



Cite this: DOI: 10.1039/d6np00016a

Structural determination of natural products at speeds and scale of high-throughput screening

Christopher C. Thornburg,^a Rohitesh Kumar,^a Dongdong Wang,^a Barry R. O'Keefe^{*bc} and Tanja Grkovic^{bc}

Covering 2000 up to 2025.

This review assesses analytical techniques available for routine structural elucidation of small molecules as part of high-throughput screening campaigns of natural product libraries. The emphasis is on the speed and scale of task management and highlights advances in technologies and library tools that enable the structural elucidation of complex molecules on small-scale and rapid timelines. We present the state of the art in the natural product field and provide a critical assessment of available techniques and tools, with a focus on the ability to analyze large numbers of samples in short timelines that can meet the speed and scale of high-throughput screening.

Received 5th February 2026

DOI: 10.1039/d6np00016a

rsc.li/npr

1. Introduction
2. Separation approaches
3. Mass spectrometry techniques
 - 3.1 Ionization sources enabling rapid NP analysis
 - 3.2 Mass analyzer platforms and their impact on throughput
4. Nuclear magnetic resonance techniques
 - 4.1 NMR magnet and probe technology
 - 4.2 Select pulse sequences and NMR experiments
 - 4.3 Non-uniform sampling
5. Other techniques
 - 5.1 Infrared spectroscopy
 - 5.2 Chiroptical spectroscopic techniques
6. Database tools
 - 6.1 Mass spectral databases
 - 6.2 NMR spectral databases
7. Meeting the speed and scale of HTS
 - 7.1 Case studies for the HT chemical analysis of NP hits from HTS
8. Conclusions and perspectives
9. Author contributions
10. Conflicts of interest
11. Data availability
12. Acknowledgements
13. References

1. Introduction

High-throughput screening (HTS) is an automated process that tests large numbers of biological and chemical samples using robotic platforms. In natural product-based HTS, libraries of crude natural product (NP) extracts and semi-purified NP fractions are screened against cell-based and cell-free disease targets to identify active molecules, which are then examined for specific activity and future drug development. A key advantage of modern HTS is the ability to use high-capacity microplates (384- or 1536-well) to achieve a throughput of >10 000 samples per day. Ideally, diverse and well-annotated libraries support the success of HTS discovery campaigns. Although some early HTS concepts were developed using NP libraries,¹ their continued use decreased due to incompatibility with robotic platforms and HTS timelines.² Notable recent examples of NPs or NP-derived compounds in pre-clinical drug development include Zotatifin, which is based on the plant natural product rocaglamide,³ and RMC-6236, which is based on the bacterial natural product sanglifhehrin.⁴ Despite these current examples, the speed of identifying activity for a natural product sample has for many years eclipsed the speed of subsequent compound identification. Recent developments are trying to address the discrepancy and bring NP chemistry efforts more in line with HTS programs.

Extracts from plants, marine biota, or microbial cultures can be incredibly complex, generally comprising dozens to thousands of chemical entities with a range of polarities and physicochemical properties. Mixtures of unknown solubility and viscosity can pose serious challenges for liquid-handling robots. Moreover, NP extracts, depending on the source biota, can contain high levels of salts, sugars, tannins, pigments, and

^aNatural Products Support Group, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702-1201, USA

^bMolecular Targets Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland, 21702-1201, USA. E-mail: okefeba@mail.nih.gov

^cNatural Products Branch, Developmental Therapeutic Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Frederick, Maryland 21702-1201, USA. E-mail: tanja.grkovic@nih.gov



lipids that can interfere with assay endpoints and yield a certain percentage of false-positive readings. But perhaps the biggest hurdle is the speed and efficiency of isolating and structurally elucidating small molecules in NP extracts. Often, isolating the active principles requires several rounds of chromatographic separation guided by repeated screening on activity before pure compounds are obtained and their structures determined. This adds to the overall cost and completion timelines of NP discovery projects. In this review, we outline technological advancements in the separation, characterization, and annotation of small molecules in NP mixtures that have enabled, among other things, their enthusiastic uptake as a resource for diverse scaffolds and structures in HTS, and we attempt to answer whether structural determination of NPs can match the speed and scale of HTS.



Christopher C. Thornburg

Christopher Thornburg earned his Doctorate in Pharmaceutical Sciences from Oregon State University, where, under the mentorship of Dr Kerry McPhail, he focused on isolating and determining the structures of biologically active marine natural products. In 2013, he became a member of the Natural Products Support Group at the Frederick National Laboratory for Cancer Research in Frederick, Maryland. By 2021, he had

advanced to the position of Senior Scientist and now directs the microbial and chemistry sections. His research focuses on integrating natural product extracts and fractions into high-throughput screening workflows for hit discovery and lead identification.



Rohitesh Kumar

Rohitesh Kumar earned his Doctorate in Chemistry from Griffith University, Australia under the supervision of Associate Prof. Rohan Davis where his doctoral research focused on the design and synthesis of libraries based on natural product scaffolds. He previously had worked at Leo pharma in Australia followed by post-doctoral positions at the University of Chemistry and Technology, Prague and the

Technical University of Denmark. He is currently a Scientist with the Natural Products Support Group at the Frederick National Laboratory for Cancer Research in Frederick, MD, a position he has held since October 2021. His research centers on the extraction and characterization of secondary metabolites from plants, marine, and fungal sources to support drug discovery.

2. Separation approaches

While most legacy NP libraries and repositories were comprised of collections of dried plant materials and frozen specimens or crude extracts, advances in chromatography automation have led to a notable increase in prefractionated NP libraries for use in drug discovery.^{2,5} A PubMed search for “NP library fraction” shows the first broad use of the term (>10 hits) in the mid-2000s, with no notable increase over the following two decades (Fig. 1A). Cross-referencing the term “HTS” shows a parallel trend, albeit with fewer counts (Fig. 1B). Most large NP libraries (defined here as >100 000 samples) use solid-phase extraction, reversed-phase high-performance liquid chromatography (RP-HPLC), or a combination of both, to separate crude extracts.^{6–8} Notably, the majority of these large libraries were generated using standardized chromatography gradients, timed HPLC collections, and high-capacity liquid-handling instrumentation. Method development and follow-up screening on prefractionated NP libraries have demonstrated successful sequestration of nuisance compounds, concentration of minor metabolites, reduction of activity to a single fraction, and faster active principle isolation timelines, underscoring the value of prefractionated NP libraries in HTS discovery campaigns.^{2,6–10}

The development of effective, automated chromatographic separation technologies has also enabled a faster identification of the individual components within NP fraction libraries. The term “dereplication,” referring to the annotation and identification of known compounds, was initially developed for NP prefractionated library collections or single discovery campaigns,^{11–13} and is now commonly used for large collections of both prefractionated NP libraries and crude extracts (Fig. 1C). The practice involves acquiring high-resolution spectrometric and spectroscopic analytical data on a semi-purified or crude NP mixture and using spectral databases and artificial intelligence (AI) tools to rapidly identify compounds present in the



Dongdong Wang

Dongdong Wang obtained his PhD in Medicinal Chemistry from Griffith University under the supervision of Professor Ronald J. Quinn, where he studied the isolation and structure elucidation of biologically active natural products. He then carried out his postdoctoral research at the Molecular Targets Program at the National Cancer Institute in Maryland, U.S.A., where he earned promotion to a Research Fellow position.

Afterwards, he joined the Natural Products Support Group at the Frederick National Laboratory for Cancer Research in Frederick, MD, in November 2025, where he is currently an NMR research scientist focusing on the structural elucidation of small molecules by using modern NMR techniques, as well as natural products drug discovery from marine, plant, and microbial sources.



sample. Initially, dereplication was used simply to weed out known structures in an effort to find new chemical entities, but owing to the advancements in analytical chemistry methodologies, it has become a much more informative technique in the NP chemist's toolbox that enables a more holistic view of the chemical space around known molecular scaffolds.

3. Mass spectrometry techniques

The analytical requirements for identifying specific compounds from hits in HTS of NP libraries—such as extracting actionable structural information from chemically complex, low-abundance mixtures within timelines that are compatible with screening—highlight the significant progress of mass spectrometry (MS) as a central analytical platform for high-throughput (HT) NP discovery.^{14,15} In current discovery workflows, MS is routinely employed to dereplicate known chemistry at early stages, prioritize novelty or actionable compound classes, identify analogue series for structure–activity relationships, and triage samples for further investigation through isolation and structural elucidation of active NPs by nuclear magnetic resonance (NMR) and/or other orthogonal techniques.^{16–18}

One of the continuing challenges in modern NP discovery is not merely HT MS data acquisition, but rather the development of integrated workflows for separation, ionization, acquisition, and informatics that support the dereplication of NP hits at HTS speed and scale.^{15,19} This section thus focuses on some of the current MS platforms and workflows for high-throughput natural product discovery (HT-NPD), whereas more comprehensive discussions of recent methodological advancements are addressed elsewhere.^{18,20–22}

3.1 Ionization sources enabling rapid NP analysis

Electrospray ionization (ESI) is the primary ionization source used in NP discovery workflows. ESI offers high sensitivity, broad metabolite coverage, and compatibility with aqueous and organic mobile phases, making it suitable for online chromatographic separation of analytes prior to MS analysis.^{23,24} Moreover, due to its soft ionization technique, ESI produces stable molecular ions and supports adduct chemistry, which aids in accurate precursor assignment and provides high-quality MS/MS spectra essential for library matching, molecular networking, and *in silico* structural prediction.^{25,26} In HT applications, ESI is combined with rapid chromatographic gradients to balance sample throughput with chemical resolution.¹⁴

Alternatively, atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI) provide complementary coverage to ESI, particularly for less polar or hydrophobic NPs, such as terpenoids, steroids, and lipophilic alkaloids.²⁷ Their higher tolerance to salts and excipients presents advantages for partially purified fractions and crude extracts prepared in microtiter plate formats.^{28,29} Multi-mode sources capable of switching between ESI and APCI/APPI without hardware modifications are increasingly implemented in HT laboratories to expand chemical space coverage while preserving throughput.²⁹

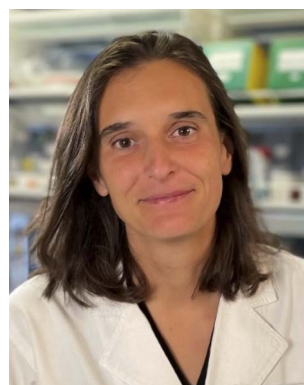
Ambient ionization techniques such as desorption electrospray ionization (DESI) and matrix-assisted laser desorption/ionization (MALDI) are used in some HTS formats because they enable direct analysis of solid and liquid samples, skipping chromatographic separation entirely.^{15,21} MALDI is especially attractive for colony-based NP screening, pathway engineering, directed evolution, and plate-based enzymatic assays. Additionally, imaging-based MALDI-MS enables spatially resolved analysis of microbial interactions and biosynthetic



Barry R. O'Keefe

Barry O'Keefe earned a B.S in Botany from Michigan State University and a PhD in Pharmacognosy from the University of Illinois at Chicago. Dr O'Keefe immediately joined the National Cancer Institute's Laboratory of Drug Discovery Research and Development to study novel proteins from natural products extracts; studies he continues as Head of the Protein Chemistry and Molecular Biology Section in the Molecular Targets Program of

the NCI Center for Cancer Research. In 2015, was appointed Chief of the Natural Products Branch in the Developmental Therapeutics Program of the NCI. He became director of the Molecular Targets Program in 2020. Dr O'Keefe's research is directed at the creation of novel biochemical and biophysical assay systems for use in the discovery of novel natural products, including proteins and peptides, with activity against cancer and infectious disease.



Tanja Grkovic

Tanja Grkovic obtained her PhD degree from the University of Auckland, New Zealand, followed by postdoctoral appointments at the Molecular Targets Laboratory, National Cancer Institute, and the Eskitis Institute for Drug Discovery. She spent six years as a Senior Scientist in the Natural Products Support Group at the Frederick National Laboratory for Cancer Research before joining the Natural Products Branch at the

National Cancer Institute in 2020 as a Staff Scientist. Her research efforts include the use of prefractionated libraries in high-throughput screens, structure elucidation of complex natural products involving a range of spectroscopic and spectrometric methods, and development of metabolomic and bioinformatic tools for the dereplication of natural products in complex mixtures.



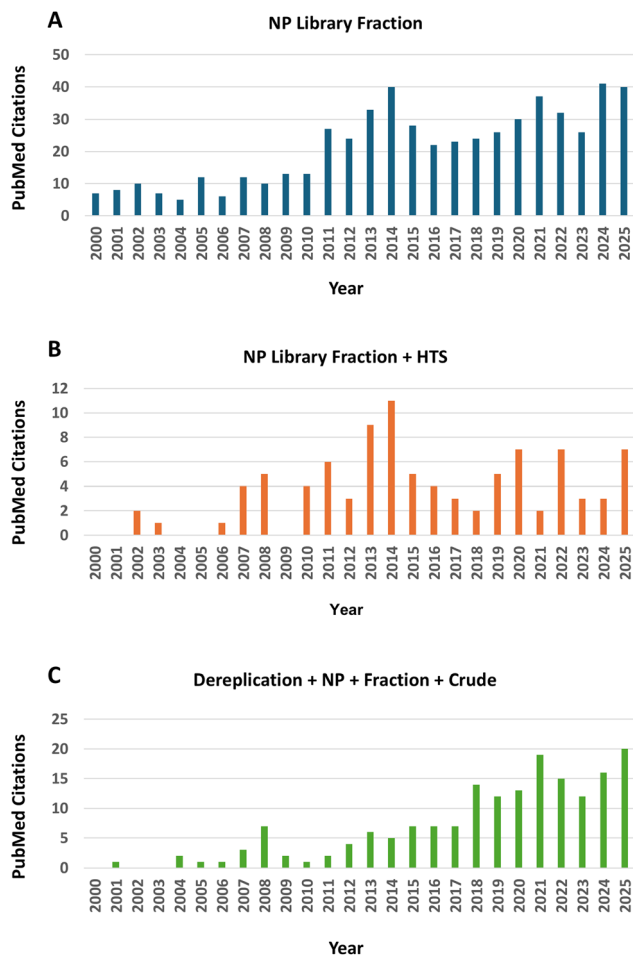


Fig. 1 A count of publications available on PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) from 2000–2025 using the terms (A) “natural product library fraction”, (B) “natural product library fraction” and “high throughput screening”, and (C) “dereplication”, “natural product”, “fraction” and “crude”.

phenotypes.^{30,31} Although recent advances in liquid-handling robotics have enabled partial automation of sample preparation, as well as the use of high-density sample arrays in HTS MALDI-MS workflows, resulting in acquisition rates of approximately 1–2 seconds per sample, sample preparation still remains a limiting factor, and matrix interference can suppress lower-molecular-weight ($m/z < 500$) analytes.^{21,30}

3.2 Mass analyzer platforms and their impact on throughput

HT-NPD fundamentally depends on high-resolution accurate-mass (HRAM) mass spectrometry platforms—primarily quadrupole time-of-flight (QTOF) and Orbitrap instruments—to facilitate elemental composition determination, spectral library matching, and computational annotation workflows.^{14,32}

QTOF instruments balance resolving power, mass accuracy, and acquisition speed, making them well-suited for HT-NPD workflows.³³ They support rapid full-scan and data-dependent MS/MS acquisition with duty cycles compatible with sub-minute chromatographic gradients.¹⁴ Their robustness and

spectral reproducibility have made QTOFs widely used workhorse instruments for untargeted metabolomics and NP dereliction, particularly when datasets are analyzed through platforms such as Global Natural Products Social (GNPS) molecular networking.³⁴

Orbitrap-based mass spectrometers offer enhanced resolving power and sub-ppm mass accuracy, facilitating precise elemental composition determination, isotopic fine-structure analysis, and *in silico* fragmentation-based annotation.³⁵ These functions are particularly advantageous in datasets that require automated molecular formula assignment, structural elucidation, or compound class prediction.³⁶ Recent advancements in scan speed and data-acquisition methodologies have substantially reduced the traditional trade-off between resolution and throughput, thereby enabling Orbitrap platforms to function effectively under short-gradient ultra-high performance liquid chromatography (UHPLC) conditions.³⁷

Ion mobility-mass spectrometry (IM-MS) is an emerging, powerful tool for natural product discovery, especially for complex plant and microbial extracts, where chemical diversity, chromatographic co-elution, and isomerism challenge traditional LC-MS readouts.^{38–40} By introducing a rapid, millisecond gas-phase separation based on ion size/shape (often reported as collision cross section, CCS) and charge, IM-MS adds a drift-time/CCS dimension to retention time and m/z , increasing the effective peak capacity (*i.e.*, the number of distinguishable features per run) without extending chromatographic gradients and enabling partial resolution of co-eluting, isobaric, and in some cases isomeric metabolites.^{41–43} The resulting CCS values provide orthogonal molecular descriptors that, together with accurate mass, retention time, and MS/MS fragmentation, can improve dereplication and minimize false annotations, and potentially reduce chimeric MS/MS spectra in data-dependent acquisition *via* drift-time separation of overlapping precursor ions.^{44–46} These attributes make IM-MS particularly well-suited for NP extracts rich in terpenoids, polyketides, alkaloids, *etc.*, where isomerism is common.³⁸ However, NP-focused CCS libraries remain incomplete, and broader adoption, especially in HTS, is limited by factors such as difficulty in distinguishing subtle isomers, platform- and calibration-dependent CCS variability, increased data complexity, and potential sensitivity issues.^{47–49} Thus, IM-MS currently functions best as a complementary method within high-throughput LC-MS/MS workflows rather than as a standalone solution.

Using HRAM with multiple ionization techniques and/or analyzing both positive and negative ions can increase confidence in the structural identification of analytes in active fractions.³² Furthermore, chromatographic-based hyphenated approaches can considerably accelerate structural elucidation without requiring large-scale isolation of analytes from complex matrices.⁵⁰ Compared with traditional HPLC, UHPLC offers significant advantages, especially in terms of speed, resolution, and sensitivity, and remains the backbone of HT-NPD analytics.⁵¹ Because UHPLC columns use smaller-particle stationary phases (<2 μm), sharper peaks and improved separation of closely related compounds can be achieved across shorter gradients.⁵² Thus, a well-equipped HT-NP laboratory can



employ sub-5-min gradients, plate-based injections, and automated QC strategies to process hundreds of samples per day and, with parallel instrumentation, thousands per day. Additional hyphenated platforms, such as GC-MS, SFC-MS, CE-MS, *etc.*, are particularly effective for volatile and semi-volatile NPs and hydrophobic metabolites and are essential for definitive structural elucidation beyond MS analysis. These platforms, as well as other hyphenated techniques, are more extensively reviewed elsewhere, as these techniques require specialized equipment and are best reserved for very specific screens and assays.^{17,50,53,54}

4. Nuclear magnetic resonance techniques

NMR spectroscopy is one of the most comprehensive methods available for studying the molecular structures of NPs. The ability to examine all paramagnetic nuclei in a molecule and establish their inter-atomic bonding patterns, coupled with its non-destructive nature allowing multiple experiments with minimal material, makes NMR particularly suitable for determining the structures and configurations of even the most complex natural products.^{55–59} The importance of NMR in NP research is further evidenced by the fact that, at present, the majority of journals publishing the structures of new NPs require NMR spectra to be supplied as images in SI, and some even require the deposition of the associated raw files.⁶⁰ Below, we outline some technological advances in the last 25 years that have brought significant improvements in both the sensitivity and resolution of NMR, and the use of this versatile technique for the structural elucidation of NPs, rapid dereplication of known structures, and HT-NPD chemical analysis.

4.1 NMR magnet and probe technology

NMR sensitivity and resolution are dependent on magnetic field strength and probe design. Since the first commercial NMR spectrometers became available in the late 1960s and early 1970s, operating at 20–200 MHz for the ¹H nucleus,^{61,62} continuing developments in superconducting field technology

have led to the recent commercial availability of NMR spectrometers operating at frequencies of >1 GHz. To gauge the NMR magnetic field strength available to a contemporary natural product chemist, we annotated the ¹H NMR field frequencies reported in publications from three issues of the *Journal of Natural Products* in 2025, 2015, and 2005 (Fig. 2). In 2025, the overwhelming majority of laboratories used high-field NMR instruments operating at 500 MHz or higher. Analysing the data for 2015 and 2005, the trend shows a gradual but consistent increase in the use of high-field NMR magnets today compared to 10 and 20 years ago. The ability to record NMR data at high-field strength affords not only enhanced sensitivity but also resolution, which can show composite signals of structurally complex molecules that would be overlapped at lower field strengths, thereby improving the speed and accuracy of dereplication, discovery of novel secondary metabolites, and the acquisition of experimental proofs necessary for structure elucidation.

Rapid advances in NMR probe technology are equally impressive. In the early days of NMR spectroscopy, standard 5 mm NMR probes were used. Nalorac 3 mm probes became available in the early 1990s,⁶³ followed by the 1.7 mm probe a few years later⁶⁴ and the 1 mm probe in 2003. For mass-limited natural product samples, smaller-diameter probes that require less volume of deuterated solvent provide an increase in sensitivity. But perhaps one of the most significant technological advances was the development of cryogenically cooled probes, which became commercially available in the late 1990s. The use of helium- and nitrogen-cooled probes, generally operated at very low temperatures—in the range of 25 K to 35 K for helium and 80 K to 83 K for nitrogen—revolutionized the capabilities of NMR, accelerated the acquisition speed of spectra, and enabled structural elucidation work with nmol or less of material.^{65,66}

With the improvements in magnetic field strengths, probe design, and console hardware of NMR spectrometers, the overall sensitivity of NMR experiments has increased by over three orders of magnitude over the past 60 years,⁶⁷ making NMR a routinely applied fast technique for obtaining critical data on chemical structures in HTS of NP research.

4.2 Select pulse sequences and NMR experiments

A modern NMR spectrometer offers over 1500 different pulse sequences. Their widespread use and availability, especially for 2D heteronuclear NMR experiments, have had a major impact on the ability to determine the structures of NPs. Here, we highlight only a few that we believe have significantly accelerated data acquisition for natural product samples and have the potential to be used in HT chemical analysis.

One of the most versatile and useful 2D heteronuclear NMR experiments used in the dereplication of natural products is the multiplicity-edited HSQC,⁶⁸ a proton-detected 2D heteronuclear NMR experiment that offers much higher sensitivity than DEPT and is a powerful, time-efficient, and routinely used tool for NP structural elucidation. When used with pattern-recognition AI tools such as Small Molecule Accurate Recognition Technology

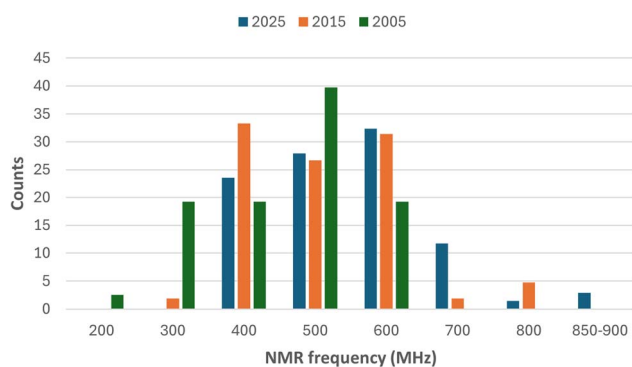


Fig. 2 NMR frequencies used for the structural elucidation of new natural products reported for issues 10, 11, and 12 in 2025, 2015, and 2005, in the *Journal of Natural Products*. Data are reported as a percentage of total counts.



(SMART)⁶⁹ and DeepSAT,⁷⁰ it provides fast and reliable compound dereplication. Moreover, its inherent sensitivity allows for short experimental times, adding to its applicability in HT-NPD chemical analysis.

Structural elucidation by NMR works best on pure or semi-purified samples. Analysis of mixtures presents challenges due to spectral complexity and signal assignment. Pulsed-field gradients can lower the detection limits by reducing artifacts and are key to diffusion-ordered nuclear magnetic resonance spectroscopy (DOSY).^{71,72} Because resonances from different compounds are distinguished by their diffusion coefficients (D), the DOSY experiment has been applied to investigate D values of structurally diverse mixtures of NPs, dereplicate known NPs, and identify new metabolites solely on parameters derived by DOSY NMR.⁷¹ The advantages of the DOSY experiment include inherent sensitivity (¹H detected experiment) and the ability to separate signals without the need for sample purification.

4.3 Non-uniform sampling

First proposed more than 40 years ago, non-uniform sampling (NUS) has been rapidly adopted and widely investigated for acquiring and processing NMR data.⁷³ In contrast to standard uniform sampling, in which time-domain NMR signals undergo a discrete Fourier transform to produce a frequency spectrum, exponential NUS retains about 25–50% of the data points that would be acquired uniformly. The use of NUS has enabled faster acquisition of high-quality NMR spectra with increased sensitivity/resolution, and has offered significant time advantages in the structure determination of small-molecule NPs.^{74,75} NUS-based enhancements for very dilute samples of complex small molecules can often be achieved in 2D NMR experiments, such as HSQC and HMBC, by lowering the detection limit and improving sensitivity by up to twofold in a given dimension,⁶⁷ enabling a broader range of experiments to be used, even with smaller sample quantities.

5. Other techniques

5.1 Infrared spectroscopy

Infrared (IR) spectroscopy measures how a compound interacts with infrared light at different wavelengths. Historically, IR spectroscopy relied on manual sample preparation on a KBr

disc, which was time-consuming and destructive to samples, limiting its routine application in HT NP chemistry workflows. The introduction of Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR) has largely overcome these limitations,⁷⁶ offering non-destructive analysis and ease of use. With the advent of modern tools such as NMR and MS, IR spectroscopy is often underestimated and underutilized, despite providing complementary analytical data. Today, FTIR spectroscopy is a cost-effective, non-destructive technique that requires minimal to no sample preparation and offers rapid data acquisition across diverse sample types, making it well-suited as an initial screening platform.^{76–78} Moreover, FTIR has one of the broadest spectral ranges among analytical techniques used for the structural elucidation of NPs, making it particularly suitable for pattern matching analysis and library-based dereplication. However, despite these advantages, FTIR is still not commonly used for NP dereplication. This is largely due to a lack of publicly available IR databases⁷⁹ that would enable rapid automated compound identification; “non-discriminatory” spectroscopic information, especially due to overlapping absorptions in similar regions;⁸⁰ and IR’s inability to provide stereochemical information, which is crucial in NP chemistry.

Nevertheless, recent work has demonstrated the growing potential of automated data analysis using IR spectra. Approaches such as spectral prediction⁸¹ and functional group classification using neural networks⁷⁹ and machine learning⁸² are rapidly evolving. For example, a transformer-based model was developed using a large experimental and simulated dataset, which was capable of predicting molecular structures directly from experimental IR data.⁷⁸ Another study has demonstrated the use of a convolutional neural network trained and validated on a large subset of IR spectra to automatically identify a wide range of functional groups without expert input.⁸³ Modern FTIR has been designed around standard format HT IR plate readers capable of analysing microgram quantities of samples on microtiter plates that can be prepared using liquid-handling robotic platforms and recorded at high throughput (Fig. 3). This was demonstrated in a recent HT IR method for peptide quantification,⁸⁴ as well as NP dereplication efforts.⁸⁵ These developments position IR as a HT tool for the future structural elucidation of NPs, complementing NMR and MS.

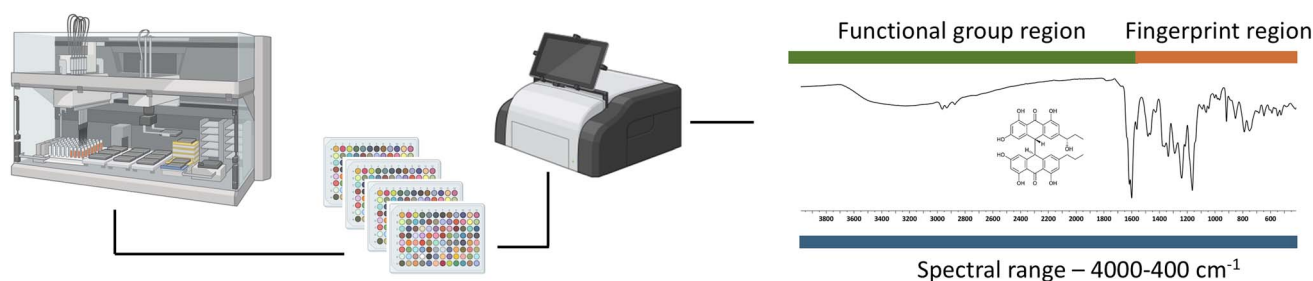


Fig. 3 Summary of modern FTIR sample preparation, acquisition and data analysis. FTIR spectrum of 1''-hydroxycrinemodin bianthrone was adapted with permission from Freire *et al.*⁸⁶ Copyright 2024, Royal Society of Chemistry. Instrument and plate images were created in BioRender (2026): <https://BioRender.com/fvp1in3>.



5.2 Chiroptical spectroscopic techniques

The use of chiroptical spectroscopic techniques such as Electronic Circular Dichroism (ECD), Vibrational Circular Dichroism (VCD), and Optical Rotation (OR) can be vital for determining the absolute configuration of chiral molecules after the planar structure has been determined.⁸⁷ These techniques measure the interaction of chiral molecules with circularly polarized light. The differential interaction of chiral molecules with light is directly related to molecular handedness, providing important information about three-dimensional arrangement, enantiomeric purity, and stereochemical information. ECD and OR are based on electronic transitions, while VCD is based on vibrational transitions of chiral molecules.⁸⁸ The complete structural assignment and absolute configuration of chiral compounds are essential, particularly in the context of biological activity, as stereochemistry can profoundly influence molecular recognition, potency, and selectivity.⁸⁹ Being inherently non-destructive, chiroptical spectroscopic methods are widely used by NP chemists for absolute configuration determination of small molecules, as well as for small peptides and proteins.⁸⁸ These techniques are particularly advantageous in NP research, where the availability of isolated compounds is often limited, making sample preservation critical. The application of chiroptical techniques in the NP field has been extensively discussed in several recent reviews.^{90,91} ECD requires UV-vis chromophores⁹² and only small sample quantities, allowing relatively short acquisition times, whereas VCD can analyze UV-silent molecules⁹³ but typically requires larger sample amounts and longer run times. Although chiroptical methods are not commonly used in HT dereplication work, this field is rapidly evolving with the inclusion of deep-learning methods⁹⁴ and time-dependent density functional theory calculations to generate theoretical chiroptical spectra (ECD, VCD,^{95,96} and OR calculations^{87,96}) with high accuracy.⁸⁷ These advances are enabling faster, more reliable stereochemical assignments, as well as the integration of FTIR data—and potentially chiroptical data in the future—into automated screening pipelines. When combined with miniaturized sample handling and HT instrumentation, such developments are poised to render chiroptical and FTIR methodologies increasingly practical tools for rapid molecular characterization, thereby complementing traditional spectroscopic dereplication techniques.

6. Database tools

At HT scale, annotation throughput often limits progress more than data acquisition.⁹⁷ Early dereplication efforts relied on limited in-house libraries or commercial resources such as the Dictionary of Natural Products,⁹⁸ MarinLit,⁹⁹ and AntiBase,¹⁰⁰ often resulting in incomplete coverage of NP chemical space.⁹⁷ In recent years, the emergence of large, community-driven NP databases has fundamentally changed this landscape. Platforms such as the GNPS molecular networking enable large-scale comparison of MS/MS spectra using molecular

networking and spectral library matching, facilitating rapid recognition of known scaffolds and related analogues.^{34,101,102}

6.1 Mass spectral databases

Beyond simply asking “what is this spectrum?”, NP dereplication increasingly focuses on “where else does it occur?” Spectral search engines, such as the GNPS-style Mass Spectrometry Search Tool (MASST),¹⁰³ along with domain-specific versions such as plantMASST and microbeMASST,^{104,105} allow users to search public repositories for a single MS/MS spectrum. This helps link unknown compounds to ecological, taxonomic, or exposure-related contexts.^{105,106} Such contextual information can significantly influence project prioritization, for example, by distinguishing strain-specific features from ubiquitous media contaminants.¹⁰⁶ When exact library hits are absent, tools that infer substructures, motifs, or chemical classes from MS/MS data can guide dereplication and isolation processes.^{107,108} Machine learning-based analogue search methods, such as MS2Query, aim to retrieve structurally similar candidates even when the exact compound is absent from existing libraries, thereby expanding MS/MS similarity beyond strict library matching (Table 1).¹⁰⁹

In parallel, structure-centric databases, such as the Natural Products Atlas,^{120,121} COCONUT,¹¹⁸ and LOTUS¹¹⁹ have expanded dramatically in size and scope (Table 1), collectively covering hundreds of thousands of NP structures with associated taxonomic, bibliographic, and, in certain instances, biological metadata. These resources facilitate not only direct dereplication but also higher-order analyses of NP chemical space, thereby enabling prioritization based on novelty, biosynthetic origin, or structural classification.¹⁶

MS-based workflows now provide the speed and information density necessary for modern HT-NPD, enabling rapid dereplication and prioritization at unprecedented scale.^{14,30} When integrated with advanced MS data-processing tools—such as feature detection, alignment, molecular networking, and *in silico* fragmentation—spectral databases enable automated annotation or triage of thousands of features from a single experimental run.^{16,34,151} Table 1 illustrates the increasingly complex ecosystem of databases and computational resources available for matching MS data to putative structures. While a detailed comparison of these tools is beyond the scope of this review, the integration of several of them into GNPS molecular networking workflows has been described by Aron *et al.*¹⁰² In addition, a growing number of dedicated reviews reflect the need to navigate this crowded landscape and provide guidance on their practical use,^{16,36,152} including those evaluated in the CASMI (Critical Assessment of Small Molecule Identification) contest.^{153,154} Overall, this represents a significant shift away from traditionally manual, late-stage dereplication paradigms.⁹⁷ However, ongoing challenges—including stereochemical ambiguity, isomer discrimination, and overall data complexity—limit the interpretability of MS data in isolation.¹⁴ Therefore, MS is most effectively employed within a tiered discovery strategy, where rapid MS-based triage is followed by targeted isolation and orthogonal structural confirmation of the



Table 1 Mass spectrometry-based databases and informatics resources for natural products dereplication

Type	Representative resources ^a	Primary function in dereplication
Spectral libraries (experimental MS/MS)	GNPS, ³⁴ MassBank, ¹¹⁰ MoNA, ¹¹¹ METLIN, ^{112,113} Wiley/NIST, ¹¹⁴ RIKEN/ReSpect, ¹¹⁵ mzCloud, ¹¹⁶ MS ³ Lib ¹¹⁷	Direct identification of known compounds through experimental MS/MS spectral matching
Structure databases (general & domain-specific)	COCONUT, ¹¹⁸ LOTUS, ¹¹⁹ NPAtlas, ^{120,121} molecules Gateway, ¹²² DNP, ⁹⁸ MarinLit, ⁹⁹ AntiBase, ¹⁰⁰ ChemSpider, ¹²³ PubChem ¹²⁴	Reference collections of reported natural product structures with literature and source annotations
<i>In silico</i> fragmentation & annotation tools	SIRIUS/CSI: FingerID, ¹²⁵ MS-FINDER, ¹²⁶ MetFrag, ¹²⁷ MetFusion, ¹²⁸ CFM-ID, ¹²⁹ ISDB, ¹³⁰ MS2Query, ¹⁰⁹ DEREPI-NT	Prediction of molecular formulas and ranking of plausible structural candidates from MS/MS data
<i>In silico</i> structure-based dereplication tools	DEREPLICATOR+ ¹³¹ MolDiscovery ¹³²	Dereplication of MS/MS spectra to predicted spectra or fragments from known natural products
Molecular networking & substructure analysis frameworks	GNPS molecular networking, ^{26,101,102} MS2LDA, ^{107,133} MolNetEnhancer, ¹³⁴ NAP ¹³⁵	Organization of MS/MS data into molecular families and substructures to support family-level dereplication
IM-MS/CCS resources	Unified CCS Compendium, ¹³⁶ METLIN-CCS, ⁴⁷ PNNL CCS, ¹³⁷ CCSbase, ¹³⁸ AllCCS ^{49,139}	Use of collision cross section values to support isomer discrimination and annotation confidence
Global MS/MS search tools	MASST (GNPS), ¹⁰³ ReDU ¹⁰⁶	Identification of where and how widely an MS/MS feature has been observed across public datasets
Bioactivity & source annotation	NPASS, ¹⁴⁰ NPBS Atlas, ¹⁴¹ ChEMBL, ¹⁴² PubChem BioAssays ¹⁴³	Literature- and assay-linked biological activity context
Raw MS data repositories	MassIVE, ¹⁴⁴ MetaboLights, ¹⁴⁵ metabolomics workbench ¹⁴⁶	Long-term deposition of raw LC-MS/MS data

^a Raw LC-MS/MS data are typically processed using vendor-specific or open-source feature extraction and alignment software such as MZmine,¹⁴⁷ XCMS,¹⁴⁸ MS-DIAL,¹⁴⁹ or OpenMS¹⁵⁰ prior to dereplication and annotation. Abbreviations. GNPS: global natural products social molecular networking; MoNA: MassBank of North America; METLIN: METAbolite LINK; NIST: national institute of standards and technology; ReSpect: RIKEN MSn Spectral database for phytochemicals; COCONUT: the COLlection of Open NatUral productTs; LOTUS: LOTUS initiative for open natural products research; NPAtlas: natural products atlas; DNP: dictionary of natural products; SIRIUS: sum formula identification by ranking isotope patterns using mass spectrometry; MetFrag: metabolite fragmenter; CFM-ID: competitive fragmentation modeling for metabolite identification; ISDB: *in silico* spectral dataBase for natural products; MS2LDA: mass spectrometry-based 2-dimensional latent dirichlet allocation; NAP: network annotation propagation; CCS: collisional cross-section; PNNL: pacific northwest national laboratory; MASST: mass spectrometry search tool; ReDU: reanalysis of data user interface; NPASS: natural product activity and species source; NPBS Atlas: natural product and biological source atlas; ChEMBL: chemogenomic database maintained by the European molecular biology laboratory-European bioinformatics institute. MassIVE: mass spectrometry interactive virtual environment.

most promising candidates using additional analytical methods.^{14,15}

6.2 NMR spectral databases

One- and two-dimensional NMR data, in the form of chemical shift, multiplicity, and integration, offer a multitude of information for spectral searches and structure identification. NMR-based dereplication on NP fractions and fraction libraries has been shown to be successful on a small scale (10 s of micrograms of a sample) and for a range of NP chemotypes.^{12,155–157}

When used as an orthogonal tool to other analytical methods, such as MS and FTIR, it can increase the confidence of assignments and annotation of hits.^{9,85} At present, there is a growing number of NP NMR databases and repositories, but not a corresponding number of spectral search and AI tools to complement the increase in the available data access (Table 2). Attempts to centralise the NMR data of all published NPs in a single database, such as NP-MRD¹⁵⁸ and NAPROC-13,¹⁵⁹ have the potential to simplify spectral searches and spur the development of new tools. Likewise, access to fast, accessible search tools, such as SMART¹⁶⁰ and DeepSAT,⁷⁰ that can batch process

spectral datasets simultaneously can effectively facilitate HT-NPD chemical searches.

Notably, while most dereplication MS- and NMR-based tools presented here excel at the identification of known natural product structures, or even predict tentative structures of related compounds, they are not capable of identifying unknown NP structures and chemotypes precisely. If the goal is to identify all, including the unknowns, and ultimately be able to compare the speed of NP-based HTS to that of screening pure compound libraries, multiple orthogonal approaches and integrated spectral libraries should be considered.

7. Meeting the speed and scale of HTS

The last 25 years have seen significant advances in analytical chemistry instrumentation, including increased sensitivity, a smaller footprint, and greater affordability. Structural elucidation is now possible, if not commonplace, to do on 10s of micrograms^{66,168,169} of a small molecule. This, coupled with the availability of spectral databases and AI tools, has enabled structural elucidation of natural products to function at an unprecedented pace.



Table 2 NMR-based databases and informatics resources for natural products dereplication

Type	Representative resources
NP spectral libraries of raw FID NMR data	NP-MRD, ¹⁵⁸ NAPROC-13 ¹⁵⁹
Other spectral raw FID NMR data libraries ^a	nmrXiv, ¹⁶¹ Harvard dataverse ¹⁶²
NP NMR spectral libraries extracted from publications	Jeol CH-NMR-NP ¹⁶³
Other NMR spectral libraries extracted from publications ^b	NMRexp, ¹⁶⁴ NMRExtractor, ¹⁶⁵ Micronmr ¹⁶⁶
<i>In silico</i> annotation & structural elucidation tools	SMART, ¹⁶⁰ DeepSAT, ⁷⁰ DP4-AI ¹⁶⁷

^a Known to contain raw fid spectra of NPs. ^b Selected based on criteria that some contain spectra of NPs. Abbreviations. FID: free induction decay; NP-MRD: natural products magnetic resonance database; NAPROC-13: NATURAL PROducts Carbon-13 NMR database; nmrXiv: nuclear magnetic resonance eXchange; Jeol CH-NMR-NP: ¹³C/¹H-NMR database for natural products; NMRexp: NMR experimental spectra database; NMRExtractor: NMR data extraction tool utilizing large language models; SMART: small molecule accurate recognition technology.

The speed and ease of data acquisition are critical to enabling these productivity advances. Fig. 4 compares the acquisition aspects of LC-MS, NMR, and FTIR spectroscopy with respect to parameters we consider most relevant to achieving the speed and scale of HTS. Many vendors of LC-MS, NMR, and FTIR equipment offer high-capacity autosamplers in 96- and 384-well plate formats, along with the automated sample acquisition software. Moreover, with the universal use of “SBS footprint” laboratory microplates, liquid handling robots can be used for automated sample preparation. While the equipment formats outlined above may meet HTS's scale requirements, they will never meet its speed requirements. The sensitivity of modern instruments has significantly shortened experimental acquisition times – but most LC-MS and NMR experiments still require 5–10 min of runtime per sample. At that pace, it would take up to 16 hours for a single 96-well plate, and years of uninterrupted experimental time to complete a library of >100 000 fractions.¹⁷⁰ The ability to process such data, especially in high-resolution LC-MS/MS, where a single run file can exceed one gigabyte, adds another timeline hurdle.

However, in an individual HTS discovery campaign, not all fractions screened need to be annotated. Focusing the value

proposition exclusively on fractions that show activity in an HTS will significantly reduce the analytical burden and shorten timelines towards compound identification. In one case study below, we summarise one such example to demonstrate that NP HTS discovery campaigns can operate at the scales and timelines of HTS screening.

7.1 Case studies for the HT chemical analysis of NP hits from HTS

The National Cancer Institute (NCI) had launched the NCI Program for Natural Product Discovery (NPNPD) to help reinvigorate drug discovery research in NPs.⁸ As an initial aim, the NPNPD has been producing and releasing prefractionated NP samples to the public for screening across all disease states. At present, the NPNPD library comprises over 700 000 partially purified NP fractions sourced from over 100 000 plant, marine, and fungal samples. To complement this large NP library, a rapid isolation and identification method for biologically active NPs has also been developed.⁹ A recently published screen⁸⁵ of 326 656 NPNPD fractions against four microbial pathogens illustrates a workflow for high-throughput analysis of HTS hits (Fig. 5). Out of 3067 confirmed hits-based on dose-

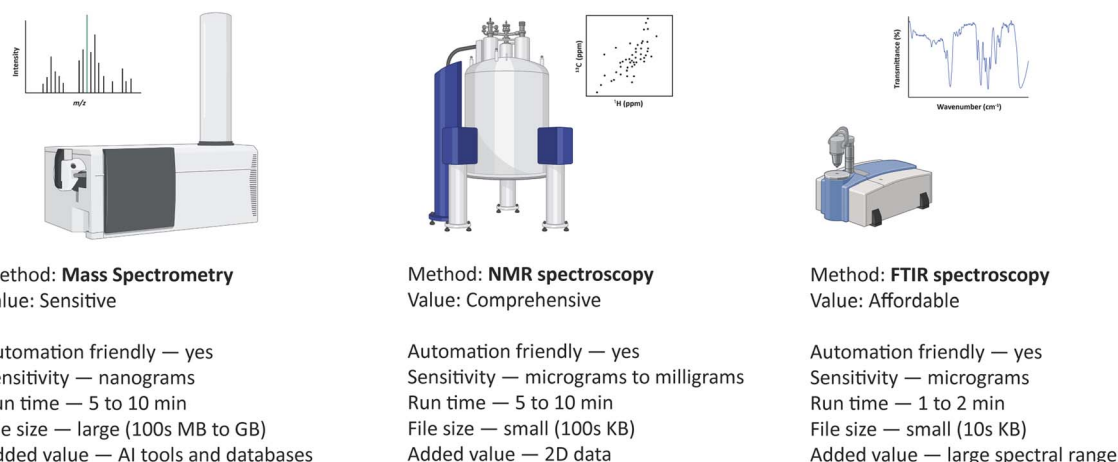


Fig. 4 Comparison of commonly used analytical methods used for the structural elucidation of NPs and their applicability to high-throughput data acquisition and processing. Run times indicated were estimated based on the following, MS: 3–8 min LC gradient with MS acquisition allowing for a column flush and equilibration post and between runs respectively; NMR: ¹H NMR experiment at 128 scans using standard zg30 Bruker pulse sequence and allowing for sample optimization such as shimming and tune and match; FTIR: 128 scans using standard Bruker transmittance experimental parameters. File sizes indicated reflect, NMR: ¹H, COSY and ¹H–¹³C HSQC data; MS: MS and MS/MS data acquisition with 1–5 ppm accuracy; FTIR: transmittance data. Created in BioRender (2026): <https://app.biorender.com/>.



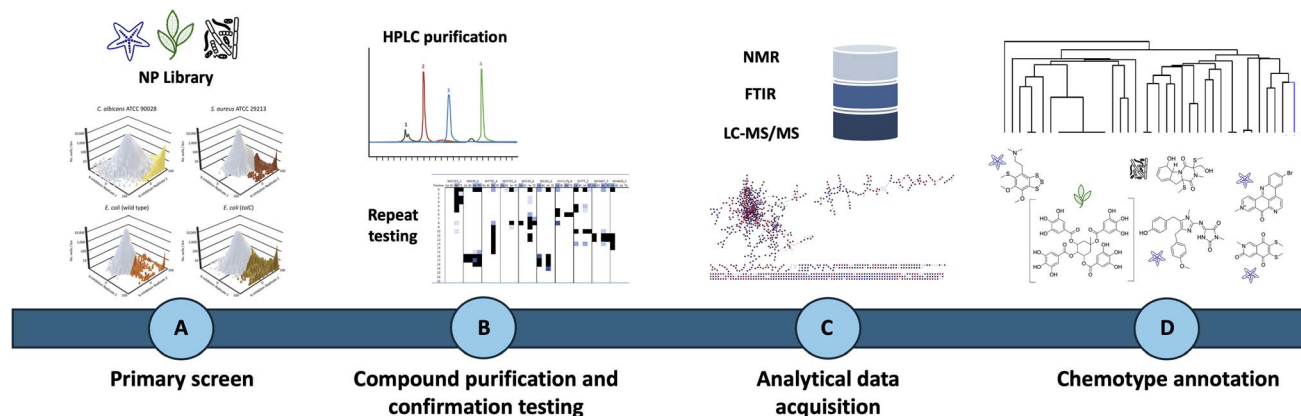


Fig. 5 An example of HT chemical analysis to support HTS screening workflows. (A) NP library primary screen results against four microbial targets; (B) second-stage HPLC purification and repeat testing; (C) HT analytical data acquisition; (D) compound identification results. The NP library and assay results are reprinted with permission from Martinez *et al.*⁸⁵ Copyright 2023, American Chemical Society.

response profile data, literature, and taxonomy information, a set of 75 fractions was prioritised for purification and active principle annotation work. Following established automation procedures,⁹ one milligram of each active fraction was subdivided into 22 subfractions (nominally 45 μg per well) *via* semi-preparative HPLC, yielding 1650 assay-ready subfractions for follow-up testing within 18 hours. From this set, approximately 250 sub-fractions were identified as active and analyzed by HR LC-MS/MS, NMR, and FTIR to gather initial chemical data, requiring approximately seven days in total (2 days for MS and NMR and 1 day for FTIR). At this stage, active subfractions were typically found to contain a single, or nearly pure, compound in approximately 80% of cases. Thus, the use of a HT, reproducible second-stage HPLC process reduced the complexity of initial hit fractions, enabling rapid chemical annotation, typically in less than a week for a set of this size. For larger hit lists of up to 500 samples, analysis can be completed within seven days by running LC-MS/MS and NMR in parallel after dividing the active sample, allowing an annotated list of identified active structural classes to be returned to the screening laboratory within two weeks. Overall, the workflow described here is one example of HT chemical analysis of screening hits at an HTS scale.

Other HT NP analyses examples, not limited to a starting point of biological activity screen, include pattern-based genome mining approaches as well as chemical reactivity-based screening. In pattern-based genome mining, genome sequences are analysed for genomic signatures of specific compound types or specific biosynthetic genes are targeted using polymerase chain reaction primers. Once the biosynthetic gene cluster of interest is identified, confirmation screening is supported by MS-based molecular networking analysis^{171–173} or for specific NP classes such as oxazoles, and piperazine acids using ^1H - ^{13}C or ^1H - ^{15}N HSQC NMR fingerprints.^{174–176} In reactivity-based screening an additional step is added where a specific functional group of interest is chemically modified and the resulting adducts detected MS metabolomics and used to guide the isolation of new NPs. This can be achieved either in tandem with genome mining, for example to find new diazo-¹⁷⁷

and azoxy-containing NPs,¹⁷⁸ on crude extracts to identify reactive carbonyl- and alkyne-containing^{179,180} NPs.

8. Conclusions and perspectives

In this review, we outlined how analytical chemistry instrumentation, such as LC-MS, NMR, and FTIR, has advanced significantly in sensitivity and automation. This, coupled with the availability of NP spectral databases and AI tools, has made dereplication and compound identification timelines more productive, focused, and ultimately much faster. As a community, we are no longer bound to having knowledge of NP structures at the very last step of lengthy purification processes and long analytical data acquisition timelines. The paradigm is shifting – the technology and tools are available for natural product science to run at the speed and scale of HTS. To best use the vast array of analytical chemistry data and machine learning resources available, NP chemists should expand the toolbox for NP dereplication by integrating analytical data streams and incorporating AI methods into structural elucidation workflows. A recent review showed that AI tools for structural elucidation of marine natural products are relatively underutilized,¹⁸¹ and most laboratories still rely on well-trained chemists to solve chemical structures by NMR. In the future, AI-generated spectral databases may soon outnumber experimentally obtained data repositories - for example Open Molecules 2025 (Om125) dataset contains 100 000 000 high-level DFT calculated quantum calculations for training AI models.¹⁸² But high-quality, well-annotated, and accessible experimental spectral data for natural product structures remain the benchmark for structural elucidation efforts. AlphaFold,¹⁸³ now capable of generating highly accurate 3-dimensional (3D) structure predictions of large biomolecules from amino acid sequences in mere minutes, was trained on the Protein Data Bank – a well-curated 3D structural database based on X-ray crystallography, electron microscopy, and NMR data. Currently, most natural product structure-predicting tools rely on a single analytical method, such as tandem mass spectra or 1- and 2-dimensional NMR data. Going forward, using multiple



data sources together with machine learning tools has the potential to increase the speed and accuracy of structure determination and annotation of natural products.

9. Author contributions

All authors collated references and jointly wrote the manuscript.

10. Conflicts of interest

The authors declare no conflicts of interest.

11. Data availability

No primary research results, software or code have been included, and no new data were generated or analyzed as part of this review.

12. Acknowledgements

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contracts 75N91019D00024 and HHSN261200800001E. The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. Some images in the TOC, Fig. 3 and 4 were created in BioRender under the following licenses: Thornburg, C. (2026) <https://BioRender.com/>, <https://app.biorender.com/>, and <https://app.biorender.com/>.

13. References

- D. A. Pereira and J. A. Williams, *Br. J. Pharmacol.*, 2007, **152**, 53–61.
- B. A. P. Wilson, C. C. Thornburg, C. J. Henrich, T. Grkovic and B. R. O'Keefe, *Nat. Prod. Rep.*, 2020, **37**, 893–918.
- J. T. Ernst, P. A. Thompson, C. Nilewski, P. A. Sprengeler, S. Sperry, G. Packard, T. Michels, A. Xiang, C. Tran, C. J. Wegerski, B. Eam, N. P. Young, S. Fish, J. Chen, H. Howard, J. Staunton, J. Molter, J. Clarine, A. Nevarez, G. G. Chiang, J. R. Appleman, K. R. Webster and S. H. Reich, *J. Med. Chem.*, 2020, **63**, 5879–5955.
- R. L. Mackman, V. A. Steadman, D. K. Dean, P. Jansa, K. G. Poullennec, T. Appleby, C. Austin, C. A. Blakemore, R. Cai, C. Cannizzaro, G. Chin, J.-Y. C. Chiva, N. A. Dunbar, H. Fliri, A. J. Highton, H. Hui, M. Ji, H. Jin, K. Karki, A. J. Keats, L. Lazarides, Y.-J. Lee, A. Licican, M. Mish, B. Murray, S. B. Pettit, P. Pyun, M. Sangi, R. Santos, J. Sanvoisin, U. Schmitz, A. Schrier, D. Siegel, D. Sperandio, G. Stepan, Y. Tian, G. M. Watt, H. Yang and B. E. Schultz, *J. Med. Chem.*, 2018, **61**, 9473–9499.
- A. L. Harvey, R. Edrada-Ebel and R. J. Quinn, *Nat. Rev. Drug Discovery*, 2015, **14**, 111–129.
- D. R. Appleton, A. D. Buss and M. S. Butler, *Chimia*, 2007, **61**, 327–331.
- D. Camp, R. A. Davis, M. Campitelli, J. Ebdon and R. J. Quinn, *J. Nat. Prod.*, 2012, **75**, 72–81.
- C. C. Thornburg, J. R. Britt, J. R. Evans, R. K. Akee, J. A. Whitt, S. K. Trinh, M. J. Harris, J. R. Thompson, T. L. Ewing, S. M. Shipley, P. G. Grothaus, D. J. Newman, J. P. Schneider, T. Grkovic and B. R. O'Keefe, *ACS Chem. Biol.*, 2018, **13**, 2484–2497.
- T. Grkovic, R. K. Akee, C. C. Thornburg, S. K. Trinh, J. R. Britt, M. J. Harris, J. R. Evans, U. Kang, S. Ensel, C. J. Henrich, K. R. Gustafson, J. P. Schneider and B. R. O'Keefe, *ACS Chem. Biol.*, 2020, **15**, 1104–1114.
- D. Camp, M. Campitelli, A. R. Carroll, R. A. Davis and R. J. Quinn, *Chem. Biodivers.*, 2013, **10**, 524–537.
- G. R. Eldridge, H. C. Vervoort, C. M. Lee, P. A. Cremin, C. T. Williams, S. M. Hart, M. G. Goering, M. O'Neil-Johnson and L. Zeng, *Anal. Chem.*, 2002, **74**, 3963–3971.
- G. Lang, N. A. Mayhudin, M. I. Mitova, L. Sun, S. van der Sar, J. W. Blunt, A. L. J. Cole, G. Ellis, H. Laatsch and M. H. G. Munro, *J. Nat. Prod.*, 2008, **71**, 1595–1599.
- H. Osada and T. Nogawa, *Pure Appl. Chem.*, 2011, **84**, 1407–1420.
- J. L. Wolfender, J. M. Nuzillard, J. J. J. van der Hooft, J. H. Renault and S. Bertrand, *Anal. Chem.*, 2019, **91**, 704–742.
- B. C. Covington and M. R. Seyedsayamdost, *Nat. Prod. Rep.*, 2025, **42**, 956–964.
- S. L. Collins, I. Koo, J. M. Peters, P. B. Smith and A. D. Patterson, *Annu. Rev. Anal. Chem.*, 2021, **14**, 467–487.
- D. G. McLaren, V. Shah, T. Wisniewski, L. Ghislain, C. Liu, H. Zhang and S. A. Saldanha, *SLAS Discovery*, 2021, **26**, 168–191.
- J. D. Williams, F. Pu, J. W. Sawicki and N. L. Elsen, *Expet Opin. Drug Discov.*, 2024, **19**, 291–301.
- S. A. Jarmusch, J. J. J. van der Hooft, P. C. Dorrestein and A. K. Jarmusch, *Nat. Prod. Rep.*, 2021, **38**, 2066–2082.
- A. Rutz, W. Bittremieux, R. Schmid, O. Cailloux, J. J. J. van der Hooft and M. A. Beniddir, *Nat. Prod. Rep.*, 2025, DOI: [10.1039/d5np00034c](https://doi.org/10.1039/d5np00034c).
- R. A. Shepherd, C. A. Fihn, A. J. Tabag, S. M. K. McKinnie and L. M. Sanchez, *Nat. Prod. Rep.*, 2025, **42**, 1037–1054.
- C. N. Naylor and G. Nagy, *Mass Spectrom. Rev.*, 2025, **44**, 581–598.
- G. R. D. Prabhu, E. R. Williams, M. Wilm and P. L. Urban, *Nat. Rev. Methods Primers*, 2023, **3**, 24.
- N. B. Cech and C. G. Enke, *Mass Spectrom. Rev.*, 2001, **20**, 362–387.
- J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.
- R. Schmid, D. Petras, L. F. Nothias, M. Wang, A. T. Aron, A. Jagels, H. Tsugawa, J. Rainer, M. Garcia-Aloy, K. Duhrkop, A. Korf, T. Pluskal, Z. Kamenik, A. K. Jarmusch, A. M. Caraballo-Rodriguez, K. C. Weldon, M. Nothias-Esposito, A. A. Aksenov, A. Bauermeister,



- A. Albarracin Orio, C. O. Grundmann, F. Vargas, I. Koester, J. M. Gauglitz, E. C. Gentry, Y. Hovelmann, S. A. Kalinina, M. A. Pendergraft, M. Panitchpakdi, R. Tehan, A. Le Gouellec, G. Aleti, H. Mannocho Russo, B. Arndt, F. Hubner, H. Hayen, H. Zhi, M. Raffatellu, K. A. Prather, L. I. Aluwihare, S. Bocker, K. L. McPhail, H. U. Humpff, U. Karst and P. C. Dorrestein, *Nat. Commun.*, 2021, **12**, 3832.
- 27 I. Marchi, S. Rudaz and J. L. Veuthey, *Talanta*, 2009, **78**, 1–18.
- 28 W. M. Niessen, P. Manini and R. Andreoli, *Mass Spectrom. Rev.*, 2006, **25**, 881–899.
- 29 T. R. Covey, B. A. Thomson and B. B. Schneider, *Mass Spectrom. Rev.*, 2009, **28**, 870–897.
- 30 R. Smith, C. Brookes, M. Morris and P. Barran, *Anal. Chem.*, 2025, **97**, 22457–22474.
- 31 J. D. Watrous and P. C. Dorrestein, *Nat. Rev. Microbiol.*, 2011, **9**, 683–694.
- 32 T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23–60.
- 33 W. B. Dunn, A. Erban, R. J. M. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka and M. R. Viant, *Metabolomics*, 2012, **9**, 44–66.
- 34 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crusemann, 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. Humpff, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. B. P, 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. 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. Alexandrov, 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. Shi, X. Liu, L. 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 and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.
- 35 R. A. Zubarev and A. Makarov, *Anal. Chem.*, 2013, **85**, 5288–5296.
- 36 I. Blazenovic, T. Kind, J. Ji and O. Fiehn, *Metabolites*, 2018, **8**, 31.
- 37 C. D. Kelstrup, D. B. Bekker-Jensen, T. N. Arrey, A. Hogrebe, A. Harder and J. V. Olsen, *J. Proteome Res.*, 2018, **17**, 727–738.
- 38 K. Masike, M. A. Stander and A. de Villiers, *J. Pharm. Biomed. Anal.*, 2021, **195**, 113846.
- 39 F. Carnevale Neto, T. N. Clark, N. P. Lopes and R. G. Linington, *J. Nat. Prod.*, 2022, **85**, 519–529.
- 40 B. A. Bell, J. M. Anderson, S. R. Rajski and T. S. Bugni, *J. Nat. Prod.*, 2025, **88**, 306–313.
- 41 V. Gabelica and E. Marklund, *Curr. Opin. Chem. Biol.*, 2018, **42**, 51–59.
- 42 J. C. May and J. A. McLean, *Anal. Chem.*, 2015, **87**, 1422–1436.
- 43 F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281–294.
- 44 A. J. Levy, N. R. Oranzi, A. Ahmadireskety, R. H. J. Kemperman, M. S. Wei and R. A. Yost, *TrAC, Trends Anal. Chem.*, 2019, **116**, 274–281.
- 45 T. Mairinger, T. J. Causon and S. Hann, *Curr. Opin. Chem. Biol.*, 2018, **42**, 9–15.
- 46 A. C. Schrimpe-Rutledge, S. D. Sherrod and J. A. McLean, *Curr. Opin. Chem. Biol.*, 2018, **42**, 160–166.
- 47 E. S. Baker, C. Hoang, W. Uritboonthai, H. M. Heyman, B. Pratt, M. MacCoss, B. MacLean, R. Plumb, A. Aisporna and G. Siuzdak, *Nat. Methods*, 2023, **20**, 1836–1837.
- 48 C. M. Nichols, J. N. Dodds, B. S. Rose, J. A. Picache, C. B. Morris, S. G. Codreanu, J. C. May, S. D. Sherrod and J. A. McLean, *Anal. Chem.*, 2018, **90**, 14484–14492.
- 49 Z. Zhou, M. Luo, X. Chen, Y. Yin, X. Xiong, R. Wang and Z. J. Zhu, *Nat. Commun.*, 2020, **11**, 4334.
- 50 A. Verma, A. Chattopadhyaya, P. Gupta, H. Tiwari, S. Singh, L. Kumar and V. Gautam, *Chem. Biodivers.*, 2025, **22**, e202500234.
- 51 S. D. Sarker and L. Nahar, in *Natural Products Isolation*, ed. S. D. Sarker and L. Nahar, Humana Press, Totowa, NJ, 2012, pp. 301–340, DOI: [10.1007/978-1-61779-624-1_12](https://doi.org/10.1007/978-1-61779-624-1_12).
- 52 M. W. Dong and K. Zhang, *TrAC, Trends Anal. Chem.*, 2014, **63**, 21–30.
- 53 S. Puri, D. Sahal and U. Sharma, *Anal. Sci. Adv.*, 2021, **2**, 579–593.
- 54 M. de Raad, C. R. Fischer and T. R. Northen, *Curr. Opin. Chem. Biol.*, 2016, **30**, 7–13.
- 55 R. E. Moore and P. J. Scheuer, *Science*, 1971, **172**, 495–498.
- 56 Z.-P. Jiang, S.-H. Sun, Y. Yu, A. Mándi, J.-Y. Luo, M.-H. Yang, T. Kurtán, W.-H. Chen, L. Shen and J. Wu, *Chem. Sci.*, 2021, **12**, 10197–10206.
- 57 D. J. Milanowski, N. Oku, L. K. Cartner, H. R. Bokesch, R. T. Williamson, J. Sauri, Y. Liu, K. A. Blinov, Y. Ding, X. C. Li, D. Ferreira, L. A. Walker, S. Khan, M. T. Davies-Coleman, J. A. Kelley, J. B. McMahan, G. E. Martin and K. R. Gustafson, *Chem. Sci.*, 2018, **9**, 307–314.
- 58 D. Wang, W. Jiang, C.-K. Kim, H. R. Bokesch, G. M. Woldemichael, B. E. Gryder, J. F. Shern, J. Khan, B. R. O'Keefe, J. A. Beutler and K. R. Gustafson, *Org. Lett.*, 2021, **23**, 3278–3281.
- 59 P.-T. Sun, Y.-G. Cao, G.-M. Xue, M. Li, C.-L. Zhang, F. Zhao, Z.-Y. Cao, D. Wang, K. R. Gustafson, X.-K. Zheng, W.-S. Feng and H. Chen, *Org. Lett.*, 2022, **24**, 1476–1480.



- 60 P. J. Proteau, *J. Nat. Prod.*, 2023, **86**, 653–654.
- 61 D. D. Laukien and W. H. Tschopp, *Concepts Magn. Reson.*, 1994, **6**, 255–273.
- 62 G. A. Morris, *J. Magn. Reson.*, 2019, **306**, 12–16.
- 63 R. C. Crouch and G. E. Martin, *J. Nat. Prod.*, 1992, **55**, 1343–1347.
- 64 G. E. Martin, R. C. Crouch and A. P. Zens, *Magn. Reson. Chem.*, 1998, **36**, 551–557.
- 65 B. D. Hilton and G. E. Martin, *J. Nat. Prod.*, 2010, **73**, 1465–1469.
- 66 T. F. Molinski, *Nat. Prod. Rep.*, 2010, **27**, 321–329.
- 67 A. Williams, G. Martin and D. Rovnyak, *Modern NMR Approaches to the Structure Elucidation of Natural Products—Volumn 1 Instrumentation and Software*, Royal Society of Chemistry, 2016.
- 68 A. G. Palmer, J. Cavanagh, P. E. Wright and M. Rance, *J. Magn. Reson.*, 1991, **93**, 151–170.
- 69 C. Zhang, Y. Idelbayev, N. Roberts, Y. Tao, Y. Nannapaneni, B. M. Duggan, J. Min, E. C. Lin, E. C. Gerwick, G. W. Cottrell and W. H. Gerwick, *Sci. Rep.*, 2017, **7**, 1–17.
- 70 H. W. Kim, C. Zhang, R. Reher, M. Wang, K. L. Alexander, L.-F. Nothias, Y. K. Han, H. Shin, K. Y. Lee, K. H. Lee, M. J. Kim, P. C. Dorrestein, W. H. Gerwick and G. W. Cottrell, *J. Cheminf.*, 2023, **15**, 71.
- 71 G. Kleks, D. C. Holland, J. Porter and A. R. Carroll, *Chem. Sci.*, 2021, **12**, 10930–10943.
- 72 R. Neufeld and D. Stalke, *Chem. Sci.*, 2015, **6**, 3354–3364.
- 73 W. T. P. Darling, S. G. Hyberts and M. Erdelyi, *Magn. Reson. Chem.*, 2025, **63**, 495–507.
- 74 D. Rovnyak, in *Annual Reports on NMR Spectroscopy: advances in non-uniform sampling NMR*, ed. W. S. Price, Academic Press, 2024, ch. 2, pp. 69–127, DOI: [10.1016/bs.arnmr.2024.01.001](https://doi.org/10.1016/bs.arnmr.2024.01.001).
- 75 S. Robson, H. Arthanari, S. G. Hyberts and G. Wagner, in *Methods Enzymol.*, ed. A. J. Wand, Academic Press, 2019, vol. 614, pp. 263–291.
- 76 M. M. Blum and H. John, *Drug Test. Anal.*, 2012, **4**, 298–302.
- 77 J. B. Johnson, K. B. Walsh, M. Naiker and K. Ameer, *Molecules*, 2023, **28**, 3215.
- 78 M. Alberts, T. Laino and A. C. Vaucher, *Commun. Chem.*, 2024, **7**, 268.
- 79 D. Punjabi, Y.-C. Huang, L. Holzhauser, P. Tremouilhac, P. Friederich, N. Jung and S. Bräse, *J. Cheminf.*, 2025, **17**, 24.
- 80 C. L. Zani and A. R. Carroll, *J. Nat. Prod.*, 2017, **80**, 1758–1766.
- 81 S. Abdul Al and A.-R. Allouche, *Chem. Phys. Lett.*, 2024, **856**, 141603.
- 82 O. Usoltsev, A. Tereshchenko, A. Skorynina, E. Kozyr, A. Soldatov, O. Safonova, A. H. Clark, D. Ferri, M. Nachttegaal and A. Bugaev, *Small Methods*, 2024, **8**, e2301397.
- 83 G. Jung, S. G. Jung and J. M. Cole, *Chem. Sci.*, 2023, **14**, 3600–3609.
- 84 N. Hendrick, D. Fraser, R. Bennett, K. Corazzata, D. A. Adpressa, A. A. Makarov and A. Beeler, *J. Pharm. Biomed. Anal.*, 2023, **229**, 115350.
- 85 L. Martínez-Fructuoso, S. J. R. Arends, V. F. Freire, J. R. Evans, S. DeVries, B. D. Peyser, R. K. Akee, C. C. Thornburg, R. Kumar, S. Ensel, G. M. Morgan, G. D. McConachie, N. Veeder, L. R. Duncan, T. Grkovic and B. R. O'Keefe, *ACS Infect. Dis.*, 2023, **9**, 1245–1256.
- 86 V. F. Freire, L. Martínez-Fructuoso, R. Kumar, R. K. Akee, C. C. Thornburg, S. Ensel, E. Okoroafor, J. R. Evans, D. Wang, B. D. Peyser, T. Grkovic and B. R. O'Keefe, *RSC Adv.*, 2024, **14**, 38200–38207.
- 87 P. J. Stephens, D. M. McCann, F. J. Devlin and A. B. Smith, *J. Nat. Prod.*, 2006, **69**, 1055–1064.
- 88 A. R. Puente, B. K. Chhetri, J. Kubanek and P. L. Polavarapu, *Symmetry*, 2024, **16**, 133.
- 89 A. McGown, J. Nafie, M. Otafah, S. Hassell-Hart, G. J. Tizzard, S. J. Coles, R. Banks, G. P. Marsh, H. J. Maple, G. E. Kostakis, I. Proietti Silvestri, P. Colbon and J. Spencer, *RSC Chem. Biol.*, 2023, **4**, 716–721.
- 90 A. Mandi and T. Kurtan, *Nat. Prod. Rep.*, 2019, **36**, 889–918.
- 91 S. P. Gaudencio, E. Bayram, L. Lukic Bilela, M. Cueto, A. R. Diaz-Marrero, B. Z. Haznedaroglu, C. Jimenez, M. Mandalakis, F. Pereira, F. Reyes and D. Tasdemir, *Mar. Drugs*, 2023, **21**, 308.
- 92 A. E. Nugroho and H. Morita, *Nat. Chem.*, 2014, **68**, 1–10.
- 93 T. B. Freedman, X. Cao, R. K. Dukor and L. A. Nafie, *Chirality*, 2003, **15**, 743–758.
- 94 D. Long, Z. Li, X. Xu, C. Liu, H. Zhang, X. Cao, L. Yang, X. Wang and F. Mo, *Sci. Data*, 2025, **12**, 1641.
- 95 Z.-Q. Huo, F. Zhu, X.-W. Zhang, X. Zhang, H.-B. Liang, J.-C. Yao, Z. Liu, G.-M. Zhang, Q.-Q. Yao and G.-F. Qin, *Mar. Drugs*, 2022, **20**, 333.
- 96 G. Mazzeo, E. Santoro, A. Andolfi, A. Cimmino, P. Troselj, A. G. Petrovic, S. Superchi, A. Evidente and N. Berova, *J. Nat. Prod.*, 2013, **76**, 588–599.
- 97 A. Mohamed, C. H. Nguyen and H. Mamitsuka, *Briefings Bioinf.*, 2016, **17**, 309–321.
- 98 Dictionary of Natural Products, 2025, **34**, 2, <https://dnp.chemnetbase.com>.
- 99 MarinLit, 2025, <https://marinlit.rsc.org>.
- 100 H. Laatsch, AntiBase Library - Wiley Identifier of Natural Products, 2025, <https://sciencesolutions.wiley.com/>.
- 101 L. F. Nothias, D. Petras, R. Schmid, K. Dührkop, 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. van der Hooft, F. Vargas, K. C. Weldon, M. Witting, H. Yang, Z. Zhang, F. Zubeil, O. Kohlbacher, S. Bocker, T. Alexandrov, N. Bandeira, M. Wang and P. C. Dorrestein, *Nat. Methods*, 2020, **17**, 905–908.



- 102 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. Tejada 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 and P. C. Dorrestein, *Nat. Protoc.*, 2020, **15**, 1954–1991.
- 103 M. Wang, A. K. Jarmusch, F. Vargas, A. A. Aksenov, J. M. Gauglitz, K. Weldon, D. Petras, R. da Silva, R. Quinn, A. V. Melnik, J. J. J. van der Hooft, A. M. Caraballo-Rodriguez, L. F. Nothias, C. M. Aceves, M. Panitchpakdi, E. Brown, F. Di Ottavio, N. Sikora, E. O. Elijah, L. Labarta-Bajo, E. C. Gentry, S. Shalpour, K. E. Kyle, S. P. Puckett, J. D. Watrous, C. S. Carpenter, A. Bouslimani, M. Ernst, A. D. Swafford, E. I. Zúñiga, M. J. Balunas, J. L. Klassen, R. Loomba, R. Knight, N. Bandeira and P. C. Dorrestein, *Nat. Biotechnol.*, 2020, **38**, 23–26.
- 104 P. W. P. Gomes, H. Mannochio-Russo, R. Schmid, S. Zuffa, T. Damiani, L. M. Quiros-Guerrero, A. M. Caraballo-Rodriguez, H. N. Zhao, H. Yang, S. Xing, V. Charron-Lamoureux, D. N. Chigumba, B. E. Sedio, J. A. Myers, P. M. Allard, T. V. Harwood, G. Tamayo-Castillo, K. B. Kang, E. Defossez, H. H. F. Koolen, M. N. da Silva, E. S. CYY, S. Rasmann, T. W. N. Walker, G. Glauser, J. M. Chaves-Fallas, B. David, H. Kim, K. H. Lee, M. J. Kim, W. J. Choi, Y. S. Keum, E. de Lima, L. S. de Medeiros, G. A. Bataglione, E. V. Costa, F. M. A. da Silva, A. R. V. Carvalho, J. D. E. Reis, S. Pamplona, E. Jeong, K. Lee, G. J. Kim, Y. S. Kil, J. W. Nam, H. Choi, Y. K. Han, S. Y. Park, K. Y. Lee, C. Hu, Y. Dong, S. Sang, C. R. Morrison, R. M. Borges, A. M. Teixeira, S. Y. Lee, B. S. Lee, S. Y. Jeong, K. H. Kim, A. Rutz, A. Gaudry, E. Bruehlhart, I. F. Kappers, R. Karlova, M. Meisenburg, R. Berdager, J. S. Tello, D. Henderson, L. Cayola, S. J. Wright, D. N. Allen, K. J. Anderson-Teixeira, J. L. Baltzer, J. A. Lutz, S. M. McMahon, G. G. Parker, J. D. Parker, T. R. Northen, B. P. Bowen, T. Pluskal, J. J. J. van der Hooft, J. J. Carver, N. Bandeira, B. S. Pullman, J. L. Wolfender, R. D. Kersten, M. Wang and P. C. Dorrestein, *bioRxiv*, 2024, DOI: [10.1101/2024.05.13.593988](https://doi.org/10.1101/2024.05.13.593988).
- 105 S. Zuffa, R. Schmid, A. Bauermeister, P. G. PW, A. M. Caraballo-Rodriguez, Y. El Abiead, A. T. Aron, E. C. Gentry, J. Zemlin, M. J. Meehan, N. E. Avalon, R. H. Cichewicz, E. Buzun, M. C. Terrazas, C. Y. Hsu, R. Oles, A. V. Ayala, J. Zhao, H. Chu, M. C. M. Kuijpers, S. L. Jackrel, F. Tugizimana, L. P. Nephali, I. A. Dubery, N. E. Madala, E. A. Moreira, L. V. Costa-Lotufu, N. P. Lopes, P. Rezende-Teixeira, P. C. Jimenez, B. Rimal, A. D. Patterson, M. F. Traxler, R. C. Pessotti, D. Alvarado-Villalobos, G. Tamayo-Castillo, P. Chaverri, E. Escudero-Leyva, L. M. Quiros-Guerrero, A. J. Bory, J. Joubert, A. Rutz, J. L. Wolfender, P. M. Allard, A. Sichert, S. Pontrelli, B. S. Pullman, N. Bandeira, W. H. Gerwick, K. Gindro, J. Massana-Codina, B. C. Wagner, K. Forchhammer, D. Petras, N. Aiosa, N. Garg, M. Liebeke, P. Bourceau, K. B. Kang, H. Gadhavi, L. P. S. de Carvalho, M. Silva Dos Santos, A. I. Perez-Lorente, C. Molina-Santiago, D. Romero, R. Franke, M. Bronstrup, A. Vera Ponce de Leon, P. B. Pope, S. L. La Rosa, G. La Barbera, H. M. Roager, M. F. Laursen, F. Hammerle, B. Siewert, U. Peintner, C. Licon-Cassani, L. Rodriguez-Orduna, E. Rampler, F. Hildebrand, G. Koellensperger, H. Schoeny, K. Hohenwallner, L. Panzenboeck, R. Gregor, E. C. O'Neill, E. T. Roxborough, J. Odoi, N. J. Bale, S. Ding, J. S. Sinninghe Damste, X. L. Guan, J. J. Cui, K. S. Ju, D. B. Silva, F. M. R. Silva, G. F. da Silva, H. H. F. Koolen, C. Grundmann, J. A. Clement, H. Mohimani, K. Broders, K. L. McPhail, S. E. Ober-Singleton, C. M. Rath, D. McDonald, R. Knight, M. Wang and P. C. Dorrestein, *Nat. Microbiol.*, 2024, **9**, 336–345.
- 106 A. K. Jarmusch, M. Wang, C. M. Aceves, R. S. Advani, S. Aguirre, A. A. Aksenov, G. Aleti, A. T. Aron, A. Bauermeister, S. Bolleddu, A. Bouslimani, A. M. Caraballo Rodriguez, R. Chaar, R. Coras, E. O. Elijah, M. Ernst, J. M. Gauglitz, E. C. Gentry, M. Husband, S. A. Jarmusch, K. L. Jones 2nd, Z. Kamenik, A. Le Gouellec, A. Lu, L. I. McCall, K. L. McPhail, M. J. Meehan, A. V. Melnik, R. C. Menezes, Y. A. Montoya Giraldo, N. H. Nguyen, L. F. Nothias, M. Nothias-Esposito, M. Panitchpakdi, D. Petras, R. A. Quinn, N. Sikora, J. J. J. van der Hooft, F. Vargas, A. Vrbanac, K. C. Weldon, R. Knight, N. Bandeira and P. C. Dorrestein, *Nat. Methods*, 2020, **17**, 901–904.
- 107 J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess and S. Rogers, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 13738–13743.
- 108 K. Dührkop, H. Shen, M. Meusel, J. Rousu and S. Böcker, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12580–12585.
- 109 N. F. de Jonge, J. J. R. Louwen, E. Chekmeneva, S. Camuzeaux, F. J. Vermeir, R. S. Jansen, F. Huber and J. J. J. van der Hooft, *Nat. Commun.*, 2023, **14**, 1752.
- 110 H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703–714.
- 111 MassBank of North America (MoNA), 2026, <https://mona.fiehnlab.ucdavis.edu/>.
- 112 J. Xue, C. Guijas, H. P. Benton, B. Warth and G. Siuzdak, *Nat. Methods*, 2020, **17**, 953–954.
- 113 C. A. Smith, G. O. Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak, *Ther. Drug Monit.*, 2005, **27**, 747–751.



- 114 Wiley Registry/NIST Mass Spectral Library, 2026, <https://www.sciencesolutions.wiley.com>.
- 115 Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M. Y. Hirai and K. Saito, *Phytochemistry*, 2012, **82**, 38–45.
- 116 mzCloud Advanced Mass Spectral Database, 2026, <https://www.mzcloud.org/>.
- 117 C. Brungs, R. Schmid, S. Heuckeroth, A. Mazumdar, M. Drexler, P. Šácha, P. C. Dorrestein, D. Petras, L.-F. Nothias, V. Veverka, R. Nencka, Z. Kameník and T. Pluskal, *Nat. Methods*, 2025, **22**, 2028–2031.
- 118 M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik and C. Steinbeck, *J. Cheminf.*, 2021, **13**, 2.
- 119 A. Rutz, M. Sorokina, J. Galgonek, D. Mietchen, E. Willighagen, A. Gaudry, J. G. Graham, R. Stephan, R. Page, J. Vondrášek, C. Steinbeck, G. F. Pauli, J.-L. Wolfender, J. Bisson and P.-M. Allard, *eLife*, 2022, **11**, e70780.
- 120 E. F. Poynton, J. A. van Santen, M. Pin, M. M. Contreras, E. McMann, J. Parra, B. Showalter, L. Zaroubi, K. R. Duncan and R. G. Linington, *Nucleic Acids Res.*, 2025, **53**, D691–D699.
- 121 J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castano-Espriu, C. Chang, T. N. Clark, J. L. Cleary Little, D. A. Delgadillo, P. C. Dorrestein, K. R. Duncan, J. M. Egan, M. M. Galey, F. P. J. Haeckl, A. Hua, A. H. Hughes, D. Iskakova, A. Khadilkar, J. H. Lee, S. Lee, N. LeGrow, D. Y. Liu, J. M. Macho, C. S. McCaughey, M. H. Medema, R. P. Neupane, T. J. O'Donnell, J. S. Paula, L. M. Sanchez, A. F. Shaikh, S. Soldatou, B. R. Terlouw, T. A. Tran, M. Valentine, J. J. J. van der Hooft, D. A. Vo, M. Wang, D. Wilson, K. E. Zink and R. G. Linington, *ACS Cent. Sci.*, 2019, **5**, 1824–1833.
- 122 M. Simone, M. Iorio, P. Monciardini, M. Santini, N. Cantù, A. Tocchetti, S. Serina, C. Brunati, T. Vernay, A. Gentile, M. Aracne, M. Cozzi, J. J. J. van der Hooft, M. Sosio, S. Donadio and S. I. Maffioli, *J. Nat. Prod.*, 2024, **87**, 2615–2628.
- 123 ChemSpider, 2026, <https://www.chemspider.com/>.
- 124 PubChem, 2026, <https://pubchem.ncbi.nlm.nih.gov/>.
- 125 K. Duhrkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu and S. Bocker, *Nat. Methods*, 2019, **16**, 299–302.
- 126 H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihiro, W. Tanaka, T. Cajka, K. Saito, O. Fiehn and M. Arita, *Anal. Chem.*, 2016, **88**, 7946–7958.
- 127 C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender and S. Neumann, *J. Cheminf.*, 2016, **8**, 3.
- 128 M. Gerlich, S. Neumann and J. Mass, *Spectrom.*, 2013, **48**, 291–298.
- 129 F. Wang, D. Allen, S. Tian, E. Oler, V. Gautam, R. Greiner, T. O. Metz and D. S. Wishart, *Nucleic Acids Res.*, 2022, **50**, W165–W174.
- 130 P. M. Allard, T. Peresse, J. Bisson, K. Gindro, L. Marcourt, V. C. Pham, F. Roussi, M. Litaudon and J. L. Wolfender, *Anal. Chem.*, 2016, **88**, 3317–3323.
- 131 H. Mohimani, A. Gurevich, A. Shlemov, A. Mikheenko, A. Korobeynikov, L. Cao, E. Shcherbin, L. F. Nothias, P. C. Dorrestein and P. A. Pevzner, *Nat. Commun.*, 2018, **9**, 4035.
- 132 L. Cao, M. Guler, A. Tagirdzhanov, Y. Y. Lee, A. Gurevich and H. Mohimani, *Nat. Commun.*, 2021, **12**, 3718.
- 133 S. Rogers, C. W. Ong, J. Wandy, M. Ernst, L. Ridder and J. J. J. van der Hooft, *Faraday Discuss.*, 2019, **218**, 284–302.
- 134 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 and J. J. J. van der Hooft, *Metabolites*, 2019, **9**, 144.
- 135 R. R. da Silva, M. Wang, L. F. Nothias, J. J. J. van der Hooft, A. M. Caraballo-Rodriguez, E. Fox, M. J. Balunas, J. L. Klassen, N. P. Lopes and P. C. Dorrestein, *PLoS Comput. Biol.*, 2018, **14**, e1006089.
- 136 J. A. Picache, B. S. Rose, A. Balinski, K. L. Leaptrot, S. D. Sherrod, J. C. May and J. A. McLean, *Chem. Sci.*, 2019, **10**, 983–993.
- 137 X. Zheng, N. A. Aly, Y. Zhou, K. T. Dupuis, A. Bilbao, V. L. Paurus, D. J. Orton, R. Wilson, S. H. Payne, R. D. Smith and E. S. Baker, *Chem. Sci.*, 2017, **8**, 7724–7736.
- 138 D. H. Ross, J. H. Cho and L. Xu, *Anal. Chem.*, 2020, **92**, 4548–4557.
- 139 H. Zhang, M. Luo, H. Wang, F. Ren, Y. Yin and Z. J. Zhu, *Anal. Chem.*, 2023, **95**, 13913–13921.
- 140 X. Zeng, P. Zhang, W. He, C. Qin, S. Chen, L. Tao, Y. Wang, Y. Tan, D. Gao, B. Wang, Z. Chen, W. Chen, Y. Y. Jiang and Y. Z. Chen, *Nucleic Acids Res.*, 2018, **46**, D1217–D1222.
- 141 T. Xu, J. Dai, Y. Li, J. Zhou, Y. Zhao, W. Chen and X. S. Xue, *J. Cheminf.*, 2025, **17**, 172.
- 142 B. Zdrzizil, E. Felix, F. Hunter, E. J. Manners, J. Blackshaw, S. Corbett, M. de Veij, H. Ioannidis, D. M. Lopez, J. F. Mosquera, M. P. Magarinos, N. Bosc, R. Arcila, T. Kiziloren, A. Gaulton, A. P. Bento, M. F. Adasme, P. Monecke, G. A. Landrum and A. R. Leach, *Nucleic Acids Res.*, 2024, **52**, D1180–D1192.
- 143 PubChem BioAssays, 2026, <https://pubchem.ncbi.nlm.nih.gov/docs/bioassays>.
- 144 MassIVE (Mass Spectrometry Interactive Virtual Environment), 2026, <https://massive.ucsd.edu>.
- 145 K. Haug, K. Cochrane, V. C. Nainala, M. Williams, J. Chang, K. V. Jayaseelan and C. O'Donovan, *Nucleic Acids Res.*, 2020, **48**, D440–D444.
- 146 M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K. S. Nair, S. Sumner and S. Subramaniam, *Nucleic Acids Res.*, 2016, **44**, D463–D470.
- 147 S. Heuckeroth, T. Damiani, A. Smirnov, O. Mokshyna, C. Brungs, A. Korf, J. D. Smith, P. Stincone, N. Dreolin, L. F. Nothias, T. Hyotylainen, M. Oresic, U. Karst, P. C. Dorrestein, D. Petras, X. Du, J. J. J. van der Hooft, R. Schmid and T. Pluskal, *Nat. Protoc.*, 2024, **19**, 2597–2641.
- 148 R. Tautenhahn, G. J. Patti, D. Rinehart and G. Siuzdak, *Anal. Chem.*, 2012, **84**, 5035–5039.
- 149 H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn and M. Arita, *Nat. Methods*, 2015, **12**, 523–526.



- 150 J. Pfeuffer, C. Bielow, S. Wein, K. Jeong, E. Netz, A. Walter, O. Alka, L. Nilse, P. D. Colaianne, D. McCloskey, J. Kim, G. Rosenberger, L. Bichmann, M. Walzer, J. Veit, B. Boudaud, M. Bernt, N. Patikas, M. Pilz, M. P. Startek, S. Kutuzova, L. Heumos, J. Charkow, J. C. Sing, A. Feroz, A. Siraj, H. Weisser, T. M. H. Dijkstra, Y. Perez-Riverol, H. Rost, O. Kohlbacher and T. Sachsenberg, *Nat. Methods*, 2024, **21**, 365–367.
- 151 H. L. Rost, T. Sachsenberg, S. Aiche, C. Bielow, H. Weisser, F. Aicheler, S. Andreotti, H. C. Ehrlich, P. Gutenbrunner, E. Kenar, X. Liang, S. Nahnsen, L. Nilse, J. Pfeuffer, G. Rosenberger, M. Rurik, U. Schmitt, J. Veit, M. Walzer, D. Wojnar, W. E. Wolski, O. Schilling, J. S. Choudhary, L. Malmstrom, R. Aebersold, K. Reinert and O. Kohlbacher, *Nat. Methods*, 2016, **13**, 741–748.
- 152 F. Yang, Z. Liang, H. Zhao, J. Zheng, L. Liu, H. Song and G. Xin, *Chin. J. Nat. Med.*, 2025, **23**, 410–420.
- 153 A. Vaniya, S. N. Samra, M. Palazoglu, H. Tsugawa and O. Fiehn, *Phytochem. Lett.*, 2017, **21**, 306–312.
- 154 I. Blazenovic, T. Kind, H. Torbasinovic, S. Obrenovic, S. S. Mehta, H. Tsugawa, T. Wermuth, N. Schauer, M. Jahn, R. Biedendieck, D. Jahn and O. Fiehn, *J. Cheminf.*, 2017, **9**, 32.
- 155 T. Grkovic, R. H. Pouwer, M.-L. Vial, L. Gambini, A. Noel, J. N. A. Hooper, S. A. Wood, G. D. Mellick and R. J. Quinn, *Angew. Chem., Int. Ed.*, 2014, **53**, 6070–6074.
- 156 M. Liu, T. Grkovic, R. J. Quinn, X. Liu, J. Han, L. Zhang and J. Han, *Synth. Syst. Biotechnol.*, 2017, **2**, 276–286.
- 157 L. Buedenbender, L. J. Habener, T. Grkovic, D. I. Kurtboke, S. Duffy, V. M. Avery and A. R. Carroll, *J. Nat. Prod.*, 2018, **81**, 957–965.
- 158 D. S. Wishart, Z. Sayeeda, Z. Budinski, A. Guo, B. L. Lee, M. Berjanskii, M. Rout, H. Peters, R. Dizon, R. Mah, C. Torres-Calzada, M. Hiebert-Giesbrecht, D. Varshavi, D. Varshavi, E. Oler, D. Allen, X. Cao, V. Gautam, A. Maras, E. F. Poynton, P. Tavangar, V. Yang, J. A. van Santen, R. Ghosh, S. Sarma, E. Knutson, V. Sullivan, A. M. Jystad, R. Renslow, L. W. Sumner, R. G. Linington and J. R. Cort, *Nucleic Acids Res.*, 2021, **50**, D665–D677.
- 159 J. L. López-Pérez, R. Therón, E. del Olmo and D. Díaz, *Bioinform.*, 2007, **23**, 3256–3257.
- 160 R. Reher, H. W. Kim, C. Zhang, H. H. Mao, M. Wang, L.-F. Nothias, A. M. Caraballo-Rodriguez, E. Glukhov, B. Teke, T. Leao, K. L. Alexander, B. M. Duggan, E. L. Van Everbroeck, P. C. Dorrestein, G. W. Cottrell and W. H. Gerwick, *J. Am. Chem. Soc.*, 2020, **142**, 4114–4120.
- 161 nmrXiv, 2026, <https://nmrxiv.org/>.
- 162 Harvard Dataverse, 2026, <https://dataverse.harvard.edu/>.
- 163 Jeol CH-NMR-NP, 2026, <https://ch-nmr-np.jeol.co.jp/en/nmrdb/>.
- 164 J.-J. Wang, Y. Jin, C.-Y. Zhi, Y.-J. Liu, X.-H. Huang, F. Xu, X. Ji, X. Fang, H. Tao, W. E. L. Zhang, G. Ke and R. Zhu, *Sci. Data*, 2025, **12**, 1954.
- 165 Q. Wang, W. Zhang, M. Chen, X. Li, Z. Xiong, J. Xiong, Z. Fu and M. Zheng, *Chem. Sci.*, 2025, **16**, 11548–11558.
- 166 Micronmr, 2026, <https://en.nmrdata.com/>.
- 167 A. Howarth, K. Ermanis and J. M. Goodman, *Chem. Sci.*, 2020, **11**, 4351–4359.
- 168 T. F. Molinski, *Curr. Opin. Drug Discovery Dev.*, 2009, **12**, 197–206.
- 169 D. S. Dalisay and T. F. Molinski, *Org. Lett.*, 2009, **11**, 1967–1970.
- 170 M. J. J. Recchia, T. U. H. Baumeister, D. Y. Liu and R. G. Linington, *Anal. Chem.*, 2023, **95**, 11908–11917.
- 171 K. R. Duncan, M. Crüsemann, A. Lechner, A. Sarkar, J. Li, N. Ziemert, M. Wang, N. Bandeira, B. S. Moore, P. C. Dorrestein and P. R. Jensen, *Chem. Biol.*, 2015, **22**, 460–471.
- 172 S. Kang, T.-H. Huynh, J. M. Kim, B. E. Heo, S. C. Jang, C. W. Ock, J. Lee, Y. Song, J. S. An, B. Shen, S. B. Kim, J. Jang, S. K. Lee, Y. J. Yoon and D.-C. Oh, *J. Am. Chem. Soc.*, 2025, **147**, 37719–37731.
- 173 D. Sweeney, A. Bogdanov, A. B. Chase, G. Castro-Falcón, A. Trinidad-Javier, S. Dahesh, V. Nizet and P. R. Jensen, *J. Nat. Prod.*, 2024, **87**, 2768–2778.
- 174 J. Park, Y.-H. Shin, S. Hwang, J. Kim, D. H. Moon, I. Kang, Y.-J. Ko, B. Chung, H. Nam, S. Kim, K. Moon, K.-B. Oh, J.-C. Cho, S. K. Lee and D.-C. Oh, *Angew. Chem., Int. Ed.*, 2024, **63**, e202402465.
- 175 M. Hagar, S. Kang, R. J. Andersen, D.-C. Oh and K. S. Ryan, *Curr. Opin. Microbiol.*, 2025, **84**, 102584.
- 176 D. Shin, W. S. Byun, S. Kang, I. Kang, E. S. Bae, J. S. An, J. H. Im, J. Park, E. Kim, K. Ko, S. Hwang, H. Lee, Y. Kwon, Y.-J. Ko, S. Hong, S.-J. Nam, S. B. Kim, W. Fenical, Y. J. Yoon, J.-C. Cho, S. K. Lee and D.-C. Oh, *J. Am. Chem. Soc.*, 2023, **145**, 19676–19690.
- 177 K. Pfeifer, D. Van Cura, K. J. Y. Wu and E. P. Balskus, *Nature*, 2026, **652**, 517–525.
- 178 A. R. Choirunnisa, K. Arima, Y. Abe, N. Kagaya, K. Kudo, H. Suenaga, J. Hashimoto, M. Fujie, N. Satoh, K. Shin-ya, K. Matsuda and T. Wakimoto, *Beilstein J. Org. Chem.*, 2022, **18**, 1017–1025.
- 179 T. Maxson, J. I. Tietz, G. A. Hudson, X. R. Guo, H.-C. Tai and D. A. Mitchell, *J. Am. Chem. Soc.*, 2016, **138**, 15157–15166.
- 180 D. Back, B. T. Shaffer, J. E. Loper and B. Philmus, *J. Nat. Prod.*, 2022, **85**, 105–114.
- 181 A. R. Carroll, B. R. Copp, T. Grkovic, R. A. Keyzers and M. R. Prinsep, *Nat. Prod. Rep.*, 2026, **43**, 89–131.
- 182 D. S. Levine, M. Shuaibi, E. W. Clark Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models*, 2026, DOI: [10.48550/arXiv.2505.08762](https://doi.org/10.48550/arXiv.2505.08762).
- 183 M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Ye, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Židek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis and S. Velankar, *Nucleic Acids Res.*, 2023, **52**, D368–D375.

