CHEMICAL DIVERSITY IN NATURAL PRODUCT LIBRARIES: A WAY TO EXPAND DISCOVERY PROGRAMS

José R. Tormo, Arnau Comajuncosa-Creus, Teresa Sardón, Germán Martínez, Fernando Reyes, Patrick Aloy and Olga Genilloud .*
Fundación MEDINA, Parque Tecnológico de Ciencias de la Salud, Avda. del Conocimiento 34, 18016 Granada, SPAIN
Institute for Research in Biomedicine (IRB), 08028 Barcelona, SPAIN
* ruben.tormo@medinaandalucia.es; www.medinadiscovery.com
Structural Bioinformatics & Network Biology Group

Introduction

Natural Product (NP) libraries have been successfully used as a source of bioactive molecules, offering a vast array of chemical structures for Drug Discovery in the past. Although nowadays its utilization has been diminished **in comparison to** the prominence of **synthetic libraries**, the chemical space covered by these two types of libraries may differ.

Overview

- **GOAL:** To compare the chemical diversity of compounds isolated from MEDINA´s NP Library, with those from different libraries representing both synthetic and natural small molecule collections.
- **METHODOLOGY:** Bioactivity signatures from the tool ChemicalChecker were used to encode the physicochemical and structural properties of small molecules. Descriptors were then plotted in two-dimensional representations (t-SNE), for a visual comparison.

MEDINA´s NP library vs Other Libraries

Microbial Strain Collections
190,000 strains
Fungi, Actinomycetes & unicellular Bacteria

Natural Product Libraries
140,000 extracts
from a large diversity of well-identified microbial strains

· Environment Microbiome · Geographical, Ecological and Taxonomical **DIVERSITY**

Crude extracts

High Microbial Diversity Module
60,539 Crudes (50% Fungi - 50% Actinomycetes)
Widely diverse selected strains (FAMES and Molecular Biology)

Activity Enriched Module
8,333 Crudes -Bioactive strains from 500,000 broad spectrum screening library
15,434 Crudes -Bioactive strain from 180,000 Gram negative screening library

Fungal Endophytes Module
12,480 Crudes -Endophytes strains from Spain (Morphology and Molecular Biology)
(62% Shrubs – 18.5% Grasses – 19.5% Xerophytic plants)

Wide range of Fungal Basidiomycota Library
6,880 Crudes 1.350 widely diverse strains from selected species

SPE extracts (solid phase extraction)

Marine Actinomycetes & Fungi Module
9,893 SPE Extracts -Widely diverse strains

Lichen associated Fungi Module
1,440 SPE Extracts (20%DMSO, 2-4xWBE)

Fractions

Module of Fractions from Actinomycetes & Fungi
10,040 Fractions containing an average of 5-7 metabolites each
926 widely diverse strains with proven production of Active Secondary Metabolites

Marine Actinomycetes & Fungi Module
4,320 Fractions -Widely diverse strains

Metabo Lights: Global database (DB) for metabolomics studies including raw experimental data and the associated metadata. This DB is cross-species and cross-technique and covers metabolite structures and their reference spectra as well as their biological roles and locations where available [1].

Collective Molecular Activities of Useful Plants (CMAUP): Collective molecular activities of 7,865 useful plants, including 2,954 medicinal plants used in 79 countries/regions, on 758 human target proteins and 3,013 Gene Ontology, 238 KEGG pathways, and their relations to 1,399 human diseases [2].

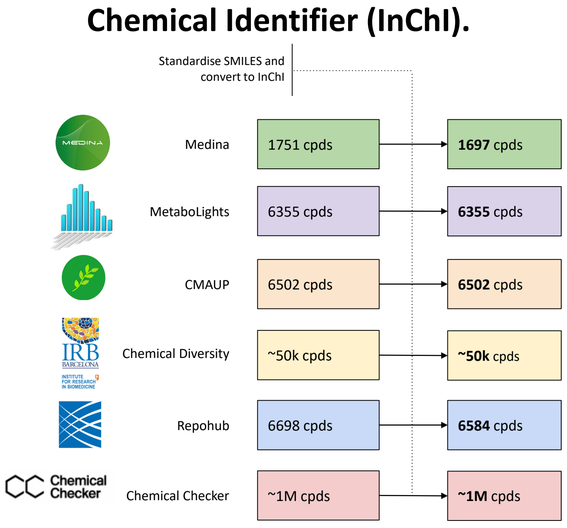
IRB: 50,000 compounds from the ChemDiv library, representative of more than 1,500,000 compounds. Including focused subsets such as Protein-Protein-Interaction (PPI), CNS, cancer, diverse set and lead-like set, among others [3].

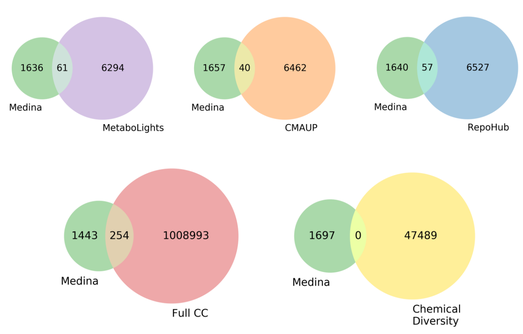
The Drug Repurposing Hub: Curated and annotated collection of FDA-approved drugs, clinical trial drugs, and pre-clinical tool compounds with a companion information resource [4].

Chemical Checker: Compound database with processed, harmonized and integrated bioactivity data on ~1,000,000 small molecules [5].

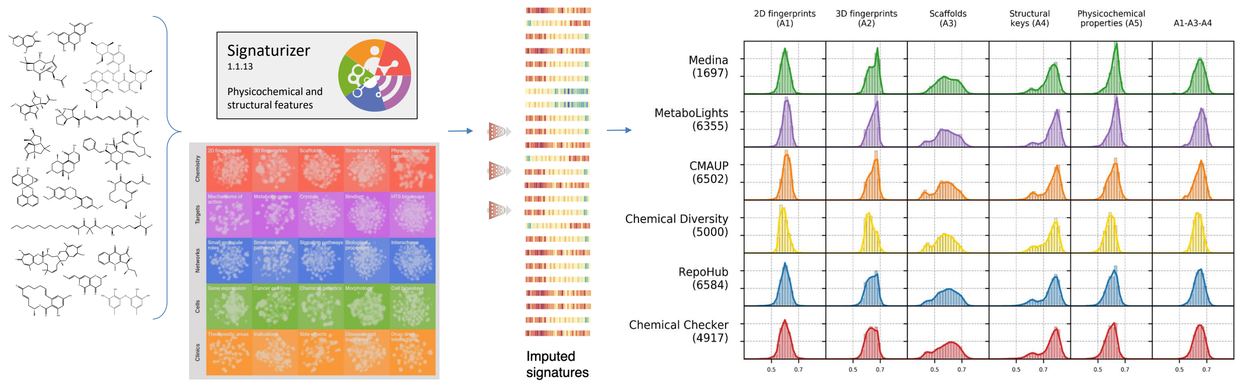
METHODS and RESULTS

1. Compounds have been standardized and converted to International Chemical Identifier (InChI).

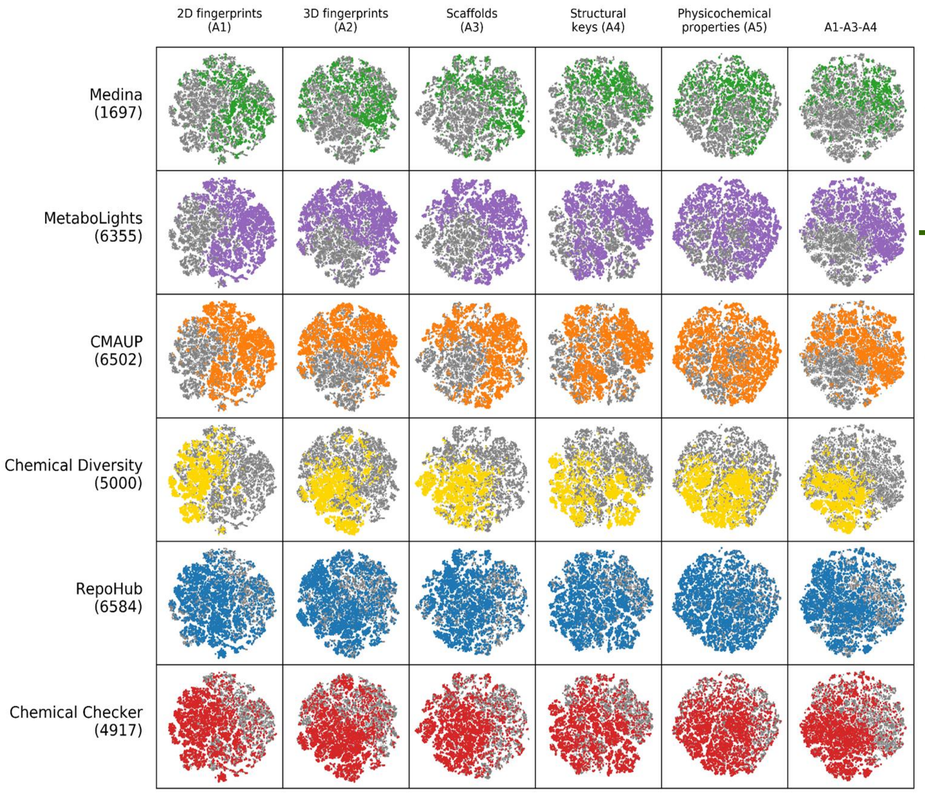




2. Different physicochemical and structural features of the studied compounds were encoded into descriptors using the Chemical Checker Signaturizer tool [3,5].



3. The generated physicochemical and structural descriptors were visually compared by means of t-SNE representations [3,5].



- Compounds identified in MEDINA´s Library **fully covers** the space attributed to other NP Libraries´ chemical diversity.
- The chemical structures of compounds identified in MEDINA´s library are **complementary** to structures present in a chemically diverse Synthetic Compound Library
- Some areas covered by **drugs** in the Repo Hub DB are also covered by compounds from MEDINA´s Library

SMILE information from 1,751 **compounds isolated from MEDINA´s library** and **identified by MS/NMR**, and SMILES from compounds included in the other studied libraries were first standardized and converted to InChI (Fig 1). A small negligible overlap between the MEDINA compounds and the other libraries was observed (Fig. 1).

To perform the structural comparisons, we used bioactivity signatures from the Chemical Checker [3,5], which encode the **physicochemical and structural properties** of small molecules. The applicability of this tool with the different compound libraries was confirmed, comparing the characteristics of the input compounds with the tool training dataset (Fig. 2).

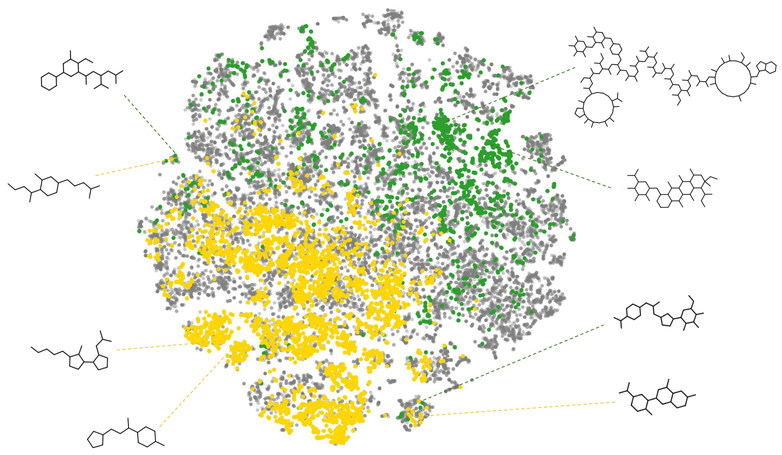
The obtained compound descriptors were then plotted in **two-dimensional representations** (t-SNE), for a visual comparison. Specifically, we used the A1-5 spaces of the Chemical Checker for a complete chemical description of the small molecules compared: **2D** fingerprints (A1), **3D** fingerprints (A2), **Scaffolds** dispersion (A3), structural **Key features (**A4) and physicochemical properties (A5) (Fig. 3).

REFERENCES

1. MetaboLights: open data repository for metabolomics. Yurekten O, Payne T, Tejera N, Amaladoss FX, Martin C, Williams M, O'Donovan C. Nucleic Acids Res, 2024. https://doi.org/10.1093/nar/gkad1045
2. Dongyue Hou, Hanbo Lin, Yuhan Feng, et al. CMAUP database update 2024: extended functional and association information of useful plants for biomedical research. Nucleic Acids Research 2024; PMID: 37897343; https://doi.org/10.1093/nar/gkad921
3. Duran-Frigola, M., Pauls, E., Guitart-Pla, O. *et al.* Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker. *Nat Biotechnol* **38**, 1087–1096 (2020). https://doi.org/10.1038/s41587-020-0502-7
4. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, Johnston SE, Vrcic A, Wong B, Khan M, Asiedu J, Narayan R, Mader CC, Subramanian A, Golub TR. The Drug Repurposing Hub: a next-generation drug library and information resource. Nature Medicine. 23, 405–408 (2017). https://www.nature.com/articles/nm.4306
5. Bertoni, M., Duran-Frigola, M., Badia-i-Mompel, P. et al. Bioactivity descriptors for uncharacterized chemical compounds. Nat Commun 12, 3932 (2021). https://doi.org/10.1038/s41467-021-24150-4

CONCLUSIONS

- **MEDINA´s Library** virtually **covers** all the chemical diversity of known **Natural Products**.
- NP and **Synthetic Compound Libraries occupy different diversity** regions of the chemical space, being complementary.
- **The combination** of Natural Products and Synthetic compounds expands the chemical diversity of the resulting library and **increases the search space in any Drug Discovery Program**.



Junta de Andalucía
UNIVERSIDAD DE GRANADA
MSD