

## Article

# Neglected Tropical Diseases: A Chemoinformatics Approach for the Use of Biodiversity in Anti-Trypanosomatid Drug Discovery

Marilia Valli <sup>1,2,\*</sup>, Thiago H. Döring <sup>1,3,†</sup>, Edgard Marx <sup>4</sup>, Leonardo L. G. Ferreira <sup>1</sup>, José L. Medina-Franco <sup>5</sup> and Adriano D. Andricopulo <sup>1,\*</sup>

<sup>1</sup> Laboratory of Medicinal and Computational Chemistry (LQMC), Center for Research and Innovation in Biodiversity and Drug Discovery (CIBFar), Institute of Physics of Sao Carlos, University of Sao Paulo (USP), Av. Joao Dagnone, n° 1100, Sao Carlos 13563-120, SP, Brazil; thiago.doring@ufsc.br (T.H.D.); leonardo@ifsc.usp.br (L.L.G.F.)

<sup>2</sup> School of Pharmaceutical Sciences of Ribeirao Preto (FCFRP), University of Sao Paulo (USP), Avenida Professor Doutor Zeferino Vaz, s/n, Ribeirao Preto 14040-903, SP, Brazil

<sup>3</sup> Department of Exact Sciences and Education (CEE), School of Technology, Exact Sciences and Education (CTE), Federal University of Santa Catarina (UFSC), Blumenau 89036-256, SC, Brazil

<sup>4</sup> Agile Knowledge Engineering and Semantic Web (AKSW), Institute of Computer Science, Leipzig University of Applied Sciences (HTWK), 04109 Leipzig, Germany; edgard.marx@htwk-leipzig.de

<sup>5</sup> DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autonoma de Mexico (UNAM), Avenida Universidad 3000, Mexico City 04510, Mexico; medinajl@unam.mx

\* Correspondence: marilia.valli@usp.br (M.V.); aandrico@ifsc.usp.br (A.D.A.)

† These authors contributed equally to this work.

**Abstract:** The development of new treatments for neglected tropical diseases (NTDs) remains a major challenge in the 21st century. In most cases, the available drugs are obsolete and have limitations in terms of efficacy and safety. The situation becomes even more complex when considering the low number of new chemical entities (NCEs) currently in use in advanced clinical trials for most of these diseases. Natural products (NPs) are valuable sources of hits and lead compounds with privileged scaffolds for the discovery of new bioactive molecules. Considering the relevance of biodiversity for drug discovery, a chemoinformatics analysis was conducted on a compound dataset of NPs with anti-trypanosomatid activity reported in 497 research articles from 2019 to 2024. Structures corresponding to different metabolic classes were identified, including terpenoids, benzoic acids, benzenoids, steroids, alkaloids, phenylpropanoids, peptides, flavonoids, polyketides, lignans, cytochalasins, and naphthoquinones. This unique collection of NPs occupies regions of the chemical space with drug-like properties that are relevant to anti-trypanosomatid drug discovery. The gathered information greatly enhanced our understanding of biologically relevant chemical classes, structural features, and physicochemical properties. These results can be useful in guiding future medicinal chemistry efforts for the development of NP-inspired NCEs to treat NTDs caused by trypanosomatid parasites.

**Keywords:** natural products; neglected tropical diseases; drug discovery; chagas disease; leishmaniasis; human African trypanosomiasis



**Citation:** Valli, M.; Döring, T.H.; Marx, E.; Ferreira, L.L.G.; Medina-Franco, J.L.; Andricopulo, A.D. Neglected Tropical Diseases: A Chemoinformatics Approach for the Use of Biodiversity in Anti-Trypanosomatid Drug Discovery. *Biomolecules* **2024**, *14*, 1033. <https://doi.org/10.3390/biom14081033>

Academic Editor: Alessandro Paiardini

Received: 2 July 2024

Revised: 2 August 2024

Accepted: 13 August 2024

Published: 20 August 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Neglected tropical diseases (NTDs) are a group of twenty diseases of poverty that impose a devastating human, social, and economic burden on more than one billion people in tropical and subtropical areas of the world [1]. The World Health Organization (WHO) 2021–2030 Road Map comprises global targets and indicators to prevent, control, eliminate, or eradicate NTDs by 2030, see ref. [2], as well as cross-cutting targets aligned with the United Nations Sustainable Development Goals (SDGs) [3].

According to the 2023 Global Report on Neglected Tropical Diseases 2023 (WHO) [4], noteworthy progress has been made since the launch of the road map. For example,

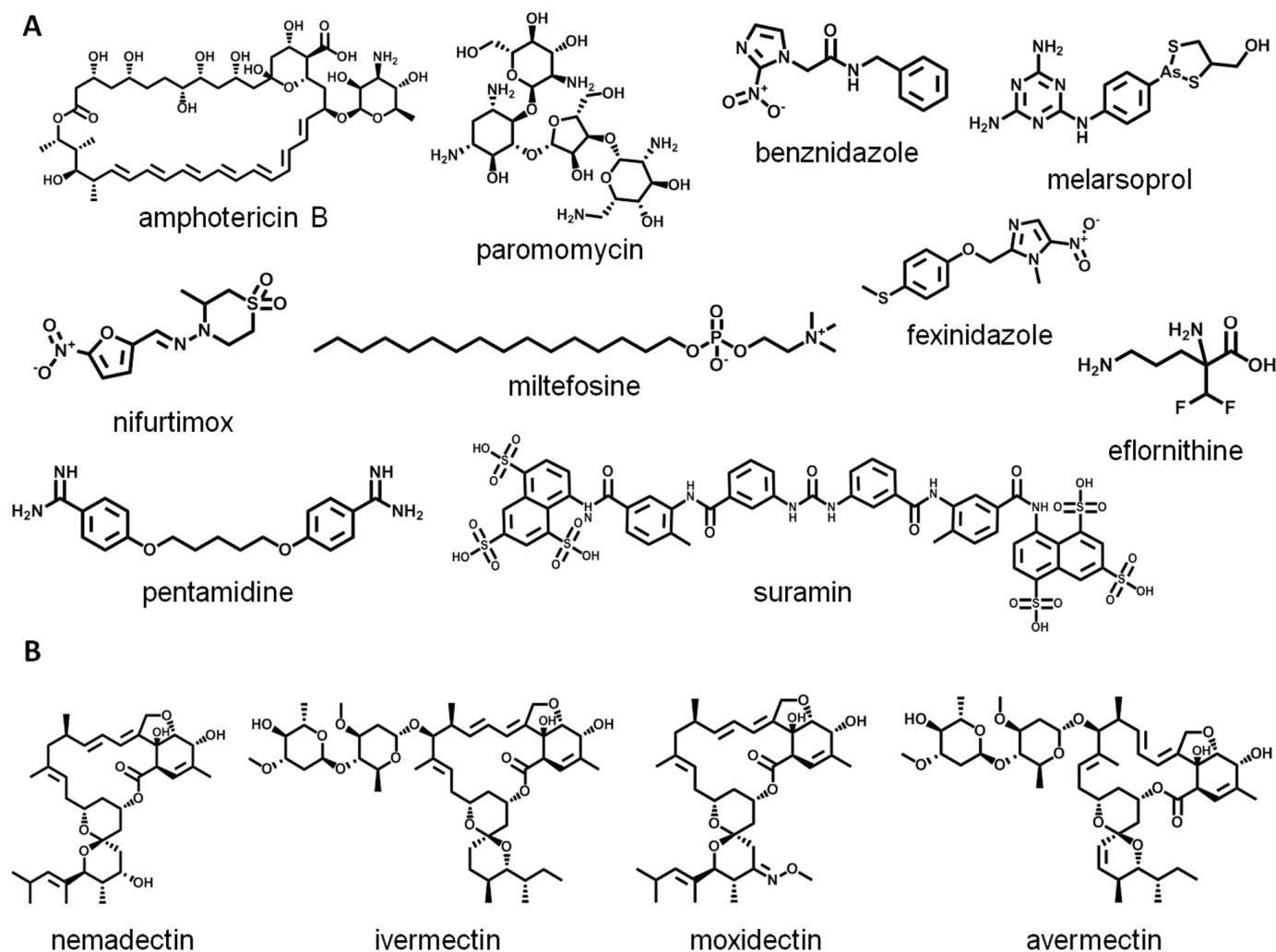
47 countries had eliminated at least one NTD by the end of December 2022. The number of people requiring interventions against NTDs has decreased by 25% over the past decade, with a reduction of 81 million people between 2020 and 2021 alone, from 1.734 billion to 1.653 billion. Nonetheless, many difficulties in achieving the targets for 2030, in addition to the COVID-19 pandemic, have revealed the scale of the task still ahead. The 2021–2022 period saw several outbreaks of NTDs, including dengue, chikungunya, leishmaniasis, Chagas disease, and scabies. The COVID-19 pandemic led to a substantial reduction in the number of people receiving interventions against NTDs. In 2020, only 798 million individuals had received treatment for at least one NTD, a reduction of 34% compared with 2019, when this figure amounted to 1.207 billion. In 2021, 90 million more people were treated, bringing the total to 888 million (+11%). Although the positive trend registered in 2021 is likely to continue, the difference from the pre-COVID-19 era is substantial, when more than one billion people were treated every year for four consecutive years between 2016 and 2019 [4].

Scientists working with NTDs are confronted with a long-standing challenge: the current treatments available have limitations in terms of safety and efficacy, among others; and inconceivably, from the 1970s to 2023, no New Chemical Entities (NCEs) were developed for this group of diseases that account for about 11% of the total disease burden in the world [5]. In this period, only new formulations or repositioned compounds were approved for these 20 conditions. In the 21st century alone, miltefosine was repurposed for leishmaniasis (2014), moxidectin for onchocerciasis (2018), fexinidazole for human African trypanosomiasis (HAT, 2021), and a pediatric formulation of benznidazole was approved for Chagas disease (2017) [1,4].

Although battling NTDs should be a priority for humanity and sustainability, there is a clear lack of investment in research and development (R&D) programs, and the NTD market is unattractive to the pharmaceutical industry [4–6]. Therefore, it is of great importance to focus on the discovery of NCEs for the treatment of NTDs. Natural products (NPs) are valuable sources for the development of drugs for a variety of human diseases. This includes NTDs, such as the anti-leishmanial agent amphotericin B (Figure 1A), extracted from *Streptomyces noclosus* and primarily used to treat fungal infections. The antimicrobial aminoglycoside paromomycin (Figure 1A), produced by *Streptomyces krestomuceticus*, is used to treat leishmaniasis. Moxidectin (Figure 1B), employed to treat onchocerciasis, is obtained from the modification of the NP nemadectin (Figure 1B), which was isolated from *Streptomyces cyaneogriseus*. Ivermectin (Figure 1B), used for the treatment of onchocerciasis, lymphatic filariasis, scabies, and other ectoparasitoses, is a dihydro analogue of the macrocyclic lactone avermectin (Figure 1B), whose analogues were obtained from *Streptomyces avermitilis*, an actinomycete present in soil.

NPs have a long history of achievements in the early stages of R&D initiatives as a source of new hits and inspiration for new lead compounds with privileged drug-like properties for NTD drug discovery. In this work, we concentrate our efforts on NTDs caused by trypanosomatid parasites, Chagas disease, HAT, and leishmaniasis, in which NCEs are needed to enable new generations of therapies to revolutionize the clinical treatment of these diseases, and to save millions of lives [7]. Chagas disease, caused by the parasite *Trypanosoma cruzi*, is endemic in 21 Latin American countries [8]. There are 6–7 million people infected worldwide, with another 75 million at risk of contamination. Only two old nitro-heterocyclic drugs—benznidazole and nifurtimox (Figure 1A)—are available, and both have several limitations. Leishmaniasis, caused by more than 20 species of *Leishmania* sp., affects 700,000 to 1 million people every year, and its visceral form is fatal if left untreated in over 95% of cases, with about 50,000–90,000 new cases each year [5]. The existing drugs have variable efficacy and serious toxicities—amphotericin B, pentavalent antimonials, and paromomycin—and only one, miltefosine (Figure 1A), is administered orally, whereas the others are given by intravenous or intramuscular injections. HAT, caused by *Trypanosoma brucei gambiense* (g-HAT) and *T. b. rhodesiense* (r-HAT), is endemic in sub-Saharan Africa. Seventy mil-

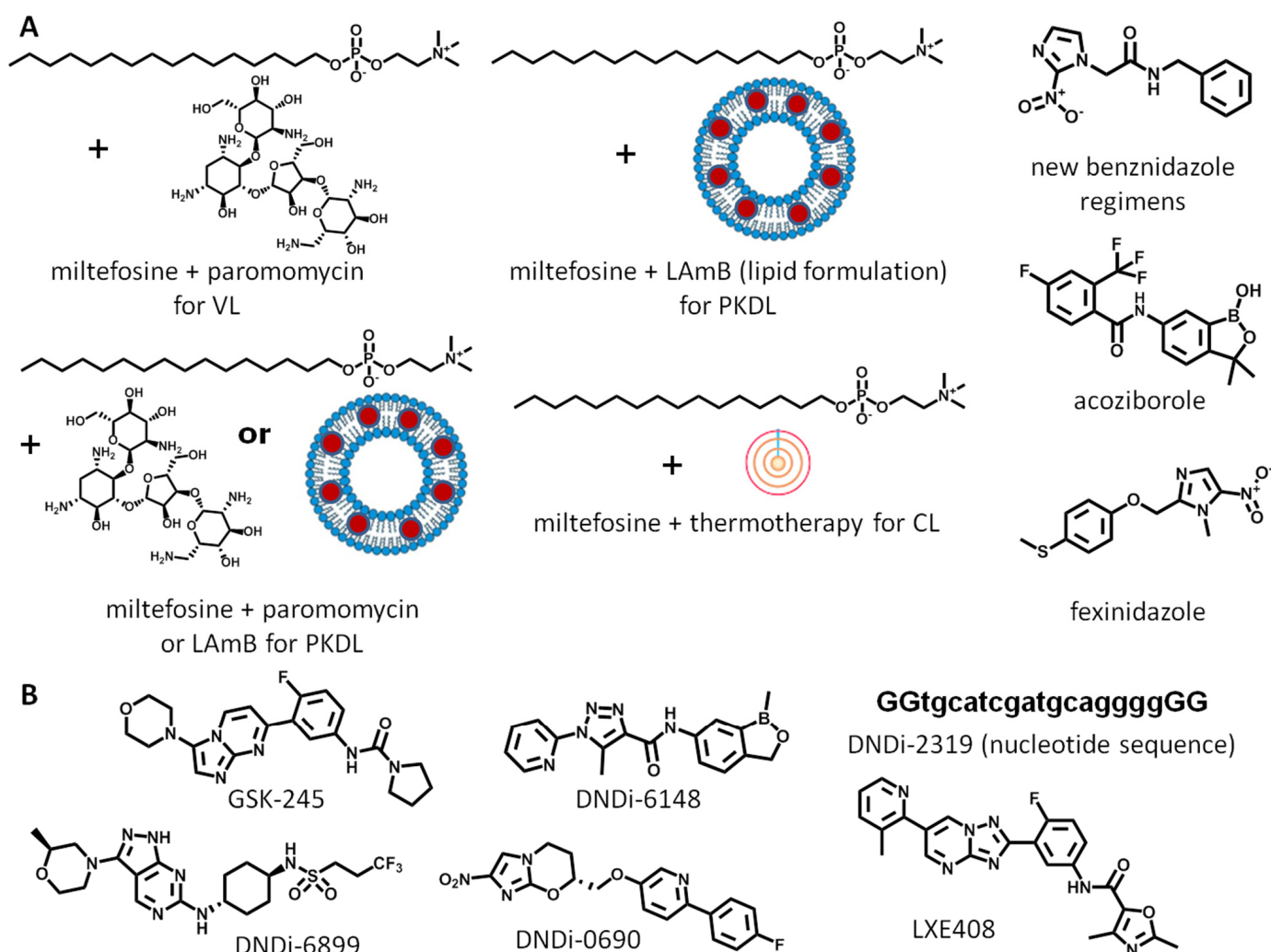
lion people are at risk of infection, [9] and the therapies available are based on highly toxic compounds: melarsoprol, eflornithine, suramin, pentamidine, and nifurtimox (Figure 1A). Fexinidazole was introduced in 2021 as the first effective oral monotherapy against g-HAT (Figure 1A).



**Figure 1.** Drugs used for NTDs: (A) drugs for trypanosomatid diseases, (B) drugs for other NTDs.

The current clinical pipeline (DNDi R&D portfolio, 2023) for anti-trypanosomatid drug discovery (Figure 2) [10], focusing on more advanced clinical trials (phase IIb/III and registration), is dominated by new formulations, new regimens, or combinations of old drugs (Figure 2) for leishmaniasis. The number of compounds is modest and represents a well-known repertoire of unsatisfactory drugs (amphotericin B, paromomycin, miltefosine, and fexinidazole; Figure 1A). For Chagas disease, only new regimens of benznidazole are under consideration (Figure 2A). For HAT, orally active acoziborole is in phase IIb/III (Figure 2A). Under registration, there is only a drug combination (miltefosine + paromomycin) for visceral leishmaniasis, and fexinidazole for HAT (Figure 2A). In the early stages of the clinical pipeline—phase I and phase IIa/proof-of-concept—there are a few NCE candidates in clinical development (Figure 2B) [10]. The situation is critical for Chagas disease; only one compound is in phase I, a benzoxaborole derivative (DNDI-6148, CPSF3 inhibitor) (Figure 2B). For leishmaniasis, six candidates are under investigation. Five candidates are in phase I: DNDI-0690 (bioactivation by NTR2), GSK-245 (proteasome inhibitor), DNDI-6148 (CPSF3 inhibitor), DNDI-6899 (CRK12 inhibitor), and DNDI-2319

(oligonucleotide). In phase II, there is only a proteasome inhibitor (LXE408) (Figure 2B). For HAT, no compounds are under consideration in the early stages.



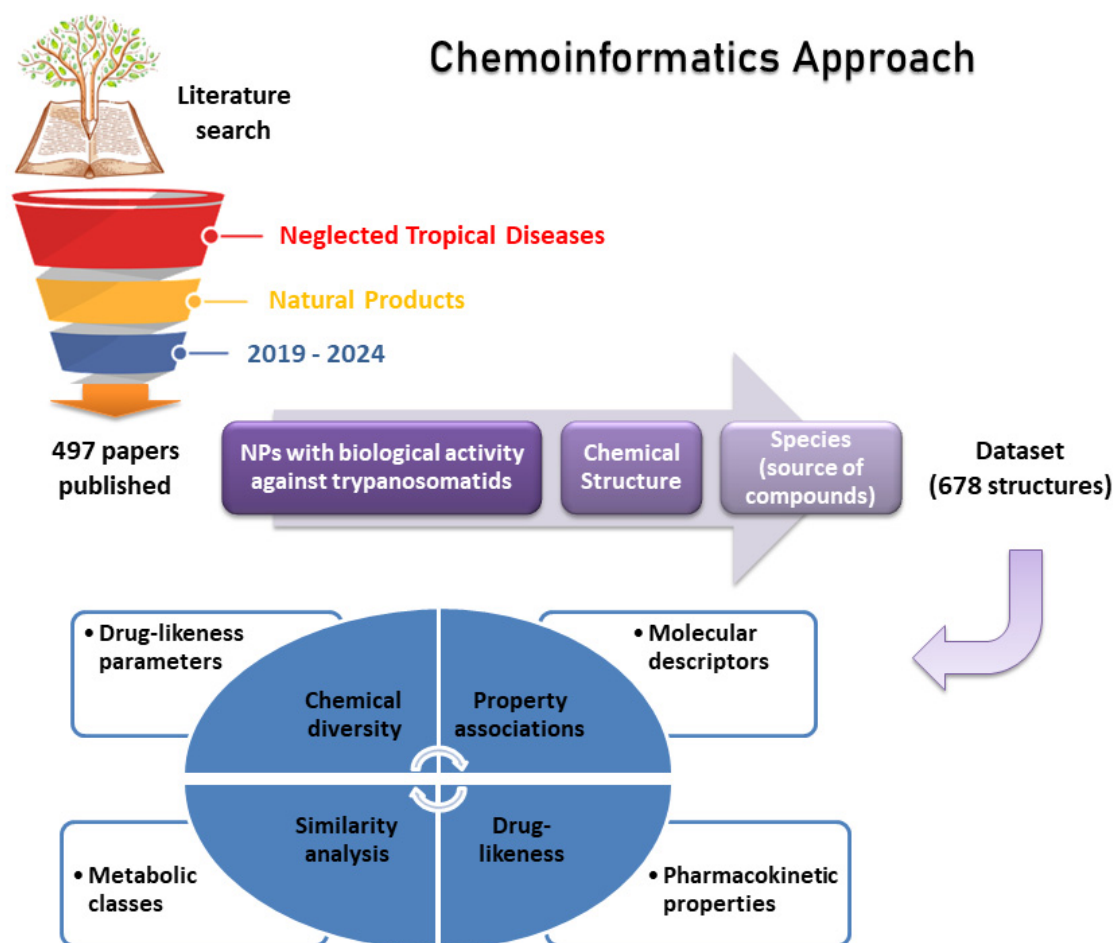
**Figure 2.** Current clinical pipeline for Chagas disease, leishmaniasis, and HAT: (A) advanced clinical trials, (B) early stages of clinical development. DNDi-2319 uppercase: phosphothionate bases; lowercase: phosphodiester bases. VL: visceral leishmaniasis; PKDL: post-kala-azar dermal leishmaniasis; CL: cutaneous leishmaniasis; LAmB: liposomal amphotericin B.

Cheminformatics have played an important role in the hit identification and hit-to-lead stages of drug discovery, allowing us to focus on privileged chemical scaffolds (lead compounds) that exhibit promising drug-like properties [11]. In this study, we examined the literature from 2019 to March 2024 to identify NP compounds with promising anti-trypanosomatid activity. As part of the literature survey, we created a database and analyzed the structural content, distribution in chemical space, and determined several molecular and physicochemical properties of pharmaceutical interest using computational tools. We also used cheminformatic approaches to reveal important insights into the understanding of the chemical classes, molecular scaffolds, and corresponding drug-like properties of small-molecule NPs. The findings of this study can be useful in guiding future medicinal chemistry efforts to develop NP-based NCEs for NTDs caused by trypanosomatid parasites.



## 2. Materials and Methods

**Literature search.** The literature search was performed on 2 April 2024 with the keywords described in Table 1 to construct the dataset used in this study. With the aim of conducting an extensive search, we used the SciFinder-n (Chemical Abstracts Service, Columbus, OH, USA) and Web of Science (Clarivate, London, UK) platforms [12,13]. A total of 497 papers published from 2019 to March 2024 were selected. The papers were individually analyzed to extract information on the bioactive compounds tested against *T. cruzi*, *T. brucei*, or eight *Leishmania* species for the creation of the dataset (Figure 3).



**Figure 3.** Strategy used to build the dataset used in this study.

**Dataset.** The complete compound dataset used in this study is available in the Supporting Information (Supporting Information). The literature search was manually analyzed to identify NPs reported to have biological activity ( $IC_{50}$  determined or percentage of inhibition greater than 50%) against trypanosomatid parasites. Using these criteria, information on 678 NPs was collected (Supporting Information).

**Molecular descriptors and pharmacokinetic properties.** Molecular descriptors, pharmacokinetic properties, and drug-likeness parameters were computed using the SwissADME platform (University of Lausanne and the SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland) [14]. Descriptors for the ring count analyses were calculated using the QikProp module in Maestro v. 11.2.013 (Schrödinger, New York, NY, USA). The clogP values for the n-octanol/water system was calculated using the implicit logP method provided by the SwissADME platform (University of Lausanne and the SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland). Metabolic classes were determined with the aid of Classyfire v. 1.0 (Edmonton, AB, Canada) [15]. The calculated data can be found in the Supporting Information.

**Structural fingerprint.** For the structural similarity analyses, the Canvas Fingerprint Similarity module was used (Maestro v. 11.2.013, Schrödinger, New York, NY, USA). The linear fingerprint type was used. The cluster was built using 64-bit precision. The atom typing scheme was distinguished by ring size, aromaticity, HBA/HBD, ionization potential, and whether the atom is terminal or halogen. Bonds were distinguished by bond order. The similarity metric applied was calculated using the Tanimoto coefficient [16] and the linkage average method.

**Molecular properties of chemical space and principal component analysis.** Four molecular properties of pharmaceutical interest were computed in DataWarrior (v. 5.5.0, Actelion/Idorsia Pharmaceuticals Ltd., Allschwil, Switzerland) [17]: nRotB, HBA, HBD, and MW (PC1 = 64.649%; PC2 = 20.397%; PC3 = 9.756%). The linear correlation between descriptors was performed using the Bravais–Pearson coefficient [18]. Candlestick charts, means, medians, and quartiles were calculated using DataWarrior (v. 5.5.0, Actelion/Idorsia Pharmaceuticals Ltd., Allschwil, Switzerland).

**Table 1.** Keywords used for the literature search and the resulting number of papers.

SciFinder-n Search			
Search	Abstract/Keywords	Publication Year	Number of Papers
Natural products	Trypanosomatida	2019–2024	20
Natural products	Trypanosoma	2019–2024	240
Natural products	Leishmania	2019–2024	329
Natural products	Leishmaniasis	2019–2024	252
Natural products	Chagas	2019–2024	120
Natural products	Neglected Tropical Diseases	2019–2024	119
Secondary metabolites	Trypanosoma	2019–2024	36
Secondary metabolites	Leishmania	2019–2024	66
Secondary metabolites	Trypanosomatida	2019–2024	2
Secondary metabolites	Leishmaniasis	2019–2024	42
Secondary metabolites	Chagas	2019–2024	9
Secondary metabolites	Neglected Tropical Diseases	2019–2024	25
Web of Science Search			
Topic	Topic	Year Published	Number of Papers
Natural products	Trypanosomatida	2019–2024	13
Natural products	Trypanosoma	2019–2024	234
Natural products	Leishmania	2019–2024	311
Natural products	Leishmaniasis	2019–2024	272
Natural products	Chagas	2019–2024	113
Natural products	Neglected Tropical Diseases	2019–2024	89
Secondary metabolites	Trypanosoma	2019–2024	50
Secondary metabolites	Leishmania	2019–2024	65
Secondary metabolites	Trypanosomatida	2019–2024	2
Secondary metabolites	Leishmaniasis	2019–2024	54
Secondary metabolites	Chagas	2019–2024	16
Secondary metabolites	Neglected Tropical Diseases	2019–2024	18

### 3. Results and Discussion

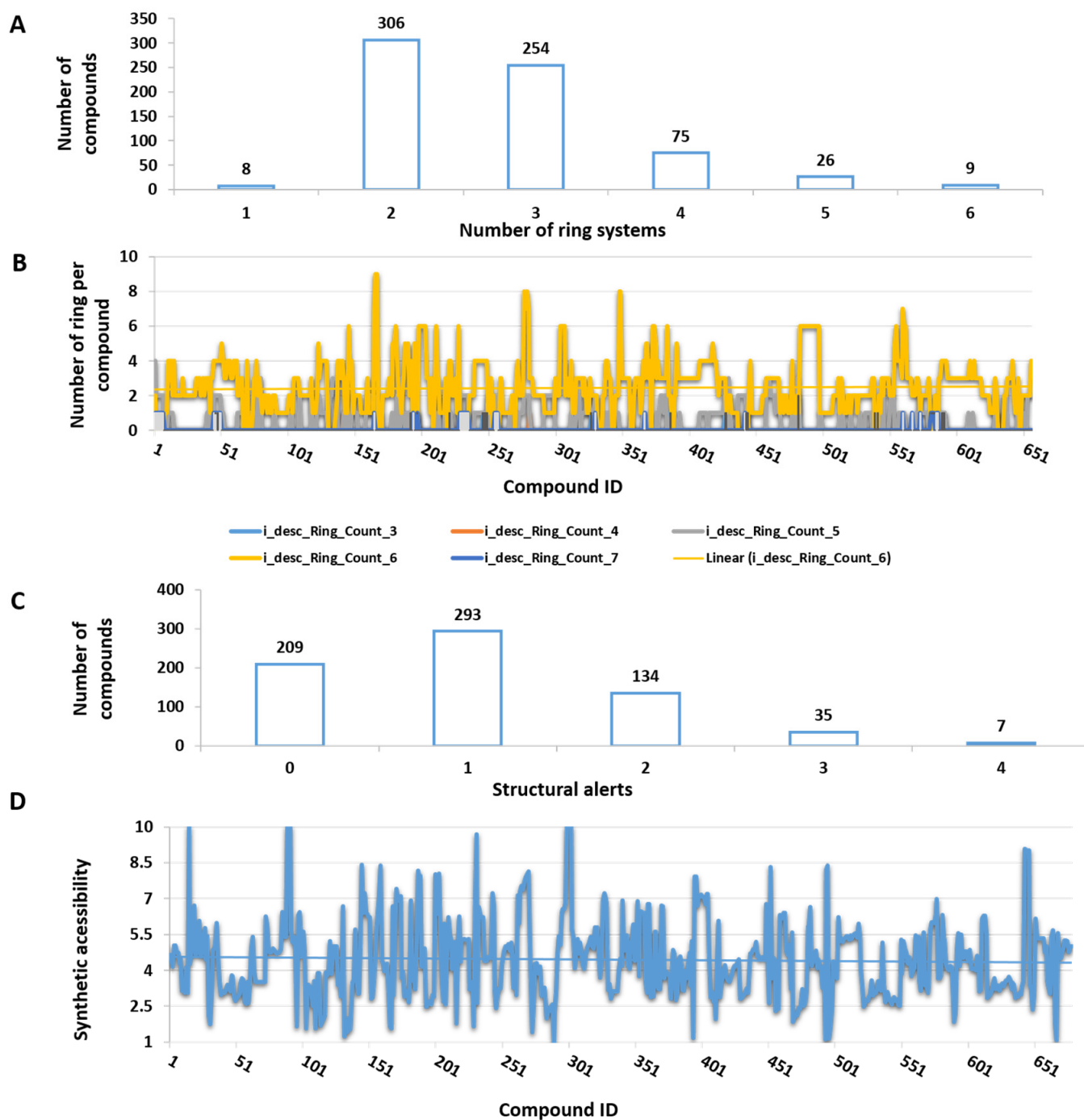
#### 3.1. Annotated Compound Database

From 2019 to March 2024, 497 research articles reported NPs with anti-trypanosomatid activity isolated from plants (73.0%), marine organisms (12.1%), fungi (4.9%), bacteria (3.1%), and animals (1.5%), as well as NP derivatives (1.9%) and compounds from NP databases (3.5%). The 497 articles selected for this study were published in the following years: 169 (24.9%) were published in 2019, 147 (21.7%) in 2020, 127 (18.7%) in 2021, 116 (17.1%) in 2022, 105 (15.5%) in 2023 and 14 (2.1%) until March 2024. The chemical and biological data collected were analysed, and a comprehensive compound dataset was generated for a unique set of 678 small-molecule bioactive NPs. Structures belonging to different metabolic classes were identified. Terpenoids represent the largest class of compounds in the dataset, with 207 structurally diverse compounds (30.5%), which is unsurprising due to their high abundance among NPs. Benzoic acids and benzenoids represent 19.5% of the dataset, with 132 structures. Furthermore, the dataset includes 63 alkaloids (9.3%), 63 polyketides (9.3%), 58 phenylpropanoids (8.6%), 44 flavonoids (6.5%), 34 peptides/peptide mimetics (5.0%), 25 lignans/neolignans (3.7%), 10 naphthoquinones (1.5%), 4 cytochalasins (0.6%), 1 xanthone (0.1%), and 1 tannin (0.1%). The remaining 32 compounds (4.7%) do not belong to any of the mentioned classes or to a single chemical class. The complete database, annotated with the chemical and biological data, is available in the Supporting Information.

#### 3.2. Ring Content, Structural Alerts, and Synthetic Accessibility

A chemoinformatic study was conducted to investigate the chemical space coverage of the dataset and to explore the biologically relevant molecular diversity for anti-trypanosomatid drug discovery. Initially, the 678 dataset compounds were grouped by ring count (Figure 4A). Most chemical structures (83% of the dataset) possess two or three ring systems, with a predominance of five- or six-membered rings (Figure 4B). Among those molecules with six-membered rings, for example, are terpenes, steroids, and flavonoids containing three rings, whereas aromatic derivatives, phenylpropanoids, and lignans bear two rings.

Next, the dataset was evaluated for the identification of structural alerts (based on Brenk filters) [19] for potentially toxic or unstable chemical moieties (Figure 4C). Most of the dataset compounds present a good drug-like profile: 502 compounds (74%) have one or no alerts; 134 (20%), 35 (5%), and 7 (1%) compounds, respectively, present two, three, and four alerts. Moreover, the synthetic accessibility of the dataset compounds was analyzed by scores varying from 1 (easiest) to 10 (most difficult), using the molecular fingerprint (FP) approach and the metric system implemented in the SwissADME web-server (University of Lausanne and the SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland) (Figure 4D) [14]. A tendency line shows that, on average, the dataset features a synthetic accessibility score between 4 and 5, which indicates an acceptable number of reaction steps to synthesize the target NP compounds. The FP method is based on the construction of a sequence of bits that determines the presence or absence of a chemical descriptor in a molecule. The final model was constructed using 1024 fragments and trained with more than 12 million structures. The model yielded a correlation coefficient value ( $r$ ) of 0.94.

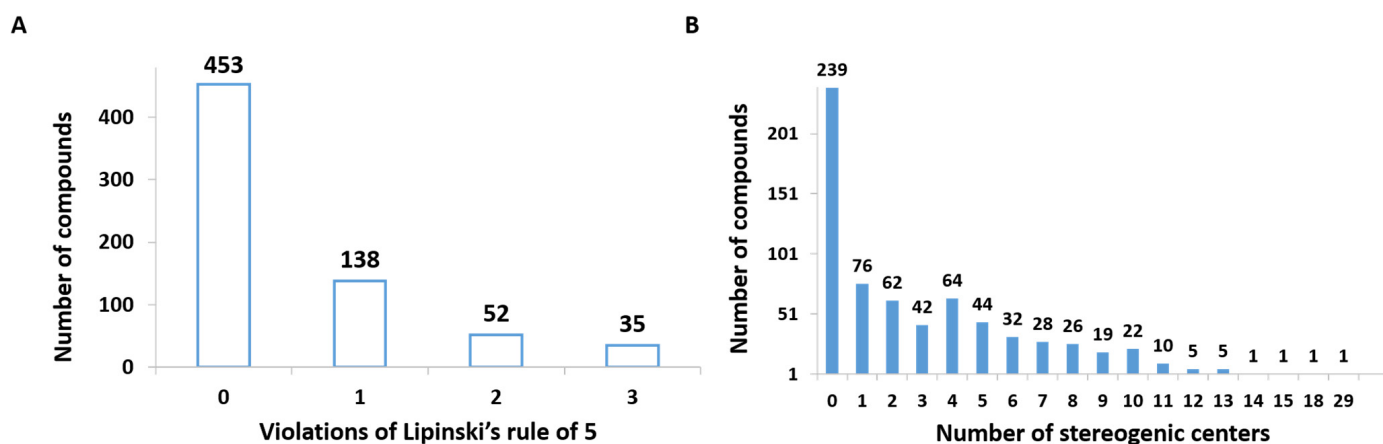


**Figure 4.** Profile of the dataset with 678 compounds regarding ring count, structural alerts, and calculated synthetic accessibility: (A) ring count considering any ring size, (B) number of rings in each structure of the dataset, (C) Brenk structural alerts, (D) synthetic accessibility scores using the SwissADME webserver (University of Lausanne and the SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland).

### 3.3. Drug-Likeness

As stated by Lipinski's rule of five (Ro5), oral drug-like compounds with good solubility and permeability should have no more than 5 hydrogen-bond donors, no more than 10 hydrogen-bond acceptors, a molecular weight (MW) no greater than 500 Da, and a calculated n-octanol/water partition coefficient (clogP) no greater than 5 [20]. With no more than one violation of the Ro5 criteria, 591 (87%) compounds of the dataset have high

potential for oral bioavailability (Figure 5A). Although there are other compound filters based on different combinations of descriptors in use today, the use of the most traditional and well-known group of Lipinski's filters is of particular interest.



**Figure 5.** Investigation of molecular properties and structural complexity of the dataset compounds: (A) violations of Lipinski's rule of five, (B) number of stereogenic centers.

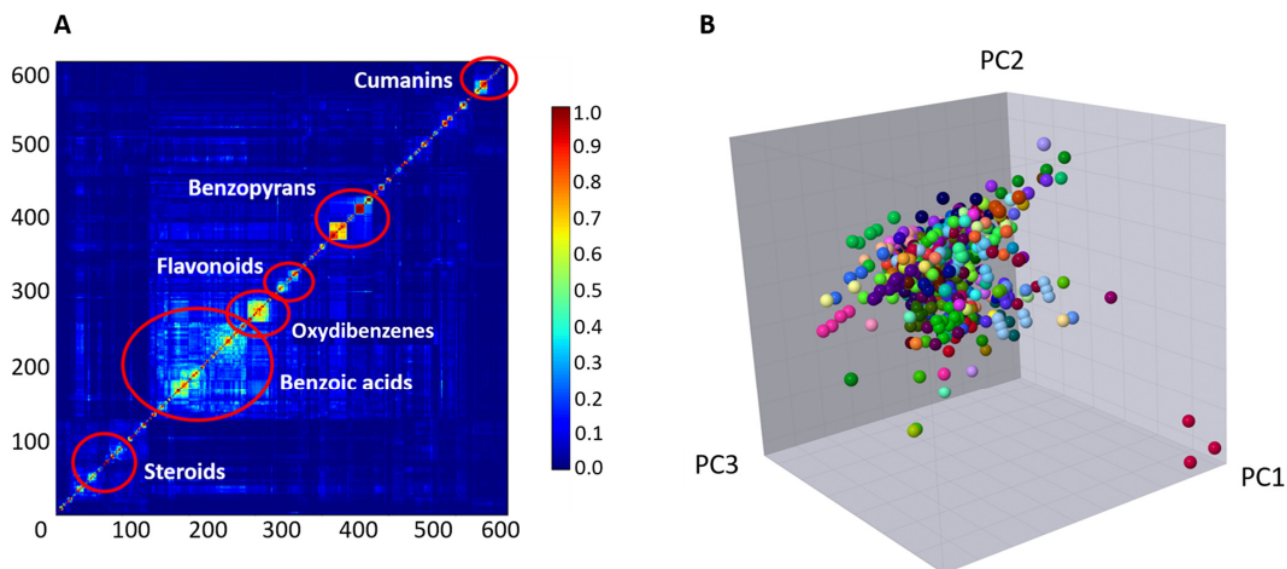
### 3.4. Stereogenic Centers

Bioactive NPs are usually associated with complex structures with high MW, moving way beyond small molecules that fall within the Ro5. As discussed in the previous section, our results indicate that 87.2% of the NPs reported in the last five years (2019 to March 2024) represent small molecules with drug-like properties for anti-trypanosomatid drug discovery. Another important finding is related to the number of stereogenic centers present in the dataset structures. Compounds with multiple chiral centers are avoided in NTD drug discovery programs due to their synthetic complexity and significant challenges for the generation of analogue series for structure–activity relationship (SAR) studies [21]. Among the NPs of the dataset, 35.3% (239 compounds) do not present stereogenic centers, 11.2% and 9.1% (76 and 62 compounds), respectively, present only 1 and 2 centers, corresponding to a total of 55.6% (Figure 5B). With considerably more complex structures, 6.8% (46 compounds) present more than 10 stereogenic centers.

### 3.5. Chemical Diversity

The chemical diversity of the dataset was assessed using a similarity chart with descriptors based on the Tanimoto coefficient and the molecular fingerprint implemented in Canvas (see Section 2) (Figure 6A) [17]. As can be seen, the overall similarity is below 30%, indicating considerable structural diversity in the dataset. The chart displays the six most important regions of similarity (red circles), representing the main classes of compounds: cumarinins, steroids, flavonoids, oxydibenzene, benzoic acids, and benzopyrans. Benzoic acids occupy the largest portion, while oxydibenzene exhibits the highest intra-class similarity. The chemical diversity was also evaluated by a three-dimensional principal component analysis (3D PCA) to reduce the dimensionality of the dataset, including the removal of descriptors that are highly correlated, while preserving as much of the relevant information as possible (Figure 6B) [22,23]. The distinct colors show the heterogeneity of the compounds in terms of their sources. According to the PCA results, for example, the regions of the plot in dark red (fungal isolates), light green (*Physalis minima*), pink (*Salileptolyngbya* sp.), and purple (*Arrabidaea brachypoda*) were found to be structurally correlated, despite their diverse sources.





**Figure 6.** Chemical diversity analysis of the dataset: (A) structural similarity chart (centroid clustered) indicating the most important regions of similarity (from dark blue to dark red, respectively, 0% to 100% similarity), (B) 3D PCA showing the chemical diversity of the NPs with their corresponding source in distinct colors. The first three components capture 94.8% of the total variance.

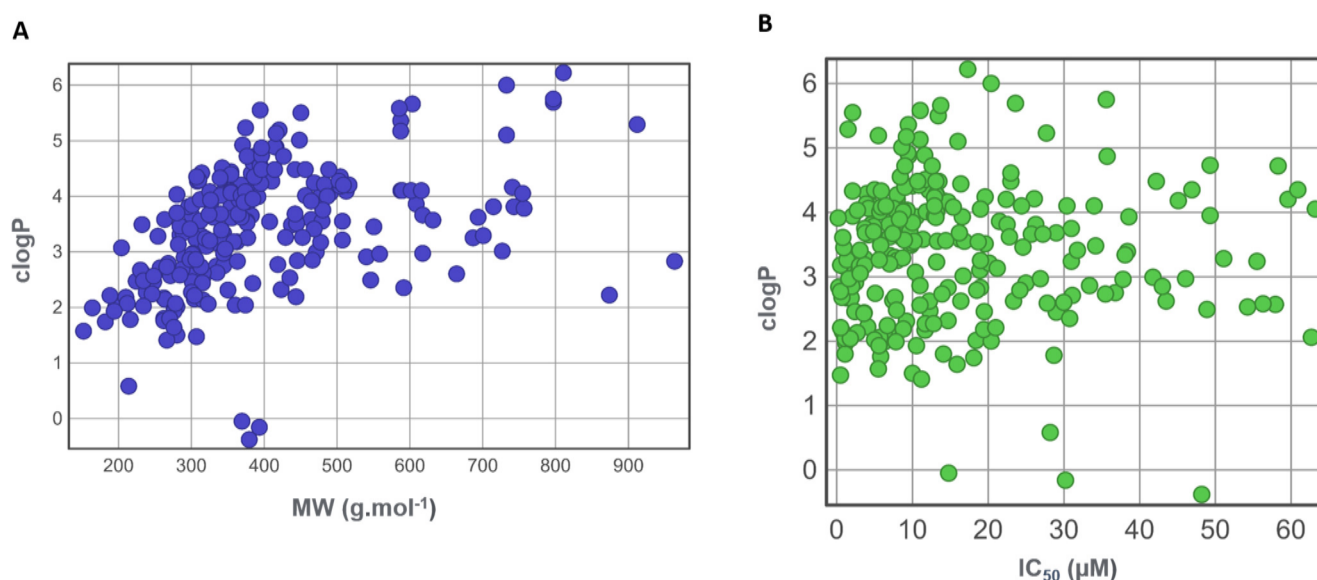
### 3.6. Property Associations

MW is one of the most important drug-like properties (Lipinski limit of 500 Da), as small-molecule drugs (organic compounds with low MW) have been the mainstay of the pharmaceutical industry for many decades. Most small molecules can be administered orally, and they can pass through cell membranes to reach intracellular targets. Lipophilicity, represented by the partition coefficient ( $p$ ), which is defined as the tendency of a neutral compound to dissolve in an immiscible biphasic system of lipids and water, is a key physicochemical property in medicinal chemistry [24]. The calculated descriptors of the logarithm  $p$  (clogP) are fundamental for predicting the permeability and absorption of bioactive compounds.

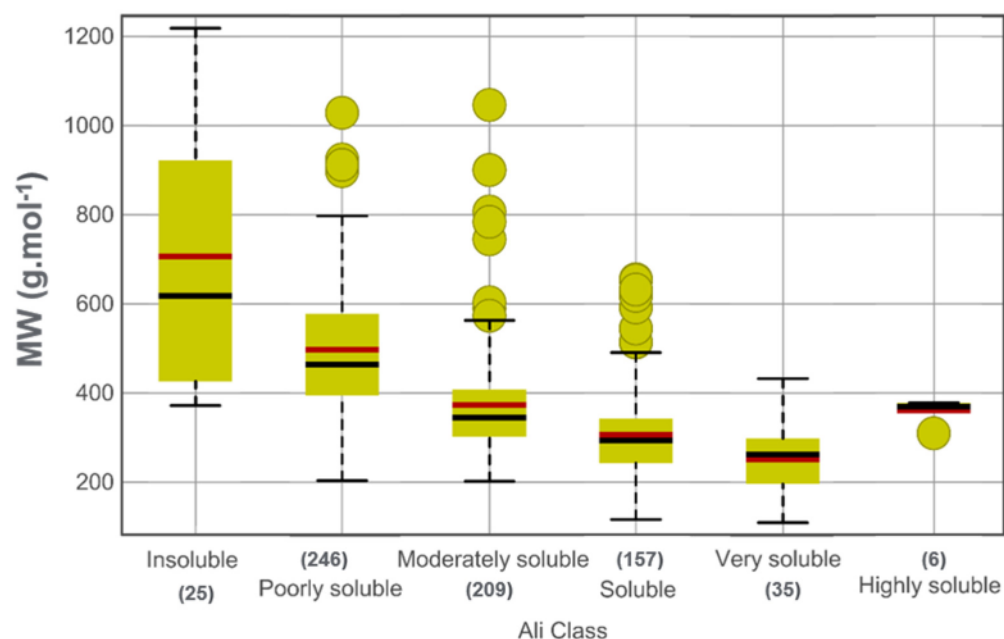
Drug candidates with higher MW and lipophilicity show poor solubility and bioavailability, leading to other problems such as challenges with metabolism, permeability, or interactions with other drugs. Given the importance of these descriptors for this unique set of NPs, their relationships were investigated using a scatter plot (Figure 7A). As can be seen, the MWs are distributed predominantly across the interval from 200 to 600 g·mol<sup>−1</sup>, whereas the clogP values are mostly scattered from 1 to 6. A strong correlation was observed between MWs and the corresponding regions of high (clogP > 4), intermediate (clogP = 2–4), and low lipophilicity (clogP < 2). Furthermore, the relationships between lipophilicity (clogP) and anti-*T. cruzi* potency (IC<sub>50</sub> values, which refer to the half-maximal inhibitory concentration) were examined for the dataset compounds (Figure 7B). The most potent compounds (IC<sub>50</sub>s < 10 μM) possess low to moderate lipophilicity (clogP from 1.8 to 3.5), which corroborates previous experimental findings [25].

Aqueous solubility is a key physicochemical property in drug discovery as it profoundly impacts bioavailability and pharmacokinetics (ADME: absorption, distribution, metabolism, and excretion) of drug candidates. It is also important in preclinical development, as the processes of hit identification, hit-to-lead, and lead optimization demand measurements of in vitro biological activity, as well as efficacy and toxicology studies in animal models [26]. The water solubility of the dataset compounds was evaluated to identify NPs with favorable oral bioavailability and pharmacokinetic characteristics (Figure 8) [14,27,28]. The results indicate that approximately 60% of the NPs have acceptable solubility (209 compounds exhibit moderate solubility, 157 are soluble, 35 are very

soluble and, 6 are highly soluble). Additionally, 246 poorly water-soluble compounds and 25 insoluble compounds were identified in the dataset.



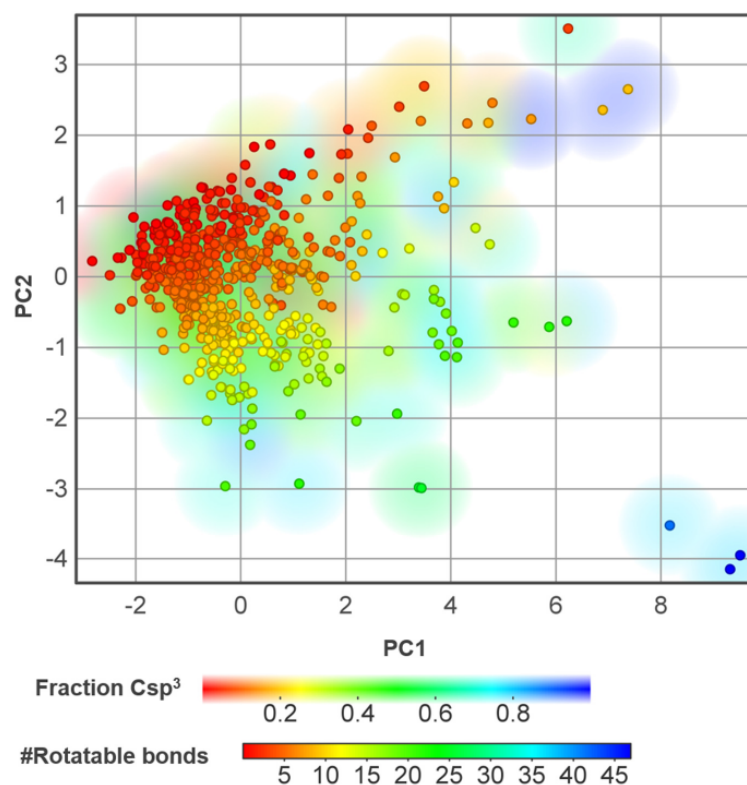
**Figure 7.** Scatter plots of associations between lipophilicity (clogP) and molecular weight (MW) and biological activity ( $IC_{50}$ ): (A) clogP versus MW for the entire dataset and (B) clogP versus  $IC_{50}$  values for a subset of 243 compounds with anti-*T. cruzi* activity.



**Figure 8.** Box plot of the distribution of MW versus water solubility scores (insoluble  $< -10 <$  poorly  $< -6 <$  moderately  $< -4 <$  soluble  $< -2 <$  very  $< 0 <$  highly). Red lines indicate the mean, black lines indicate the median, and dots indicate the outliers. Dashed lines indicate the upper and lower quartiles.

A PCA analysis was carried out using the dataset of NPs for the following molecular descriptors: rotatable bonds (nRotB), hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD), and MW. The contribution of hydrogen bonding capacity, the number of rotatable bonds, and the associated molecular conformational changes of small molecules are responsible for substantial differences in efficacy and pharmacokinetic properties. A

molecule's flexibility and rotatable bonds affect its ability to bind tightly to its targets, which is observed for rigid molecules with too few rotatable bonds. In addition, according to Veber's rule of drug-likeness, compounds with more than 10 rotatable bonds are likely to exhibit low oral bioavailability [29]. In medicinal chemistry, it is important to design molecules (lead optimization stages) with an appropriate number of rotatable bonds that balance flexibility and rigidity as well as the number of HBA and HBD (hydrogen bonding capacity) for optimal binding and improved ADME characteristics. In general, the dataset compounds possess similar characteristics in terms of nRotB, HBA, HBD, and MW (Figure 9). The analysis revealed that approximately 50% of the dataset compounds have 10 or fewer rotatable bonds (nRotB, solid dots).



**Figure 9.** 2D PCA performed using rotatable bonds (nRotB), hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD), and molecular weight (MW). Solid dot colors represent nRotB and smooth colors represent the fraction of  $sp^3$  hybridized carbon atoms related to the total carbon count (Csp<sup>3</sup>). # = number.

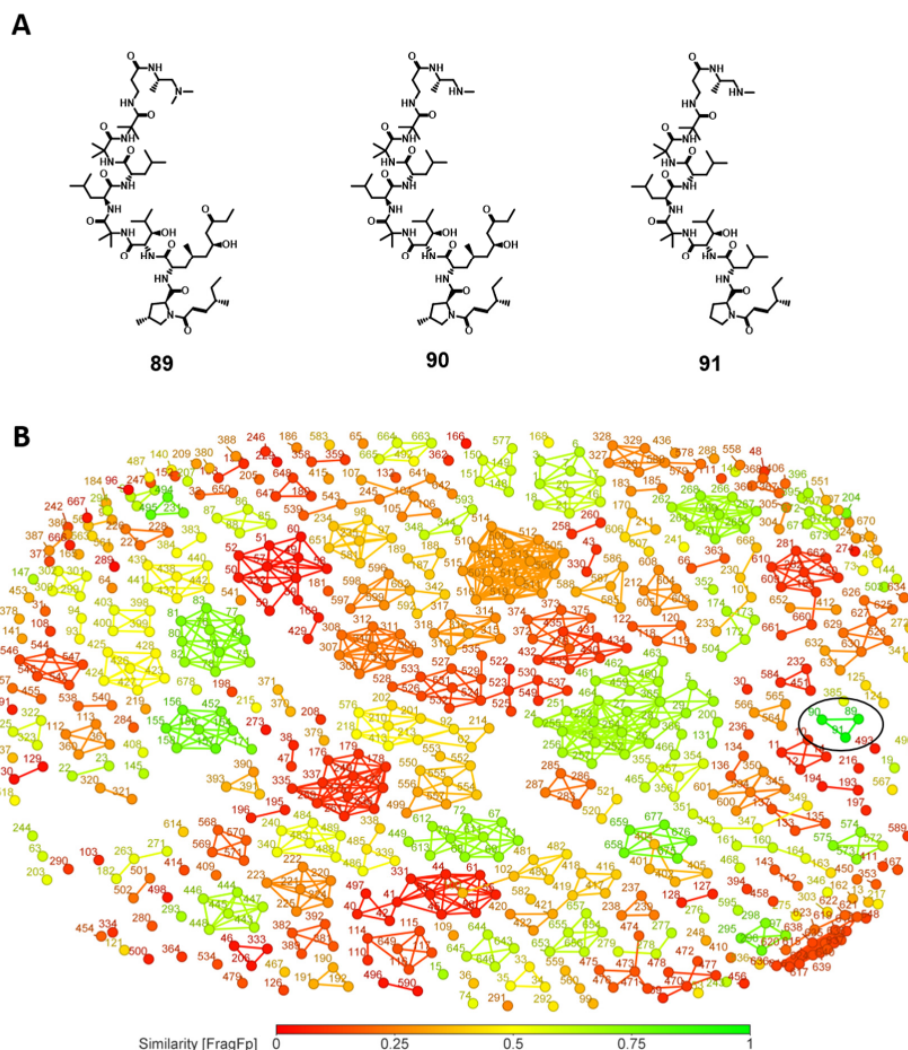
The increase in the degree of saturation, defined as the fraction of  $sp^3$  hybridized carbon atoms in relation to the total carbon count (Csp<sup>3</sup>), has been correlated with the probability of a compound translating from the discovery phase to clinical development [30]. Increasing Csp<sup>3</sup> was found to reduce molecular planarity and packing, which, in turn, enhances water solubility. Regarding this parameter, most of the dataset compounds feature a Csp<sup>3</sup> fraction > 0.25 (Figure 9).

### 3.7. Similarity Analysis of Potent Compounds

In this study, the FragFp descriptor was selected to build similarity charts. Similarity charts show similarities between two structures using specified fragment-based descriptors. FragFp includes a dictionary with 512 substructure fragments, and the more fragments two molecules have in common, the higher is the score [31]. The most potent compound (IC<sub>50</sub> = 5 nM) against the amastigote form of *T. cruzi*, leucinoastatin F (91, Figure 10A) [32] was used as a reference to investigate the degree of structural similarity to the other compounds of the dataset (Figure 10B). In this similarity analysis, which encodes the fragments into struc-

tural fingerprints, two other leucinostatin analogues (89 and 90, Figure 10A) were identified with similarity greater than 95% (Figure 10B). Both leucinostatin A (89) and leucinostatin B (90) are potent anti-*T. cruzi* agents, with IC<sub>50</sub> values of 7.1 nM and 12 nM, respectively.

A number of other structures, shown as nodes color-coded in green (Figure 10B), exhibited a degree of similarity to leucinostatin F of more than 70% (91). Nonetheless, they have the great advantage of containing active compounds with superior drug-like properties (MW ≤ 500, clogP ≤ 5, HBD ≤ 5, HBA ≤ 10, and nRotB ≤ 10). These could be explored in the design of novel antitrypanosomal drugs, including the sesquiterpene 76, the small peptides 154 and 155, the meroterpenoid 33 isolated from *Memnoniella dichroa*, and the polyketide strasseriolide 355 isolated from *Strasseria geniculata* (Figure 11).



**Figure 10.** (A) Structures of leucinostatin A (89), leucinostatin B (90), and leucinostatin F (91), (B) similarity network for leucinostatin F (91).

Similarly, the analysis with the most potent anti-*T. brucei* compound, the steroid 97 (Figure 12A), with an IC<sub>50</sub> value of 2.9 nM, revealed three compounds (98, 234, and 235) with a fingerprint-based similarity greater than 90% (Figure 12B) [33,34]. For instance, compound 98 has similarity of 96.7% and an IC<sub>50</sub><sup>*T.brucei*</sup> of 520 nM. All compounds 97, 98, 234, and 235 follow Lipinski's and Veber's rules. Compound 126 (IC<sub>50</sub> = 12 nM) is a sesquiterpene derivative isolated from *Dorema glabrum* and it also presented a relevant IC<sub>50</sub> for *L. donovani* (700 nM), demonstrating the potential of this compound for drug discovery efforts on both parasites.



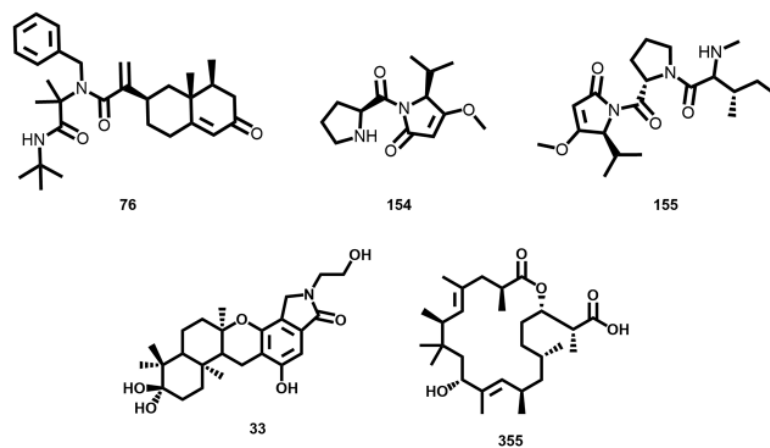


Figure 11. Structure of drug-like compounds 76, 154, 155, 33, and 355.

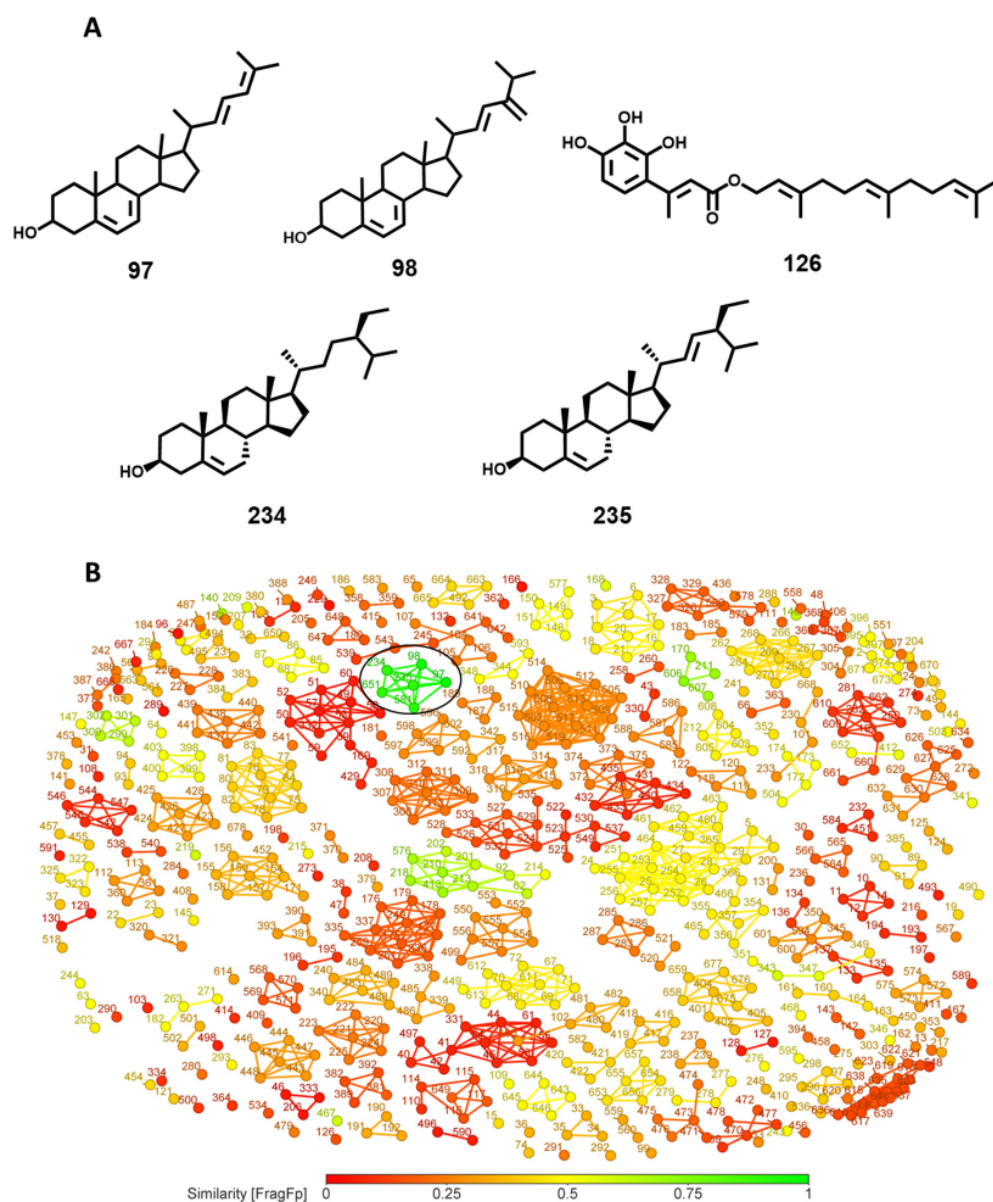
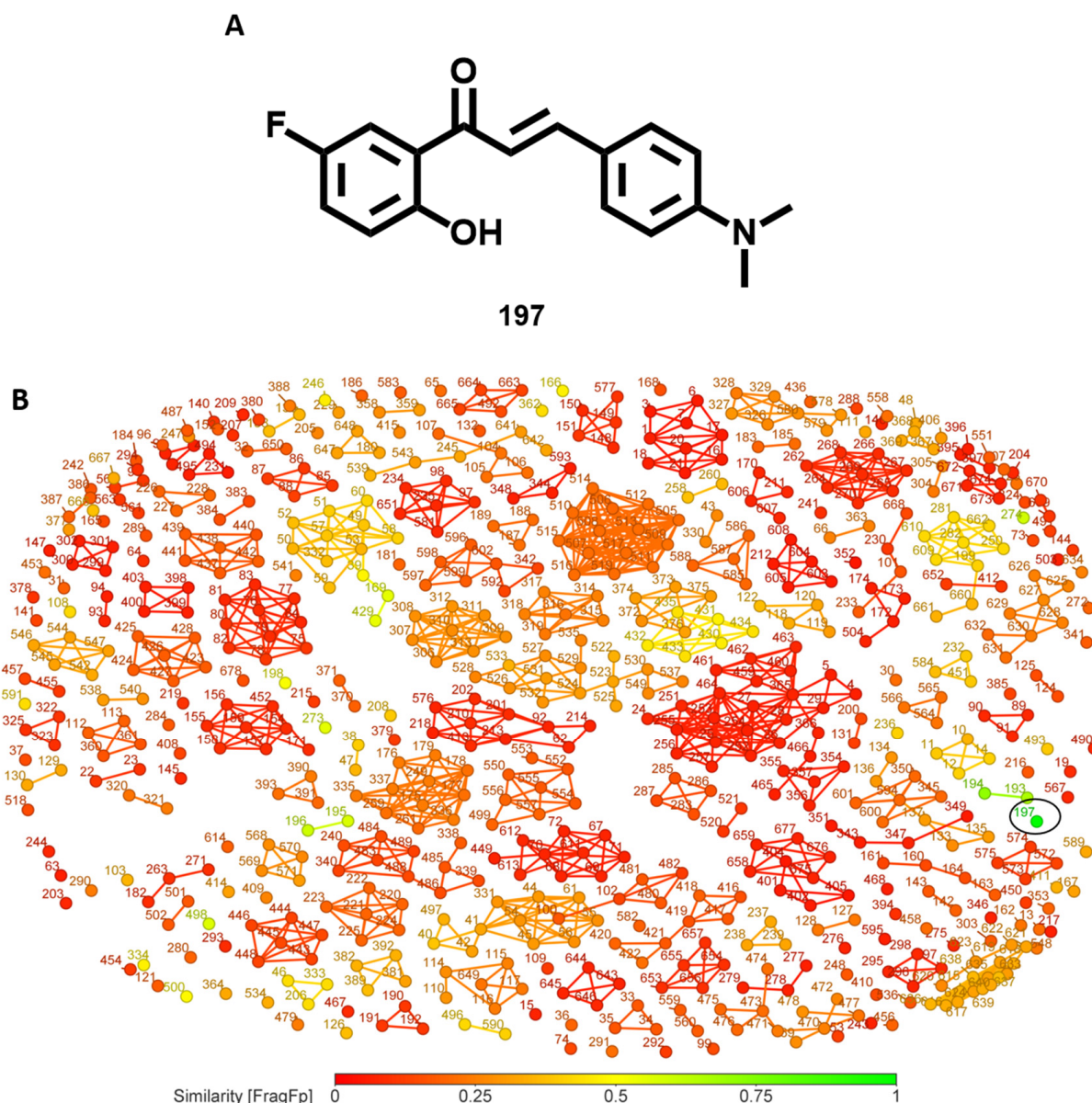


Figure 12. (A) Structure of CHT (97), ERGT (98), sesquiterpene (126),  $\beta$ -sitosterol (234) and stigmasterol (235), (B) similarity network for compound 97.

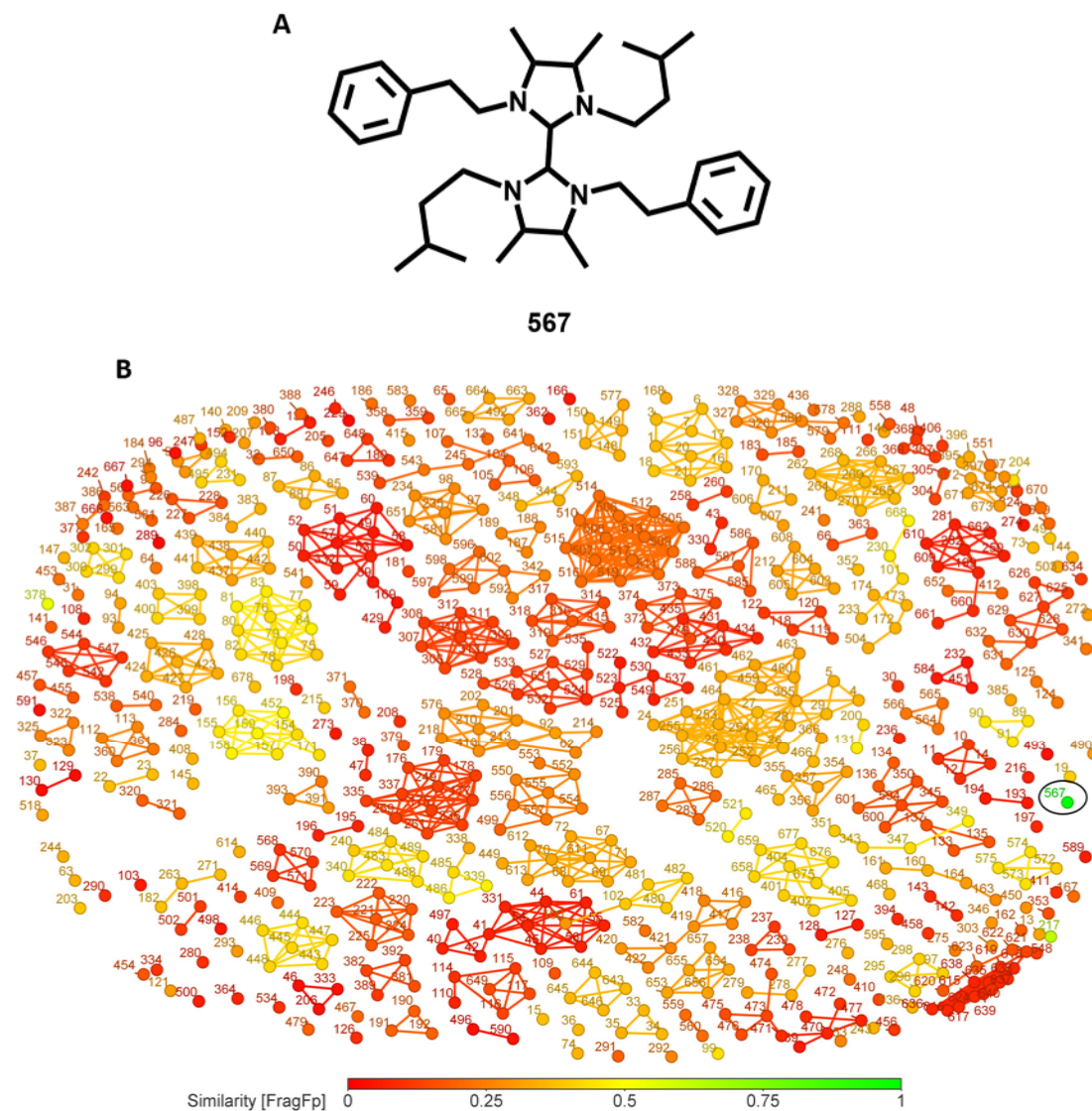


From a series of chalcone/flavonoid derivatives (193–199, Supporting Information) with potent activity against *Leishmania*, [35] compound 197 (Figure 13A) was selected as the reference for the similarity analysis. This compound, with an  $IC_{50}$  of 500 nM (against *L. amazonensis*), possesses rather low structural similarity (<60%) compared to the rest of the compounds in the database (Figure 13B). Given its drug-like properties and high anti-*Leishmania* potency, compound 197 could be used for similarity searches in other compound databases, providing good starting points for SAR studies.



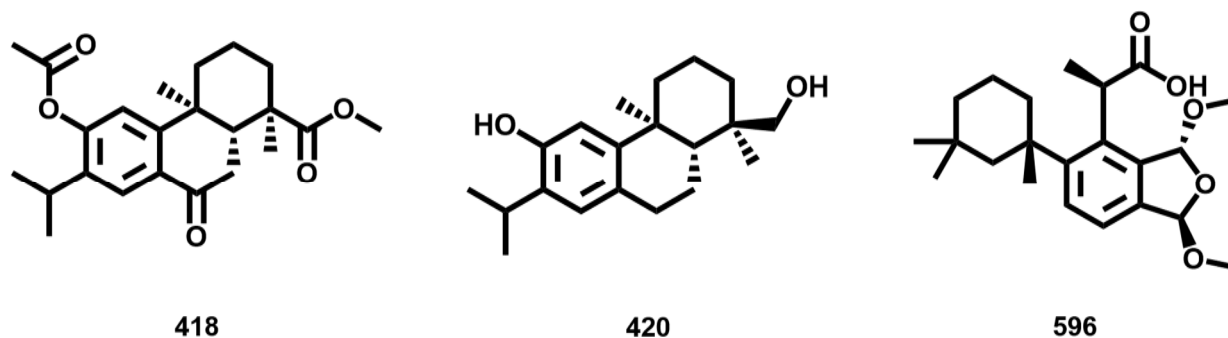
**Figure 13.** (A) Structure of compound 197, a chalcone derivative, (B) similarity chart for compound 197.

The imidazole alkaloid 567 (Figure 14A) isolated from the bacteria *Paenibacillus* sp., presented an  $IC_{50}$  of 750 nM against *L. major*. The majority of the dataset presents a similarity below 40% with this structure and no other structure in the dataset was linked in the similarity cluster (Figure 14B). Thus, compound 567 has a promising potential for SAR exploration given that no analogues were identified and tested for *L. major*.



**Figure 14.** (A) Structure of compound 567 (B) similarity chart for compound 567.

Both diterpenes **418** and **420** (Figure 15) extracted from *Abies* genus showed  $IC_{50}$  values of 700 nM against *L. infantum*. The diterpene **596** (Figure 15), extracted from the marine species *Dendrilla antarctica* has an  $IC_{50}$  of 800 nM against *L. donovani*. Similarity with most of the dataset compounds is below 50% for these structures, which can represent suitable starting points for future SAR exploration.



**Figure 15.** Structure of drug-like terpenoids **418**, **420**, and **596**.

#### 4. Conclusions

The development of new bioactive molecules is facilitated by a deeper understanding of the relevant chemical space, which can improve the success rate of drug discovery efforts. Chagas disease, HAT, and leishmaniasis are NTDs for which innovation is urgently needed for the next generation of drugs. In this work, compounds from diverse natural sources have been collected through an in-depth survey of the recent literature. A chemoinformatic analysis of the 678 compounds identified several promising hits and lead candidates for anti-trypanosomatid drug discovery. Plants, the primary source of NPs explored in drug research, were the most abundant source of bioactive compounds. The secondary metabolites that were investigated comprise six major regions of structural similarity: benzopyrans, oxydibenzene, flavonoids, steroids, benzoic acids, and cumarin derivatives. This finding shows that a few chemical classes can be privileged structural scaffolds for anti-trypanosomatid drug discovery. The steroid metabolic class was the most promising for both anti-*Trypanosoma* and anti-*Leishmania* activity. Peptides also showed promising anti-*Trypanosoma* activity, however, it was underexplored for *Leishmania*. Studies involving NPs with anti-*T. cruzi* and anti-*Leishmania* properties are more prevalent than those with *T. brucei* activity. These data reveal a tendency of the research community toward trypanosomatid diseases that are widely present across different regions of the world. Considering the lack of innovation in the NTD pharmaceutical pipeline, the results reported in this work provide valuable information to guide further NP-based drug discovery efforts on trypanosomatid diseases.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biom14081033/s1>.

**Author Contributions:** M.V., T.H.D., E.M., L.L.G.F., J.L.M.-F. and A.D.A. performed the analyses, wrote and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Sao Paulo Research Foundation (FAPESP) grants #2020/11967-3 (DFG/FAPESP), #2022/08333-8 (DAAD/FAPESP), #2013/07600-3 (CIBFar-CEPID), #2014/50926-0, #465637/2014-0 (INCT BioNat CNPq/FAPESP), National Council for Scientific and Technological Development (CNPq), and Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil. The authors acknowledge the scholarship conferred to MV: FAPESP #2019/05967-3 and CNPq #382044/2023-1.

**Data Availability Statement:** All data generated in this study can be found in the Supporting Information.

**Conflicts of Interest:** The authors declare no conflicts of interest.

#### References

1. Ferreira, L.L.G.; Andricopulo, A.D. Drugs and vaccines in the 21st century for neglected diseases. *Lancet Infect. Dis.* **2019**, *19*, 125–127. [CrossRef]
2. World Health Organization. Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases, 2021–2030. Available online: <https://www.who.int/publications/i/item/9789240010352> (accessed on 1 June 2024).
3. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. Available online: <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf> (accessed on 1 June 2024).
4. World Health Organization. Global Report on Neglected Tropical Diseases. 2024. Available online: <https://www.who.int/teams/control-of-neglected-tropical-diseases/global-report-on-neglected-tropical-diseases-2024> (accessed on 1 June 2024).
5. Ferreira, L.L.G.; de Moraes, J.; Andricopulo, A.D. Approaches to advance drug discovery for neglected tropical Diseases. *Drug Discov. Today* **2022**, *27*, 2278–2287. [CrossRef] [PubMed]
6. Mukherjee, S. The United States Food and Drug Administration (FDA) regulatory response to combat neglected tropical diseases (NTDs): A review. *PLoS Negl. Trop. Dis.* **2023**, *17*, e0011010. [CrossRef] [PubMed]



7. De Rycker, M.; Wyllie, S.; Horn, D.; Read, K.D.; Gilbert, I.H. Anti-trypanosomatid drug discovery: Progress and challenges. *Nat. Rev. Microbiol.* **2023**, *21*, 35–50. [CrossRef]
8. World Health Organization. Chagas Disease (Also Known as American Trypanosomiasis). Available online: [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)) (accessed on 1 June 2024).
9. Nambala, P.; Mulindwa, J.; Chammudzi, P.; Senga, E.; Lemelani, M.; Zgambo, D.; Matovu, E.; MacLeod, A.; Musaya, J. Persistently High Incidences of *Trypanosoma brucei rhodesiense* Sleeping Sickness With Contrasting Focus-Dependent Clinical Phenotypes in Malawi. *Front. Trop. Dis.* **2022**, *3*, 824484. [CrossRef]
10. Drugs for Neglected Diseases Initiative. R&D Portfolio June 2023—12 Treatments Delivered. Available online: <https://dndi.org/wp-content/uploads/2023/07/DNDi-RD-Portfolio-June-2023.pdf> (accessed on 1 June 2024).
11. Ntie-Kang, F.; Nyongbela, K.D.; Ayimele, G.A.; Shekfeh, S. “Drug-likeness” properties of natural compounds. *Phys. Sci. Rev.* **2019**, *4*, 20180169. [CrossRef]
12. SciFinder-n. Chemical Abstracts Service. Available online: [https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder-n?gad\\_source=1](https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder-n?gad_source=1) (accessed on 1 June 2024).
13. Clarivate. Web of Science. Available online: <https://www.webofscience.com/wos/woscc/smart-search> (accessed on 1 June 2024).
14. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef] [PubMed]
15. Djoumbou, F.Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated Chemical Classification With A Comprehensive, Computable Taxonomy. *J. Cheminform.* **2016**, *8*, 61. [CrossRef]
16. Rácz, A.; Bajusz, D.; Héberger, K. Life beyond the Tanimoto coefficient: Similarity measures for interaction fingerprints. *J. Cheminform.* **2018**, *10*, 48. [CrossRef]
17. 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]
18. Artusi, R.; Verderio, P.; Marubini, E. Bravais-Pearson and Spearman correlation coefficients: Meaning, test of hypothesis and confidence interval. *Int. J. Biol. Markers* **2002**, *17*, 148–151. [CrossRef]
19. Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I.H.; Frearson, J.; Wyatt, P.G. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem* **2008**, *3*, 435–444. [CrossRef]
20. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25. [CrossRef]
21. Méndez-Lucio, O.; Medina-Franco, J.L. The many roles of molecular complexity in drug discovery. *Drug Discov. Today* **2017**, *22*, 120–126. [CrossRef] [PubMed]
22. Greenacre, M.; Groenen, P.J.F.; Hastie, T.; D’Enza, A.I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100. [CrossRef]
23. Kusaka, Y.; Hasegawa, T.; Kaji, H. Noise Reduction in Solid-State NMR Spectra Using Principal Component Analysis. *J. Phys. Chem. A* **2019**, *123*, 10333–10338. [CrossRef]
24. Miller, R.R.; Madeira, M.; Wood, H.B.; Geissler, W.M.; Raab, C.E.; Martin, I.J. Integrating the Impact of Lipophilicity on Potency and Pharmacokinetic Parameters Enables the Use of Diverse Chemical Space during Small Molecule Drug Optimization. *J. Med. Chem.* **2020**, *63*, 12156–12170. [CrossRef]
25. Júnior, C.d.O.R.; Martinez, P.D.G.; Ferreira, R.A.A.; Koovits, P.J.; Soares, B.M.; Ferreira, L.L.; Michelan-Duarte, S.; Chelucci, R.C.; Andricopulo, A.D.; Matheussen, A.; et al. Hit-to-lead optimization of a 2-aminobenzimidazole series as new candidates for chagas disease. *Eur. J. Med. Chem.* **2023**, *246*, 114925. [CrossRef]
26. Barrett, J.A.; Yang, W.; Skolnik, S.M.; Belliveau, L.M.; Patros, K.M. Discovery solubility measurement and assessment of small molecules with drug development in mind. *Drug Discov. Today* **2022**, *27*, 1315–1325. [CrossRef]
27. O’Donovan, D.H.; De Fusco, C.; Kuhnke, L.; Reichel, A. Trends in Molecular Properties, Bioavailability, and Permeability across the Bayer Compound Collection. *J. Med. Chem.* **2023**, *66*, 2347–2360. [CrossRef]
28. Ali, J.; Camilleri, P.; Brown, M.B.; Hutt, A.J.; Kirton, S.B. In silico prediction of aqueous solubility using simple QSPR models: The importance of phenol and phenol-like moieties. *J. Chem. Inf. Model.* **2012**, *52*, 2950–2957. [CrossRef] [PubMed]
29. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [CrossRef] [PubMed]
30. Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752–6756. [CrossRef] [PubMed]
31. Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280. [CrossRef] [PubMed]
32. Bernatchez, J.A.; Kil, Y.S.; da Silva, E.B.; Thomas, D.; McCall, L.I.; Wendt, K.L.; Souza, J.M.; Ackermann, J.; McKerrow, J.H.; Cichewicz, R.H.; et al. Identification of Leucenostatin from *Ophiocordyceps* sp. as Antiparasitic Agents against *Trypanosoma cruzi*. *ACS Omega* **2022**, *7*, 7675–7682. [CrossRef] [PubMed]
33. Chaudhuri, M.; Singha, U.K.; Vanderloop, B.H.; Tripathi, A.; Nes, W.D. Steroidal Antimetabolites Protect Mice against *Trypanosoma brucei*. *Molecules* **2022**, *27*, 4088. [CrossRef] [PubMed]

34. Amang À Ngnoung, G.A.; Sidjui, L.S.; Leutcha, P.B.; Nganso Ditchou, Y.O.; Tchokouaha, L.R.Y.; Herbette, G.; Baghdikian, B.; Kowa, T.K.; Soh, D.; Kemzeu, R.; et al. Antileishmanial and Antiplasmodial Activities of Secondary Metabolites from the Root of *Antrocaryon klaineianum* Pierre (Anacardiaceae). *Molecules* **2023**, *28*, 2730. [[CrossRef](#)]
35. Lourenço, E.M.G.; Di Iório, J.F.; da Silva, F.; Fialho, F.L.B.; Monteiro, M.M.; Beatriz, A.; Perdomo, R.T.; Barbosa, E.G.; Oses, J.P.; de Arruda, C.C.P.; et al. Flavonoid Derivatives as New Potent Inhibitors of Cysteine Proteases: An Important Step toward the Design of New Compounds for the Treatment of Leishmaniasis. *Microorganisms* **2023**, *11*, 225. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.