Navigating the Chemical Space and Chemical Multiverse of a Unified Latin American Natural Product Database: LANaPDB

Alejandro Gómez-García[1], Daniel A. Acuña Jiménez[2], William J. Zamora[2,3,4], Haruna L. Barazorda-Ccahuana[5], Miguel Á. Chávez-Fumagalli[5], Marilia Valli[6], Adriano D. Andricopulo[6], Vanderlan da S. Bolzani[7], Dionisio A. Olmedo[8], Pablo N. Solís[8], Marvin J. Núñez[9], Johny R. Rodríguez Pérez[10,11], Hoover A. Valencia Sánchez[10], Héctor F. Cortés Hernández[10], José L. Medina- Franco[1*]

[1] DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico.
[2] CBio3 Laboratory, School of Chemistry, University of Costa Rica, San Pedro, San José 11501-2060, Costa Rica.
[3] Laboratory of Computational Toxicology and Artificial Intelligence (LaToxCIA), Biological Testing Laboratory (LEBi), University of Costa Rica, San Pedro, San José 11501-2060, Costa Rica.
[4] Advanced Computing Lab (CNCA), National High Technology Center (CeNAT), Pavas, San José 1174-1200, Costa Rica
[5] Computational Biology and Chemistry Research Group, Vicerrectorado de Investigación, Universidad Católica de Santa Maria, Arequipa, Peru.
[6] Laboratory of Medicinal and Computational Chemistry (LQMC), Centre for Research and Innovation in Biodiversity and Drug Discovery (CIBFar), São Carlos Institute of Physics (IFSC), University of São Paulo (USP), Av. João Dagnone, 1100, 13563-120, São Carlos, SP, Brazil.
[7] Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products (NuBBE), Department of Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Av. Prof. Francisco Degni, 55, 14.800-900, Araraquara, SP, Brazil.
[8] Center for Pharmacognostic Research on Panamanian Flora (CIFLORPAN), College of Pharmacy, Av. Manuel E. Batista and Jose De Fabrega, Panama City, Panama.
[9] Natural Product Research Laboratory, School of Chemistry and Pharmacy, University of El Salvador, Final Ave. Mártires Estudiantes del 30 de Julio, 01101, San Salvador, El Salvador.
[10] GIFES Research Group, School of Chemistry technology, Universidad Tecnológica de Pereira, Pereira, 660003, Colombia.
[11] GIEPRONAL Research Group, School of Basic Sciences, Technology and Engineering, Universidad Nacional Abierta y a Distancia, Dosquebradas, 661001, Colombia.

* Correspondence: medinajl@unam.mx; Tel. +52-55-5622-3899

Abbreviations: ADMET, absorption, distribution, metabolism, excretion, and toxicity; AI, artificial intelligence; CADD, computer-aided drug design; COCONUT, The Collection of Open Natural Products; HBA, hydrogen bond acceptors; HBD, hydrogen bond donors; GSK, GlaxoSmithKline's; HTS, high-throughput screening; LANaPDB, Latin American Natural Products Database; MW, molecular weight; NP, natural product; NPs, natural products; $NuBBE_{DB}$, Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products Database; Rb, rotatable bonds; Ro5, rule of 5; t-SNE, t-distributed stochastic neighbor embedding; TPSA, topological polar surface area; TMAP, tree MAP; VS, virtual screening.

Abstract

The number of databases of natural products (NPs) have increased substantially. Latin America is extraordinarily rich in biodiversity enabling the identification of novel NPs, which has encouraged both the development of databases and the implementation of those that are being created or are under development. In a collective effort from several Latin American countries, herein we introduce the first version of Latin American Natural Products Database (LANaPDB), a public compound collection that gathers the chemical information of NPs contained in diverse databases from this geographical region. The current version of LANaPD unifies the information from six countries and contains 12,959 chemical structures. The structural classification showed that the most abundant compounds are the terpenoids 63.2%, phenylpropanoids 18% and the alkaloids 11.8%. From the analysis of the distribution of properties of pharmaceutical interest, it was observed that many LaNaPDB compounds satisfy some drug-like rules of thumb for physicochemical properties. The concept of the chemical multiverse was employed to generate multiple chemical spaces from two different fingerprints and two dimensionality reduction techniques. Comparing LaNaPDB with FDA-approved drugs and the major open-access repository of NPs, COCONUT it was concluded that the chemical space covered by LaNaPDB completely overlaps with COCONUT and in some regions with FDA-approved drugs. LANaPD will be updated adding more compounds from each database plus the addition of databases from other Latin American countries. The database is freely available at https://github.com/alexgoga21/LaNaPDB.

Keywords: chemical multiverse, chemical space, chemoinformatics, databases, diversity, drug discovery, Latin America, natural products, virtual screening.

1. Introduction

Historically, natural products (NPs) have been the biggest source of bioactive compounds for medicinal chemistry. For instance, in cancer research, in the lapse of time 1946 to 1980, seventy-five small molecules were approved worldwide, of which 53% were unaltered NPs or natural product (NP) derivatives. Moreover, from 1981 to 2019, of the 185 small molecules approved to treat cancer, 64.9% were unaltered NPs and synthetic drugs with a NP pharmacophore [1]. Another example is the actual development of new promising

antibiotics against drug-resistant bacteria from NPs [2]. Furthermore, in a recent review it was shown that 697 natural steroidal alkaloids were isolated and characterized with various biological activities, from 1926 to 2021 [3]. The bioactive compounds encompass marine [4,5], fungal [6,7], bacteria [8], plants [9] and endogenous substances produced by humans and animals, sources [10], including venoms and poisons produced by different animals [11]. Even, as recently reviewed, the fruit peels are a source of bioactive compounds that in many instances display better biological and pharmacological applications than the compounds of other sections of the fruit [12].

To date, the discovery process of more than seventy commercialized drugs has included the rational use of at least a computational method [13]. Computer-aided drug design (CADD) has the potential to reduce the billionaire cost and decrease the time through the drug design process, e.g., the hit identification rate for high-throughput screening (HTS) to discover novel inhibitors for the enzyme protein tyrosine phosphatase-1B is only 0.021% and the one for molecular docking is 34.8% [14]. A crucial resource in CADD are the databases of chemical compounds including NP databases. From the compound databases, it is possible to identify potential hit molecules through several virtual screening (VS) techniques [15,16], including the training of artificial intelligence (AI) algorithms [17]. When the compound databases are annotated with biological activity (or other property of relevance), it is possible to use the data to perform structure-activity (property) relationships and develop predictive models. From 2003 to 2018, 104 research articles reported the identification of potential drug candidates from NP databases by using computational tools [18].

Between 2000 and 2019, one-hundred twenty-three commercial and public NP databases have been published. Among them, ninety-eight are still somehow accessible (online or under request access), ninety-two are free access, and only fifty contain molecular structures that can be retrieved for a chemoinformatic analysis [19]. Examples of the most representative open-access NP databases include: The Collection of Open Natural Products (COCONUT) [20] which is a major repository containing more than 411,000 NPs collected from 50 open access NP databases. The Universal Natural Product Database [21] is a compilation that tries to gather all the known NPs; it has more than 229,000 NPs. It is not yet accessible through the link in the original publication, nevertheless, it is contained and maintained on the ISDB website [22]. SuperNatural Ⅱ [23] database contains over 325,000 NPs, nonetheless, it does not provide a bulk download. ZINC [24] database has over 80,000 NPs, approximately 48,000 purchasable. Moreover, it contains some NP

databases that are no longer accessible through the link provided in the original publication, e.g., Herbal Ingredient Targets [25] and Herbal Ingredients in vivo Metabolism database [26], which contain NPs mostly from Chinese plants. Moreover, there are NP databases which contain compounds isolated and characterized in certain geographical areas. That is the case of China, where have been published multiple compound databases containing only NP of this country [27–33], nevertheless, TCM@Taiwan [34] is the largest, which contains 58,000 compounds. There are two databases of NPs from India, IMPPAT [35] composed of approximately 10,000 phytochemicals extracted from 1,700 medicinal plants, and MedPServer [36], containing 1,124 NPs. Regarding NPs from Africa, there are several NP databases [37–42], nonetheless, AfroDB [43] is the most extensive, containing over one thousand NPs. Recently, was published Phyto4Health [44], a NP database with 3,128 NPs isolated from medicinal plants of Russia.

Latin America contains at least a third of the global biodiversity [45], in fact, half of the countries have been classified as megadiverse: Bolivia, Brazil, Colombia, Costa Rica, Ecuador, Mexico, Peru and Venezuela) [46]. Therefore, Latin America represents a large source of bioactive molecules and potential drug candidates (Figure 1). There have been published databases containing NPs from some Latin American countries such as NaturAr [47] (Argentina), $NuBBE_{DB}$ [48,49], SistematX [50,51], UEFS [52] (Brasil), CIFPMA [53,54] (Panama), PeruNPDB [55], (Peru), UNIIQUIM [56] and BIOFACQUIM [57,58] (Mexico). Recently, it was reviewed the present state of the art in developing Latin American NP databases and their practical applications to the drug discovery area [59]. Multiple drug candidates have been identified from the Latin American NP databases as therapeutic agents for diseases caused by infectious agents (Chagas disease [60,61], tuberculosis [62], Leishmaniasis [63,64], schistosomiasis [65], coronavirus disease [66], human immunodeficiency virus infection and acquired immunodeficiency syndrome, hepatitis B and C) [67], pain [68], obesity, diabetes, hyperlipoproteinemia, cancer, and age-related diseases [69,70].

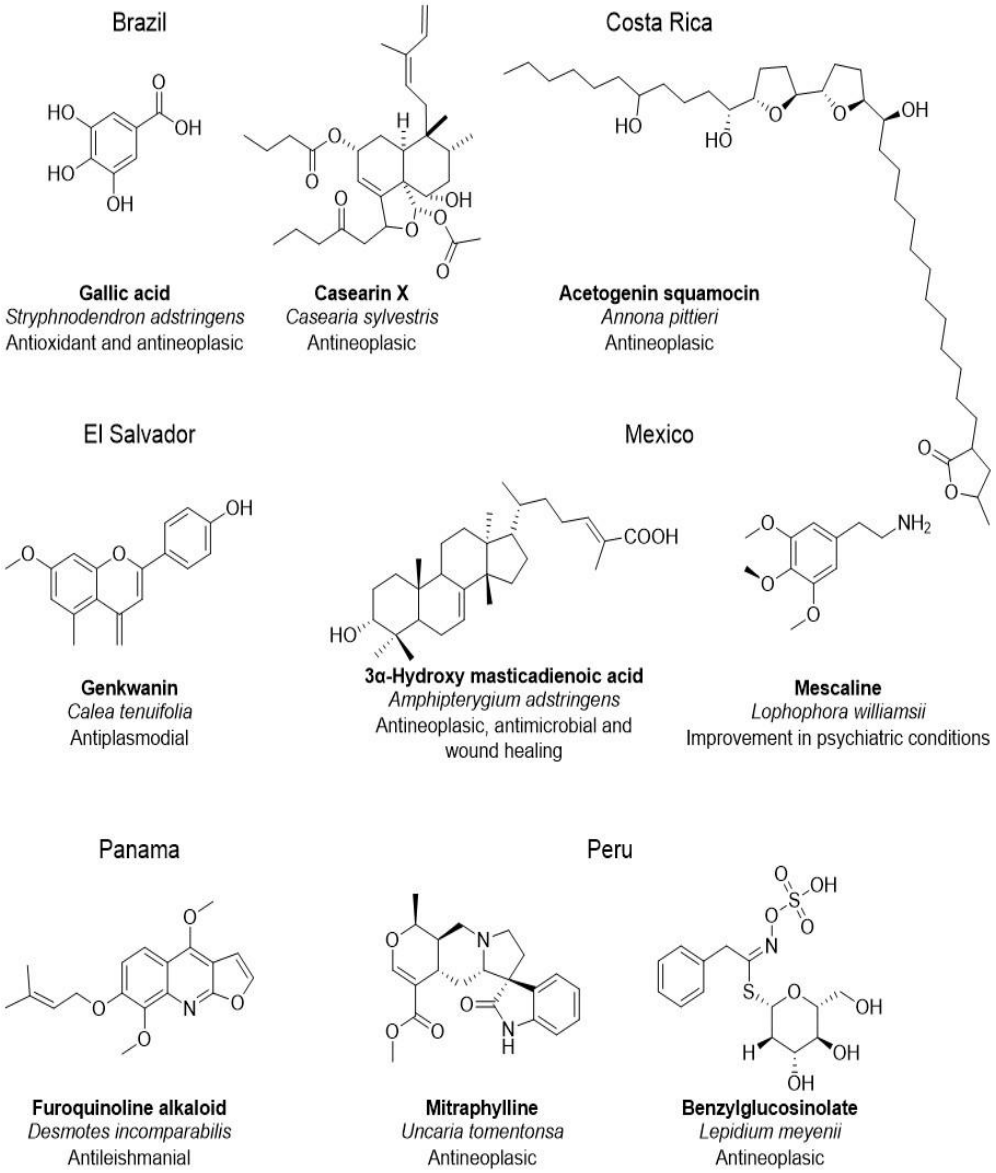

Figure 1. Active compounds of representative medicinal plants from Latin American countries and some of its therapeutic effects described in the literature. Brazil (gallic acid [71] and casearin x [72]), Costa Rica (acetogenin squamocin [73]) El Salvador (Genkwanin [74]), Mexico (3α-hydroxy masticadienoic acid [75] and mescaline [76]), Panama (furoquinoline alkaloid [77]) and Peru (mitraphylline [78] and benzylglucosinolate [79]).

The long-term goal of the project is to collect, unify, and standardize the Latin American NP collections available in the public domain into one public database. In this study, we report significant advances towards this goal by the assembly of the first version of the unified database herein called Latin American Natural Products Database (LANaPD). We report its curation, standardization, and a comprehensive analysis of nine compound databases, totaling 12,959 unique molecules. As part of this study, analyzed the structural content (scaffolds and ring systems), structural diversity, and complexity of the compounds in LANaPD. We also represent coverage in the chemical space of compounds in LANaPD using the concept of chemical multiverse [80].

2. Results and Discussion

2.1. Dataset curation

From nine Latin American NP databases of six different countries (Table **1**) was constructed the first version of LANaPDB that currently contains 12959 compounds in total.

Table 1. Latin American natural product databases analyzed in this work.

Database	Number of compounds[a]	Country	Source	General description	Ref.
$NuBBE_{DB}$	2223	Brazil	Plants Microorganisms Terrestrial and marine animals	Natural products of Brazilian biodiversity. Developed by the São Paulo State University and the University of São Paulo.	[48],[49]
SistematX	9514	Brazil	Plants	Database composed of secondary metabolites and developed at the Federal University of Paraiba.	[50],[51]
UEFS	503	Brazil	Plants	Natural products that have been separately published, but there is no common publication nor public database for it. Developed at the State University of Feira de Santana.	[52]
NAPRORE-CR	359	Costa Rica	Plants Microorganisms	Developed in the CBio3 and LaToxCIA Laboratories of the University of Costa Rica.	*
LAIPNUDELSAV	214	El Salvador		Developed by the Research Laboratory in Natural Products of the University of El Salvador.	*
UNIIQUIM	1112	Mexico	Plants	Natural products isolated and characterized at the Institute of Chemistry of the National Autonomous University of Mexico.	[56]
BIOFACQUIM	553	Mexico	Plants Fungus Propolis Marine animals	Natural products isolated and characterized in Mexico at the School of Chemistry of the National Autonomous University of Mexico and other Mexican institutions.	[57],[58]
CIFPMA	363	Panama	Plants	Natural products that have been tested in over twenty-five *in vitro* and *in vivo* bioassays, for different therapeutic targets. Developed at the University of Panama.	[53],[54]
PeruNPDB	280	Peru	Animals Plants	Created and curated at the Catholic University of Santa Maria.	[55]

The URL of the websites where the natural product databases of Latin America are allocated is in the supplementary material (Table S3).

[a] Number of compounds contained in each database previous to the curation process.

*Actually, there is not a publication associated with the database.

2.2. Structural classification

The compounds were classified in a total of seven different pathways, fifty-three superclasses, and 336 classes (Figure 2). The three predominant pathways are terpenoids 63.2%, shikimates and phenylpropanoids 18% and alkaloids 11.8%. The main superclasses are diterpenoids 34.3%, sesquiterpenoids 17.6% and flavonoids 10.3%. The prevalent classes are kaurane and phyllocladane diterpenoids 6.99%, colensane and clerodane diterpenoids 5.91% and germacrane sesquiterpenoids 5.36%. The results are in accordance with expectations because the terpenoids are the most diverse group of secondary metabolites derived from natural sources [81].

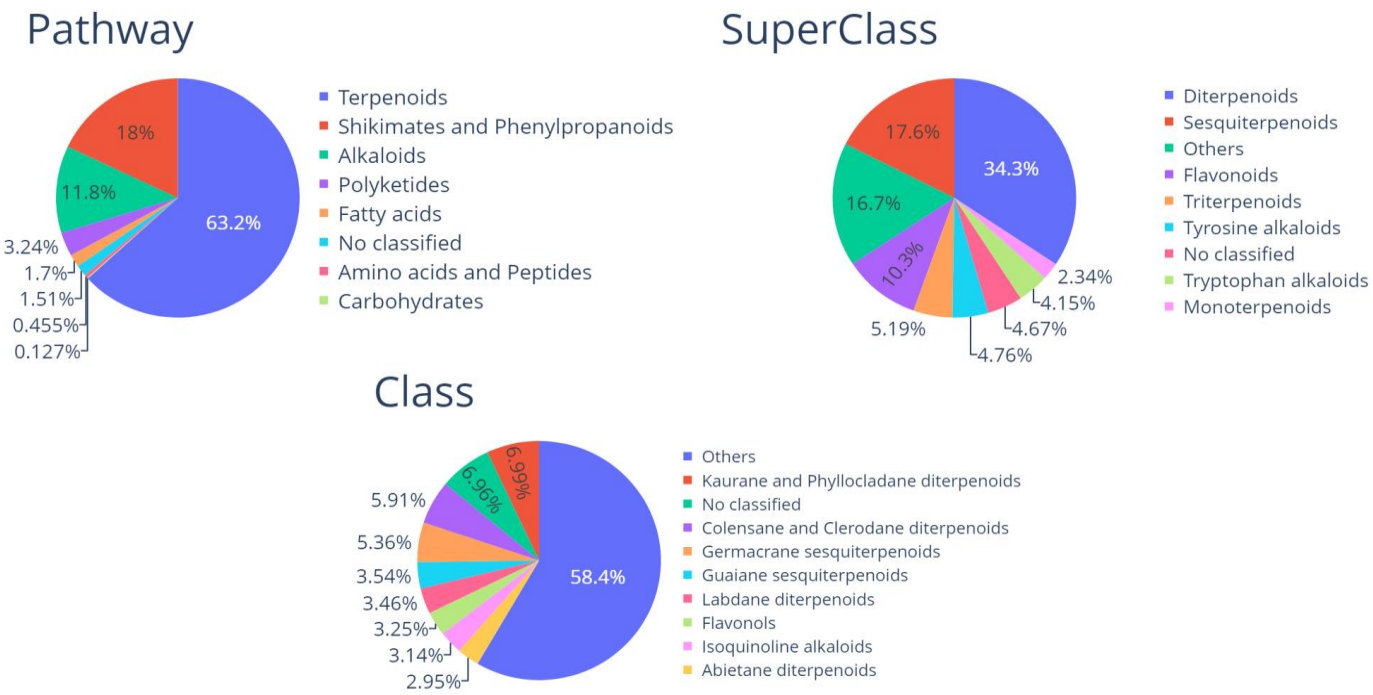

Figure 2. Structural classification of the compounds in the current (first) version of LANaPDB.

2.3. Physicochemical properties

The violin plots show the distribution of six physicochemical properties of pharmaceutical interest: SlogP [82], molecular weight (MW), topological polar surface area (TPSA) [83], rotatable bonds (Rb), hydrogen bond acceptors (HBA), and hydrogen bond donors (HBD) (Figures 3 and 4). In the violin plots is marked with a horizontal line the limits of the following rules of thumb of drug-likeness: Lipinski's rule of 5 (Ro5) [84,85], Veber's rules [86], GlaxoSmithKline's (GSK) 4/400 rule [87] and Pfizer 3/75 rule [88] (Table S1). Having physicochemical properties in the limits of either Lipinski's, Veber's or GSK rules is usually related with a good

oral bioavailability. The fulfillment of these rules of thumb is associated with the improvement of the following parameters: aqueous solubility and intestinal permeability (Lipinski's Ro5), passive membrane permeation (Veber's rules), absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile (GlaxoSmithKline's 4/400 rule) and toxicity (Pfizer 3/75 rule).

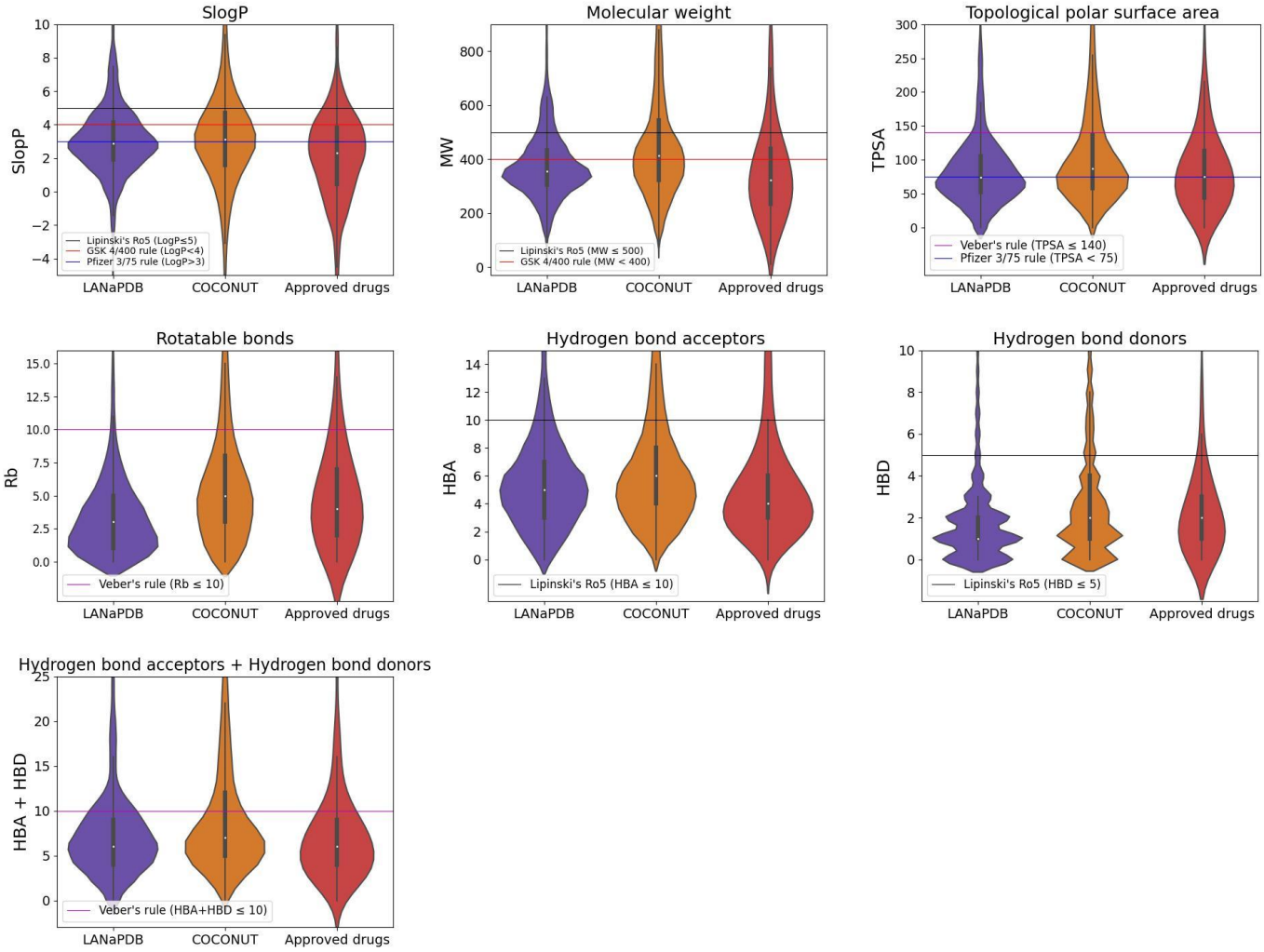

Figure 3. Violin plots summarizing the distribution of the representing physicochemical properties of pharmaceutical interest of the compounds of three databases LaNaPDB, COCONUT, and FDA-approved small molecule drugs.

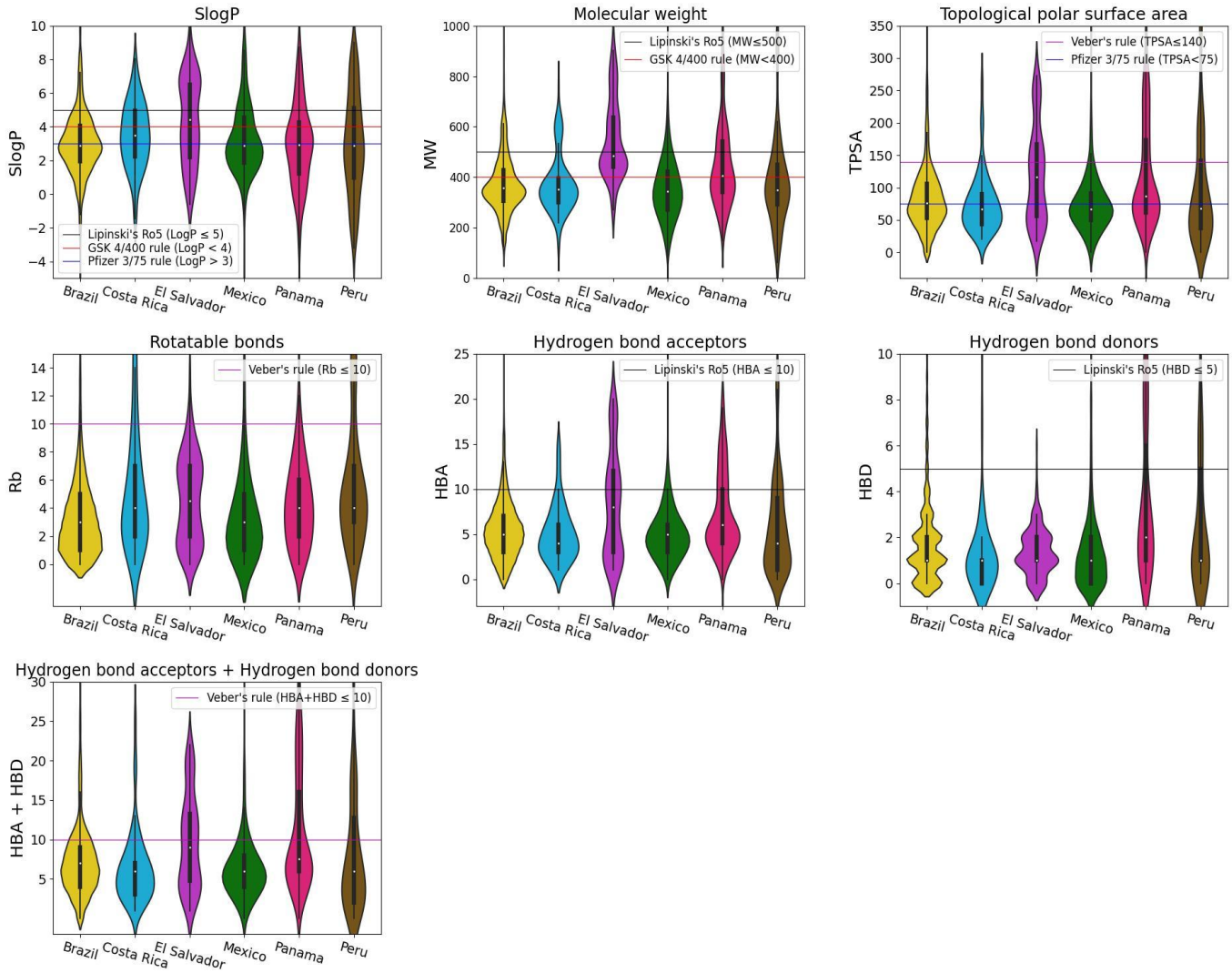

Figure 4. Violin plots summarizing the distribution of the physicochemical properties of pharmaceutical interest of the compounds in LaNaPDB. The databases that encompassess every country: Brazil ($NuBBE_{DB}$, SistematX and UEFS), Costa Rica (NAPRORE-CR), El Salvador (LAIPNUDELSAV), Mexico (UNIIQUIM and BIOFACQUIM), Panama (CIFPMA), and Peru (PeruNPDB).

NPs contain complex structures and are large and diverse, therefore, compared with synthetic drugs, it is not easy that they satisfy most of the criteria of Lipinski's Ro5 [89] or the other drug-likeness parameters mentioned above. Nevertheless, it is shown in the violin plots that a broad range of the LaNaPDB compounds satisfy most of the rules of thumb of Table S1 for the physicochemical properties of pharmaceutical interest. The LaNaPDB and COCONUT compound distribution of the physicochemical properties is in general similar (Figure 3). Also as expected, COCONUT covers the broadest area of the chemical space, because it is the largest database (411,000 compounds) (Figure 5). Many compounds of LaNaPDB fulfill the rules of thumb associated with drug-likeness (Figure 3) and part of the LaNaPDB chemical space overlaps with the chemical space comprised by the approved drugs (Figure 5).

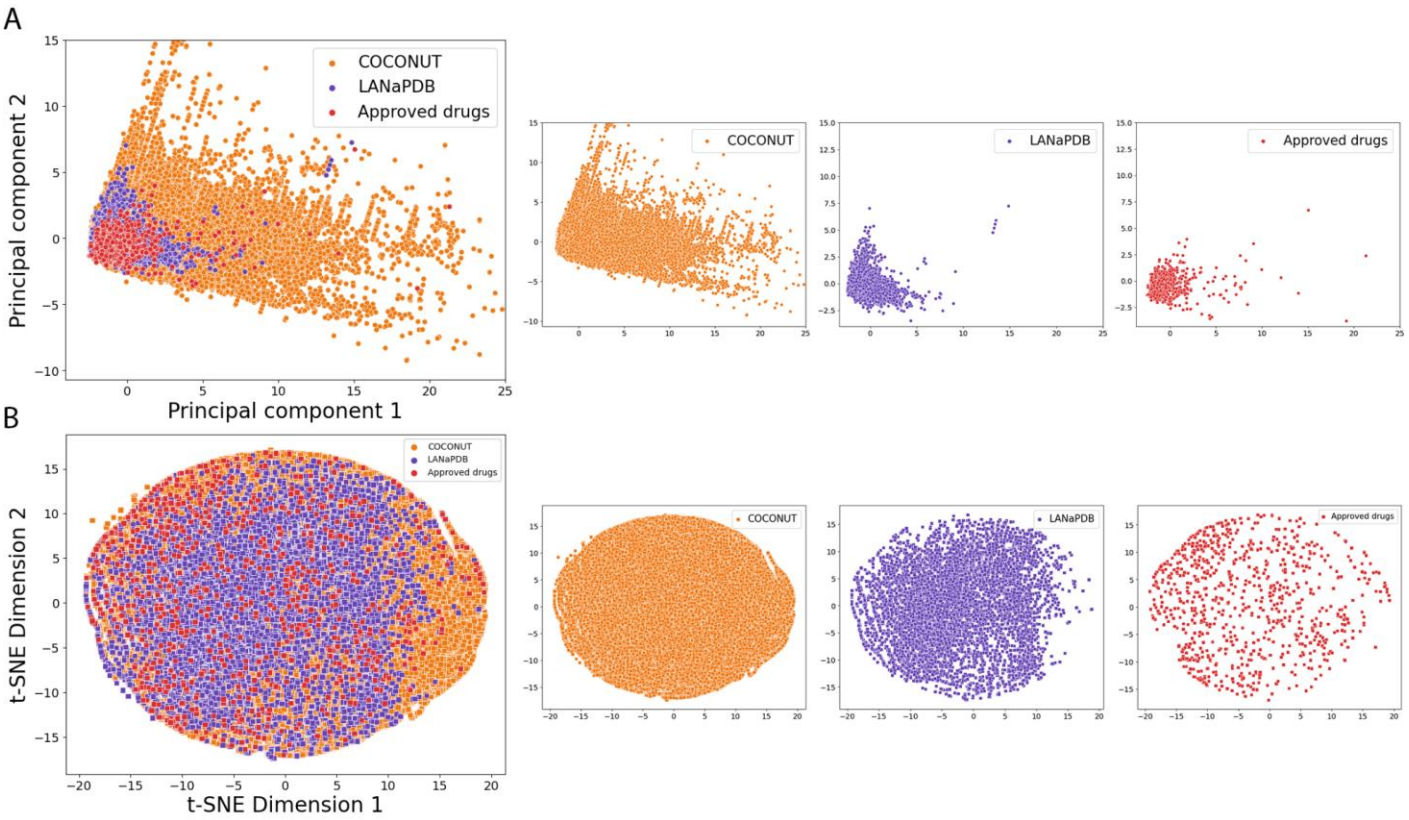

Figure 5. Visual representation of the chemical space based on six physicochemical properties of pharmaceutical interest of LaNaPDB and its comparison with COCONUT and approved drugs. The chemical space was generated with **A**) principal component analysis (PCA), the first two principal components capture 89.3% of the total variance; **B**) t-distributed stochastic neighbor embedding (t-SNE).

The distribution of the physicochemical properties of the NPs in the countries with more compounds (Brazil and Mexico) is in general, more focused in certain regions, compared with the NPs from countries with less compounds (Costa Rica, El Salvador, Panama, and Peru) which is broader (e.g. SlogP, Brazil vs Peru from Figure 4). The chemical space represented by the six physicochemical properties is overlapped among the NPs from the six Latin American countries (Figure 6). In the principal component analysis (PCA), the first two principal components are enough to represent most of the explained variance percentage: 89.3% in the LaNaPDB, COCONUT and approved drugs comparison and 84.6% in the Latin American countries' comparison (Table S2). Moreover, TPSA, MW, HBD and HBA are the descriptors with more contribution to the principal component 1. The descriptors with more contribution to the principal component 2 are SlogP and Rb.

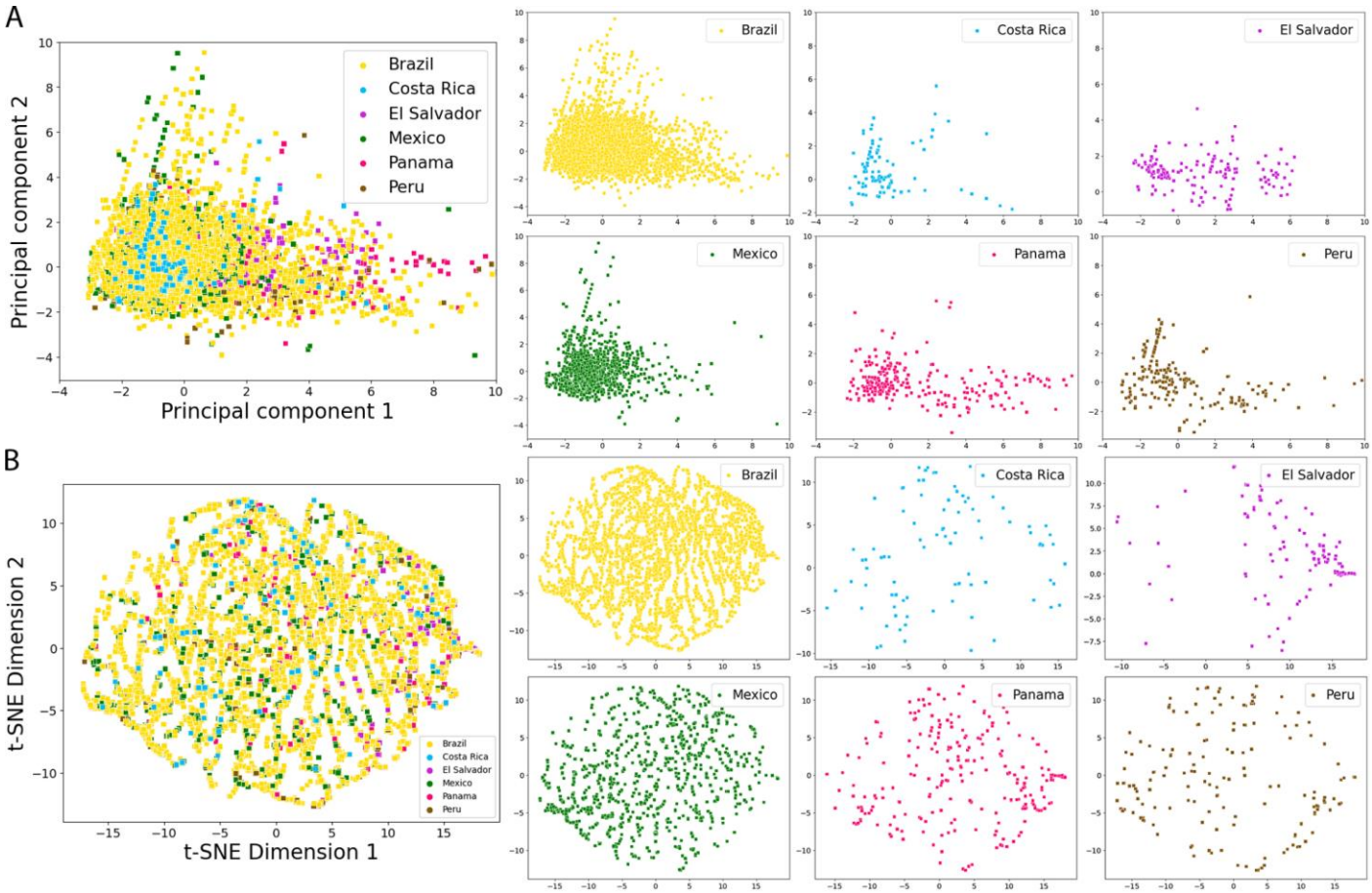

Figure 6. Visual representation of the chemical space based on six physicochemical properties of pharmaceutical interest of LaNaPDB and individual Latin American natural product databases. The chemical space was generated with **A**) principal component analysis (PCA) the first two principal components capture 84.6% of the total variance; **B**) t-distributed stochastic neighbor embedding (t-SNE).

2.4. Molecular fingerprints

Figures 7 and 8 show the visual representation of the chemical multiverse of LANaPDB generated with t-distributed stochastic neighbor embedding (t-SNE) and tree MAP (TMAP) [90] and two fingerprints of different design: MACCS keys (166-bits) (Figure 7A and Figure 8A) and MAP4 (Figure 7B and Figure 8B). As discussed recently, the chemical multiverse can be defined as a group of chemical spaces, each generated with a diverse set of descriptors [80]. A chemical multiverse is a natural extension of the concept of chemical space and its advantage is that it provides a more complete description of the chemical space of a set of compounds as opposed to using only one representation. Based on the visual representation of the chemical multiverse it is concluded that t-SNE has a better performance with MACCS keys (166-bits) fingerprint over MAP4 fingerprint, separating the NPs on clusters according to the structural features (Figure 7). The efficacy of TMAP to separate compounds in clusters from MACCS keys (166-bits) and MAP4 fingerprints is similar

with both fingerprints (Figure 8). Moreover, TMAP performed better than t-SNE in the NPs cluster creation with both fingerprints. A interactive version of the scatter plot created with TMAP from MAP4 fingerprints (Figure 8B) is freely available at https://github.com/alexgoga21/LaNaPDB/blob/main/Interactive%20TMAP_MAP4.html. To open the interactive map, download the file and open it in web explorer. Since TMAP performed better than t-SNE, and MACCS keys (166-bits) and MAP4 fingerprints showed a similar efficacy in the TMAP, the comparison of LaNaPDB with the reference databases was made with TMAP and MACCS keys (166-bits) fingerprint (Figure 9). It can be observed that LANaPDB overlaps with COCONUT in well-defined areas, nevertheless, the approved drugs are more dispersed and some of them overlap with compound in LANaPDB (Figure 9).

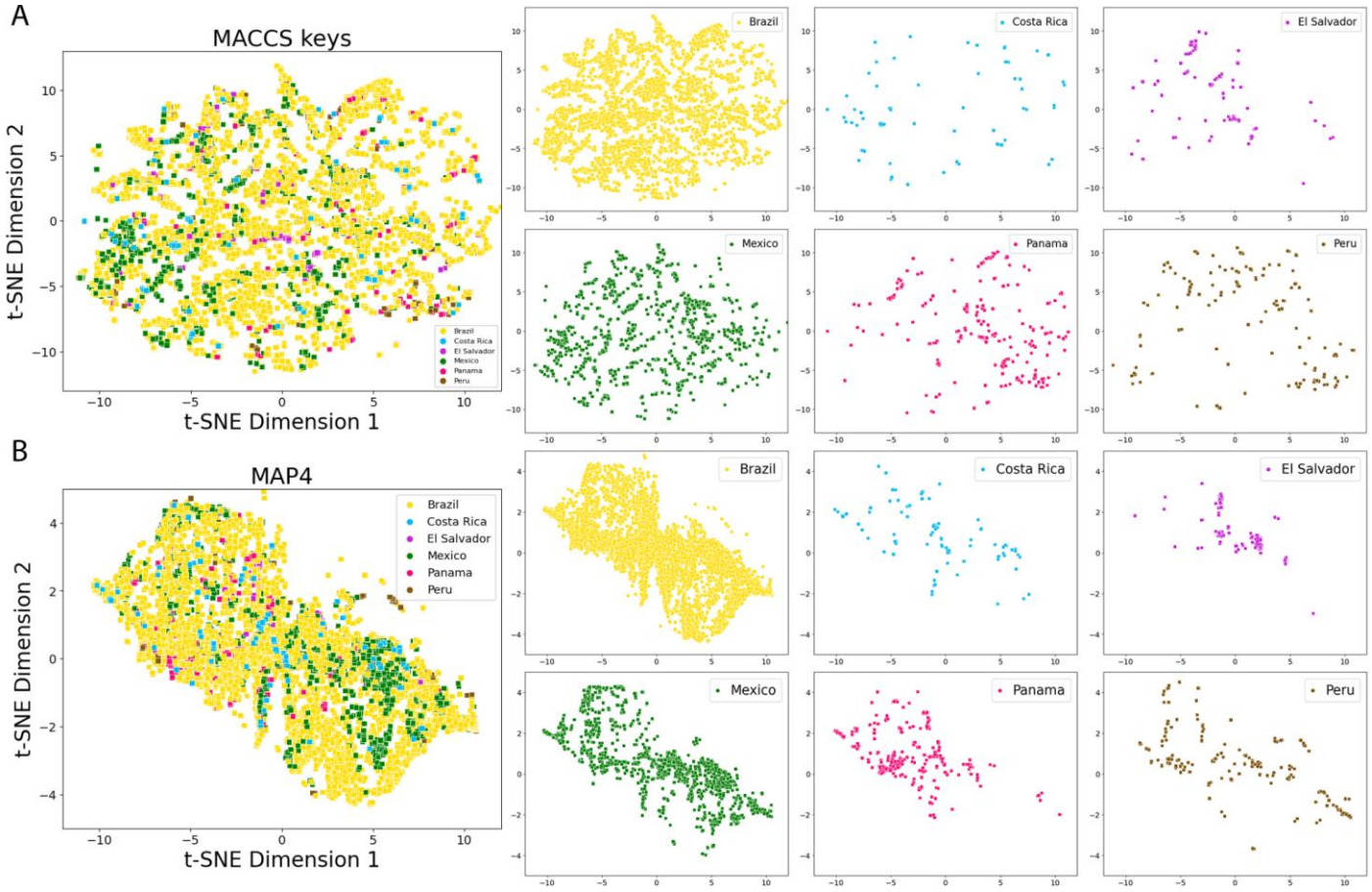

Figure 7. Visual representation of the chemical multiverse of LANaPDB and individual Latin American natural product databases. The chemical multiverse is a group of chemical spaces, each generated with a different set of descriptors. Chemical space comprised by **A**) (t-SNE)-MACCS keys (166-bits) fingerprint; **B**) (t-SNE)-MAP4 fingerprint.

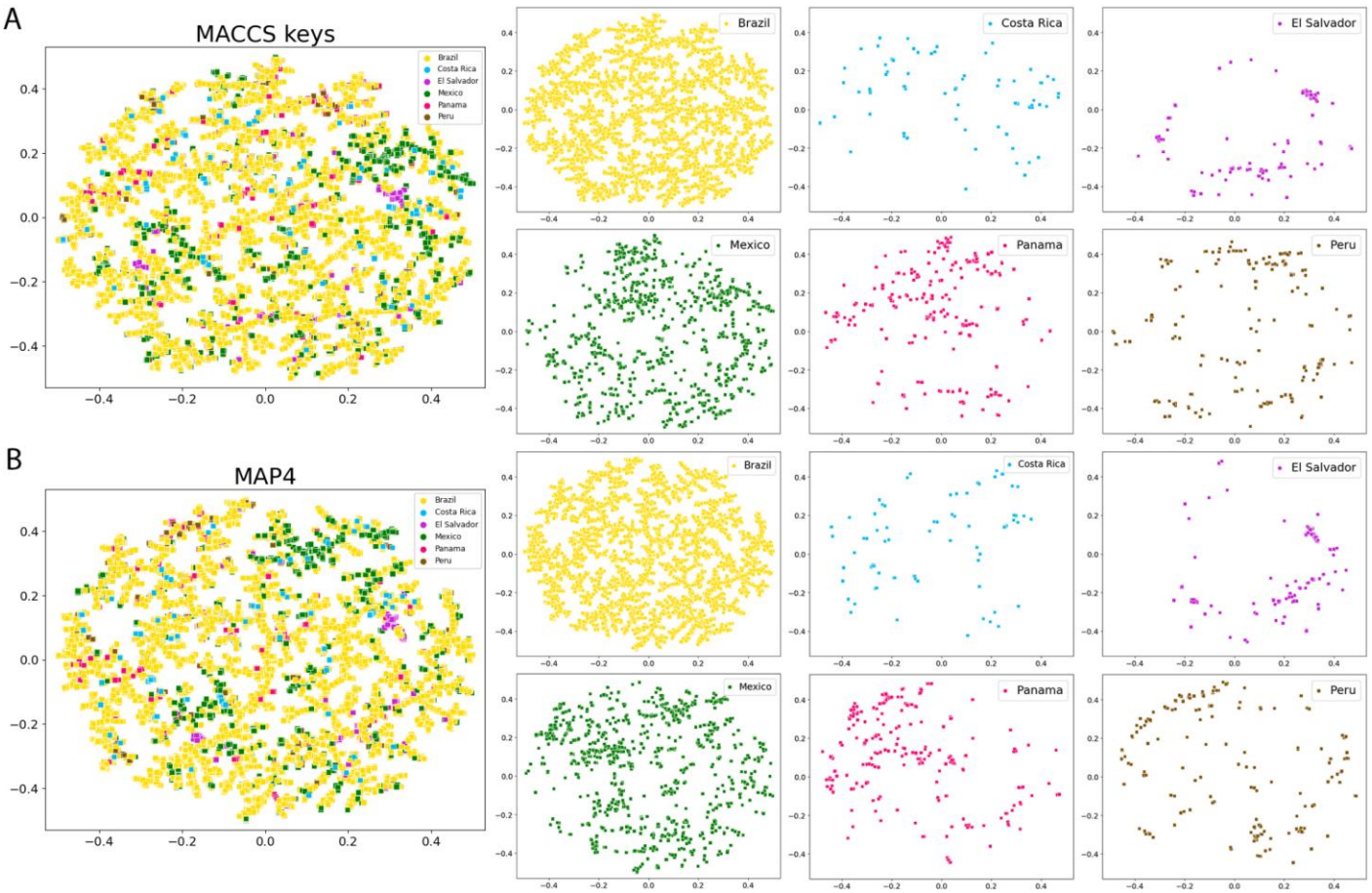

Figure 8. Visual representation of the chemical multiverse of LANaPDB and individual Latin American natural product databases. The chemical multiverse is a group of chemical spaces, each generated with a different set of descriptors. Chemical space comprised by **A**) (TMAP)-MACCS keys (166-bits) fingerprint; **B**) (TMAP)-MAP4 fingerprint (interactive version: https://github.com/alexgoga21/LaNaPDB/blob/main/Interactive%20TMAP_MAP4.html).

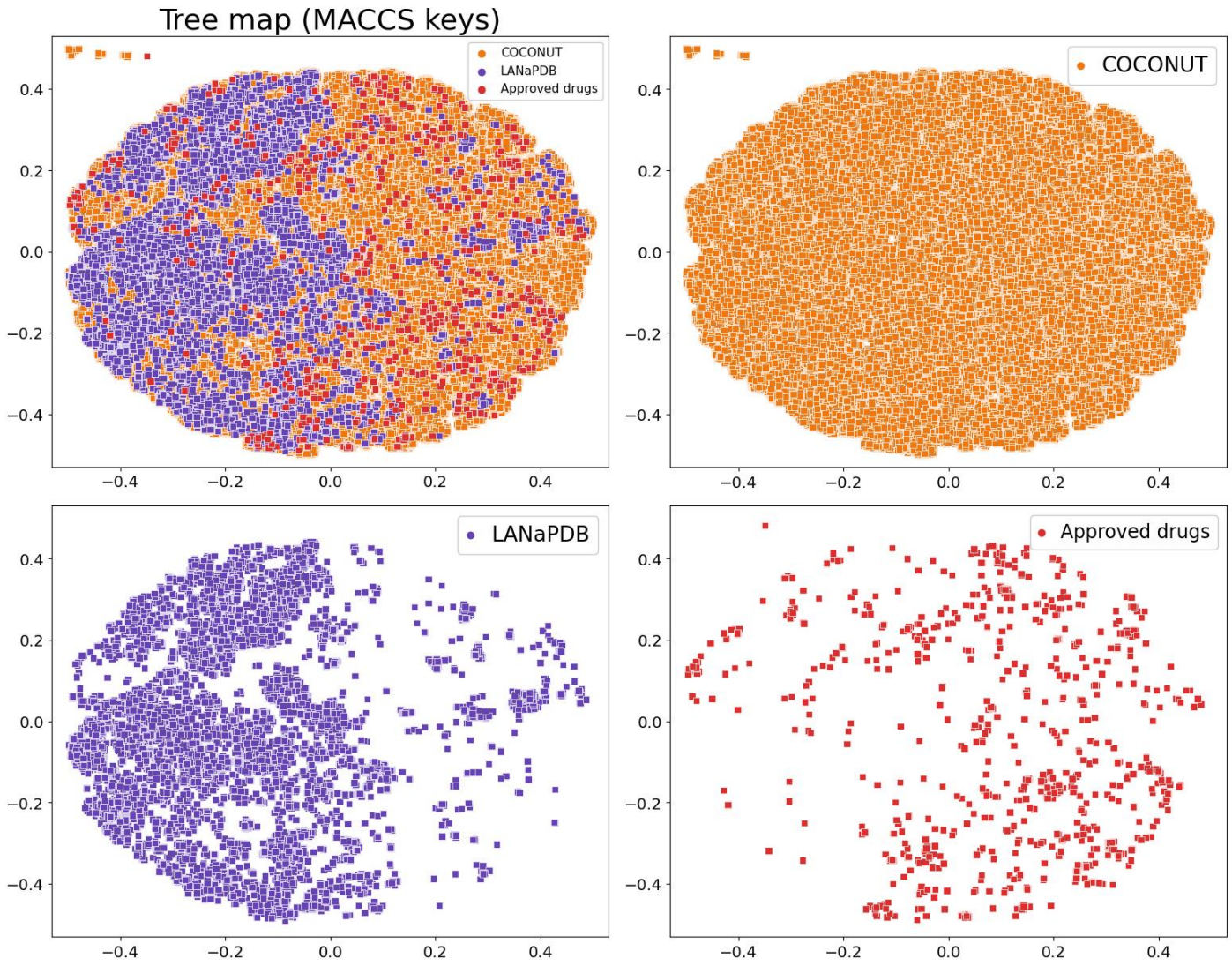

Figure 9. Tree map from MACCS keys (166-bits) of LaNaPDB and the comparison with COCONUT and approved drugs.

3. Materials and methods

The visual representations of the different chemical spaces that consider, either physicochemical properties or molecular fingerprints were illustrated with scatter plots (Figures 5-9). Every point in the scatter plots represents a unique compound. The scatter plots were created in the python programming language (version 3.10.7), employing the seaborn module (0.12.2) [91].

3.1. Dataset curation

The Latin American NP databases of Table **1** were used to construct the unified NP database LANaPDB. The process was carried out in the python programming language (version 3.10.7), employing the RDKit (version

2022.03.5) [92] and MolVS (version 0.1.1) [93] modules. The standardization process of MolVS was applied, which consist in the remotion of explicit hydrogens, disconnection of covalent bonds between metals and organic atoms, application of normalization rules (transformations to correct common drawing errors and standardization of functional groups), reionization (ensure the strongest acid groups protonate first in partially ionized molecules) and recalculation of the stereochemistry. The salts were removed, keeping the largest fragment, which was neutralized, and the remaining partially ionized fragments were reionized. The canonical tautomer was determined, and, from the InChIKey strings of the canonical tautomer, the duplicate compounds were removed.

3.2. Structural classification

Compounds in LANaPD were classified with NPClassifier [94] which is a freely available deep neural network-based structural classification tool for NPs. NPClassifier establishes a classification system based on the literature from the specialized metabolism of plants, marine organisms, fungi, and microorganisms. The categories used in NPClassifier are defined at three hierarchical levels: Pathway (nature of the biosynthetic pathway), Superclass (chemical properties or chemotaxonomic information), and Class (structural details).

3.3. Physicochemical properties

Employing the software KNIME [95] version 4.7.1, with the RDKit nodes, six physicochemical properties of pharmaceutical interest were calculated: SlogP [82], MW, TPSA [83], Rb, HBA, HBD. Violin plots were constructed to summarize the distribution of each property individually. In each violin plot we highlighted the limit of drug-like rules of thumb (Table S1). To generate a visual representation of the chemical space of the compound libraries based on the six properties, we reduced the data dimensionality to two dimensions employing PCA and t-SNE with the python module Scikit-learn version 1.2.2 [96]. PCA: principal component one and principal component two to represent the six physicochemical properties. t-SNE hyperparameters: perplexity=40 and number of iterations=300. The distribution of the individual properties and the two-dimensional representation of the chemical space were conducted to analyze and compare the properties of the NPs among the six Latin American countries and with two other reference datasets, COCONUT [20] and FDA-approved small molecule drugs version 5.1.10 (released by DrugBank in January 2023) [97].

3.4. Molecular fingerprints

A fingerprint encodes the structural information of a molecule in a vector [98]. Two different fingerprints were determined for each molecule: MACCS keys (166-bits) fingerprint and MAP4 fingerprint. MACCS keys (166-bits) fingerprints were calculated with KNIME [95] version 4.7.1, employing the Chemistry Development Kit (CDK) nodes [99]. MAP4 fingerprints were determined with the python programming language following the instructions of the creators of this fingerprint [100]. To allow a 2D representation of the molecules, two different techniques for dimensionality reduction were employed: t-SNE and TMAP [90]. For t-SNE, the same hyperparameters of section 2.3 were used. Employing the TMAP with MACCS keys (166-bits), LaNaPDB was compared with the two reference datasets used in section 2.3. From the MAP4 fingerprints an interactive TMAP was created in the python programming language (version 3.9.17) with the faerun module (version 0.4.2).

4. Conclusions

Here we report progress towards the assembly of the first version of a unified Latin American Natural Products Database. The current version has 12,959 compounds from nine compound databases of six different Latin American countries. The database is freely available and the information of each compound in this first version includes the structures in SMILES format, the structural classification and six physicochemical properties of pharmaceutical interest. The LaNaPDB compounds obtained in plants, terrestrial and marine animals, fungi and bacteria. Moreover, the most abundant NPs were the terpenoids 63.2%, followed by the shikimates and phenylpropanoids 18% and the alkaloids 11.8%. Although it is not easy that NPs fulfill most of the drug-likeness parameters compared with synthetic drugs, many LaNaPDB compounds satisfy some drug-like rules of thumb for physicochemical properties. Moreover, the chemical space covered by LaNaPDB completely overlaps with COCONUT and in some regions with the FDA-approved drugs. The concept of the chemical multiverse was used to generate multiple chemical spaces from two different dimensionality reduction techniques (t-SNE and TMAP) and two fingerprints (MACCS keys (166-bits) and MAP4). Besides, MAP4 performed better than t-SNE to separate the compounds in clusters according to their structural features. All the resources used for the assembly, curation, analysis and graphics creation are freely available.

LANaPDB is part of one of the strategic actions to contribute to the further development of chemoinformatics and related disciplines in Latin America and strengthen the interactions between Latin America and other geographical regions [101]. We encourage the community to visit the websites where the individual NP databases of the different Latin American countries are reported (Table S3).

We anticipate that LANaPDB will continue growing and evolving with the update of more compounds from each existing database plus the addition of databases from other Latin American countries. One of the first steps in this direction is the integration of a larger set of NAPRORE-CR and the incorporation of natural products database NPDB-EjeCol from Colombia. Another perspective is the implementation of the database in a free-web server. Likewise, LANaPD could be integrated with other large public databases of natural products such as COCONUT or LOTUS.

Supplementary material

Table S1: Rules of thumb - guides - associated with drug-likeness. **Table S2**: Analysis metrics of the principal component analysis. **Table S3**: Websites of the natural product databases of Latin America.

Author contributions

Conceptualization, J.L.M.F.; methodology, J.L.M.F., A.G.G.; software, J.L.M.F.; validation, W.J.Z., M.A.C.F., D.A.O.; formal analysis, J.L.M.F., A.G.G.; investigation, A.G.G., W.J.Z., H.L.B.C, M.A.C.F., M.V., A.D.A., V.S.B., D.A.O., P.N.S, M.J.N., J.R.R.P, H.A.V.S, H.F.C.H, J.L.M.F.; resources, J.L.M.F.; data curation, A.G.G.; writing—original draft preparation, J.L.M.F., A.G.G.; writing—review and editing, A.G.G., W.J.Z., H.L.B.C, M.A.C.F., M.V., A.D.A., V.S.B., D.A.O., P.N.S, M.J.N., J.R.R.P, H.A.V.S, H.F.C.H, J.L.M.F.; visualization, A.G.G.; supervision, J.L.M.F.; project administration, J.L.M.F.; funding acquisition, J.L.M.F., W.J.Z., M.A.C.F., M.V., A.D.A., V.S.B., D.A.O., M.J.N. All authors have read and agreed to the published version of the manuscript.

Funding

We thank the innovation space UNAM-HUAWEI the computational resources to use their supercomputer under project No. 7 “Desarrollo y aplicación de algoritmos de inteligencia artificial para el diseño de fármacos

aplicables al tratamiento de diabetes mellitus y cáncer". We also thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grants #2019/05967-3 (Scholarship MV), #2020/11967-3 (DFG/FAPESP), under the project DINOBBIO (DFG Project #459288952) https://dinobbio.aksw.org, #2022/08333-8 (DAAD/FAPESP), #2013/07600-3 (CIBFar-CEPID), #2014/50926-0 (INCT BioNatCNPq/FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for grant support and research fellowships. Vice Chancellor for Research of the University of Costa Rica for grant via the research project 115-C2-126. Vice-rectory of Research and Postgraduate Studies of the University of Panama for University Research Funds CUFI-VIP-01–14–2019–05 and SNI sponsor.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data is contained within the article, supplementary material and the following github repository https://github.com/alexgoga21/LaNaPDB.

Acknowledgments

A G.-G. thanks the Consejo Nacional de Ciencia y Tecnología (CONACyT) for the PhD scholarship 912137. We thank Ana L. Chávez-Hernández for sharing some scripts. W.J.Z thanks to the students of the chemoinformatics course (II semester 2022) of the University of Costa Rica for initiating the construction of NAPRORE-CR.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
2. Porras-Alcalá, C.; Moya-Utrera, F.; García-Castro, M.; Sánchez-Ruiz, A.; López-Romero, J.M.; Pino-González, M.S.; Díaz-Morilla, A.; Kitamura, S.; Wolan, D.W.; Prados, J.; Melguizo, C.; Cheng-Sánchez, I.; Sarabia, F. The Development of the Bengamides as New Antibiotics against Drug-Resistant Bacteria. *Mar. Drugs* **2022**, *20*, doi:10.3390/md20060373.
3. Xiang, M.-L.; Hu, B.-Y.; Qi, Z.-H.; Wang, X.-N.; Xie, T.-Z.; Wang, Z.-J.; Ma, D.-Y.; Zeng, Q.; Luo, X.-D. Chemistry and bioactivities of natural steroidal alkaloids. *Nat. Prod. Bioprospect.* **2022**, *12*, 23, doi:10.1007/s13659-022-00345-0.
4. Li, X.-W. Chemical ecology-driven discovery of bioactive marine natural products as potential drug leads. *Chin. J. Nat. Med.* **2020**, *18*, 837–838, doi:10.1016/S1875-5364(20)60024-3.
5. Banerjee, P.; Mandhare, A.; Bagalkote, V. Marine natural products as source of new drugs: an updated patent review (July 2018-July 2021). *Expert Opin. Ther. Pat.* **2022**, *32*, 317–363, doi:10.1080/13543776.2022.2012150.
6. Singh, A.; Singh, D.K.; Kharwar, R.N.; White, J.F.; Gond, S.K. Fungal Endophytes as Efficient Sources of Plant-Derived Bioactive Compounds and Their Prospective Applications in Natural Product Drug Discovery: Insights, Avenues, and Challenges. *Microorganisms* **2021**, *9*, doi:10.3390/microorganisms9010197.
7. Tiwari, P.; Bae, H. Endophytic fungi: key insights, emerging prospects, and challenges in natural product drug discovery. *Microorganisms* **2022**, *10*, doi:10.3390/microorganisms10020360.
8. Foxfire, A.; Buhrow, A.R.; Orugunty, R.S.; Smith, L. Drug discovery through the isolation of natural products from Burkholderia. *Expert Opin. Drug Discov.* **2021**, *16*, 807–822, doi:10.1080/17460441.2021.1877655.
9. Porras, G.; Chassagne, F.; Lyles, J.T.; Marquez, L.; Dettweiler, M.; Salam, A.M.; Samarakoon, T.; Shabih, S.; Farrokhi, D.R.; Quave, C.L. Ethnobotany and the role of plant natural products in antibiotic drug discovery. *Chem. Rev.* **2021**, *121*, 3495–3560, doi:10.1021/acs.chemrev.0c00922.
10. Zhang, L.; Song, J.; Kong, L.; Yuan, T.; Li, W.; Zhang, W.; Hou, B.; Lu, Y.; Du, G. The strategies and

techniques of drug discovery from natural products. *Pharmacol. Ther.* **2020**, *216*, 107686, doi:10.1016/j.pharmthera.2020.107686.

11. Bordon, K. de C.F.; Cologna, C.T.; Fornari-Baldo, E.C.; Pinheiro-Júnior, E.L.; Cerni, F.A.; Amorim, F.G.; Anjolette, F.A.P.; Cordeiro, F.A.; Wiezel, G.A.; Cardoso, I.A.; Ferreira, I.G.; de Oliveira, I.S.; Boldrini-França, J.; Pucca, M.B.; Baldo, M.A.; Arantes, E.C. From animal poisons and venoms to medicines: achievements, challenges and perspectives in drug discovery. *Front. Pharmacol.* **2020**, *11*, 1132, doi:10.3389/fphar.2020.01132.
12. Hussain, H.; Mamadalieva, N.Z.; Hussain, A.; Hassan, U.; Rabnawaz, A.; Ahmed, I.; Green, I.R. Fruit peels: food waste as a valuable source of bioactive natural products for drug discovery. *Curr. Issues Mol. Biol.* **2022**, *44*, 1960–1994, doi:10.3390/cimb44050134.
13. Sabe, V.T.; Ntombela, T.; Jhamba, L.A.; Maguire, G.E.M.; Govender, T.; Naicker, T.; Kruger, H.G. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur. J. Med. Chem.* **2021**, *224*, 113705, doi:10.1016/j.ejmech.2021.113705.
14. Doman, T.N.; McGovern, S.L.; Witherbee, B.J.; Kasten, T.P.; Kurumbail, R.; Stallings, W.C.; Connolly, D.T.; Shoichet, B.K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213–2221, doi:10.1021/jm010548w.
15. Kar, S.; Roy, K. How far can virtual screening take us in drug discovery? *Expert Opin. Drug Discov.* **2013**, *8*, 245–261, doi:10.1517/17460441.2013.761204.
16. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395, doi:10.1124/pr.112.007336.
17. Vijayan, R.S.K.; Kihlberg, J.; Cross, J.B.; Poongavanam, V. Enhancing preclinical drug discovery with artificial intelligence. *Drug Discov. Today* **2022**, *27*, 967–984, doi:10.1016/j.drudis.2021.11.023.
18. de Sousa Luis, J.A.; Barros, R.P.C.; de Sousa, N.F.; Muratov, E.; Scotti, L.; Scotti, M.T. Virtual screening of natural products database. *Mini Rev. Med. Chem.* **2021**, *21*, 2657-2730 doi:10.2174/1389557520666200730161549.
19. Sorokina, M.; Steinbeck, C. Review on natural products databases: where to find data in 2020. *J. Cheminform.* **2020**, *12*, 20, doi:10.1186/s13321-020-00424-9.
20. Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M.A.; Steinbeck, C. COCONUT online: Collection of

Open Natural Products database. *J. Cheminform.* **2021**, *13*, 2, doi:10.1186/s13321-020-00478-9.

21. Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H.-Z.; Xu, X. Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS ONE* **2013**, *8*, e62839, doi:10.1371/journal.pone.0062839.
22. ISDB. A database of In-Silico predicted MS/MS spectrum of Natural Products. Available online: http://oolonek.github.io/ISDB/ (accessed on 12 June 2023).
23. Banerjee, P.; Erehman, J.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II--a database of natural products. *Nucleic Acids Res.* **2015**, *43*, D935-9, doi:10.1093/nar/gku886.
24. Sterling, T.; Irwin, J.J. ZINC 15--Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337, doi:10.1021/acs.jcim.5b00559.
25. Ye, H.; Ye, L.; Kang, H.; Zhang, D.; Tao, L.; Tang, K.; Liu, X.; Zhu, R.; Liu, Q.; Chen, Y.Z.; Li, Y.; Cao, Z. HIT: linking herbal active ingredients to targets. *Nucleic Acids Res.* **2011**, *39*, D1055-9, doi:10.1093/nar/gkq1165.
26. Kang, H.; Tang, K.; Liu, Q.; Sun, Y.; Huang, Q.; Zhu, R.; Gao, J.; Zhang, D.; Huang, C.; Cao, Z. HIM-herbal ingredients in-vivo metabolism database. *J. Cheminform.* **2013**, *5*, 28, doi:10.1186/1758-2946-5-28.
27. Li, B.; Ma, C.; Zhao, X.; Hu, Z.; Du, T.; Xu, X.; Wang, Z.; Lin, J. YaTCM: Yet another Traditional Chinese Medicine Database for Drug Discovery. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 600–610, doi:10.1016/j.csbj.2018.11.002.
28. Ru, J.; Li, P.; Wang, J.; Zhou, W.; Li, B.; Huang, C.; Li, P.; Guo, Z.; Tao, W.; Yang, Y.; Xu, X.; Li, Y.; Wang, Y.; Yang, L. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* **2014**, *6*, 13, doi:10.1186/1758-2946-6-13.
29. Kim, S.-K.; Nam, S.; Jang, H.; Kim, A.; Lee, J.-J. TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. *BMC Complement. Altern. Med.* **2015**, *15*, 218, doi:10.1186/s12906-015-0758-5.
30. Xu, H.-Y.; Zhang, Y.-Q.; Liu, Z.-M.; Chen, T.; Lv, C.-Y.; Tang, S.-H.; Zhang, X.-B.; Zhang, W.; Li, Z.-Y.; Zhou, R.-R.; Yang, H.-J.; Wang, X.-J.; Huang, L.-Q. ETCM: an encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res.* **2019**, *47*, D976–D982, doi:10.1093/nar/gky987.
31. Fang, X.; Shao, L.; Zhang, H.; Wang, S. CHMIS-C: a comprehensive herbal medicine information

system for cancer. *J. Med. Chem.* **2005**, *48*, 1481–1488, doi:10.1021/jm049838d.

32. Qiao, X.; Hou, T.; Zhang, W.; Guo, S.; Xu, X. A 3D structure database of components from Chinese traditional medicinal herbs. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 481–489, doi:10.1021/ci010113h.
33. Huang, J.; Zheng, Y.; Wu, W.; Xie, T.; Yao, H.; Pang, X.; Sun, F.; Ouyang, L.; Wang, J. CEMTDD: The database for elucidating the relationships among herbs, compounds, targets and related diseases for Chinese ethnic minority traditional drugs. *Oncotarget* **2015**, *6*, 17675–17684, doi:10.18632/oncotarget.3789.
34. Chen, C.Y.-C. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE* **2011**, *6*, e15939, doi:10.1371/journal.pone.0015939.
35. Mohanraj, K.; Karthikeyan, B.S.; Vivek-Ananth, R.P.; Chand, R.P.B.; Aparna, S.R.; Mangalapandi, P.; Samal, A. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Sci. Rep.* **2018**, *8*, 4329, doi:10.1038/s41598-018-22631-z.
36. Potshangbam, A.M.; Polavarapu, R.; Rathore, R.S.; Naresh, D.; Prabhu, N.P.; Potshangbam, N.; Kumar, P.; Vindal, V. MedPServer: A database for identification of therapeutic targets and novel leads pertaining to natural products. *Chem. Biol. Drug Des.* **2019**, *93*, 438–446, doi:10.1111/cbdd.13430.
37. Bultum, L.E.; Woyessa, A.M.; Lee, D. ETM-DB: integrated Ethiopian traditional herbal medicine and phytochemicals database. *BMC Complement. Altern. Med.* **2019**, *19*, 212, doi:10.1186/s12906-019-2634-1.
38. Ntie-Kang, F.; Onguéné, P.A.; Scharfe, M.; Owono Owono, L.C.; Megnassan, E.; Mbaze, L.M.; Sippl, W.; Efange, S.M.N. ConMedNP: a natural product library from Central African medicinal plants for drug discovery. *RSC Adv.* **2014**, *4*, 409–419, doi:10.1039/C3RA43754J.
39. Ibezim, A.; Debnath, B.; Ntie-Kang, F.; Mbah, C.J.; Nwodo, N.J. Binding of anti-Trypanosoma natural products from African flora against selected drug targets: a docking study. *Med. Chem. Res.* **2017**, *26*, 562–579, doi:10.1007/s00044-016-1764-y.
40. Onguéné, P.A.; Ntie-Kang, F.; Mbah, J.A.; Lifongo, L.L.; Ndom, J.C.; Sippl, W.; Mbaze, L.M. The potential of anti-malarial compounds derived from African medicinal plants, part III: an in silico evaluation of drug metabolism and pharmacokinetics profiling. *Org. Med. Chem. Lett.* **2014**, *4*, 6, doi:10.1186/s13588-014-0006-x.

41. Ntie-Kang, F.; Nwodo, J.N.; Ibezim, A.; Simoben, C.V.; Karaman, B.; Ngwa, V.F.; Sippl, W.; Adikwu, M.U.; Mbaze, L.M. Molecular modeling of potential anticancer agents from African medicinal plants. *J. Chem. Inf. Model.* **2014**, *54*, 2433–2450, doi:10.1021/ci5003697.

42. Ntie-Kang, F.; Amoa Onguéné, P.; Fotso, G.W.; Andrae-Marobela, K.; Bezabih, M.; Ndom, J.C.; Ngadjui, B.T.; Ogundaini, A.O.; Abegaz, B.M.; Meva'a, L.M. Virtualizing the p-ANAPL library: a step towards drug discovery from African medicinal plants. *PLoS ONE* **2014**, *9*, e90655, doi:10.1371/journal.pone.0090655.

43. Ntie-Kang, F.; Zofou, D.; Babiaka, S.B.; Meudom, R.; Scharfe, M.; Lifongo, L.L.; Mbah, J.A.; Mbaze, L.M.; Sippl, W.; Efange, S.M.N. AfroDb: a select highly potent and diverse natural product library from African medicinal plants. *PLoS ONE* **2013**, *8*, e78085, doi:10.1371/journal.pone.0078085.

44. Ionov, N.; Druzhilovskiy, D.; Filimonov, D.; Poroikov, V. Phyto4Health: Database of Phytocomponents from Russian Pharmacopoeia Plants. *J. Chem. Inf. Model.* **2023**, *63*, 1847–1851, doi:10.1021/acs.jcim.2c01567.

45. Raven, P.H.; Gereau, R.E.; Phillipson, P.B.; Chatelain, C.; Jenkins, C.N.; Ulloa Ulloa, C. The distribution of biodiversity richness in the tropics. *Sci. Adv.* **2020**, *6*, doi:10.1126/sciadv.abc6228.

46. Mittermeier, R.A.; Turner, W.R.; Larsen, F.W.; Brooks, T.M.; Gascon, C. Global Biodiversity Conservation: The Critical Role of Hotspots. In *Biodiversity Hotspots*; Zachos, F. E., Habel, J. C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp. 3–22 ISBN 978-3-642-20991-8.

47. NaturAr. Available online: https://naturar.quimica.unlp.edu.ar/es/ (accessed on 9 December 2022).

48. Valli, M.; dos Santos, R.N.; Figueira, L.D.; Nakajima, C.H.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Development of a natural products database from the biodiversity of Brazil. *J. Nat. Prod.* **2013**, *76*, 439–444, doi:10.1021/np3006875.

49. Pilon, A.C.; Valli, M.; Dametto, A.C.; Pinto, M.E.F.; Freire, R.T.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* **2017**, *7*, 7215, doi:10.1038/s41598-017-07451-x.

50. Scotti, M.T.; Herrera-Acevedo, C.; Oliveira, T.B.; Costa, R.P.O.; Santos, S.Y.K. de O.; Rodrigues, R.P.; Scotti, L.; Da-Costa, F.B. SistematX, an Online Web-Based Cheminformatics Tool for Data Management of Secondary Metabolites. *Molecules* **2018**, *23*, doi:10.3390/molecules23010103.

51. Costa, R.P.O.; Lucena, L.F.; Silva, L.M.A.; Zocolo, G.J.; Herrera-Acevedo, C.; Scotti, L.; Da-Costa, F.B.; Ionov, N.; Poroikov, V.; Muratov, E.N.; Scotti, M.T. The sistematx web portal of natural products: an update. *J. Chem. Inf. Model.* **2021**, *61*, 2516–2522, doi:10.1021/acs.jcim.1c00083.

52. UEFS Natural Products. Available online: http://zinc12.docking.org/catalogs/uefsnp (accessed on 2 December 2022).

53. Olmedo, D.A.; González-Medina, M.; Gupta, M.P.; Medina-Franco, J.L. Cheminformatic characterization of natural products from Panama. *Mol. Divers.* **2017**, *21*, 779–789, doi:10.1007/s11030-017-9781-4.

54. A. Olmedo, D.; L. Medina-Franco, J. Chemoinformatic approach: the case of natural products of panama. In *Cheminformatics and its applications [working title]*; IntechOpen, 2019.

55. Barazorda-Ccahuana, H.L.; Ranilla, L.G.; Candia-Puma, M.A.; Cárcamo-Rodriguez, E.G.; Centeno-Lopez, A.E.; Davila-Del-Carpio, G.; Medina-Franco, J.L.; Chávez-Fumagalli, M.A. PeruNPDB: the Peruvian Natural Products Database for in silico drug screening. *Sci. Rep.* **2023**, *13*, 7577, doi:10.1038/s41598-023-34729-0.

56. UNIIQUIM. Available online: https://uniiquim.iquimica.unam.mx/ (accessed on 6 December 2022).

57. Pilón-Jiménez, B.A.; Saldívar-González, F.I.; Díaz-Eufracio, B.I.; Medina-Franco, J.L. BIOFACQUIM: A mexican compound database of natural products. *Biomolecules* **2019**, *9*, doi:10.3390/biom9010031.

58. Sánchez-Cruz, N.; Pilón-Jiménez, B.A.; Medina-Franco, J.L. Functional group and diversity analysis of BIOFACQUIM: A Mexican natural product database. *F1000Res.* **2019**, *8*, doi:10.12688/f1000research.21540.2.

59. Gómez-García, A.; Medina-Franco, J.L. Progress and impact of latin american natural product databases. *Biomolecules* **2022**, *12*, doi:10.3390/biom12091202.

60. do Carmo Santos, N.; da Paixão, V.G.; da Rocha Pita, S.S. New Trypanosoma cruzi Trypanothione Reductase Inhibitors Identification using the Virtual Screening in Database of Nucleus Bioassay, Biosynthesis and Ecophysiology (NuBBE). *AIA* **2019**, *17*, 138–149, doi:10.2174/2211352516666180928130031.

61. Acevedo, C.H.; Scotti, L.; Scotti, M.T. In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018**, *13*, 634–

645, doi:10.1002/cmdc.201700743.

62. Antunes, S.S.; Won-Held Rabelo, V.; Romeiro, N.C. Natural products from Brazilian biodiversity identified as potential inhibitors of PknA and PknB of M. tuberculosis using molecular modeling tools. *Comput. Biol. Med.* **2021**, *136*, 104694, doi:10.1016/j.compbiomed.2021.104694.

63. Herrera-Acevedo, C.; Dos Santos Maia, M.; Cavalcanti, É.B.V.S.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Selection of antileishmanial sesquiterpene lactones from SistematX database using a combined ligand-/structure-based virtual screening approach. *Mol. Divers.* **2021**, *25*, 2411–2427, doi:10.1007/s11030-020-10139-6.

64. Barazorda-Ccahuana, H.L.; Goyzueta-Mamani, L.D.; Candia Puma, M.A.; Simões de Freitas, C.; de Sousa Vieria Tavares, G.; Pagliara Lage, D.; Ferraz Coelho, E.A.; Chávez-Fumagalli, M.A. Computer-aided drug design approaches applied to screen natural product's structural analogs targeting arginase in Leishmania spp. *F1000Res.* **2023**, *12*, 93, doi:10.12688/f1000research.129943.1.

65. Menezes, R.P.B. de; Viana, J. de O.; Muratov, E.; Scotti, L.; Scotti, M.T. Computer-Assisted Discovery of Alkaloids with Schistosomicidal Activity. *Curr. Issues Mol. Biol.* **2022**, *44*, 383–408, doi:10.3390/cimb44010028.

66. Rodrigues, G.C.S.; Dos Santos Maia, M.; de Menezes, R.P.B.; Cavalcanti, A.B.S.; de Sousa, N.F.; de Moura, É.P.; Monteiro, A.F.M.; Scotti, L.; Scotti, M.T. Ligand and Structure-based Virtual Screening of Lamiaceae Diterpenes with Potential Activity against a Novel Coronavirus (2019-nCoV). *Curr. Top. Med. Chem.* **2020**, *20*, 2126–2145, doi:10.2174/1568026620666200716114546.

67. Przybyłek, M. Application 2D Descriptors and Artificial Neural Networks for Beta-Glucosidase Inhibitors Screening. *Molecules* **2020**, *25*, doi:10.3390/molecules25245942.

68. Martinez-Mayorga, K.; Marmolejo-Valencia, A.F.; Cortes-Guzman, F.; García-Ramos, J.C.; Sánchez-Flores, E.I.; Barroso-Flores, J.; Medina-Franco, J.L.; Esquivel-Rodriguez, B. Toxicity Assessment of Structurally Relevant Natural Products from Mexican Plants with Antinociceptive Activity. *J. Mex. Chem. Soc.* **2017**, *61*, doi:10.29356/jmcs.v61i3.344.

69. Barrera-Vázquez, O.S.; Gómez-Verjan, J.C.; Magos-Guerrero, G.A. Chemoinformatic Screening for the Selection of Potential Senolytic Compounds from Natural Products. *Biomolecules* **2021**, *11*, doi:10.3390/biom11030467.

70. Herrera-Acevedo, C.; Perdomo-Madrigal, C.; Herrera-Acevedo, K.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Machine learning models to select potential inhibitors of acetylcholinesterase activity from SistematX: a natural products database. *Mol. Divers.* **2021**, *25*, 1553–1568, doi:10.1007/s11030-021-10245-z.

71. de Souza Ribeiro, M.M.; dos Santos, L.C.; de Novais, N.S.; Viganó, J.; Veggi, P.C. An evaluative review on Stryphnodendron adstringens extract composition: Current and future perspectives on extraction and application. *Industrial Crops and Products* **2022**, *187*, 115325, doi:10.1016/j.indcrop.2022.115325.

72. Li, R.; Morris-Natschke, S.L.; Lee, K.-H. Clerodane diterpenes: sources, structures, and biological activities. *Nat. Prod. Rep.* **2016**, *33*, 1166–1226, doi:10.1039/c5np00137d.

73. Parra, J.; Ford, C.D.; Murillo, R. PHYTOCHEMICAL STUDY OF ENDEMIC COSTA RICAN ANNONACEAE SPECIES Annona pittieri AND Cymbopetalum costaricense. *J. Chil. Chem. Soc.* **2021**, *66*, 5047–5050, doi:10.4067/S0717-97072021000105047.

74. Köhler, I.; Jenett-Siems, K.; Siems, K.; Hernández, M.A.; Ibarra, R.A.; Berendsohn, W.G.; Bienzle, U.; Eich, E. In vitro antiplasmodial investigation of medicinal plants from El Salvador. *Z Naturforsch, C, J Biosci* **2002**, *57*, 277–281, doi:10.1515/znc-2002-3-413.

75. Sotelo-Barrera, M.; Cília-García, M.; Luna-Cavazos, M.; Díaz-Núñez, J.L.; Romero-Manzanares, A.; Soto-Hernández, R.M.; Castillo-Juárez, I. Amphipterygium adstringens (Schltdl.) Schiede ex Standl (Anacardiaceae): An Endemic Plant with Relevant Pharmacological Properties. *Plants* **2022**, *11*, doi:10.3390/plants11131766.

76. Agin-Liebes, G.; Haas, T.F.; Lancelotta, R.; Uthaug, M.V.; Ramaekers, J.G.; Davis, A.K. Naturalistic Use of Mescaline Is Associated with Self-Reported Psychiatric Improvements and Enduring Positive Life Changes. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 543–552, doi:10.1021/acsptsci.1c00018.

77. Cubilla-Rios, L.; Chérigo, L.; Ríos, C.; Togna, G.D.; Gerwick, W.H. Phytochemical analysis and antileishmanial activity of Desmotes incomparabilis, an endemic plant from Panama. *Planta Med.* **2008**, *74*, doi:10.1055/s-0028-1084096.

78. García Giménez, D.; García Prado, E.; Sáenz Rodríguez, T.; Fernández Arche, A.; De la Puerta, R. Cytotoxic effect of the pentacyclic oxindole alkaloid mitraphylline isolated from Uncaria tomentosa bark on human Ewing's sarcoma and breast cancer cell lines. *Planta Med.* **2010**, *76*, 133–136, doi:10.1055/s-

0029-1186048.

79. Gonzales, G.F.; Valerio, L.G. Medicinal plants from Peru: a review of plants as potential agents against cancer. *Anticancer Agents Med Chem* **2006**, *6*, 429–444, doi:10.2174/187152006778226486.

80. Medina-Franco, J.L.; Chávez-Hernández, A.L.; López-López, E.; Saldívar-González, F.I. Chemical multiverse: an expanded view of chemical space. *Mol. Inform.* **2022**, *41*, e2200116, doi:10.1002/minf.202200116.

81. Isah, M.B.; Tajuddeen, N.; Umar, M.I.; Alhafiz, Z.A.; Mohammed, A.; Ibrahim, M.A. Terpenoids as Emerging Therapeutic Agents: Cellular Targets and Mechanisms of Action against Protozoan Parasites. In; Studies in natural products chemistry; Elsevier, 2018; Vol. 59, pp. 227–250 ISBN 9780444641793.

82. Wildman, S.A.; Crippen, G.M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873, doi:10.1021/ci990307l.

83. Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3717, doi:10.1021/jm000942e.

84. 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.* **2001**, *46*, 3–26, doi:10.1016/S0169-409X(00)00129-0.

85. Lipinski, C.A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337–341, doi:10.1016/j.ddtec.2004.11.007.

86. 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, doi:10.1021/jm020017n.

87. Gleeson, M.P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008**, *51*, 817–834, doi:10.1021/jm701122q.

88. Hughes, J.D.; Blagg, J.; Price, D.A.; Bailey, S.; Decrescenzo, G.A.; Devraj, R.V.; Ellsworth, E.; Fobian, Y.M.; Gibbs, M.E.; Gilles, R.W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875, doi:10.1016/j.bmcl.2008.07.071.

89. Ntie-Kang, F.; Nyongbela, K.D.; Ayimele, G.A.; Shekfeh, S. “Drug-likeness” properties of natural compounds. *Physical Sciences Reviews* **2019**, *4*, doi:10.1515/psr-2018-0169.

90. Probst, D.; Reymond, J.-L. Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminform.* **2020**, *12*, 12, doi:10.1186/s13321-020-0416-x.

91. Waskom, M. seaborn: statistical data visualization. *JOSS* **2021**, *6*, 3021, doi:10.21105/joss.03021.

92. Open-source chemoinformatics and machine learning. RDKit: Open-Source Cheminformatics Software. Available online: https://www.rdkit.org (accessed on 8 February 2023).

93. MolVS. Molecule Validation and Standardization. Available online: https://molvs.readthedocs.io/en/latest/index.html (accessed on 9 February 2023).

94. Kim, H.W.; Wang, M.; Leber, C.A.; Nothias, L.-F.; Reher, R.; Kang, K.B.; van der Hooft, J.J.J.; Dorrestein, P.C.; Gerwick, W.H.; Cottrell, G.W. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **2021**, *84*, 2795–2807, doi:10.1021/acs.jnatprod.1c00399.

95. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME - the Konstanz information miner. *SIGKDD Explor. Newsl.* **2009**, *11*, 26, doi:10.1145/1656274.1656280.

96. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *arXiv* **2012**, doi:10.48550/arxiv.1201.0490.

97. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082, doi:10.1093/nar/gkx1037.

98. Bajusz, D.; Rácz, A.; Héberger, K. Chemical data formats, fingerprints, and other molecular descriptions for database analysis and searching. In *Comprehensive medicinal chemistry III*; Elsevier, 2017; pp. 329–378 ISBN 9780128032015.

99. Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal,

T.; Rojas-Chertó, M.; Spjuth, O.; Torrance, G.; Evelo, C.T.; Guha, R.; Steinbeck, C. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform.* **2017**, *9*, 33, doi:10.1186/s13321-017-0220-4.

100. Capecchi, A.; Probst, D.; Reymond, J.-L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *J. Cheminform.* **2020**, *12*, 43, doi:10.1186/s13321-020-00445-4.
101. Miranda-Salas, J.; Peña-Varas, C.; Valenzuela Martínez, I.; Olmedo, D.A.; Zamora, W.J.; Chávez-Fumagalli, M.A.; Azevedo, D.Q.; Castilho, R.O.; Maltarollo, V.G.; Ramírez, D.; Medina-Franco, J.L. Trends and challenges in chemoinformatics research in Latin America. *Artif. Intell. Life Sci.* **2023**, *3*, 100077, doi:10.1016/j.ailsci.2023.100077.