doi:10.2533/chimia.2026.676 Chimia 80 (2026) 676–684 © J.-L. Wolfender, L. Le Grégam, E. F. Queiroz

From LC-MS Metabolomic Profiling to AI-Enabled Navigation of NPs Chemical Space: Recent Advances and Emerging Directions

Jean-Luc Wolfender[a,b]*, Loïc Le Grégam[a,b], and Emerson Ferreira Queiroz[a,b]

Abstract: Natural products (NPs) occupy a unique region of chemical space and continue to provide privileged scaffolds for drug discovery, chemical ecology, and biotechnology. However, the exploration of NP diversity remains constrained by the complexity of biological extracts, the low annotation rate of LC-MS features, and the rediscovery of known metabolites. Recent advances in LC-HRMS/MS-based metabolomics, coupled with expanding spectral and structural databases, molecular networking, and *in silico* annotation methods, have considerably improved the characterization of complex metabolomes and facilitated large-scale exploration of natural chemical space. Simultaneously, artificial intelligence (AI) and machine-learning approaches are enhancing metabolite annotation, chemical classification, and prioritization, while knowledge graph technologies enable the integration of chemical, taxonomic, biological, and ecological information into interoperable, AI-ready frameworks. This mini-review highlights recent developments that support a transition from conventional metabolite annotation towards effective navigation of NP chemical space. We discuss the role of open data resources, computational metabolomics, knowledge-driven data integration, and emerging AI-assisted tools in accelerating NP discovery. We also examine how biotransformation and chemoenzymatic diversification can extend the boundaries of natural chemical space by generating nature-inspired molecules and previously inaccessible scaffolds. Together, these approaches are paving the way towards a more predictive, integrative, and data-driven paradigm for NP research and bioactive compound discovery.

Keywords: Artificial intelligence · Biotransformation · Knowledge graphs · Metabolomics · Natural products

Jean-Luc Wolfender is Professor at the School of Pharmaceutical Sciences of the University of Geneva, Switzerland, where he leads the Bioactive Natural Products Unit. His research focuses on the chemical and biological investigation of natural products, integrating advanced analytical chemistry, metabolomics, and bioinformatics. His group develops and applies cutting-edge approaches including LC-HRMS, NMR spectroscopy, molecular networking, and computational annotation to explore the chemodiversity of plants, microorganisms, and complex biological extracts. His work supports the discovery of bioactive metabolites, natural-product quality assessment, and the integration of artificial intelligence and knowledge-based methods to accelerate natural-products research and chemical-space exploration.

Loïc Le Grégam is a Postdoctoral Researcher in bioinformatics at the School of Pharmaceutical Sciences, University of Geneva, Switzerland, where his research focuses on computational mass spectrometry and metabolomics data integration. Within the Bioactive Natural Products Unit, he develops automated systems for assessing confidence in metabolite annotations and increasing annotation rates of natural products in complex samples, combining knowledge graphs, semantic web technologies, advanced computational MS methods and artificial intelligence. His work is part of the MetabolinkAI consortium (www.metabolinkai.net), which aims to accelerate discoveries in biomedical and environmental sciences through AI and comprehensive knowledge graphs.

Emerson Ferreira Queiroz is a Senior Lecturer and Research Associate at the School of Pharmaceutical Sciences, University of Geneva, Switzerland. His research focuses on the discovery and characterization of bioactive natural products through the integration of natural products chemistry, metabolomics, and advanced analytical technologies. He develops innovative workflows combining preparative separation methods and high-resolution analytical platforms for the efficient isolation and structural elucidation of natural compounds. His work also explores strategies to expand the chemical diversity of natural products through chemoenzymatic and biotransformation approaches, using abundant plant metabolites as sustainable substrates for the generation of novel bioactive molecules with potential pharmaceutical applications.

1. Introduction

Natural products (NPs), defined here as structurally characterized molecules isolated from biological sources, remain central to chemistry, pharmacy, agriculture, and health because they provide

*Correspondence: Dr. J.-L. Wolfender, E-mail: Jean-Luc.Wolfender@unige.ch
[a]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, CMU, CH-1211 Geneva, Switzerland; [b]School of Pharmaceutical Sciences, University of Geneva, CMU, CH-1211 Geneva, Switzerland

chemically diverse bioactive scaffolds.[1] They occupy a distinct and extensive chemical space, shaped by biosynthetic constraints and evolutionary selection. The documented NP space is already substantial. The proprietary Dictionary of Natural Products (DNP) contains nearly 300,000 compounds with associated structural, biological, bibliographic, and source-organism information,[2] while open databases such as COCONUT and LOTUS[3] collectively provide access to more than 125,000 plant-derived NPs.

Yet, the known NP space likely represents only a small fraction of nature's true chemical diversity. Current estimates suggest that plant chemical space (based on 400,000 species) may comprise millions of metabolites, with recent metabolomics-based projections indicating at least 1.5 million distinct compounds, although such estimates remain dependent on sampling depth and modelling assumptions.[1,4] Defining the actual boundaries of natural chemical space therefore remains a major challenge, closely linked to database coverage, taxonomic sampling, and the extent to which metabolomes can be characterized.[1,4,5]

Although several major pharmaceutical companies reduced their investment in NP-based drug discovery during the early 2000s due to challenges associated with compound isolation, supply, and compatibility with conventional high-throughput screening platforms,[6] NPs continue to play a pivotal role in modern drug discovery. Their unparalleled chemical diversity has provided numerous clinically relevant drugs and serves as a valuable source of novel chemical scaffolds and optimized derivatives. Beyond their therapeutic potential, NPs are increasingly recognized as powerful molecular probes for investigating biological pathways and elucidating complex cellular processes.[7] Concurrently, significant progress in screening technologies has transformed the evaluation of bioactive compounds. The shift from traditional phenotypic screening toward target-based high-throughput and high-content screening approaches has enabled the systematic assessment of compound activity, toxicity, and mechanisms of action at unprecedented scale and resolution. In addition, advances in metabolomics and other omics technologies have enhanced the ability to characterize NP-induced biological responses within complex cellular and organismal systems, thereby facilitating the identification of molecular targets and accelerating the translation of NP-derived compounds into therapeutic candidates.[8]

The main challenges in exploring the NP chemical space are well recognized: frequent rediscovery of known metabolites, the high complexity of extracts (often containing thousands of signals), the time and cost associated with isolation and structure elucidation[9] and the persistently low annotation rate of Liquid Chromatography–Mass Spectrometry (LC-MS) features (typically below 10%).[10] In parallel, access to natural resources is increasingly shaped by legal and ethical frameworks, including the Nagoya Protocol, which provides a foundation for more equitable and sustainable use of natural resources, while also supporting innovation and economic growth.[11]

Recent advances in Liquid Chromatography–High-Resolution Tandem Mass Spectrometry (LC-HRMS/MS)-based metabolomics, combined with complementary approaches such as advanced chemoinformatics tools, open-access NP and spectroscopic databases, and rapid, efficient targeted isolation at the microgram scale coupled with highly sensitive nuclear magnetic resonance (NMR) techniques, have transformed the field. These developments enable comprehensive mapping of natural extracts, facilitating efficient dereplication, prioritization of novel molecular scaffolds, and the generation of rich metadata describing chemical diversity, biological relevance, and ecological context.[9,12]

Comprehensive metabolic profiling now enables researchers to explore a virtual chemical space encompassing both NP analogues and *in silico*-predicted metabolites, extending well beyond the boundaries of traditional phytochemical investigations.[13] By integrating experimental metabolomic data with computational exploration, this approach facilitates the rapid detection and annotation of previously uncharacterized metabolites, while expanding opportunities for the discovery of novel bioactive structures.[14]

Concurrently, the integration of artificial intelligence (AI) and machine learning into computational metabolomics and chemoinformatics is substantially enhancing annotation and discovery workflows.[15] These advances include MS/MS-based compound classification,[16] knowledge-guided molecular network construction,[17] structure prediction,[15] and bioactivity-informed dereplication directly from LC-MS/MS data,[17] enabling more efficient navigation of the NP chemical space and accelerating the identification of promising lead compounds.[15]

This mini-review aims to illustrate how the integration of LC-MS-based metabolomics, artificial intelligence, and structured data resources (including spectral libraries, NP databases, and taxonomic and bioactivity metadata) enables a transition from conventional annotation workflows to true navigation within NP chemical space. This integrated strategy accelerates dereplication, limits rediscovery, and enhances access to novel chemical entities for bioactive compound discovery. Recent contributions from our group and collaborators highlight the potential of these emerging approaches.

2. Data Resources and Knowledge Bases for NPs

The rapid development of AI-driven approaches for NP research relies heavily on access to high-quality structural and spectroscopic databases, ideally centralized and openly accessible. In LC-MS-based metabolomics, these resources now constitute a critical information infrastructure. However, they remain largely distributed across numerous platforms, with heterogeneous formats and standards that still limit interoperability.

Databases play a central role in LC-MS/MS dereplication and annotation workflows, yet their coverage of NP chemodiversity remains incomplete and fragmented. Spectral repositories such as Global Natural Products Social Molecular Networking (GNPS),[18] which supports MS/MS data analysis and sharing through molecular networking, and MassBank, which provides high-quality reference spectra, have become essential resources for untargeted metabolomics.[19] Complementing these efforts, open NP structure databases such as LOTUS, to which we have contributed, consolidate structure-organism relationships extracted from the literature and connect chemical diversity with taxonomic and biological information.[3]

The chemical space covered by LOTUS is illustrated in Fig. 1 using a Tree MAP (TMAP) representation (https://github.com/reymond-group/tmap), where compounds are organized according to Tanimoto structural similarity and colored according to NPClassifier[20] superclasses. This visualization highlights the extensive chemodiversity of NPs and provides a framework for navigating NP chemical space.

Despite the availability of numerous specialized resources, the NP data landscape remains highly fragmented, with multimodal data often scattered across platforms and lacking standardized formats, thereby limiting interoperability and AI-ready integration.[21]

To address this challenge, OctoChemDB[22] has recently emerged as an identification-oriented aggregation layer integrating exact masses, MS/MS spectra, chemical classifications, and source-organism metadata. By consolidating complementary information from multiple resources, it provides a more comprehensive and customizable reference framework, particularly valuable for NPs that remain underrepresented in conventional spectral libraries. As such, it represents an important integrative component within a still fragmented and heterogeneous mass spectrometry data ecosystem.[23]

Within the same effort to improve spectroscopic data standardization and accessibility, we launched the noBS (natural

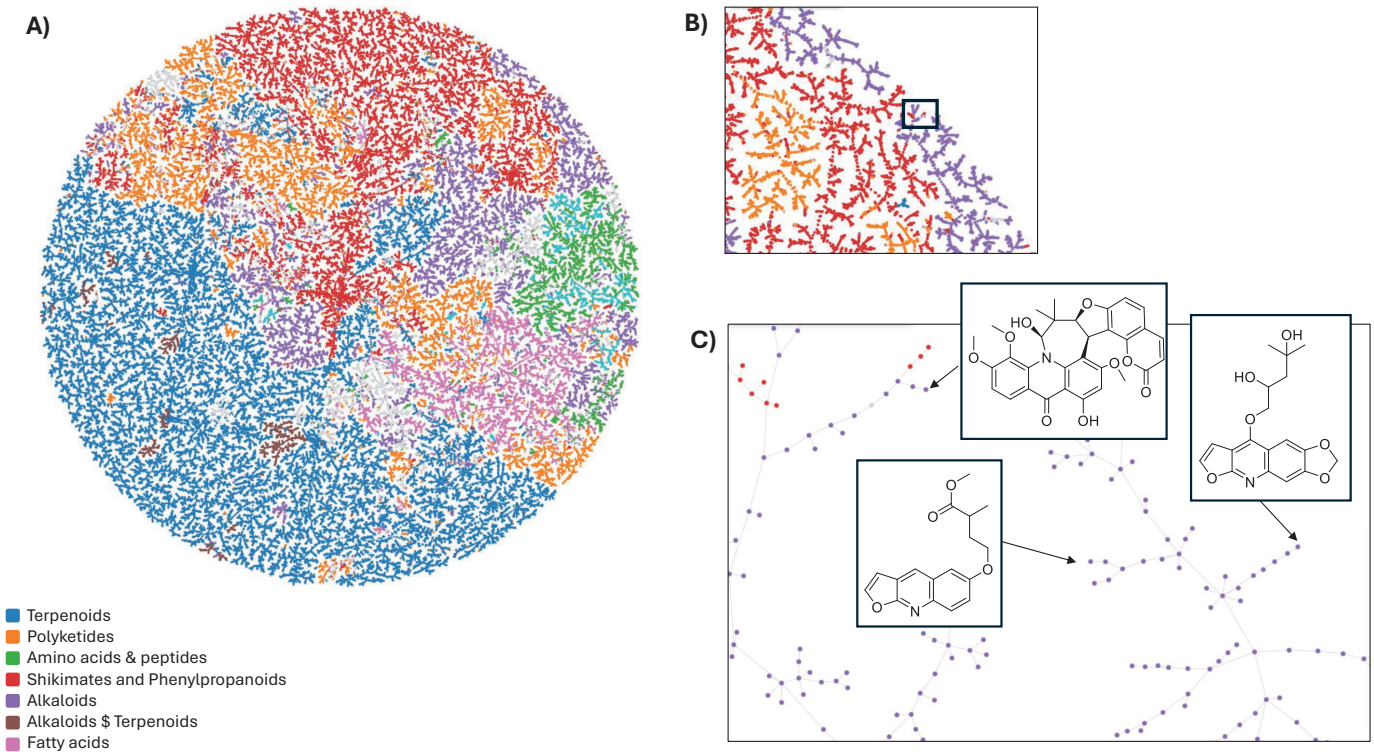

Fig. 1. TMAP visualization of the LOTUS natural product database. Each node represents a natural product, organized according to structural similarity and colored according to its NPClassifier superclass. The map highlights the extensive chemodiversity represented within LOTUS and facilitates exploration of natural product chemical space. (A) Overview of the complete LOTUS chemical space.[3] (B) Zoomed-in view of a region containing predominantly shikimates and phenylpropanoids (red), as well as alkaloids (purple) and polyketides (orange). (C) Further magnification of selected branches illustrating alkaloid (purple) clusters and their structural relationships across the chemical space. The presence of a dollar sign ($) between chemical groups indicates a dual annotation by NPClassifier, suggesting that the compound belong to two distinct chemical classes.

products Biodata Sharing) initiative (https://nobs.naturalproducts.net) together with a consortium of laboratories specializing in NP research. As a nascent community-driven effort, noBS focuses on isolated and fully characterized NPs, with the objective of aggregating fully assigned NMR and MS datasets. This would make them available in searchable, machine-readable formats compatible with computational workflows in metabolomics and AI-driven applications.

Overall, the development of interoperable databases integrating chemical structures, spectral information, and textual knowledge is becoming a prerequisite for advanced metabolite annotation and efficient exploration of NP chemical space. Continuous enrichment of these resources with high-quality, rigorously curated data will be essential to further improve annotation coverage, reliability, and accuracy, while enabling the next generation of AI-assisted approaches for NP research.

3. MS-Based Metabolomics and Metabolite Annotation

Natural extract profiling increasingly relies on untargeted LC-HRMS/MS workflows designed to capture the broad metabolite diversity present within complex samples. In this context, data-dependent acquisition (DDA) enables the simultaneous collection of HRMS and MS/MS data and has become a standard approach for annotation-oriented metabolomics. Coupling chromatographic separation with high-resolution, accurate-mass spectrometry improves the resolution of sample complexity, enhances metabolome coverage, and provides high quality spectral information required for large-scale metabolite annotation.[12]

Metabolite annotation in complex extracts from plants and microorganisms initially relies on the comparison of experimental spectra against libraries of experimentally acquired MS/MS spectra as well as in-silico MS/MS fragmentation databases. This strategy remains the cornerstone of dereplication for known compounds but is limited by the still incomplete coverage of spectral databases, particularly for specialized NPs. Platforms such as GNPS and feature-based molecular networking[18] organize spectra according to fragmentation similarity, facilitating the annotation of molecular families and the propagation of information from nodes (MS feature: *m*/*z* at RT) already identified through reference standards or library matching. These experimentally driven approaches are especially valuable for mapping the chemical space of complex extracts, visualizing structural analogues, and accelerating dereplication. However, a substantial fraction of detected features often remains unannotated when the corresponding metabolites are absent from reference libraries.[18,19,24]

To overcome the limitations of spectral library matching, *in silico* MS/MS annotation exploits fragmentation spectra to infer molecular formulas, rank candidate structures, and predict compound classes. Within this framework, SIRIUS has become a key platform for molecular formula annotation and structure elucidation from MS/MS data,[25] while CANOPUS enables compound-class prediction even for metabolites that are absent from spectral libraries.[16,26]

These approaches are particularly valuable for plant extracts and other natural-product-rich samples, where chemical diversity is high and confident metabolite identification remains challenging.[26] In such cases, class-level annotation can still provide biologically and chemically meaningful information. For NPs, this interpretation is further strengthened by the use of NPClassifier, a taxonomy specifically designed for NPs that is more informative than generic chemical classification systems such as ClassyFire.[20,26]

The SIRIUS ecosystem now integrates NPClassifier through CANOPUS, facilitating natural-product-oriented classification of annotated candidates. More recently, annotation workflows have

expanded beyond compound-class prediction by combining complementary strategies that support both dereplication and NP discovery. CANOPUS enables chemical-class annotation from MS/MS spectra,[16] SNAP-MS assigns putative NP compound families to molecular-network subnetworks,[27a] and MassKG supports knowledge-based *in silico* MS/MS annotation for rapid dereplication.[27b] Together, these complementary approaches provide increasingly informative chemical contextualization of complex NP datasets.[26]

The complementarity between experimental data and *in silico* approaches now forms the foundation of modern metabolomics workflows, extending annotation capabilities far beyond the compounds represented in reference libraries. Integrative platforms such as MolNetEnhancer exemplify this evolution by combining molecular networking, library matching, *in silico* annotation, and chemical classification within a unified framework.[28]

Reducing LC-MS data redundancy is another essential step for transforming thousands of detected features into a coherent representation of unique metabolites. Isotopes, adducts, in-source fragments, multimers, and contaminants can constitute a substantial proportion of the detected signals and artificially fragment the underlying chemical information. Although current curation tools greatly reduce this complexity, their performance remains dataset-dependent, making molecule-centered approaches (*i.e.* grouping multiple degenerate LC-MS features back to their common underlying metabolite) still necessary to improve dereplication and navigation of chemical space.[29]

In this context, we developed a taxonomically informed annotation workflow that combines conventional MS/MS similarity scores with a metascore based on the taxonomic proximity of source organisms.[26] By re-ranking candidate structures according to their taxonomic plausibility, this strategy improves annotation confidence while highlighting the importance of databases linking chemical structures to their biological origins. We subsequently extended this concept to an automated workflow combining untargeted LC-MS with charged aerosol detection (CAD), enabling the grouping of redundant signals, estimation of relative metabolite abundance, and generation of a simplified representation of complex extracts suitable for virtual annotation and metabolite prioritization.[26]

Using these metabolomics profiling and data deconvolution pipelines, we performed high-throughput LC-HRMS profiling of several thousand plant samples within an SNSF-Sinergia project, resulting in the creation of an open dataset comprising 1,600 extracts spanning a broad range of chemical diversity.[30] This resource, composed of extensively annotated mass spectra enriched with chemical, taxonomic, and biological metadata, provides a valuable foundation for NP discovery, exploration of plant chemical space, and prioritization of isolation strategies. Beyond these applications, it also offers a unique opportunity to investigate the ecological determinants of chemical diversity. Within the framework of the 'Functional Traits 2.0' concept, which considers metabolites as functional traits in their own right, extensive metabolome annotation is essential for linking chemical diversity to organismal function and ecological processes.[31]

Building upon this approach, the same chemical-space navigation strategies can be applied to explore chemotaxonomic relationships and the ecological drivers of metabolic diversity. For example, a study of Corsican bryophytes, based on untargeted ultra high performance liquid chromatography (UHPLC)-HRMS/MS profiling of 63 species, demonstrated that extensive metabolite annotation combined with molecular networking can clearly discriminate mosses from liverworts, identify biosynthetic signatures characteristic of specific lineages, and facilitate the discovery of novel metabolites.[32]

Overall, while confident annotation at the individual metabolite level remains beyond reach for a large proportion of detected features, the integration of orthogonal metadata and compound-class annotation already provides a reliable framework for interpreting complex metabolomes and generating

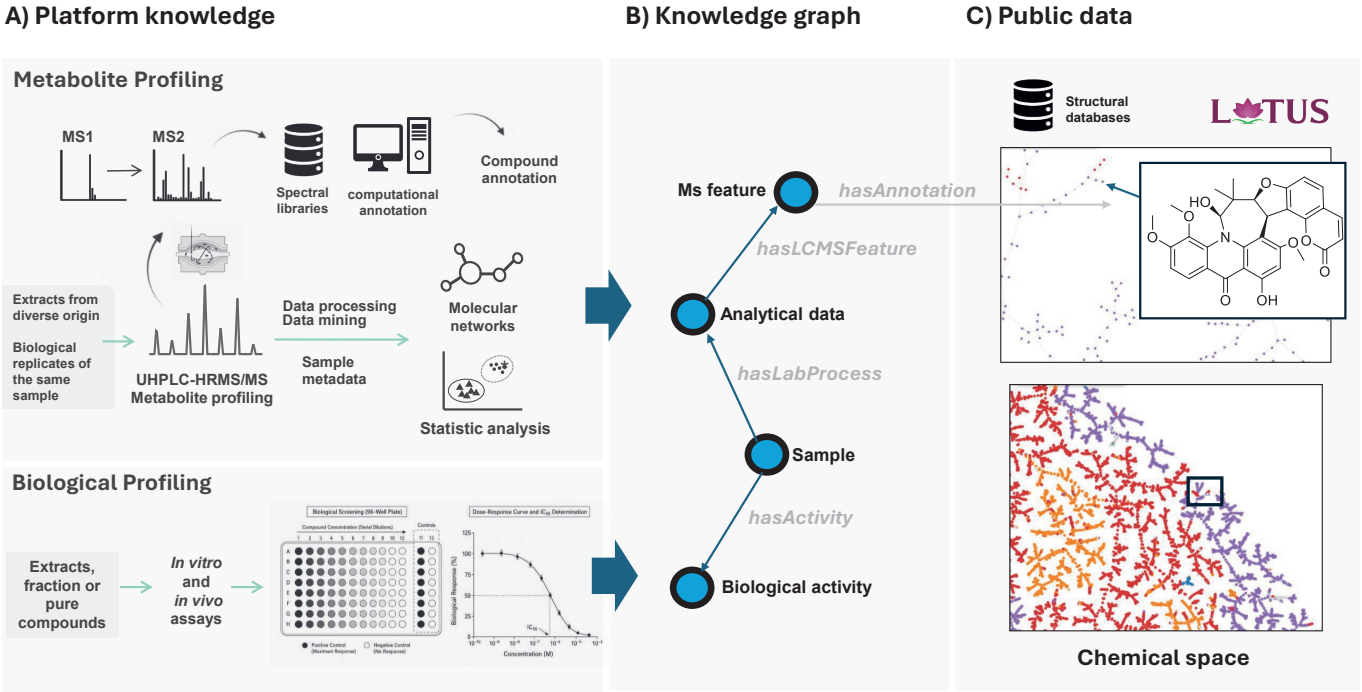

Fig. 2. Schematic overview of a typical metabolomics workflow, knowledge graph integration, and projection onto publicly available natural product chemical space. (A) Summary of the metabolomics profiling workflow, including LC-MS/MS data acquisition with data-dependent acquisition (DDA), generation of MS^1 and MS^2 spectra, and data processing through molecular networking (adapted from Ref. [13]). Parallel evaluation of biological activities. (B) Integration of analytical, biological, and metadata into a knowledge graph with nodes representing entities and edges representing semantic relationships. (C) Integration of public data from databases into knowledge graphs through semantic relationships, enabling interoperable exploration of chemical, biological, and taxonomic knowledge.

AI-ready datasets for NP research, chemical ecology, and drug discovery.

4. Data Structuring for Effective AI

The increasing volume and complexity of data generated by metabolomics profiling, deconvolution, and annotation workflows require data structures that support efficient integration, mining, and contextualization, ultimately improving biological interpretation. However, the heterogeneous nature of metabolomics data combined with the diversity of complementary knowledge sources (scientific literature, public databases, published datasets, taxonomic information, ...) poses a major challenge for conventional data integration approaches. In this context, knowledge graphs (KGs) provide a particularly relevant framework by linking spectra, metabolites, annotations, taxa, and metadata within a queryable semantic network that supports heterogeneous data integration. This graph-structured, semantically enriched data model enables researchers to explore complex relationships through explicit semantic links and logical reasoning. Furthermore, this same structure serves as a substrate on which machine learning models can learn embeddings (vector representations) and predict latent relationships not explicitly encoded in the graph.[21,33] In NP research, such unified KGs- integrating chemical structures, metabolomics data, genomic information (biosynthetic gene clusters, BGCs), biological screening results, and expert annotations - are increasingly viewed as core data infrastructures that enable AI systems to uncover global patterns, prioritize candidate compounds, and support expert reasoning during chemical space exploration.[15,21] Fig. 2 summarizes the workflow, integrating high-throughput metabolomics with spectral databases and *in silico* annotation tools. The resulting data are organized in a knowledge graph framework, enabling efficient exploration of NP chemical space and future AI-driven analyses.

Recent developments such as MetaboKG illustrate the value of knowledge graphs for metabolomics by integrating MS/MS spectra, analytical features, workflow outputs, annotations, confidence information, and repository metadata within a semantically integrated RDF knowledge graph that explicitly connects annotations to their underlying analytical evidence, samples, studies, and workflows. This interconnected structure supports data reuse, interoperability, traceability, and reproducible exploration of biochemical relationships across public metabolomics datasets while promoting FAIR data principles.[34]

Within this framework, we contributed to the development of the Experimental NPs Knowledge Graph (ENPKG),[35] which exemplifies the transition toward knowledge-driven metabolomics. Unlike conventional workflows based on global LC-MS/MS feature alignment, ENPKG adopts a sample-centric strategy, allowing new datasets to be progressively incorporated without requiring complete reprocessing. The workflow combines structural annotation tools (ISDB-LOTUS, SIRIUS, CSI:FingerID, and CANOPUS), taxonomic standardization, and enrichment through public resources such as Wikidata and ChEMBL before integrating the resulting information into a queryable resource description framework (RDF) KG.

Initially applied to the collection of 1,600 plant extracts systematically profiled by LC-HRMS/MS metabolomics,[30] and further enriched with public datasets, ENPKG enabled the construction of a network containing hundreds of millions of relationships explicitly connecting spectra, metabolites, taxa, and biological activities. This organization facilitates the simultaneous exploration of chemical, taxonomic, and biological dimensions and accelerates the identification of compounds of interest. As a proof of concept, the approach led to the identification of rotenoids active against *Trypanosoma cruzi* and triterpenoid derivatives active against *Leishmania donovani*.[36] More broadly, ENPKG demonstrates the potential of KGs as a key infrastructure for metabolomics data reuse, large-scale integration, and the future development of AI-driven approaches for exploring chemodiversity.[35]

The value of this strategy has been confirmed in several recent studies that used ENPKG to directly guide discovery campaigns. In the case of rotenoids, integrating chemical, biological, and taxonomic information within the KG enabled the rapid prioritization of the most promising candidates. This approach resulted in the isolation and biological evaluation of 41 compounds, including 11 rotenoids and several previously undescribed molecules, further demonstrating the ability of ENPKG to accelerate the identification and characterization of novel bioactive NPs.[36]

We also demonstrated that ENPKG can be used to select extracts based on their chemical proximity to a bioactive reference extract.[37] By integrating CANOPUS-derived compound-class annotations with similarity metrics such as the Jaccard index, we were able to perform semantic searches within the ENPKG dataset and identify metabolomic profiles exhibiting strong compositional similarities to samples of interest. This strategy highlighted *Gnetum edule*, a tropical woody vine, as a particularly promising source of oligomeric stilbenes. Subsequent phytochemical investigations led to the isolation of eleven stilbene oligomers, including six new compounds, and revealed significant anti-infective activity, particularly for (-)-gnetuhainin M.

Beyond KG construction itself, the emergence of large language models (LLMs) is opening new avenues for exploiting structured metabolomics knowledge. In this context, we participated in the development of MetaboT, a multi-agent framework dedicated to the interactive analysis of MS-based metabolomics data.[33b] Built on specialized agents capable of interpreting natural-language queries, querying databases, annotating spectra, and reasoning over chemical and biological relationships, MetaboT dynamically leverages spectral libraries, *in silico* annotation tools, and KGs. It transforms natural-language requests into automated analytical workflows, combining knowledge-based reasoning with the generative capabilities of LLMs. This approach facilitates compound identification, exploration of chemical families, structural analogue searches, and hypothesis generation, illustrating the convergence of metabolomics, KGs, and artificial intelligence toward tools capable of actively assisting researchers in data interpretation.

This convergence between metabolomics, knowledge representation, and artificial intelligence is paving the way for a new generation of research environments capable of actively exploiting data to drive scientific discovery. In this perspective, we are participating in the MetaboLinkAI project (2025–2029) (SNSF project: MetaboLinkAI, N°10002786), which combines KGs, artificial intelligence, and metabolomics data to develop use cases in NP research and biomedicine, build an interoperable knowledge hub,[38] and create an AI-powered research assistant dedicated to data exploration and hypothesis generation.

5. Nature-Inspired Expansion of Chemical Space

While metabolomics, knowledge graphs, and artificial intelligence now enable more efficient exploration of NP chemical space, a further step consists in actively extending its boundaries. Beyond the discovery of metabolites already produced by living organisms, biotransformation, biocatalysis, and chemoenzymatic diversification approaches make it possible to generate new nature-inspired molecules and access previously unexplored regions of chemical space.[39]

This strategy is based on the concept that NPs should be viewed not only as discovery targets but also as evolutionary platforms for the generation of new analogues and molecular scaffolds. This concept is currently being explored within the framework of a SNSF-funded project ‘Improving Natural Products

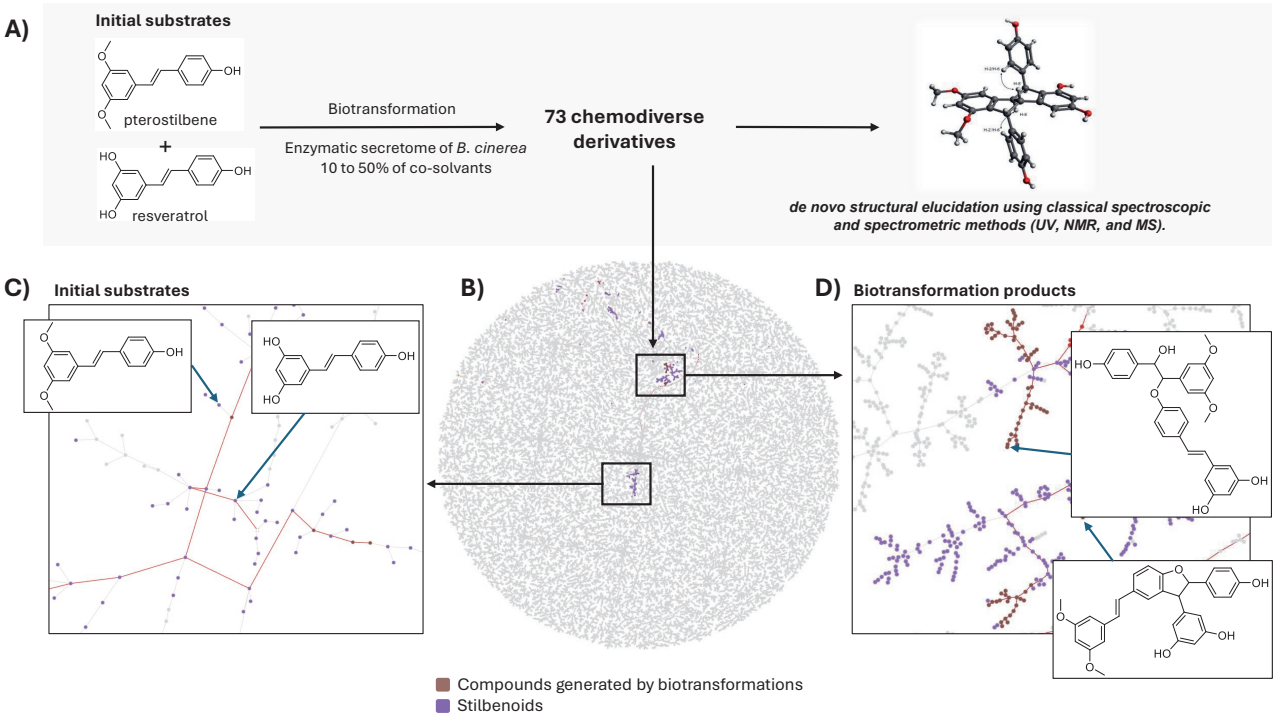

Fig. 3. Overview of the structural modification workflow and biotransformation process, highlighting the localization of newly generated structures within chemical space relative to their parent natural products. (A) Schematic representation of the biotransformation process leading to the formation of new metabolites. (B) TMAP visualization of the chemical space presented in Fig. 1, highlighting stilbenoids. (C) Zoomed-in view of the TMAP showing initial substrates used for subsequent biotransformations (brown) in comparison with naturally occurring stilbenoids already present in the chemical space. (D) Zoomed-in view of newly biotransformed compounds (brown) in comparison with naturally occurring stilbenoids already present in the chemical space.

Chemical Biodiversity for Drug Discovery by Fungal-Assisted Biotransformation' (Grant No. 205320_215078). Thanks to the remarkable selectivity of enzymes, transformations such as oxidation, coupling, esterification, and rearrangement reactions can be achieved with high chemo- and stereoselectivity, providing access to architectures that are often difficult to obtain through conventional synthetic approaches.[39] Thus by rapidly generating libraries of structural derivatives from known metabolites, these approaches expand the accessible chemical space while preserving the complexity and biological relevance characteristic of NPs. They therefore provide new opportunities for structure-activity relationship studies and the discovery of bioactive molecules.

In this context, integrating chemoenzymatic diversification with metabolomics, knowledge-graph frameworks, and artificial intelligence opens the way to a more rational exploration and expansion of NP chemical space aimed at the discovery of novel bioactive scaffolds.

With this objective in mind, we are currently investigating fungal secretomes as tools for molecular diversification. The secretome, comprising the ensemble of extracellular enzymes secreted by an organism, represents a rich and dynamic catalytic system.[40] In the filamentous fungus *Botrytis cinerea*, the secretome contains numerous oxidative enzymes capable of transforming phenolic compounds.[41] We are therefore evaluating its potential to generate new stilbene-derived scaffolds, a family of NPs known for its biological properties but still relatively underexplored from the standpoint of structural diversification.

Unlike approaches relying on whole microorganisms or isolated enzymes, the direct use of the secretome mobilizes a complex ensemble of enzymatic activities that can act cooperatively and sequentially, promoting access to molecular diversity that is difficult to achieve with simplified catalytic systems. Such approaches can generate highly complex metabolite mixtures from a single substrate through multiple, concurrent biotransformation pathways, yielding diverse combinations of chemical modifications and reaction products. This expanded chemodiversity provides a valuable opportunity to explore regions of NP chemical space that are otherwise difficult to access through conventional synthetic or biocatalytic methods. Ultimately, we aim to integrate these biocatalytic diversification strategies with the metabolomics and AI-driven approaches described above in order to extend, explore, and exploit NP chemical space in a more rational manner for the discovery of new bioactive scaffolds.

To expand the accessible chemodiversity, we employed the *B. cinerea* secretome to transform abundant NPs, particularly stilbenes. Reactions were performed in aqueous media, with selected co-solvents used when necessary to improve substrate solubility and diversify biocatalytic outcomes. Combined with UHPLC-HRMS profiling, this strategy revealed the appearance of compounds displaying unusual molecular formulas, suggesting the formation of new architectures, notably through stilbene dimerization processes.[42] It led to the isolation and structural characterization of 73 derivatives belonging to six structural families, including 55 previously undescribed compounds (Fig. 3).[42]

Biological evaluation of this library revealed several promising activities, including selective inhibition of the Wnt signaling pathway, a key regulator involved in numerous cancer-related processes.[42] Several derivatives also exhibited strong activity against *Staphylococcus aureus*, including methicillin-resistant strains (MRSA), with some trans-δ-viniferin dimers displaying activities comparable to those of reference antibiotics.[43] Moreover, the introduction of halogen atoms (Cl, Br, and I), which are rarely encountered in terrestrial plant metabolites, further expanded the accessible chemical space and generated derivatives with notable antibacterial and antiviral activities.[43a,44]

Analysis of these biotransformations also identified a key structural motif consisting of a *para*-hydroxyphenyl group conjugated to a double bond, which appears particularly favorable

for oxidative couplings catalyzed by the *B. cinerea* secretome.[45] Building on this observation, we extended the approach to other natural-product-derived substrates, including ferulic and caffeic acids. Their transformation yielded 23 derivatives, including 11 new compounds and several unprecedented scaffolds arising from dimerization, trimerization, and tetramerization processes.[46]

These findings highlight the potential of complex enzyme-mediated biotransformations as a powerful approach to expand the chemical space of NPs, creating nature-inspired molecular diversity and generating novel bioactive scaffolds. We anticipate that, once fully integrated with metabolomics, KG data model, and artificial intelligence tools, such approaches could enable a more rational exploration of chemodiversity. They are expected to help prioritize experimental efforts toward the most original and promising structures, thereby facilitating the discovery of bioactive natural-product-inspired molecules targeting therapeutically relevant biological mechanisms.

6. Conclusions

The convergence of MS-based metabolomics, the KG data model, and artificial intelligence is profoundly changing the scale at which NPs can be explored. High-throughput LC-MS/MS platforms now provide access to an ever-growing fraction of the chemodiversity produced by living organisms, while machine-learning approaches increasingly improve metabolite annotation, classification, and prediction of biological properties.[15,47]

Despite these advances, achieving reliable and confident metabolite annotation remains a major challenge in NP metabolomics. Limitations in spectral coverage, reference data availability, annotation confidence, and the inherent constraints of MS/MS fragmentation, together with the instrument-dependent nature of tandem mass spectra, continue to hamper accurate compositional assessments of complex biological samples. Consequently, further progress in AI-driven metabolomics and NP discovery depends not only on methodological developments but also on the availability of high-quality, standardized, and interoperable datasets.

The implementation of findable, accessible, interoperable, and reusable (FAIR) principles has therefore emerged as a prerequisite for fully realizing the potential of AI in NP and drug discovery.[48] In this context, several community-driven initiatives are helping to build a global knowledge infrastructure. LOTUS links chemical structures to producing organisms, OctoChemDB federates chemical, biological, and spectral information from multiple resources, noBS promotes the sharing of fully assigned reference spectroscopic datasets, and the Earth Metabolome Initiative (https://www.earthmetabolome.org/) aims to map the planet's metabolic diversity on an unprecedented scale.

Together, these initiatives are laying the foundations of an integrated and FAIR-compliant ecosystem in which chemical structures, spectral data, taxonomy, bioactivities, and ecological metadata can be interconnected through AI-ready KGs. Such infrastructures are essential for improving metabolite annotation confidence, enabling more accurate compositional assessments, and supporting the development of robust machine-learning models for NP research.

Beyond accelerating NP discovery, comprehensive metabolome coverage of complex natural extracts offers unprecedented opportunities to better understand the relationships between metabolite composition, chemical ecology, and chemotaxonomy. It also provides a framework for rationalizing the traditional and contemporary use of medicinal plants by linking their demonstrated biological activities to their underlying metabolite profiles. In this context, AI has the potential to uncover hidden relationships across diverse phytochemical compositions associated with a given therapeutic indication or pharmacological property. By integrating large-scale metabolomics, bioactivity, taxonomic, and ecological datasets, AI approaches can reveal patterns that are difficult to identify through conventional reductionist strategies, including the contribution of minor constituents, multi-target effects, and potential synergistic interactions among metabolites. Such holistic analyses may ultimately provide a more comprehensive understanding of the molecular basis of efficacy in complex natural extracts and guide the identification of bioactive chemical signatures associated with specific biological functions.

Within this framework, artificial intelligence is not intended to replace the experimental approaches that are indispensable in NP research for the chemical and biological characterization of compounds, nor the human expert knowledge for data interpretation. Rather, it provides a powerful methodological framework for integrating and mining metabolomics, genomics, bioactivity, and environmental data, thereby facilitating the exploration of biological chemical space, accelerating hypothesis generation, and supporting the discovery of novel bioactive compounds.

Combined with emerging strategies for expanding this chemical space through biocatalysis and nature-inspired diversification, this convergence is paving the way for a more rational discovery of novel bioactive molecules and molecular scaffolds. NP research is thus evolving toward an increasingly predictive, collaborative, and cumulative discipline, rooted in the observation of living systems, amplified by omics technologies and artificial intelligence, and supported by open data infrastructures collectively developed by the communities of chemistry, biology, computer science, and health sciences.

More broadly, these developments bring us closer to a long-standing objective: to explore, understand, and now rationally expand the chemical space of life. The challenge is no longer limited to navigating known chemodiversity, but also to pushing its boundaries in order to access new molecular territories that may help address major challenges in health, sustainability, and society. NP research is therefore extending far beyond the molecules currently observed in nature to encompass their potential extensions and unexplored derivatives, enabling an ever-deeper exploration of natural chemical space and its fringes.

Acknowledgments

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the ANR-SNSF project 'MetaboLinkAI' (Grant No. IC00I0L-231646), funded by the Swiss National Science Foundation (SNSF). The authors also gratefully acknowledge financial support from SNSF for the project 'Improving Natural Products Chemical Biodiversity for Drug Discovery by Fungal-Assisted Biotransformation' (Grant No. 205320_215078), led by Katia Gindro (Agroscope) and Emerson F. Queiroz (University of Geneva). The authors gratefully acknowledge Arnaud Gaudry for providing the TMAP generation code. The authors acknowledge the use of Microsoft Copilot (Microsoft, https://copilot.microsoft.com/) for language editing and academic style improvement, and Consensus (https://consensus.app/) to assist with literature discovery and information gathering. All retrieved sources were independently evaluated and interpreted by the authors. The authors take full responsibility for the content, accuracy, and conclusions of this paper.

Received: July 15, 2026

[1] C. E. Hart, Y. Gadiya, T. Kind, C. A. Krettler, M. Gaetz, B. B. Misra, D. Healey, A. Allen, V. Colluru, D. Domingo-Fernández, *Gigascience* **2025**, *14*, giaf033, https://doi.org/10.1093/gigascience/giaf033.

[2] F. Chassagne, G. Cabanac, G. Hubert, B. David, G. Marti, *Phytochem. Rev.* **2019**, *18*, 601, https://doi.org/10.1007/s11101-019-09606-2.

[3] A. Rutz, M. Sorokina, J. Galgonek, D. Mietchen, E. Willighagen, A. Gaudry, J. G. Graham, R. Stephan, R. Page, J. Vondrasek, C. Steinbeck, G. F. Pauli, J. L. Wolfender, J. Bisson, P. M. Allard, *eLife* **2022**, *11*, https://doi.org/10.7554/eLife.70780.

[4] D. Boldini, D. Ballabio, V. Consonni, R. Todeschini, F. Grisoni, S. A. Sieber, *J. Cheminf.* **2024**, *16*, 35, https://doi.org/10.1186/s13321-024-00830-3.

[5] R. G. Linington, *ACS Cent. Sci*. **2025**, *11*, 1536, https://doi.org/10.1021/acscentsci.5c00804.

[6] B. David, J. L. Wolfender, D. A. Dias, *Phytochem. Rev.* **2015**, *14*, 299, https://doi.org/10.1007/s11101-014-9367-z.

[7] T. Rodrigues, D. Reker, P. Schneider, G. Schneider, *Nat. Chem*. **2016**, *8*, 531, https://doi.org/10.1038/nchem.2479.

[8] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, The International Natural Product Sciences Taskforce, C. T. Supuran, *Nat. Rev. Drug Discov*ery **2021**, *20*, 200, https://doi.org/10.1038/s41573-020-00114-z.

[9] D. P. Demarque, R. G. Dusi, F. D. M. de Sousa, S. M. Grossi, M. R. S. Silverio, N. P. Lopes, L. S. Espindola, *Sci. Rep*. **2020**, *10*, 1051, https://doi.org/10.1038/s41598-020-58046-y.

[10] N. F. de Jonge, K. Mildau, D. Meijer, J. J. R. Louwen, C. Bueschl, F. Huber, J. J. J. van der Hooft, *Metabolomics* **2022**, *18*, 103, https://doi.org/10.1007/s11306-022-01963-y.

[11] S. P. Gaudêncio, B. K. Stroil, L. Novoveska, C. Rebours, X. T. Schneider, *Nat. Prod. Rep*. **2026**, *43*, 1403, https://doi.org/10.1039/d5np00072f.

[12] J. L. Wolfender, J. M. Nuzillard, J. J. J. van der Hooft, J. H. Renault, S. Bertrand, *Anal. Chem*. **2019**, *91*, 704, https://doi.org/10.1021/acs.analchem.8b05112.

[13] J. L. Wolfender, A. Gaudry, A. Rutz, L. M. Quiros-Guerrero, L. F. Nothias, E. Ferreira Queiroz, E. Defossez, P. M. Allard, *CHIMIA* **2022**, *76*, 954, https://doi.org/10.2533/chimia.2022.954.

[14] M. A. Beniddir, K. B. Kang, G. Genta-Jouve, F. Huber, S. Rogers, J. J. J. van der Hooft, *Nat. Prod. Rep*. **2021**, *38*, 1967, https://doi.org/10.1039/d1np00023c.

[15] M. W. Mullowney, K. R. Duncan, S. S. Elsayed, N. Garg, J. J. J. van der Hooft, N. I. Martin, D. Meijer, B. R. Terlouw, F. Biermann, K. Blin, J. Durairaj, M. Gorostiola Gonzalez, E. J. N. Helfrich, F. Huber, S. Leopold-Messer, K. Rajan, T. de Rond, J. A. van Santen, M. Sorokina, M. J. Balunas, M. A. Beniddir, D. A. van Bergeijk, L. M. Carroll, C. M. Clark, D. A. Clevert, C. A. Dejong, C. Du, S. Ferrinho, F. Grisoni, A. Hofstetter, W. Jespers, O. V. Kalinina, S. A. Kautsar, H. Kim, T. F. Leao, J. Masschelein, E. R. Rees, R. Reher, D. Reker, P. Schwaller, M. Segler, M. A. Skinnider, A. S. Walker, E. L. Willighagen, B. Zdrazil, N. Ziemert, R. J. M. Goss, P. Guyomard, A. Volkamer, W. H. Gerwick, H. U. Kim, R. Muller, G. P. van Wezel, G. J. P. van Westen, A. K. H. Hirsch, R. G. Linington, S. L. Robinson, M. H. Medema, *Nat. Rev. Drug. Discovery* **2023**, *22*, 895, https://doi.org/10.1038/s41573-023-00774-7.

[16] K. Duhrkop, L. F. Nothias, M. Fleischauer, R. Reher, M. Ludwig, M. A. Hoffmann, D. Petras, W. H. Gerwick, J. Rousu, P. C. Dorrestein, S. Bocker, *Nat. Biotechnol*. **2021**, *39*, 462, https://doi.org/10.1038/s41587-020-0740-8.

[17] L. F. Nothias, M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A. Sarvepalli, P. Leyssen, D. Touboul, J. Costa, J. Paolini, T. Alexandrov, M. Litaudon, P. C. Dorrestein, *J. Nat. Prod*. **2018**, *81*, 758, https://doi.org/10.1021/acs.jnatprod.7b00737.

[18] L. F. Nothias, D. Petras, R. Schmid, K. Duhrkop, J. Rainer, A. Sarvepalli, I. Protsyuk, M. Ernst, H. Tsugawa, M. Fleischauer, F. Aicheler, A. A. Aksenov, O. Alka, P. M. Allard, A. Barsch, X. Cachet, A. M. Caraballo-Rodriguez, R. R. Da Silva, T. Dang, N. Garg, J. M. Gauglitz, A. Gurevich, G. Isaac, A. K. Jarmusch, Z. Kamenik, K. B. Kang, N. Kessler, I. Koester, A. Korf, A. Le Gouellec, M. Ludwig, H. C. Martin, L. I. McCall, J. McSayles, S. W. Meyer, H. Mohimani, M. Morsy, O. Moyne, S. Neumann, H. Neuweger, N. H. Nguyen, M. Nothias-Esposito, J. Paolini, V. V. Phelan, T. Pluskal, R. A. Quinn, S. Rogers, B. Shrestha, A. Tripathi, J. J. J. van der Hooft, F. Vargas, K. C. Weldon, M. Witting, H. J. Yang, Z. Zhang, F. Zubeil, O. Kohlbacher, S. Bocker, T. Alexandrov, N. Bandeira, M. X. Wang, P. C. Dorrestein, *Nat. Methods* **2020**, *17*, 905, https://doi.org/10.1038/s41592-020-0933-6.

[19] M. X. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Criisemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Q. Dai, R. Neupane, J. Gurr, A. M. C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrovr, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K. M. Waters, W. Y. Shi, X. T. Liu, L. X. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutierrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol*. **2016**, *34*, 828, https://doi.org/10.1038/nbt.3597.

[20] H. W. Kim, M. X. Wang, C. A. Leber, L. F. Nothias, R. Reher, K. B. Kang, J. J. J. van der Hooft, P. C. Dorrestein, W. H. Gerwick, G. W. Cottrell, *J. Nat. Prod*. **2021**, *84*, 2795, https://doi.org/10.1021/acs.jnatprod.1c00399.

[21] D. Meijer, M. A. Beniddir, C. W. Coley, Y. M. Mejri, M. Ozturk, J. J. J. van der Hooft, M. H. Medema, A. Skiredj, *Nat. Prod. Rep.* **2025**, *42*, 654, https://doi.org/10.1039/d4np00008k.

[22] R. Silvestre, R. Martinent, L. Menin, N. Gasilova, V. Mutel, C. Portmann, L. Patiny, *Anal. Chem*. **2026**, *98*, 6102, https://doi.org/10.1021/acs.analchem.5c06761.

[23] R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T. S. Dyrlund, R. Bushuiev, K. J. Murray, N. Hoffmann, M. Lu, A. Sarvepalli, Z. Zhang, M. Fleischauer, K. Duhrkop, M. Wesner, S. J. Hoogstra, E. Rudt, O. Mokshyna, C. Brungs, K. Ponomarov, L. Mutabdzija, T. Damiani, C. J. Pudney, M. Earll, P. O. Helmer, T. R. Fallon, T. Schulze, A. Rivas-Ubach, A. Bilbao, H. Richter, L. F. Nothias, M. Wang, M. Oresic, J. K. Weng, S. Bocker, A. Jeibmann, H. Hayen, U. Karst, P. C. Dorrestein, D. Petras, X. Du, T. Pluskal, *Nat. Biotechnol*. **2023**, *41*, 447, https://doi.org/10.1038/s41587-023-01690-2.

[24] A. T. Aron, E. C. Gentry, K. L. McPhail, L. F. Nothias, M. Nothias-Esposito, A. Bouslimani, D. Petras, J. M. Gauglitz, N. Sikora, F. Vargas, J. J. J. van der Hooft, M. Ernst, K. B. Kang, C. M. Aceves, A. M. Caraballo-Rodriguez, I. Koester, K. C. Weldon, S. Bertrand, C. Roullier, K. Sun, R. M. Tehan, P. C. Boya, M. H. Christian, M. Gutierrez, A. M. Ulloa, J. A. Tejeda Mora, R. Mojica-Flores, J. Lakey-Beitia, V. Vasquez-Chaves, Y. Zhang, A. I. Calderon, N. Tayler, R. A. Keyzers, F. Tugizimana, N. Ndlovu, A. A. Aksenov, A. K. Jarmusch, R. Schmid, A. W. Truman, N. Bandeira, M. Wang, P. C. Dorrestein, *Nat. Protoc*. **2020**, *15*, 1954, https://doi.org/10.1038/s41596-020-0317-5.

[25] K. Duhrkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Bocker, *Nat. Methods* **2019**, *16*, 299, https://doi.org/10.1038/s41592-019-0344-8.

[26] A. Rutz, J. L. Wolfender, *J. Agric. Food Chem.* **2023**, *71*, 18010, https://doi.org/10.1021/acs.jafc.3c03099.

[27] a) N. J. Morehouse, T. N. Clark, E. J. McMann, J. A. van Santen, F. P. J. Haeckl, C. A. Gray, R. G. Linington, *Nat. Commun.* **2023**, *14*, 308, https://doi.org/10.1038/s41467-022-35734-z; b) B. Zhu, Z. Li, Z. Jin, Y. Zhong, T. Lv, Z. Ge, H. Li, T. Wang, Y. Lin, H. Liu, T. Ma, S. Wang, J. Liao, X. Fan, *Comput. Struct. Biotechnol. J.* **2024**, *23*, 3327, https://doi.org/10.1016/j.csbj.2024.09.001.

[28] a) M. Ernst, K. B. Kang, A. M. Caraballo-Rodriguez, L. F. Nothias, J. Wandy, C. Chen, M. Wang, S. Rogers, M. H. Medema, P. C. Dorrestein, J. J. J. van der Hooft, *Metabolites* **2019**, *9*, 144, https://doi.org/10.3390/metabo9070144; b) M. A. Hoffmann, L. F. Nothias, M. Ludwig, M. Fleischauer, E. C. Gentry, M. Witting, P. C. Dorrestein, K. Duhrkop, S. Bocker, *Nat. Biotechnol*. **2022**, *40*, 411, https://doi.org/10.1038/s41587-021-01045-9.

[29] a) N. G. Mahieu, G. J. Patti, *Anal. Chem*. **2017**, *89*, 10397, https://doi.org/10.1021/acs.analchem.7b02380; b) W. J. Nash, J. B. Ngere, L. Najdekr, W. B. Dunn, *Anal. Chem*. **2024**, *96*, 10935, https://doi.org/10.1021/acs.analchem.4c00966.

[30] P. M. Allard, A. Gaudry, L. M. Quiros-Guerrero, A. Rutz, M. Dounoue-Kubo, T. W. N. Walker, E. Defossez, C. Long, A. Grondin, B. David, J. L. Wolfender, *Gigascience* **2022**, *12*, giac124, https://doi.org/10.1093/gigascience/giac124.

[31] T. W. N. Walker, J. M. Alexander, P. M. Allard, O. Baines, V. Baldy, R. D. Bardgett, P. Capdevila, P. D. Coley, B. David, E. Defossez, M. J. Endara, M. Ernst, C. Fernandez, D. Forrister, A. Gargallo-Garriga, V. E. J. Jassey, S. Marr, S. Neumann, L. Pellissier, J. Peñuelas, K. Peters, S. Rasmann, U. Roessner, J. Sardans, F. Schrodt, M. C. Schuman, A. Soule, H. Uthe, W. Weckwerth, J. L. Wolfender, N. M. van Dam, R. Salguero-Gómez, *J. Ecol*. **2022**, *110*, 4, https://doi.org/10.1111/1365-2745.13826.

[32] A. Pannequin, A. Muselli, L. Marcourt, E. Ferreira Queiroz, L. M. Quiros-Guerrero, Y. Asakawa, M. Dounoue-Kubo, J. L. Wolfender, *Front. Plant Sci*. **2024**, *15*, 1470307, https://doi.org/10.3389/fpls.2024.1470307.

[33] a) M. Delmas, O. Filangi, N. Paulhe, F. Vinson, C. Duperier, W. Garrier, P. E. Saunier, Y. Pitarch, F. Jourdan, F. Giacomoni, C. Frainay, *Bioinformatics* **2021**, *37*, 3896, https://doi.org/10.1093/bioinformatics/btab627; b) M. Bekbergenova, L. Pradi, B. Navet, E. Tysinger, F. Michel, M. Feraud, Y. Taghzouti, Y. Z. Chen, O. Kirchhoffer, F. Mehl, M. Legrand, T. Jiang, M. Pagni, S. Hassoun, J.-L. Wolfender, W. Bittremieux, F. Gandon, L.-F. Nothias, *J. Cheminformatics* **2026**, https://doi.org/10.1186/s13321-026-01257-8.

[34] M. Feraud, D. Boukhajou, F. L. Gandon, L.-F. Nothias, *arXiv* **2026**, https://doi.org/10.48550/arXiv.2605.24706.

[35] A. Gaudry, M. Pagni, F. Mehl, S. Moretti, L. M. Quiros-Guerrero, L. Cappelletti, A. Rutz, M. Kaiser, L. Marcourt, E. F. Queiroz, J. R. Ioset, A. Grondin, B. David, J. L. Wolfender, P. M. Allard, *Acs Cent. Sci*. **2024**, *10*, 494, https://doi.org/10.1021/acscentsci.3c00800.

[36] A. Gaudry, L. Marcourt, M. Kaiser, J. Fluckiger, B. David, A. Grondin, J. R. Ioset, P. Maser, E. F. Queiroz, P. M. Allard, J. L. Wolfender, *RSC Adv.* **2025**, *15*, 15240, https://doi.org/10.1039/d4ra08652j.
[37] O. A. Kirchhoffer, L. Quiros-Guerrero, J. Nitschke, L. F. Nothias, F. Burdet, L. Marcourt, N. Hanna, F. Mehl, B. David, A. Grondin, E. F. Queiroz, M. Pagni, T. Soldati, J. L. Wolfender, *RSC Adv*. **2025**, *15*, 13010, https://doi.org/10.1039/d4ra08421g.
[38] D. Tandon, T. Mendes De Farias, P.-M. Allard, E. Defossez, *bioRxiv* **2026**, https://doi.org/10.1101/2025.08.20.671289.
[39] J. Schwarz, K. Rosenthal, R. Snajdrova, M. Kittelmann, S. Lutz, *CHIMIA* **2020**, *74*, 368, https://doi.org/10.2533/chimia.2020.368.
[40] C. Filiatrault-Chastel, S. Heiss-Blanquet, A. Margeot, J. G. Berrin, *Biotechnol. Adv*. **2021**, *52*, 107833, https://doi.org/10.1016/j.biotechadv.2021.107833.
[41] K. Gindro, S. Schnee, D. Righi, L. Marcourt, S. Nejad Ebrahimi, J. M. Codina, F. Voinesco, E. Michellod, J. L. Wolfender, E. F. Queiroz, *J. Nat. Prod*. **2017**, *80*, 887, https://doi.org/10.1021/acs.jnatprod.6b00760.
[42] R. Huber, A. Koval, L. Marcourt, M. Heritier, S. Schnee, E. Michellod, L. Scapozza, V. L. Katanaev, J. L. Wolfender, K. Gindro, E. F. Queiroz, *Front. Chem*. **2022**, *10*, https://doi.org/10.3389/fchem.2022.881298.
[43] a) R. Huber, L. Marcourt, M. Heritier, A. Luscher, L. Guebey, S. Schnee, E. Michellod, S. Guerrier, J. L. Wolfender, L. Scapozza, T. Kohler, K. Gindro, E. F. Queiroz, *Sci. Rep*. **2023**, *13*, 15986, https://doi.org/10.1038/s41598-023-43000-5; b) R. Huber, L. Marcourt, L. M. Quiros-Gurerrero, A. Luscher, S. Schnee, E. Michellod, V. Ducret, K. Perron, J. L. Wolfender, K. Gindro, E. F. Queiroz, *Front. Chem*. **2022**, *10*, 912396, https://doi.org/10.3389/fchem.2022.912396; c) D. Righi, R. Huber, A. Koval, L. Marcourt, S. Schnee, A. Le Floch, V. Ducret, R. Perozzo, C. C. de Ruvo, N. Lecoultre, E. Michellod, S. N. Ebrahimi, E. Rivara-Minten, V. L. Katanaev, K. Perron, J. L. Wolfender, K. Gindro, E. F. Queiroz, *J. Nat. Prod.* **2020**, *83*, 2347, https://doi.org/10.1021/acs.jnatprod.0c00071.
[44] a) A. C. Zwygart, C. Medaglia, R. Huber, R. Poli, L. Marcourt, S. Schnee, E. Michellod, B. Mazel-Sanchez, S. Constant, S. Huang, M. Bekliz, S. Clement, K. Gindro, E. F. Queiroz, C. Tapparel, *Biomed. Pharmacother*. **2023**, *163*, 114825, https://doi.org/10.1016/j.biopha.2023.114825; b) J. Nitschke, R. Huber, S. Vossio, D. Moreau, L. Marcourt, K. Gindro, E. F. Queiroz, T. Soldati, N. Hanna, *Front. Microbiol*. **2024**, *15*, https://doi.org/10.3389/fmicb.2024.1439814.
[45] R. Huber, L. Marcourt, F. Félix, S. Tardy, E. Michellod, L. Scapozza, J. L. Wolfender, K. Gindro, E. F. Queiroz, *Front. Chem*. **2024**, *12*, https://doi.org/10.3389/fchem.2024.1390066.
[46] R. Huber, L. Marcourt, A. Koval, S. Schnee, D. Righi, E. Michellod, V. L. Katanaev, J. L. Wolfender, K. Gindro, E. F. Queiroz, *Front. Plant Sci*. **2021**, *12*, 805610, https://doi.org/10.3389/fpls.2021.805610.
[47] N. J. Brittin, J. M. Anderson, D. R. Braun, S. R. Rajski, C. R. Currie, T. S. Bugni, *J. Nat. Prod*. **2025**, *88*, 361, https://doi.org/10.1021/acs.jnatprod.4c01123.
[48] H. O. Brinkhaus, K. Rajan, J. Schaub, A. Zielesny, C. Steinbeck, *Curr. Opin. Struct. Biol*. **2023**, *79*, 102542, https://doi.org/10.1016/j.sbi.2023.102542.

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (https://chimia.ch/chimia/about).

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2026.676