

Review

Colombian Contributions Fighting Leishmaniasis: A Systematic Review on Antileishmanials Combined with Chemoinformatics Analysis

Jeysson Sánchez-Suárez ¹, Freddy A. Bernal ^{2,†} and Ericsson Coy-Barrera ^{2,*}

¹ Bioprospecting Research Group, School of Engineering, Universidad de La Sabana, Chía 250001, Colombia; jeyssonsasu@unisabana.edu.co

² Bioorganic Chemistry Laboratory, Universidad Militar Nueva Granada, Cajicá 250247, Colombia; Freddy.Bernal@hki-jena.de

* Correspondence: ericsson.coy@unimilitar.edu.co

† Current address: Transfer Group Anti-infectives, Leibniz Institute for Natural Product Research and Infection Biology, HKI, Beutenbergstraße 11a, 07745 Jena, Germany.

Academic Editors: Susana Santos Braga, Carlos J. P. Monteiro and Carlos Silva

Received: 15 November 2020; Accepted: 30 November 2020; Published: 3 December 2020



Abstract: Leishmaniasis is a parasitic morbid/fatal disease caused by *Leishmania* protozoa. Twelve million people worldwide are appraised to be currently infected, including *ca.* two million infections each year, and 350 million people in 88 countries are at risk of becoming infected. In Colombia, cutaneous leishmaniasis (CL) is a public health problem in some tropical areas. Therapeutics is based on traditional antileishmanial drugs, but this practice has several drawbacks for patients. Thus, the search for new antileishmanial agents is a serious need, but the lack of adequately funded research programs on drug discovery has hampered its progress. Some Colombian researchers have conducted different research projects focused on the assessment of the antileishmanial activity of naturally occurring and synthetic compounds against promastigotes and/or amastigotes. Results of such studies have separately demonstrated important hits and reasonable potential, but a holistic view of them is lacking. Hence, we present the outcome from a systematic review of the literature (under PRISMA guidelines) on those Colombian studies investigating antileishmanials during the last thirty-two years. In order to combine the general efforts aiming at finding a lead against *Leishmania panamensis* (one of the most studied and incident parasites in Colombia causing CL) and to recognize structural features of representative compounds, fingerprint-based analyses using conventional machine learning algorithms and clustering methods are shown. Abstraction from such a meta-description led to describe some function-determining molecular features and simplify the clustering of plausible isofunctional hits. This systematic review indicated that the Colombian efforts for the antileishmanials discovery are increasingly intensified, though improvements in the followed pathways must be definitively pursued. In this context, a brief discussion about scope, strengths and limitations of such advances and relationships is addressed.

Keywords: leishmania parasites; leishmanicidal; neglected tropical diseases; chemoinformatics; machine learning; Colombia; *Leishmania panamensis*

1. Introduction

Leishmaniasis is a vector-borne parasitic zoonosis caused by protozoa of the genus *Leishmania*, which is considered as an important neglected tropical disease (NTD). Clinically, this disease is classified as cutaneous (CL), mucosal (ML), or visceral leishmaniasis (VL). Central and South America are among the most affected regions, registering an annual incidence of 54,950 cases between 2001 and

2018 [1]. Among the 18 American countries where leishmaniasis is considered an endemic disease, Brazil, Colombia and Peru are those with the highest number of cases, involving 16432, 6362 and 6321 respectively, for 2018 [1]. In Colombia, the incidence rate of this disease was 26.2 cases per 100,000 population, with 98.6% of the cases related to CL [2]. Such an incidence was due to the presence of several parasites species, including *L. venezuelensis*, *L. equatorensis*, *L. lainsoni*, *L. colombiensis*, *L. mexicana*, *L. amazonensis*, *L. infantum*, *L. guyanensis*, *L. braziliensis* and *L. panamensis* [3–6], with the last three *Leishmania* species being the most representative etiological agents [4,6].

Despite the efforts involved in the development of new chemotherapeutic options/alternatives, the use of pentavalent antimony compounds still remains the first-line treatment today [7]. These drugs are recognized by their side effects [8–10], which implies an additional effort to monitor patients under treatment [11]. Moreover, the increasing number of therapeutic failure (mainly associated with parasite drug resistance) [12,13] establishes the need to persist in the search for more effective and safer antileishmanial agents. In this context, the accumulated knowledge about leishmaniasis pathophysiology, parasite biology and advancements in high-throughput screening, big-data analysis, analytical platforms, extraction/isolation and organic synthesis, open up new opportunities regarding the fight against NTDs such as leishmaniasis.

In 2013, the state-of-the-art regarding leishmaniasis research in Latin America showed that Brazil and Colombia were the most-contributing countries. However, Colombia's scientific production was far from that of Brazil (almost six-fold lower production) [14]. After our current search, we found that this scenario has not practically changed, due to the fact that the burden/liability has been assumed by a small set of research groups basically disconnected from industry partners. Despite this, Colombian research on antileishmanials has led to the discovery of interesting chemical entities (both from natural and synthetic origin) with prospective activity against promastigotes and/or amastigotes of different *Leishmania* species. Their results have separately demonstrated high potential involving possible hits, but a holistic and comprehensive overall view of the outcome of those studies remains to be uncovered. Such a view would allow understanding/delineating the current status and future directions on further drug discovery-based initiatives against *Leishmania* parasites.

Increasing efforts are constantly paid by researchers around the world to improve the current drug discovery pipelines. Hence, the advances of computational methods and their application to them constitute a significant component. Accordingly, the use of novel and better algorithms have led to a large number of publications showing the extent of their applicability in computer-aided drug design projects [15–17]. The impact of chemoinformatics, understood nowadays as a discipline intersecting chemistry and computer science [18], has been moreover boosted by the development of machine learning algorithms in recent years [19–21]. Its use has extended across all levels of a typical drug discovery pipeline.

The so-called *in silico* methods have been thoroughly applied to a wide range of scientific problems, including the search for new treatments against infectious diseases, and more specifically, NTDs such as leishmaniasis, as extensively reviewed [22–25]. Chemoinformatics has also greatly influenced the renaissance of natural products, not only as a tool for identification of their vast biological potential but also aiming to fight NTDs [26–31]. Hence, a notorious use of chemoinformatics within the forthcoming Colombian research projects focused on antiparasitic agents is no less than expected.

As an endeavor to describe and characterize the status to date of the antileishmanial-focused Colombian studies, we present herein a systematic and comprehensive review of Colombian studies that have performed *in vitro* leishmanicidal trials. An approach to the chemical space conformed by the compounds involved in those research projects is disclosed for the first time. Finally, machine learning models are established for those compounds acting on amastigotes of *L. panamensis* (causative agent of CL in Colombia), including a particular emphasis on their interpretability and its relationship with important structural features.

2. Results and Discussion

2.1. Study Characteristics

The systematic review was performed under the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines: the flowchart is outlined in Figure 1. After removal of duplicate studies, 1029 articles were screened and 900 (87.46%) were then excluded. At this point, we intended to delineate how the leishmaniasis research has been approached in Colombia. Hence, we classified the excluded studies within the following categories: entomology (vector studies), therapeutics (e.g., pharmacokinetics/pharmacodynamics studies), parasite biology, disease pathophysiology, epidemiology, diagnosis, case report, miscellaneous (i.e., studies with mixed goals and those with goals outside the previously described categories) and unrelated (i.e., those studies that, although were retrieved by the search, actually did not focus on the leishmaniasis subject). Most of the excluded papers were centered on entomological approaches (22%), followed by the ones with therapeutics goals (18%). The full results of this classification are shown in Supplementary Figure S1. Regarding the 129 papers that passed this phase, each one was read and analyzed at the full-text level (access to all the selected papers was granted) and 29 of them were afterward excluded according to inclusion/exclusion criteria, which resulted in 100 studies for the qualitative synthesis. In this phase, we could identify that 46 studies provided outputs comparable to each other, which enabled performing a quantitative synthesis.

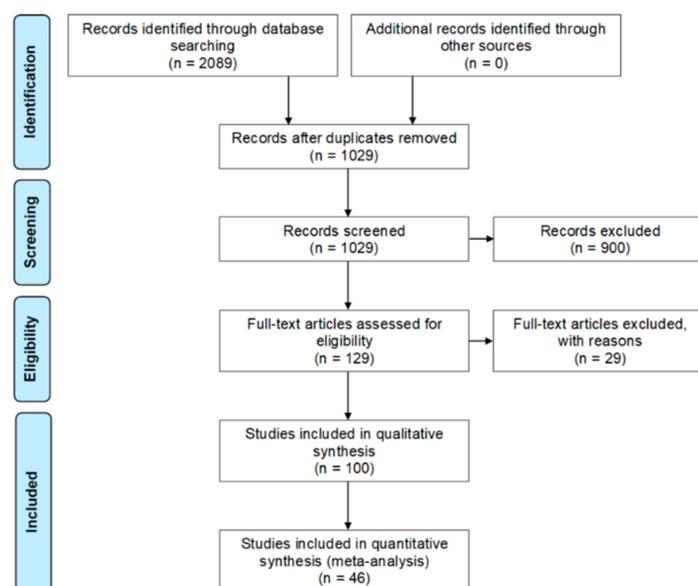


Figure 1. PRISMA flow diagram of this systematic review. Adapted from Moher et al. [32]. Compliance with the items in the statement guideline is presented in Supplementary Table S1.

2.2. General Findings

Publications showed an increasing trend between 1998 and 2020 (Figure 2A) as an indicator of the research importance on the discovery of new antileishmanial compounds. Remarkably, one institution leads the academic production in this field (*ca.* 40%), despite the wide distribution of leishmaniasis through several regions within the country (Figure 2B). Owing to the importance of this disease as a public health problem in Colombia, this fact calls for the local investment in other regions to improve the research capabilities on antileishmanials, even more considering the positive effect by the drug development from public research on the burden of neglected diseases [33]. It is worth noting that, despite that leishmaniasis' global burden has noticeably decreased between 2007 and 2017, CL—the most prevalent leishmaniasis form in Colombia—and ML did not follow this

trend, showing a significant disability-adjusted life year (DALY) rate increase (31.5%) in the same period [34]. The above-mentioned events and the therapeutic failure-related drawbacks of current antileishmanial drugs justify the persistent need for the discovery/development of more efficient and safer chemotherapy to treat CL. However, such a pipeline is generally prolonged (*ca.* 10 years) and very expensive (a general cost among \$1.4–2.9 billion) [35,36], generating a clear imbalance between the required investment and the local budget, which is a common limitation in developing countries. The condition in the case of NTDs-related research is even worse, since no big pharmaceutical company is committed to participate in an antileishmanial discovery program due to the unattractive incentives to cover the costs for the development of drugs against NTDs [37]. An alternative is related to promote effective partnerships with non-profit organizations, such as Drugs for Neglected Diseases *initiative* (DNDi) as a well-known example, whose modus operandi has covered those gaps to drive various compounds into lead optimization and pre-clinical phases [38].

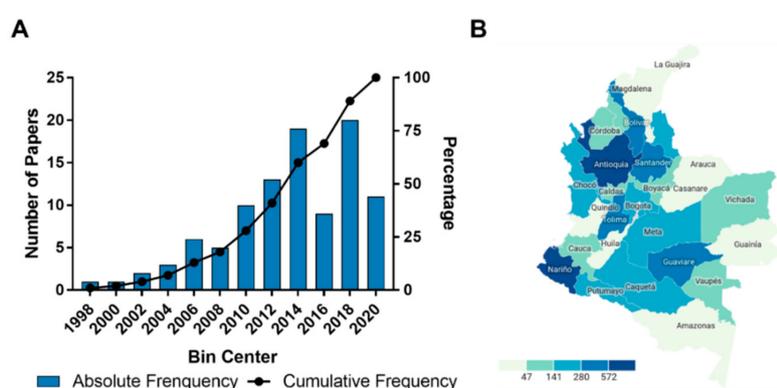


Figure 2. Evolution of the included scientific publications and geographical distribution of reported leishmaniasis cases in Colombia. (A) Time course of publications showing absolute (left-hand, *y*-axis) and cumulative frequency (right-hand, *y*-axis). (B) Number of reported cutaneous leishmaniasis (CL) cases for 2019 distributed through the political-administrative Colombian regions [39].

Although Colombia is a megadiverse country, most of the evaluated substances were obtained from synthetic approaches (Figure 3A). Regarding natural products, they were mostly obtained from terrestrial organisms (93%), while aquatic ecosystems (both marine and freshwater) remain as unexplored habitats (Figure 3B). Most of the studies (82.9%) used plants as a source of substances to be evaluated in antileishmanial assays, but the number of such records seems to be very low in comparison to other countries. These deeds call into question the efficiency of the natural products research in terms of antileishmanials discovery. Hence, a lack of organized/systematic programs, appropriately funded and directed to explore the potential of Colombian biodiversity under collaborative research networks, was evidenced.

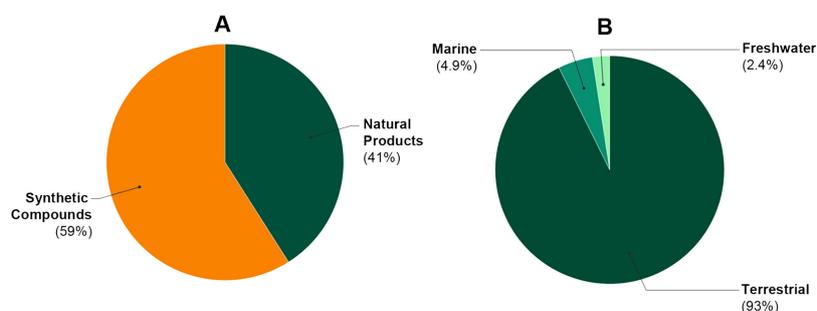


Figure 3. Classification of evaluated substances and environments explored for the research of antileishmanial agents in Colombia. (A) Publications classified according to the type of substance tested. (B) Publications evaluating natural products subdivided into the habitat source.

In addition, other sources should also be examined to look for novel, effective leishmanicidal agents. For instance, marine organisms have been considered as promising suppliers of compounds with novel structures and noteworthy antiparasitic activity [40–42], including antileishmanial properties [43–45]. In this sense, it may be worthwhile considering marine specimens in future studies, especially marine microorganisms, whose chemical composition and biological activity—to our knowledge—remain to be deeply explored. Among the studies employing natural products, 22 evaluated the antileishmanial activity of crude extracts or solvent/chromatographic fractions (Table 1). Activity of the ethanol extract of *Bomarea setacea* aerial part can be highlighted, since it exhibited half-maximal effective concentrations (EC₅₀) between 4.9 and 5.1 µg/mL against *L. amazonensis*, *L. braziliensis* and *L. donovani* promastigotes [46]. Only four studies showed promising activities against the clinically relevant stage of the parasite (i.e., intracellular amastigotes), comprising EC₅₀ values under 10 µg/mL [47–50]. From those studies, the dichloromethane extract from leaves of *Conobea scoparioides* was found to be the most promising extract (EC₅₀ = 1.30 µg/mL, selectivity index = 48.8) against *L. panamensis* [47]. This plant has been traditionally used for the treatment of leishmaniasis in Colombia [47]. However, despite the encouraging result, we did not find post-studies about the putative antileishmanial compounds isolated from *C. scoparioides*.

Table 1. Summary of the antileishmanial potential of crude extracts retrieved from the articles included in this systematic review.

<i>Leishmania</i> Species ^a	Parasites Form ^b	EC ₅₀ (µg/mL) ^c	E/F ^d	Source ^e	Ref ^h
<i>L. braziliensis</i> ; <i>L. infantum</i> ; <i>L. panamensis</i>	Promastigote	N/A ^e	1	<i>Annona spraguei</i>	[51]
<i>L. braziliensis</i> ; <i>L. panamensis</i>	Promastigote	N/A ^e	3	<i>Annona muricata</i>	[52]
<i>L. amazonensis</i> ; <i>L. braziliensis</i> ; <i>L. infantum</i> ; <i>L. panamensis</i>	Intracellular amastigote	1.30	88	<i>Conobea scoparioides</i>	[47]
<i>L. amazonensis</i> ; <i>L. braziliensis</i> ; <i>L. donovani</i>	Promastigote	10.70	36	<i>Rollinia pittieri</i>	[53]
<i>L. amazonensis</i> ; <i>L. braziliensis</i> ; <i>L. donovani</i>	Promastigote	4.90	26	<i>Bomarea setacea</i>	[46]
<i>L. braziliensis</i>	Intracellular amastigote	12.40	1	<i>Ervatamia coronaria</i>	[54]
<i>L. panamensis</i>	Intracellular amastigote	6.25	8	<i>Xylopiia discreta</i>	[48]
<i>L. braziliensis</i>	Promastigote	17.40	13	<i>Rosmarinus officinalis</i>	[55]
<i>L. amazonensis</i>	Axenic amastigotes	9.00	94	<i>Renealmia alpina</i>	[56]
<i>L. mexicana</i>	Axenic amastigotes	>50.00	452 ^f	<i>Several</i> ^g	[57]
<i>L. panamensis</i>	Intracellular amastigote	15.40	6	<i>Physalis peruviana</i>	[58]
<i>L. panamensis</i> ; <i>L. braziliensis</i> ; <i>L. major</i> ; <i>L. guyanensis</i>	Promastigote	42.23	10	<i>Origanum vulgare</i>	[59]
<i>L. panamensis</i>	Intracellular amastigote, Axenic amastigotes	38.50	6	<i>Piper daniel-gonzalezii</i>	[60]
<i>L. panamensis</i>	Intracellular amastigote	18.50	13	<i>Heliotropium indicum</i>	[61]
<i>L. panamensis</i> ; <i>L. major</i>	Promastigote, Intracellular amastigote	6.16	3	<i>Zanthoxylum monophyllum</i>	[49]
<i>L. panamensis</i>	Intracellular amastigote	48.07	1	<i>Artemisia annua</i>	[62]
<i>L. braziliensis</i> ; <i>L. panamensis</i>	Promastigote, Intracellular amastigote	9.19	4	<i>Lippia alba</i>	[50]
<i>L. panamensis</i>	Intracellular amastigote	30.70	8	<i>Pilocarpus alvaradoi</i>	[63]
<i>L. braziliensis</i>	in vivo on golden hamsters	N/A ^f	1	<i>Arnica montana</i>	[64]
<i>L. panamensis</i>	Promastigote , Intracellular amastigote	23.42	2	<i>Sarconesiopsis magellanica</i>	[65]
<i>L. panamensis</i>	Promastigote	N/A ^f	4	<i>Galleria mellonella</i>	[66]
<i>L. infantum</i> ; <i>L. braziliensis</i>	Promastigote	47.70	12	<i>Enterobacter hormaechei</i>	[67]

^a *Leishmania* species used in the antileishmanial assays. For studies involving more than one species, the most sensitive species are highlighted in bold font. ^b Parasite forms (i.e., promastigote, intracellular amastigote or axenic amastigote) employed in the antileishmanial assays. For studies involving more than one parasite form, the highlighted one in bold font indicates the result presented in this table. ^c Leishmanicidal half-maximal effective concentration (EC₅₀). The lowest EC₅₀ value reported in each study. ^d E/F = Number of Extracts/Fractions Assayed; ^e Scientific name of the source of the most active crude extract/fraction. ^f In these studies, the EC₅₀ values were not calculated/informed. ^g This study screened several plants from different Latin American countries. None of the 17 Colombian screened plants were found to be active in the range of test concentrations. ^h Ref = Cited reference.

Antileishmanial assays can be performed on different *Leishmania* species as well as parasite forms as experimental models, and this fact limits the comparison possibilities owing to the differential response of the test parasite. Thus, in order to know the distribution of *Leishmania* species and forms within the group of included papers, we examined the experimental model employed in the assay. As expected, the main model used to evaluate the antileishmanial potential was intracellular amastigotes (50%), followed by the evaluation on promastigotes (35%) (Figure 4A). Regarding *Leishmania* species, most of the studies were conducted on *L. panamensis* (Figure 4B). This fact can be considered reasonable due to the fact that *L. panamensis* is the etiological agent most frequently reported in Colombia [3,6]. However, recent studies reported that *L. braziliensis* and *L. guyanensis* have relevant epidemiological reports and wide spatial distribution in Colombia [68,69]. Considering Colombia as a country with a significant number of pathogenic *Leishmania* species in circulation, these facts are very critical, bearing in mind that parasite sensitivity to antileishmanial agents relies on *Leishmania* species [70,71]. For instance, a study evaluated dehydroabietylamine derivatives and found species-dependent susceptibility for several of them [72], being consistent with other reports [73,74]. Indeed, owing to the role of intra- and inter-species *Leishmania* susceptibility [71,75,76], the incorporation of drug-resistant strains during antileishmanial research programs have been strongly recommended [77]. Presumably, the availability of high-throughput infection models has limited the Colombian research using other *Leishmania* species, since we found that 43.0% of the studies involved the green fluorescent protein (GFP)-transfected *L. panamensis* strain [78], representing 61.0% of such studies.

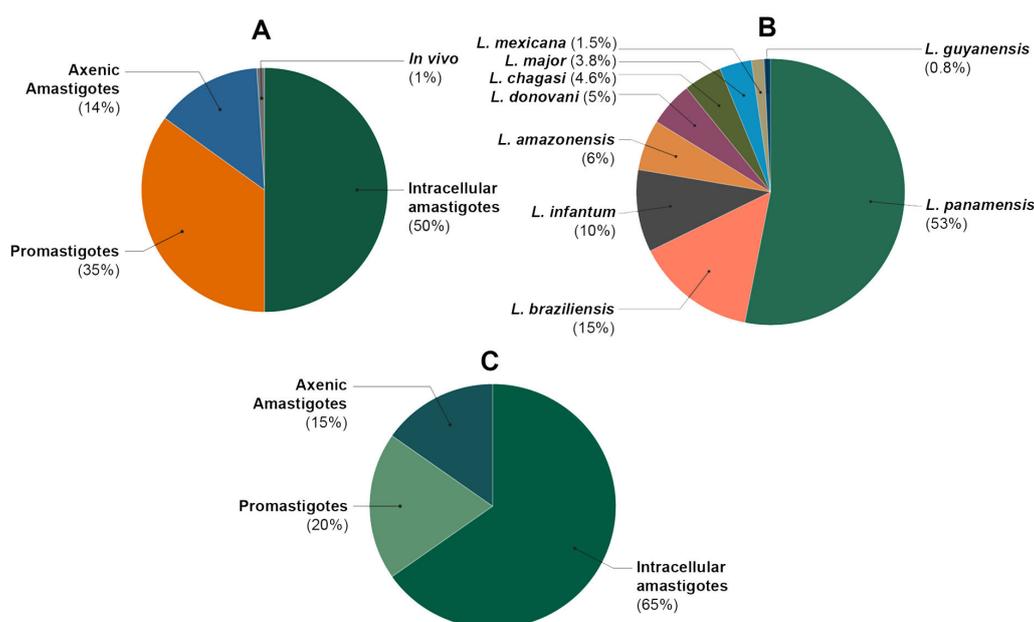


Figure 4. Characteristics of antileishmanial assays reported in those papers included in the present review. (A) Distribution of antileishmanial in vitro (subdivided into parasite forms, i.e., intracellular amastigotes, promastigotes and axenic amastigotes) and in vivo assays. (B) Distribution of *Leishmania* species involved in the antileishmanial in vitro assays. (C) Distribution of parasite forms used in antileishmanial in vitro assays against *L. panamensis*.

Additionally, we discerned that most of the reviewed basic studies were initiated from an exploratory basis, starting from in vitro studies to find bioactives against one or two *Leishmania* parasites and using synthetic compounds or natural substances (i.e., crude extracts, solvent/chromatographic fractions and isolated compounds). Some studies (<10%) continued in order to involve in vivo trials and/or expand the structural alternatives by synthesizing more compounds with related structures. However, there is still a lack of studies continuing down to further development, since none of the tested compounds have entered biomedical or clinical phases, possibly due to budget and/or continuity

issues or disconnection with pharmaceutical partners whose financial capacity and/or infrastructure might impede such a kind of innovative, required solutions.

On the other hand, a set of 836 compounds were retrieved from 75 articles. We have mined 1060 records from these papers (Table 2), since some studies involved evaluations against different *Leishmania* species, even different antileishmanial models (i.e., promastigotes, axenic amastigotes and intracellular amastigotes). However, intracellular amastigotes were found to be the most used antileishmanial assay model (65%, involving 46 articles) among in vitro trials against *L. panamensis* (Figure 4C). This valuable information was further exploited for the first time to perform a chemoinformatics-based analysis using these leishmanicidal results (as EC₅₀ values).

Table 2. Summary of antileishmanial activity of those compounds retrieved from the articles included in this systematic review.

Parasite Form ^a	Antileishmanial Activity Category ^b	Number of Records ^c
Intracellular amastigotes	High	127
	Intermediate	68
	Low	221
	Not Determined	162
	Not Available	112
Axenic amastigotes	High	28
	Intermediate	27
	Low	37
	Not Determined	32
	Not Available	28
Promastigotes	High	29
	Intermediate	44
	Low	81
	Not Determined	50
	Not Available	14
Total ^d		1060

^a Parasite form employed in the antileishmanial assay. ^b Categories according to the resulting negative decadic logarithm of the half-maximal effective concentration in mol/L (pEC₅₀) for each compound: High = pEC₅₀ ≥ 5.00 (EC₅₀ ≤ 10.0 μM), Intermediate = 4.60 ≤ pEC₅₀ < 5.00 (25.0 μM > EC₅₀ ≥ 10.0 μM), Low = pEC₅₀ < 4.6 (EC₅₀ > 25.1 μM), Not Determined = compounds included into the respective study, but the EC₅₀ value was over the maximum evaluated concentration, Not Available = compounds included into the respective study, but the antileishmanial assay did not return an EC₅₀. ^c Number of records regarding those test compounds with a pEC₅₀ value within the respective antileishmanial activity category and parasite form. ^d The raw data of this table is presented in Supplementary Table S2.

2.3. Chemoinformatics Analyses on Retrieved Compounds

A custom-made library was then compiled from the systematic review-derived records. Such a library gathered 836 compounds, containing 90.3% synthetic compounds and 9.7% natural products. Activity and structural details (as EC₅₀ in μM and Simplified Molecular-Input Line-Entry System (SMILES) codes, respectively) of these compounds can be found in Supplementary Table S2. The chemical space of the whole compound set was firstly examined in order to perform a structural filtering and qualitative characterization according to structural fragments/scaffolds. Therefore, a preliminary structural similarity analysis between test compounds was performed using the FragFp descriptor available in DataWarrior [79]. Compound 431 (having a linked thiophen-epoxybenzo[b]azepine moiety) was arbitrarily selected as a reference compound to calculate such a descriptor. The resulting similarity chart is presented in Figure 5. After such an analysis, some clusters were revealed with FragFp values, according to the heatmap based on the structure similarity index, between 0 (red = very different) and 1 (green = very similar). Such a scale indicated that our custom-made library gathered compounds related to 53 different FragFp-derived main clusters, which comprise small subsets (covering 6–23 compounds) of very structurally similar compounds depending on compound origin (e.g., the synthetic

approach or natural source used in the respective study). Hence, several attempts to study/model a statistically validated structure–activity relationship were unsuccessful. However, this structure similarity analysis clustered the test compounds into several classes with FragFp ≥ 0.4 , indicating that the library involves particular scaffolds that deserve a more robust analysis.

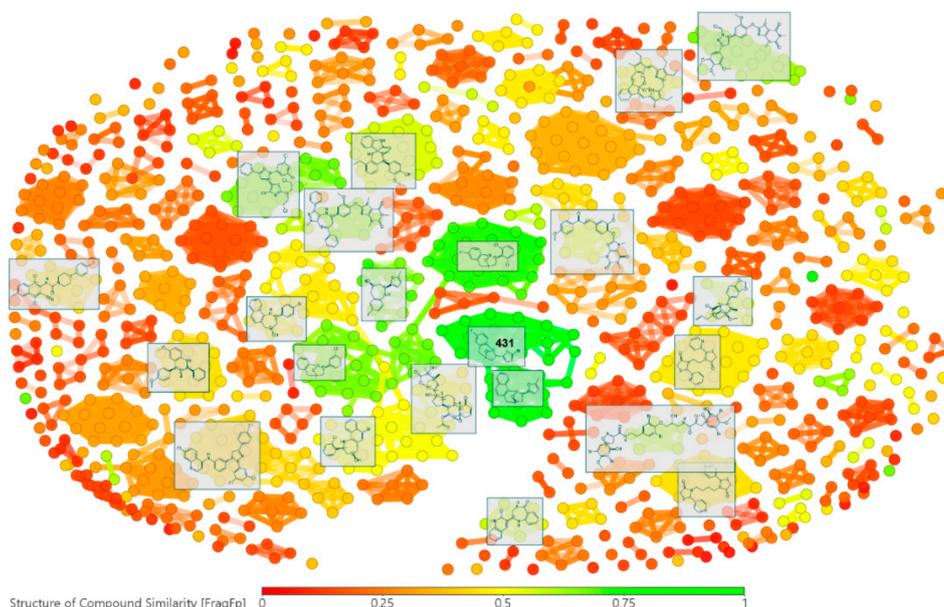


Figure 5. Similarity chart of the custom-made library. This plot was obtained after structure similarity analysis using the substructure fragment dictionary-based binary fingerprint descriptor (FragFp) [79]. The library comprised 836 compounds retrieved from the systematic review-derived information about Colombian research on antileishmanials. Structures of some compounds (enclosed in boxes) for those clusters possessing FragFP ≥ 0.4 were arbitrarily selected to illustrate the compound subset, according to the detailed information presented in Supplementary Table S2. Compound **431** was randomly designated as the reference compound to calculate the FragFp descriptor.

Regarding the antileishmanial activity of retrieved compounds (Table 2), Figure 6A shows the general distribution of activity values for pure compounds expressed as negative decadic logarithm of the half-maximal effective concentration in mol/L (pEC_{50}). As can be seen, despite the relatively large number of publications included, most of the tested compounds from these Colombian campaigns are rather poorly active. Only a limited fraction (18%) of the compounds displays EC_{50} values in the range 1–10 μ M ($5 < pEC_{50} < 6$), whereas the number of substances with activity in the sub-micromolar range ($pEC_{50} \geq 6$) is below 5%. This observation let us infer that the efforts made so far to find antileishmanial agents in Colombia should be considerably strengthened, if actual positive results are wanted. Otherwise, the studies will remain belonging to exploratory-oriented basic science rather than needed/applied medicinal and natural product chemistry projects.

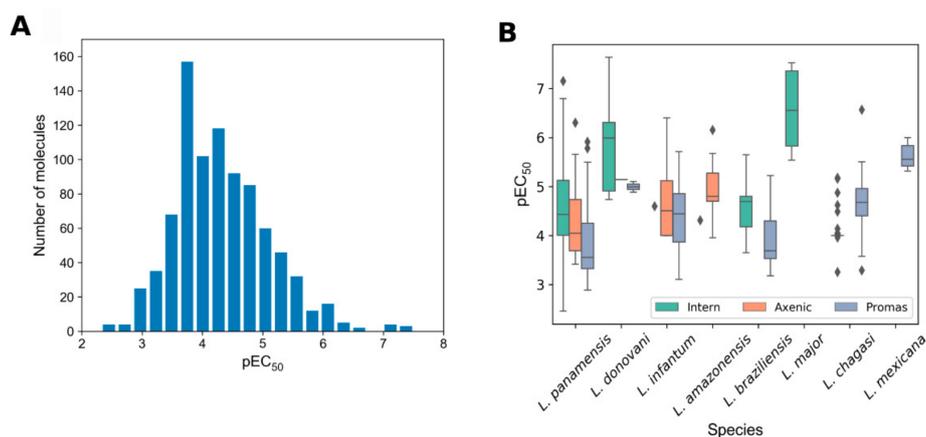


Figure 6. Antileishmanial activity of pure compounds expressed as pEC_{50} . **(A)** Number of compounds associated with activity ranges and **(B)** distribution of activity according to parasite species and form.

As mentioned before, several *Leishmania* species have been included within the considered reports. Analysis of the activity of the studied compounds by *Leishmania* species (Figure 6B) reinforces the previously discussed necessity of further and more efficiently driven projects (as observed, most of the boxes appeared in a range of pEC_{50} between 4 and 6). Interestingly, there are some compounds with remarkable activity against intracellular amastigotes of *L. major*. Similarly, some compounds have displayed exceptionally high activity against intracellular amastigotes of *L. donovani* and *L. panamensis* (mainly found within the whiskers zone of the corresponding distribution). Following the well-recognized criteria for hit compounds for infectious diseases (including VL) [80], solely those compounds with EC_{50} below 10 μ M could be considered as hits. Nonetheless, lack of information regarding structure–activity relationships, tractability of the chemotype, conformity with the rule of five and selectivity (>10-fold) may be considered as the major issue to define them as truly hit compounds.

2.3.1. Chemical Space of Selected Compounds

The elevated number of papers determining activity against *L. panamensis* might be seen as directly related to the high incidence of reported cases of infection by this species in Colombia, as mentioned earlier. Thus, we decided to focus our attention on the compounds described therein, looking for possibly relevant structural information that could be used as a first-line tool for further investigations. The dataset was therefore filtered, keeping only entries for pure compounds that were tested against *L. panamensis*. More specifically, and in order to reach some degree of comparability, exclusively those entries with activity determinations on intracellular amastigotes were used in the subsequent analyses. This form of the parasite/model was preferred over axenic amastigotes and promastigotes within the publications (see above), supporting our decision for deeper examination. A total of 484 compounds were considered for further studies, comprising 9.1% natural products and 90.9% synthetic compounds.

Structural information of the resulting group of compounds was extracted by using the Molecular ACCess System (MACCS) keys (167 bits) [81] and the Morgan fingerprints (radius = 2, 1024 bits) [82] implemented in RDKit [83]. These two sets of general molecular fingerprints were used aiming at a general glance of the corresponding chemical space. Owing to the typically high performance of t-distributed Stochastic Neighbor Embedding (t-SNE) as a dimensionality reduction method [84], we decided to look for possible compound clustering using such an algorithm and the fingerprints as independent inputs (Figure 7A). Although not directly comparable, both fingerprints offered some degree of clustering after t-SNE. Principal component analysis (PCA) was not able to represent clusters of similar compounds in a simple representation (data not shown). The partial generation of clusters of compounds by t-SNE would indicate actual structural relations among them, at least to some extent, as described by each fingerprint, i.e., some structural features seem to appear reiteratively within some groups of compounds. However, the limited clustering (high dispersion) also indicates a

quite significant structural diversity within the whole group of compounds (wide chemical space). Hierarchical clustering analysis (HCA) proved to be consistent with the spatial distribution provided by t-SNE (see Supplementary Figure S2).

After exhaustive analysis of the clusters obtained by HCA and extraction of the corresponding maximum common substructure (MCS), only three of the clusters were completely coincident between fingerprints, whereas a fourth was partly in agreement (the MCS from the cluster using Morgan fingerprint was part of a larger substructure found when using MACCS keys). This result highlights the well-known differences among fingerprints, which would translate to changes in outcomes coming from direct comparisons of fingerprints. The four common clusters are depicted in Figure 7A by different colors.

Representative compounds from each cluster are also included in Figure 7B, whose MCSs are highlighted. As expected, the relative location of each common cluster in the scatter plots is different. However, it is noteworthy to mention that three of them are quite well separated from the others, suggesting very particular features compared to the rest of the compounds. Interestingly, the compounds in the cluster in red were not particularly separated from other compounds compared to those previously mentioned. The seemingly marked lack of resolution of this particular cluster (especially when using Morgan fingerprints) might be due to high structural diversity of its compounds, whereupon the fingerprint features would be rather strongly shared (overlapped), ending up with many common bits with other clusters.

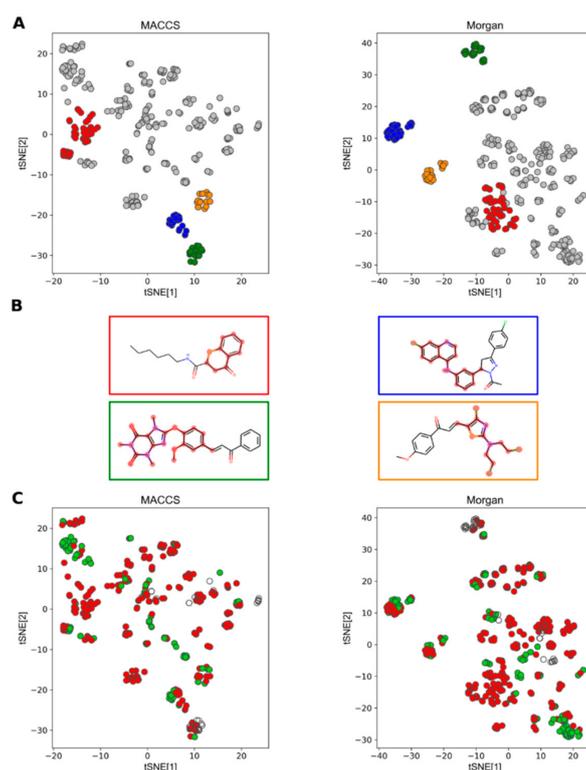


Figure 7. Representation of the chemical space of compounds tested against intracellular amastigotes of *L. panamensis*. (A) t-distributed Stochastic Neighbor Embedding (t-SNE) plot showing four common clusters to Molecular ACCess System (MACCS) (left) and Morgan (right) fingerprints, highlighted by red, blue, orange and dark green dots. Light gray dots represent the rest of the compounds. (B) Chemical structures of representative compounds of each selected cluster (enclosed in boxes colored according to the previous plot) with maximum common substructure (MCS) highlighted by pink contours. (C) t-SNE plot with dots colored by activity. Green dots: active compounds ($pEC_{50} \geq 5$), red dots: intermediate and poorly active compounds ($pEC_{50} < 5$), empty dots: compounds without EC_{50} determination.

Looking for insights into the possible effect of structural diversity on the antileishmanial activity, the t-SNE-derived scatter plots were colored by activity threshold (actives in green, $pEC_{50} \geq 5.0$; Figure 7C). It is evident that the most active compounds are not concentrated in any specific cluster, i.e., none of the scaffolds so far analyzed in Colombian studies are clearly favored over the others. Particularly, the cluster in red (Figure 7A) is mainly constituted by poorly active compounds (red in Figure 7C), while it is difficult to establish the potential of compounds in the cluster in green (Figure 7A) owing to the lack of EC_{50} determinations for some of them (empty circles in Figure 7C).

2.3.2. Machine Learning

After filtering off entries whose EC_{50} determinations were not available (e.g., only biological determination of percentage of inhibition at specific compound concentrations reported), a final set of 428 compounds was obtained. Owing to the inherent structural similarities among some compounds but also the huge differences in other cases (as shown above), and to the restricted capacity of the linear algorithms to provide reliable models (as mentioned before), machine learning was selected as a tool to analyze this dataset. Two different extensively used supervised learning algorithms were chosen to accomplish the task: random forest (RF) and support vector machines (SVM). Both MACCS and Morgan fingerprints were independently used for building the models. Preliminary evaluation of the classification variants of the selected algorithms showed decent performance (data not shown) and encouraged us to use actual activity values rather than an arbitrarily defined categorical dependent variable. Having decided for regression models, a coarse-to-fine scheme was followed for the optimization of the corresponding hyperparameters. In case of RF models, the number of trees in the forest, the minimum number of samples required to be at a leaf node, the minimum number of samples required to split an internal node, the maximum number of features to consider for the best split and the number of samples to draw from the training set during bootstrap were considered for optimization. The dataset was randomly split into training and test set (80:20%), ensuring maximum coverage of activity range for the latter. For both fingerprints, models offered maximum performance using 1 sample as a minimum to be at a leaf node and the total number of features as maximum. Those models were named M1 and M2, for MACCS and Morgan, respectively. While M1 used 306 trees, 7 samples to split a node and 75% of samples drawn during bootstrap, M2 required 127 trees, 2 samples and 94% of the samples, respectively.

In case of SVM models (M3 and M4 for MACCS and Morgan, respectively), the optimization was performed considering variations in the kernel functions, the kernel coefficient, the epsilon-tube and the regularization parameter C . The optimized models made use of the Radial Basis Function (RBF) kernel. The best performance for M3 was achieved with $\epsilon = 0.1$, $C = 2.5$ and $\gamma = 0.04$. M4 performed better when it used the combination of hyperparameters $\epsilon = 0.08$, $C = 2.8$ and $\gamma = 0.05$.

All the models were trained and tested for predictivity using ten repetitions. Table 3 shows the corresponding validation parameters as a mean of the ten runs. As can be seen, the generated models offered barely acceptable cross-validation (CV) scores, with limited prediction power. Nonetheless, M1–M4 outperformed classical linear models. The limited robustness for M1–M4 was not less than expected coming from such a diverse dataset. It is impossible to properly ensure comparability of activity data due to presumable changes in the specific procedures, despite using the same parasite forms/models (i.e., not all the compounds were experimentally tested at the same time and under the same exact conditions or not even in the same laboratory). Moreover, we did not take into account the implicit data error (which is sometimes not adequately informed, either), whose impact on computational modeling was already highlighted long ago [85]. It must be noted however that the data error is still not included in most of the Quantitative Structure-Activity Relationships/Quantitative Structure-Property Relationships (QSAR/QSPR) studies published in scientific journals. Regardless, both algorithms provided comparable results in terms of internal and external validation (Table 3).

Table 3. Statistical performance of machine learning models.

Validation Parameter ^a	M1 ^{b,d}	M2 ^{b,e}	M3 ^{c,d}	M4 ^{c,e}
R ² _{train}	0.813	0.925	0.847	0.849
MAE _{train}	0.250	0.155	0.202	0.183
R ² _{CV}	0.621	0.621	0.600	0.592
MAE _{CV}	0.359	0.354	0.370	0.358
R ² _{test}	0.670	0.666	0.583	0.689
MAE _{test}	0.322	0.324	0.352	0.339

^a R²: coefficient of determination, MAE: mean absolute error, CV: 10-fold cross-validation. ^b RF = Random Forest.

^c SVM = Support Vector Machines. ^d MACCS = Molecular ACCess System. ^e Morgan.

Although RF using Morgan (M2) fingerprints displayed significantly higher R² than that from MACCS (M1) during training, both internal and external validation coefficients were indistinguishable between models. In the case of SVM, the use of Morgan fingerprints (M4) demonstrated better predictability of the external set of compounds, albeit comparably low performance during CV. Beyond the isolated statistical values, and in spite of their closely related performance, M2 afforded the lowest deviations in predicted antileishmanial activity, represented by the lowest dispersion of data points around the regression line between experimental and predicted values (Supplementary Figure S3; all the corresponding activity predictions are included in Supplementary Table S3).

With the limited but still acceptable capacity of the obtained models, we were interested in deciphering the governing structure–activity relationships behind them. Although the machine learning algorithms are typically known for their black box nature, recent advances have been made in order to extract information regarding feature importance, like the use of the SHAP (SHapley Additive exPlanations) theory and the derived Shapley values [86,87]. Taken from game modeling, the SHAP theory helps to explain the contribution each single feature has on the outcome obtained. Implementation of this theory to gain detailed information from machine learning models has already been shown for drug discovery projects [88,89], making it possible to define the most important fingerprint bits contributing to the variance in activity. We applied the SHAP theory to the optimized models. In the case of RF models (M2 and M4), the Gini importance was also analyzed. The results are shown in Figure 8. There was an overall agreement between Gini and SHAP values for both M1 and M2, e.g., features 99 and 125 were consistently the top two in M1 (Figure 8A,C), whereas for M2 features, 1 and 259 appeared the most relevant by both methods (Figure 8B,D). High correlation between Gini and SHAP values have already been observed and reported [88]. Interestingly, the observed profound effect of feature 99 on the prediction of activity also prevailed in M3 (Figure 8E), suggesting some similarity between algorithms. The SHAP values also allowed inferring a significant effect of features 125 and 95 on M3 predictions, which were within the top ten features affecting M1 as well. This marked coincidence of features would indicate that both algorithms were able to identify basically the same structural features (held by the MACCS fingerprints) as responsible for the variance in activity. The SHAP values were also in agreement with the Gini importance values for M2, e.g., features 1, 259, 352 and 547 were ranked as the top four in both cases (Figure 8B,D). However, analysis of the corresponding SHAP values for M4 revealed a completely different distribution of features affecting the outcome of the model (Figure 8F). Only features 352 and 547 remained as part of the top ten, although with less importance. Being affected by several features at comparable costs implied that there was not any specific feature with a clear strong impact on M4 predictions, as it was observed above for the other models.

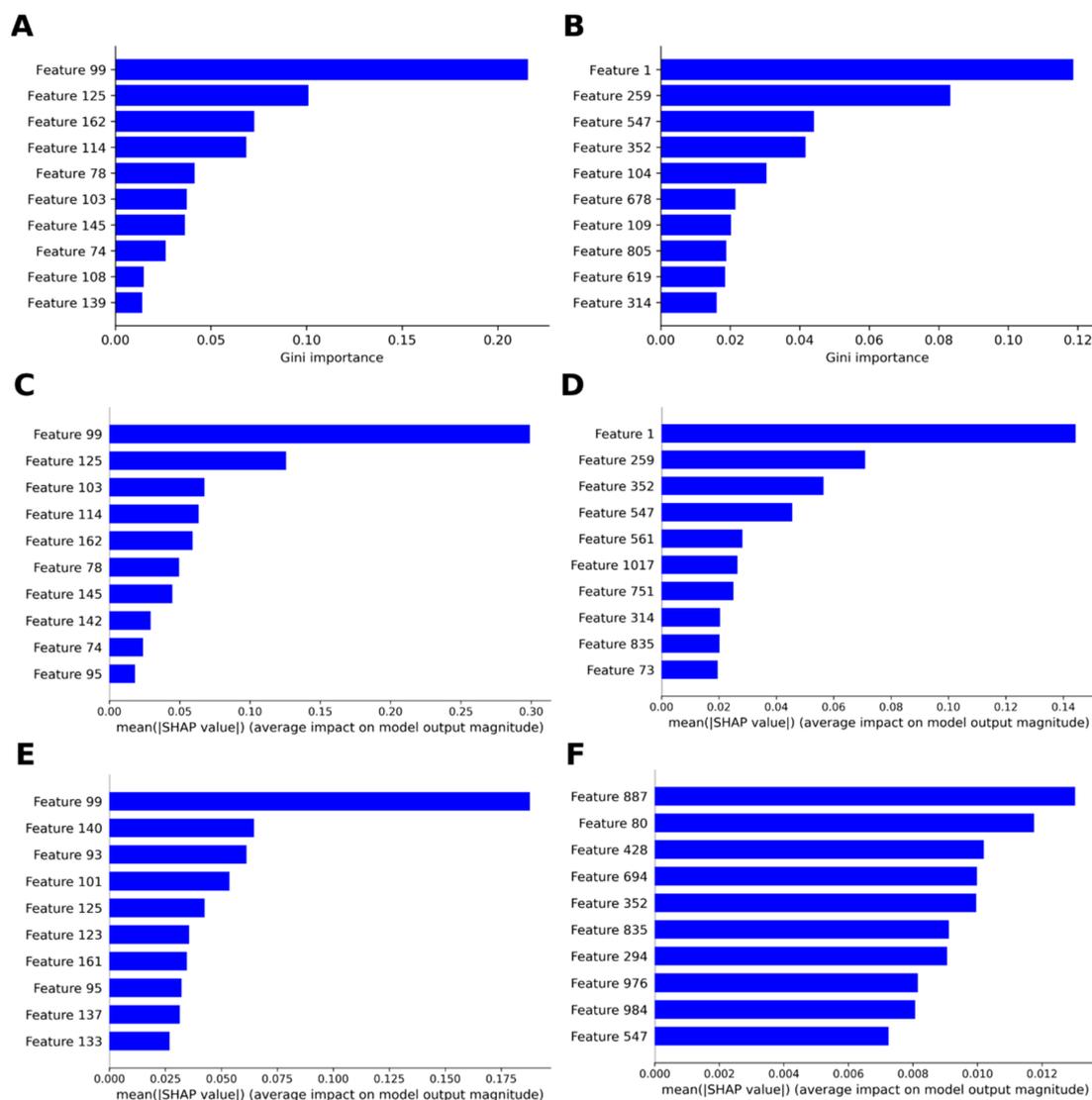


Figure 8. Relevant structural features for machine learning models. (A) Gini importance for M1, (B) Gini importance for M2, (C) SHapley Additive exPlanations (SHAP) values for M1, (D) SHAP values for M2, (E) SHAP values for M3 and (F) SHAP values for M4. Only the ten top features are shown.

Detailed analysis of the definition of the MACCS keys with higher SHAP values revealed that both M1 and M3 relied on similar structural patterns overall. Features 99 and 125, found in both cases, and 162 and 101, being exclusive for each model respectively, are related to the presence of C = C and aromatic rings. For M1, feature 114, which represents the presence of ethyl units bound to any atom, was also important, while the presence of methyl groups bound to heteroatoms was relevant for M3 (feature 93). Particularly interesting was the fact that M3 predictions were affected by the number of oxygen atoms in the molecule (feature 140 for O > 3). In contrast, the presence of chlorine atoms (feature 103) was important for M1.

Analyzing the individual contributions of each feature to the general outcome for M2 showed that the presence of feature 1 in the compounds was deleterious for the activity (Figure 9A). A similar general result was observed for features 352 and 751, although at considerably lesser extent. In contrast, the presence of feature 259 significantly favored the predicted activity values. Features 547, 561 and 1017 are other examples of features positively contributing to the activity. A comparable analysis in case of M4 was not straightforward due to the already mentioned high number of features responsible

for the activity. However, absence of features 352 and 984 seemed beneficial for the activity, whereas features like 887 and 835 appeared to play a positive role (Figure 9B).

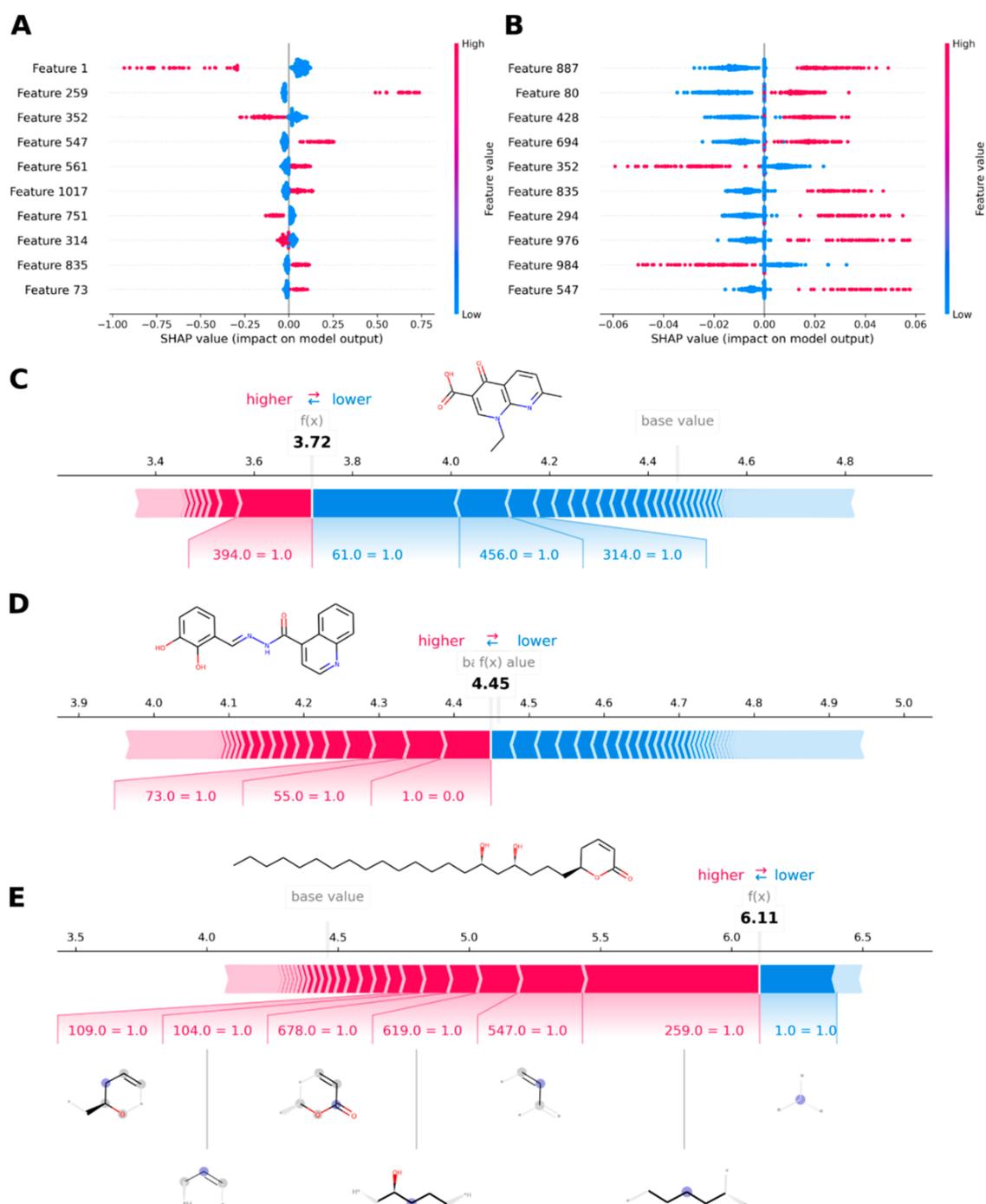


Figure 9. Contribution of structural features to M2 and M4. (A) SHAP values for M2 and (B) SHAP values for M4, colored by feature value (red: presence, blue: absence). (C) Force plot for a representative poorly active compound (190), (D) force plot for a representative compound with intermediate activity (586) and (E) force plot for a representative compound with high activity (164) as predicted by M2. Plot (E) includes the most important fingerprint bits affecting the prediction.

A more comprehensive analysis of the underlying structure–activity relationships for the compounds in the present dataset is not practically achievable because of the strong structural differences among compounds. Nevertheless, several additional insights could be retrieved from in-depth exploration of the individual Shapley values. Thus, taking advantage of the likelihood of drawing Morgan fingerprints offered by RDKit, representative compounds with low (compound **190**), intermediate (compound **586**) and high (compound **164**) antileishmanial activity were further studied. Model M2 was chosen for this analysis based on its apparently low deviation in predictions and clearly outlined important features. Figure 8C–E shows the corresponding force plots for those compounds. It can be observed how the activity of the inactive compound (**190**, Figure 9C) is strengthened by the presence of feature 394, while features 61, 456 and 314 could be responsible for the low value as they are pushing it down. Surprisingly, only the latter feature is part of the top ten features affecting the general outcome of the model. On the other hand, the activity of compound **586** (Figure 9D) was apparently caused by the presence of features 73 and 55. Moreover, absence of feature 1 significantly contributed to the activity of this specific compound, too, being the most important feature for M2, as previously noted. In the case of the active compound (**164**, Figure 9E), the activity was dominated by the presence of several features, including 109, 104, 678, 619, 547 and 259. To make predictions, the model predominantly used most of those features. In addition, the presence of feature 1 in this compound decreased the predicted value, as expected from the general trend observed. Features 109, 547 and 678 correspond to the 5,6-dihydro-2H-pyran-2-one moiety (Figure 9E). Meanwhile, features 259 and 619 are related to the aliphatic chains in vicinity of the hydroxyl groups. Particularly, feature 1 in this compound structure refers to the hydroxylated chiral carbons. Presumably, M2 might have learned some effects on activity due to chirality of those stereocenters.

2.3.3. Drug-Likeness Filtering

In order to get an idea of some interesting scaffolds to be considered in future investigations, further analyses were carried out on the group of active compounds. As a first step, their drug-likeness was partially assessed checking for the presence of undesirable moieties according to the filters for Pan-Assay INterference compounds (PAINs) implemented in the FAFDrug4 web server [90]. Despite that 85% of the active compounds passed the three filters available in the server, more than 60% of them might still be considered as potentially reactive substances containing groups susceptible to covalent binding (e.g., 23% of the active compounds contain Michael acceptor groups). Only a reduced set of twenty-eight compounds with confirmed activity on intracellular amastigotes passed the mentioned filters. However, more than half of these compounds (57%) showed compliance with the rule of five (Ro5) as well (one violation of the Ro5 was mainly found for the rest).

On the other hand, selectivity index (SI), defined as the ratio between cytotoxicity and antileishmanial activity, was considered for the last filtering step. This process revealed that whereas sixteen compounds (57%) showed $SI > 2$, only three of them (11%) displayed actual interesting selectivity values as to be considered for further development (Table 4). Figure 10 shows some of the best candidates after the aforementioned filtering.

Interestingly, a similarity search in SciFinder revealed that the scaffolds comprised by the above-mentioned compounds (Figure 10) have been rather uniquely considered as antileishmanial agents in Colombian research projects. Thus, in spite of the somewhat common nature of most of those scaffolds, their specific combinations have not yet become part of other studies focused on antileishmanial substances. No additional reports were found for the combination of chloroquine and pyrazole scaffolds nor for the combination of indolinone and tetrahydroquinoline scaffolds (as in compounds **511** and **465**, respectively). Similarly, no further studies on leishmanicidal sulfonylhydrazides of beyerene- or stevioside-like diterpenes (e.g., **84**) have been published. On the other hand, compound **191** represents a group of substances (styrylquinolines) quite commonly included in medicinal chemistry projects. Nonetheless, studies on their potential as antitrypanosomatid agents has been limited as well. To the best of our knowledge, there is only one recent study focused

on the leishmanicidal properties of a group of related compounds (4-aminostyrylquinolines) [91]. Decent activity against amastigotes of *L. pifanoi* and moderate selectivity indexes were therein reported. Additionally, assessment of the antileishmanial potential of alkenylquinolines was previously reported [92]. In this case, rather poorly active compounds were evinced, limiting the possible identification of interesting candidates. Seemingly, most of those compounds showed better antitrypanosomal activity.

Table 4. Selected candidates after Pan-Assay INterference compounds (PAINs) filtering.

Compound	Species	EC ₅₀ ^a	Cellular Line	IC ₅₀ ^b	SI ^c	Ro5 ^d Violations
3	<i>L. panamensis</i>	4.03	U-937	15.6	3.9	0
84	<i>L. braziliensis</i>	2.26	U-937	6.99	3.1	1
85	<i>L. braziliensis</i>	2.53	U-937	8.75	3.5	2
191	<i>L. panamensis</i>	0.57	U-937	8.02	14	0
192	<i>L. panamensis</i>	7.48	U-937	17.7	2.4	0
341	<i>L. panamensis</i>	4.81	U-937	12.7	2.6	0
343	<i>L. panamensis</i>	6.18	U-937	21.3	3.4	0
345	<i>L. panamensis</i>	5.90	U-937	15.6	2.6	0
465	<i>L. braziliensis</i>	3.30	BMDM	724	91	1
487	<i>L. panamensis</i>	7.07	U-937	19.7	2.8	2
489	<i>L. panamensis</i>	3.70	U-937	9.62	2.6	2
490	<i>L. panamensis</i>	3.41	U-937	9.69	2.8	1
499	<i>L. panamensis</i>	6.50	U-937	16.7	2.6	2
511	<i>L. panamensis</i>	4.47	U-937	322	72	2
566	<i>L. panamensis</i>	5.51	U-937	13.1	2.4	1
804	<i>L. panamensis</i>	0.60	U-937	3.87	3.9	2

^a EC₅₀ = half-maximal effective concentrations (expressed in μM) determined in intracellular amastigotes of the respective *Leishmania* species; ^b IC₅₀ = half-maximal inhibitory concentrations (expressed in μM) determined in human monocytes (U-937) or bone marrow-derived macrophages (BMDM) as listed in the respective cell line; ^c SI = selectivity index; ^d Ro5 = rule of five.

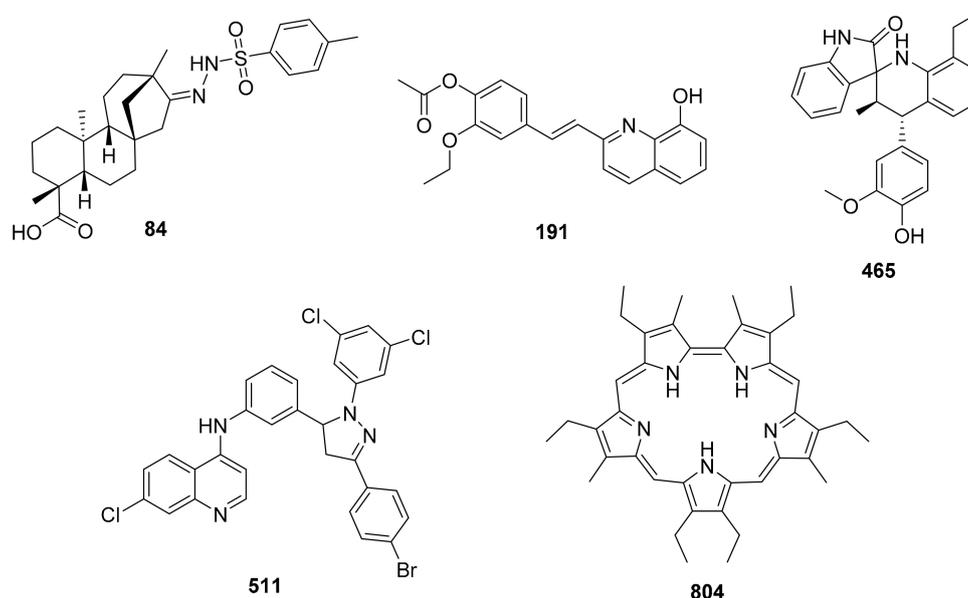


Figure 10. Chemical structures of the best antileishmanial candidates from Colombian campaigns.

3. Methods

3.1. Systematic Review

3.1.1. Search and Eligibility Criteria

The search was carried out in Scopus, Web of Science, PubMed and Scielo databases (last search on 6 July 2020) using an appropriated search equation for each database with the keywords and Boolean operators as follows: leishmani* OR antileishmani* OR leishmanicid*. The search results were then refined according to the filter tools available in each database to select the documents affiliated to Colombia, excepting the Scielo database. In this database, the search equation was “(leishmani* OR antileishmani* OR leishmanicid*) AND (colombia)”, as this database does not provide a filter option by country affiliation. After data retrieval from databases, the inclusion criteria were defined for those articles containing the following characteristics/information: (1) original articles, (2) in vitro or in vivo antileishmanial activity, (3) assays with synthetic compounds, (4) assays with pure isolated compounds of natural origin and (5) tested crude extracts. Retrieved studies were excluded if they involved only known/recognized antileishmanial drugs, or if accessibility to full-text versions was not accomplished.

3.1.2. Study Selection

The selection of studies was made in two phases [93]. First, the search results were uploaded to the Rayyan web application [94]. Two reviewers independently screened the titles according to the inclusion criteria. Articles marked as “included” by the two reviewers were promptly selected for the next phase. In the cases of unmatched marks (i.e., articles that were marked as “included,” “excluded” or “maybe” by only one reviewer), the papers were analyzed by the two reviewers; if the disagreement persisted, the final decision was made by the third author. Then, the full-text version of the initially filtered articles was read and selected applying the inclusion/exclusion criteria.

3.1.3. Data Collection

A preliminary data collection form was built according to the review goals, and its suitability was evaluated in a pilot procedure with ten randomly selected papers. Then, the final version of the data collection form was used for the survey/examination of each paper that passed the title-screening phase. The three authors cured the resulting spreadsheet. Since the analysis and characterization of the reported antileishmanial compounds are one of the main objectives of this review, a chemoinformatics approach to the chemical space represented by the compiled compound library was further accomplished. This analysis was based on fingerprints using conventional machine learning algorithms and clustering methods.

3.2. Chemoinformatics Analysis

3.2.1. Data Preparation

After compilation of the systematic review-derived antileishmanial records, the structures of retrieved compounds were individually sketched in MarvinSketch (ChemAxon, Budapest, Hungary) and converted into SMILES as a line notation for chemical structure. This notation uses the American Standard Code for Information Interchange (ASCII) character encoding. Once the custom-made library was completed, a structure filtering analysis was firstly performed using the substructure fragment dictionary-based binary fingerprint descriptor (FragFp), incorporated in DataWarrior v5.0.0 [95]. A structure comparison between compound sets can be achieved with this descriptor (analogous to Molecular Design Limited (MDL) keys), as it considers structural moieties through 512 predefined fragments into a dictionary [79]. Each fragment is contained into one bit of the FragFp descriptor; therefore, a bit is defined as 1 if a respective fragment occurred in the structure at least one time. Thus, a list of all dictionary fragments (as part of the substructure query) is generated and the overall,

comparison led to the substructure filtering. This filtering is visualized through a similarity chart (e.g., a scatter plot) according to the FragFp index.

Compounds' structures were additionally characterized by MACCS keys [81] and Morgan fingerprints [82] as selected molecular representations. MACCS keys consist of 166 bits accounting for either absence or presence of specific structural patterns. Morgan fingerprints encode for structural features on a radial basis, i.e., having a circular shape, and they account for structural fragments. In this work, a radius of 2 and length of 1024 bits were chosen. Both sets of fingerprints were calculated using RDKit 2020.03.1 [83], where Morgan fingerprints are defined as a modification of the extended connectivity fingerprints (ECFP) [96]. Activity data were transformed into the corresponding pEC_{50} (negative decadic logarithm of the EC_{50} in mol/L). Finally, the PAINs and Ro5-based filtering was accomplished using the FAF-Drugs4 web server [90].

3.2.2. Chemical Space by t-SNE

t-distributed Stochastic Neighbor Embedding (t-SNE) is an unsupervised, non-linear technique that allows the visualization of high-dimensional data [84]. It works in three steps: first, similarities among samples in the high-dimensional space are defined, by measuring the corresponding probabilities using Gaussian distributions; secondly, similarities among samples in the low-dimensional space (typically two-dimensional (2D)) are calculated, but in this case using Student's t-distribution with one degree of freedom (known as Cauchy distribution) instead. Finally, and in order to genuinely recreate the high-dimensional distribution, optimization of the distributions is conducted by gradient descent using Kullback–Liebler divergence as the loss function. The perplexity and maximum number of iterations were manually tuned. t-SNE implementation in Scikit-learn [97] was used in the present work.

3.2.3. Random Forest

As a supervised machine learning algorithm, random forest (RF) makes use of groups of decision trees, and is therefore part of the ensemble learning methods [98,99]. Each tree is trained from a bootstrapped sample of data, where typically a random subset of features is considered for node splitting. The final predictions correspond to the average of all the predictions made from those individual learners. Node splitting is controlled by reduction of the Gini index (Gini "impurity"). The sum of the reduction in Gini impurity is termed as Gini importance [100]. The number of trees in the forest (5–1000), the minimum number of samples required to be at a leaf node (1–10), the minimum number of samples required to split an internal node (2–16), the maximum number of features to consider for the best split (total number of features, base 2 log of the total, and squared root of the total) and the number of samples to draw from the training set during bootstrap (5–95%) were subjected to optimization in this work. RF regression models were built using Scikit-learn.

3.2.4. Support Vector Machines

Support Vector Machines (SVM) is another supervised machine learning algorithm, whose principle is to define hyperplanes for effective segregation of the data typically into classes of objects [101]. Those data points closest to the hyperplanes are called support vectors. The best hyperplanes are selected by minimization of the margin (gap between the support vectors delimiting it). Implementation of SVM usually requires so-called kernel functions that help finding the hyperplanes by increasing the dimensionality of the data (transformation from lower to higher dimensional space). The regularization parameter C trades off misclassification error and decision boundary (margin size). The kernel function (linear, polynomial, sigmoid and radial basis function (RBF)), the kernel coefficient gamma (1×10^{-6} to 1), the epsilon-tube (0.1–0.5) and the regularization parameter C (1–100) were optimized during the present work. SVM in regression models (SVR) were built with Scikit-learn.

3.2.5. Hyperparameter Optimization

Both RF and SVM models were submitted to hyperparameter optimization in a coarse-to-fine approach. The process was carried out in two instances. The first one consisted of random sampling of the corresponding hyperparameter grid. The best combination of hyperparameters was selected based on the coefficient of determination obtained during a 5-fold cross-validation (CV) scheme. Afterwards, an exhaustive evaluation of hyperparameter combinations in the proximity of the best set detected in the previous step was performed. The same scoring function was used in the last step. The whole process was achieved applying the corresponding Scikit-learn implementations.

3.2.6. Final Models

Final RF and SVM regression models were built using the corresponding optimized hyperparameter sets. Each model was fitted ten times and predictions were obtained accordingly. Performance of the models was evaluated by the coefficient of determination (R^2) and the mean absolute error (MAE) calculated for the respective activity predictions, and expressed as an average. 10-fold CV assessed internal validity of the models.

3.2.7. Analysis of Contributions by SHAP Values

The concept of Shapley values was developed early in cooperative game theory [102], where the calculation of the contribution of each single player to the global outcome is highly important. Thus, properly rewarding each player, in order to provide a unique result prediction, is what the Shapley values represent. This theory was recently extended aiming at a measure of feature importance in different predictive models [86]. The introduced SHAP (SHapley Additive exPlanation) values help explain how the predicted output changes according to the appearance of any feature. Their application for machine learning models' explanations has been proven [87], including their outstanding potential for machine learning-based drug discovery projects [88,89]. Computation of SHAP values was carried out using an open implementation under Python [87,103].

4. Conclusions

Although Colombia is one of the countries with more pathogenic *Leishmania* species in circulation, scientific research was found focused on *L. panamensis*. Since other species such as *L. braziliensis* and *L. guyanensis* are becoming significant etiological agents (associated with an important number of CL cases), these *Leishmania* species should be included in the forthcoming research programs on antileishmanial discovery. Furthermore, our findings highlight the need for involvement of more research centers, allowing strategic collaborations that can lead to more multidisciplinary approaches. Indeed, considering the CL burden in Colombia, along with the demand for more effective and safer chemotherapeutic options, it is essential that public investment (national and regional) maintain and even prolong/improve the effective financial support for the research on leishmaniasis control (particularly the development of antileishmanial agents). Pondering on studies involving the GFP-transfected *L. panamensis* strain, and the difficulties and challenges associated with the leishmanicidal screening assays, the development of more GFP-transfected species (e.g., *L. braziliensis*, *L. guyanensis*) would positively impact the antileishmanial-oriented studies. Regarding natural products, it was expected to find a higher number of studies. The internal regulations for accessing genetic sources has probably limited bioprospecting studies in this field. In this regard, this assumption should be exhaustively evaluated. In any case, we concluded that nature remains as an underexplored resource concerning its leishmanicidal potential and the limited number of studies show a biased attentiveness on plants.

Moreover, a holistic analysis of the activity data retrieved from our systematic literature revision exposed that more than 50% of them have low to non-existent activity against *Leishmania* parasites. Consequently, future investigations on small molecules targeting this disease should be guided by

medicinal chemistry principles using rational approaches. Additionally, the bottlenecks and gaps across the antileishmanial development pipelines can be overcome with public–private partnerships, combining knowledge from academia and infrastructure and financial support from pharmaceutical companies within an efficient and effective scientific and technical cooperation.

Profound cheminformatics analyses indicated apparently high chemical diversity within the group of compounds with measured activity against intracellular amastigotes of *L. panamensis* (the largest available group of compounds). Interestingly, this chemical space could be condensed within a relatively small number of compound clusters without any visibly privileged scaffold. Furthermore, classical machine learning algorithms facilitated uncovering some underlying structure–activity relationships, affording a set of models with decent capacity to predict antileishmanial activity. Combination of these models with SHAP theory could be a valuable tool in further research to be developed in Colombia, as foreseeing the possible activity and associating it with SAR information might serve as a simple but effective initial guidance.

Supplementary Materials: The following are available online, Figure S1: Distribution of the Colombian scientific literature on leishmaniasis, Figure S2: t-SNE plot using MACCS and Morgan fingerprints colored by HCA, Figure S3: Experimental versus predicted activity (pEC_{50}) for machine learning models, Table S1: PRISMA checklist, Table S2: List of compounds and their antileishmanial activity retrieved from the reviewed literature, Table S3: Antileishmanial activity predicted by machine learning models.

Author Contributions: J.S.-S. and E.C.-B.: Conceptualization. J.S.-S., F.A.B. and E.C.-B.: Investigation, Data curation, Writing—original draft preparation, Writing—review and editing and Visualization. J.S.-S.: PRISMA guidelines addressing and validation. E.C.-B.: Structural similarity analysis. F.A.B.: Clustering, machine learning modeling and Drug-likeness filtering. E.C.-B.: Resources, Supervision, and Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The present study is a product derived from the Project IMP-CIAS-2924 funded by Vicerrectoría de Investigaciones at Universidad Militar Nueva Granada (UMNG)—Validity Period 2019.

Acknowledgments: Authors thank UMNG for the financial support and ChemAxon (<https://www.chemaxon.com>) for the academic licensing of MarvinSketch and Standardizer (Product version 20.11.0). This study was performed as part of the activities of the Research Network Natural Products against Neglected Diseases (ResNet NPND), <https://resnetnpnd.org/>.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Pan American Health Organization. *Leishmaniasis: Epidemiological Report in the Americas*; Technical Report No. 8-2019; PAHO: Washington, DC, USA, 2019; pp. 1–10.
2. Pan American Health Organization. *Colombia: Cutaneous and Mucosal Leishmaniasis*; PAHO: Washington, DC, USA, 2019; Volume 14, p. e0224351.
3. Ramírez, J.D.; Hernández, C.; León, C.M.; Ayala, M.S.; Flórez, C.; González, C. Taxonomy, diversity, temporal and geographical distribution of Cutaneous Leishmaniasis in Colombia: A retrospective study. *Sci. Rep.* **2016**, *6*, 28266. [[CrossRef](#)] [[PubMed](#)]
4. Montalvo, A.M.; Fraga, J.; Montano, I.; Monzote, L.; van der Auwera, G.; Marín, M.; Muskus, C. Identificación molecular de aislamientos clínicos de *Leishmania* spp. procedentes de Colombia con base en el gen hsp70. *Biomédica* **2016**, *36*. [[CrossRef](#)] [[PubMed](#)]
5. Montalvo, A.M.; Fraga, J.; Tirado, D.; Blandón, G.; Alba, A.; van der Auwera, G.; Vélez, I.D.; Muskus, C. Detection and identification of *Leishmania* spp.: Application of two hsp70-based PCR-RFLP protocols to clinical samples from the New World. *Parasitol. Res.* **2017**, *116*, 1843–1848. [[CrossRef](#)] [[PubMed](#)]
6. Salgado-Almario, J.; Hernández, C.A.; Ovalle-Bracho, C. Geographical distribution of *Leishmania* species in Colombia, 1985–2017. *Biomédica* **2019**, *39*, 278–290. [[CrossRef](#)] [[PubMed](#)]
7. Pan American Health Organization. *Manual of Procedures for Surveillance and Control of Leishmaniasis in the Americas*; PAHO: Washington, DC, USA, 2019; ISBN 978-92-75-32063-1.

8. An, I.; Harman, M.; Esen, M.; Celik, H. The effect of pentavalent antimonial compounds used in the treatment of cutaneous leishmaniasis on hemogram and biochemical parameters. *Cutan. Ocul. Toxicol.* **2019**, *38*, 294–297. [[CrossRef](#)] [[PubMed](#)]
9. Lyra, M.R.; Passos, S.R.L.; Pimentel, M.I.F.; Bedoya-Pacheco, S.J.; Valete-Rosalino, C.M.; Vasconcellos, E.C.F.; Antonio, L.F.; Saheki, M.N.; Salgueiro, M.M.; Santos, G.P.L.; et al. Pancreatic toxicity as an adverse effect induced by meglumine antimoniate therapy in a clinical trial for cutaneous leishmaniasis. *Rev. Instit. Med. Trop. São Paulo* **2016**, *58*. [[CrossRef](#)]
10. Marques, S.A.; Merlotto, M.R.; Ramos, P.M.; Marques, S.A. American tegumentary leishmaniasis: Severe side effects of pentavalent antimonial in a patient with chronic renal failure. *An. Bras. Dermatol.* **2019**, *94*, 355–357. [[CrossRef](#)]
11. Brito, N.C.; Rabello, A.; Cota, G. Efficacy of pentavalent antimoniate intralesional infiltration therapy for cutaneous leishmaniasis: A systematic review. *PLoS ONE* **2017**, *12*, e0184777. [[CrossRef](#)]
12. Croft, S.L.; Sundar, S.; Fairlamb, A.H. Drug Resistance in Leishmaniasis. *Clin. Microbiol. Rev.* **2006**, *19*, 111–126. [[CrossRef](#)]
13. Ponte-Sucré, A.; Gamarro, F.; Dujardin, J.-C.; Barrett, M.P.; López-Vélez, R.; García-Hernández, R.; Pountain, A.W.; Mwenechanya, R.; Papadopoulou, B. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0006052. [[CrossRef](#)]
14. Perilla-Gonzalez, Y.; Gomez-Suta, D.; Osorio, N.D.; Hurtado-Hurtado, N.; Baquero-Rodriguez, J.D.; Lopez-Isaza, A.F.; Lagos-Grisales, G.J.; Villegas, S.; Rodríguez-Morales, A.J. Study of the scientific production on leishmaniasis in Latin America. *Recent Pat. Anti-Infect. Drug Discov.* **2014**, *9*, 216–222. [[CrossRef](#)] [[PubMed](#)]
15. Martinez-Mayorga, K.; Madariaga-Mazón, A.; Medina-Franco, J.L.M.; Maggiora, G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opin. Drug Discov.* **2020**, *15*, 293–306. [[CrossRef](#)] [[PubMed](#)]
16. Cavasotto, C.N.; Aucar, M.G.; Adler, N.S. Computational chemistry in drug lead discovery and design. *Int. J. Quantum Chem.* **2019**, *119*, e25678. [[CrossRef](#)]
17. Makhouri, F.R.; Ghasemi, J.B. Combating Diseases with Computational Strategies Used for Drug Design and Discovery. *Curr. Top. Med. Chem.* **2019**, *18*, 2743–2773. [[CrossRef](#)]
18. Gillet, V.J. Applications of Chemoinformatics in Drug Discovery. In *Biomolecular and Bioanalytical Techniques*; Wiley: Hoboken, NJ, USA, 2019; pp. 17–36.
19. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. [[CrossRef](#)]
20. Lo, Y.-C.; Rensi, S.E.; Torng, W.; Altman, R.B. Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* **2018**, *23*, 1538–1546. [[CrossRef](#)]
21. Réda, C.; Kaufmann, E.; Delahaye-Duriez, A. Machine learning applications in drug development. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 241–252. [[CrossRef](#)]
22. Halder, A.K.; Cordeiro, M.N.D.S. Advanced in Silico Methods for the Development of Anti- Leishmaniasis and Anti-Trypanosomiasis Agents. *Curr. Med. Chem.* **2020**, *27*, 697–718. [[CrossRef](#)]
23. Scotti, L.; Ishiki, H.M.; Júnior, F.M.; da Silva, M.S.; Scotti, M. Artificial Neural Network Methods Applied to Drug Discovery for Neglected Diseases. *Comb. Chem. High Throughput Screen.* **2015**, *18*, 819–829. [[CrossRef](#)]
24. Njogu, P.M.; Guantai, E.M.; Pavadai, E.; Chibale, K. Computer-Aided Drug Discovery Approaches against the Tropical Infectious Diseases Malaria, Tuberculosis, Trypanosomiasis, and Leishmaniasis. *ACS Infect. Dis.* **2016**, *2*, 8–31. [[CrossRef](#)]
25. Ferreira, L.G.; Andricopulo, A.D. Chemoinformatics Strategies for Leishmaniasis Drug Discovery. *Front. Pharmacol.* **2018**, *9*, 1278. [[CrossRef](#)] [[PubMed](#)]
26. Romano, J.D.; Tatonetti, N.P. Informatics and Computational Methods in Natural Product Drug Discovery: A Review and Perspectives. *Front. Genet.* **2019**, *10*, 368. [[CrossRef](#)] [[PubMed](#)]
27. Rodrigues, T. Harnessing the potential of natural products in drug discovery from a cheminformatics vantage point. *Org. Biomol. Chem.* **2017**, *15*, 9275–9282. [[CrossRef](#)] [[PubMed](#)]
28. Olğaç, A.; Orhan, I.E.; Banoglu, E. The potential role of in silico approaches to identify novel bioactive molecules from natural resources. *Futur. Med. Chem.* **2017**, *9*, 1663–1684. [[CrossRef](#)]

29. Pereira, F.; Aires-De-Sousa, J. Computational Methodologies in the Exploration of Marine Natural Product Leads. *Mar. Drugs* **2018**, *16*, 236. [CrossRef]
30. Herrera-Acevedo, C.; Scotti, L.; Alves, M.F.; Diniz, M.D.F.F.M.; Scotti, M.T. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. *Molecules* **2017**, *22*, 79. [CrossRef]
31. Scotti, L.; Ishiki, H.; Mendonca, F.; Silva, M.S.; Scotti, M. In-silico Analyses of Natural Products on *Leishmania* Enzyme Targets. *Mini-Rev. Med. Chem.* **2015**, *15*, 253–269. [CrossRef]
32. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]
33. Trouiller, P.; Olliaro, P.; Torreele, E.; Orbinski, J.; Laing, R.; Ford, N. Drug development for neglected diseases: A deficient market and a public-health policy failure. *Lancet* **2002**, *359*, 2188–2194. [CrossRef]
34. Murray, C.J.L.; Barber, R.M.; Foreman, K.J.; Ozgoren, A.A.; Abd-Allah, F.; Abera, S.F.; Aboyans, V.; Abraham, J.P.; Abubakar, I.; Abu-Raddad, L.J.; et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: Quantifying the epidemiological transition. *Lancet* **2015**, *386*, 2145–2191. [CrossRef]
35. Di Masi, J.A.; Grabowski, H.G.; Hansen, R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33. [CrossRef]
36. Light, D.W.; Lexchin, J.R. Pharmaceutical research and development: What do we get for all that money? *BMJ* **2012**, *345*, e4348. [CrossRef] [PubMed]
37. Surur, A.S.; Fekadu, A.; Makonnen, E.; Hailu, A. Challenges and Opportunities for Drug Discovery in Developing Countries: The Example of Cutaneous Leishmaniasis. *ACS Med. Chem. Lett.* **2020**, *11*, 2058–2062. [CrossRef] [PubMed]
38. Alcântara, L.M.; Ferreira, T.C.; Gadelha, F.R.; Miguel, D.C. Challenges in drug discovery targeting TriTryp diseases with an emphasis on leishmaniasis. *Int. J. Parasitol. Drugs Drug Resist.* **2018**, *8*, 430–439. [CrossRef] [PubMed]
39. Instituto Nacional de Salud Estadísticas de Vigilancia Rutinaria. Available online: http://portalsivigila.ins.gov.co/VigilanciaRutinaria/rutinaria_2019.xlsx (accessed on 29 October 2020).
40. Watts, K.R.; Tenney, K.; Crews, P. The structural diversity and promise of antiparasitic marine invertebrate-derived small molecules. *Curr. Opin. Biotechnol.* **2010**, *21*, 808–818. [CrossRef] [PubMed]
41. Abdelmohsen, U.R.; Balasubramanian, S.; Oelschlaeger, T.A.; Grkovic, T.; Pham, N.B.; Quinn, R.J.; Hentschel, U. Potential of marine natural products against drug-resistant fungal, viral, and parasitic infections. *Lancet Infect. Dis.* **2017**, *17*, e30–e41. [CrossRef]
42. Imperatore, C.; Gimmelli, R.; Persico, M.; Casertano, M.; Guidi, A.; Saccoccia, F.; Ruberti, G.; Luciano, P.; Aiello, A.; Parapini, S.; et al. Investigating the Antiparasitic Potential of the Marine Sesquiterpene Avarone, Its Reduced Form Avarol, and the Novel Semisynthetic Thiazinoquinone Analogue Thiazovarone. *Mar. Drugs* **2020**, *18*, 112. [CrossRef]
43. Oliveira, M.; Barreira, L.; Gangadhar, K.N.; Rodrigues, M.J.; Santos, T.; Varela, J.; Custódio, L. Natural products from marine invertebrates against *Leishmania* parasites: A comprehensive review. *Phytochem. Rev.* **2016**, *15*, 663–697. [CrossRef]
44. Yamthe, L.R.T.; Appiah-Opong, R.; Fokou, P.V.T.; Nolé, T.; Boyom, F.F.; Nyarko, A.K.; Wilson, M. Marine Algae as Source of Novel Antileishmanial Drugs: A Review. *Mar. Drugs* **2017**, *15*, 323. [CrossRef]
45. Álvarez-Bardón, M.; Pérez-Pertejo, Y.; Ordóñez, C.; Sepúlveda-Crespo, D.; Carballeira, N.M.; Tekwani, B.L.; Sankaranarayanan, M.; Martínez-Valladares, M.; García-Estrada, C.; Reguera, R.M.; et al. Screening Marine Natural Products for New Drug Leads against Trypanosomatids and Malaria. *Mar. Drugs* **2020**, *18*, 187. [CrossRef]
46. Alzate, F.; Jimenez, N.; Weniger, B.; Bastida, J.; Gimenez, A. Antiprotozoal Activity of Ethanol Extracts of Some Bomarea Species. *Pharm. Biol.* **2008**, *46*, 575–578. [CrossRef]
47. Weniger, B.; Robledo, S.; Arango, G.J.; Deharo, E.; Aragón, R.; Muñoz, V.; Callapa, J.; Lobstein, A.; Anton, R. Antiprotozoal activities of Colombian plants. *J. Ethnopharmacol.* **2001**, *78*, 193–200. [CrossRef]
48. Lopez, R.; Cuca, L.; Delgado, G. Antileishmanial and immunomodulatory activity of *Xylopiya discreta*. *Parasite Immunol.* **2009**, *31*, 623–630. [CrossRef] [PubMed]

49. Enciso, N.A.C.; Coy-Barrera, E.; Patiño, O.J.; Cuca, L.E.; Delgado, G. Evaluation of the Leishmanicidal Activity of Rutaceae and Lauraceae Ethanol Extracts on Golden Syrian Hamster (*Mesocricetus auratus*) Peritoneal Macrophages. *Indian J. Pharm. Sci.* **2014**, *76*, 188–197.
50. Neira, L.F.; Mantilla, J.C.; Stashenko, E.; Escobar, P. Toxicidad, genotoxicidad y actividad anti-Leishmania de aceites esenciales obtenidos de cuatro (4) quimiotipos del género *Lippia*. *Bol. Latinoam. Caribe Plantas Med. Aromat.* **2018**, *17*, 68–83.
51. Saez, J.; Granados, H.; Torres, B.; Velez, I.D.; Munoz, D. Leishmanicidal activity of *Annona* aff. *spraguei* seeds. *Fitoterapia* **1998**, *69*, 478–479.
52. Jaramillo, M.; Arango, G.; González, M.; Robledo, S.; Velez, I. Cytotoxicity and antileishmanial activity of *Annona muricata* pericarp. *Fitoterapia* **2000**, *71*, 183–186. [[CrossRef](#)]
53. Osorio, E.; Arango, G.J.; Jimenez, N.; Alzate, F.; Ruiz, G.; Gutiérrez, D.; Paco, M.A.; Giménez, A.; Robledo, S. Antiprotozoal and cytotoxic activities in vitro of Colombian Annonaceae. *J. Ethnopharmacol.* **2007**, *111*, 630–635. [[CrossRef](#)]
54. Rodríguez, A.M.; Camargo, J.R.; García, F.J.B. Actividad in vitro de la mezcla de alcaloides de *Ervatamia coronaria* (Jacq) Staff. Apocynaceae sobre amastigotes de *Leishmania braziliensis*. *Rev. Bras. Farm.* **2008**, *18*, 350–355. [[CrossRef](#)]
55. Arévalo, Y.; Robledo, S.; Muñoz, D.L.; Granados-Falla, D.; Cuca, L.E.; Delgado, G. Evaluación in vitro de la actividad de aceites esenciales de plantas colombianas sobre *Leishmania Braziliensis*. *Rev. Colomb. Cienc. Quim. Farm* **2009**, *38*, 131–141.
56. Céline, V.; Adriana, P.; Eric, D.; Joaquina, A.; Yannick, E.; Augusto, L.F.; Rosario, R.; Dionicia, G.; Michel, S.; Denis, C.; et al. Medicinal plants from the Yanesha (Peru): Evaluation of the leishmanicidal and antimalarial activity of selected extracts. *J. Ethnopharmacol.* **2009**, *123*, 413–422. [[CrossRef](#)]
57. Calderon, A.I.; Romero, L.I.; Ortega-Barría, E.; Solis, P.N.; Zacchino, S.; Gimenez, A.; Pinzón, R.; Cáceres, A.; Tamayo, G.; Guerra, C.; et al. Screening of Latin American plants for antiparasitic activities against malaria, Chagas disease, and leishmaniasis. *Pharm. Biol.* **2010**, *48*, 545–553. [[CrossRef](#)] [[PubMed](#)]
58. Martínez, W.; Ospina, L.F.; Granados, D.; Delgado, G. In vitro studies on the relationship between the anti-inflammatory activity of *Physalis peruviana* extracts and the phagocytic process. *Immunopharmacol. Immunotoxicol.* **2009**, *32*, 63–73. [[CrossRef](#)]
59. Sanchez-Suarez, J.; Riveros, I.; Delgado, G. Evaluation of the Leishmanicidal and Cytotoxic Potential of Essential Oils Derived from Ten Colombian Plants. *Iran. J. Parasitol.* **2013**, *8*, 129–136.
60. Cardona Galeano, C.W.; Robledo-Restrepo, C.S.M.; Rojano, C.B.A.; Alzate-Guarin, C.F.; Muñoz-Herrera, D.L.; Saez-Vega, C.J. Leishmanicidal and antioxidant activity of extracts of *Piper daniel-gonzalezii* trel. (piperaceae). *Rev. Cubana Plantas Med.* **2013**, *18*, 268–277.
61. Espitia-Baena, J.E.; Robledo-Restrepo, S.M.; Cuadrado-Cano, B.S.; Duran-Sandoval, H.R.; Gómez-Estrada, H.A. Perfil fitoquímico, actividad anti-Leishmania, hemolítica y toxicológica de *Cordia dentata* Poir. y *Heliotropium indicum* L. *Rev. Cubana Plantas Med.* **2014**, *19*, 208–224.
62. Mesa, L.E.; Vasquez, D.; Lutgen, P.; Restrepo, A.M.; Robledo, S.; Velez, I.D.; Ortiz, I. In vitro and in vivo antileishmanial activity of *Artemisia annua* L. leaf powder and its potential usefulness in the treatment of uncomplicated cutaneous leishmaniasis in humans. *Rev. Soc. Bras. Med. Trop.* **2017**, *50*, 52–60. [[CrossRef](#)]
63. Marin, F.J.; Torres, O.L.; Robledo, S.; Doria, M.E. Estudio Fitoquímico y Evaluación de la Actividad Antioxidante y Leishmanicida de la Especie *Pilocarpus alvaradoi* (Rutaceae). *Inform. Tecnol.* **2018**, *29*, 177–186. [[CrossRef](#)]
64. Robledo, S.; Velez, I.D.; Schmidt, T.J. Arnica Tincture Cures Cutaneous Leishmaniasis in Golden Hamsters. *Molecules* **2018**, *23*, 150. [[CrossRef](#)]
65. Laverde-Paz, M.J.; Echeverry, M.C.; Patarroyo, M.A.; Bello, F.J. Evaluating the anti-leishmania activity of *Lucilia sericata* and *Sarconesiopsis magellanica* blowfly larval excretions/secretions in an in vitro model. *Acta Trop.* **2018**, *177*, 44–50. [[CrossRef](#)]
66. Patiño-Márquez, I.A.; Patiño-González, E.; Hernández-Villa, L.; Ortiz-Reyes, B.; Manrique-Moreno, M. Identification and evaluation of *Galleria mellonella* peptides with antileishmanial activity. *Anal. Biochem.* **2018**, *546*, 35–42. [[CrossRef](#)] [[PubMed](#)]
67. Vivero, R.J.; Mesa, G.B.; Robledo, S.M.; Herrera, C.X.M.; Cadavid-Restrepo, G. Enzymatic, antimicrobial, and leishmanicidal bioactivity of gram-negative bacteria strains from the midgut of *Lutzomyia evansi*, an insect vector of leishmaniasis in Colombia. *Biotechnol. Rep.* **2019**, *24*, e00379. [[CrossRef](#)] [[PubMed](#)]

68. Patino, L.H.; Mendez, C.; Rodriguez, O.; Romero, Y.; Velandia, D.; Alvarado, M.; Pérez, J.; Duque, M.C.; Ramírez, J.D. Spatial distribution, Leishmania species and clinical traits of Cutaneous Leishmaniasis cases in the Colombian army. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005876. [CrossRef]
69. Ovalle-Bracho, C.; Londoño-Barbosa, D.; Salgado-Almario, J.; González, C. Evaluating the spatial distribution of *Leishmania* parasites in Colombia from clinical samples and human isolates (1999 to 2016). *PLoS ONE* **2019**, *14*, e0214124. [CrossRef] [PubMed]
70. Croft, S.L.; Yardley, V.; Kendrick, H. Drug sensitivity of *Leishmania* species: Some unresolved problems. *Trans. R. Soc. Trop. Med. Hyg.* **2002**, *96*, S127–S129. [CrossRef]
71. Alcântara, L.M.; Ferreira, T.C.; Fontana, V.; Chatelain, E.; Moraes, C.B.; Freitas-Junior, L.H. A Multi-Species Phenotypic Screening Assay for Leishmaniasis Drug Discovery Shows That Active Compounds Display a High Degree of Species-Specificity. *Molecules* **2020**, *25*, 2551. [CrossRef]
72. Dea-Ayuela, M.A.; Bilbao-Ramos, P.; Bolas-Fernández, F.; González, M.A. Synthesis and antileishmanial activity of C7- and C12-functionalized dehydroabietylamine derivatives. *Eur. J. Med. Chem.* **2016**, *121*, 445–450. [CrossRef]
73. Emami, S.; Tavangar, P.; Keighobadi, M. An overview of azoles targeting sterol 14 α -demethylase for antileishmanial therapy. *Eur. J. Med. Chem.* **2017**, *135*, 241–259. [CrossRef]
74. Faiões, V.D.S.; da Frota, L.C.R.M.; Cunha-Júnior, E.F.; Barcellos, J.C.F.; da Silva, T.; Netto, C.D.; da Silva, S.A.G.; da Silva, A.J.M.; Costa, P.R.R.; Torres-Santos, E.C. Second-generation pterocarpanquinones: Synthesis and antileishmanial activity. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2018**, *24*, 35. [CrossRef]
75. Fernández, O.L.; Diaz-Toro, Y.; Ovalle, C.; Valderrama, L.; Muvdi, S.; Rodríguez, I.; Gomez, M.A.; Saravia, N.G. Miltefosine and Antimonial Drug Susceptibility of *Leishmania viannia* Species and Populations in Regions of High Transmission in Colombia. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2871. [CrossRef]
76. Franco-Muñoz, C.; Manjarrés-Estremor, M.; Ovalle-Bracho, C. Intraspecies differences in natural susceptibility to amphotericin B of clinical isolates of *Leishmania* subgenus Viannia. *PLoS ONE* **2018**, *13*, e0196247. [CrossRef]
77. Hefnawy, A.; Berg, M.; Dujardin, J.-C.; de Muylder, G. Exploiting Knowledge on Leishmania Drug Resistance to Support the Quest for New Drugs. *Trends Parasitol.* **2017**, *33*, 162–174. [CrossRef] [PubMed]
78. Muñoz, D.L.; Robledo, S.M.; Kolli, B.K.; Dutta, S.; Chang, K.P.; Muskus, C. *Leishmania (Viannia) panamensis*: An in vitro assay using the expression of GFP for screening of antileishmanial drug. *Exp. Parasitol.* **2009**, *122*, 134–139. [CrossRef]
79. Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. [CrossRef] [PubMed]
80. Katsuno, K.; Burrows, J.N.; Duncan, K.; van Huijsduijnen, R.H.; Kaneko, T.; Kita, K.; Mowbray, C.E.; Schmatz, D.; Warner, P.; Slingsby, B.T. Hit and lead criteria in drug discovery for infectious diseases of the developing world. *Nat. Rev. Drug Discov.* **2015**, *14*, 751–758. [CrossRef]
81. *MACCS Structural Keys*; Accelrys: San Diego, CA, USA, 2011.
82. Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113. [CrossRef]
83. RDKit: Cheminformatics and Machine Learning Software. Available online: <http://rdkit.org> (accessed on 21 May 2020).
84. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
85. Brown, S.P.; Muchmore, S.W.; Hajduk, P.J. Healthy skepticism: Assessing realistic model performance. *Drug Discov. Today* **2009**, *14*, 420–427. [CrossRef]
86. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Systems (NIPS 2017), Long Beach, California, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4768–4777, ISBN 978-1-5108-6096-4.
87. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]
88. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J. Med. Chem.* **2019**, *63*, 8761–8777. [CrossRef]

89. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *J. Comput. Mol. Des.* **2020**, *34*, 1013–1026. [[CrossRef](#)] [[PubMed](#)]
90. Lagorce, D.; Bouzlama, L.; Becot, J.; Miteva, M.A.; Villoutreix, B.O. FAF-Drugs4: Free ADME-tox filtering computations for chemical biology and early stages drug discovery. *Bioinformatics* **2017**, *33*, 3658–3660. [[CrossRef](#)] [[PubMed](#)]
91. Staderini, M.; Piquero, M.; Abengózar, M.Á.; Nachér-Vázquez, M.; Romanelli, G.; López-Alvarado, P.; Rivas, L.; Bolognesi, M.L.; Menéndez, J.C. Structure-activity relationships and mechanistic studies of novel mitochondria-targeted, leishmanicidal derivatives of the 4-aminostyrylquinoline scaffold. *Eur. J. Med. Chem.* **2019**, *171*, 38–53. [[CrossRef](#)] [[PubMed](#)]
92. Fakhfakh, M.A.; Fournet, A.; Prina, E.; Mouscadet, J.F.; Franck, X.; Hocquemiller, R.; Figadère, B. Synthesis and biological evaluation of substituted quinolines: Potential treatment of protozoal and retroviral co-infections. *Bioorg. Med. Chem.* **2003**, *11*, 5013–5023. [[CrossRef](#)]
93. Mateen, F.J.; Oh, J.; Tergas, A.I.; Bhayani, N.H.; Kamdar, B.B. Titles versus titles and abstracts for initial screening of articles for systematic reviews. *Clin. Epidemiol.* **2013**, *5*, 89–95. [[CrossRef](#)]
94. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 1–10. [[CrossRef](#)]
95. López-López, E.; Naveja, J.J.; Medina-Franco, J.L. DataWarrior: An evaluation of the open-source drug discovery tool. *Expert Opin. Drug Discov.* **2019**, *14*, 335–341. [[CrossRef](#)]
96. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [[CrossRef](#)]
97. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, O.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
98. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
99. Fawagreh, K.; Gaber, M.M.; Elyan, E. Random forests: From early developments to recent advancements. *Syst. Sci. Control. Eng.* **2014**, *2*, 602–609. [[CrossRef](#)]
100. Nembrini, S.; König, I.R.; Wright, M.N. The revival of the Gini importance? *Bioinformatics* **2018**, *34*, 3711–3718. [[CrossRef](#)] [[PubMed](#)]
101. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
102. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games*; Kuhn, H.W., Tucker, A.W., Eds.; Princeton University Press: Princeton, NJ, USA, 1953; Volume 2, pp. 307–317.
103. Lundberg, S.M. A Game Theoretic Approach to Explain the Output of Any Machine Learning Model. Available online: <https://github.com/slundberg/shap#citations> (accessed on 29 October 2020).

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).