PUMA v2: An Open-Access Platform for Molecular Diversity Analysis and Chemical Space Exploration

Daniel A. Acuña-Jiménez[1], Diego Alvarez Saravia[2], David Medina-Ortiz[2], José L. Medina-Franco[1]

[1]*DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, México.*

[2]*Departamento de Ingeniería en Computación, Universidad de Magallanes, Avenida Bulnes 01855, Punta Arenas, Chile.*

Molecular diversity analysis and exploration of the chemical space are central to chemoinformatics, enabling informed decision-making in the study of potentially bioactive molecules with novel properties [1]. Although traditionally associated with drug discovery projects, these approaches have become increasingly relevant in other areas, including natural products [2], food [3], and agricultural chemistry [4], among others. In many contexts rapid, accurate and reliable assessment of the chemical diversity of compound libraries is highly desirable. Such analyses benefit from a varied assortment of molecular representations, similarity metrics and dimensionality reduction techniques to capture complementary aspects of chemical diversity. To this end, accessible tools that integrate these approaches, particularly for users without programming or computational expertise, are highly valued. In this presentation, we describe a significant update to the Platform for Unified Molecular Analysis (PUMA) [5], an intuitive and freely available web-based tool for chemoinformatics-driven diversity analysis and chemical space visualization of user-supplied datasets. The updated version incorporates an automated data curation workflow, expanded descriptor sets to better assess drug-likeness and synthetic feasibility, additional molecular fingerprint and scaffold-based representations, and further visualization methods for analysis of fingerprint-based diversity. Notably, we introduce a novel representation of scaffold-based diversity based on Jensen-Shannon Divergence. The presentation will highlight the utility of the updated version of PUMA as a specialized tool aimed at students and researchers with little or no programming experience, that requires minimal user setup As part of the presentation, a case study analyzing the contents and diversity of compound sets related to diverse domains of chemical interest, such as drugs, food chemicals, pesticides, cosmetics, and natural products, will be shown. The updated version of PUMA (v2) will be freely available through its interactive web interface, as well as via an API and a Python library for local use.

Bibliography:

[1] Willett P. Chemoinformatics: a history. WIREs Comput Mol Sci 2011;1:46–56. https://doi.org/10.1002/wcms.1.

[2] Chen Y, Garcia de Lomana M, Friedrich N-O, Kirchmair J. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products. J Chem Inf Model 2018;58:1518–32. https://doi.org/10.1021/acs.jcim.8b00302.

[3] Avellaneda-Tamayo JF, Chávez-Hernández AL, Prado-Romero DL, Medina-Franco JL. Chemical Multiverse and Diversity of Food Chemicals. J Chem Inf Model 2024;64:1229–44. https://doi.org/10.1021/acs.jcim.3c01617.

[4] Foil DH, König J, Herrmann K, Jacob RA, Kneuer C, Kirchmair J. Comparative Chemical Space Analysis of Pesticides and Substances with Genotoxicity Data. Chem Res Toxicol 2025;38:1871–88. https://doi.org/10.1021/acs.chemrestox.5c00198.

[5] González-Medina M, Medina-Franco JL. Platform for Unified Molecular Analysis: PUMA. J Chem Inf Model 2017;57:1735–40. https://doi.org/10.1021/acs.jcim.7b00253.