%)
Entrenamiento	0.82	91.47	86.59	91.03	5.43
Predicción	0.72	86.76	74.07	90.91	4.88

**Leyenda:** Q exactitud, C coeficiente de correlación de Matthews; FPr Razón de falsos positivos.



**Figura 23.**Curvas ROC para compuestos activos (A) e inactivos (A) de la SE para el modelo IBK.



**Figura 24.**Curvas ROC para compuestos activos (A) e inactivos (B) de la SP para el modelo IBK.

Al comparar con modelos similares encontrados en la literatura pudimos observar que en el esquema algorítmico utilizado para la técnica de red neuronal MLP obtenida para promastigotes de *L. amazonensis* y amastigotes de *L. infantum* se empleó un intervalo de incremento de 0.9 y 1, con valores de *M* iguales a 0.4 y 0.8 respectivamente y un número de capas ocultas de 20 y 13.(79, 80) Los valores de exactitud son similares en todos los modelos mientras que la especificidad en *Leishmania amazonensis* para la SE resultó ser superior a las obtenidas en el resto de los modelos.

### ***Técnica de máquina vectorial de soporte***

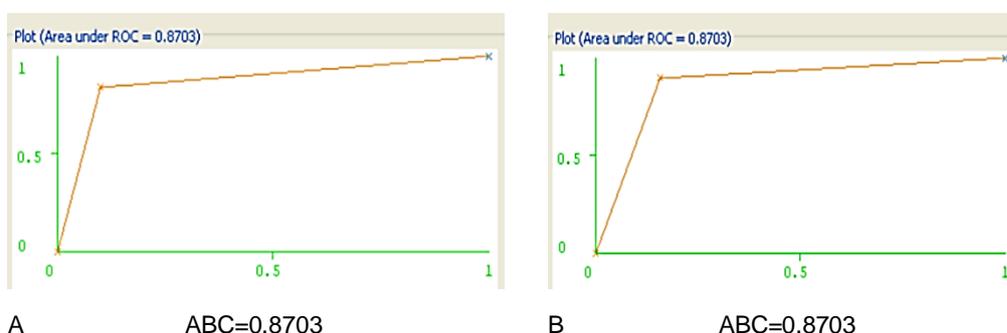
El modelo final para la técnica máquina vectorial de soporte (SMO) estuvo conformado por 8 variables: T(N..N) que representa la distancia topológica entre N...N, tres descriptores que cuentan los grupos funcionales (nRNHR el número de aminas secundarias alifáticas, nArOH el número de hidroxilos aromáticos y nPyrazines el número de pirazinas), dos descriptores de huellas binarias 2D que miden la presencia

o ausencia de enlaces O-O y O-Br a la distancia de cuatro y ocho carbonos respectivamente (B04[O-O], B08[O-Br]), así como dos descriptores de huellas frecuenciales 2D que calculan la frecuencia de enlace F04[O-O]O-O y F09[N-N]N-Na la distancia de cuatro y nueve carbonos respectivamente. Se utilizó como esquema algorítmico: `[WEKA.classifiers.functions.SMO-C10.0-L0.001-P1.0E-12-N0-V-1-W1-K.WEKA.classifiers.functions.supportVector.RBFKernel-C 250007-G 1.0 -t SVM_training.arff-T SVM_Test.arff-o]` en la cual C es complejidad, *numFolds(-1)* (o incremento numérico) nos indica el número de pliegues de validación cruzada utilizados para generar datos de entrenamiento para modelos logísticos y G es el valor gamma. La serie de entrenamiento reportó un 87.68% de instancias clasificadas correctamente como activas e inactivas con respecto al total y la serie de predicción o un 82.35%. El resto de los parámetros estadísticos quedaron plasmados en la tabla IV. Cabe resaltar que dentro de los parámetros estadísticos obtenidos el más significativo fue el de la exactitud para la SE (87,68%). En la figura 25 y 26 (79, 80) se presentan los valores de Área Bajo la Curva (ABC) ROC para los compuestos activos e inactivos de las SE y SP.

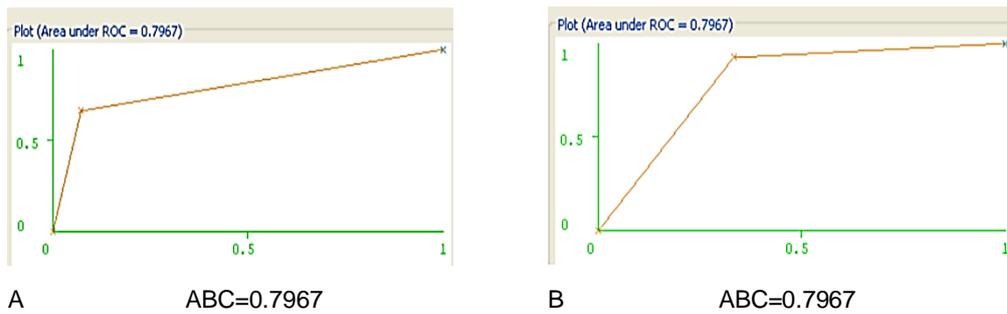
**Tabla IV.** Parámetros estadísticos para el modelo máquina vectorial de soporte.

Serie	C	Q (%)	Sensibilidad(%)	Especificidad(%)	FP <sub>r</sub> (%)
Entrenamiento	0.74	87.68	84.15	84.15	10.08
Predicción	0.63	82.35	66.67	85.71	7.32

**Legenda:** Q exactitud, C coeficiente de correlación de Matthews; FPr Razón de falsos positivos.



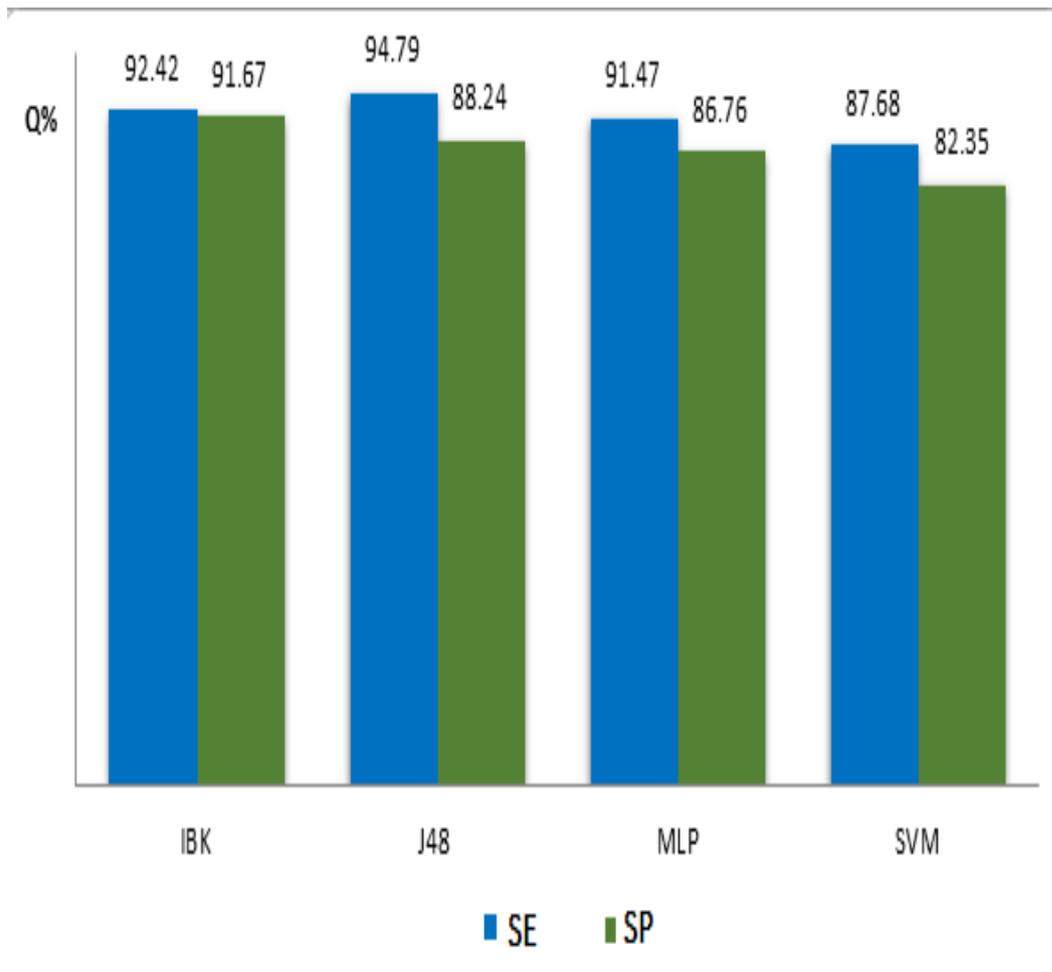
**Figura 25.** Curvas ROC para compuestos activos (A) e inactivos (B) de la SE para el modelo IBK.



**Figura 26.**Curvas ROC para compuestos activos(a) e inactivos(b) de la SP para el modelo IBK.

Los estudios realizados anteriormente con *L. amazonensis* y *L. infantum* varían en complejidad con valores de 90.0 y 50.0, y en el valor gamma siendo estos de 0.9 y 3.9 respectivamente. El modelo para *L. major* presenta un 87.68% de exactitud en la SE, elevado en comparación con este parámetro estadístico para la SE (81,47%) de *L. infantum* en un 6% y menor en un 4% respecto al alcanzado para *L. amazonensis* en la SE.

En la figura 27 se muestran los porcentajes de clasificación para las series de entrenamiento y validación cruzada en los modelos finales obtenidos en este trabajo, resultando superiores los del modelo IBK seguidos por la técnica J48. Teniendo en cuenta los cuatro modelos finales obtenidos, el valor más significativo de exactitud pertenece a la técnica del árbol con un 94.79% de compuestos correctamente clasificados para la SE, en el caso de la sensibilidad su valor más elevado fue de 93.90% de la SE correspondiente a la técnica del vecino más cercano, los mejores resultados de especificidad corresponden a los modelos del árbol SE (93.83%), SP (88.0) y MLP SE (91.03%), SP (90.91%) y los falsos positivos alcanzaron para la técnica del árbol solo un 3.88%(SE).



**Figura 27.** Porcentajes de clasificación para las series de entrenamiento y serie de validación cruzada en los modelos finales obtenidos.

### Validación de los modelos

Tras efectuar la validación de los cuatro modelos derivados de las diferentes técnicas por dos métodos diferentes (empleando una serie externa y validación cruzada) se obtuvieron resultados que avalan la robustez y capacidad predictiva de cada uno de ellos, los cuales se muestran a continuación (tabla V)

Los valores de exactitud para las series de predicción externa y la validación cruzada oscilan entre el 91 y el 81 % en todos los casos, la sensibilidad es también elevada logrando alcanzar un 100% para la serie de validación externa con la técnica IBk, resultando los valores de falsos positivos más bajos para el árbol J48 así como también más altos los parámetros de especificidad.

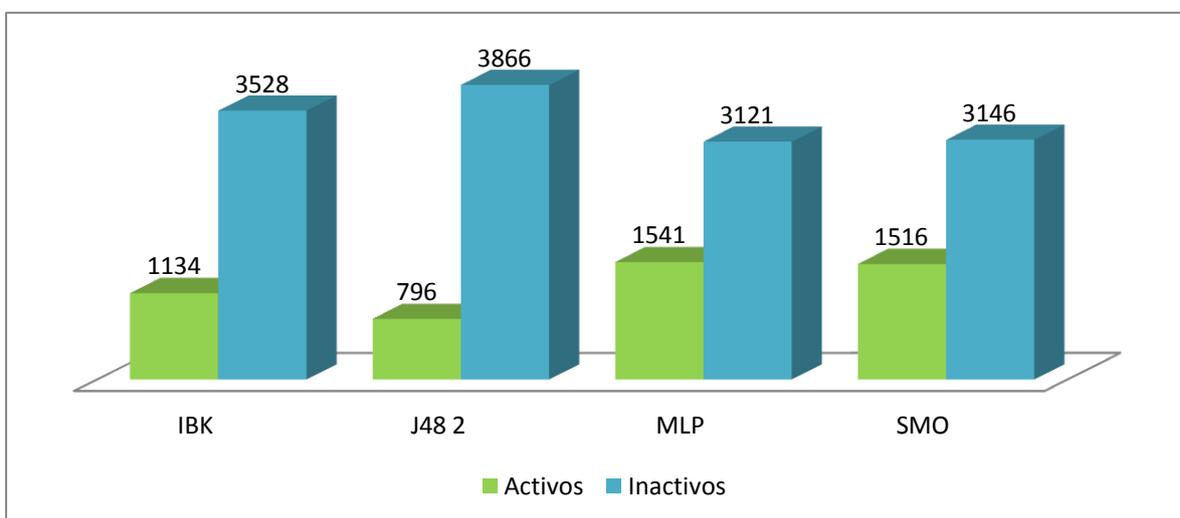
**Tabla V.** Parámetros estadísticos obtenidos para el modelo en la SPE y la VC

Serie	C	Q (%)	Sensibilidad(%)	Especificidad(%)	FP <sub>A</sub> (%)
<b>IBk</b>					
Predicción Externa	0.84	91.67	100.00	82.61	13.79
Validación cruzada	0.74	87.20	90.24	79.57	14.73
<b>J48</b>					
Predicción Externa	0.70	85.42	84.21	80.00	13.79
Validación cruzada	0.74	87.68	82.93	85.00	9.30
<b>MLP</b>					
Predicción Externa	0.60	81.25	73.68	77.78	13.79
Validación cruzada	0.68	84.36	85.37	76.92	16.28
<b>SMO</b>					
Predicción Externa	0.60	81.25	68.42	81.25	10.34
Validación cruzada	0.68	84.36	84.15	77.53	15.50

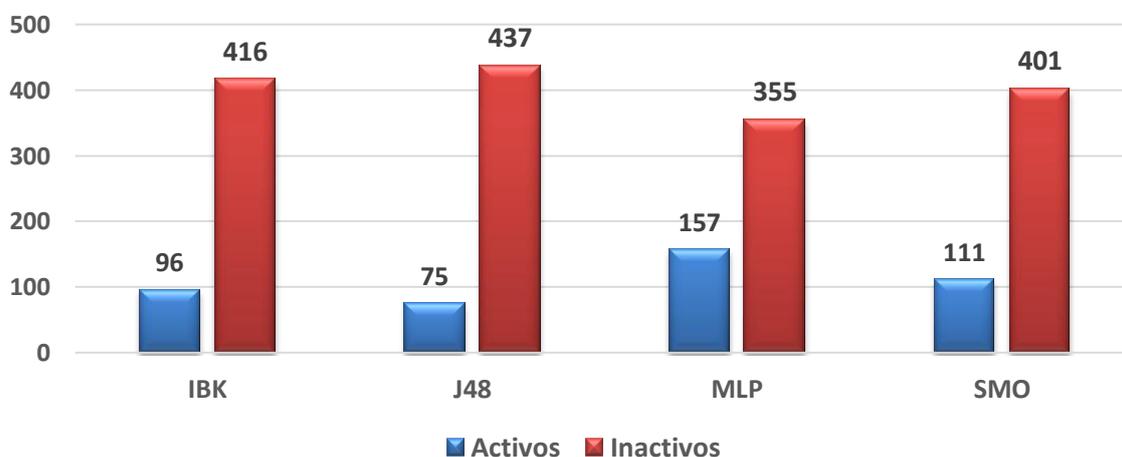
### 3.2 Cribado de Base de datos y ensamblado.

En este trabajo fueron empleados para el cribado virtual un total de 5174 compuestos, de los cuales 4662 pertenecen al DrugBank, 492 son compuestos de síntesis y 20 son aislados a partir de la planta *Boldoa purpurascens* Cav. A continuación exponemos los resultados obtenidos los cuales permitieron la identificación de nuevas entidades químicas con potencial actividad contra promastigotes de *L. major* con una amplia variedad estructural por cada uno de los cuatros modelos (IBK, J48, MLP, SVM). En la

figura 28 se representa el total de compuestos identificados como activos e inactivos por cada técnica luego del cribado de la base de datos DrugBank y en la figura 29 los resultados del cribado del resto de los compuestos. Fueron identificados como potencialmente antileishmaniásicos una gran cantidad de entidades químicas con diferentes esqueletos estructurales, ver anexo 1.



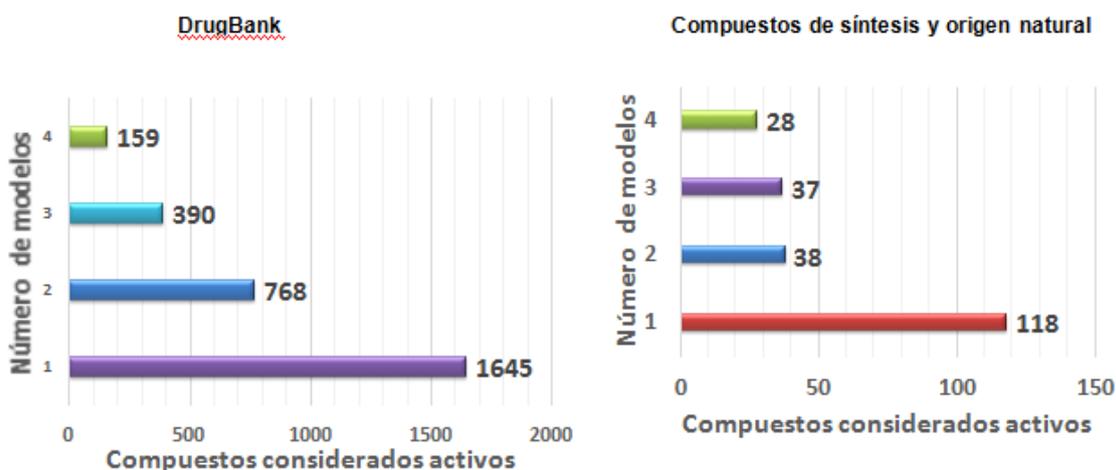
**Figura 28.** Resultados del cribado de la base de datos DrugBank para los cinco modelos.



**Figura 29.** Resultados del cribado los compuestos de síntesis y de origen natural.

Con el objetivo de lograr reducir aún más el número de compuestos identificados como potencialmente activos se procedió a realizar un ensamblado de voto, teniendo en cuenta el resultado del cribado obtenido por cada técnica y otorgando igual peso a cada uno de los modelos, de esta forma aquellos compuestos identificados como activos por todas las técnicas fueron seleccionados (165). De esta forma se

analizaron los posibles ensamblados a partir de los modelos obtenidos por las cuatro técnicas de inteligencia artificial, resultando mejor aquel que incluía los resultados de los cuatro modelos en conjunto, es decir, serían identificados como potencialmente activos aquellos compuestos predichos por la totalidad de los modelos. En la figura 30 puede observarse el número de compuestos predichos como activos por dichos modelos luego del cribado del DrugBank y el resto de los compuestos.



**Figura 30.** Número de compuestos considerados activos por los diferentes modelos. Ensamblado empleando voto mayoritario simple.

Como puede observarse en la figura anterior se logra reducir el número de compuestos predichos como activos a medida que se incrementa el número de modelos que los predicen como tales, pues cada modelo capta por separado una gran fracción de la información química incluida en la base de datos a través de los DMs que usa pero con la complicidad del resto de los modelos se logra obtener una gran congruencia y mejora del resultado final, siendo la categorización más notoria la que resulta elegida por el conjunto de todos los modelos como decisión final. Los compuestos identificados de esta forma como potencialmente activos incluyeron De ellos 159 de la base de datos DrugBank, 14 nuevos compuestos de síntesis y 14 compuestos de origen natural. Su identificación constituye un paso crucial en la investigación de estructuras antileishmaniásicas, lo cual podrá corroborarse en próximos estudios de manera experimental. El protocolo más popular es el voto de la pluralidad, donde la clase con la generalidad de los votos es la escogida.

## CONCLUSIONES

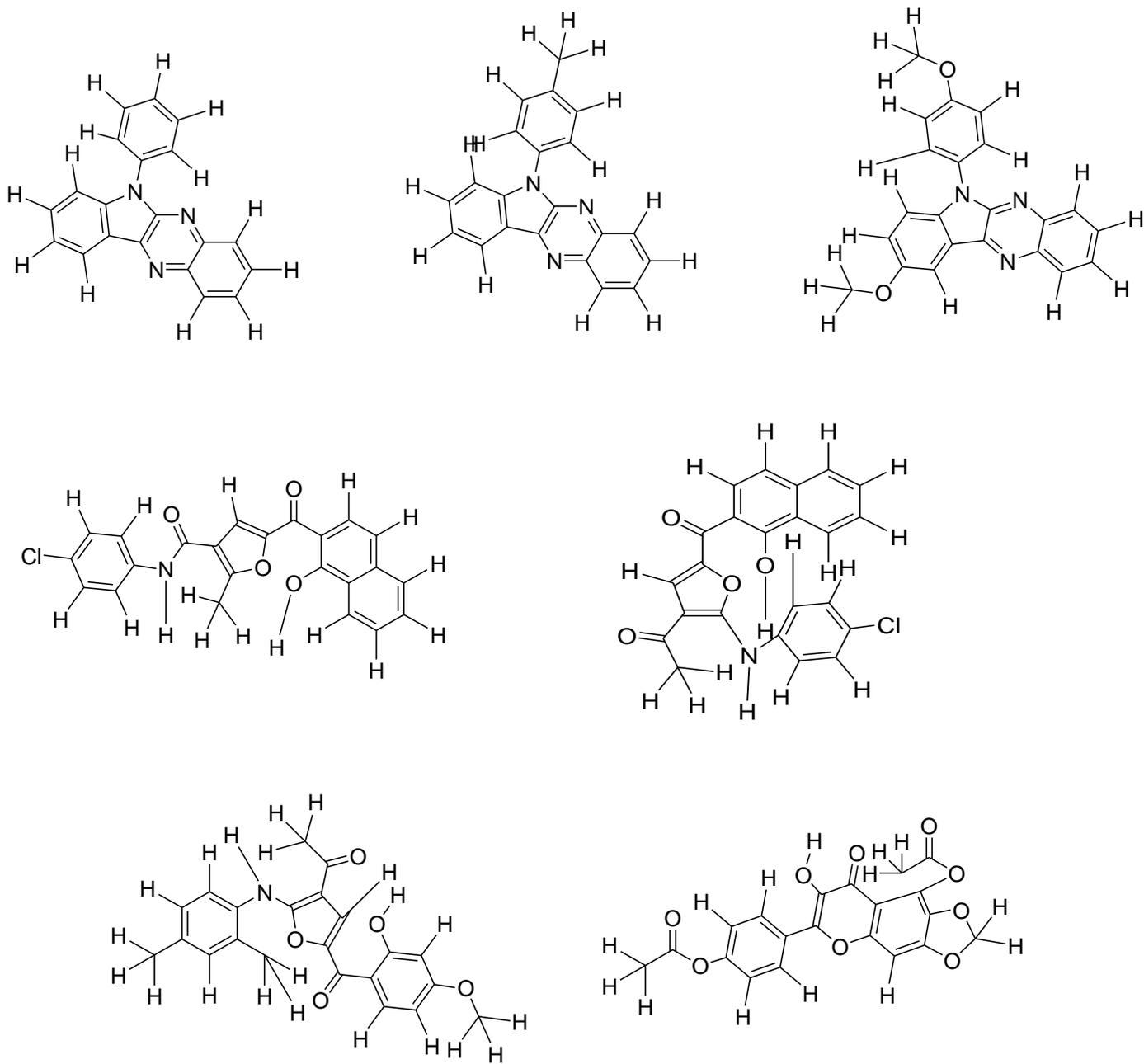
1. La base de datos obtenida incluyó 336 compuestos evaluados experimentalmente contra la forma promastigote del parásito *Leishmania major* con una amplia variabilidad estructural.
2. Los cuatro modelos QSAR desarrollados con técnicas de inteligencia artificial (IBk, J48 1, J48 2, MLP y SMO) alcanzaron buenos porcentajes de clasificación, sensibilidad y especificidad tanto para las series de entrenamiento como las de predicción.
3. La validación externa e interna de los modelos obtenidos para promastigotes demostró su alto grado de predictibilidad, robustez y reproducibilidad.
4. Los modelos obtenidos y el método de ensamblado empleado permitieron la identificación de 187 nuevos compuestos con potencial actividad contra *Leishmania major* a través del cribado virtual de una base de datos internacional DrugBank, compuestos de síntesis y de origen natural.

## RECOMENDACIONES

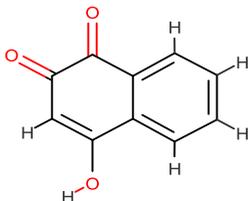
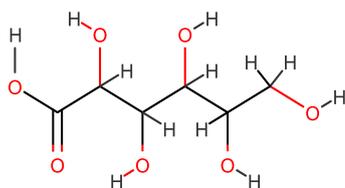
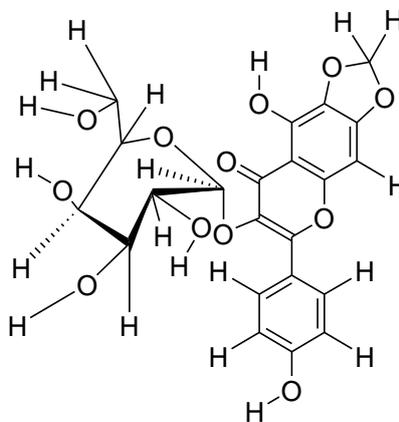
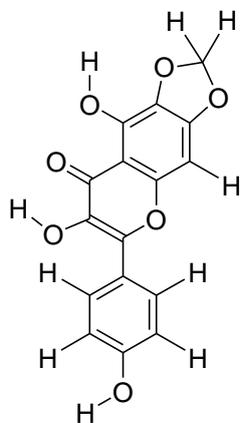
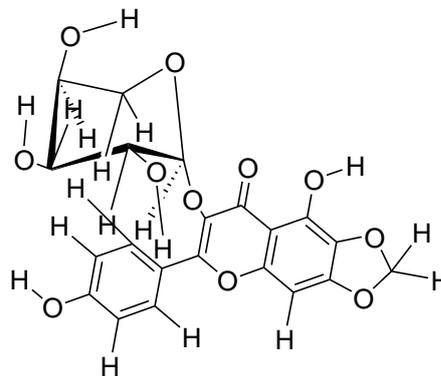
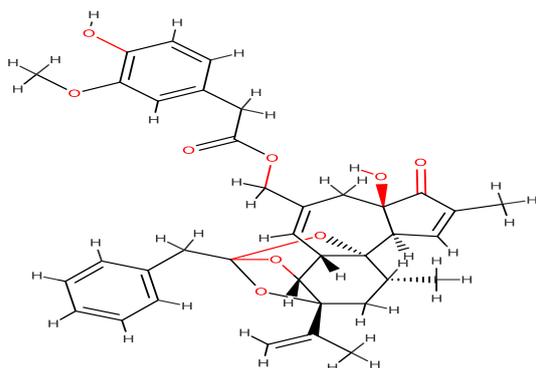
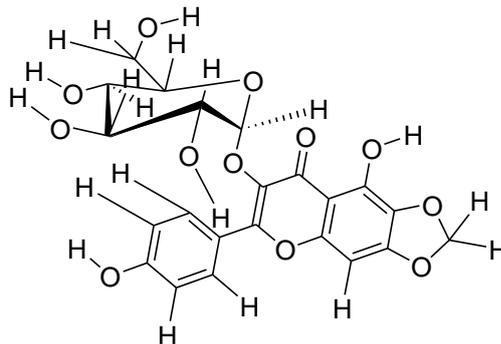
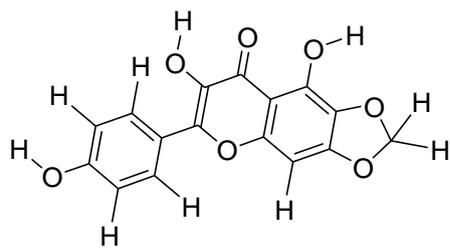
1. Evaluar experimentalmente los compuestos identificados como potencialmente activos para corroborar su actividad.
2. Obtener nuevos modelos de clasificación para la forma amastigote del parásito *L. major*.

# ANEXO1

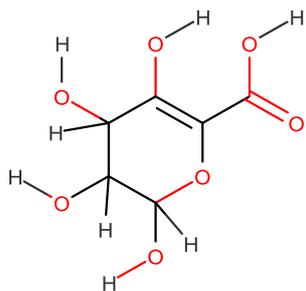
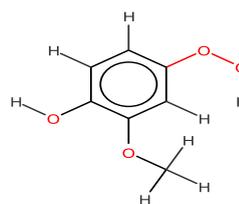
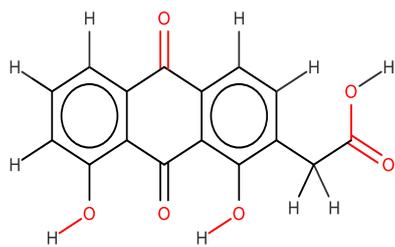
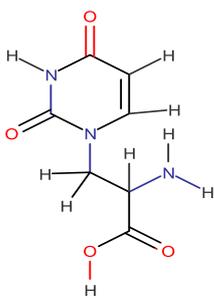
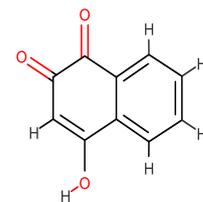
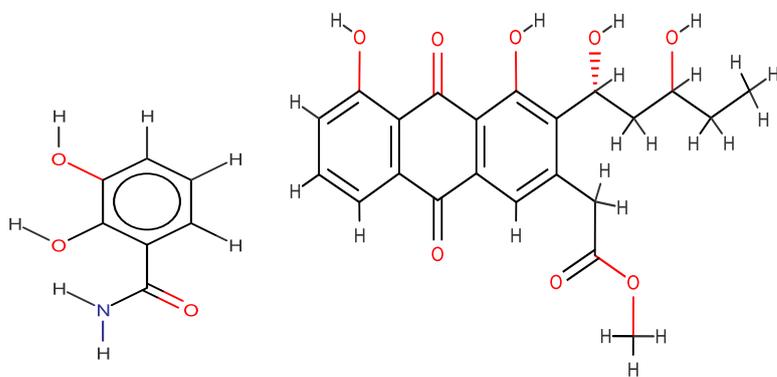
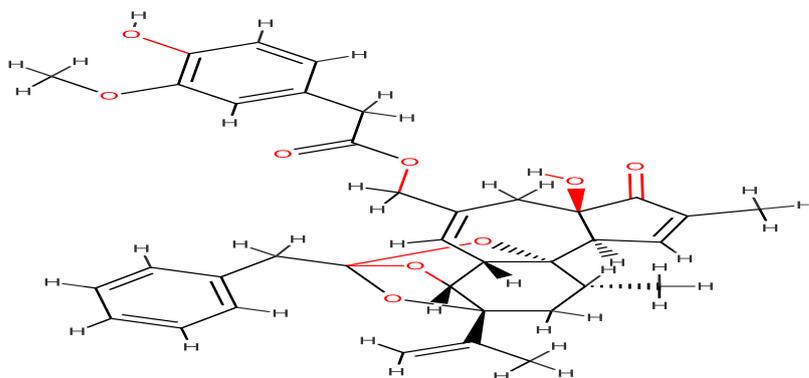
## Representación de los compuestos identificados como potencialmente activos



# ANEXO1 Continuación



# ANEXO1 Continuación



## **Bibliografía**

- 1. Miraballes Ferrer MC. Leishmaniasis: enfermedad emergente en Uruguay. 2013.**
- 2. Estrada E. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. Journal of chemical information and computer sciences. 1996;36(4):844-9.**
- 3. M. H, R. G. ChemBiol Curr Opin 1999.**
- 4. Marrero-Ponce Y, Siverio-Mota D, Gálvez-Llompert M, Recio MC, Giner RM, García-Domènech R, et al. Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: the nitroindazolinone chemotype. European journal of medicinal chemistry. 2011;46(12):5736-53.**
- 5. Chang P, Bray RS. Leishmaniasis: Elsevier Science Publishers BV (Biomedical Division); 1986.**
- 6. Herwaldt BL, Berman JD. Recommendations for treating leishmaniasis with sodium stibogluconate (Pentostam) and review of pertinent clinical studies. The American journal of tropical medicine and hygiene. 1992;46(3):296-306.**
- 7. Desjeux P. Leishmaniasis: public health aspects and control. Clinics in dermatology. 1996;14(5):417-23.**
- 8. Foroumadi A, Pournourmohammadi S, Soltani F, Asgharian-Rezaee M, Dabiri S, Kharazmi A, et al. Synthesis and in vitro leishmanicidal activity of 2-(5-nitro-2-furyl) and 2-(5-nitro-2-thienyl)-5-substituted-1, 3, 4-thiadiazoles. Bioorganic & medicinal chemistry letters. 2005;15(8):1983-5.**
- 9. Anzola NS. Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario. ODEON. 2016(9):113-72.**
- 10. SERRANO-MARTIN X, GONZÁLEZ-LUNA D, FUENTES J, CAMPOS J, FRANCO MA, PÉREZ JL, et al. Evaluación de drogas en Leishmania sp.: Estudios genéticos y moleculares de nuevos blancos. Estudios in vitro e in vivo de la resistencia a drogas. Mem IBE. 2005;4:117-20.**
- 11. DE TDLE. GESTIÓN PARA LA VIGILANCIA ENTOMOLÓGICA Y CONTROL DE LA TRANSMISIÓN DE LA ENFERMEDAD DE CHAGAS.**
- 12. F. FAC. Determinación de los factores de riesgo y su relación con Leishmaniasis Cutanea en los pacientes que acuden al laboratorio del Centro de Salud Satelital Central del Tena Provincial de Napo en el periodo diciembre 2012-julio 2013: Universidad Técnica de Ambato; 2014.**
- 13. M. G, C. R. Leishmania: Barcelona. Available from: workforce.cup.edu/buckelew/ Leishmania.**

14. Chaben MF. Leishmaniasis 2017. Available from: <http://www.anlis.gov.ar/inp/wp-content/uploads>.
15. Control de las leishmaniasis Ginebramarzo 2010. Available from: [http://apps.who.int/iris/bitstream/10665/82766/1/WHO TRS 949 spa.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/82766/1/WHO_TRS_949_spa.pdf?ua=1).
16. Sánchez DPC. Moléculas bioactivas contra Leishmania (viannia) panamensis. Actividad y optimización molecular. Medellín: Universidad de Antioquía; 2006.
17. Almeida Zurita PM. Análisis comparativo del nivel de conocimientos sobre Leishmaniasis Cutánea entre el personal médico de zonas endémicas del primer nivel de salud (área Pedro Vicente Maldonado y área Santo Domingo de los Tsáchilas): PUCE; 2016.
18. Zuluaga M, Robledo SM. Las células de Langerhans en la inmunidad a leishmaniasis. Biomédica. 2004;24(3):302-17.
19. Belkaid Y, Piccirillo CA, Mendez S, Shevach EM, Sacks DL. CD4+ CD25+ regulatory T cells control Leishmania major persistence and immunity. Nature. 2002;420(6915):502.
20. Salud OPdl. Leishmaniasis: Informe epidemiológico de las Américas. OPS Washington DD; 2013.
21. Yuil JMR, Sousa O. Inmunología en la infección por Leishmania: conceptos actuales. Rev méd cient. 2010;23(1):11-23.
22. Rosal Rabes Td, Baquero-Artigao F, García Miguel M. Leishmaniasis cutánea. Pediatría Atención Primaria. 2010;12(46):263-71.
23. Mateos DM. Estrategias celulares y moleculares en el cribado de nuevos agentes anti-Leishmani. Madrid: Universidad de Alcalá; 2010.
24. Alvarez AMM. Tipificación de especies de Leishmania de importancia médica basada en el gen que codifica la proteína HSP70 citoplasmática [Tesis presentada en opción al grado científico de Doctor en Ciencias de la Salud]. La Habana: Instituto de Medicina Tropical "Pedro Kourí" 2011.
25. Olmo F, Guardia J, Marin C, Messouri I, Rosales M, Urbanová K, et al. Prospects of an alternative treatment against Trypanosoma cruzi based on abietic acid derivatives show promising results in Balb/c mouse model. European journal of medicinal chemistry. 2015;89:683-90.
26. Balmaseda NF. Identificación de nuevos compuestos con potencial actividad antileishmaniásica mediante estudios in silico. Santa Clara Universidad Central Marta Abreu de las Villas 2015.
27. Álvarez ER. La anfotericina B en el tratamiento de la Leishmaniosis. Madrid: Universidad Complutense de Madrid; 2015 Junio.

29. Henaó HH, Osorio Y, Saravia NG, Gómez A, Travi B. Efficacy and toxicity of pentavalent antimonials (Glucantime® and Pentostam®) in an American cutaneous leishmaniasis animal model: luminometry application. *Biomédica*. 2004;24:393-402.
30. Pérez SM. Manual práctico de tratamiento de la Leishmaniasis. MINISTERIO DE SALUD Y PREVISIÓN SOCIAL  
UNIDAD NACIONAL DE VIGILANCIA EPIDEMIOLOGICA 2002;Publicación Técnica N° 10.
31. Ramos J, Segovia M. Estado actual del tratamiento farmacológico de la leishmaniasis. *Revista española de Quimioterapia*. 1997;10:26-35.
32. Rabes TdR, Baquero-Artigao F, Fernández CG, Miguela MJG, Laguna RdL. Treatment of cutaneous leishmaniasis with liposomal amphotericin B. *Anales de pediatría*. 2010;73(2):101-2.
33. Roy P, Das S, Auddy RG, Mukherjee A. Biological targeting and drug delivery in control of Leishmaniasis. *J Cell Anim Biol*. 2012;6(6):73-87.
34. J. LS. Identificación de nuevos compuestos líderes con actividad Antileishmaniásica a través de estudios "in silico: Universidad Central Marta Abreu de las Villas; 2010.
35. del Rosal Rabes T, Baquero-Artigao F, Miguel MG. Leishmaniasis cutánea. *Revista pediatría de atención primaria*. 2010;12(46):263-71.
36. Sundar S, Olliaro PL. Miltefosine in the treatment of leishmaniasis: Clinical evidence for informed clinical risk management. *Ther Clin Risk Manag*. 2007;3(5):733-40.
37. Hann M, Green R. *Current Opinion in Chemical Biology*. 1999.
38. Sayers JL. Identificación de nuevos compuestos líderes con actividad Antileishmaniásica a través de estudios "*in silico*" [Tesis para optar por el Título de Licenciada en Ciencias Farmacéuticas ]: UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS; 2009-2010
39. Sayers JL. Identificación de nuevos compuestos líderes con actividad Antileishmaniásica a través de estudios "in silico": Universidad Central Marta Abreu de las Villas. 2010.
40. Marrero-Ponce Y, Siverio-Mota D, Gálvez-Llompert M, Recio MC, Giner RM, García-Domènech R. Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: The nitroindazolinone chemotype. *European journal of medicinal chemistry*. 2011;46(12):5736-53.
41. Sánchez PL, Fernández-Baza A, Fernández-Avilés F. ¿Qué son las redes temáticas de investigación cooperativa en salud (RETICS)? . *Cardiocre*. 2011;46(4):131-5.

42. Todeschini R, Consonni V. Handbook of Molecular Descriptors. Federal Republic of Germany.2000.
43. González AIA. Modelación de la actividad antimalárica de compuestos orgánicos frente al parásito protozoario Plasmodium falciparum [Trabajo de Diploma]. Santa Clara: Universidad Central Marta Abreu de Las Villas; 2015-2016.
44. Anzola NS. Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario. ODEON. 2015(9):113-72.
45. Martínez EA, Perez-Riverol DY. Estimación del punto isoelectrico de péptidos empleando descriptores moleculares y máquinas de soporte vectorial.
47. Sucar LE. Métodos de Inteligencia Artificial. p. 20.
48. Lopez J, Bravo EFC. Una aproximación práctica a las Redes Neuronales Artificiales. Primera ed: Universidad del Valle; December 2009.
49. Ramírez V, Solano R. Inteligencia artificial avanzada árboles de clasificación. Inteligencia artificial avanzada árboles de clasificación2009.
50. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic acids research. 2011;39(suppl 1):D1035-D41.
61. Mills N. ChemDraw Ultra 10.0 CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. www. cambridgesoft. com. Commercial Price: 1910fordownload, 2150 for CD-ROM; Academic Price: 710fordownload, 800 for CD-ROM. Journal of the American Chemical Society. 2006;128(41):13649-50.
62. Ruggiu F, Marcou G, Varnek A, Horvath D. ISIDA Property-Labelled Fragment Descriptors. Molecular Informatics. 2010;29(12):855-68.
63. Mollerach R, Fink J, SA NA. Utilizacion del programa dragon para el cálculo de mecanismos de reactividad.
64. Xu J, Hagler A. Chemoinformatics and drug discovery. Molecules. 2002;7(8):566-600.
65. Johnson RA, Wichern DW. Applied multivariate statistics. Englewood Cliffs, NJ: Prentice Hall; 1988.
66. van de Waterbeemd H, Rose S, Butler S. Quantitative approaches to structure-activity relationships. The practice of medicinal chemistry. 1996;3.
67. Nascimento DS, Coelho AL. Bagging heterogêneo evolutivo: Caracterização e análise comparativa com ensembles homogêneas de redes neurais RBF. Anais do IX Simpósio Brasileiro de Automação Inteligente. 2009:6.

68. Ariew R. Maquinilla de afeitar de Ockham: Un análisis histórico y filosófico del principio de parsimonia de Ockham. Universidad de Illinois: Champane-Urbana1976.
69. Wold S, Eriksson, L. Statistical Validation of QSAR Results. Chemometrics Methods in Molecular design. Weinheim (Germany): van de Waterbeemd, H., VCH; 1995. p. 309-18.
70. Castillo-Garit JA. Parametrización de la estructura química empleando formas lineales: Aplicaciones al Diseño Molecular Asistido por Ordenador [En opción al grado científico de Doctor en Ciencias Farmacéuticas]. Santa Clara, Cuba: Universidad Central Marta Abreu de las Villas; 2008.
71. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research. 2008;36(suppl 1):D901-D6.
72. Alfredo MM, Yanetsy MT, Yosvani MP, Vicente Jesús AR, José Antonio EG-T, Alicia GB, et al., editors. TAMIZAJE FARMACOLÓGICO EN LA BÚSQUEDA DE POTENCIALES FÁRMACOS ANTIMALÁRICOS INTEGRANDO NUEVOS MODELOS IN SILICO Y CORROBORACIÓN EXPERIMENTAL. 8th Cuban Congress on Microbiology and Parasitology, 5th National Congress on Tropical Medicine and 5th International Symposium on HIV/aids infection in Cuba; 2014.
73. Opitz D, Maclin R. Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research. 1999:169-98.
74. Goebel K, Yan W, editors. Choosing classifiers for decision fusion. Proceedings of the Seventh International Conference on Information Fusion; 2004.
75. Hansen RA, Deutscher T. An empirical investigation of attribute importance in retail store selection: College of Administrative Science, Ohio State University; 1976.
76. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H. Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. Bioinformatics. 2000;16:412-24.
77. Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning. 1999;36(1-2):105-39.
78. Dietterich TG. Ensemble methods in machine learning. Multiple classifier systems: Springer; 2000. p. 1-15.
79. Flores Balmaseda N. Identificación de nuevos compuestos con potencial actividad antileishmaniásica mediante estudios in silico: Universidad Central" Marta Abreu" de Las Villa. Centro de Bioactivos Químicos; 2015.
80. Flores Balmaseda N, Marrero Ponce Y, Louvina Sayer J, Merchán T, Sebastián D. Identificación de compuestos líderes con potencial actividad

**antileishmaniásica: una contribución a la farmacéutica médica. Medicentro Electrónica. 2017;21(1):39-45.**