The Brazilian Compound Library (BraCoLi) database: a repository of chemical and biological information for drug design

Gabriel Corrêa Veríssimo[1]*, **Valtair Severino dos Santos Júnior**[1]*, Ingrid Ariela do Rosário de Almeida[1], Marina Sant'Anna Mitraud Ruas[1], Lukas Galuppo Coutinho[1], Renata Barbosa de Oliveira[1], Ricardo José Alves[1], Vinícius Gonçalves Maltarollo[1‡]

1. Departamento de Produtos Farmacêuticos, Faculdade de Farmácia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

***The authors equally contributed to the manuscript.**

‡**Correspondence author:** viniciusmaltarollo@gmail.com

ABSTRACT

The Brazilian Compound Library (BraCoLi) is a novel open access and manually curated electronic library of compounds developed by Brazilian research groups to support further computer-aided drug design works. Herein, the first version of the database is described comprising 1,176 compounds. Also, the chemical diversity and drug-like profiles of BraCoLi were defined to analyze its chemical space. A significant amount of the compounds fitted Lipinski and Veber's rules, alongside other drug-likeness properties. A comparison using principal component analysis showed that BraCoLi is similar to other databases (FDA-approved drugs and $NuBBE_{DB}$) regarding structural and physicochemical patterns. Furthermore, a scaffold analysis showed that BraCoLi presents several privileged chemical skeletons with great diversity. Despite the similar distribution in the structural and physicochemical spaces, similarity analysis indicated that compounds present in the BraCoLi are generally different from the two other databases showing an interesting innovative aspect, which is a desirable feature for novel drug design purposes.

Keywords: BraCoLi, drug design, database, chemical library, medicinal chemistry, cheminformatics.

INTRODUCTION

The application of computational tools as an ally in drug design was an important milestone in medicinal chemistry. This approach is now known as computer-aided drug design (CADD) and it is widely used to optimize the discovery and design of new drug candidates[1,2]. Molecular docking, structure-activity relationship (SAR) studies, and virtual screening are a few examples of available computational approaches that can be employed in drug design[3].

Currently, CADD data comes from individual works available in the literature or some websites. These strategies demand a large amount of chemical information to automatize the screening of novel bioactive compounds. Therefore, chemical databases were built as a resource to obtain this type of data more easily[4,5]. They were architected to store, organize, and enable the search for readily available and quantitative information for biological applications, as well as physicochemical and molecular properties of ligands and targets[6,7]. We listed some examples of free-to-access virtual chemical libraries that are extensively used on CAAD nowadays (**Table 1**).

Table 1. Open-access databases containing information about small molecules and their biological activities useful for CADD. The databases are shown in descending order of the number of compounds. Reported numbers were obtained in July 2021.

Database	Entries	Link	Reference
ZINC	750 M	http://zinc20.docking.org/	[8]
PubChem	110 M	https://pubchem.ncbi.nlm.nih.gov/	[9]
ChEMBL	2.1 M	https://www.ebi.ac.uk/chembl/	[10,11]
BindingDB	992 K	https://www.bindingdb.org/bind/index.jsp	[12]
TCM Database	37 K	http://tcm.cmu.edu.tw/about01.php?menuid=1	[13]
DrugBank	15 K	https://go.drugbank.com/	[14]
Drug Repurposing Hub	6.8 K	https://clue.io/repurposing	[15]
NuBBEDB	2.2 K	https://nubbe.iq.unesp.br/	[16,17]
AntibioticDB	1 K	https://www.antibioticdb.com/	[18]
AfroDB	954	http://zinc.docking.org/catalogs/afronp/	[19]
BIOFACQUIM	421	https://biofacquim.herokuapp.com/	[20]

Thus, the process of building libraries is critical since the data must be reliable to enable chemoinformatic experiments[21]. It is important to emphasize that *in silico* approaches help and speed up the search for new bioactive compounds. This can reduce the amount of compounds to be tested in *in vitro* and *in vivo* assays, anticipating adequate pharmacokinetic profiles, high selectivity, and low toxicity predictions[22,23]. In this way,

virtual libraries contribute to increase the success rate in the process of selecting new leads and to gather information with parsimony, ensuring the quality, variety, and consistency of the curated data[2,21,24].

Inspired by these examples, the Brazilian Compound Library (BraCoLi) was built as a manually curated and open-access electronic database containing biological and chemical information of synthetic and natural semisynthetic molecules from Brazilian research groups. In this work, a description and a cheminformatic characterization of BraCoLi based on chemical features and drug-like profiling, comparing with other databases is presented. The comparison was based on molecular, pharmaceutical and physicochemical properties of interest in drug design. The data was initially compilated from our research group on Pharmaceutical and Medicinal Chemistry from the Universidade Federal de Minas Gerais *(*UFMG), encouraging us to provide the dataset to the scientific community.

RESULTS AND DISCUSSION

Description of BraCoLi database and biological applications

To compose the first version of the BraCoLi database, 31 peer-reviewed thesis and papers that evaluated any biological activity of pure and characterized compounds from our research group were analyzed. 1,176 unique compounds derived from natural scaffolds or completely synthetic compounds were gathered. For each entry, the molecular formula, molecular weight (in g/mol), melting points (in °C) and, when available, biological information were reported. The structures are displayed in 3D lowest energy conformers and are available in Mol2 and SDF file formats. In addition, XLSX and PDF files with chemical and biological information and references regarding the compounds are also provided. All files are available for download at https://www.farmacia.ufmg.br/qf/downloads/. A flowchart of the strategy for the development of BraCoLi database is presented in **Figure 1**.

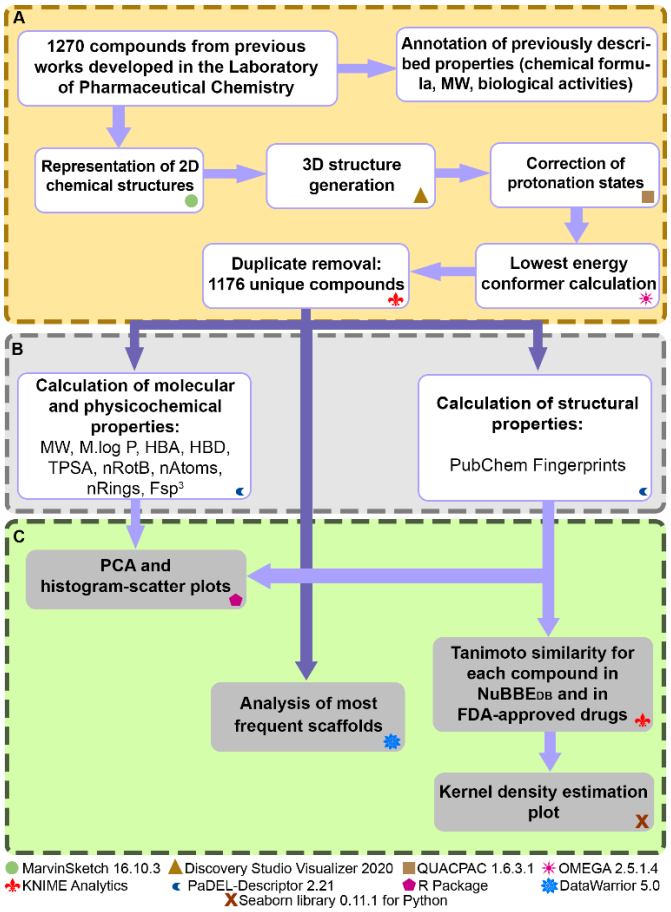


Figure 1. Flowchart highlighting the executed steps in BraCoLi construction. The main steps were grouped in **a)** dataset curation followed by structure treatment of the compounds, **b)** molecular and physicochemical properties prediction, and **c)** chemical space and diversity comparisons.

The substances showed a broad range of activity, with reported antibacterial, antifungal, antileishmanial, antimalarial, antioxidant, antitrypanosomal, antiviral, and cytotoxic activities. From BraCoLi database, two classes of compounds with great advance in drug development could be highlighted. Firstly, 2-thiazolylhydrazone derivatives (**Figure 2**) such as **RN104** and **RI76**, have shown promising *in vitro* and *in vivo* antifungal potential against both standard strains and clinical isolates of *Candida* and *Cryptococcus* species, [25–32]. Anti-diabetes and antioxidant activities were also reported for these analogs. These compounds have been evaluated in preclinical assays, including (i) analytical characterization[33], (ii) *in vivo*, *in vitro* and *in silico* pharmacokinetic and toxicity profiles[34,35], (iii) stability studies[36], and (iv) tests with different formulations to improve solubility[37].

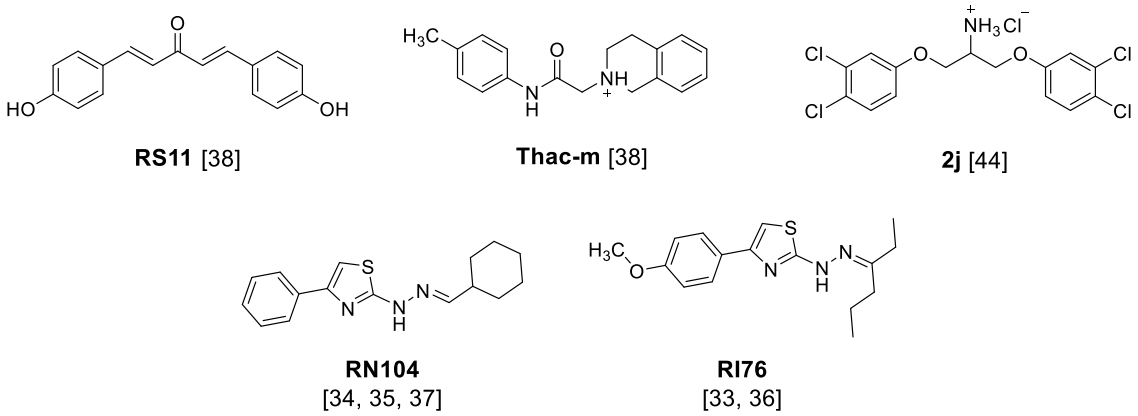


Figure 2. Structures of prominent bioactive compounds from BraCoLi database. Compounds **RS11** and **Thac-m** were selected by Asse Junior et al. (2020)[38] and presented MIC values of 125-250 µM against *S. aureus* and MRSA strains. **2j**, a 1,3-bis(aryloxy)propan-2-amine derivative, presented fungicidal activity against *Candida spp*. in infected *Drosophila melanogaster* flies[39]. **RI17** and **RN104** are potent 2-thiazolylhydrazone antifungal agents that were evaluated in the beforehand preclinical assays.

Other important set of substances includes the 1,3-bis(aryloxy)propan-2-amines and 1,3-bis(aryloxy)propan-2-ols derivatives such as compound **2j** (**Figure 2**) with a broad spectrum of activities. There are reports describing their *in vitro* antibacterial[40], antifungal[41], antileishmanial[42,43] and antitrypanosomal[44] activities, as well as a pharmacokinetic and pharmacodynamic characterization in *D. melanogaster* model of candidiasis[39] and a patent deposited in Brazil[45].

In addition, a preliminary unpublished version of the database has been applied to develop new antibacterial leads, exemplifying the application of BraCoLi in cheminformatics. Asse Junior and co-workers (2020) carried out a virtual screening to select potential Enoyl-ACP reductase (FabI) inhibitors. The authors carried out a ligand-based virtual screening via chemical similarity models using the in-house dataset alongside ZINC, FDA-approved drugs, TCM, and $NuBBE_{DB}$ databases. Four compounds were selected from BraCoLi and 2 of them (**Figure 2**, **RS11** and **Thac-m**) presented antibacterial activity against standard strains of *Staphylococcus aureus* and MRSA as well as clinical isolates[38].

Chemical space and drug-like profiling of BraCoLi

The chemical space of BraCoLi database was compared to 728 drugs approved by FDA between 1900 and 2017[46] and 2,223 compounds retrieved from NuBBEDB in terms of

chemical features and drug-like profiles. Firstly, nine molecular and physicochemical properties aiming to compare their drug-likeness were calculated: molecular weight (MW), logarithm *n*-octanol/water partition-coefficient calculated using the Moriguchi method (MLogP), number of hydrogen bond acceptors (HBA) and donors (HBD), topological polar surface area (TPSA), number of rotatable bonds (nRotB), fraction of sp^3 carbons (Fsp^3), number of atoms (nAtoms) and the number of rings (nRings). The drug-like potential of the compounds was analyzed based on two drug-likeness empirical rules: Lipinski's (Ro5) (MlogP ≤ 5, MW ≤ 500 Da, HBA ≤ 10, HBD ≤ 5)[47] and Veber's rules (TPSA ≤ 140 Å², nRotB ≤ 10)[48]. For the other properties, the value ranges provided by Ghose's (20 ≤ nAtoms ≤ 70)[49] and Muegge's (nRings ≤ 7)[50] rules were employed as references. These rules are applied to predict oral bioavailability of substances according to physicochemical fitness to the stablished ranges for each property.

The compounds presented a remarkably similar Gaussian distribution for MW and MLogP (**Figure 3a**). These databases present both hydrophilic and hydrophobic compounds (1 < MLogP < 7), indicating a good to moderate solubility in water probability whereas a good absorption in TGI. Most of the molecules (980 entries or 82.98% for MW and 1,160 entries or 98.22% for MLogP) fitted the quartile between the maximum values provided by Ro5. The average values for MW and MLogP were 378.58 Da and 2.58, respectively. Still in Ro5 discussion, **Figure 3b** shows a high population in the quartile between the adequate ratio of hydrogen-bond acceptors and donors, where 926 compounds (78.40%) showed no more than 10 HBA, and 1,103 (93.40%) presented no more than 5 HBD. Furthermore, 900 compounds (76.21%) fitted both conditions. Finally, **Figure 3c** represents a comparison between the two Veber's rules: number of rotatable bonds and topological polar surface area. Both parameters are related to the flexibility and capability of penetration in the cell membrane. Most molecules (887 entries or 75.11% for TPSA and 924 entries or 78.24% for nRotB) fitted the maximum values stablished by Veber and co-workers. The mean value for TPSA was 106.10 Å². Also, nRotB showed an average value of 7.77. The highest densities of points fitted the Lipinski and Veber's rules ranges (MW < 500 Da, 1 < MLogP < 5, HBA < 10, HBD < 5, nRotB < 10, TPSA < 140 Å²). At final count, 862 substances showed no violations and 133 showed one violation to Ro5 (totalizing 995 entries or 84.25%), 815 compounds (69%) fitted Veber's parameters, and 814 compounds (68.92%) fitted both empirical rules, showing a proper drug-like profile of the dataset.

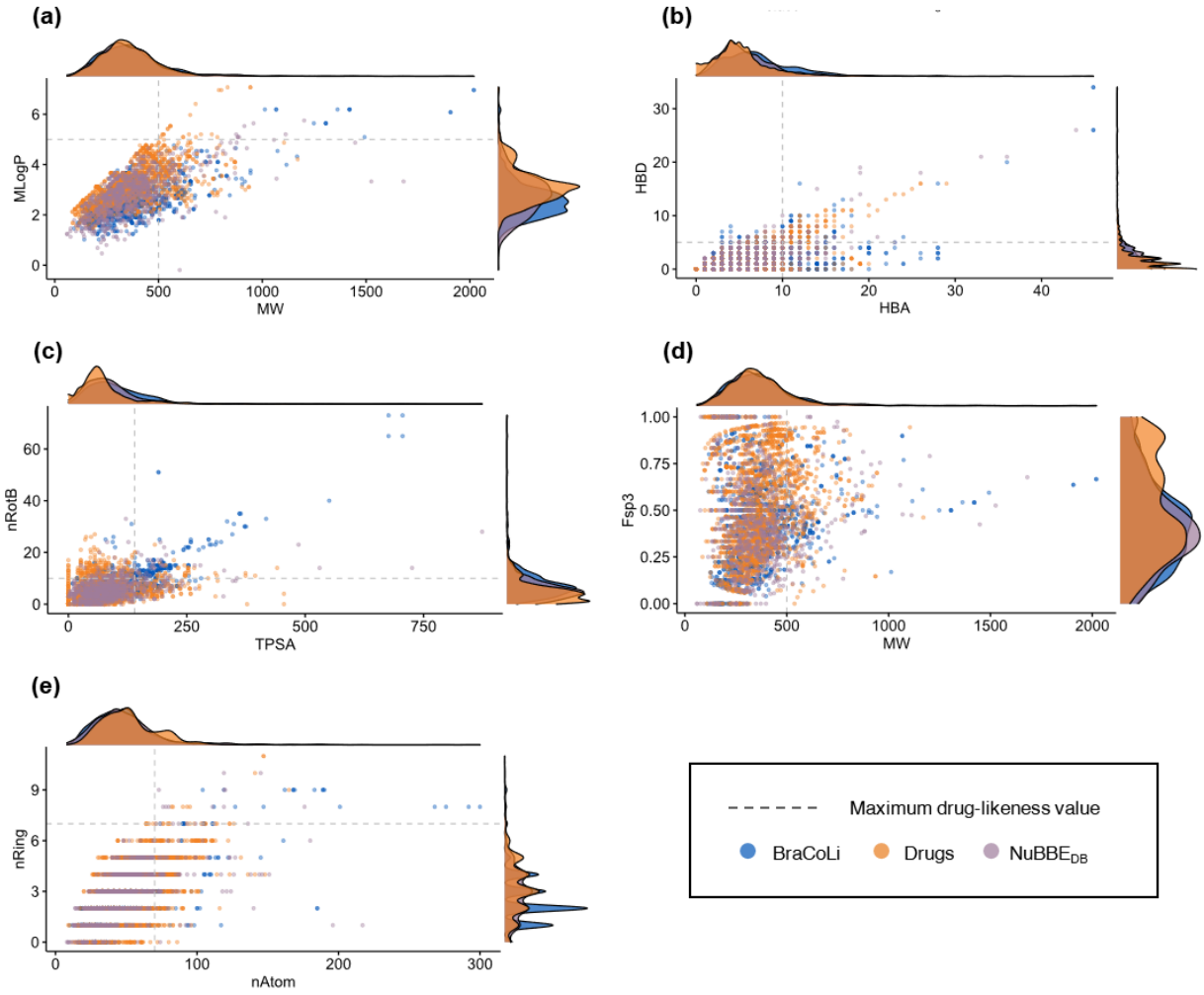


Figure 3. Scatter plots of the compounds of the BraCoLi (blue), FDA-approved drugs (orange), and NuBBEDB (lilac) according to molecular and physicochemical properties of relevance for drug-like profiling. The gray dashed line shows the range of each property according to Lipinski's and Veber's rules. The visual representations are **a)** MlogP vs MW, **b)** HBD vs HBA, **c)** nRotB vs TPSA, **d)** Fsp^3 vs MW, and **e)** nRing vs nAtoms. The graphics were obtained using R package.

Further comparisons were carried out to evaluate other physicochemical parameters. **Figure 3d** shows the comparison of the fraction of sp^3 carbons (Fsp^3) values to MW. Fsp^3 are related to the flexibility of the molecules, such as nRotB, and also represents the hybridization ratio of the structure. The average value of Fsp^3 was 0.43, meaning an approximately Csp^3 ratio of 1/2.3, indicating that the dataset contains more rigid than flexible structures. The parameters quantity of atoms (nAtoms) and quantity of rings (nRings) were also compared, as shown in **Figure 3e**, since these properties are related to the size of the molecules. As expected, the parameters presented high correlation to each other. The mean values calculated for nAtoms and nRings were 48.66 atoms and

2.45 rings, fitting the ranges predicted by the empirical rules, and 999 entries (84.59%) fit both rules simultaneously. It could be seen in all scatter plots from **Figure 3**, BraCoLi presented similar distributions to FDA-approved drugs and NuBBEDB in all comparisons.

The BraCoLi database, FDA-approved drugs, and NuBBEDB were compared regarding their chemical and structural spaces (**Figure 4**). The chemical spaces were generated employing principal component analysis (PCA), using two major approaches: (i) drug-like profiles in terms of the nine physicochemical properties evaluated beforehand (nAtoms, HBA, HBD, Fsp^3, MLogP, MW, nRings, nRotB, TPSA) (**Figure 4a**); and (ii) molecular fingerprints, according to PubChem fingerprints set (**Figure 4b**). Both plots are represented by the first two principal components (PC1 and PC2), where PC1 showed most contribution to the PCA (94.3% for drug-like-based PCA and 19.2% for fingerprint-based PCA). All three chemical sets show a similar distribution in the PCA plots, as expected from the drug-likeness analysis, indicating that the compounds present a comparable predicted pharmacokinetic profile. This indicate that, even they represent different datasets, BraCoLi presents an interesting applicability to discover lead candidates with adequate drug-like profiles in comparison to other largely used databases.

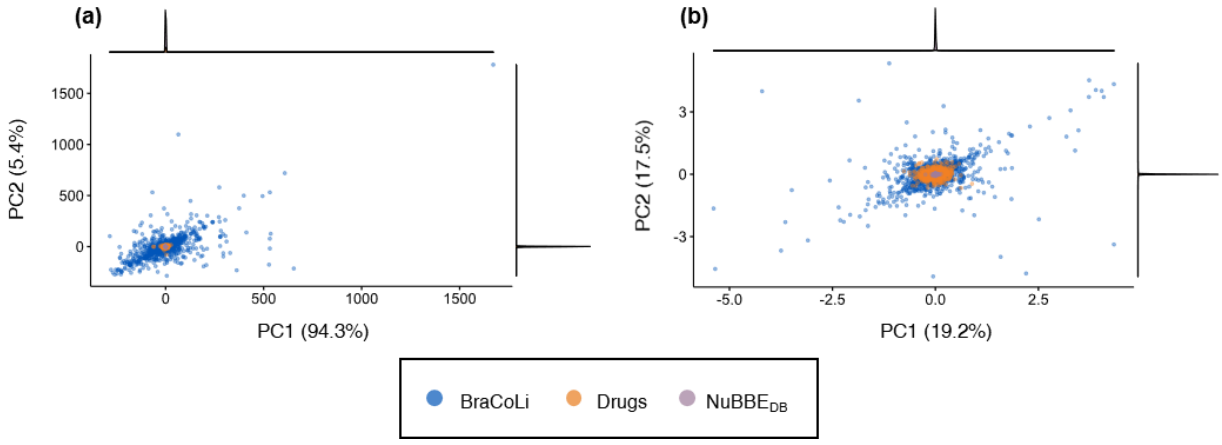


Figure 4. Chemical space visualization for the BraCoLi (blue), FDA-approved drugs (orange), and NuBBEDB (lilac) generated by PCA. The comparisons were based on **a)** drug-like profiles and **b)** molecular fingerprints. The graphics were obtained using R package.

Although these drug-likeness rules are still universally applied in the early stages of drug design, it is important to stress that some approved drugs violate them. Even Lipinski

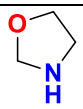
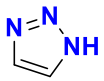
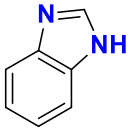
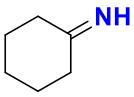
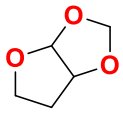
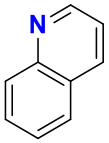
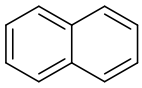
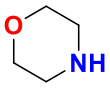
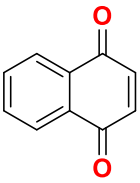
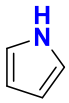
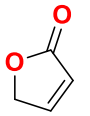
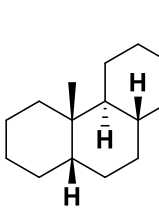
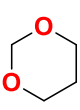
(2004) states that some scaffolds do not fit the Ro5 four parameters, especially natural products or derivatives and molecules that are recognized by active transport systems[51]. Pathania and Singh (2020) discuss in an editorial paper when is the ideal stage of drug development to take account on pharmacokinetic optimization and how empirical rules are helpful. According to the authors, 15 out of 26 FDA-approved small molecules in 2020 do not fit one or more drug-likeness rules. They suggest to apply those predictions after the evaluation of biological activity[52]. Other works also accent the necessity to revise those empirical rules after several years and expand the chemical space to fit new bioactive molecules with adequate experimental drug-like profiles[53–57]. Obviously, it is a compelling starting point that two-thirds of BraCoLi database present an adequate drug-like prediction, which can facilitate the screening of potential bioactive compounds, but it is necessary to balance both pharmacokinetic and pharmacodynamic profiles, since a molecule with a good pharmacokinetic profile does not necessarily present potent bioactivity and vice-versa[52,55].

Chemical scaffolds

To verify the chemical diversity of the dataset, the most frequent scaffolds were analyzed using the DataWarrior software. The software accounts how many times a certain ring system appears in the input dataset. We retrieved the twenty most frequent scaffolds in **Table 2**. As shown, most of them are heterocycles containing oxygen, nitrogen and/or sulfur in 5- or 6-membered rings. As discussed by Jampilek (2019), heterocycles are a very versatile group of structures with important applications in medicinal chemistry. Due to their privileged fragments, they present a broad spectrum of bioactivities as well as they can be easily modified or simplified to optimize pharmacodynamic and pharmacokinetic profiles[58].

We performed the same protocol using the $NuBBE_{DB}$ and FDA-approved drugs datasets and compared their top 20 most frequent scaffolds with BraCoLi's top 20. $NuBBE_{DB}$ showed five similar scaffolds (benzene, oxane, oxole, oxalane and naphthalene), where FDA-approved drugs showed nine similar scaffolds (benzene, pyridine, oxane, 1,3-thiazole, indole, oxalane, quinoline, morpholine, and naphthalene).

Table 2. Top 20 most frequent chemical scaffolds presented in the BraCoLi database and their respective frequencies.

				
benzene	***oxane***	***oxazolidine***	***1,2,3-triazole***	***1,3-thiazole***
1164 entries (98.9%)	455 entries (38.7%)	76 entries (6.46%)	76 entries (6.46%)	60 entries (5.1%)
				
benzimidazole	***cyclohexanimine***	***oxole***	***tetrahydrofuro[2,3-d][1,3]dioxole***	***quinoline***
50 entries (4.3%)	46 entries (3.9%)	44 entries (3.7%)	40 entries (3.4%)	37 entries (3.1%)
				
naphthalene	***morpholine***	***pyridine***	***naphthoquinone***	***oxalane***
36 entries (3.1%)	32 entries (2.7%)	32 entries (2.7%)	28 entries (2.4%)	24 entries (2%)
				
azole	***butanolide***	***cardenolide steroid***	***indole***	***1,3-dioxane***
22 entries (1.9%)	21 entries (1.8%)	20 entries (1.7%)	20 entries (1.7%)	17 entries (1.4%)

Finally, we compared the similarity between BraCoLi and the two datasets using Tanimoto coefficient calculated with PubChem fingerprints. This analysis helped us to verify how chemically distinctive BraCoLi is when compared to other scaffolds based on a mathematical similarity. The Tanimoto similarity values were plotted into a kernel density distribution (**Figure 5**). As shown, the distributions are quite different, as BraCoLi showed low to medium similarity to $NuBBE_{DB}$ and FDA-approved drugs datasets, where their Tanimoto similarities values for the highest densities were around 0.25 and 0.4, respectively. Despite the three chemical sets showed similar distributions in the drug-like profile analysis (**Figures 3** and **4**), they are chemically diverse.

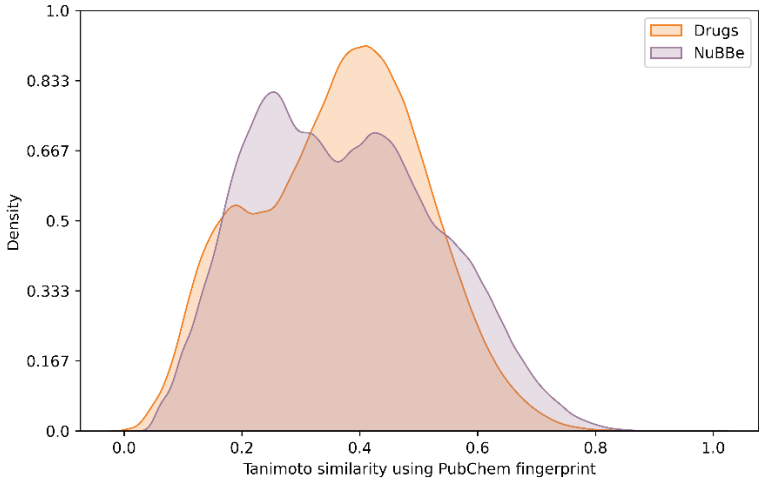


Figure 5. Kernel distribution plot for Tanimoto similarity comparison between BraCoLi with FDA-approved drugs (orange) and NuBBEDB (lilac).

CONCLUSION

After years cataloguing this information, a total of 1,176 compounds were gathered to build the BraCoLi database. As stablished in the cheminformatic characterization, the dataset present rich chemical diversity, broad spectrum of bioactivities and drug-like potential. More than two thirds of the compounds fitted both Lipinski's Ro5 and Veber's rules. These structures can be now explored by other medicinal chemistry groups to support–their research. The dataset update is planned when novel substances are obtained by our group and other Brazilian laboratories to expand BraCoLi database, being a novel platform to compilate and organize information on Brazilian-developed bioactive compounds for computational studies and experimental assays. Also, we are developing a database website to improve the visualization of the data.

MATERIALS AND METHODS

Curation and preparation of the biological and chemical data

The compounds were curated from prior works developed in the Laboratory of Pharmaceutical Chemistry (Faculty of Pharmacy, Federal University of Minas Gerais). Their chemical formula, molecular weight, and biological assays data were annotated. The 2D chemical structures were generated in Marvin Sketch 16.10.3 (Chemaxon, 2015). After, the structures were converted to a 3D format and had their conformation energy minimized using Discovery Studio Visualizer (BIOVIA, 2020). Also, any lacking hydrogen atoms were added to the structures. The most stable conformers were generated by OMEGA 2.5.1.4[59]. Ionization states in physiological pH (7.4) were corrected using fixpka software implemented in QUACPAC 1.6.3.1 (OpenEye Scientific

Software, 2016), in which the total energy was minimized using MMFF94 force field[60]. The 3D structures dataset is available in SDF and Mol2 file format.

Drug-like profiling and cheminformatic characterization

Molecular and physicochemical properties were calculated using PaDEL descriptor software[61]. R package was used to carry out statistical analysis. Principal component analysis (PCA) was carried out using prcomp function and histogram-scatter plots were generated via the function scatterhist.

Chemical diversity and substructures scaffolds

The chemical substructures were generated using the function "Analyse Scaffolds" and "Plain ring systems" filter criteria in software DataWarrior 5.2.1[62]. Applying the option "Split multiple values row", the most frequent and distinctive rings after the software counted the frequency of appearance of each substructure were manually verified. Ionized conjugate acids or bases were not differentiated from the non-ionized groups.

ACKNOWLEDGMENTS

The authors would like to thank FAPEMIG, CNPq, CAPES, and PRPq-UFMG for financial support and scholarships. Also, we would like to thank OpenEye Scientific Software for OMEGA and QUACPAC academic licenses.

REFERENCES

(1) Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25* (6), 1375. https://doi.org/10.3390/molecules25061375.

(2) Zhao, L.; Ciallella, H. L.; Aleksunes, L. M.; Zhu, H. Advancing Computer-Aided Drug Discovery (CADD) by Big Data and Data-Driven Machine Learning Modeling. *Drug Discovery Today* **2020**, *25* (9), 1624–1638. https://doi.org/10.1016/j.drudis.2020.07.005.

(3) Ferreira, L. L. G.; Andricopulo, A. D. Chemoinformatics Approaches to Structure- and Ligand-Based Drug Design. *Front. Pharmacol.* **2018**, *9*. https://doi.org/10.3389/fphar.2018.01416.

(4) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432* (7019), 862–865. https://doi.org/10.1038/nature03197.

(5) Brown, N.; Ertl, P.; Lewis, R.; Luksch, T.; Reker, D.; Schneider, N. Artificial Intelligence in Chemistry and Drug Design. *J Comput Aided Mol Des* **2020**, *34* (7), 709–715. https://doi.org/10.1007/s10822-020-00317-x.

(6) Nicola, G.; Liu, T.; Gilson, M. K. Public Domain Databases for Medicinal Chemistry. *J. Med. Chem.* **2012**, *55* (16), 6987–7002. https://doi.org/10.1021/jm300501t.

(7) Bajorath, J. Extending Accessible Chemical Space for the Identification of Novel Leads. *Expert Opinion on Drug Discovery* **2016**, *11* (9), 825–829. https://doi.org/10.1080/17460441.2016.1210126.

(8) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065–6073. https://doi.org/10.1021/acs.jcim.0c00675.

(9) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Research* **2021**, *49* (D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971.

(10) Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P. ChEMBL Web Services: Streamlining Access to Drug Discovery Data and Utilities. *Nucleic Acids Research* **2015**, *43* (W1), W612–W620. https://doi.org/10.1093/nar/gkv352.

(11) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Research* **2019**, *47* (D1), D930–D940. https://doi.org/10.1093/nar/gky1075.

(12) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Research* **2016**, *44* (D1), D1045–D1053. https://doi.org/10.1093/nar/gkv1072.

(13) Chen, C. Y.-C. TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening In Silico. *PLOS ONE* **2011**, *6* (1), e15939. https://doi.org/10.1371/journal.pone.0015939.

(14) 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.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Research* **2018**, *46* (D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037.

(15) Corsello, S. M.; Bittker, J. A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J. E.; Johnston, S. E.; Vrcic, A.; Wong, B.; Khan, M.; Asiedu, J.; Narayan, R.; Mader, C. C.; Subramanian, A.; Golub, T. R. The Drug Repurposing Hub: A next-Generation Drug Library and Information Resource. *Nature Medicine* **2017**, *23* (4), 405–408. https://doi.org/10.1038/nm.4306.

(16) 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* (3), 439–444. https://doi.org/10.1021/np3006875.

(17) Pilon, A. C.; Valli, M.; Dametto, A. C.; Pinto, M. E. F.; Freire, R. T.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S. NuBBE DB : An Updated Database to Uncover Chemical and Biological Information from Brazilian Biodiversity. *Scientific Reports* **2017**, *7* (1), 7215. https://doi.org/10.1038/s41598-017-07451-x.

(18) Farrell, L. J.; Lo, R.; Wanford, J. J.; Jenkins, A.; Maxwell, A.; Piddock, L. J. V. Revitalizing the Drug Pipeline: AntibioticDB, an Open Access Database to Aid Antibacterial Research and Development. *J Antimicrob Chemother* **2018**, *73* (9), 2284–2297. https://doi.org/10.1093/jac/dky208.

(19) 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* (10), e78085. https://doi.org/10.1371/journal.pone.0078085.

(20) 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* (1), 31. https://doi.org/10.3390/biom9010031.

(21) Zhao, L.; Wang, W.; Sedykh, A.; Zhu, H. Experimental Errors in QSAR Modeling Sets: What We Can Do and What We Cannot Do. *ACS Omega* **2017**, *2* (6), 2805–2812. https://doi.org/10.1021/acsomega.7b00274.

(22) Anuraj, N. Experimental and Computational Approaches to Improve Binding Affinity in Chemical Biology and Drug Discovery. *Current Topics in Medicinal Chemistry* **2020**, *20* (19), 1651–1660.

(23) Martinez-Mayorga, K.; Madariaga-Mazon, A.; Medina-Franco, J. L.; Maggiora, G. The Impact of Chemoinformatics on Drug Discovery in the Pharmaceutical Industry. *Expert Opinion on Drug Discovery* **2020**, *15* (3), 293–306. https://doi.org/10.1080/17460441.2020.1696307.

(24) Fourches, D.; Muratov, E.; Tropsha, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189–1204. https://doi.org/10.1021/ci100176x.

(25) Pereira de Sá, N.; Lino, C. I.; Fonseca, N. C.; Borelli, B. M.; Ramos, J. P.; Souza-Fagundes, E. M.; Rosa, C. A.; Santos, D. A.; Barbosa de Oliveira, R.; Johann, S. Thiazole Compounds with Activity against Cryptococcus Gattii and Cryptococcus Neoformans in Vitro. *European Journal of Medicinal Chemistry* **2015**, *102*, 233–242. https://doi.org/10.1016/j.ejmech.2015.07.032.

(26) Sá, N. P. de; Lima, C. M. de; Lino, C. I.; Barbeira, P. J. S.; Baltazar, L. de M.; Santos, D. A.; Oliveira, R. B. de; Mylonakis, E.; Fuchs, B. B.; Johann, S. Heterocycle Thiazole Compounds Exhibit Antifungal Activity through Increase in the Production of Reactive Oxygen Species in the Cryptococcus Neoformans-Cryptococcus Gattii Species Complex. *Antimicrobial Agents and Chemotherapy* **2017**, *61* (8). https://doi.org/10.1128/AAC.02700-16.

(27) Cruz, L. I. B.; Lopes, L. F. F.; De Camargo Ribeiro, F.; De Sá, N. P.; Lino, C. I.; Tharmalingam, N.; De Oliveira, R. B.; Rosa, C. A.; Mylonakis, E.; Fuchs, B. B.; Johann, S. Anti-Candida Albicans Activity of Thiazolylhydrazone Derivatives in Invertebrate and Murine Models. *Journal of Fungi* **2018**, *4* (4), 134. https://doi.org/10.3390/jof4040134.

(28) Sá, N. P.; Pôssa, A. P.; Pilar, P.; Jaqueline, M. S. F.; Nayara, C. F.; Cleudiomar, I. L.; Lana, B. C.; Renata, B. de O.; Carlos, A. R.; Beatriz, M. B.; Eleftherios, M.; Beth, B. F.; Susana, J. Antifungal Activity Directed Toward the Cell Wall by 2-Cyclohexylidenhydrazo- 4-Phenyl-Thiazole Against Candida Albicans. *Infectious Disorders - Drug Targets* **2019**, *19* (4), 428–438.

(29) Lino, C. I.; Gonçalves de Souza, I.; Borelli, B. M.; Silvério Matos, T. T.; Santos Teixeira, I. N.; Ramos, J. P.; Maria de Souza Fagundes, E.; de Oliveira Fernandes, P.; Maltarollo, V. G.;

Johann, S.; de Oliveira, R. B. Synthesis, Molecular Modeling Studies and Evaluation of Antifungal Activity of a Novel Series of Thiazole Derivatives. *European Journal of Medicinal Chemistry* **2018**, *151*, 248–260. https://doi.org/10.1016/j.ejmech.2018.03.083.

(30) Sá, N. P.; Lima, C. M.; A dos Santos, J. R.; Costa, M. C.; de Barros, P. P.; Junqueira, J. C.; Vaz, J. A.; Oliveira, R. B.; Fuchs, B. B.; Mylonakis, E.; Rosa, C. A.; Santos, D. A.; Johann, S. A Phenylthiazole Derivative Demonstrates Efficacy on Treatment of the Cryptococcosis & Candidiasis in Animal Models. *Future Science OA* **2018**, *4* (6), FSO305. https://doi.org/10.4155/fsoa-2018-0001.

(31) de Sá, N. P.; de Barros, P. P.; Junqueira, J. C.; Vaz, J. A.; de Oliveira, R. B.; Rosa, C. A.; Santos, D. A.; Johann, S. Thiazole Derivatives Act on Virulence Factors of Cryptococcus Spp. *Medical Mycology* **2019**, *57* (1), 84–91. https://doi.org/10.1093/mmy/myx158.

(32) Sá, N. P. de; Barros, P. P. de; Junqueira, J. C.; Valério, A. D.; Lino, C. I.; Oliveira, R. B. de; Rosa, C. A.; Johann, S. Antivirulence Activity and in Vivo Efficacy of a Thiazole Derivative against Candidiasis. *Journal of Medical Mycology* **2021**, *31* (2), 101134. https://doi.org/10.1016/j.mycmed.2021.101134.

(33) Franco, P. H. C.; Braga, S. F. P.; Oliveira, R. B. de; César, I. C. Purity Determination of a New Antifungal Drug Candidate Using Quantitative 1H NMR Spectroscopy: Method Validation and Comparison of Calibration Approaches. *Magnetic Resonance in Chemistry* **2020**, *58* (1), 97–105. https://doi.org/10.1002/mrc.4936.

(34) Silva, I. R.; Braga, A. V.; Gloria, M. B. de A.; Machado, R. de R.; César, I. C.; Oliveira, R. B. Preclinical Pharmacokinetic Study of a New Thiazolyl Hydrazone Derivative with Antifungal Activity in Mice Plasma by LC-MS/MS. *Journal of Chromatography B* **2020**, *1149*, 122180. https://doi.org/10.1016/j.jchromb.2020.122180.

(35) Tonholo, D. R.; Maltarollo, V. G.; Kronenberger, T.; Silva, I. R.; Azevedo, P. O.; Oliveira, R. B.; Souza, L. C. R.; Tagliati, C. A. Preclinical Toxicity of Innovative Molecules: In Vitro, in Vivo and Metabolism Prediction. *Chemico-Biological Interactions* **2020**, *315*, 108896. https://doi.org/10.1016/j.cbi.2019.108896.

(36) Franco, P. H. C.; Vieira, J. G.; Ramos, C. A. de O.; Johann, S.; Oliveira, R. B. de; César, I. C. Stability-Indicating Method for the Novel Antifungal Compound RI76: Characterization and in Vitro Antifungal Activity of Its Active Degradation Product. *Biomedical Chromatography* **2021**, *35* (3), e5014. https://doi.org/10.1002/bmc.5014.

(37) Silva, I. R.; Kronenberger, T.; Gomes, E. C. L.; César, I. C.; Oliveira, R. B.; Maltarollo, V. G. Improving the Solubility of an Antifungal Thiazolyl Hydrazone Derivative by Cyclodextrin Complexation. *European Journal of Pharmaceutical Sciences* **2021**, *156*, 105575. https://doi.org/10.1016/j.ejps.2020.105575.

(38) Asse Junior, L. R.; Kronenberger, T.; Magalhães Serafim, M. S.; Sousa, Y. V.; Franco, I. D.; Valli, M.; Silva Bolzani, V. da; Monteiro, G. C.; Bruno Prates, J. L.; Kroon, E. G.; Fernandes Mota, B. E.; Santos Ferreira, D. dos; de Oliveira, R. B.; Maltarollo, V. G. Virtual Screening of Antibacterial Compounds by Similarity Search of Enoyl-ACP Reductase (FabI) Inhibitors. *Future Medicinal Chemistry* **2019**, *12* (1), 51–68. https://doi.org/10.4155/fmc-2019-0158.

(39) Dalla Lana, D. F.; Kaminski, T. F. A.; Lavorato, S. N.; Merkel, S.; Zanette, R. A.; da Rosa, P. D.; Staudt, K. J.; de Araújo, B. V.; da Costa, B.; Quatrin, P. M.; Bazana, L. C. G.; Ferreira, F. A.; Caurio, C. F. B.; de Andrade, S. F.; Alves, R. J.; Fuentefria, A. M. In Vitro Pharmacokinetics/Pharmacodynamics Modeling and Efficacy against Systemic Candidiasis in Drosophila Melanogaster of a Bisaryloxypropanamine Derivative. *Medical Mycology* **2021**, *59* (1), 58–66. https://doi.org/10.1093/mmy/myaa030.

(40) Serafim, M. S. M.; Lavorato, S. N.; Kronenberger, T.; Sousa, Y. V.; Oliveira, G. P.; Santos, S. G. dos; Kroon, E. G.; Maltarollo, V. G.; Alves, R. J.; Mota, B. E. F. Antibacterial Activity of Synthetic 1,3-Bis(Aryloxy)Propan-2-Amines against Gram-Positive Bacteria. *MicrobiologyOpen* **2019**, *8* (11), e814. https://doi.org/10.1002/mbo3.814.

(41) Lana, D. F. D.; Lavorato, S. N.; Giuliani, L. M.; Cruz, L.; Lopes, W.; Vainstein, M. H.; Fontana, I. C.; Zimmer, A. R.; Freitas, M. de A.; Andrade, S. F. de; Alves, R. J.; Fuentefria, A. M. Discovery of a Novel and Selective Fungicide That Targets Fungal Cell Wall to Treat Dermatomycoses: 1,3-Bis(3,4-Dichlorophenoxy)Propan-2-Aminium Chloride. *Mycoses* **2020**, *63* (2), 197–211. https://doi.org/10.1111/myc.13027.

(42) Lavorato, S. N.; Duarte, M. C.; Lage, D. P.; Tavares, C. A. P.; Coelho, E. A. F.; Alves, R. J. 1,3-Bis(Aryloxy)Propan-2-Ols as Potential Antileishmanial Agents. *Chemical Biology & Drug Design* **2017**, *90* (5), 981–986. https://doi.org/10.1111/cbdd.13024.

(43) Lavorato, S. N.; Duarte, M. C.; Lage, D. P.; Tavares, C. A. P.; Coelho, E. A. F.; Alves, R. J. Synthesis and Antileishmanial Activity of 1,3-Bis(Aryloxy)Propan-2-Amines. *Med Chem Res* **2017**, *26* (5), 1052–1072. https://doi.org/10.1007/s00044-017-1805-1.

(44) Lavorato, S. N.; Sales Júnior, P. A.; Murta, S. M. F.; Romanha, A. J.; Alves/, R. J.; Lavorato, S. N.; Sales Júnior, P. A.; Murta, S. M. F.; Romanha, A. J.; Alves/, R. J. In Vitro Activity of 1,3-Bisaryloxypropanamines against Trypanosoma Cruzi-Infected L929 Cultures. *Memórias do Instituto Oswaldo Cruz* **2015**, *110* (4), 566–568. https://doi.org/10.1590/0074-02760150007.

(45) Coelho, E. A. F.; Alves, R. J.; Romanha, A. J.; Lavorato, S. N. Diarilaminas, composições farmacêuticas contendo as diarilaminas e usos. B.R. Patent BR1020140290788, November 21, 2014.

(46) Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62* (4), 1701–1714. https://doi.org/10.1021/acs.jmedchem.8b00686.

(47) 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 Settings1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews* **2001**, *46* (1), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0.

(48) 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* (12), 2615–2623. https://doi.org/10.1021/jm020017n.

(49) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1* (1), 55–68. https://doi.org/10.1021/cc9800071.

(50) Muegge, I.; Heald, S. L.; Brittelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44* (12), 1841–1846. https://doi.org/10.1021/jm015507e.

(51) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discovery Today: Technologies* **2004**, *1* (4), 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007.

(52) Pathania, S.; Singh, P. K. Analyzing FDA-Approved Drugs for Compliance of Pharmacokinetic Principles: Should There Be a Critical Screening Parameter in Drug Designing

Protocols? *Expert Opinion on Drug Metabolism & Toxicology* **2020**, *0* (0), 1–4. https://doi.org/10.1080/17425255.2021.1865309.

(53) Varma, M. V. S.; Obach, R. S.; Rotter, C.; Miller, H. R.; Chang, G.; Steyn, S. J.; El-Kattan, A.; Troutman, M. D. Physicochemical Space for Optimum Oral Bioavailability: Contribution of Human Intestinal Absorption and First-Pass Elimination. *J. Med. Chem.* **2010**, *53* (3), 1098–1108. https://doi.org/10.1021/jm901371v.

(54) Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *J. Med. Chem.* **2014**, *57* (2), 278–295. https://doi.org/10.1021/jm400887j.

(55) Lipinski, C. A. Rule of Five in 2015 and beyond: Target and Ligand Structural Limitations, Ligand Chemistry Structure and Drug Discovery Project Decisions. *Advanced Drug Delivery Reviews* **2016**, *101*, 34–41. https://doi.org/10.1016/j.addr.2016.04.029.

(56) Furukawa, A.; Schwochert, J.; Pye, C. R.; Asano, D.; Edmondson, Q. D.; Turmon, A. C.; Klein, V. G.; Ono, S.; Okada, O.; Lokey, R. S. Drug-Like Properties in Macrocycles above MW 1000: Backbone Rigidity versus Side-Chain Lipophilicity. *Angewandte Chemie* **2020**, *132* (48), 21755–21761. https://doi.org/10.1002/ange.202004550.

(57) Protti, Í. F.; Rodrigues, D. R.; Fonseca, S. K.; Alves, R. J.; Oliveira, R. B. de; Maltarollo, V. G. Do Drug-Likeness Rules Apply to Oral Prodrugs? *ChemMedChem* **2021**, *n/a* (n/a). https://doi.org/10.1002/cmdc.202000805.

(58) Jampilek, J. Heterocycles in Medicinal Chemistry. *Molecules* **2019**, *24* (21), 3839. https://doi.org/10.3390/molecules24213839.

(59) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572–584. https://doi.org/10.1021/ci100031x.

(60) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *Journal of Computational Chemistry* **1999**, *20* (7), 720–729. https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X.

(61) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *Journal of Computational Chemistry* **2011**, *32* (7), 1466–1474. https://doi.org/10.1002/jcc.21707.

(62) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55* (2), 460–473. https://doi.org/10.1021/ci500588j.