NANPDB: A database of Natural Products from Northern Africa

Telukunta, KK[a]; Ntie-Kang, F[b]; Döring, K[a]; Simoben, C.V[b]; Moumbock, A.F.A[c]; Malange, Y.I[c]; Njume, L.E[d]; Yong, J.N[c]; Sippl[b], W; Günther S[a,e]

[a]*Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, University of Freiburg, Germany*
[b]*Dept. of Pharmaceutical Chemistry, Martin-Luther University of Halle-Wittenberg, Germany*
[c]*Dept. of Chemistry and* [d]*Chemical & Bioactivity Information Centre, Dept. of Chemistry, Faculty of Science, University of Buea, Cameroon*
[e]*Freiburg Institute for Advanced Studies (FRIAS, University of Freiburg, Germany)*

kiran.telukunta@pharmazie.uni-freiburg.de

What is NANPDB ?

Northern African Natural Products Database (NANPDB)[1] is an online accessible database of Natural products(NPs), which are main sources of drugs and drug leads and play an important role in drug discovery by providing novel scaffolds[2,3]. In NANPDB we have covered from Northern African region which is spreading over 9 million kms[2,4] of Africa continent. It is a significant region as concluded in a recent UNO survey that this part of the world is a huge repository of bioactive NPs. For the Database, we have collected information (see Figure 4), which covered literature sources for the period from 1962 to 2016. The data consists of ~4500 NPs isolated from plants, animal (e.g. Coral), fungal, and bacterial sources.

Current data of NANPDB

unique SMILES	plant families	source organisms	biological activities	modes of action
4469	146	617	98	37
unique PubChem IDs	**kingdoms** (majorly from plantae, animal, bacteria and fungi)	**cited references**	**PubMed references**	**compound classes**
2059	5	787	324	95

Compound classes

A significant percentage of identified compounds are terpenoids, flavonoids, and alkaloids. About one-fifth of the compounds have been tested to be active in at least one bioassay.

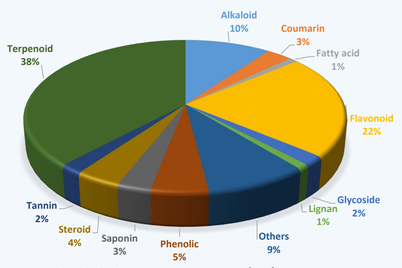

Figure 1: Main compound classes

Biological activities

The majority of compounds among recorded 98 bioactivities are anti-infective, cytotoxic and potential anticancer drugs. Other class includes compounds exhibiting inhibitory activities against recorded antileukemic, kinase inhibitors, and tumour anti-initiating compounds.

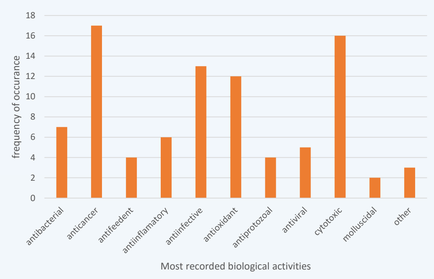

Figure 2: Frequently occured biological activities

Simplified database schema

About 2000 compounds were present in PubChem[5]. Similarly, half of 781 literature references are currently listed in PubMed, indicating unique nature of data in NANPDB. The gives also Toxicity predictions based on the pkCSM model[6]. For the taxonomic information, compounds are linked to NCBI database. To describe source species, NANPDB has been linked to the Prota[7] and the Tropicos[8] database. The complete data is organized into optimized schema as shown in Figure 3. Database is implemented on PostgreSQL 9.5 version and the Python-Django-Mezzanine Web-Framework.

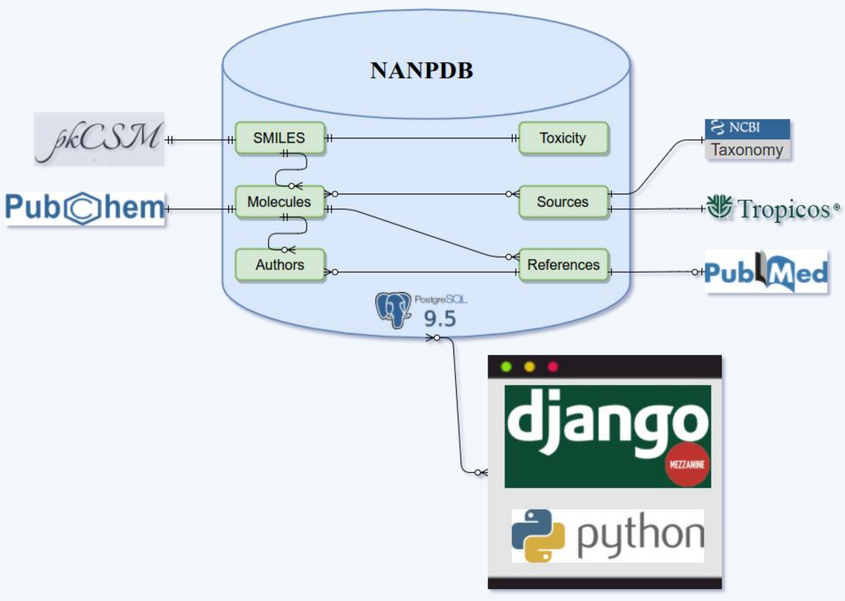

Figure 3: Schematic diagram of database and web server

"Northern African natural products ready for Drug Discovery"

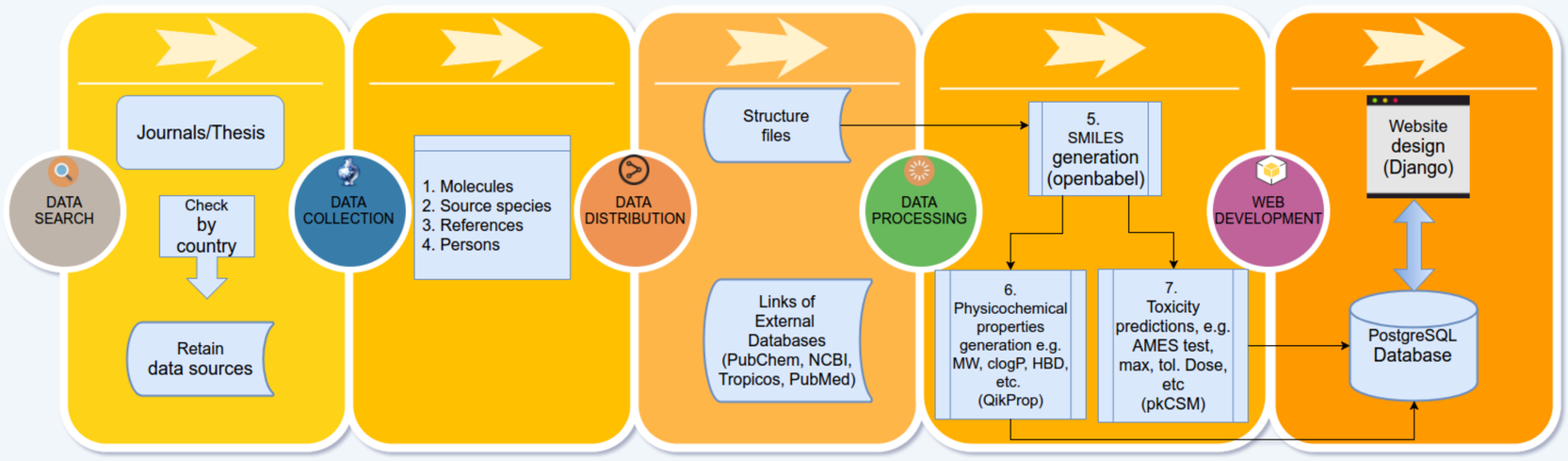

Figure 4: Workflow of NANPDB

NANPDB vs Other NP Databases

There are ~95% unique compounds in NANPDB (Figure 5) when compared to NPs of bacterial origin (StreptomeDB 2.0[9], BioPhytMol[10]), along with NPs that target the specific diseases tuberculosis and cancer (NPACT[11]). Comparing physicochemical properties involved in Lipinski's rule of five indicates that around ~75% of compounds had MW ≤ 500Da. Average MWs of NANPDB (420) are in the range of average MWs of DrugBank[12] (341), BioPhytMol (347), StreptomeDB 2.0 (514), and NPACT (442). On the other hand, mean cLogP values of both NANPDB and FDA-approved drugs were equal to ~2. Also, ~57% of NANPDB compounds showed no violations of any Lipinski rule, ~75% of NPs showed less than two violations. And ~87% of NANPDB compounds respecting Lipinski's criterion (cLogP ≤ 5), when compared with ~83% (DrugBank), ~89% (StreptomeDB 2.0), ~78%(NPACT), and ~69% (BioPhytMol). Indicating good drug-likeliness of the NANPDB compounds.

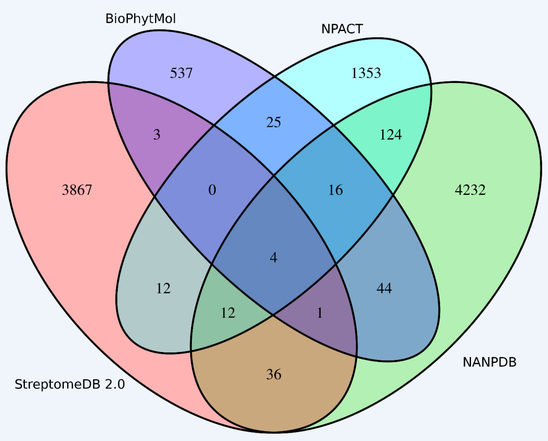

Figure 5: NANPDB and other Databases

Website features

- **Browsable lists**: Lists contain information in alphabetical browsable order. These lists exist for Compounds, Species, Families, References, and Authors.
- **Compound & Species Cards**: Pages displaying complete information of the selected compound or species.
- **Query searches**: Compound name or ID or a keyword can be given to retrieve all compounds of the given query.
- **Similarity & Structure Search**: A full molecule or substructure can be drawn or SMILES format can be directly given to identify the most similar compounds of the given structure or substructure.

www.african-compounds.org/nanpdb

[1] Ntie-Kang, F, et al.: *J. Nat. Products* **2017**, PMID: 28641017
[2] Harvey, A. L, et al.: *J. Nat. Rev. Drug Discovery* **2015**, 14, 111-129
[3] Rodrigues, T, et al.: *G. Nat. Chem.* **2016**, 8, 531-541
[4] UNO finding: http:// millenniumindicators.un.org/unsd/methods/m49/m49regin.htm
[5] Kim, S, et al.: *Nucleic Acids Res.* **2016**, 44, D1202-D121310.1093
[6] Pires, D. E. V, et al.: *J. Med. Chem.* **2015**, 58, 40667-4072
[7] Prota Africa: Plant resources of Tropic Africa (http://www.prota4u.org/)
[8] Tropicos.org. Missouri Botonical Garden (http://www.tropicos.org)
[9] Klementz, D, et al.: *Nucleic Acids Res.* **2016**, 44, D509-D51410.1093
[10] Sharma, A, et al.: *J. Cheminf.* **2014**, 6, 4610.1186/s13321-014-0046-2
[11] Mangal, M, et al.: *Nucleic Acids Res* **2013**, 41, D1124-D112910.1093
[12] DrugBank version 5.0

Pharmazeutische Bioinformatik
http://www.pharmazeutische-bioinformatik.de

Alexander von Humboldt Stiftung/Foundation

FRIAS FREIBURG INSTITUTE FOR ADVANCED STUDIES

Scan & Visit NANPDB

UNI FREIBURG