Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

LANaPD: Towards a Unified Latin America Natural Products Database[+]

José L. Medina-Franco*

Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México.

* Correspondence: medinajl@unam.com.mx; Tel.: +5255-5622-3899

[+]This manuscript is dedicated to all people affected directly or indirectly by the COVID-19 pandemic.

ABTRACT

Around the World, the number of compound databases of natural products in the public domain is rising. This is in line with the increasing synergistic combination of natural product research and chemoinformatics. Towards this global endeavor, countries in Latin America are assembling, curating, and analyzing the contents and diversity of natural products available in their geographical regions. In this manuscript we collect and analyze the efforts that countries in Latin America have made so far to build natural product databases. We further encourage the scientific community in particular in Latin America, to continue their efforts to building quality natural product databases and, whenever possible, to make them publicly accessible. It is proposed that all compound collections could be assembled into a unified resource called LANaPD: Latin America Natural Products Database. Opportunities and challenges to build, distribute, and maintain LANaPD are also discussed.

Keywords: chemoinformatics; chemical space; database; LANaPD; molecular diversity; drug discovery; natural sources

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

1. Introduction

For centuries, natural products (NPs) have been the basis for the prevention and treatment of diseases. Up to date, NPs continue to have a profound impact in drug discovery [1,2]. There are compound from natural sources or are natural product derivatives that are now drugs approved for clinical use. Amid the current pandemic of COVID-19, a notable example of a promising NP is chloroquine phosphate that is an analogue of the alkaloid quinine, originally extracted from the bark of cinchona trees [3] (Figure 1). In addition, NPs have largely contributed to compounds that to start compounds that are later optimized in terms of potency or pharma kinetic or pharmacodynamics properties or serve as source of inspiration to synthesize organic compounds. It is well known that the heath-related benefits are somehow associated with their structural uniqueness, diversity, and complexity, as compared to other drugs from different sources [4,5].

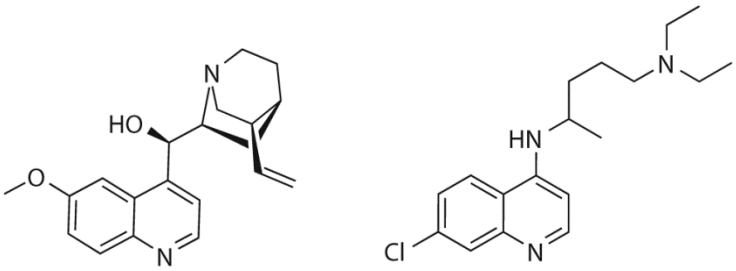


Figure 1. Chemical structures of the natural product quinine and it structural analogue chloroquine with high relevance amid the current COVID-19 pandemic.

Chemical informatics (also termed in the literature *chem(i)oinformatics*) is increasingly contributing to drug discovery at different levels [6-8]. For instance, a key contribution of this research discipline is designing, building, and curating compound databases. Indeed, compound databases play a significant role in drug discovery [9] and

different collections, in particular in the public domain, have been reviewed elsewhere [10]. Additional major contributions of chemoinformatics to drug discovery projects are assisting screening compounds (for example, filtering *in silico* compounds libraries to select compounds for experimental testing), and analyzing the outcome of experimental screening assays, either, in small, medium, and/or high-throughput format [6,7,11]. Based on the experimental results, cheminformatics help to generate and/or refine hypothesis of the mechanism of action of bioactive molecules at the molecular level, and/or build models to predict the outcome of untested compounds e.g., part of a new cycle of *in silico* screening. All these and several other chemoinformatics tools have been successfully applied for organic compounds, including NPs and food chemicals [12,13]. Recently, it has been proposed the systematic application of chemoinformatics resources to further advance the field of Medicinal Organometallic Chemistry [14].

Contributions of informatics to advance NP research in general, and NP-based drug discovery in particular, are increasing [15,16]. One of such key contributions has been the organization and analysis of chemical information of NPs, with or without biological activity, in compound databases. Over the past five years, several reviews of NP databases have been published [17-21]. Some of these reviews include chemoinformatic analysis of the contents, diversity, and coverage of the compounds in chemical space. Of note, there has been a rapid increase in the number of publicly accessible NP databases. One of the first reviews was published in 2012 [18] that included five NPs datasets (commercial and non-commercial with chemical structures available on the web). Recently, it was published the COlleCtion of Open NatUral producTs (COCONUT) database that collects over 120 databases collecting more than 400,000 non-redundant NPs and are freely accessible [22]. As part of the global efforts, different countries around the world are analyzing the information of NPs in their countries of origin. Examples are AfroDd, from Africa [23] and VIETHERB from Vietnam [24]. As part of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

such global efforts, different Latin America countries are building their own compound databases using chemoinformatics resources [25],.

The primary goal of this manuscript is to discuss the recent progress of countries in Latin America to put together, curate, and analyze compound databases of NP molecules contained in their geographical region. Indeed, Latin American countries are traditionally rich in their unique biodiversity and herbal medicine has a strong tradition and use in the region. Herein we also propose to join efforts and assemble a unified Latin America Natural Products Database (LANaPD).

2. Natural Product Databases in Latin America

Thus far, Brazil, Mexico, and Panama have published NP databases (Table 1). In some cases, the chemical structures are already available in the public domain and/or comprehensive analyses of their content and diversity have been released. In this section we discuss the progress on the development of such compounds databases. For each one we describe briefly the research group and institution developing the database, the contents and number of compounds currently available, accessibility and capabilities to browse the contents, and where available, we summarize recent analysis of the chemical diversity and coverage of the chemical space and other uses.

Table 1. Overview of current natural product databases developed in Latin America.

Database	Country	Brief description	URL	Ref
$NuBBE_{DB}$	Brazil	Over 2000 secondary metabolites of plants, fungi, insects, marine organisms, and bacteria.	https://nubbe.iq.unesp.br/portal/nubbe-search.html (English and Portuguese)	[26] [27]
CIFPMA	Panama	Over 450 natural products	Online database under construction.	[28] [29]

		from Panama.		
UNIQUIM	Mexico	Compounds from plants, fungi, marine organisms, and insects.	https://uniiquim.iquimica.unam.mx/ (Spanish only)	NA
BIOFACQUIM	Mexico	Over 550 compounds from Mexican biodiversity mainly from plants, fungi, and Mexican propolis.	1st version at *BIOFACQUIM Explorer*: https://biofacquim.herokuapp.com/ (English only) 2nd version: http://doi.org/10.6084/m9.figshare.11312702	[30] [31]

2.1 NuBBE$_{DB}$

2.1.1 Developers

This public database was launched in 2013 as a joint effort of the Brazilian research groups Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products (NuBBE) of the São Paulo State University, and the Laboratory of Computational and Medicinal Chemistry of the University of São Paulo.

2.1.2 Contents

The first release of NuBBE$_{DB}$ contained approximately 640 compounds collected from publications of the NuBBE research group [26]. Four years later, the same group published an update expanding the number of compounds to more than 2000, thus increasing representation of the large biodiversity in Brazil. The update also had significant enhancements to the web-site interface [27]. Compounds in NuBBE$_{DB}$ are secondary metabolites of plants, fungi, insects, marine organisms, and bacteria.

Compounds in NuBBE$_{DB}$ are annotated with chemical, biological, pharmacological, and spectroscopic data. Chemical information includes IUPAC name, chemical structure, drug-like physicochemical properties, and metabolic class. The biological information

comprises species, geographical location, and biological activities. The spectroscopic data includes molar mass and nuclear magnetic resonance, NMR data.

2.1.3 Accessibility and searching capabilities

$NuBBE_{DB}$ is accessible and searchable at the web-site interface (link in Table 1). It is also available at ChemSpider and ZINC 15 [32] where it can be found, for instance, as a natural product catalog. $NuBBE_{DB}$ has been recently included the COCONUT database [22].

The user can download the entire database or perform on-line searches. It has in-built a broad range of searching and filtering criterions (Figure 2). For instance, is possible to search by species, geographical region in Brazil, source, biological properties, chemical structure, chemical drug-like descriptors, spectroscopic data (specifically, NMR information), and bibliographic information.

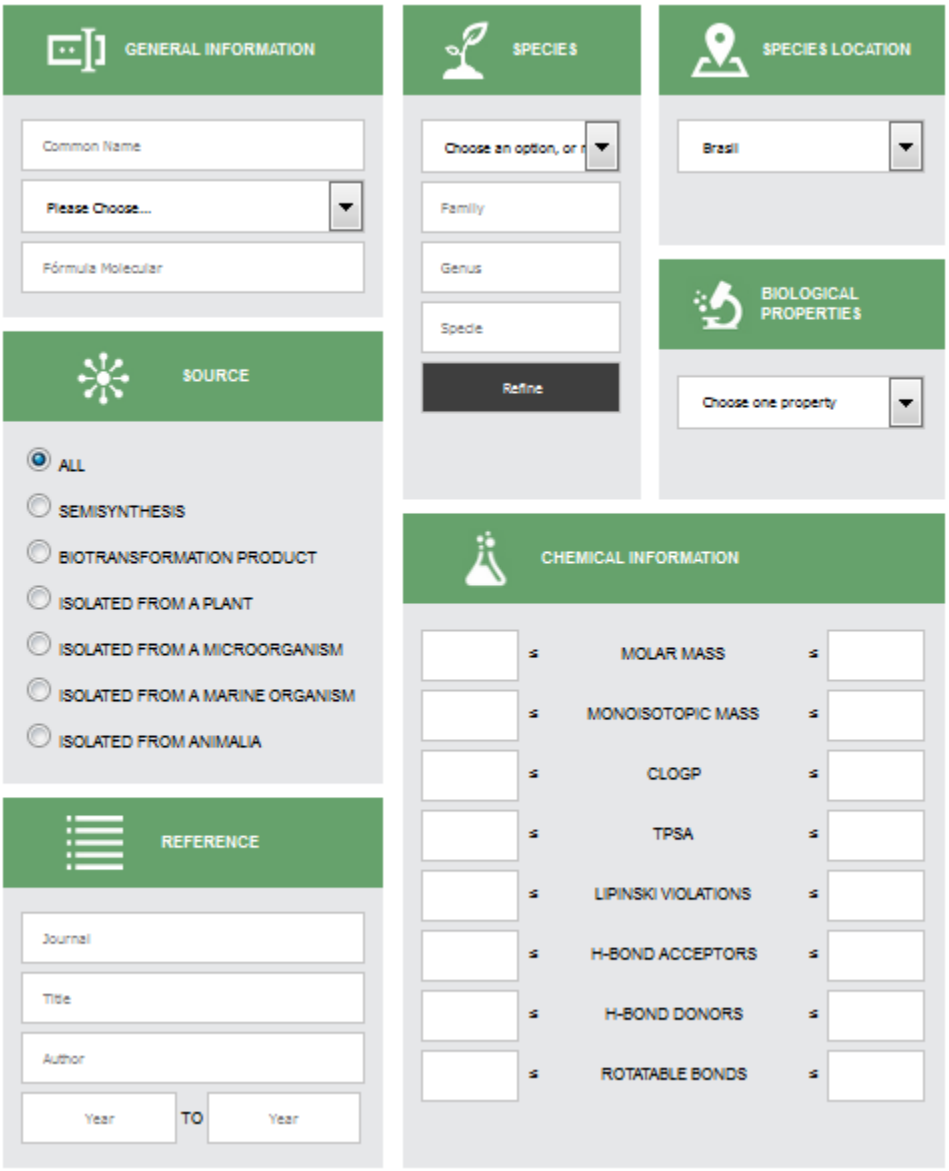
2.1.4 Diversity analysis and other applications

The most recent published version of $NuBBE_{DB}$ was analyzed based on structural diversity and complexity of the chemical structures. To this end, several chemoinformatic tools were employed. As part of the study, the contents and diversity profile $NuBBE_{DB}$ was compared to other commercial and non-commercial NP collections whose chemical structures are freely available. The reference collections included the Universal Natural Product Database, with more than 200000 molecules [33], and ChEMBL [34]. It was concluded that compounds in $NuBBE_{DB}$ are diverse in terms of molecular fingerprints, chemical scaffolds, and drug-like properties. Using stablished chemoinformatic tools, the study supported that several compounds in $NuBBE_{DB}$ are promising candidates for drug discovery and medicinal chemistry [35]. Interestingly, the study also revealed that 12% of the chemical scaffolds in $NuBBE_{DB}$ are not present in ChEMBL. Also, an *in silico*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

ADMET profiling of $NuBBE_{DB}$ has been published recently [36]. As discussed hereunder, chemoinformatic comparisons of $NuBBE_{DB}$ and other NP databases in Latin America have been performed. $NuBBE_{DB}$ has been successfully used in several drug discovery and dereplication studies as reviewed in [27].



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

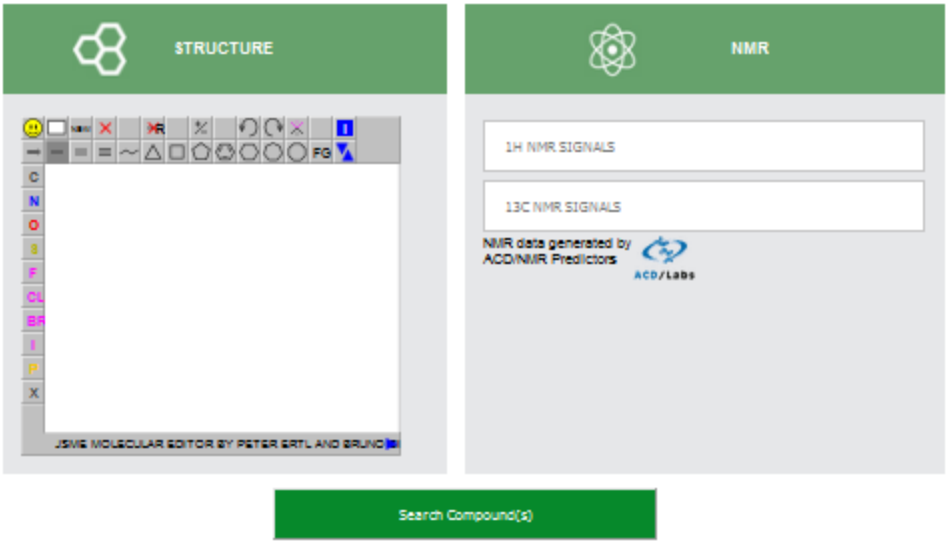


Figure 2. Screenshot of part of the *NuBBE*$_{DB}$ web-interface.

2.2 CIFPMA

2.2.1 Developers

Over the past few years, the Center for Pharmacognostic Research on Panamanian Flora, College of Pharmacy of the University of Panama (CIFLORPAN, for its acronym in Spanish) has been building The Natural Products Database from The University of Panama, Republic of Panama: CIFPMA. This dataset was first disclosed in 2017 [28].

2.2.2 Contents

The first disclosure of CIFPMA contained 354 compounds [28] and recently was updated to 454 molecules [29]. CIFPMA has compounds that have been tested biologically under more than 25 *in vitro* and *in vivo* bioassays. Examples of target therapeutic indications are anti-HIV, antioxidants, and anticancer.

2.2.3 Accessibility and searching capabilities

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

A web-site is under construction. Currently, the chemical structures would be available upon request.

2.2.4 Diversity analysis and other applications

The content, diversity analysis, as systematic structure-structure activity relationship studies of compounds in CIFPMA have been reported [28,29]. The first version with 354 molecules was compared to $NuBBE_{DB}$, molecules from the Traditional Chinese Medicine database, compounds with drug indications in ChEMBL, and other reference libraries of NPs. It was concluded that metabolites in CIFPMA have large scaffold diversity and also has several unique scaffolds. The high scaffold diversity is in agreement with the broad range of biological activities [28]. The most recent version of CIFPMA was compared to other NPs databases including $NuBBE_{DB}$ and BIOFACQUIM, drugs approved for clinical use, and synthetic compounds [29]. The comparison was made based on drug-like physicochemical properties, structural fingerprints, and molecular scaffolds. It was concluded that NP databases have higher structural complexity than synthetic compounds. It was also concluded that while compounds from synthetic origin have a larger proportion of aromatic atoms [29].

2.3 UNIIQUIM

2.3.1 Developers

For more than five years, the Informatics Unit of the Institute of Chemistry (UNIIQUIM, for its acronym in Spanish) of the National Autonomous University of Mexico (UNAM) has been assembling and curating an open database with NP from Mexico, mainly isolated at published by researchers of the Natural Products Department of the Institute of Chemistry, UNAM.

2.3.2 Contents

This is a database intended to collect part of the large biodiversity of Mexico that has been published by the Natural Products Department of the Institute of Chemistry, UNAM. Compounds in UNIIQUIM are NP isolated in Mexico from plants, fungi, marine organisms, and insects found in Mexico. The total number of compounds is not totally clear from the website that is available only in Spanish (Table 1).

Compounds in UNIIQUIM are annotated with chemical and biological data, when available. Chemical information includes molecular formula, IUPAC names, CAS number, and the chemical structure. Each compound record is linked to the reported biological activity, if reported in the publication source.

2.2.3 Accessibility and searching capabilities

UNIIQUIM database is accessible at the web-site interface (Table 1) that is currently available in Spanish (an English version will be released). It is not possible to download the entire database. The user can browse the contents by displaying either of two look-up tables: list of chemical compounds, and list of organisms (Figure 3). The user can select the desired chemical compound or organism for specific information. It is also possible to search the database by bibliographic information.

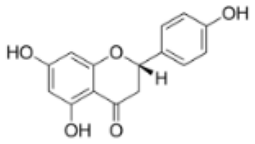
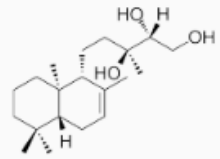
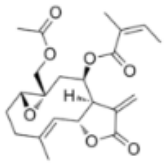
2.2.4 Diversity analysis and other applications

To the best of our knowledge there are no reports of published applications of UNIIQUIM. The contents was first reviewed in [37]. It is anticipated that the database will be cited in the near future.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

Lista de Compuestos

Compuesto	Fórmula molécular	CAS	Estructura
(-)-Naringenina	$C_{15}H_{12}O_5$	480-41-1	 Naringenina
(+)-*ent*-Labd-7-en-13*S*,14*S*,15-triol	$C_{20}H_{36}O_3$	156317-09-8	 (+)-*ent*-Labd-7-en-13*S*,14*S*,15-triol
(*Z*)-18-Metilsferocefalina	$C_{22}H_{28}O_7$	79384-03-5	 (*Z*)-18-Metilsferocefalina

Lista de Organismos

Organismo	Artículo Relacionado	Año
Montanoa sp.	Structure and stereochemistry of two coexisting sesquiterpene lactones 8α-[(*S*)-2'-methylbutyroxy]custunolide and 8α-isobutyroxycustunolide	1989
Stevia tephrophylla Blake	A one step transformation of 4α,5β-epoxygermacranolide into pseudoguaianolide	1989
Montanoa tomentosa	X-Ray crystal structure of grandiflorenic acid [(-)-kaura-9(11)-16-dien-19-oic acid] methyl ester, a compound formerly considered as an oily derivative	2009
Ambrosia peruviana	Modificación química del ácido ambrósico (sesquiterpeno, inhibidor de la germinación) para aumentar su actividad biológica	1994
Stevia lucida Lay. var. *lucida*	Structure and stereochemistry of 7β-[(+)-camphorsulfonyl]-9α-hydroxylongipin-2-en-1-one	1988
Stevia polycephala	Complete ^{1}H and ^{13}C NMR assignment of stephalic acid	1999

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0012.v1

Peer-reviewed version available at *Future Science OA* **2020**, FSO597; doi:10.2144/fsoa-2020-0068

Figure 3. Screenshot of *UNIIQUIM* web-interface with a focus on searching capabilities based on look-up tables of chemistry and organism information (current web-site available in Spanish only).

2.4 BIOFACQUIM

2.4.1 Developers

For the past two years, the Computer-Aided Design at the School of Chemistry group (DIFACQUIM, for its acronym in Spanish) at UNAM is building and curating a NP database containing compounds isolated in Mexico. The final goal is capturing, as much as possible, the Mexican biodiversity.

2.4.2 Contents

The first version of BIOFACQUIM was released in 2019 and contained 423 molecules gathered from publications of the School of Chemistry for a 10-year period [30]. The same year, the database was updated with 148 structures to reach 553 compounds including molecules isolated not only in that institution but also by research groups in other Mexican institutions. As other NP databases discussed herein, BIOFACQUIM continue to be updated. Most of the compounds in BIOFACQUIM were isolated from plant, bacteria, and Mexican propolis.

Molecules in BIOFACQUIM are annotated with the chemical name and structure, bibliographic information, kingdom, genus, and species of the NP, and geographical location of the collection. If the biological information is included in the original publication, the activity data is included in the compound record.

2.4.3 Accessibility and searching capabilities

The first version of BIOFACQUIM is accessible and searchable at the "BIOFACQUIM Explorer" web-site (link in Table 1) (Figure 4). It is also available at ZINC 15 and is part

of the COCONUT database [22]. The second version of BIOFACQUIM is freely accessible at Fighsare [31] [38] and is part of the D-TOOLS initiative aimed to provide open databases and chemoinformatic resources [25].

2.2.4 Diversity analysis and other applications

A comprehensive diversity analysis of the first release of BIOFACQUIM was published recently, along with the disclosure of the database itself [30]. It was concluded that compounds in this database have a broad coverage of the chemical space, overlapping with drug-like space as compared to approved drugs. Furthermore, the analysis also revealed structures with high chemical similarity to drug in clinical use. Recently, the chemical fragments in BIOFACQUIM were compared with those fragments available in ChEMBL 25, and a therein constructed assembled dataset with 169,839 unique structures of NPs [31]. It was concluded that, as expected, the chemical diversity of BIOFACQUIM increased in terms of chemical scaffolds and structural fingerprints relative to the first version. It was also concluded that, despite the relative few number of compounds in BIOFACQUIM as compared to the reference databases, there are a significant number of compounds, scaffolds, and functional groups in BIOFACQUIM that are not present in the reference datasets [31].

Figure 4. Screenshot of *BIOFACQUIM Explorer* website.

3. Towards LANaPD

Herein it is proposed building a unified database of NPs that represent the biodiversity of Latin America. Challenging tasks that can be overcome, one more difficult than others are discussed hereunder. Recent guidelines to assemble databases of NP have been published, in particular when intended to be used in virtual screening [21].

3.1 Collection and standardization

The first step towards creating LANaPD is putting together all NP databases, processing, and curating them using standard protocols. Although this step is not straightforward it is feasible. It would be advisable that a research group would be in charge of this endeavor using publicly accessible tools and scripts or workflows available in public repositories

such as Github. Examples of freely accessible workflows to curate compound database are available [39,40]. COCONUT database (*vide supra*) is an example of a large-scale database assembled and curated from several different sources around the world [22]. However, as discussed above, COCONUT is not focused on specific geographical regions and it does not contain all public databases from Latin America.

3.2 Accessibility

Ideally, LANaPD can be made accessible to the public. This can be done generating a web-server dedicated to the database following the Findable, Accessible, Interoperable, and Reusable (FAIR) principles [41]. Another option to deploy the database is using a public repository such as Figshare (https://figshare.com/) or ZENDO (https://zenodo.org/) where uploads are assigned a Digital Object Identifier (DOI) making them easily and uniquely citeable. LANaPD could be also accessible through other major databases broadly used so far like the ZINC 15 database [32]. NP databases such as BIOFACQUIM and AfroDB, for example, are accessible through ZINC 15 database.

3.3 Maintenance

Updating and maintaining compound databases is of critical importance for the sustained and timely use of the information. This is also a challenging step, in particular for public databases, because of issues of sustained funding that experience basically all research groups and consortiums. For instance, it is well-known that several web-servers in the public domain are discontinued after certain time [42]. In the NPs area an example is the Universal Natural Products Database [33] that, at that time, was the largest non-commercial and openly available database and contained 197,201 NPs from plants, animals, and microorganisms. The web-site hosting the database is no longer accessible. One of the workaround to address this problem is making use of repositories

with permanent link with a DOI number such as Fighshare or ZENDO (*vide supra*). Else, be successful in getting financial resources to sustain the web-site. An excellent example of such open compound database is ZINC 15 hosted by a research group at the University of California in San Francisco [32]. Other examples of public databases with sustained financial support are PubChem [43], ChEMBL [34], and DrugBank [44].

4. Conclusions

In line with the continued significance of NP to drug discovery and the accessibility of informatics resources, Latin American countries are developing compound databases with compounds available in their geographical region. Such efforts are part of a larger and global scale of research groups developing NP databases available in the public domain, representing the biodiversity of other countries. This far, Brazil, Mexico, and Panama have developed their databases releasing to the public the compounds and/or information of their contents. Other countries such as Colombia are also currently building a large database that will be released soon. The largest database this far is $NuBBE_{DB}$ from Brazil with over 2000 compounds. Building and maintaining all these databases are ongoing projects and the databases continue to grow as collecting the large biodiversity available in Latin American countries is challenging. It is expected that, putting all resources together in a single and unified compound database that can be called LANaPD, will have a significant contribution to NP research and NP-based drug discovery not only in Latin America but worldwide.

Acknowledgments

Discussions with members of the DIFACQUIM research group are greatly acknowledged.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629-661.
2. Newman, D.J. From natural products to drugs. *Physical Sciences Reviews* **2018**, *4*, 20180111.
3. Editorial. Redeploying plant defences. *Nature Plants* **2020**, *6*, 177-177.
4. Feher, M.; Schmidt, J.M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem Inf. Comput. Sci.* **2003**, *43*, 218-227.
5. Stratton, C.F.; Newman, D.J.; Tan, D.S. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4802-4807.
6. Chen, H.M.; Kogej, T.; Engkvist, O. Cheminformatics in drug discovery, an industrial perspective. *Mol. Inf.* **2018**, *37*.
7. Martinez-Mayorga, K.; Madariaga-Mazon, A.; Medina-Franco, J.L.; Maggiora, G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Exp. Opin. Drug Discov.* **2020**, *15*, 293-306.
8. Medina-Franco, J. New approaches for the discovery of pharmacologically-active natural compounds. *Biomolecules* **2019**, *9*, 115.
9. Miller, M.A. Chemical database techniques in drug discovery. *Nat Rev Drug Discov* **2002**, *1*, 220-227.
10. JingFang, Y.; Di, W.; Chenyang, J.; Mengyao, W.; GeFei, H.; GuangFu, Y. Freely accessible chemical database resources of compounds for in silico drug discovery. *Curr. Med. Chem.* **2019**, *26*, 7581-7597.
11. Duffy, B.C.; Zhu, L.; Decornez, H.; Kitchen, D.B. Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series. *Bioorg. Med. Chem.* **2012**, *20*, 5324-5342.

12. Naveja, J.J.; Rico-Hidalgo, M.P.; Medina-Franco, J.L. Analysis of a large food chemical database: Chemical space, diversity, and complexity. *F1000Research* **2018**, *7(Chem Inf Sci)*, 993.
13. Peña-Castillo, A.; Méndez-Lucio, O.; Owen, J.R.; Martínez-Mayorga, K.; Medina-Franco, J.L. Chemoinformatics in food science. In *Applied chemoinformatics*, Engel, T.; Gasteiger, J., Eds. Wiley: 2018; pp 501-525.
14. Medina-Franco, J.; Cruz-Lemus, Y.; Percastre-Cruz, Y. Chemoinformatic resources for organometallic drug discovery. *Computational Molecular Bioscience* **2020**, *10*, 1-11.
15. Pereira, F.; Aires-de-Sousa, J. Computational methodologies in the exploration of marine natural product leads. *Marine Drugs* **2018**, *16*, 236.
16. Prieto-Martínez F.D.; Norinder U.; J.L., M.-F. Cheminformatics explorations of natural products. In *Progress in the chemistry of organic natural products* Kinghorn A.; Falk H.; Gibbons S.; Kobayashi J.; Asakawa Y.; JK., L., Eds. Springer, Cham: 2019; Vol. 110.
17. Fullbeck, M.; Michalsky, E.; Dunkel, M.; Preissner, R. Natural products: Sources and databases. *Nat. Prod. Rep.* **2006**, *23*, 347-356.
18. Yongye, A.B.; Waddell, J.; Medina-Franco, J.L. Molecular scaffold analysis of natural products databases in the public domain. *Chem. Biol. Drug Des.* **2012**, *80*, 717-724.
19. Chen, Y.; de Bruyn Kops, C.; Kirchmair, J. Data resources for the computer-guided discovery of bioactive natural products. *J. Chem. Inf. Model.* **2017**, *57*, 2099-2111.
20. Chen, Y.; Garcia de Lomana, M.; Friedrich, N.-O.; Kirchmair, J. Characterization of the chemical space of known and readily obtainable natural products. *J. Chem. Inf. Model.* **2018**, *58*, 1518-1532.
21. Koulouridi, E.; Valli, M.; Ntie-Kang, F.; Bolzani, V.D.S. A primer on natural product-based virtual screening. *Physical Sciences Reviews* **2019**, *4*, 20180105.
22. Sorokina, M.; Steinbeck, C. On the redundancy of natural products public databases and where to find data in 2020 - a review on natural products databases. *Preprints* **2019**, 2019120332.
23. 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.

24. Nguyen-Vo, T.H.; Le, T.; Pham, D.; Nguyen, T.; Le, P.; Nguyen, A.; Nguyen, T.; Nguyen, T.N.; Nguyen, V.; Do, H., *et al.* Vietherb: A database for vietnamese herbal species. *J. Chem. Inf. Model.* **2019**, *59*, 1-9.
25. Naveja, J.J.; Oviedo-Osornio, C.I.; Trujillo-Minero, N.N.; Medina-Franco, J.L. Chemoinformatics: A perspective from an academic setting in latin america. *Mol. Divers.* **2018**, *22*, 247-258.
26. 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.
27. 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.
28. 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.
29. Olmedo, D.A.; Medina-Franco, J.L. Chemoinformatic approach: The case of natural products of Panama. In *Cheminformatics and its applications*, **2019**; IntechOpen, DOI: 10.5772/intechopen.87779. Available from: https://www.intechopen.com/online-first/chemoinformatic-approach-the-case-of-natural-products-of-panama.
30. Pilon-Jimenez, B.A.; Saldivar-Gonzalez, F.I.; Diaz-Eufracio, B.I.; Medina-Franco, J.L. Biofacquim: A Mexican compound database of natural products. *Biomolecules* **2019**, *9*, 31.
31. Sánchez-Cruz, N.; Pilón-Jiménez, B.; Medina-Franco, J. Functional group and diversity analysis of biofacquim: A mexican natural product database. *F1000Research* **2019**, *8*, 2071.
32. Sterling, T.; Irwin, J.J. Zinc 15 – ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324-2337.
33. 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.
34. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, María P.; Mosquera, Juan F.; Mutowo, P.; Nowotka, M., *et al.* Chembl: Towards direct deposition of bioassay data. *Nucl. Acids. Res.* **2019**, *47*, D930-D940.

35. Saldívar-González, F.I.; Valli, M.; Andricopulo, A.D.; da Silva Bolzani, V.; Medina-Franco, J.L. Chemical space and diversity of the nubbe database: A chemoinformatic characterization. *J. Chem. Inf. Model.* **2019**, *59*, 74-85.
36. Fatima, S.; Gupta, P.; Sharma, S.; Sharma, A.; Agarwal, S.M. ADMET profiling of geographically diverse phytochemical using chemoinformatic tools. *Fut. Med. Chem.* **2020**, *12*, 69-87.
37. Medina-Franco, J.L. Discovery and development of lead compounds from natural sources using computational approaches. In *Evidence-based validation of herbal medicine*, Mukherjee, P., Ed. Elsevier: 2015; pp 455-475.
38. José Luis, M.-F.; Norberto, S.-C.; B. Angélica, P.-J. *Biofacquim_v2.Sdf*. 2019.
39. Marcou, G.; Varnek, A. Data curation. In *Tutorials in chemoinformatics*, Varnek, A., Ed. Wiley: 2017; pp 1-36.
40. Naveja J.J.; Saldívar-González F.I.; Sánchez-Cruz N.; Medina-Franco J.L. Cheminformatics approaches to study drug polypharmacology. In *Multi-target drug design using chem-bioinformatic approaches. Methods in pharmacology and toxicology*, Roy, K., Ed. Humana Press: New York, NY, 2018.
41. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Da Silva Santos, L.B.; Bourne, P.E., *et al.* The fair guiding principles for scientific data management and stewardship. *Scientific Data* **2016**, *3*, 160018.
42. Villoutreix, B.O.; Lagorce, D.; Labbé, C.M.; Sperandio, O.; Miteva, M.A. One hundred thousand mouse clicks down the road: Selected online resources supporting drug discovery collected over a decade. *Drug Discovery Today* **2013**, *18*, 1081-1089.
43. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B., *et al.* Pubchem 2019 update: Improved access to chemical data. *Nucleic Acids Research* **2019**, *47*, D1102-D1109.
44. 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., *et al.* Drugbank 5.0: A major update to the drugbank database for 2018. *Nucl. Acids Res.* **2018**, *46*, D1074-D1082.