Chemical multiverse and diversity of food chemicals

Juan F. Avellaneda-Tamayo, Ana L. Chávez-Hernández,
Diana L. Prado-Romero, José L. Medina-Franco*

DIFACQUIM Research Group
School of Chemistry, Universidad Nacional Autónoma de México
September 19 2023

XXIX Symposium on Bioinformatics and Computer-Aided Drug Discovery

Introduction

- Canadian Institute of Health Research (CIHR), the Canada Foundation for Innovation, and the Metabolomics Innovation Centre (TMIC).
- 70,926 compounds from 797 foods*.
- FooDB contains more than 100 separate data fields.
- PCP, biological, organoleptic.

CIHR IRSC
Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

INNOVATION
Canada Foundation for Innovation | Fondation canadienne pour l'innovation

TMIC
The Metabolomics Innovation Centre

FooDB. Available at https://foodb.ca/ (accessed on September 18th, 2023)

Introduction

- 28,883 FOODB compounds, comparing vs. GRAS, approved drugs, and NPs from ZINC.
- ChemMaps chemical similarity
- PCP related to oral biodisponibility
- Molecular complexity by sp^3 hybridized carbon atoms → highest complexity.
- Scaffold content → abundance of acyclic compounds. Scaffolds comparable to NPs and drug type. High content of polyphenols.
- Global diversity with Consensus Diversity Plot.

Naveja JJ, Rico-Hidalgo MP, Medina-Franco JL. F1000Res. 7 (2018), 993.

Objectives

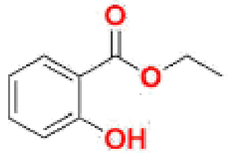
- Chemical content, diversity, chemical space, and potential as bioactive compounds from food chemicals.

Methods

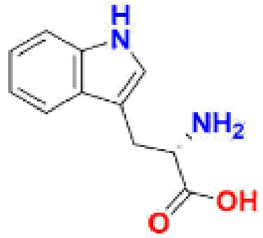
1. Data curation (FooDB, FDA-approved drugs, UNPD-A, commercially available FooDB on ZINC20).
2. Overlap of compounds and scaffolds.
3. Molecular descriptors and FPs.
4. Scaffold content and diversity.
5. Natural product likeness score.
6. Structural diversity Tanimoto's coefficient.
7. Chemical multiverse visualization.
8. Chemical profiling by NPClassifier.

A.L. Chávez-Hernández and J.L. Medina-Franco, Artificial Intelligence in the Life Sciences 3 (2023), pp. 100066.

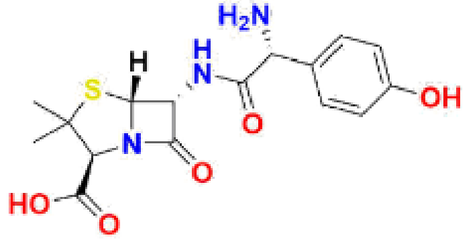
Main results: overlapping compounds



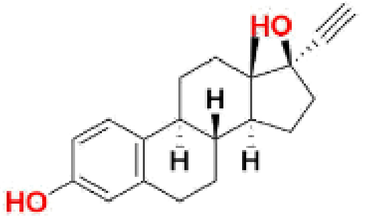
Ethyl salicylate



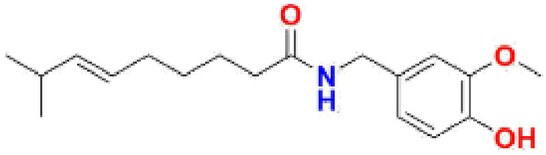
Tryptophan



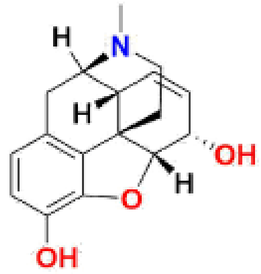
Amoxicillin



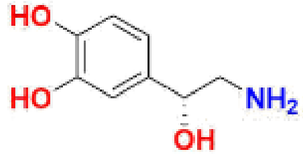
Ethinylestradiol



Capsaicin



Morphine



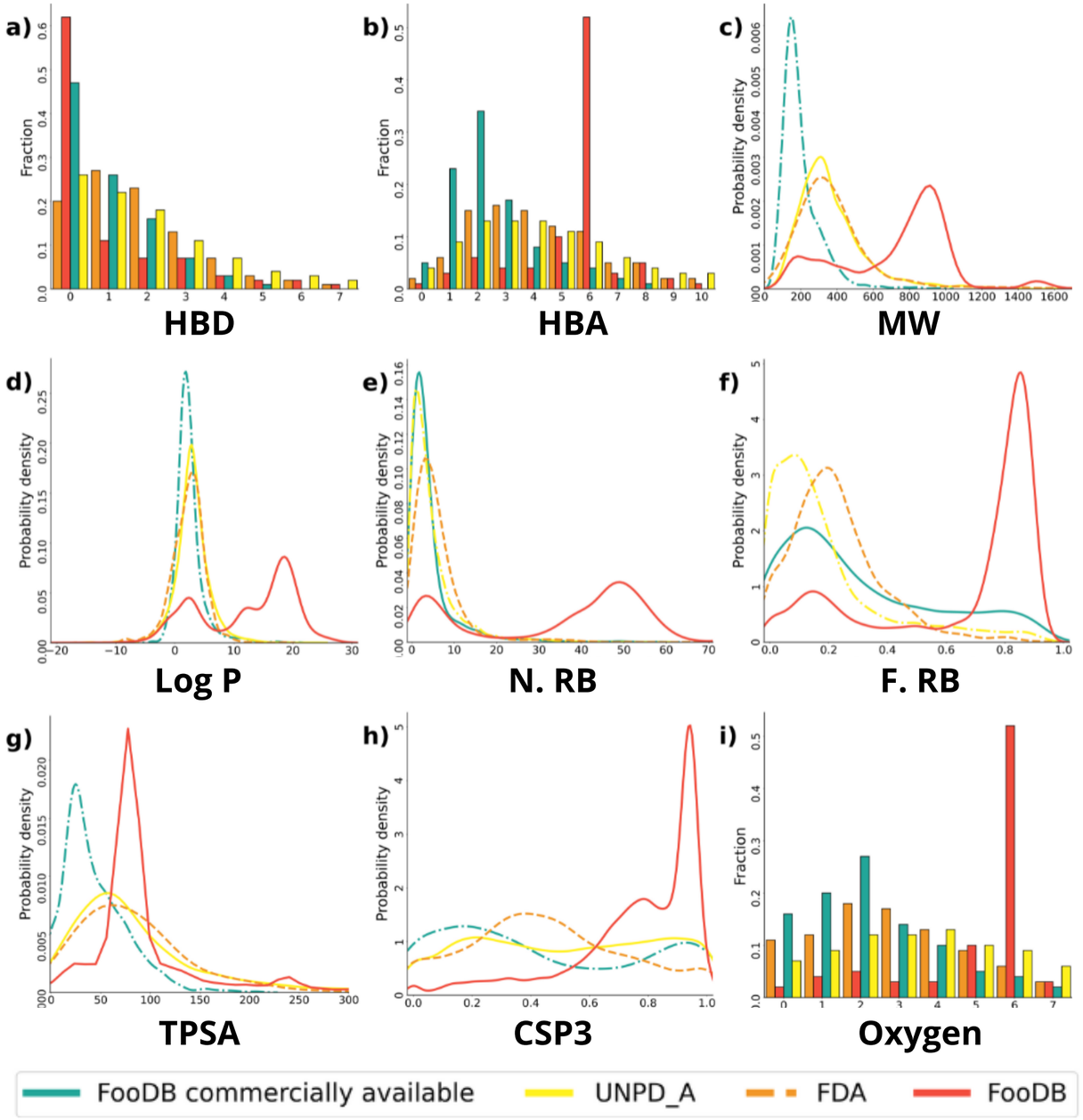
Norepinephrine

Cycloserine

- Sixty compounds were common to the three databases.
- Nutrients used in the clinic for nutrient deficiencies.
- Specific drugs are also present.

P. Anand and K. Bley, Br. J. Anaesth. 107 (2011), pp. 490–502.

Main results: molecular descriptors



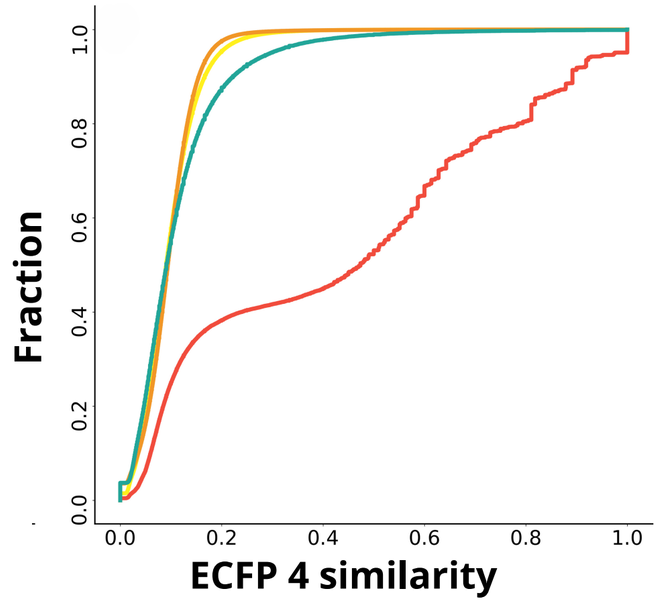
Tendency for **larger and less polar molecules** in food chemicals.

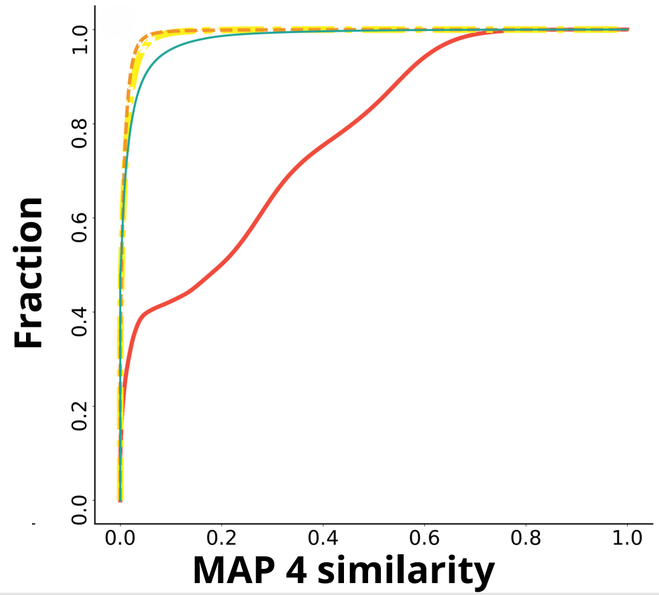
Tendency for **alquilic molecules**, high fraction of sp^3 and rotational bonds, less content of aromatic and alicyclic rings (relation to scaffold content).

Changes in distribution vs. previous version of FooDB.

F.I. Saldívar-González, M. Valli, A.D. Andricopulo, V. da Silva Bolzani and J.L. Medina-Franco, J. Chem. Inf. Model. 59 (2019), pp. 74–85.

Main results: distribution of pairwise Tanimoto similarity





FooDB commercially available UNPD_A FDA FooDB

Data set	Mean (median) similarity			
	MACCS 167 bits	ECFP4 1024 bits	ECFP6 1024 bits	MAP4 2048 bits
FooDB	**0.65 (0.62)**	**0.44 (0.47)**	**0.40 (0.42)**	**0.23 (0.20)**
UNPD-A	0.35 (0.34)	0.10 (0.09)	0.08 (0.08)	0.01 (0.00)
FDA	0.30 (0.30)	0.10 (0.09)	0.08 (0.08)	0.01 (0.00)
FooDB commercial	0.26 (0.22)	0.12 (0.09)	0.09 (0.08)	0.02 (0.00)

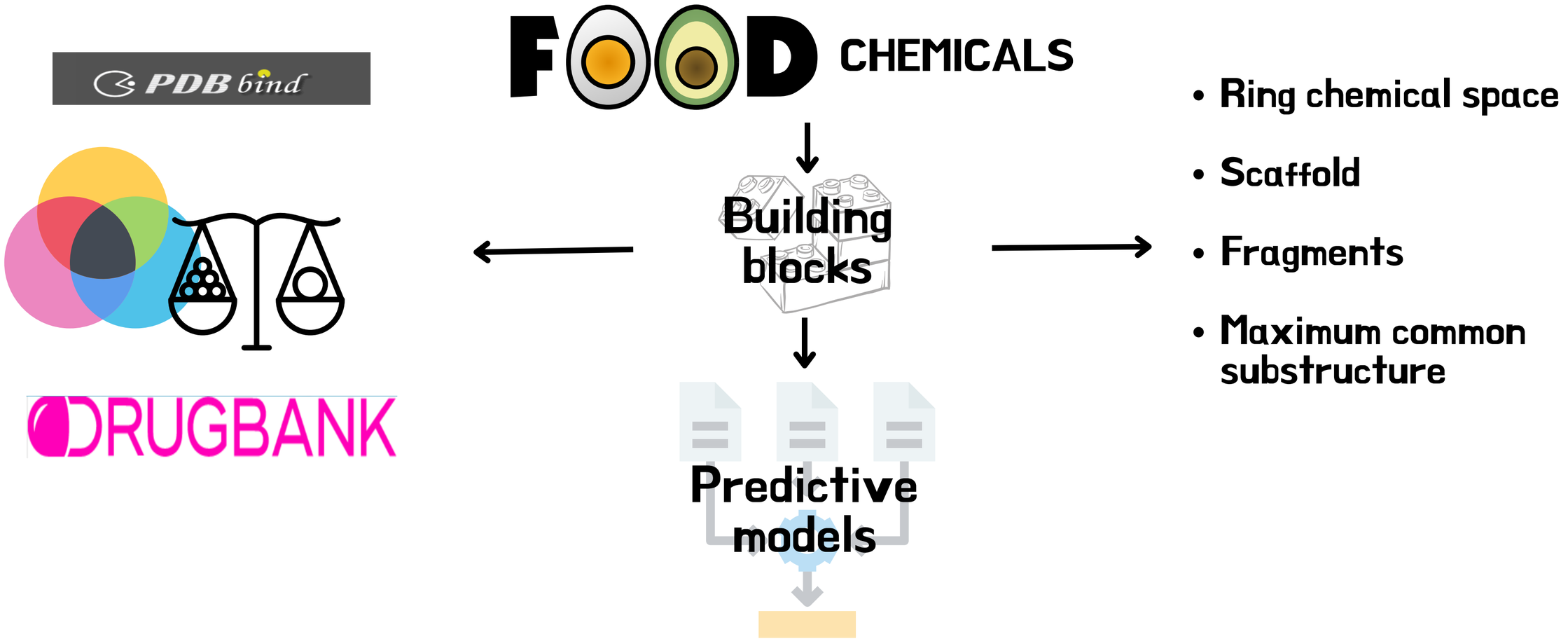
N. Singh, R. Guha, M.A. Giulianotti, C. Pinilla, R.A. Houghten and J.L. Medina-Franco, J. Chem. Inf. Model. 49 (2009), pp. 1010–1024.

Conclusive ideas

- Large **overlap** among food chemicals, natural products, and approved drugs.
- **Broad** and **specific** distribution of properties of food chemicals (drug type, constitutionals, complexity).
- Lower scaffold and fingerprint-based **diversity** regardless of the fingerprint.
- Wider coverage of chemical **multiverse**.
- The most represented biosynthetic pathway predicted was **fatty acids**.
- Further studies are adequate for the development of **machine learning models** and **bioactive** compounds based on food chemicals.

Perspectives

- J. Avellaneda, A. Chávez, D. Prado, J. Medina, Chemical multiverse and diversity of food chemicals, **SAR and QSAR in Environmental Research,** (2023) under review.



Acknowledgments

CONAHCYT
CONSEJO NACIONAL DE HUMANIDADES CIENCIAS Y TECNOLOGÍAS

Scholarship CVU 1270553

dgapa

PAPIIT Grant No. IV200121

DIFACQUIM
UNAM

medinajl@unam.mx

@difacquim

@difacquim_unam

@avellanedajuan