Annals of West University of Timişoara, ser. Biology, 2024, vol. 27 (1), pp.31-50

DATABASES FOR PLANTS THAT ARE USED IN COMPUTATIONAL BIOLOGY. *ARTEMISIA VULGARIS* CASE STUDY

Andrijana PUJICIC, Adriana ISVORAN*
[1]West University of Timisoara, Faculty of Chemistry, Biology, Geography, Department of Biology, Pestalozzi 16, Timisoara, 300115, Romania
**Corresponding author's e-mail: adriana.isvoran@e-uvt.ro*
Received 4 June 2024; accepted 30 July 2024

ABSTRACT
This study is intended to be a brief synthesis of the free accessible databases that are useful for implementing computational studies involving phytochemical constituents of general interest. For every database, the types of data it contains and the search capabilities it offers are described. Furthermore, the way of using every database and the provided information are illustrated when searching for the Artemisia vulgaris (or for other plant belonging to Artemisia sp. when there are dot data for A. vulgaris) and/or for one of its identified compounds. The study also emphasizes that phytochemical databases are rich source of data for implementing computational studies involving plant constituents.
KEY WORDS: *databases, phytochemical constituents, computational study, medicinal plant, Artemisia vulgaris*

1. INTRODUCTION

Databases are digital libraries that contain scientific information collected from the literature that disseminates the results of scientific experiments and computer analyses (Xiong, 2006). Biological data that are synthesized in biological databases can appear in various forms: genetic sequences, protein sequences, protein structures, protein functions, plant constituents, properties of chemical compounds with biological activity, medical findings, toxicological data, ecotoxicological data, etc. Each database has its own focus, containing specific types of biological data or broader collections of related information, allowing researchers to explore and analyze the data in multiple contexts. This emphasizes that biological databases help many scientists, educators, students, and other types of biological data stakeholders to deal with well-organized biological information that is continuously updated and usually freely and easily accessible.

The databases used in computational biology include information from a variety of research fields: genomic databases, protein databases, transcriptomics databases, metabolomics databases, pathway and interaction databases. There also are more specialized databases such as: Cancer Genome Atlas (TCGA) providing genomic data

related to cancer, ClinVar exposing clinically relevant variants in the human genome, database of Short Genetic Variants (dbSNP) also exposing genetic variation due to single nucleotide polymorphisms, information regarding the biologically active components of plants from different geographical areas and their biological activity. These databases are important tools for researchers and educators because they provide them with the opportunity to access a wide variety of biologically relevant data, which contributes to the understanding and explanation of many biological phenomena, from the structure and interactions of biomolecules to the evolution of species, respectively allow the development of scientific knowledge through the use and analysis of existing information (Kaisar *et al*., 2022). Regarding the use of existing databases in computational biology studies, the current information allows predictions about the biological, toxicological and ecotoxicological effects of various classes of chemicals, the development of new drugs, the design of chemicals with predetermined properties, etc.

There are numerous databases containing information regarding the plant compounds. In 2020, there were over 120 distinct databases and collections of natural products, 50 of them being open access (Sorokina & Steinbeck, 2020).

Artemisia vulgaris, known as common mugwort, is a plant used in food preparation and the food industry, in traditional medicine and pharmacy, having many beneficial effects. These uses are due to the biologically active components it contains (Ekiert *et al*., 2020).

The aim of this study is to present up-to-date and free accessible databases containing information on numerous biologically active plant compounds of general interest and to illustrate their use to obtain information on the constituents of *Artemisia vulgaris* (or of other *Artemisia sp*.) and their biological/medical applications. Some databases contain information regarding phytochemicals that are used in therapies of specific diseases (cancer, malaria) or for computational toxicology studies, but they are not considered in the present review.

The plant of interest in this study was *Artemisia dracunculus*, the plant used in the pharmaceutical industry and traditional medicine. With the help of the databases used for plant research, we identified the main chemical compounds that *Artemisia dracunculus* contains and obtained information about their biological and medicinal use.

2. DATABASES CONTAINING INFORMATION ON BIOLOGICALLY ACTIVE PLANT COMPONENTS OF GENERAL INTEREST AND THEIR POSSIBLE THERAPEUTIC EFFECTS

2.1. Indian Medicinal Plants, Phytochemistry and Therapeutics database (IMPPAT)

The Indian Medicinal Plants, Phytochemistry and Therapeutics free accessible database version 2 (IMPPAT2.0, https://cb.imsc.res.in/imppat/) contains information about 4010 Indian medicinal plants containing 17967 phytochemicals with known 1095 therapeutic uses associated and on 1133 traditional Indian medicinal formulations at the level of plant parts such as stem, root or leaves (Mohanraj *et al*., 2018; Vivek-Ananth *et al*., 2023). For the 17967 phytochemicals, IMPPAT2.0 also contains the 2D and 3D chemical structures and molecular properties and there is a subset of 1335 drug-like phytochemicals filtered using several drug-likeness rules. Moreover, molecular complexity and molecular scaffolding were constructed based on the structural diversity of the phytochemical space of the compounds contained in this database. Comparison of the phytochemicals contained in the IMPPAT2.0 database with compounds contained in other databases showed that the phytochemicals in IMPPAT2.0 have similar scaffold diversity to many natural product libraries and are enriched with specific protein binders (Vivek-Ananth *et al*., 2023).

The way to use the IMPPAT database involves entering in the search line the name for the plant of interest, *A. vulgaris* in this case, and the search returns a list of the phytochemicals contained in the plant of interest, the part of the plant in which it is found, an identifier of each substance in the database and a link to the bibliographic reference(s) from which the information was extracted (Figure 1). The search for *A. vulgaris* illustrates the presence of numerous chemical components, with myrtenol and ascaridole being among the components that are found in the aerial parts of the plant. The database also provides information on the therapeutic effects of the plants, part of the result obtained for *A. vulgaris* being reproduced in Figure 2. Thus, it is observed in this figure that the plant *A. vulgaris* can be used for therapeutic purposes as an anticoagulant, for the treatment of asthma, diseases of the nervous system and for the treatment of spasms.

The database can also be searched for a chemical compound using either its name or its identifier from IMPPAT2.0 (if known), in which case the search leads to the identification of plants containing that chemical compound, as shown in Figure 3 for the alpha-pinene. The search shows that the *A. vulgaris* plant contains the investigated

compound, which is found in the aerial parts of the plant, respectively in the stem and flowers.

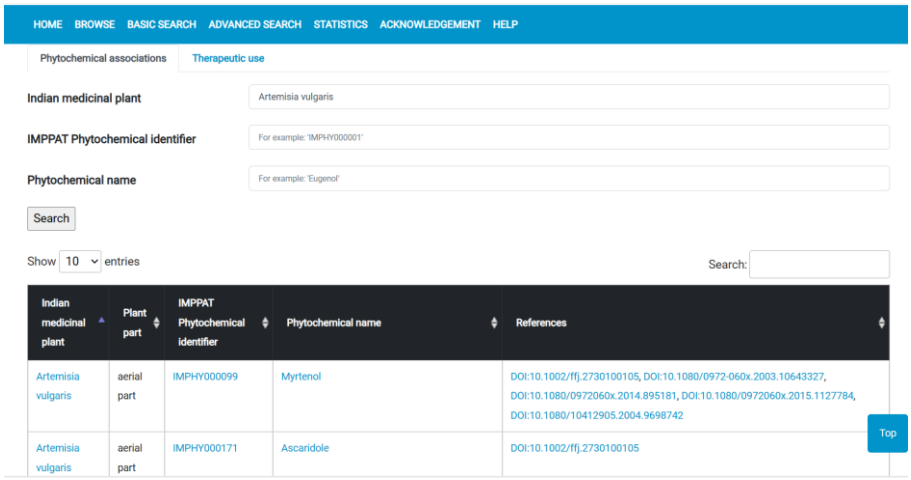


FIG.1. Illustration of how to use the IMPPAT2.0 database to obtain information on the chemicals contained in *Artemisia vulgaris*

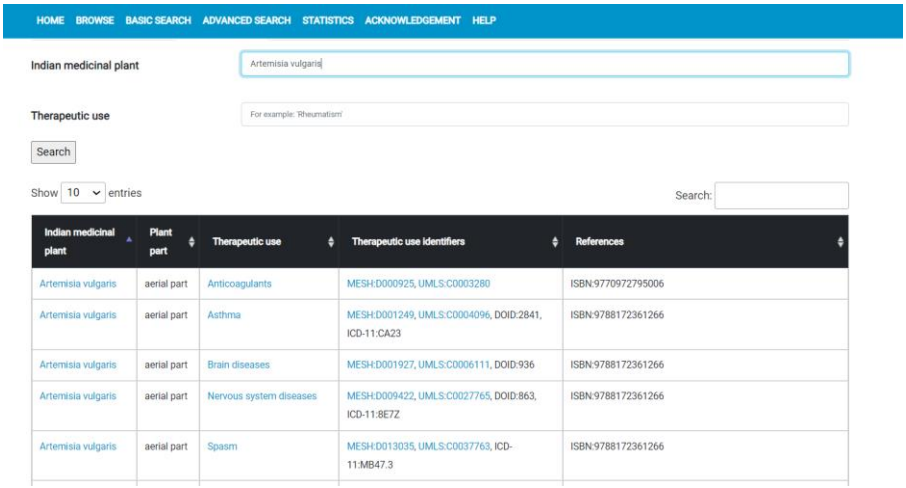


FIG. 2. Illustration of how to use the IMPPAT2.0 database to obtain information on the therapeutic effects of *Artemisia vulgaris*

Indian medicinal plant	Plant part	IMPPAT Phytochemical identifier	Phytochemical name	References
Artemisia vulgaris	aerial part	IMPHY012061	alpha-Pinene	DOI:10.1002/ffj.2730100105, DOI:10.1080/0972-060x.2003.10643327, DOI:10.1080/0972060x.2010.10643792, DOI:10.1080/0972060x.2014.895181, DOI:10.1080/0972060x.2015.1127784, DOI:10.1080/10412905.2004.9698742, DOI:10.1080/10412905.2006.9699118
Artemisia vulgaris	flower	IMPHY012061	alpha-Pinene	DOI:10.1080/10412905.2003.9698615
Artemisia vulgaris	fruit	IMPHY012061	alpha-Pinene	DOI:10.1080/10412905.2003.9698615
Artemisia vulgaris	leaf	IMPHY012061	alpha-Pinene	DOI:10.1080/0972060x.2019.1602083, DOI:10.1080/10412905.1992.9698101, DOI:10.1080/10412905.1997.10554252, DOI:10.1080/10412905.2003.9698615
Artemisia vulgaris	seed	IMPHY012061	alpha-Pinene	DOI:10.1002/ffj.1246
Artemisia vulgaris	stem	IMPHY012061	alpha-Pinene	DOI:10.1002/ffj.1246
Artemisia vulgaris		IMPHY012061	alpha-Pinene	DOI:10.1002/ffj.1246

FIG. 3. Illustration of how to use the IMPPAT 2.0 database to obtain information on plants containing the chemical compound alpha-pinene

Since its appearance in 2018, the IMPPAT database has been widely used in various studies. A search for the keyword "IMPPAT" in the scientific literature database ScienceDirect leads to a number of 82 articles (excluding the 2 articles describing the database) with an increasing trend from year to year (Figure 4) and taking into account the fact that the year 2024 is not over.

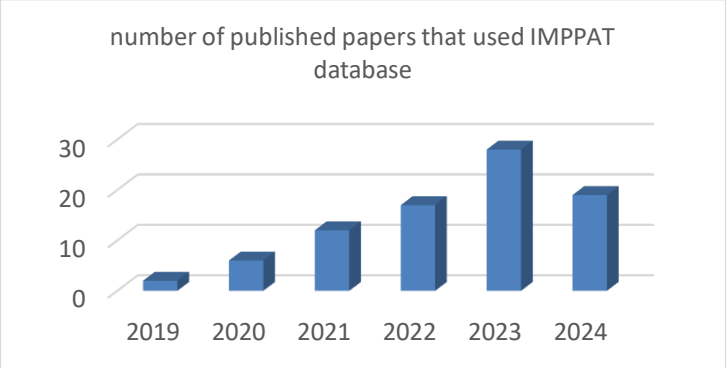


FIG. 4. Number of publications since 2019 in ScienceDirect citing the IMPPAT database

Of the 82 published articles, 59 are research articles, and of the latter 48 (81%) contain computational studies using information from the IMPPAT database. This proves the usefulness of this database for computational studies. There are also 7 citations of the IMPPAT database regarding *Artemisia sp.*

2.2. Collective Molecular Activities of Useful Plants database (CMAUP)

Collective Molecular Activities of Useful Plants database (CMAUP, https://bidd.group/CMAUP/) contains information regarding 7865 plants with identifiable taxonomic information at the level of species and genus. The 7865 plants include 2954 medicinal plants and 170 food plants collected on 79 countries/regions, and there also are 758 human target proteins of these plants (Zeng *et al*., 2019; Hou *et al.*, 2024). Information contained in this database also includes several scales of collective molecular activities of useful plants, and data obtained using bioinformatics and/or cheminformatics tool regarding, for example, gene ontology, bioavailability, absorption, distribution, metabolism, excretion and toxicity (ADMET) properties obtained using computational tools.

There are several ways to search for information in this database: by keywords (plant name, plant use classes, etc.), by chemical components to identify the plants that contain them, by the target protein investigated, by the disease to be treated, according to the genetic ontology, respectively according to the geographical location (Figure 5). The CMAUP database does not contain information on *A. vulgaris*, but searching for the keyword "Artemisia" yields information on molecular targets, genomic ontology, and associated diseases in addition to genus, family, and plant category (Figure 6). The information contained in this database reveals that plants of *Artemisia sp*. belong especially to the category of medicinal plants providing compounds with medicinal action. This underlines the importance of the information as it allows relevant computational studies for pharmaceutical applications of the components of these plants.

Since its appearance in 2019, the CMAUP database has accumulated a number of 23 citations, of which 9 (35%) refer to computational studies. These citations were obtained as follows: 2 in 2020, 3 in 2021, 6 in 2022, 5 in 2023 and 5 in the first 4 months of 2024, the increasing trend of citations being noted in this case as well. Given the recent enrichment of the content of the CMAUP database with detailed information at the molecular level that correlates plant ingredients with their therapeutic effects, we expect the number of citations to increase significantly in the next period.

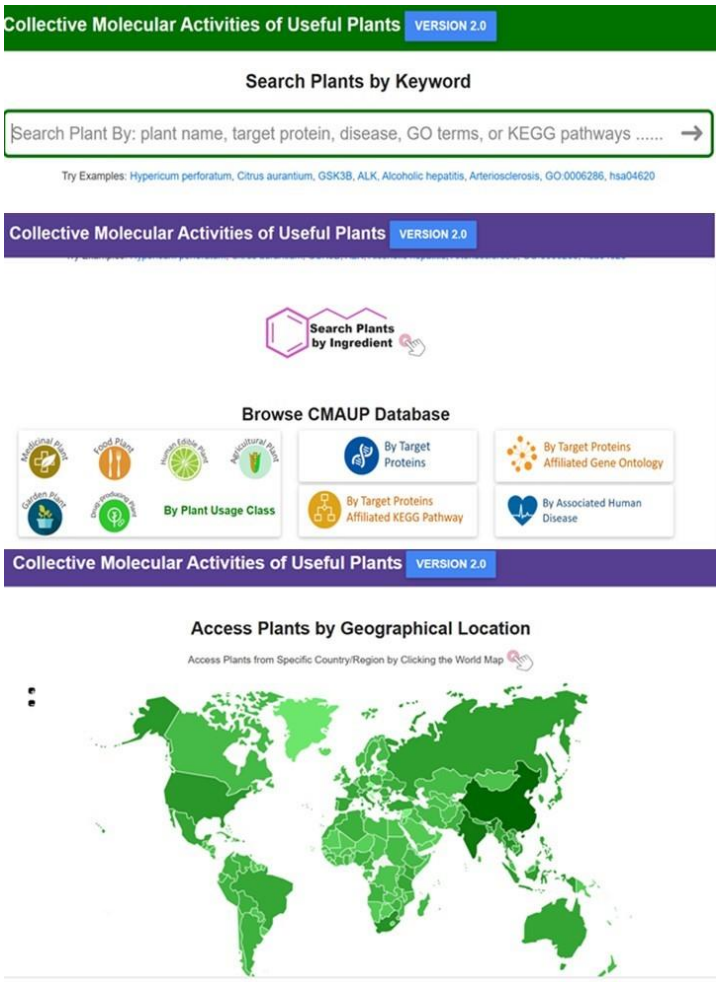


FIG. 5. Illustration of how to search for information in the CMAUP database

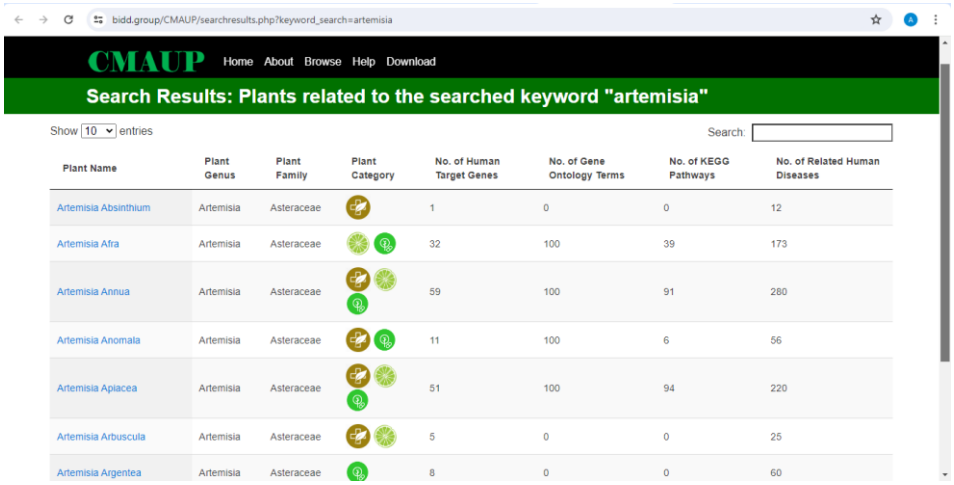


FIG.6. The result obtained when searching for the keyword "Artemisia" in the CMAUP database

2.3. Medicinal Plant Database of Bangladesh (MPDB)

Medicinal Plant Database of Bangladesh, version 2.0 (MPDB2.0 (https://www.medicinalplantbd.com/) contains information regarding more than 500 Bangladesh plants with medicinal use (plants of 122 family, 381 genus and 557 species). MPDB2.0 has a user-friendly interface and renders information about plants belonging to Asteraceae, Euphorbiaceae, Fabaceae, Lamiaceae, Poaceae and Solanaceae families that have potent medicinal and/or nutritive properties: local, scientific and family names, utilized parts of plants, active compounds, the diseases against which it is used, and a link(s) to the article(s) contained in the PubMed database from which the information presented is extracted. (Ashraf *et al*., 2014; Hussain *et al*., 2021).

The way to use the MPBD2.0 database involves searching for the plant of interest by its scientific name, the local name, plant genus, name of a chemical constituent, disease to be treated, plant parts used. How to use this database to obtain information about *A. vulgaris* is highlighted in Figure 7.

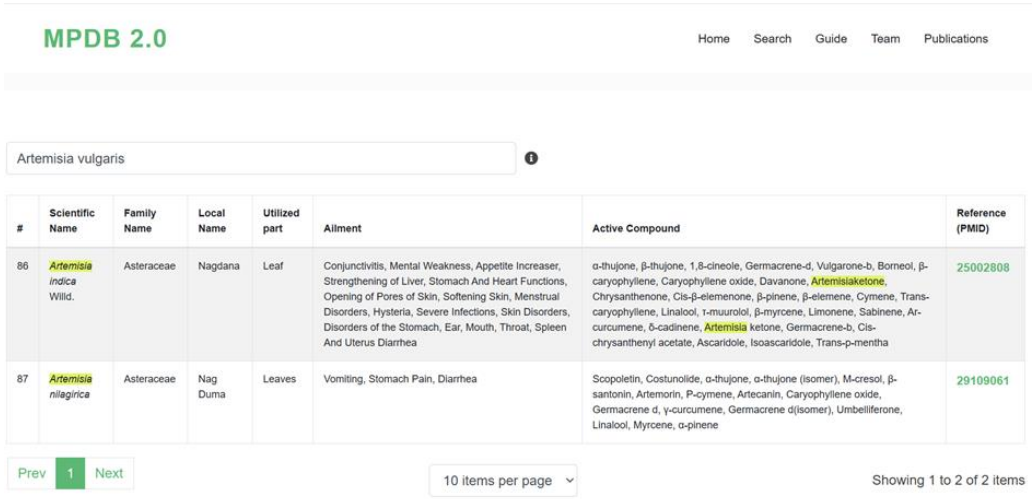


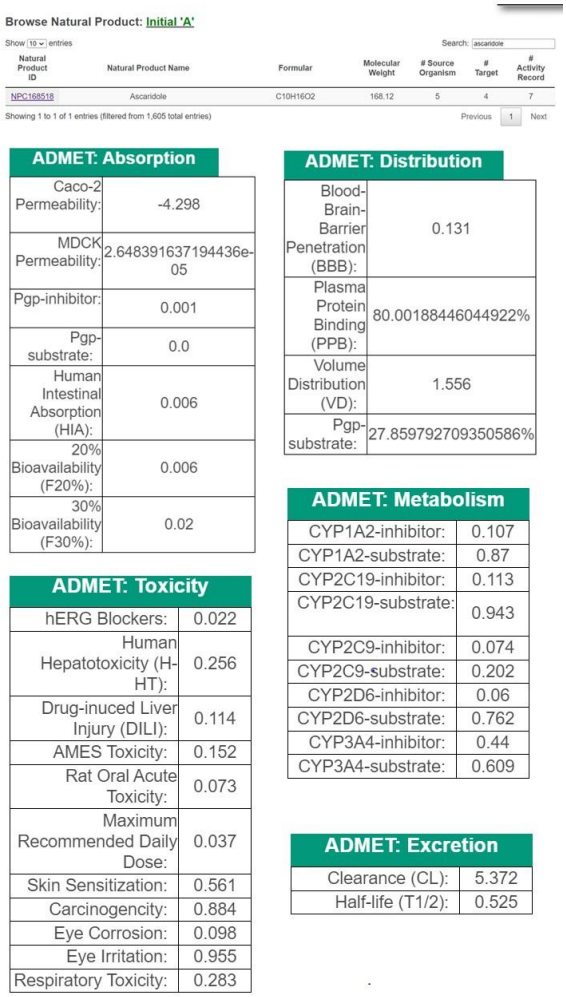
FIG.7. Information provided by the MPDB 2.0 database for *A. vulgaris*

Because the acronym MPDB is used in science to symbolize several names, it is complicated to identify the number of studies that cite the database. However, the information in the MPDB provides access to a collection of phytochemicals derived from Bangladesh medicinal plants, indicating it as an essential tool for researchers working in the field of new drug discovery, traditional medicine and phytochemistry. Thus, MPDB supports computer-aided drug discovery techniques, molecular docking and virtual screening, with the database's precise 3D structures enabling researchers to predict potential therapeutic interactions with target proteins. MPDB makes a significant contribution to the field of drug discovery based on the identification of the therapeutic potential of plant chemicals.

2.4. Natural Product Activity & Species Sources database (NPASS)

Natural Product Activity & Species Sources database, version 2.0, (NPASS2.0, http://bidd2.nus.edu.sg/NPASS/), contains 94413 unique natural products from 32287 organisms, 958866 activity records on 7753 targets including proteins (enzymes, transporters, ion channels, membrane receptors) and nucleic acids (Zeng *et al.*, 2018; Zhao *et al.*, 2023). For the compounds, NPASS2.0 database reveals the structures, physicochemical properties, biological activities, similarity with clinical drugs, etc. The way to use the NPASS2.0 database is seen in Figure 8. The name of the chemical

compound is entered in the search bar. In this case, we searched for the chemical compound ascaridole, identified using the IMPPAT database as being present in A. vulgaris.



Browse Natural Product: Initial 'A'

Show 10 entries — Search: ascaridole

Natural Product ID	Natural Product Name	Formular	Molecular Weight	# Source Organism	# Target	# Activity Record
NPC168518	Ascaridole	C10H16O2	168.12	5	4	7

Showing 1 to 1 of 1 entries (filtered from 1,605 total entries) — Previous 1 Next

ADMET: Absorption

Caco-2 Permeability:	-4.298
MDCK Permeability:	2.648391637194436e-05
Pgp-inhibitor:	0.001
Pgp-substrate:	0.0
Human Intestinal Absorption (HIA):	0.006
20% Bioavailability (F20%):	0.006
30% Bioavailability (F30%):	0.02

ADMET: Distribution

Blood-Brain-Barrier Penetration (BBB):	0.131
Plasma Protein Binding (PPB):	80.00188446044922%
Volume Distribution (VD):	1.556
Pgp-substrate:	27.859792709350586%

ADMET: Metabolism

CYP1A2-inhibitor:	0.107
CYP1A2-substrate:	0.87
CYP2C19-inhibitor:	0.113
CYP2C19-substrate:	0.943
CYP2C9-inhibitor:	0.074
CYP2C9-substrate:	0.202
CYP2D6-inhibitor:	0.06
CYP2D6-substrate:	0.762
CYP3A4-inhibitor:	0.44
CYP3A4-substrate:	0.609

ADMET: Toxicity

hERG Blockers:	0.022
Human Hepatotoxicity (H-HT):	0.256
Drug-inuced Liver Injury (DILI):	0.114
AMES Toxicity:	0.152
Rat Oral Acute Toxicity:	0.073
Maximum Recommended Daily Dose:	0.037
Skin Sensitization:	0.561
Carcinogencity:	0.884
Eye Corrosion:	0.098
Eye Irritation:	0.955
Respiratory Toxicity:	0.283

ADMET: Excretion

Clearance (CL):	5.372
Half-life (T1/2):	0.525

FIG. 8. Illustration of the results provided by NPASS2.0 database when searching for the chemical compound ascaridole identified in the plant of interest *Artemisia vulgaris*

By searching for the compound, we reach the page with the compound ID. Pressing the button with the compound ID takes us to the database page that displays more information about the chemical compound: IUPAC name, InCHI (International Chemical Identifier) code, SMILES formula, structure of the compound, chemical classification and physicochemical properties, fulfillment of several empirical rules regarding medicinal chemistry, chemically structural similarity (data not shown), and the ADME profile (Figure 8). Figure 8 shows that ascaridole reveals a good ADME profile. Regarding the toxicity of ascaridole, it may produce eye irritation and skin sensitization and there is a reasonable probability to be a carcinogenic compound. Furthermore, NPASS database provides information regarding the biological activity of ascaridole (Figure 9).

Show 10 entries Search:

Target ID	Target Type	Target Name	Target Organism	Activity Type	Activity Relation	Value	Unit	Reference
NPT4910	Organism	Aphelenchoides besseyi	Aphelenchoides besseyi	LC50	=	365.0	ug.mL-1	PMID[545689]
NPT450	Organism	Meloidogyne incognita	Meloidogyne incognita	LC50	=	52.0	ug.mL-1	PMID[545689]
NPT4910	Organism	Aphelenchoides besseyi	Aphelenchoides besseyi	mortality	<	50.0	%	PMID[545689]
NPT450	Organism	Meloidogyne incognita	Meloidogyne incognita	mortality	=	100.0	%	PMID[545689]
NPT580	Organism	Trypanosoma cruzi	Trypanosoma cruzi	MLC	=	23.0	uM	PMID[545688]
NPT580	Organism	Trypanosoma cruzi	Trypanosoma cruzi	Activity	=	8.5	%	PMID[545688]
NPT580	Organism	Trypanosoma cruzi	Trypanosoma cruzi	Activity	=	22.8	n.a.	PMID[545688]
NPT20529	NON-MOLECULAR	NON-PROTEIN TARGET	n.a.	IC80	>	500.0	uM	PMID[545687]
NPT20529	NON-MOLECULAR	NON-PROTEIN TARGET	n.a.	IC50	>	500000.0	nM	PMID[545687]
NPT20529	NON-MOLECULAR	NON-PROTEIN TARGET	n.a.	Activity	=	74.8	%	PMID[545687]

Showing 1 to 10 of 10 entries Previous 1 Next

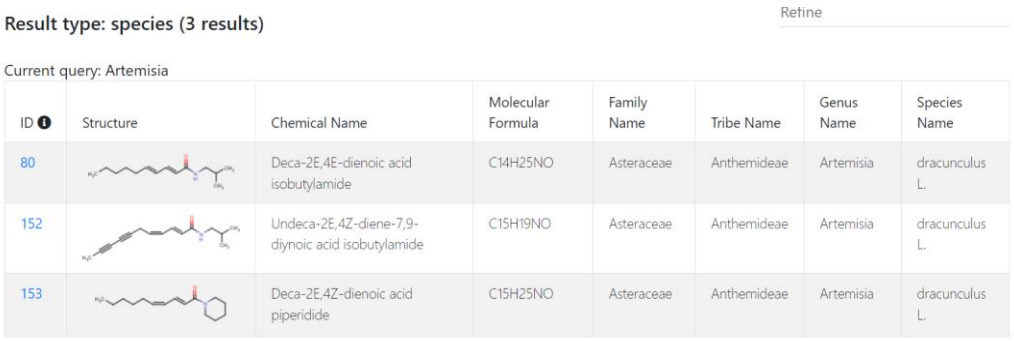
FIG. 9. Illustration of the information provided by NPASS2.0 database when searching for ascaridole and regarding its biological activity

The relevance of the NPASS database in research is great, in recent years it has been used in both experimental and computational studies. Currently, NPASS database accumulated 83 citations, 29 of each belong to pharmacology, toxicology and pharmaceutical science.

2.5. Alkamid database

The Alkamid database (https://alkamid.ugent.be/) is a more specialized resource containing information on plants containing N-alkylamides (NAAs), a group of compounds with potential applications in pharmacology, immunology and plant biology and which are present in over 25 plant families. This database provides ethnopharmacological and biofunctional data for NAAs, their physicochemical properties and plant origin. The database allows searching by NAA, by functionality, or by item specifications (author, title, year). Starting from a plant, which can be specified by family, genus or species, the Alkamid database provides the NAAs that occur in that plant. For any NAA, the Alkamid database provides information on significant chemical identifiers (common and IUPAC names, SMILES string, structural formula), most important physicochemical properties, known functionalities (anti-inflammatory, antimicrobial, antiprotozoal, insecticidal, cytotoxicity, etc.), and literature data (Boonen *et al.*, 2012).

When searching for the keyword ”Artemisia” in this database, the database provides information only for NAAs that are found on *Artemisia dracunculus*. The information regarding the properties, functionalities and literature for every NAA can be accessed by accessing its ID (Figure 10).


Result type: species (3 results)

Refine

Current query: Artemisia

ID	Structure	Chemical Name	Molecular Formula	Family Name	Tribe Name	Genus Name	Species Name
80		Deca-2E,4E-dienoic acid isobutylamide	C14H25NO	Asteraceae	Anthemideae	Artemisia	dracunculus L.
152		Undeca-2E,4Z-diene-7,9-diynoic acid isobutylamide	C15H19NO	Asteraceae	Anthemideae	Artemisia	dracunculus L.
153		Deca-2E,4Z-dienoic acid piperidide	C15H25NO	Asteraceae	Anthemideae	Artemisia	dracunculus L.

FIG. 10. Information provided by alkamid database when searching for ”Artemisia”, i.e. n-alkylamides contained in *Artemisia dracunculus*

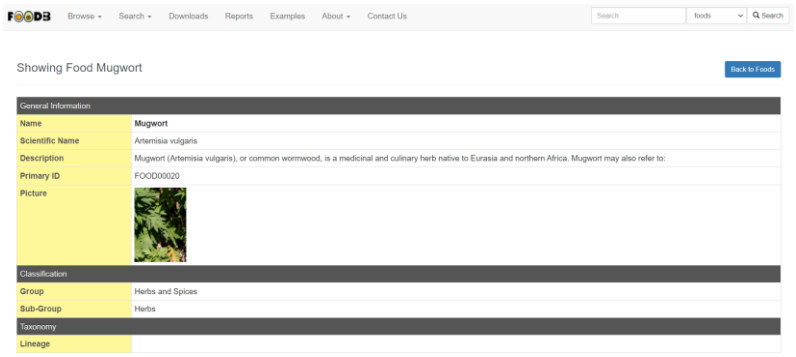
Like the other databases, ALKAMID has been used in computational studies, of the 9 citations of this database, 3 corresponding to its use in computational approaches.

2.6. FooDB- Food database

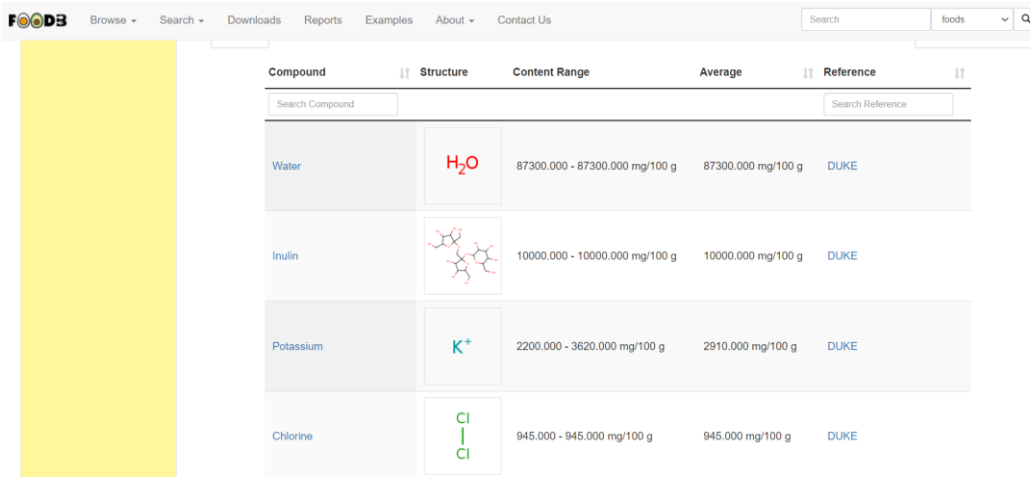
FooDB (https://foodb.ca/ - accessed on 02.02.2024) is a comprehensive and freely accessible resource that provides information on the chemical composition of various foods and food ingredients. For every constituent of a food, information provided is: nomenclature, structure, chemical class, molecular properties, physical properties, food source(s), physiological effects, probable health effects, and its concentrations in various foods. All this information is obtained from the scientific literature. The database covers a wide range of food components, including amino acids, fatty acids, flavor compounds, food additives, minerals, vitamins, phytochemicals.

The way to use the FooDB database involves searching by the name of the food, the name of the chemical compound or the name of the investigated nutrient. Figure 11 shows the information provided by the FooDB database for a search by the name of the plant of interest, *A. vulgaris*.

For a given chemical compound, it is possible to access multiple information: names (common name and IUPAC, synonyms), chemical class, physical and chemical properties, spectra, associated foods, biological effects and interactions, physiological pathways, references and external links (data not shown), such as to understand its role in different foods. The FoodB database has also been widely used in various studies. Searching for the keyword "FooDB" in the ScienceDirect database leads to a number of 368 articles of which 102 (28%) containing computational studies that used information from this database. Of these citations, 11 refer to studies performed on plants from *Artemisia sp*. The number of citations has an increasing trend from year to year, with 24 citations only in the first 4 months of 2024 (Figure 12).



A



B

FIG. 11. The information provided by FooDB for a search for *Artemisia vulgaris:* short information about this plant (A) and a print screen of a part of information regarding its chemical compounds (B)

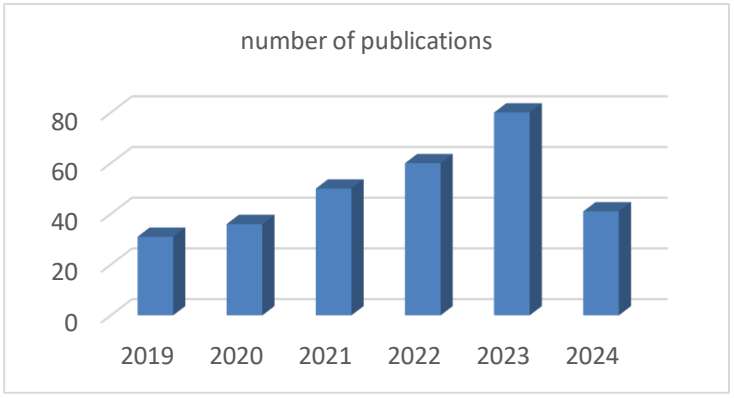


FIG. 12. Number of publications in the ScienceDirect database citing the FooDB database

These data illustrate that FooDB serves as a valuable tool for researchers, and it is also useful for computational studies.

2.7. Dr. Duke's Phytochemical and Ethnobotanical Databases

Managed by the United States Department of Agriculture, this database provides extensive information on the chemical constituents of plants, along with their ethnobotanical uses (https://phytochem.nal.usda.gov/ - accessed on 18.05.2024). The search in these databases is done according to the common or scientific name of the plant of interest and the provided information refers to chemical constituents, biological activity, and ethnobotanical use (Figure 13).

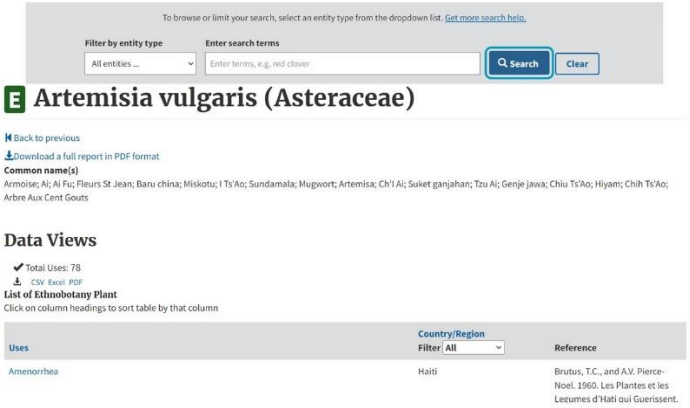


FIG. 13. The way to use Dr. Duke's Phytochemical and Ethnobotanical Databases

Search results can be viewed online or downloaded in PDF or CSV formats. The databases are of interest to pharmaceutical, nutritional, and biomedical research, as well as alternative therapies and herbal products.

2.8. OTHER DATABASES THAT CAN BE USED IN COMPUTATIONAL STUDIES

There are a large number of databases that contain information on the properties, structures and possible toxicological effects of biologically active compounds that can be found in plants, these databases being also utilizable in computational studies. Such databases are: ChEMBL (https://www.ebi.ac.uk/chembl/, Zdrazil *et al*., 2024), Chemical Entities of Biological Interest (ChEBI, https://www.ebi.ac.uk/chebi/, Hastings *et al.*, 2016), ChemSpider (https://www.chemspider.com/ - accessed in 16.04.2024), PubChem (https://pubchem.ncbi.nlm.nih.gov, Kim *et al*., 2023), OpenFoodTox (https:

//www.efsa.europa.eu/en/discover/infographics/openfoodtox-chemical-hazards-database - accessed on 16.04.2024), Traditional Chinese Medicine Database@Taiwan (TCM, http://tcm.cmu.edu.tw/ - accessed on 19.05.2024). Here, we limit ourselves to describing the PubChem database as it better suits the purpose of this study.

In the field of computational biology and chemistry, there are also needed databases containing information on macromolecules that may represent molecular targets of biologically active plant compounds. Within this study we only focus on databases providing information about proteins, UniProt (The UniProt Consortium, 2023) and Protein Data Bank (Berman *et al*., 2000).

2.8.1. PubChem database

PubChem is one of the most comprehensive public databases synthesizing information about chemical compounds, especially small molecules, from many data sources. Information for each chemical compound refers to nomenclature (including IUPAC name), structure, physicochemical properties, interactions, pharmacological properties, toxicological information (both human and environmental), etc. Thus, the PubChem database allows, when necessary, to complete the spectrum of information for the biologically active compounds present in plants. Data contained in PubChem have proven useful for numerous AI and machine learning studies (Kim *et al*., 2023).

The way to use the PubChem database involves entering in the search line either the name of the chemical compound of interest, or its SMILES formula or International Chemical Identifier. We illustrate how to use the database for the alpha-pinene compound found in *A. vulgaris*, (Figure 14). By accessing the file corresponding to alpha-pinene, the information contained in this database about this compound will be available. Thus, existing data in PubChem can be used in numerous drug discovery studies containing plant-derived compounds, with bioactivity data being used to build computational models for predicting bioactivity or toxicity, or to discover multi-target ligands.

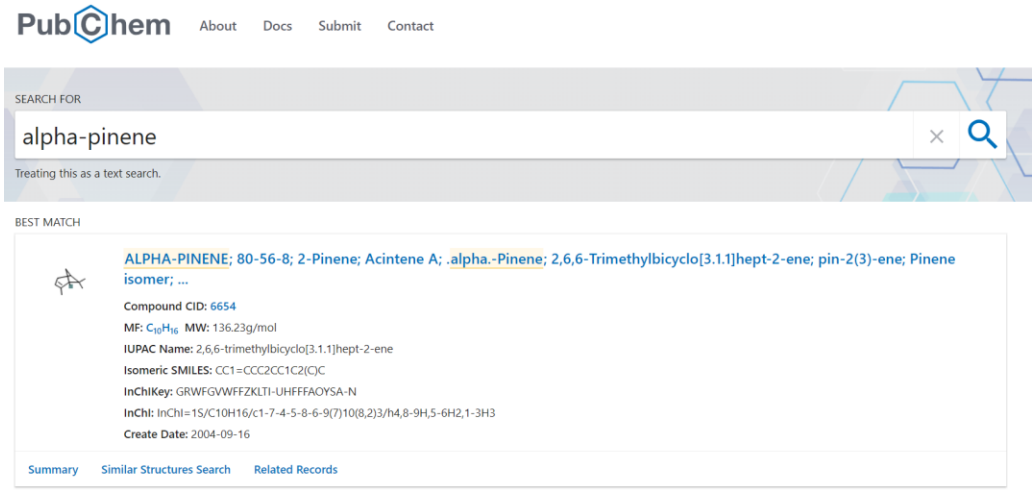


FIG. 14. Illustration of how to use the PubChem database to obtain information about the compound alpha-pinene

For example, the PubChem database was used to extract the SMILES formulas and physicochemical properties of acyclic monoterpenes identified as components of essential oils extracted from plants, and this information was then used to obtain predictions regarding the adverse effects of these components on humans (Dascalu *et al.*, 2023).

2.8.2. UniProt – The Universal Protein Resource

The UniProt database (https://www.uniprot.org/, The UniProt Consortium, 2023) provides valuable information on the sequences, cellular localization and biological functions of proteins that are molecular targets for phytochemicals identified in plants, thus supporting research in biotechnology, biological, biomedical, and pharmaceutical fields, as the understanding of protein functions is essential for the progress in these fields. Data presented in UniProt are extracted from multiple selected resources and are interpreted and standardized. The database is continuously enriched in content including newly available information regarding protein sequences and functions. UniProt is an important resource for computational studies as it integrates with various other databases and bioinformatics tools.

2.8.3. Protein Data Bank (PDB)

The Protein Data Bank (PDB, https://www.rcsb.org, Berman *et al.*, 2000) is a database that contains structures of biological macromolecules and the complexes formed by them with various ligands. The structures found in the PDB are determined either by X-ray diffraction, nuclear magnetic resonance (NMR), electron microscopy (EM), and molecular modeling. The data presented in a PDB file contain, in addition to the spatial coordinates of the constituent atoms of the system, general information necessary for all stored structures and information specific to the structure determination method. This database is heavily used in the fields of computational biology, bioinformatics and cheminformatics because it provides target protein structures for biologically active compounds, structures that are subsequently used in molecular modeling that provides information on interactions between plant chemicals and target proteins for disease treatment.

3. CONCLUSIONS

Databases play a crucial role in computational studies of phytochemicals, providing researchers with access to vast amounts of structured information about these compounds. The chemical structures of phytochemicals, their molecular formulas, physico-chemical properties, spectral data, but not only, are data that can be used for identifying and/or characterizing phytochemicals present in plant extracts or natural products. Many databases contain data regarding (or allow prediction of) information about the biological activities of phytochemicals, such as their pharmacological effects, toxicity and mechanisms of action. Therefore, researchers can also use these databases to discover correlations, patterns, and/or relationships between chemical structures and biological activities, facilitating the development and/or improvement of predictive models for identifying compounds in plants that are efficient in drug discovery and design. Phytochemicals contained in databases can be also used in virtual screening against molecular targets involved in disease pathways and to identify compounds with therapeutic potential. Consequently, using these databases, researchers in the field of computational biology and chemistry can conduct studies meant to identify phytochemicals with specific biological activities relevant to their research interests.

REFERENCES

- Ashraf M.A., Khatun A., Sharmin T., Mobin F., Tanu A.R., Morshed T., Fakir T.A., Begum R.A., Nabi A.N. 2014. MPDB 1.0: a medicinal plant database of Bangladesh. *Bioinformation* 10: 384–386.

- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.
- Boonen J., Bronselaer A., Nielandt J., Veryser L., De Tré G., De Spiegeleer B. 2012. Alkamid database: Chemistry, occurrence and functionality of plant N-alkylamides. *J Ethnopharmacol* 142: 563-590.
- ChemSpider (https://www.chemspider.com/ - accessed in 16.04.2024)
- Dascalu D., Isvoran A., Ianovici N. 2023. Molecules, Predictions of the Biological Effects of Several Acyclic Monoterpenes as Chemical Constituents of Essential Oils Extracted from Plants. *Molecules* 28: e4640.
- Dr. Duke's Phytochemical and Ethnobotanical Databaseshttps://phytochem.nal.usda.gov/ - accessed in 18.05.2024
- Ekiert H., Pajor J., Klin P., Rzepiela A., Ślesak H., Szopa A. 2020. Significance of *Artemisia Vulgaris L.* (Common Mugwort) in the History of Medicine and Its Possible Contemporary Applications Substantiated by Phytochemical and Pharmacological Studies. *Molecules* 25: e4415.
- FooDB (https://foodb.ca/ - accesat la 02.02.2024
- Hastings J., Owen G., Dekker A., Ennis M., Kale N., Muthukrishnan V., Turner S., Swainston N., Mendes P., Steinbeck C. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* 44: D1214-D1219.
- Hou D., Lin H., Feng Y., Zhou K., Li X., Yang Y., Wang S., Yang X., Wang J., Zhao H., Zhang X., Fan J., Lu S., Wang D., Zhu L., Ju D., Chen Y.Z., Zeng X. 2024. CMAUP database update 2024: extended functional and association information of useful plants for biomedical research. *Nucleic Acids Res* 5: 1508-1518.
- Hussain N., Chanda R., Abir R.A., Mou M.A., Hasan K., Ashraf M.A. 2021. MPDB 2.0: a large scale and integrated medicinal plant database of Bangladesh. *BMC Research Notes* 14: e301.
- Kaisar M., El-Attar N., Elmashad S., Abdelkader W. 2022. Perspectives review and Challenges in Biological Databases Integration. *Alfarama Journal of Basic & Applied Sciences* 3: 346-362.
- Kim S., Chen J., Cheng T., Gindulyte A., He J., He S., Li G., Shoemaker B.A., Thiessen P.A., Yu B., Zaslavsky L., Zhang J., Bolton E.E. 2023. PubChem 2023 update. *Nucleic Acids Res* 51: D1373-D1380.
- Mohanraj K., Karthikeyan B.S., Vivek-Ananth R.P., Chand R.P.B., Aparna S.R., Mangalapandi P., Samal A. 2018. IMPPAT: A Curated Database of Indian Medicinal Plants, Phytochemistry and Therapeutics. *Sci. Rep* 8: e4329.
- OpenFoodTox (https: //www.efsa.europa.eu/en/discover/infographics/openfoodtox-chemical-hazards-database - accessed in 16.04.2024
- Sorokina M., Steinbeck C. 2020. Review on natural products databases: where to find data in 2020. *J Cheminform* 12: 1-51.
- Traditional Chinese Medicine Database@Taiwan (TCM, http://tcm.cmu.edu.tw/ - accessed in 19.05.2024)
- UniProt Consortium. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 51: D523-D531.
- Vivek-Ananth R.P., Mohanraja K., Sahooa A.K., Samal A. 2023. IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. *ACS Omega* 8: 8827–8845.
- Xiong J. 2006. *Essential Bioinformatics*, ed. 1, Editura Cambridge University, Cambridge, 331 p.
- Zdrazil B., Felix E., Hunter F., Manners E.J., Blackshaw J., Corbett S., De Veij M., Ioannidis H., Mendez Lopez D., Mosquera J.F., Magarinos M.P., Bosc N., Arcila R., Kizilören T., Gaulton A., Bento A.P., Adasme M.F., Monecke P., Landrum G.A., Leach A.R. 2024. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* 52: D1180–D1192.
- Zeng X., Zhang P., He W., Qin C., Chen S., Tao L., Wang Y., Tan Y., Gao D., Wang B., Chen Z., Chen W., Jiang Y.Y., Chen Y.Z. 2018. NPASS: Natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Research* 46: D1217–D1222.

- Zeng X., Zhang P., Wang Y., Qin C., Chen S., He W., Tao L., Tan Y., Gao D., Wang B., Chen Z., Chen W., Jiang Y.Y., Chen Y.Z. 2019. CMAUP: a database of collective molecular activities of useful plants. *Nucleic Acids Research* 47: D1118–D1127.
- Zhao H., Yang Y., Wang S., Yang X., Zhou K., Xu C., Zhang X., Fan J., Hou D., Li X., Lin H., Tan Y., Wang S., Chu X.Y., Zhuoma D., Zhang F., Ju D., Zeng X., Chen Y.Z. 2023. NPASS database update 2023: quantitative natural product activity and species source database for biomedical research. *Nucleic Acids Res* 51: 621-628.