Bioactivity prediction of coffee seed-derived natural products using a structure-based approach

Emanuele Carosati[1], Nicola de Zordi [2], Diana Lorena Prado Romero[1]

[1]*Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via Licio Giorgieri 1, 34127, Trieste (TS), Italy.*

[2]*SAM srl, Società Agricola Moldoi srl, Via Moldoi 53, 32037, Sospirolo (BL), Italy.*

Since the availability of data concerning natural products is continuously increasing, specialized databases are being developed. These include region-centered databases, like SANCDB from South Africa or LANaPDB from Latin America, that collect information from compounds isolated and characterized in certain geographical areas. Large-scale databases, like COCONUT are also provided, with thousands of natural-origin molecules. Nevertheless, bioactivity prediction methods for natural products have not yet reached the same level of development as those for small molecules, despite the accessibility of data.

Structure-based approaches provide an alternative to current AI-driven chemoinformatic methods, as they rely on the 3D structure of targets rather than known ligands. While traditionally dependent on experimentally resolved protein structures, recent advances like AlphaFold have expanded the applicability of such methods. This data accessibility could also enhance bioactivity prediction against more than one target at the same time. Although time could represent a disadvantage, multi-target screening tools like BioGPS [1, 2], allow the rapid estimation of molecular affinity. Thus, the calculation is made across a broad set of protein pockets using GRID-derived Molecular Interaction Fields (MIFs).

Given that BioGPS can be extended to analyzing more than one molecule at a time, and information on plant-derived compounds is attainable, a case study with molecules found in coffee is presented. The BioGPS method was applied on a list of chemicals found in coffee seeds across thousands of predicted pockets from human proteins. As a result, a score describing the complementarity for each molecule-pocket pair was calculated. Data was normalized with a recently developed approach (ZZscore) [3, 4], and target predictions yielded lists that underwent pathway enrichment analysis, using bioinformatic tools commonly used in omics research.

By leveraging OpenTargets data [5], the potential disease associations of molecular components from coffee were explored. With this methodology, the aim was to identify potential diseases that could be modulated by the phytocomplex. The results, supported by literature evidence, suggest new avenues for exploring coffee bioactivity. Available bioactivity data from ChEMBL [6] is also being analyzed to compare the target molecule associations. The design of a database with BioGPS scores and the OpenTargets database is also being considered, this could facilitate plant-disease associations to give new uses for known medicinal plants or combinations of them.

Bibliography :

[1] S. L ; C. S ; B. M ; G. L ; C. G. Proteins 83 (2015) 517-532.
[2] https://www.moldiscovery.com/software/biogps
[3] R. O. M ; G. N ; C. V ; B. V ; C. E ; T. E. J. Neurochem. 169 (2025) e70093.
[4] C. E ; M. L. B ; S. A ; C. G ; M. M ; M. G ; M. C ; G. M. S ; N. F. K ; V. S ; S. G ; C. M ; F. M ; C. I ; B. R. ACS Omega. 11 (2026) 9757-9773.
[5] https://www.opentargets.org/
[6] https://www.ebi.ac.uk/chembl/