Automated Analysis of Structure-Multiple Property Relationships: Impact on SMARts

Jesús Israel Espinoza Castañeda - israelch317@gmail.com
Dr. José L. Medina-Franco - medinajl@unam.mx

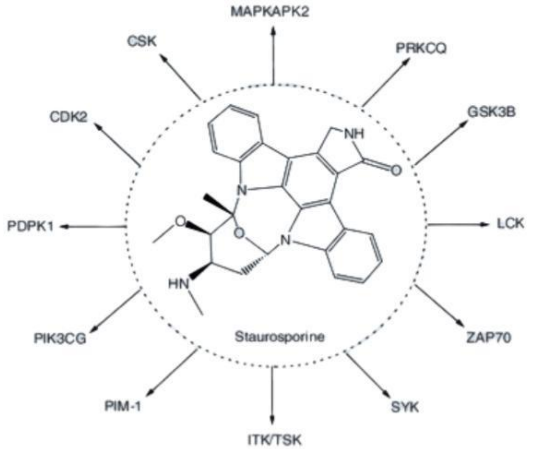
XXX Symposium on Bioinformatics and Computer-Aided Drug Discovery

DIFACQUIM Research Group
School of Chemistry, Universidad Nacional Autónoma de México
September 2024

Introduction

Structure-Activity Relationships (**SAR**) is a fundamental concept in medicinal chemistry and drug discovery. **SAR information** of compounds data sets is crucial to extract information about chemical modifications and their impact on physicochemical properties or activity

In order to increased awareness of polypharmacology a standard practice in drug discovery is explore **Structure-Multiple Activity Relationships (SMARts)**

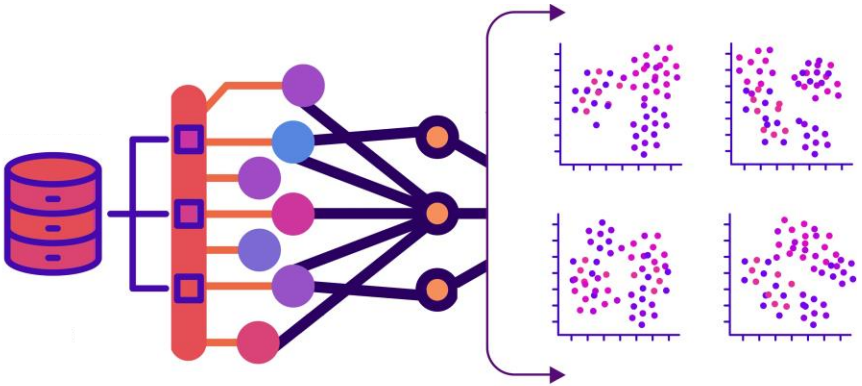


Reddy S. and Zhang S. (2014). Expert Rev. Clin. Pharmacol. 6(1), 41-47

Edgar López-López and José L. Medina-Franco. (2024). *Drug Discovery Today*, 29(7)

Objective

Automate the generation of a chemical multiverse by implementing various types of descriptors to provide a global view of SMARts.



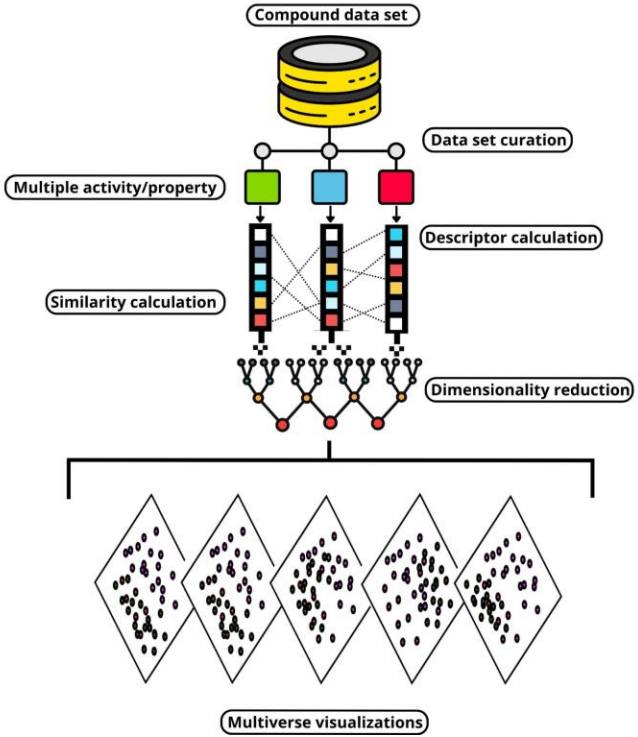
Methods

MAYA (**M**ultiple **A**ctivit**y** **A**nalyzer) is developed as a user-friendly, open-source tool that automates the construction of chemical spaces by integrating 3 representations of a set of molecules described by their SMILES notation and an associated activity/property.

Workflow

Curation protocol

1. Screening of allowed elements
2. Normalization rules
3. Selection of largets fragments
4. Charge corrections
5. Ionization of strong acids
6. Selection of the most chemically reasonable tautomers
7. Elimination of repeated SMILES



Calculation descriptors

- Molecular Acces System (MACCS keys) - 166-bits
- Extended Conectivity Fingerprints (ECFP 4/6) - 1024-bits
- Drug-likeness descriptors (MW, Log(P), TPSA, HBD, HBA and RB)

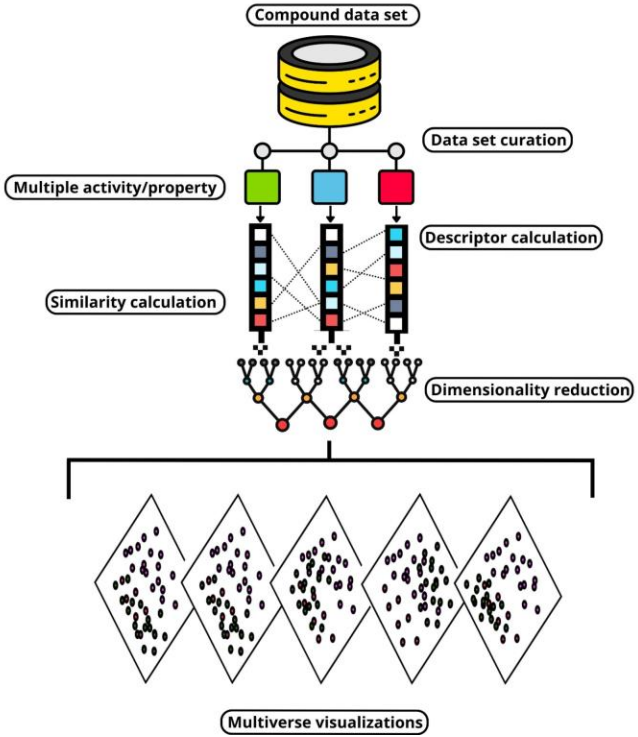
RDKit

Workflow

Similarity calculations

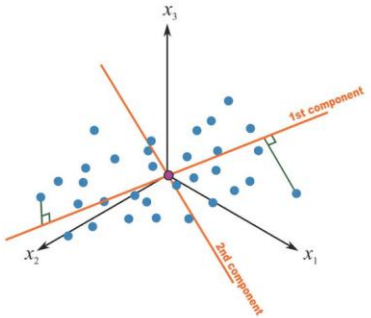
Pairwise similarity calculations are performed with the Tanimoto coefficient

$$T(a,b) = \frac{N_c}{N_a + N_b - N_c}$$



Visulization of chemical space

Principal Component Analysis (**PCA**) and t-Distributed Stochastic Neighbor Embedding (t-SNE) are implemented to represent the variables in a 2D dimensional space



Script

MAYA is an automated application for chemical multiverse visualizations since it enables the fast generation of chemical spaces of the same data set

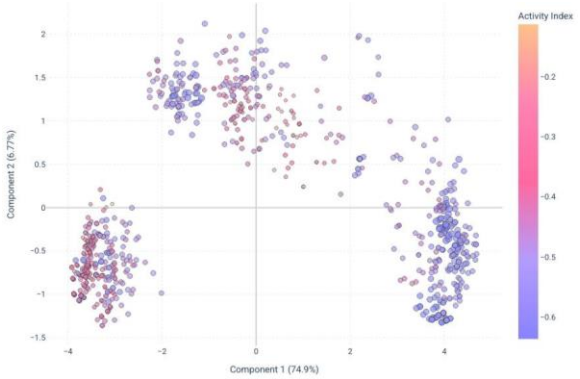
Visualizations

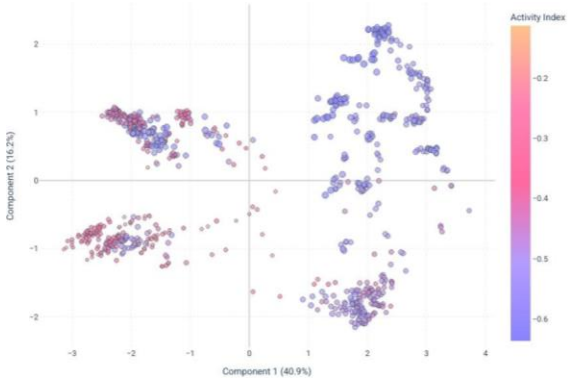
MACCS Keys

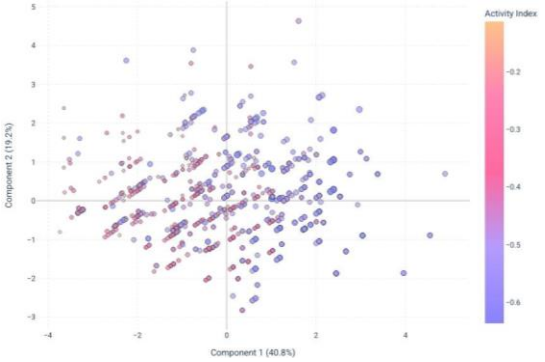
ECFP-4

Drug-likeness descriptors

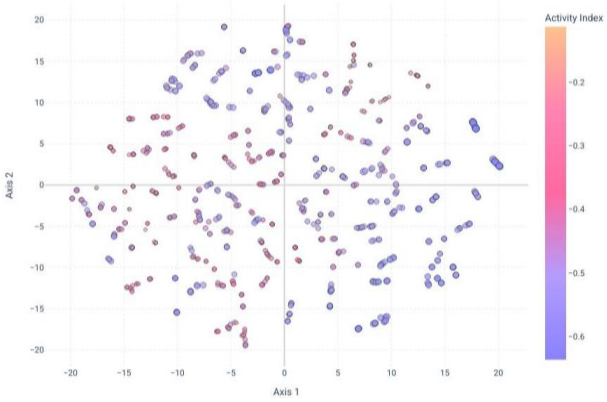
PCA

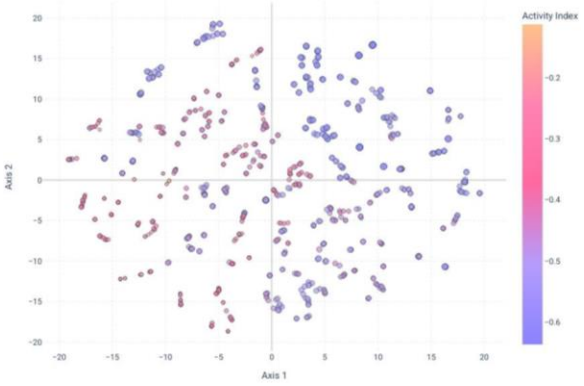


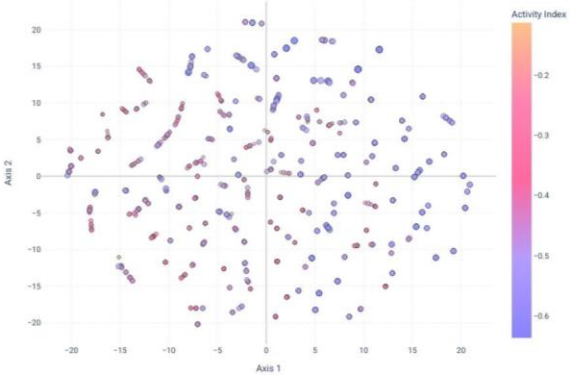




T-SNE







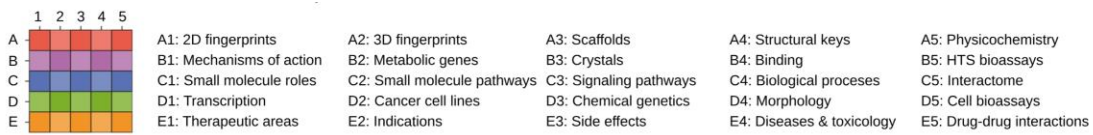
Conclusions

- **MAYA** is an open-access tool to analyze the chemical space and chemical multiverse of compound data sets annotated with multiple biological activities. Generating a so called ‘Chemical Megaverse’ defined as a group of chemical spaces of the set of chemical compounds

- This approach not only facilitates data analysis and representation but also provides a **flexible platform** and transparency in research. The developed code has the potential to **accelerate** drug development and improve precision in drug repositioning.

Perspectives

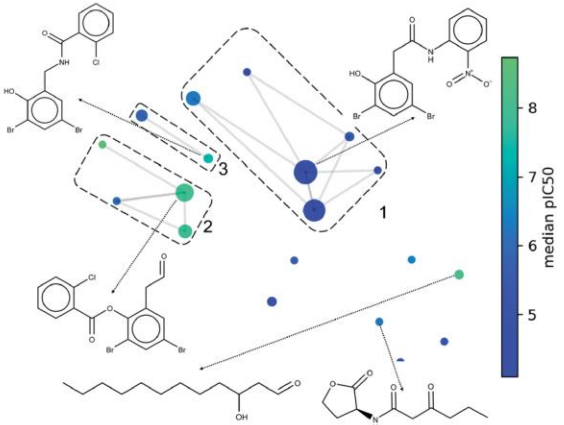
- Incorporation of others types of descriptors and **implementing similarity indices** other than the Tanimoto coefficient.



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

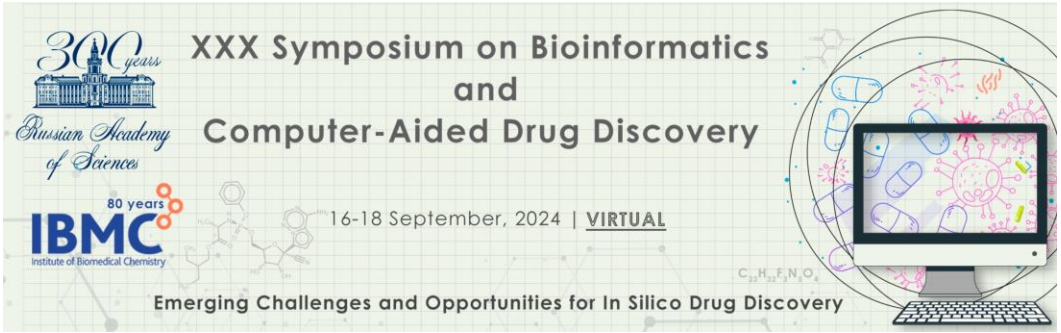
- Similarly, other **data visualization techniques** could be incorporated.

Constellation plots



RSC Adv., 2022,**12**, 6783-6790

Multiple Activity Analyzer



Thank you

CONAHCYT
CONSEJO NACIONAL DE HUMANIDADES CIENCIAS Y TECNOLOGÍAS

Scholarship No. CVU 1340927

PAPIIT Grant No. IG200124

DIFACQUIM
COMPUTER-AIDED DRUG DESIGN AT UNAM