



A BIOSYNFORMATIC MOLECULAR DESCRIPTOR FOR NATURAL PRODUCT RESEARCH

Lucina-May Nollen^{1,2}, David Meijer¹, Maria Sorokina³, Justin J.J. van der Hoof^{1,4}

¹ Bioinformatics Group, Wageningen University & Research, Netherlands

² Current address: Leiden Academic Centre of Drug Research, Leiden University, Netherlands

³ DS&AI, Bayer Pharmaceuticals, Germany

⁴ Department of Biochemistry, University of Johannesburg, South Africa

Natural products drive drug discovery, yet standard fingerprints overlook their **biosynthetic context**.

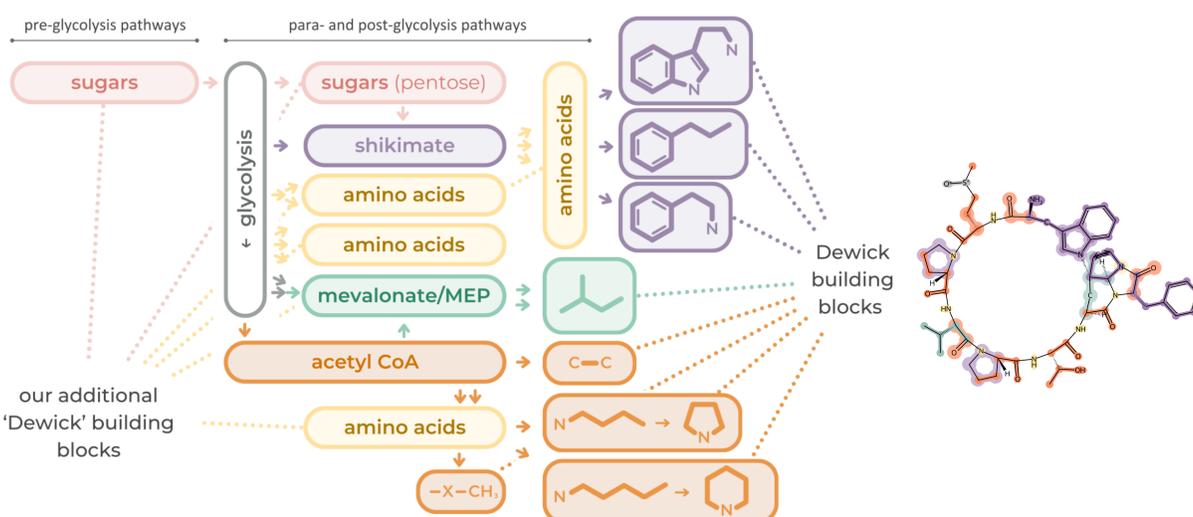
Biosynfoni fills this gap as a **compact, interpretable fingerprint** designed for natural product chemistry.

FINGERPRINT DESIGN

39 SMARTS keys

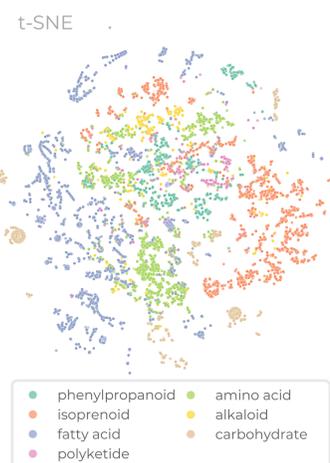
- **core biosynthetic units** (e.g., acetate, terpenoid, shikimate)
- **common decorations** (e.g., hydroxyls, halogens, rings)
- 97% natural product heavy atom coverage

count-based encoding preserves quantitative substructure information

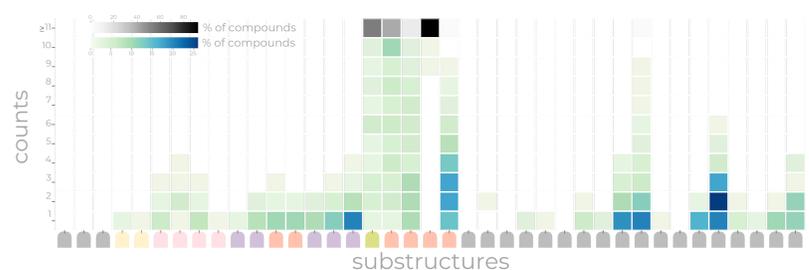
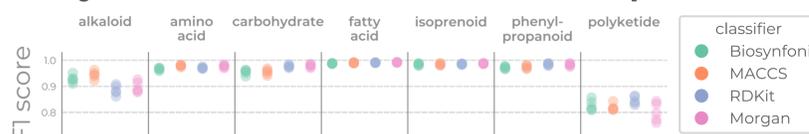


VISUALISATION, CLASSIFICATION & INTERPRETATION

chemical space visualisations show tight natural-product **class clustering**

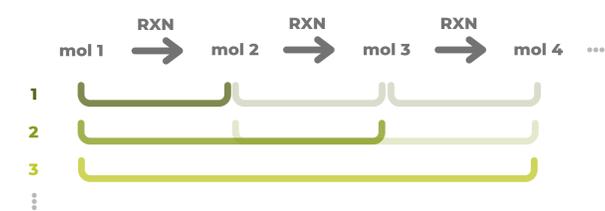


comparable classification performance, with **biosynthetic information in feature importances**

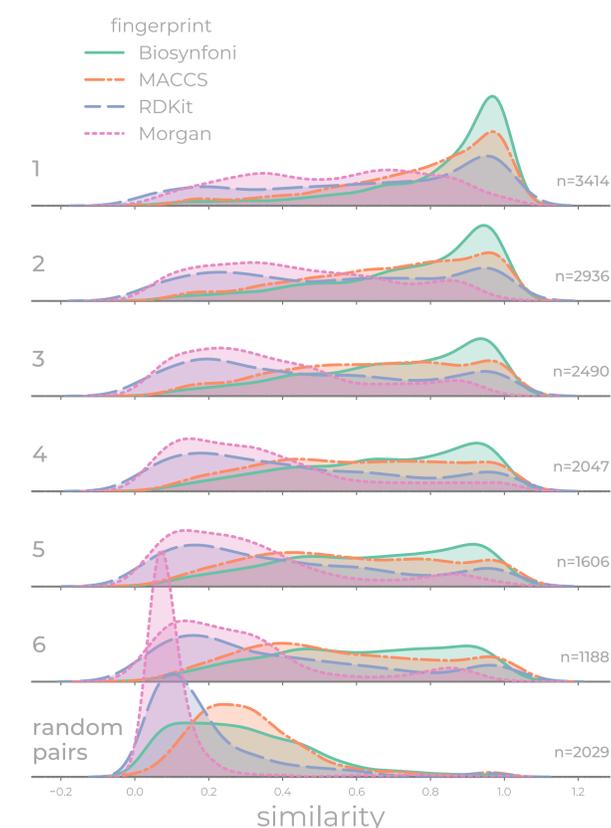


molecular QR-code visualisation of compound set substructure distribution allow interpretation of **biosynthetic differences** across metabolite groups

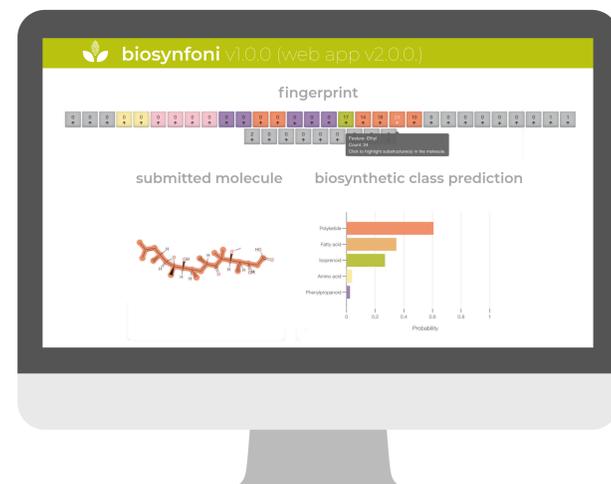
BIOSYNTHETIC SIMILARITY



smooth decreases in similarity score as enzymatic steps increase and clear **distinction of biosynthetically related and unrelated** compounds provide improved **resolution of biosynthetic relationships**



classification web app



REFERENCES

1. Dewick, P. M. Medicinal Natural Products: A Biosynthetic Approach. (Wiley, 2009).
2. Boldini, D. et al. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. *J Cheminform* **16**, 35 (2024).
3. Chandrasekhar, V. et al. COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Research* **53**, D634–D643 (2025).

Biosynfoni reflects biosynthetic information compactly and explicitly, enabling lightweight, **interpretable** natural product analyses that integrate smoothly into the computational workflows of natural product chemists.

