Title

Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy

Journal

Trends in Pharmacological Sciences, 38(2)

ISSN

0165-6147

Authors

Quinn, Robert A

Nothias, Louis-Felix

Vining, Oliver

et al.

Publication Date

2017-02-01

DOI

10.1016/j.tips.2016.10.011

CellPress

Review

Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy

Robert A. Quinn,[1,4] Louis-Felix Nothias,[1,4] Oliver Vining,[2] Michael Meehan,[1] Eduardo Esquenazi,[2] and Pieter C. Dorrestein[1,3,*]

Molecular networking is a tandem mass spectrometry (MS/MS) data organizational approach that has been recently introduced in the drug discovery, metabolomics, and medical fields. The chemistry of molecules dictates how they will be fragmented by MS/MS in the gas phase and, therefore, two related molecules are likely to display similar fragment ion spectra. Molecular networking organizes the MS/MS data as a relational spectral network thereby mapping the chemistry that was detected in an MS/MS-based metabolomics experiment. Although the wider utility of molecular networking is just beginning to be recognized, in this review we highlight the principles behind molecular networking and its use for the discovery of therapeutic leads, monitoring drug metabolism, clinical diagnostics, and emerging applications in precision medicine.

Introduction to Molecular Networking

Mass spectrometry (MS) (see Glossary)-based profiling of human samples for the identification of disease began with the analysis of human urine and breath in the late 1960s [1,2]. Since these early days, the technology has advanced to the point that the instruments are being applied in real time during surgery to determine tumor phenotype [3]. MS has also become essential for research on active biological small molecules from nature and led to the development of new methodological approaches to explore the immense diversity of these molecules, also called **natural products**. These molecules are produced by diverse organisms found around the globe and have become some of our most essential medicines (e.g., penicillin, vancomycin, rapamycin, taxol, lovastatin). Despite the broad applicability of **untargeted MS** analysis and its widespread use, it remains challenging to annotate most of the chemical signatures detected by untargeted MS, limiting the utility of the data collected [4]. These chemical signatures can include unknown analogs of known compounds, completely unique chemical entities with no known relatives, or spectral signatures resulting from in-source **fragmentation** and ionization adducts. **Molecular networking** [5] is a computational strategy that aids visualization and interpretation of the chemical repertoire that can be detected using MS. It is has great potential to aid both MS-based disease diagnostics and drug development, including metabolism, because it harnesses the power of untargeted MS data that has historically been underutilized.

Molecular networking is a visualization strategy for untargeted MS. Untargeted MS is one of the key metabolite discovery and annotation strategies in **metabolomics**. Molecular networking

Trends

Molecular networking has been used to develop the world's largest repository and data analysis tool for tandem mass spectrometry (MS/MS) data, named Global Natural Products Social Molecular Networking (GNPS). GNPS is being used to decipher the metabolomic 'dark matter' of our world, in everything from plant extracts and microbial cultures to a variety of human and environmental samples, by propagating spectral library-based annotation and showing that there are chemical relationships between detected molecules across many sample types.

Molecular networking is being used to identify compounds related to medically important drugs that can be developed as new therapies.

Molecular networking can identify a broad diversity of unknown natural products with potential medical relevance, even from organisms that have already been extensively characterized.

GNPS and molecular networking are beginning to show cross-associations between the chemistries of seemingly unrelated biological systems. For example, platelet-activating factor, a bioactive lipid involved in human inflammation, was shown to also be involved in immune defense in corals by molecular networking-based annotation of untargeted MS data.

Molecular networking is beginning to be used in clinical medicine. MS/MS data can be collected and analyzed

provides a visual overview of all of the ions of molecules that were detected and fragmented during an MS experiment and the chemical relationships between them. This fragmentation-based MS method is referred to as **MS/MS**. Current methods using molecular networking of MS/MS data show that we can annotate only approximately 1.8% of the data compared with existing MS/MS reference libraries in the public domain [4,6]. The inability to annotate data is a significant bottleneck for the MS community that generally results in reporting of only the known molecules in an untargeted dataset and ignoring the rest. Although not directly annotated, there are many related ions that give rise to similar MS/MS spectra in a metabolomics experiment. Molecular networking enables the visualization of this chemical similarity, which is otherwise left unanalyzed.

Molecular networking uses a vector-based **computational algorithm** to compare the degree of spectral similarity between every MS/MS spectra in a dataset [5,7]. Output can be visualized as networks of MS/MS spectra called **molecular networks** [5,7]. When molecular networking was introduced, a commentary highlighted that this method is the chemical analog of metagenomics and that it would 'usher a new era in therapeutic discovery' [8]. Since its introduction in the field of metabolomics in 2012 [5], it has been successfully applied to benefit drug discovery pipelines, for the identification of novel virulence factors [9] and natural products, and in drug metabolism studies [10–12]. Molecular networking has led to the development of Global Natural Products Social Molecular Networking (GNPS) (http://gnps.ucsd.edu), a molecular networking and data-sharing web-based platform [6] allowing an ever-growing community of users to take advantage of this **bioinformatics** strategy and perform data-driven, crowd-sourced analysis. As database and algorithm, GNPS and molecular networking are analogous to GenBank and **Basic Local Alignment Search Tool (BLAST)** [13], which have had a significant impact on the field of DNA and RNA sequence-based science.

Several powerful metabolomics workflows, such as XC-MS online [14] or, more recently, W4M [15], MetaboAnalyst [16], and many other tools, use feature detection/alignment algorithms and various statistical approaches to discover biomarkers that are meaningful in a biological context. These bioinformatics pipelines are experiencing widespread use [17–19]. The molecular networking-based metabolomics workflow differs from previously established statistical workflows in that it analyzes chemical relationships between every MS/MS spectrum, visualizing the entire metabolome detected in a sample. Future workflows that integrate molecular networking with statistical comparison of samples or cohorts are being developed and becoming available through GNPS [6]. Here we introduce molecular networking and then focus on the range of its applications to highlight how it can be used as an innovative tool for drug discovery, clinical applications, and **precision medicine**.

Molecular Network Generation and Visualization

Molecular networking of MS/MS data is a graph-based workflow that aims to organize large MS datasets by mining spectral similarity between the MS/MS fragmentation patterns of different, but structurally related precursor ions (Figure 1, Key Figure) [5,7,8,20]. First, MS/MS data are simplified to reduce the downstream computational load and to enhance the efficiency of the spectral similarity algorithm [20,21]. In particular, low-intensity fragment ions and the precursor ion are removed from the MS/MS spectra. Moreover, spectra with the same precursor ion **mass-to-charge ratio (*m/z*)** and that have similar MS/MS spectra are merged into a single consensus spectrum [20,21]. These necessary data preprocessing steps tend to improve the spectral quality of lower-intensity precursor ions but cause most structural isomers to be merged into a single consensus spectrum. These consensus MS/MS spectra are then simplified as vectors in a multidimensional normalized space where each dimension corresponds to an *m/z* value and its respective intensity. These vectors are then used to calculate a **cosine score** (normalized dot product) between every possible pair of consensus MS/MS spectra, which allows the determination of the degree of spectral similarity between them (ranging from 0 to 1, 1

within hours to identify the molecular signatures of disease and metabolites of microbial, host, and xenobiotic origin.

[1]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA
[2]Sirenas, San Diego, CA, USA
[3]Departments of Pharmacology and Pediatrics, University of California, San Diego, San Diego, CA, USA
[4]These authors contributed equally

*Correspondence:
pdorrestein@ucsd.edu
(P.C. Dorrestein).

representing identical spectra [20]). Note that several parameters can be modified depending on the mass spectrometer used, such as the precursor ion mass tolerance for the consensus spectrum and the fragment ion mass tolerance (typically ± 0.02 Da for high-resolution instruments, in the range *m/z* 100–1500 Da). Moreover, the number of minimum-match fragment ions can be adjusted to meet the specificity of the fragmentation behavior of the analyzed molecules.

The output of these vector-based comparisons can then be visualized as graphs of spectral similarity called spectral networks [7] or molecular networks [5], where each **node** is a consensus MS/MS spectrum and **edges** between nodes indicate the degree of similarity between consensus spectra (above the similarity score threshold defined by the user, usually 0.7).

Early studies on molecular networking used MATLAB (The Mathworks, Inc.®) scripts installed on a desktop computer for the computation of similarity scores and visualization was achieved using Cytoscape® software [22]. With the introduction of the GNPS web platform (http://gnps.ucsd.edu), users can upload and store MS/MS data online, generate molecular networks, capture knowledge of the networks as an individual or as a community, and add sample information and other metadata to understand the network [6]. Subclusters of a molecular network represent molecules that are structurally related and are referred to as molecular families [23]. Data visualization of molecular families can now be performed directly online on GNPS, whereas visualization of an entire molecular network and its constituent molecular families is best done off-line with Cytoscape [22] or another network visualization tool. In Cytoscape, attributes of molecular networks can be tuned to help data interpretation, such as edge thickness, which can be proportional to the cosine score, node size, which can be correlated to precursor ion intensity, or node color, which can be used to map various metadata associated with samples (group mapping).

Originating from the assumption that related molecules produce similar fragmentation patterns in MS/MS [5,8], molecular networking produces an MS/MS spectral similarity map that allows the visualization of structurally related molecules. The main strength of this approach is that it can be used for the exploration of thousands to millions (and potentially billions) of MS/MS spectra without any prior knowledge regarding the chemical composition of the samples [24]. Moreover, GNPS can automatically perform a spectral library search for known molecules in the molecular networks, if their MS/MS spectra are available in public MS/MS spectral libraries [6]. This process, called **dereplication** by the natural product community and also known as identifying 'known unknowns' in metabolomics, is critical not only for the molecular annotation of MS/MS spectra, but also for the propagation of these annotations through the networks, allowing detection of analogs and the discovery of novel chemical products (Figure 2) [25]. It is in this way that molecular networking is analogous to the BLAST [13] algorithms used to identify similar nucleic acid sequences, a method that led to the explosion in the use of sequence data generated by the field of genomics.

Thus, the interpretation of MS/MS molecular networks can be conceptualized in two complementary paradigms: (i) an *ab initio* paradigm where data is organized without any prior knowledge of its chemical composition; and (ii) an *incrementum* paradigm in which molecular annotation is propagated through the molecular networks by 'seeding' confident annotation of molecular features.

Molecular Networking for Drug Discovery Leads

Natural products are a prolific source for new therapeutic leads [26]. Several studies report the successful application of molecular networking for the detection and isolation of bioactive compounds. Application of molecular networking to the natural products produced by marine *Vibrio* species led to the discovery of a series of antibacterial amino-polyketide derivatives named

Glossary

Basic Local Alignment Search Tool (BLAST): a method of comparing the similarity of nucleic acid sequences.
Bioinformatics: computational tools that can help in the analysis of large datasets generated on biological samples, most often nucleic acid sequencing and MS data.
Computational algorithm: sequence of operations that is computationally integrated.
Cosine score: expresses the angles between a pair of vectored MS/MS spectra.
Cystic fibrosis (CF): a genetic disease caused by mutations in the CF transmembrane conductance regulator gene that result in mucus accumulation in multiple organs and chronic lung infections.
Dereplication: the annotation of known molecules in a biological sample. In untargeted MS it is typically performed by spectral matching with spectral library spectra. In targeted MS it can be done by comparison with the properties of reference standards.
Edge: in a molecular network, the link between two metabolite nodes represented by the existence of a significantly spectral similarity, expressed as the cosine score value between a pair of MS/MS spectra.
Fragmentation/fragmented: dissociation of a precursor ion; can occur in-source and/or in the collision chamber, such as collision-induced dissociation (CID) or high-energy collision dissociation (HCD).
Liquid chromatography (LC): separates molecules based on their chemical interaction with the stationary phase.
Liquid chromatography–tandem mass spectrometry (LC-MS/MS): a hyphenated method comprising a liquid chromatography system coupled to a tandem mass spectrometer.
Mass spectrometry (MS): an analytical instrument that produces a beam of gas-phase ions from samples, sorts the resulting mixture of ions according to their *m/z* value and provides output signals from which the *m/z* and intensity (abundance) of each detected ionic species may be determined.
Mass-to-charge ratio (*m/z*): of an ion, measured by a mass spectrometer.

vitroprocines [27] as well as the anti-inflammatory and analgesic sphongonucleosides [28]. A joint metabolomics and genomics approach allowed the detection of unreported analogs and the biosynthetic intermediates of four non-ribosomal peptide synthase-derived molecular families with antibacterial activities in *Streptomyces roseosporus* [29]. The columbamides, a new class of trichlorinated acyl-amides from *Moorea* species with cannabinomimetic activity, were also discovered by a similar combined approach [30].

When larger numbers of organisms and their metabolomics datasets and genomes are available, pattern-based genome mining becomes possible. In effect, this enables the correlation of molecular families with the presence or absence of genes (or gene clusters) that encode them. The unique nature of molecular networking is that the molecules do not have to be identical but belong to the same family, as biosynthetic evolutionary processes often expand the diversity of bioactive molecules through attaching novel structural groups to base chemical backbones, enabling the discovery of related molecules. The first paper describing pattern-based genome mining was used to explore the molecular composition of thirty-five closely related strains of *Salinispora* (Actinomycetes) [31]. In this paper, pattern-based genome mining identified a unique biogenetic gene cluster observed in only a single strain. The MS/MS molecular networking used in this study led to the isolation and characterization of retimycin A, a new member of the quinomycin family of antibiotics.

Recently, molecular networking was also employed to explore the induced metabolome of *Burkholderia thailandensis* in response to the antibiotic trimethoprim. Over 100 novel compounds not seen in standard growth conditions were observed [32]. This example illustrates the power of molecular networking to reveal unknown bioactive compounds, even from microorganisms that have been extensively studied by conventional means.

The use of molecular networking in industrial settings is also becoming more prevalent. In the biotechnology sector, Sirenas LLC often relies on molecular networking to help expedite the discovery of novel small-molecule therapeutic leads from complex marine sources. The company has successfully used molecular networking in the discovery of potent cytotoxins that can serve as payloads for antibody–drug conjugates (ADCs). ADCs combine potent cytotoxic small molecules with highly specific monoclonal antibodies (mAbs) targeted to tumor-specific antigens, forging an important tool in personalized cancer treatment [33]. Sirenas has integrated molecular networking with chemometrics-based bioactivity predictions into their drug discovery platform. This approach has been successful in identifying new classes of molecules, including a new member of the well-studied dolastatin 10 structural class (the origin of the most successful ADC payload, auristatin, which is found in over 20 clinical-stage cancer products and the FDA-approved product Adcetris). The novel discovery exhibits increased potency and a more desirable physicochemical profile. To find this new molecule with molecular networking, a marine fraction library was screened against several cancer cell lines and analyzed by **liquid chromatography (LC)-MS/MS**. All metabolites observed in the fraction library were assigned 'activity scores' to prioritize efforts on only the molecules most likely to exhibit cytotoxic activity. Briefly, these scores were generated by determining the relative abundance of each metabolite through the integration of LC-MS features, which provides a comprehensive semiquantitative database of metabolites across the fraction library. Importantly, a single metabolite can be present in varying abundances across multiple fractions, allowing rank-order correlations to be made between the relative abundance of a metabolite and the strength of biological activity observed for each fraction in which it is present. Global analysis of the fraction library yields a correlative value, or activity score, for each metabolite tested in a given assay. This score falls between 0 and 1, with a score of 1 representing perfect correlation between the abundance of a metabolite and the biological activity of the fraction. Mapping these scores to the corresponding metabolites in a molecular network provides a method to visualize the structural and bioactivity

Metabolomics: the study of the whole set of small molecules from an organism.
Molecular networking: also called spectral networking; a method for the analysis and visualization of MS/MS data using vector-based spectral matching.
Molecular network: also called spectral network; a visual representation of the degree of spectral similarity between a dataset of MS/MS spectra.
Natural products: small-molecule secondary metabolites produced by a biological organism.
Node: in a molecular network, represents a cluster of highly similar MS/MS spectra (also called consensus MS/MS spectra) as determined by vector-based spectral matching.
Precision medicine: a method of medical practice involving a focused approach to diagnosis and treatment based on large amounts of data from a single individual.
Tandem mass spectrometry (MS/MS): an MS method that combines several mass analyzers into a single instrument allowing the filtering and fragmentation of a precursor ion and measurement of its product ions.
Targeted MS: MS method aimed at detecting specific ions and/or precursor-to-product transitions.
Untargeted MS: MS methods that are able to acquire the mass spectra of any detected ions that meet predefined characteristics (such as threshold level or *m/z* range).
Xenobiotics: metabolites of exogenous sources found in the human body.

Key Figure

Molecular Networking: A Graph-Based Tool to Explore Spectral Similarity in Liquid Chromatography–Tandem Mass Spectrometry (LC-MS/MS) Data from Molecular Mixtures

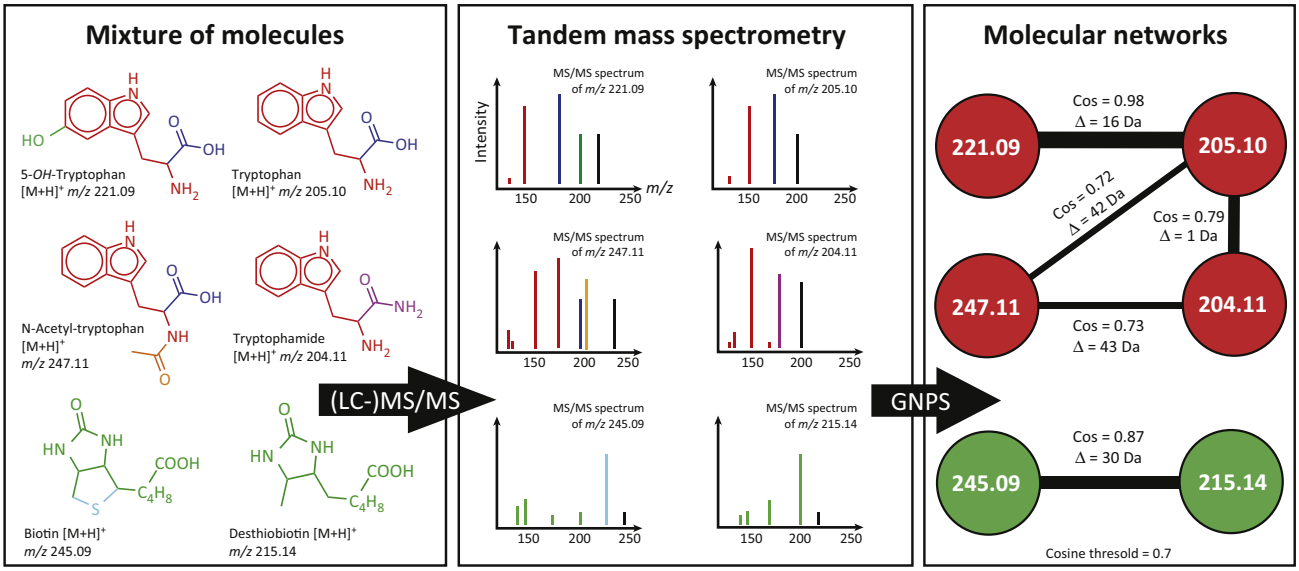


Figure 1. Molecular networks from LC-MS/MS spectra of tryptophan and biotin derivatives (from EMBL MCF spectral library on GNPS). The interactive view of the molecular network can be visualized directly on GNPS via the following link http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0415c5ae1d65449297b8aef26a480af9.

relationships between molecules (Figure 2). Interrogation of this network revealed a molecular family exhibiting strong activity scores and several member nodes were dereplicated as dolastatin 10 [34], as well as the related symplostatins 1 and 3 [35,36], and malevamide D [37] family members (Figure 2), along with a novel analog (SMD5041). Relying on mass-directed isolation, ~15 μg of pure compound was recovered from the original extract, allowing verification of an MS/MS-proposed structure by NMR. Secondary assays confirmed the predicted activity, with an IC_{50} of 1 nM against bt474 cancer cells. Importantly, isolation of this extremely minor metabolite would not have been possible via traditional bioassay-guided fractionation, as the requirement for iterative biological testing would have consumed all available material.

Synthesis of SMD5041 has been completed alongside several analogs, many of which retained low-nanomolar activity. Currently, clinically relevant ADCs using this molecular network-identified payload are undergoing preclinical evaluation. Development of these analogs, in addition to isolation and characterization of the natural product, represents a novel intellectual property space within a group of compounds that has been actively pursued in the pharmaceutical industry for over three decades.

The examples above demonstrate that in only a few years molecular networking has proved to be a valuable tool for drug lead discovery from nature. First, molecular networking can prioritize and rationalize the search for molecules in a biomass collection by accelerating the dereplication/annotation steps, putative recognition of bioactive molecules, and the detection of unique chemical space or niches. Second, its integration with genomic analysis can lead to the

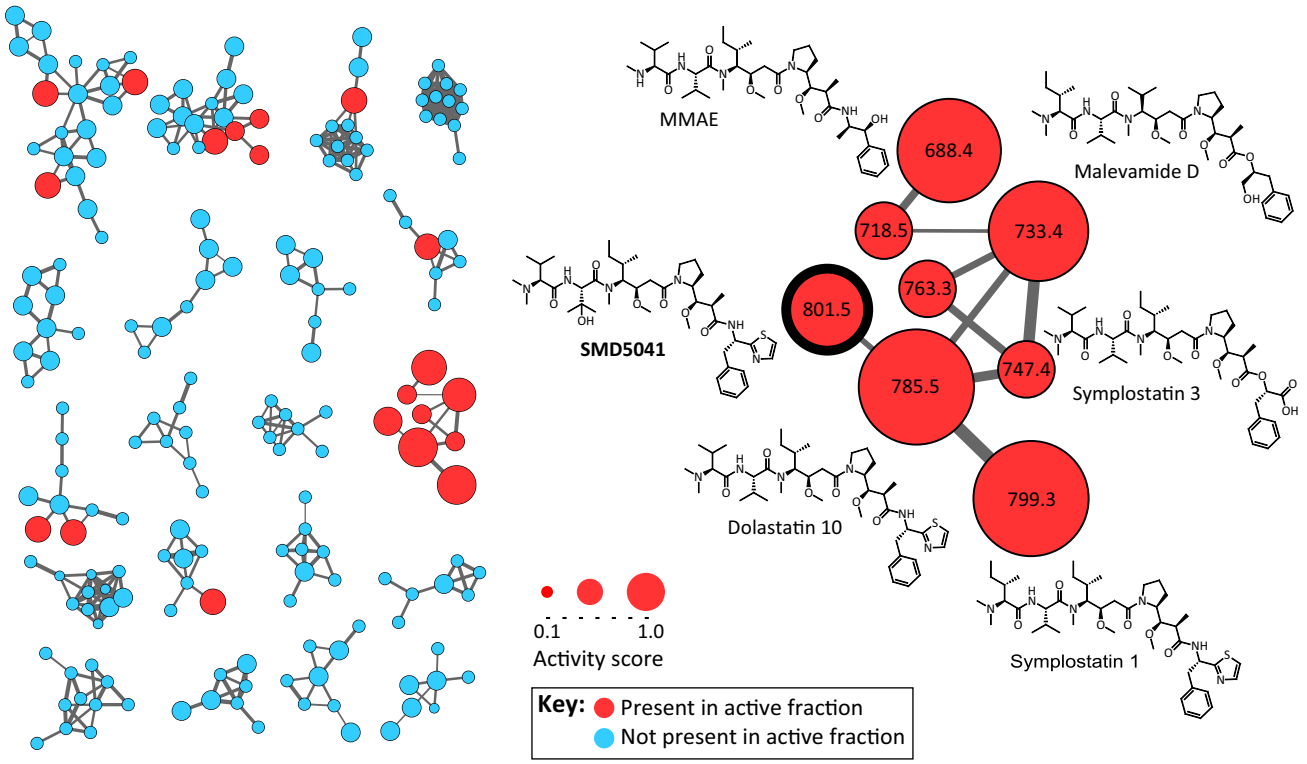


Figure 2. Molecular Networking in Natural Product Drug Discovery. Extracts of marine organisms are analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) and subjected to biological assay. Individual metabolites in the spectral network are colored according to their presence or absence in active samples (red and blue nodes, respectively). Statistical analysis of MS data for each sample is used to generate an 'activity score' represented as node size, with larger nodes corresponding to greater likelihood of contribution to observed biological activity. Expanded view of the dolastatin 10 tetrapeptide molecular family shows a new, highly active structural analog.

identification of biogenetic gene clusters of molecular families. The analysis of a biosynthetic pathway is the starting point for metabolic engineering efforts to optimize the production yield and is the biosynthetic platform for synthetic biology [38,39], such as that used for the industrial semisynthesis of artemisinin in *Saccharomyces cerevisiae* by Sanofi [40].

Molecular Networking to Visualize Medications and Drug Metabolism

Drug metabolism can significantly affect disease treatment because slight chemical changes of molecules administered as drugs can have profound physiological effects, including making an inactive drug active and vice versa. These transformations can be performed by host enzymes such as cytochromes [41], but more recently microbial enzymes from human-associated microbiota have been identified as mediators of drug metabolism [42–44]. This creates the potential for the discovery of an array of unique chemical structures of **xenobiotic** molecules generated by microbial and host enzymes and the novel application of molecular networking to aid pharmacology and medicine.

MS is currently a central tool in the study of drug metabolism [45,46]. Typically, **targeted MS** approaches are used for the quantitation of a drug and its metabolized products within biofluids or samples [47–49]. Although still at the proof-of-principle stage, the use of molecular networking for the discovery of drug metabolism was recently introduced [50]. This approach allows one to capture related molecules in an untargeted fashion, enabling the discovery of metabolized derivatives of compounds that could have been overlooked by a targeted approach in a complex

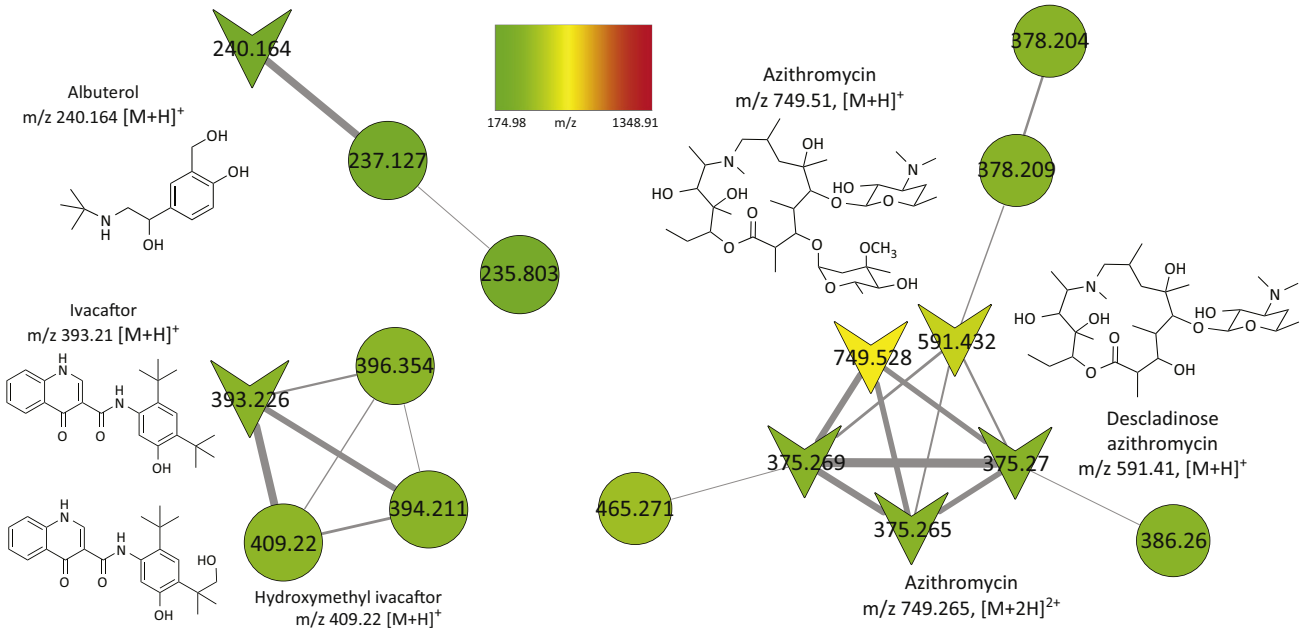


Figure 3. Detection of Medications and the Metabolism of Medications in Cystic Fibrosis Lungs. Molecular network clusters of known molecules and their metabolized derivatives are shown. Nodes are colored according to increasing mass-to-charge ratio (*m/z*) and shaped as arrowheads if they were known spectra annotated through GNPS or circles if they were unknown or inferred relatives of known drugs. Hydroxymethyl ivacaftor is produced by the liver enzyme cytochrome P450.

biological sample. For example, the diversity of xenobiotics in **cystic fibrosis (CF)** mucus has been described using molecular networking [50]. CF is a genetic disease that results in mucus accumulation in the lungs and susceptibility to chronic bacterial, fungal, and viral infections [51]. CF patients are administered myriad medicines, from antibiotics to anti-inflammatories to antidepressants. Many of these molecules were detected by untargeted LC-MS/MS analysis of lung mucus samples and annotated by spectral searching of GNPS libraries, which are largely populated with MS/MS spectra from commercial drugs [6]. Moreover, molecular networking was used to capture and annotate unknown biotransformed analogs. Figure 3 illustrates some of the chemical transformations seen in actual data generated from CF lung mucus. Albuterol, an anti-inflammatory drug, and antibiotics such as azithromycin show marked transformations, appearing as connected nodes in clusters containing known medicines. Some of these molecules represent previously observed breakdown and/or metabolic products of drugs, such as descladinose azithromycin, where a sugar is removed, and hydroxymethyl ivacaftor, a natural metabolic byproduct of the metabolism of this CF drug [52], but many of these nodes represent unknown metabolites. The LC-MS/MS and molecular networking approach is compatible with all sample types, including biopsies, fecal samples, urine, plasma or blood, and even skin [53]. Molecular networking should enable a more comprehensive understanding of drug metabolism and lead to the discovery of unknown molecular byproducts and metabolites of drugs that can help in deciphering their mechanism of action and side effects. This unique capability of molecular networking to visualize the chemical relatives of known pharmaceuticals represents a useful application in the pharmaceutical and medical fields.

The Potential of Molecular Networking in Clinical Diagnostics and Precision Medicine

Metabolic and infectious diseases result in an altered physiological state that can be detected through changes in small molecules. For example, elevated glucose and/or uric acid content has long been used as a signature of metabolic syndrome [54–56]. More recently, many diseases are

being diagnosed with tandem MS, such as inborn metabolic disorders of newborns including phenylketonuria, maple syrup urine disease, tyrosinemia, citrullinemia, and others [57]. Clinical outcomes in some infectious diseases, such as dengue fever, can be distinguished by numerous small molecules identified by LC-MS/MS [58], GC-MS detection of volatile metabolites in breath gas or breath condensate has seen great interest for the identification of infectious diseases of the airway [59], and in 2014 the FDA approved the clinical use of MS in typing of microbes [60]. MS is an especially attractive clinical tool for the detection of chemical changes associated with disease due to the speed of its use and the breadth of metabolites that can be screened. Its application has become commonplace in clinical diagnostics, to the extent that it is even being used in the operating room [3].

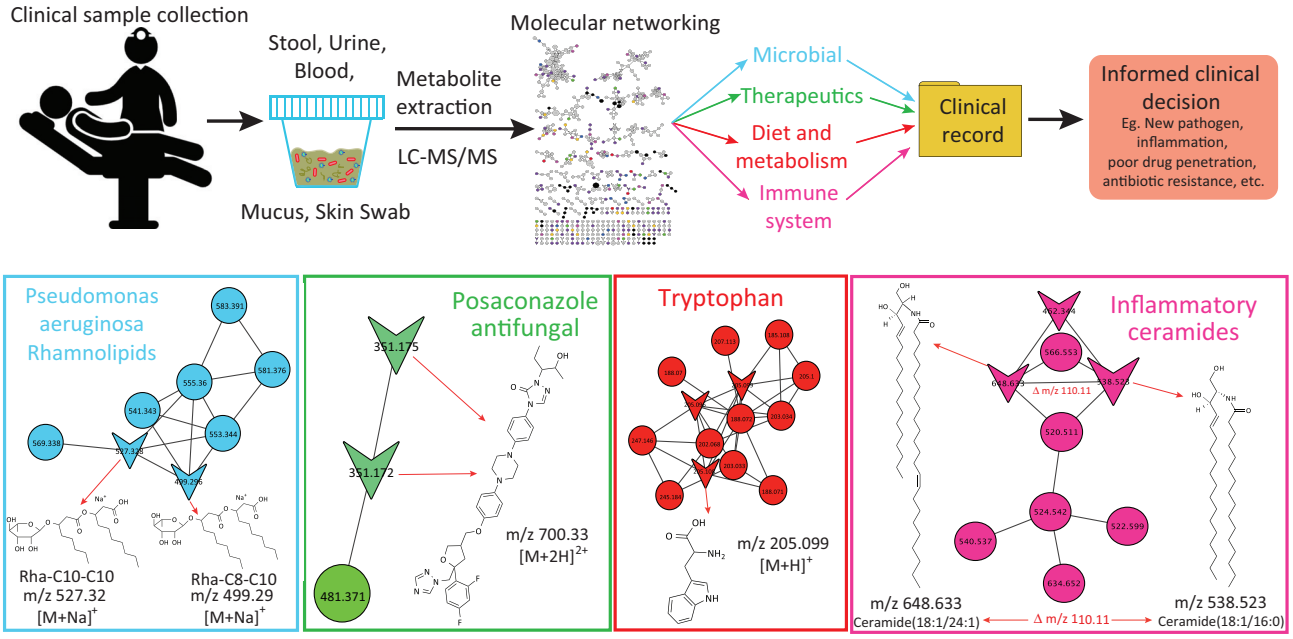
The potential power of MS-based diagnostics could be even greater because many of the spectral signatures in an MS experiment remain unknown, leaving their diagnostic capacity rarely explored. Untargeted fecal, urine, saliva, or skin MS data are very complex; within a single sample there are several thousand molecular signals. Molecular networking represents a tool with the ability to expand the potential of MS/MS-based diagnostics by visualizing the complete repertoire of chemical signals within a clinical sample and better aiding their annotation as biomarkers. Current MS approaches for clinical diagnostics target specific molecules; thus, different experimental procedures are required to detect multiple molecules of interest. Molecular networking can better visualize the known spectra in a single MS experiment and their relatives, including those from host, diet, medications, personal lifestyle, pathogens, and other exogenous sources [50]. Furthermore, molecular networking through GNPS can now be completed within hours of sample collection, making the timeframe for LC-MS/MS data generation and analysis clinically relevant [61]. The algorithm enables rapid interpretation of metabolites that are present in complex clinical samples and facilitates an easier path to the annotation of unknown molecular signatures that may be associated with a disease state.

CF is the first example in which molecular networking is being used to aid disease diagnostics. Some CF bacterial pathogens produce specialized metabolites detectable by molecular networking that can be used as potentially diagnostic markers of infection [9]. The CF bacterial pathogen *Pseudomonas aeruginosa* produces quorum-sensing molecules called quinolones, redox-active metabolites called phenazines, and the rhamnolipids, all of which can aid in the detection of this pathogen in CF patients [50,62,63]. Molecular networking allows the visualization of microbial metabolites, including related ones not normally screened for diagnostics. The approach has been used to identify microbial metabolites, host inflammatory markers, and drug metabolism in CF samples [50]. Other studies are beginning to use similar approaches to identify metabolites specific to particular pathogens as biomarkers for infection, including the agent of tuberculosis [64]. The method has also been employed to analyze the chemical repertoire on human skin, identifying the microbial metabolism of host compounds [53]. The current challenge with this approach is that many bacterial pathogens do not have a well-characterized specialized metabolome, another area in which the algorithm could aid the field. However, some are beginning to use molecular networking to help annotate the specialized metabolites produced by important human-associated microorganisms. Colibactins, for example, are small molecules produced by *Escherichia coli* that were found to be associated with increased inflammation and colon cancer risk [65]. Molecular networking was used to explore their biosynthesis and mode of action. Clearly, better annotation and documentation of the metabolites produced by important human-associated or pathogenic microbes, whether known or unknown, will greatly aid the application of molecular networking as diagnostic tool for microbial infections and other clinical manifestations that are microbe associated.

Precision medicine is a patient management approach that takes into account the individual patient's phenotypic and genotypic disease in the context of his or her environment and lifestyle

to better tune treatments towards the individual. A new initiative on precision medicine was launched by the US Government in 2015 aiming to make this approach a common method of treating disease [66]. The expansion of precision medicine has been in part due to the development of omics approaches, such as genomics and metabolomics, and the availability of databases in which to compare data from an individual with those from all others with similar disease characteristics [66]. We speculate that molecular networking has significant potential for use as a precision medicine tool because it can serve as a general scanning approach for detecting, visualizing, and identifying the molecules on and in a patient in the context of his or her disease. Chemical changes within individuals can point to aspects of their particular disease and inform a clinician about drug metabolism, drug penetration, inflammatory load, metabolic activity, microbial infections, and other medically relevant aspects of a patient's metabolome [50]. Molecular networking can aid the translation of MS data to precision medicine because it better enables data visualization and the identification of molecules associated with a specific individual's disease state or pathogens of interest [50]. The visual aspects of molecular networking are especially well designed for detecting changes in metabolite signatures in a personalized manner, because nodes can be highlighted by various metadata attributes making it easy to identify molecules unique to an individual during times of disease or symptom worsening [67]. The visual nature of the node-mapping approach can also aid longitudinal studies of patient health, as one can easily observe changes in a patient through time as unique nodes in a molecular network.

An example of a molecular networking approach to precision medicine is shown in Figure 4, using actual data from a CF patient to illustrate the method's potential. Clinical samples, in this case sputum from a CF lung, can be collected at the bedside and then extracted with various solvents in the laboratory. MS/MS acquisition of metabolite signatures of the sample can be completed within minutes using today's mass spectrometers and molecular networks can be



Trends in Pharmacological Sciences

Figure 4. Implementation of Molecular Networking in Clinical Medicine. Samples of any type are collected and subjected to metabolite extraction and liquid chromatography–tandem mass spectrometry (LC-MS/MS). Molecular networks can then be mined based on metadata for microbial, inflammatory, drug metabolism, and other chemical signatures. Clusters of molecules reveal microbial, host, and drug chemistry, including virulence factors such as rhamnolipids, drug penetration at the site of action, and host metabolism and inflammatory metabolites, such as ceramides and other bioactive lipids.

built within minutes to hours of data acquisition via GNPS. In this example a patient presents with an increase in symptom severity and the clinicians are interested in identifying the signatures of metabolites related to microbial infections, inflammatory load, basic metabolism, and therapeutic penetration. In this example *P. aeruginosa* rhamnolipids are detected, indicating a potential role of this bacterium in the malady being assessed. Fungicides given to the patient are also observed, indicating that these drugs are present at the site of infection where they are meant to act but may not be effective in this instance if the infection is primarily bacterial. Bioactive lipids indicate a possibly heightened inflammatory response, telling the clinician that steroid or other anti-inflammatory therapies may be needed. This example illustrates how molecular networking can provide a chemical picture of a patient's disease, aiding clinical decisions for personalized medicine.

Concluding Remarks and Future Perspectives

Although molecular networking has only recently been implemented in drug discovery and metabolomics following the first paper on microbial metabolite networks in 2012 [5], the breadth of its applications has been ever expanding and will continue to increase now that GNPS is available [6]. There are two main bottlenecks preventing molecular networking from reaching its full potential (see Outstanding Questions). The first is the lack of efficient integration with existing LC-MS detection tools. The second is the remaining challenge in annotating detected MS/MS spectra. Because GNPS has a community-based platform, annotations are crowd sourced, taking advantage of the knowledge available in the MS community. This is beginning to increase the number of annotations, but better informatics tools are still required. Recent tools are now able to propose the correct molecular formula in automated manner with a high success rate [68] and tools that provide insight into candidate molecules based on *in silico* approaches, like CSI: FingerID [69], MetFrag [70], ISDB-UNPD [71], and CFM-ID [72], are advancing rapidly. The expected union of these approaches in one comprehensive and accessible workflow will unleash a tremendous amount of chemical information on our biological world. Ultimately, this will accelerate the discovery of new drugs by revealing the chemical language of an entire microbial ecosystem. Moreover, the visual nature of molecular networks enables rapid interpretation of MS data, crucial for translation in a clinically relevant time frame [61].

As MS methods become increasingly more tractable and amenable to direct clinical application, molecular networking will become an indispensable tool for such translation. To truly harness the power of this tool, scientists from a broad range of fields must be encouraged to use it. Fundamentally, one must understand that molecular networking organizes and visualizes the chemical information from MS/MS data and that its potential applications far exceed the chemistry laboratory and MS experts. Nonetheless, it should be taken into account that the interpretation of molecular networking cannot surpass classical challenges in MS and metabolomics. Thus, the novice must collaborate with MS experts and educate themselves [73] to avoid common pitfalls that can occur upstream during sample preparation (polymers released from plastics or the matrix effect [74,75]), batch design (batch effects [76] and carryover [77]), and MS data acquisition (in-source fragmentation, ionization method, fragmentation behavior of molecules, etc.) and downstream during data analysis, such as adjustment of molecular networking parameters (https://bix-lab.ucsd.edu/display/Public/Molecular+Networking+Documentation), validation of automatic annotation, and chimeric spectra resulting from isobaric compounds [78].

Now that molecular networking is readily accessible through GNPS, the true power of the algorithm can be better realized. The billions of spectra generated with MS/MS around the world can be now compared with each other to identify novel compounds related to known drugs and completely new chemical families. Integration with other bioinformatics tools such as those described above will be a methodological tipping point in natural product discovery,

Outstanding Questions

Can molecular networking help us better visualize the chemical communication between organisms that are clinically relevant? Who produces what is a fundamentally hard question when investigating human biology and its interface with the microbiome. A major area of interest in molecular networking is its ability to help us visualize the chemical crosstalk within complex biological samples and to track the spectral signatures to their microbial or host origin.

What information is contained in a molecular network that has not yet been mined? There is potential to use molecular networking to provide a systems-level view of biological transformations or enzyme activity within a biological system. While this approach has not yet been systematized, it could be a novel method to monitor pathway and enzymatic activity within a biological system by looking at the atomic differences between MS/MS spectra of molecules detected to infer the chemical reactions involved (e.g., acetylation, methylation, phosphorylation, lipid chain saturation and desaturation)

Can molecular networking census the world's medical chemistry as a stepping stone towards chemical epidemiology? Broad use of the molecular networking algorithm through publically available databases such as GNPS will provide a census of the chemistry of our planet and allow cross-comparison between datasets to ask broad questions about ubiquity, prevalence, correlation, and the strength of association between particular molecules and various human diseases.

metabolomics-based clinical screening, and medicine, finally harnessing the power of the highly advanced mass spectrometers available to scientists and clinicians today.

Disclaimer Statement

O.V. and E.E. are employees of Sirenas and P.C.D. is an advisor to this company, which uses molecular networking for the discovery of molecules of marine origin in its drug development pipeline.

References

1. Horning, E.C. and Horning, M.G. (1970) Metabolic profiles: chromatographic methods for isolation and characterization of a variety of metabolites in man. *Methods Med. Res.* 12, 369–371
2. Pauling, L. *et al.* (1971) Quantitative analysis of urine vapor and breath by gas–liquid partition chromatography. *Proc. Natl. Acad. Sci. U. S. A.* 68, 2374–2376
3. Balog, J. *et al.* (2013) Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci. Transl. Med.* 5, 194ra93
4. da Silva, R.R. *et al.* (2015) Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12549–12550
5. Watrous, J. *et al.* (2012) Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1743–1752
6. Wang, M. *et al.* (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837
7. Guthals, A. *et al.* (2012) The spectral networks paradigm in high throughput mass spectrometry. *Mol. Biosyst.* 8, 2535–2544
8. Traxler, M.F. and Kolter, R. (2012) A massively spectacular view of the chemical lives of microbes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 10128–10129
9. Moree, W.J. *et al.* (2012) Interkingdom metabolic transformations captured by microbial imaging mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 109, 13811–13816
10. Vizcaino, M.I. and Crawford, J.M. (2016) Secondary metabolic pathway-targeted metabolomics. *Methods Mol. Biol.* 1401, 175–195
11. Li, D. *et al.* (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4147–E4155
12. Edlund, A. *et al.* (2015) Meta-omics uncover temporal regulation of pathways across oral microbiome genera during *in vitro* sugar metabolism. *ISME J.* 9, 2605–2619
13. Altschul, S.F. *et al.* (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410
14. Tautenhahn, R. *et al.* (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* 84, 5035–5039
15. Giacomoni, F. *et al.* (2015) Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* 31, 1493–1495
16. Xia, J. *et al.* (2015) MetaboAnalyst 3.0 – making metabolomics more meaningful. *Nucleic Acids Res.* 43, W251–W257
17. Johnson, C.H. *et al.* (2015) Bioinformatics: the next frontier of metabolomics. *Anal. Chem.* 87, 147–156
18. Misra, B.B. and van der Hooft, J.J.J. (2016) Updates in metabolomics tools and resources: 2014-2015. *Electrophoresis* 37, 86–110
19. Alonso, A. *et al.* (2015) Analytical methods in untargeted metabolomics: state of the art in 2015. *Front. Bioeng. Biotechnol.* 3, 23
20. Frank, A.M. *et al.* (2011) Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. *Nat. Methods* 8, 587–591
21. Frank, A.M. *et al.* (2008) Clustering millions of tandem mass spectra. *J. Proteome Res.* 7, 113–122
22. Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504
23. Nguyen, D.D. *et al.* (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2611–E2620
24. Garg, N. *et al.* (2014) Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. *Int. J. Mass Spectrom.* 377, 719–727
25. Yang, J.Y. *et al.* (2013) Molecular networking as a dereplication strategy. *J. Nat. Prod.* 76, 1686–1699
26. Cragg, G.M. and Newman, D.J. (2013) Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* 1830, 3670–3695
27. Liaw, C-C. *et al.* (2015) Vitroprocines, new antibiotics against *Acinetobacter baumannii*, discovered from marine *Vibrio* sp. QWI-06 using mass-spectrometry-based metabolomics approach. *Sci. Rep.* 5, 12856
28. Bertin, M.J. *et al.* (2015) Spongosine production by a *Vibrio harveyi* strain associated with the sponge *Tectitethya crypta*. *J. Nat. Prod.* 78, 493–499
29. Liu, W-T. *et al.* (2014) MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *J. Antibiot. (Tokyo)* 67, 99–104
30. Kleigrewe, K. *et al.* (2015) Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J. Nat. Prod.* 78, 1671–1682
31. Duncan, K.R. *et al.* (2015) Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem. Biol.* 22, 460–471
32. Okada, B.K. *et al.* (2016) Mapping the trimethoprim-induced secondary metabolome of *Burkholderia thailandensis*. *ACS Chem. Biol.* 11, 2124–2130
33. Perez, H.L. *et al.* (2014) Antibody–drug conjugates: current status and future directions. *Drug Discov. Today* 19, 869–881
34. Pettit, G.R. *et al.* (1987) The isolation and structure of a remarkable marine animal antineoplastic constituent: dolastatin 10. *J. Am. Chem. Soc.* 109, 6883–6885
35. Harrigan, G.G. *et al.* (1998) Symplostatin 1: a dolastatin 10 analogue from the marine cyanobacterium *Symploca hydnoides*. *J. Nat. Prod.* 61, 1075–1077
36. Luesch, H. *et al.* (2002) Symplostatin 3, a new dolastatin 10 analogue from the marine cyanobacterium *Symploca* sp. VP452. *J. Nat. Prod.* 65, 16–20
37. Horgen, F.D. *et al.* (2002) Malevamide D: isolation and structure determination of an isodolastatin H analogue from the marine cyanobacterium *Symploca hydnoides*. *J. Nat. Prod.* 65, 487–491
38. Harvey, A.L. *et al.* (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* 14, 111–129
39. Kim, E. *et al.* (2015) Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* 11, 649–659
40. Paddon, C.J. and Keasling, J.D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* 12, 355–367
41. Guengerich, F.P. (2006) Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J.* 8, E101–E111
42. Wallace, B.D. *et al.* (2015) Structure and inhibition of microbiome β-glucuronidases essential to the alleviation of cancer drug toxicity. *Chem. Biol.* 22, 1238–1249

43. Spanogiannopoulos, P. *et al.* (2016) The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* 14, 273–287
44. Haiser, H.J. *et al.* (2013) Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341, 295–298
45. Prakash, C. *et al.* (2007) Analytical strategies for identifying drug metabolites. *Mass Spectrom. Rev.* 26, 340–369
46. Kind, T. and Fiehn, O. (2010) Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* 2, 23–60
47. Kostiainen, R. *et al.* (2003) Liquid chromatography/atmospheric pressure ionization–mass spectrometry in drug metabolism studies. *J. Mass Spectrom.* 38, 357–372
48. Ma, S. and Zhu, M. (2009) Recent advances in applications of liquid chromatography–tandem mass spectrometry to the analysis of reactive drug metabolites. *Chem. Biol. Interact.* 179, 25–37
49. Tolonen, A. *et al.* (2009) Liquid chromatography–mass spectrometry in *in vitro* drug metabolite screening. *Drug Discov. Today* 14, 120–133
50. Quinn, R.A. *et al.* (2015) Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *ISME J.* 10, 1483–1498
51. Lynch, S.V. and Bruce, K.D. (2013) The cystic fibrosis airway microbiome. *Cold Spring Harb. Perspect. Med.* 3, a009738
52. Robertson, S.M. *et al.* (2015) Clinical drug–drug interaction assessment of ivacaftor as a potential inhibitor of cytochrome P450 and P-glycoprotein. *J. Clin. Pharmacol.* 55, 56–62
53. Bouslimani, A. *et al.* (2015) Molecular cartography of the human skin surface in 3D. *Proc. Natl. Acad. Sci. U. S. A.* 112, E2120–E2129
54. Facchini, F. *et al.* (1991) Relationship between resistance to insulin-mediated glucose uptake, urinary uric acid clearance, and plasma uric acid concentration. *JAMA* 266, 3008–3011
55. Matsuura, F. *et al.* (1998) Effect of visceral fat accumulation on uric acid metabolism in male obese subjects: visceral fat obesity is linked more closely to overproduction of uric acid than subcutaneous fat obesity. *Metabolism* 47, 929–933
56. Yoo, T.W. *et al.* (2005) Relationship between serum uric acid concentration and insulin resistance and metabolic syndrome. *Circ. J.* 69, 928–933
57. Chace, D.H. *et al.* (2003) Use of tandem mass spectrometry for multianalyte screening of dried blood specimens from newborns. *Clin. Chem.* 49, 1797–1817
58. Voge, N.V. *et al.* (2016) Metabolomics-based discovery of small molecule biomarkers in serum associated with dengue virus infections and disease outcomes. *PLoS Negl. Trop. Dis.* 10, e0004449
59. Whiteson, K.L. *et al.* (2014) Breath gas metabolites and bacterial metagenomes from cystic fibrosis airways indicate active pH neutral 2,3-butanedione fermentation. *ISME J.* 8, 1247–1258
60. Singhal, N. *et al.* (2015) MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front. Microbiol.* 6, 791
61. Quinn, R.A. *et al.* (2016) From sample to multi-omics conclusions in under 48 hours. *mSystems* 1, e00038-16
62. Hunter, R.C. *et al.* (2012) Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. *Am. J. Respir. Cell Mol. Biol.* 47, 738–745
63. Singh, P.K. *et al.* (2000) Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* 407, 762–764
64. Lau, S.K.P. *et al.* (2015) Identification of specific metabolites in culture supernatant of *Mycobacterium tuberculosis* using metabolomics: exploration of potential biomarkers. *Emerg. Microbes Infect.* 4, e6
65. Vizcaino, M.I. and Crawford, J.M. (2015) The colibactin warhead crosslinks DNA. *Nat. Chem.* 7, 411–417
66. Collins, F.S. and Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793–795
67. Quinn, R.A. *et al.* (2016) Metabolomics of pulmonary exacerbations reveals the personalized nature of cystic fibrosis disease. *PeerJ* 4, e2174
68. Böcker, S. and Dührkop, K. (2016) Fragmentation trees reloaded. *J. Cheminform.* 8, 5
69. Dührkop, K. *et al.* (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12580–12585
70. Ruttkies, C. *et al.* (2016) MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation. *J. Cheminform.* 8, 3
71. Allard, P-M. *et al.* (2016) Integration of molecular networking and *in-silico* MS/MS fragmentation for natural products dereplication. *Anal. Chem.* 88, 3317–3323
72. Allen, F. *et al.* (2014) CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* 42, W94–W99
73. Greaves, J. and Roboz, J. (2013) *Mass Spectrometry for the Novice,* CRC Press
74. Vuckovic, D. (2012) Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal. Bioanal. Chem.* 403, 1523–1548
75. Bylda, C. *et al.* (2014) Recent advances in sample preparation techniques to overcome difficulties encountered during quantitative analysis of small molecules from biofluids using LC-MS/MS. *Analyst* 139, 2265
76. Wehrens, R. *et al.* (2016) Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* 12, 88
77. Koal, T. and Deigner, H-P. (2010) Challenges in mass spectrometry based targeted metabolomics. *Curr. Mol. Med.* 10, 216–226
78. Nothias-Scaglia, L-F. *et al.* (2015) Molecular networks, a comprehensive bioinformatics approach to interpreting tandem mass spectrometry data. *Spectra Anal.* 307, 73–78