Department of Pharmacy, National Autonomous
University of Mexico, Mexico City, Mexico

CHEMOINFORMATIC ANALYSIS OF NATURAL PRODUCTS FROM MEXICO

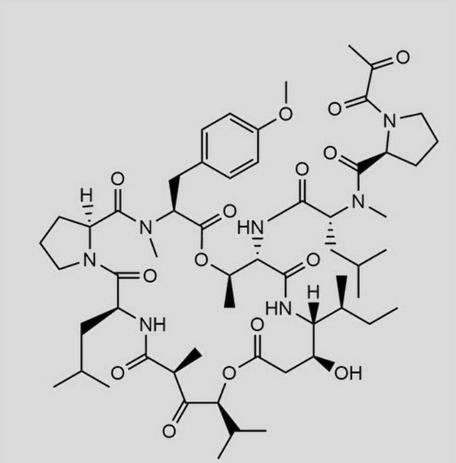
Carlos Daniel Ramírez Márquez
DIFACQUIM research group

Introduction

- store
- process
- organize
- disseminate

small molecules and their properties

Chemical databases



Natural products (NPs)

From 1981 to 2019
1881 approved drugs
3.8% NPs
18.1% NPs Derivatives

Natural products database developed in Mexico

- 500 compounds

Objective: Update and characterize the Mexican natural products database BIOFACQUIM[1] using cheminformatics tools.

(1) Pilón-Jiménez, B. A.; Saldívar-González, F. I.; Díaz-Eufracio, B. I.; Medina-Franco, J. L. BIOFACQUIM: A Mexican Compound Database of Natural Products. *Biomolecules* **2019**, *9* (1), 31. https://doi.org/10.3390/biom9010031.

Methods

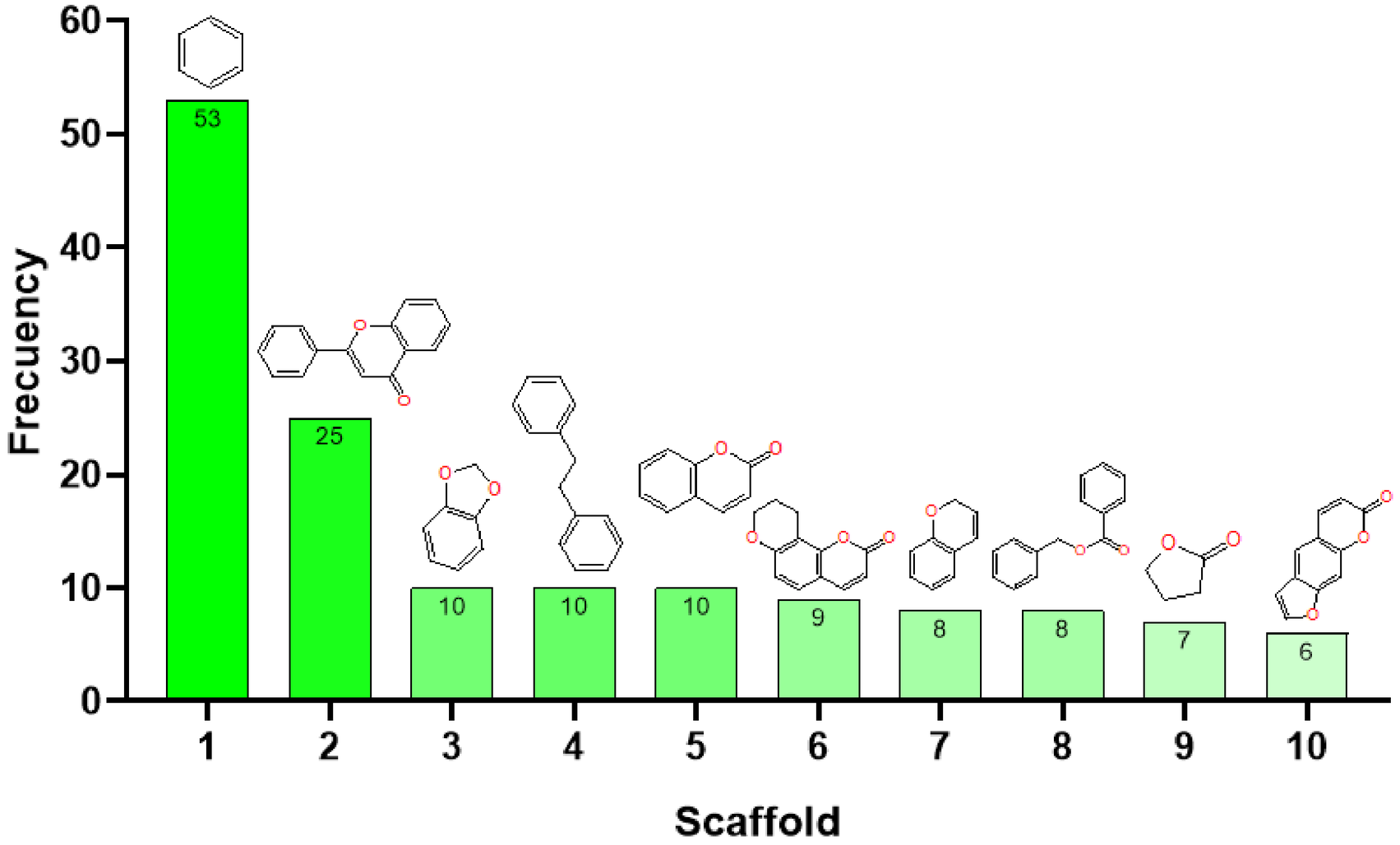
Update of BIOFACQUIM 01	Bibliographic research. Information: Structure, SMILES, bibliographic source , information about the natural source (kingdom, genus and species), and location. Curation.
Scaffold content 02	To identify the most frequent scaffolds in BIOFACQUIM using Bemis and Murcko definition[2]. Software Osirirs Data warrior was used.
Characterization with physicochemical properties 03	Calculation of six physicochemical properties: SLogP, MW, HBA, HBD, RB, TPSA. Statistical comparison with 8 reference databases.
Chemical space[3] visualization 04	Principal component analysis to generate a physicochemical properties based visual representation of the chemical space.

(2) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893. https://doi.org/10.1021/jm9602928
(3) Medina-Franco, J. L.; Chávez-Hernández, A. L.; López-López, E.; Saldívar-González, F. I. Chemical Multiverse: An Expanded View of Chemical Space. **2022**. https://doi.org/10.26434/chemrxiv-2022-3h0ls.

Results and analysis - Scaffold content

BIOFACQUIM was updated to **605 different compounds**

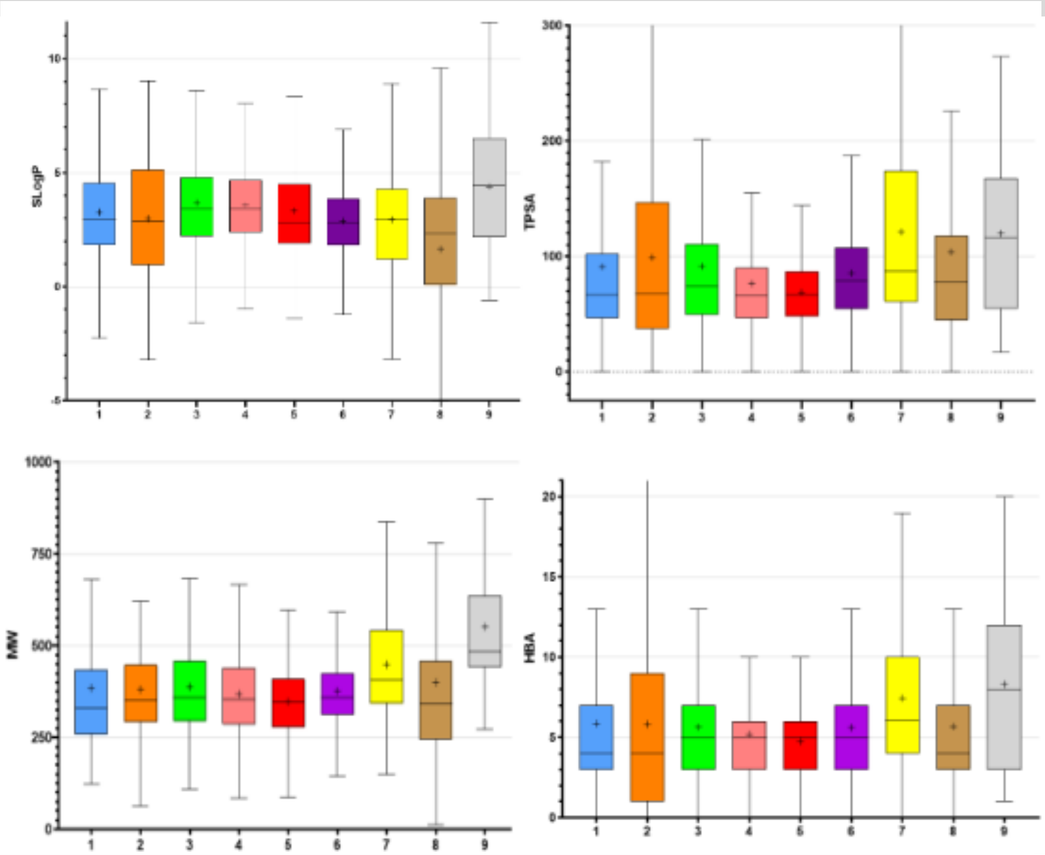
311 different Scaffolds



Most frecuent Scaffolds in BIOFACQUIM

- Benzene
- Flavone
- 1,2-benzodioxole
- Dibencyl
- Coumarine
- 9,10-dihydro-2H,8H-pyrano[2,3-f]chromen-2-one
- 2H-chormene
- Benzyl benzoate
- γ-Butyrolactone
- Psoralene

Results and analysis - Characterization with physicochemical properties



BIOFACQUIM:

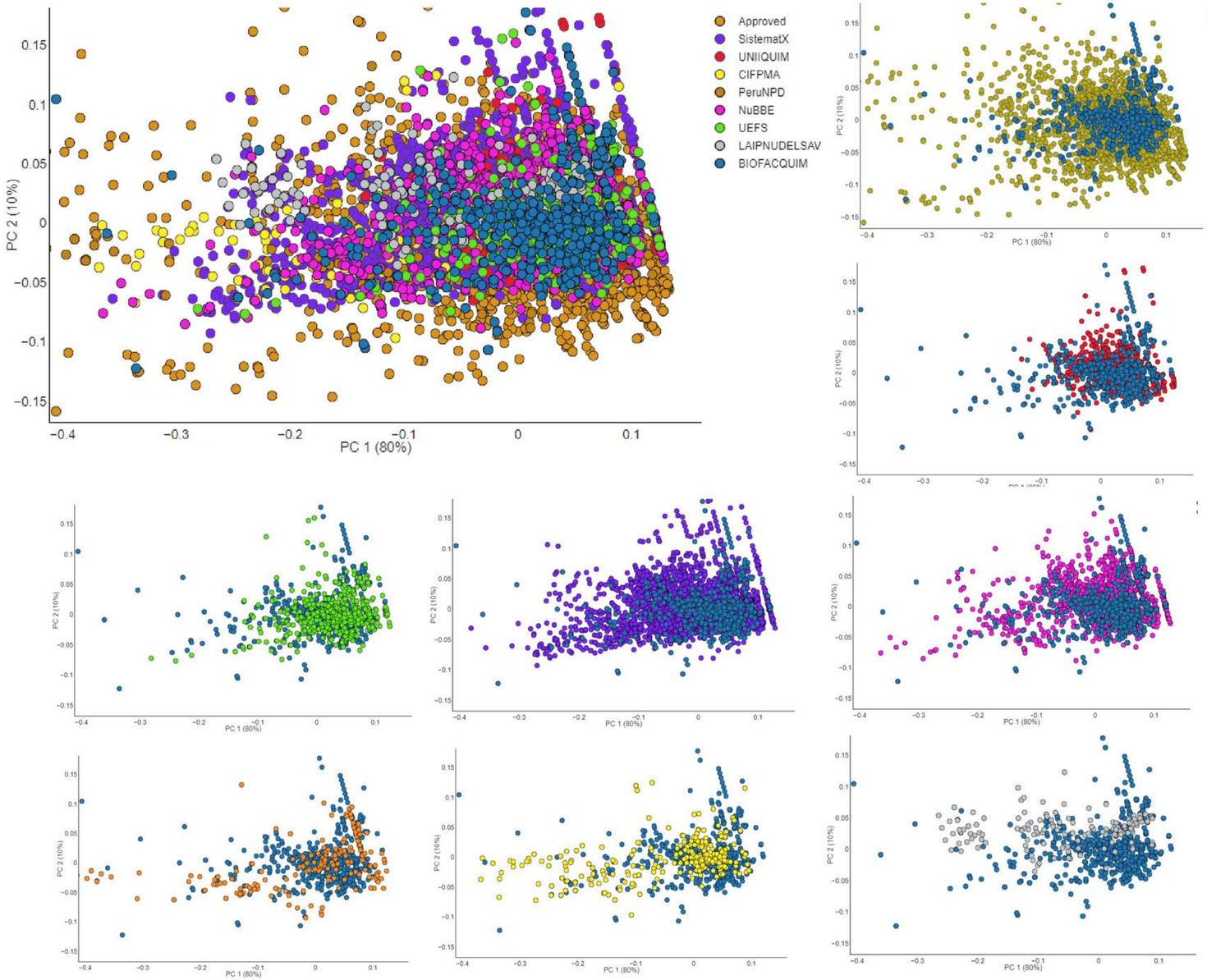
- It includes more lipophilic compounds than Approved drugs.
- It is the second database with the largest distribution of MW, TPSA, and RB.
- It also has the largest HBA distribution.

SLogP	(1) BIOFQ	(2) PeruNPDB	(3) UEFS	(4) NuBBE	(5) UNIIQUIM	(6) SistematX	(7) CIFPMA	(8) Approved	(9) LAPIDELS ALV
Min	-8.57	-3.22	-2.61	-5.61	-3.22	-7.07	-4.76	-35.16	-0.59
Median	2.92	2.90	3.41	3.43	2.81	2.80	2.91	2.34	4.43
Max	17.02	9.03	14.85	16.85	20.88	13.79	10.61	57.75	11.59
Mean	3.27	2.98	3.69	3.57	3.33	2.88	2.94	1.64	4.39
Std dev(σ)	2.82	2.98	2.35	2.27	2.17	1.84	2.52	4.14	2.59

MW	(1) BIOFQ	(2) PeruNPDB	(3) UEFS	(4) NuBBE	(5) UNIIQUIM	(6) SistematX	(7) CIFPMA	(8) Approved	(9) LAPIDELS ALV
Min	122.00	62.02	109.10	84.09	86.07	84.09	150.10	12.00	272.10
Mediana	330.20	350.60	359.10	354.20	346.20	358.10	406.20	342.00	484.30
Max	2794.00	964.30	1045.00	1170.00	874.80	1078.00	1045.00	4111.00	945.30
Media	385.10	381.10	388.40	368.60	347.60	376.50	448.70	399.90	551.80
DE (σ)	267.10	176.70	146.20	132.30	102.90	110.50	166.30	307.30	157.30

(1) BIOFACQUIM, (2) PeruNPDB, (3) UEFS, (4) NuBBE, (5) UNIIQUIM, (6) SistematX, (7) CIFPMA, (8) Approved drugs, (9) LAIPDELSALV

Results and analysis - Chemical space



- The first two principal components capture 90% of the total variance.
- Descriptors with the largest contribution:

  PC 1: HBA and TPSA

  PC 2: SLogP and MW

- The chemical space of BIOFACQUIM overlaps with the chemical space of Approved drugs and the NPs reference databases.

Principal component	PC 1	PC 2
Eigenvalue	0.008	0.001
Eigenvalue (%)	80%	90%
SLogP	0.125	0.62
TPSA	-0.444	-0.131
MW	-0.392	0.514
RB	-0.198	0.492
HBD	-0.391	-0.295
HBA	-0.665	-0.072

Conclusions

BIOFACQUIM

- Contains 605 different NPs isolated and charaterized in Mexico.
- Scaffold analysis identified 311 different scaffolds. Benzene, flavone, benzodyoxole, dibenzyl and coumarin are the five most frequent.
- From physicochemical properties analysis was determined that BIOFACQUIM includes more lipophilic compounds than Approved drugs, it is the second database with the largest distribution of MW, TPSA, and RB. And it also has the largest HBA distribution.
- The chemical space of BIOFACQUIM overlaps with the chemical space comprised by Approved drugs and the NPs reference databases.