Untargeted metabolomics for the study of antiinfective plants

Joshua J. Kellogg
Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, United States

Introduction

Natural products represent a tremendously rich resource for drug discovery paradigms or as dietary supplements to improve human health. They arise from a variety of sources, including bacteria, fungi, marine organisms, and plants (Newman & Cragg, 2020). In their natural state, natural products are complex mixtures containing many structurally diverse components (Enke & Nagels, 2011). For the majority of natural products, the identities of potential bioactive constituents are often not known, and also likely to differ depending on the biological activity evaluated (Alvarez-Zapata et al., 2015; Simmler et al., 2016). A primary obstacle facing researchers is unraveling the intricacies of such complex mixtures and assigning structures and biological activities to the components. While traditional methods such as bioassay-guided fractionation (BGF) have had decades of success, newer methodologies based upon metabolomics are poised to delve deeper into the chemical make-up of natural products and are able to couple with biological activity evaluations in order to enhance bioactive compound discovery efforts. With this chapter, we review various strategies that can be employed and discuss their advantages and disadvantages in hunting for potential antiinfective compounds from plants.

Plants as sources of antiinfective agents

Natural products have long served as the foundation for antiinfective bioactive molecule discovery, though the majority of discoveries have arrived from microbial natural sources (Newman & Cragg, 2020). However, as analytical instrumentation has improved, along with the implementation of high-throughput techniques that can accommodate...
more complex botanical samples (Harvey, Edrada-Ebel, & Quinn, 2015; Shen, 2015), plants have gained heightened attention for their potential bioactive secondary metabolites. Indeed, plants hold significant interest as potential sources of antiinfective agents due to their long history of biochemical defenses against microbial pathogens, (Adedeji & Babalola, 2020; Zaynab et al., 2018). Plants also possess a diverse array of structurally unique secondary metabolites from the estimated 374,000 species worldwide (Christenhusz & Byng, 2016), many of which are prominently featured in the traditional ecological knowledge and medicinal pharmacopeia of communities across the globe (Porras et al., 2020; Willis, 2017). However, with such diversity of both species and chemistry, accessing the relevant bioactive metabolites from a complex matrix can be challenging. Traditionally, iterative separative techniques have been employed to try and tease out the molecular underpinnings of bioactivity.

**Bioassay-guided fractionation**

Identifying novel chemical structures from natural sources can be a complex and lengthy process. The long-standing approach to isolation and identification of bioactive constituent(s) from plants (or other natural sources) is a robust repertoire of techniques broadly referred to as “bioassay-guided fractionation” (BGF) (Kinghorn et al., 1998). Bioassay-guided fractionation is an iterative process, with successive rounds of chromatographic separation and bioassay until a single bioactive entity (i.e., an isolated pure compound) is achieved (Fig. 10.1). This approach has remained the “gold standard” of bioactive discovery from natural resources for decades (Selander et al., 2015) and has resulted in the discovery of a host of critically important pharmaceutical agents, including camptothecin and taxol (paclitaxel) (Oberlies & Kroll, 2004; Wani, Taylor, Wall, Coggon, & McPhail, 1971), artemisinin (Tu, 2011), and vinblastine (Noble, 1990). However, a bioassay-guided fractionation approach possesses several limitations. It is possible that, at the end of multiple rounds of chromatographic separation on different resins and orthogonal techniques, isolation of a single bioactive agent fails due to chemical degradation on the chromatography matrices during separations, or “disappears” due to irreversible binding to the resin (Qiu et al., 2013). In addition, it is often possible that the observed activity evidenced by a complex sample is due to multiple compounds working in concert (e.g., synergy or additivity), and the biological effect is not observed once the purification separates them (Caesar & Cech, 2019). Because the process of refractionation and iterative bioassay is time-, labor-, and resource-consuming, some have begun to consider this approach too risky to pursue (Li & Vederas, 2009). There has thus been considerable interest in developing approaches that can help identify candidate metabolites earlier in the isolation workflow to prioritize and guide isolation efforts.

**Metabolomics**

The metabolome is defined as the complement of low-molecular-weight metabolites (ca. <1200 Da) present in a sample (e.g., environmental sample, biological fluid, cell, or organism) representing a snapshot of the system at a particular set of physiological
conditions. This could represent a multitude of different possible stimuli, including genetic variations, pathogenic infestation, or other abiotic or biotic stresses. Similarly, the goal of metabolomics is the chemical profiling of a phenotype through the qualitative and/or quantitative analysis of all measurable metabolites in a complex system and the measurement of the change in the metabolite profile due to some type of challenge or perturbation (Oliver, Winson, Kell, & Baganz, 1998). In ca. two decades since the term “metabolomics” was coined (Oliver et al., 1998), metabolomics has developed into an important analytical tool for a wide variety of applications, including studies on diseases (Gowda et al., 2008), toxicity (Ramirez et al., 2018), quality control (Shu et al., 2017), and natural products (Kellogg, Paine, McCune, Oberlies, & Cech, 2019).

Metabolomics strategies can be broadly divided into untargeted and targeted approaches, each with their own advantages and limitations. Untargeted metabolomics focuses on the analysis of all detectable metabolites within a sample, regardless of whether the chemical identity can be ascertained. However, the presence of such a large number of chemical constituents (especially unknown metabolites) makes quantification difficult (Cajka & Fiehn, 2016).
By contrast, targeted metabolomics centers around the measurement of a defined group of metabolites, often with known chemical structures. While this does not provide as wide metabolite coverage as untargeted approaches, targeted methods have the potential to reliably quantify the targeted group of metabolites, via internal standards or multipoint calibration curves (Fig. 10.2). This allows for more quantitative measurements of the metabolome subset.

### Methods of detection

There are a variety of analytical instruments currently employed to provide a chemical profile that forms the basis for metabolomics analyses. Studies have been carried out using several forms of spectroscopy, such as ultraviolet-visible spectroscopy (Anđelković et al., 2017) and Fourier-transformed infrared spectroscopy (Liu, Finley, Betz, & Brown, 2018). However, the two main analytical techniques currently employed for metabolomic studies are mass spectrometry (MS) and nuclear magnetic resonance (NMR). NMR-based metabolomic acquisition offers an unbiased assessment of a complex sample’s composition, able to measure all metabolites therein, allowing for the simultaneous identification and quantification of diverse metabolites (Pauli et al., 2014). MS-based metabolomic methods have the advantages of greater sensitivity compared to NMR spectroscopy by several orders of magnitude and the ability to couple directly to separation methods such as gas chromatography (GC) or liquid chromatography (LC). However, a disadvantage of MS analysis is that ionization is required to detect sample components, yet not every molecule is ionized in a mass spectrometer (Cech & Yu, 2013), and those metabolites which are ionized are not universally uniformly ionized, which complicates quantification of metabolites in the sample. The advancements in these two analytical techniques have driven many developments in metabolomics profiling, and modern systems are able to provide exceptional sensitivity,

---

**FIGURE 10.2** Untargeted and targeted metabolomics approaches. Untargeted and targeted metabolomics approaches differ in two critical areas: the number of metabolites detected (x-axis) and the ability to quantify the observed metabolites (y-axis). Source: From Cajka, T., & Fiehn, O. (2016). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Analytical Chemistry, 88, 524–545. https://doi.org/10.1021/acs.analchem.5b04491.
signal-to-noise ratio, and levels of structural characterization. However, the purpose of this chapter is not to extensively review analytical techniques; the relative advantages of these approaches have been discussed elsewhere (Emwas et al., 2019; Jorge, Mata, & António, 2016; Markley et al., 2017; Moco et al., 2007).

Data analysis

Metabolomic analysis results in the generation of large datasets comprising both major and minor components from the sample, which requires statistical tools to parse the data and provide significant analysis. The most employed statistical tool for characterization and comparison of metabolomics data sets is the unsupervised method principal component analysis (Rajalahti & Kvalheim, 2011), while supervised methods for statistical analysis include soft independent modeling by class analogy (Wallace, Todd, Harnly, Cech, & Kellogg, 2020). However, it should be noted that these analyses consider only the chemical data from metabolomic profiling. Incorporating a dependent variable, such as bioactivity, requires different multivariate statistical analyses [e.g., partial least squares (PLS); see below for more details].

Many metabolomics studies focus on identifying potentially relevant features from a complex profile (or eliminating potentially irrelevant signals or noise). Advances in statistical analysis allow metabolomics studies to be utilized in more detailed or complex inquiries. One recent informatics tool that allows for these new inquiries is molecular networking, where molecules are clustered based upon their fragmentation MS/MS (aka MS², or MS⁹) ion patterns (Watrous et al., 2012; Yang et al., 2013). The MS/MS fragmentation ions from each precursor feature are compared against each other and assigned a similarity score (“cosine score”); the underlying principle is that molecules with similar backbone structures contain comparable substructures upon fragmentation. The data are converted into a network diagram, with nodes representing MS/MS spectra of precursor ions at distinct m/z values, while the connections between nodes indicate a similarity in the fragmentation pattern (cosine score). Nodes are clustered based upon potential structural similarities as well as specific chemical properties (Watrous et al., 2012). These clusters can be supplemented with standards or database entries to facilitate annotation and structural identification (Guthals, Watrous, Dorrestein, & Bandeira, 2012).

Biochemometrics

The profiling characteristics of metabolomics studies possess several advantages when screening for bioactive metabolites, compared to the traditional bioassay-guided fraction approach (Calderon, 2017; Roberts et al., 2019). First, metabolomic profiling does not require multiple iterative separation and purification steps, and thus unstable compounds are more likely to be detected and measured. In addition, since there is less emphasis on chromatographic separation of the sample pre-analysis, additive or synergistic effects are more likely to be detected (Britton, Kellogg, Kvalheim, & Cech, 2018; Caesar & Cech, 2019). In addition, even in untargeted approaches, qualitative or relative quantitative differences in the metabolomic profiles of different samples can suggest potential chemical...
signals underlying the observed phenotypic changes (Prince & Pohnert, 2010). Thus, metabo-
olomic fingerprinting has increasingly been integrated with bioactivity data to statistically model changing behavior as the chemical composition varies across different taxa, products, or fractions. These methods, collectively termed “biochemometrics” (Kellogg et al., 2016; Martens, Bruun, Adt, Sockalingum, & Kohler, 2006), have become a primary driver in improving the efficiency of bioactive molecule discovery from natural products and other resources (Wyss, Llivina, & Calderón, 2019). For the integration of these two disparate datasets into a single predictive model, multivariate statistical methods are required. Several different statistical approaches have been employed for this purpose, including Pearson correlations (Inui, Wang, Pro, Franzblau, & Pauli, 2012; Nothias et al., 2018; Richards et al., 2018), PLS (Britton et al., 2018; Kellogg et al., 2016; Kvalheim et al., 2011), PLS-discriminant analysis (PLS-DA, OPLS-DA) (Alvarez-Zapata et al., 2015; Chagas-Paula, Zhang, Da Costa, & Édrada-Ebel, 2015; Wen et al., 2018), and hierarchical cluster analysis (Patras et al., 2011). PLS has emerged as one of the foremost statistical modeling approaches for biochemometrics studies, and often generates multidimensional models, which can be difficult to deconstruct. To aid in the interpretation of the model, several visualization metrics and plots have been developed to interpret PLS models, with the variable importance in projection (VIP) method, the S-plot, and the selectivity ratio being the leading metrics (Farrés, Platikanov, Tsakovski, & Tauler, 2015; Kellogg et al., 2016; Rajalahti, Arneberg, Berven et al., 2009; Rajalahti, Arneberg, Kroksveen et al., 2009).

Metabolomics-driven antiinfective discovery from plants

The versatility and robustness of metabolomic and biochemometric analyses have improved the discovery of phytochemicals that target various infective agents without the intensive process of BGF. From the Pauli group, one study used Pearson correlation to identify constituents from the Alaskan ethnobotanical Oplopanax horridus (Sm.) Miq. (devil’s club) possessing antituberculosis activity. Biochemometric analysis using orthogonal chromatographic methods (countercurrent and GC) coupled to MS identified the c. 100 most active constituents from the plant extract through Pearson correlations, which subsequently resulted in the annotation of 29 bioactive structures from three dominant structural classes which were present in the 19 initial fractions produced (Inui et al., 2012; Li et al., 2013; Qiu et al., 2013) (Fig. 10.3). Another respiratory-focused experiment investigated the antitussive and expectorant properties of Tussilago farfara, a traditional herbal medicine found in both European and Chinese pharmacopeia. This study examined multiple tissue types from T. farfara, including roots, leaves, and flower buds, and combined an in vivo expectorant test with 1H NMR metabolomic profiling and PCA analysis. It was revealed that the roots had no discernable antitussive or expectorant effects, while the leaves and flower buds both possessed efficacy. The biochemometric analysis from the untargeted metabolome indicated significantly different chemical profiles between the leaves, flower buds, and roots and highlighted three metabolites correlating with active constituents: chlorogenic acid, 3,5-dicaffeoylquinic acid, and rutin (Li et al., 2013).

Metabolomic profiling of essential oils has also revealed potential antibacterial compounds. Torch ginger (Etlingera elatior (Jack) R.M.Sm.) is an edible plant rich in
phytochemicals with well-documented pharmacological properties (Chan, Lim, & Wong, 2011). Torch ginger flower oil was subjected to metabolomics profiling using both gas chromatography-mass spectrometry (GC-MS) and $^1$H-NMR approaches. Evaluating the antibacterial activity of the torch ginger flower oil via an agar diffusion assay demonstrated strong antibacterial activity against Gram-negative and Gram-positive bacterial strains Salmonella typhimurium, Staphylococcus aureus, and Escherichia coli, with sub-mg/mL MIC. The metabolomics analysis revealed 33 compounds using GC-MS, 15 of which were previously known for their antimicrobial activity. In addition, 16 metabolites were identified from the $^1$H-NMR analysis and eight of those had antibacterial activity (Anzian et al., 2020). This study highlighted the potential benefits of using multiple analytical instruments to broaden coverage of the metabolome. Cinnamomum camphora (L.) J. Presl is one of the oldest herbal medicines used as a traditional medicine, and its essential oil contains a multitude of potentially bioactive compounds. The essential oil from cinnamon was profiled by GC-MS, and the methicillin-resistant Staphylococcus aureus (MRSA) metabolic profile in the presence of the essential oil was also analyzed with GC-MS-based metabolomics. Potential metabolites with antibacterial activity were believed to be linalool, eucalyptol, $\alpha$-terpineol, isoborneol, $\beta$-phellandrene, and camphor (Chen et al., 2020). This study was also of note as it used metabolomics on both sides of the discovery process; profiling not only the botanical source of the antiinfective compound(s) but also looking at the response from the infectious agent, finding that 74 bacterial metabolites demonstrated significant differences (including 29 upregulated and 45 downregulated metabolites) across seven metabolic pathways (Chen et al., 2020).

Two recent studies looked at gastrointestinal nematodiasis, which can have substantial effects on ruminant health, and is experiencing a rise in resistance to commonly used antihelmintics.

FIGURE 10.3 Biochromatogram of a fractionated O. horridus (Sm.) Miq. extract. The x-axis represents the countercurrent chromatographic fractions, while the y-axis indicates the antituberculosis activities of the fractions. The bioactivity observed at 50 $\mu$g/mL (black line) yielded 19 biopeaks (shaded), which represent the active components from devil’s club and underwent further GC-MS analysis and biochemometrics using Pearson’s correlation. Source: From Inui, T., Wang, Y., Pro, S.M., Franzblau, S.G., & Pauli, G.F. (2012). Unbiased evaluation of bioactive secondary metabolites in complex matrices. Fitoterapia, 83, 1218–1225.
The first study used $^1$H-NMR metabolomics to provide preliminary identification of antihelminthic compounds from *Lysiloma latisiliquum* (L.) Benth., an ethnoveterinary plant from Mexico (Hernández-Bolio, Ruiz-Vargas, & Peña-Rodríguez, 2019). A study by Hernández-Bolio, Kutzner, Eisenreich, de Jesús Torres-Acosta, and Peña-Rodríguez (2018) used OPLS-DA to identify the glycosylated compounds quercitrin and arbutin as possessing activity against the nematode *Haemonchus contortus* (Hernández-Bolio, Kutzner, Eisenreich, de Jesús Torres-Acosta, & Peña-Rodríguez, 2018). In addition, Borges et al. (2019) investigated the ovicidal activity of ethanol extracts from 17 plants collected from the Pantanal wetland in the state of Mato Grosso do Sul, Brazil. These plants were evaluated using the egg hatchability test of *Haemonchus placei*, and ethanol extracts were analyzed via HPLC-MS with Partial least squares regression discriminate analysis (PLS-DA) (Borges et al., 2019; Peña-Espinoza et al., 2020) (Fig. 10.4) and a univariate

correlation mapping was used to detect compounds that positively correlated with ovicidal activity. Using multiple plant taxa enabled the discrimination of active plant compounds from the plant extracts. Ten metabolites were identified which had the strongest correlation with ovicidal activity, which spanned four different structural classes (phenylpropanoids, triterpene saponins, brevipolide, and flavonoid). And Peña-Espinoza et al. (2020) investigated the potential anthelmintic properties of chicory using untargeted LC–MS metabolomics, finding several sesquiterpene lactones that correlated with bioactivity. Notably, 11,13-dihydro-lactucopicrin was identified as the most correlative metabolite against the pig nematode Ascaris suum (Peña-Espinoza et al., 2020).

Metabolomics approaches have also been used to elucidate how external stimuli impact bioactive secondary metabolism in plants, both to understand biosynthetic mechanisms as well as potentially increase the supply of complex molecules for drug discovery research (Atanasov et al., 2015; Harvey et al., 2015). As an example, Psiadia arguta Voigt (Asteraceae) is a plant endemic to Mauritius, traditionally used to treat multiple ailments including as an expectorant or for the treatment of bronchitis and asthma. Preliminary biological screenings had suggested antimalarial (Plasmodium falciparum) activity from P. arguta leaves, and a phytochemical investigation of this plant led to the isolation and characterization of five antiplasmodial molecules (Mahadeo et al., 2019). However, due to its low occurrence naturally, and its protection from the collection due to its threatened status, Mahadeo et al. (2020) employed a 1H NMR-based metabolomic approach to study the accumulation of antiplasmodial compounds during the growth of the plant. Young plants of P. arguta were cultured in vitro and then micropropagated plants at different stages of development were acclimatized in order to identify factors influencing the production of bioactive compounds. The metabolomics analysis revealed that four bioactive compounds (labdan-13(E)-en-8α-ol-15-yl acetate, labdan-8α-ol-15-yl acetate, labdan-13(E)-ene-8α-ol-15-diol, and (8R,13S)-labdan-8,15-diol) accumulated in the P. arguta leaves when the plants were subjected to biotic stress (Mahadeo et al., 2020).

The molecular networking approach has been adapted further to create a biochemometric modeling method, termed “bioactive molecular networking,” where the chemical and bioactivity analyses are synchronized to visualize bioactive metabolites in the context of their structural relationship to one another. Nothias et al. developed a biochemometric workflow to overlay bioactivity results within a molecular network. They applied this approach to discover novel antiviral compounds from Euphorbia dendroides L. (Nothias et al., 2018). What was especially remarkable about this approach was that the bioactive molecular networks, built upon the untargeted metabolomics permitted the detection of bioactive molecules which were still unknown even after performing a classical BGF procedure.

The brown alga Fucus vesiculosus L. has demonstrated consistent antimicrobial activity of the extract against human pathogenic bacteria; at the same time, untargeted metabolomics analyses have suggested a variety of metabolites in the algal extract. Buedenbender, Astone, and Tasdemir (2020) also applied a “bioactive molecular networking” approach using the bioactive hexane and butanol partitions from an F. vesiculosus extract, and networked to identify the compounds responsible for antibacterial activity against MRSA. The first bioactive cluster identified by the bioactive molecular network consisted of galactolipids and allowed for subsequent targeted isolation efforts of six monogalactosyldiacylglycerol (MGDG) derivatives and one digalactosyldiacylglycerol (DGDG) (Buedenbender, Astone, & Tasdemir, 2020) (Fig. 10.5). Two of the MGDGs as well as the DGDG exhibited activity against MRSA. A second compound
structural class with enhanced bioactivity was phlorotannins, with phlorethol-type phlorotannins evidencing especially high correlations with antimicrobial activity, in which two active phlorotannins were isolated based on the bioactive molecular network approach (Buedenbender et al., 2020). This study further highlighted the analytical potential of combining molecular networking with bioactivity assaying as a complementary tool for the identification and targeted isolation of bioactive compounds from plant structures.
A study by Caesar, Kellogg, Kvalheim, Cech, & Cech (2018) investigated the antimicrobial properties of *Angelica keiskei* (Miq.) Koidz. (Apiaceae), or ashitaba, which is native to the southernmost islands of Japan and has been traditionally used to extend life expectancy, increase vitality, and treat a broad range of diseases and infections. Authors used a method integrating biochemometrics, and molecular networking was devised and applied to *Angelica keiskei* to comprehensively evaluate its antimicrobial activity against *Staphylococcus aureus*. This approach highlighted potential bioactive compounds and provided structural information on these structures. A set of chalcone analogs were prioritized for isolation, yielding 4-hydroxyderricin (MIC $\leq 4.6 \mu M$, IC$_{50} = 2.0 \mu M$), xanthoangelol (MIC $\leq 4.0 \mu M$, IC$_{50} = 2.3 \mu M$), and xanthoangelol K (IC$_{50} = 168 \mu M$), the latter of which had not been previously reported to possess antimicrobial activity (Caesar et al. 2018). The amalgamation of the two approaches, biochemometrics and molecular networking enabled a more complete understanding of the compounds responsible for *A. keiskei*’s antimicrobial activity.

Bioactive molecular networking also was employed to investigate chicory (*Cichorium intybus* L.), a plant rich in sesquiterpene lactones and demonstrated anthelmintic activity in livestock. From six different sources of chicory material, including industrial byproducts (fresh chicory root pulp, fresh leaves from chicory cv. Choice, and four samplings of fresh leaves from chicory cv. Spadona). The resulting extracts were tested for anthelmintic activity against the nematode model *Caenorhabditis elegans* as well as the pig nematode *Ascaris suum*. Untargeted metabolomics revealed that the chicory root pulp had a distinctly different chemical fingerprint compared to the fresh leaf chicory extracts. Molecular networking confirmed several sesquiterpene lactones and associated derivatives that may be responsible of its potent antihelminthic activity. Bioactivity-based molecular networking of chicory root pulp and the most potent forage chicory extracts used a Pearson correlation metric between a feature’s relative abundance (LC-MS peak area) and the EC$_{50}$ from the corresponding *A. suum* assays, and the subsequent analysis revealed a high predicted bioactivity for the guaianolide sesquiterpene 11,13-dihydro-lactucopicrin (Peña-Espinoza et al., 2020) (Fig. 10.6). Thus the study highlighted the potential of the agricultural or industrial by-product chicory root pulp as a livestock nutraceutical antihelminthic, as well as a source of new antiparasitic compounds (Peña-Espinoza et al., 2020).

## Challenges and future directions

Metabolomics has evolved significantly since its inception two decades ago, but still faces challenges. Three areas include improving the coverage of the metabolome, annotation, and identification of peaks in untargeted metabolomics, and questions surrounding multiple bioactive components working in combination.

## Metabolome coverage

Given the incredible diversity in the structural chemistry and complexity, as well as the wide dynamic range of abundance of plant metabolites, the methods for detection are not as facile as those for genomic or transcriptomic analysis. Technological advancements continue to
improve the sensitivity of analytical instruments, including the evolution of hyphenated methods for MS and NMR and could result in significant improvements in metabolome coverage. However, the most likely means to improve coverage is to utilize current techniques and instrumentation; however, this would require a considerable effort, by which multiple samples from a particular (or multiple) species would be evaluated by using different extraction techniques and conditions (Yanes, Tautenhahn, Patti, & Siuzdak, 2011; Yuliana, Khatib, Verpoorte, & Choi, 2011), separation technologies (Naser et al., 2018), MS ionization and detection methods (Nordström, Want, Northen, Lehtio, & Siuzdak, 2008), and NMR experiments (Bingol, 2018; Emwas et al., 2019) to access the myriad different chemical structures. This would also be facilitated by the open exchange of metabolomics data and results across and between labs, enhancing cross-comparability of results; this has begun with open-access databases of metabolomics data, including the National Institute of Health’s National Metabolomics Data Repository (https://www.metabolomicsworkbench.org/) and the MassIVE public repository in the Global Natural Product Social Molecular Networking (GNPS) system (https://gnps.ucsd.edu/).

FIGURE 10.6 Bioactivity-based molecular networking of sesquiterpene lactones and derivatives in six chicory (Cichorium intybus L.) extracts based on their anthelmintic activity against the pig nematode Ascaris suum. Each nodes and edges represent one molecular ion and the pairwise spectral comparison between them, respectively. The node pie charts represent the relative quantitation (based on peak area) for the compound between the different chicory extracts, and node sizes represent the predicted bioactivity score of the feature, which was designated as the Pearson correlation coefficient ($r$) between the molecule relative abundance (peak area) and the EC50 value in the A. suum assays. The node highlighted in yellow represents dihydro-lactucopicrin, which evidenced a statistically significant high bioactivity score ($r < 0.85$ and a significance of $P < .03$). Compound annotations were determined by searching against the Global Natural Product Social Molecular Networking libraries. Source: From Peña-Espinoza, M., Valente, A.H., Bornancin, L., Simonsen, H. T., Thamsborg, S.M., Williams, A.R., & López-Muñoz, R. (2020). Anthelmintic and metabolomic analyses of chicory (Cichorium intybus) identify an industrial by-product with potent in vitro antinematodal activity. Veterinary Parasitology, 280, 109088.
There is one emergent strategy that warrants mention here: metabolite imaging. Metabolite imaging techniques provide a high spatial resolution of metabolite abundances. Spatial analysis of plant tissues has the potential to provide unique insights into the locations of biosynthesis, storage, and action of botanical natural products, including antiinfective agents, and produce insights into plant biology. In mass spectrometry imaging (MSI), the mass spec either samples small, discrete locations on the surface, analyzing the $m/z$ values and then constructing a two-dimensional representation of the data, or larger sections that are analyzed multiple times and then stitched together as a single computational reconstruction of the sample (Boughton, Thinagaran, Sarabia, Bacic, & Roessner, 2016; Dreisewerd, 2003; Gonzalez et al., 2012; Kertesz & Van Berkel, 2010; Poulin & Pohnert, 2019; Stopka et al., 2019) (Fig. 10.7). Ionization of the samples can be achieved using a variety of methods, including desorption electrospray ionization (DESI), matrix-assisted laser desorption ionization (MALDI)-MSI,

**FIGURE 10.7** Scheme illustrating a mass spectrometry imaging workflow showing two main approaches, the microprobe (1) and microscope (2) technique. (1) In the microprobe approach, discrete locations on the sample surface are sampled and the $m/z$ of the resulting ions is measured, after which the resulting mass spectra for each location (given as $x,y$ coordinates) are reconstructed to form a final dataset. (2) With the microscope approach, a broadly focused laser permits sampling wide areas of tissue and the resulting ions are detected using both a position and time-sensitive time-of-flight (TOF) detector, which permits determination of the $m/z$ and the spatial distribution of each ion within the sample area. Multiple samplings are required to cover very large surface areas, after which the data are computationally reconstructed to complete the dataset. Source: From Boughton, B.A., Thinagaran, D., Sarabia, D., Bacic, A., & Roessner, U. (2016). Mass spectrometry imaging for plant biology: A review. Phytochemistry Reviews, 15, 445–488.
matrix-free laser desorption ionization MS (LDI–MS), secondary ion MS (SIMS), droplet-based liquid microjunction surface sampling probe (droplet-LMJ-SSP) (commonly known as the droplet probe), and laser-ablation electrospray ionization (LAESI)-MS (Gonzalez et al., 2012; Kertesz & Van Berkel, 2010; Poulin & Pohnert, 2019; Stopka et al., 2019). The most common mass spectrometry imaging ionization techniques are assisted laser desorption ionization, softer ionization methods that are used for direct measurement of molecular ions, some of which (e.g., MALDI and LAESI) are dependent upon a matrix layer deposited onto a sample surface to enable desorption of analytes from solid into the gas phase and to promote ionization (Dreisewerd, 2003).

As the spatial resolution of these imaging methods has advanced, along with the increased sensitivity and dynamic range of mass spectrometers, the size of samples has shrunk, enabling the investigation of metabolites or profiles in individual or small clusters of cells: the resolution of matrix-assisted laser desorption/ionization-mass spectrometry imaging (MALDI-MSI) has reached 5–20 μm (Dalisay et al., 2015; Soltwisch et al., 2015), while liquid metal ion guns for SIMS imaging have reduced the beam width to 200 nm (Fletcher & Vickerman, 2010). Imaging mass spec has enabled deeper insight into biosynthetic pathways as well as localization of metabolites within plant tissue. For example, the biosynthetic steps of the naphthodianthrone hypericin from Hypericum perforatum L. are still unestablished. Revuru, Bálintová, Henzelyová, Čellárová, and Kusari (2020) hypothesized that skyrin serves as a precursor in the biosynthesis of hypericin. They established the spatial distribution of skyrin and how it correlated to the distribution of hypericin using MALDI-high resolution mass spectrometry (MALDI-HRMS) imaging. The imaging across five different Hypericum species revealed a species-specific distribution and localization pattern of skyrin, which was similar to hypericin in the leaf tissue, and suggests an alternative biosynthetic pathway of hypericin and analogs (Revuru, Bálintová, Henzelyová, Čellárová, & Kusari, 2020).

However, many of these approaches have been limited to targeted metabolite analysis, rather than an untargeted metabolomics method. With advances in sampling methods and instrumentation, there has been a turn towards more metabolomics-based investigations of plant systems. Cahill, Riba, and Kertesz (2019) used a single-cell printer technology and liquid vortex capture-MS to investigate the lipid composition of the microalgae Chlamydomonas reinhardtii and Euglena gracilis. The imaging analysis discovered multiple diacylglyceroltrimethylhomo-Ser, phosphatidylcholine, MGDG, and DGDG lipids in single cells. The approach was also able to differentiate mixed cultures of the two algae by their lipidomic profile and specific levels of diacylglyceroltrimethylhomo-Ser and phosphatidylcholine lipids (Cahill, Riba, & Kertesz, 2019). To meet the challenges of limited sampling from small laser beam widths, Hansen and Lee (2018) developed a “multiplex MS imaging” technique to improve coverage and facilitate better annotation of the metabolites detected, while also developing new matrices capable of increasing the sensitivity from the small sampling sizes (Hansen & Lee, 2018). Feenstra, Hansen, and Lee (2015) sought to increase the diversity of chemical compounds that can be imaged and identified, using multiple ionization matrices to overcome the desorption/ionization bias for different metabolite classes and coupled it with dual polarity ionization and tandem MSI. They found that the use of multiple matrixes along with dual ionization polarities allowed the visualization of multiple compound classes, and the data-dependent MS²...
spectra permitted the identification of the compounds directly from the tissue. The team employed a test case of germinated corn seed, identifying 166 unique ions from the MS\(^2\) spectra, 52 of which were identified as unique compounds. And, based upon the scans, the authors estimated over 500 metabolites could be potentially identified and visualized (Feenstra, Hansen, & Lee, 2015).

In addition, one in situ analysis technique incorporates chromatography, using a micro-extraction technique featuring a droplet-liquid microjunction-surface sampling probe (Kertesz & Van Berkel, 2013; Oberlies et al., 2019). This “droplet probe” has been used to perform microextraction on the surface of fungal (Knowles et al., 2019; Sica et al., 2015) and botanical samples (Kao, Henkin, Soejarto, Kinghorn, & Oberlies, 2018) for targeted as well as untargeted metabolomics. Kao et al. (2018) used the droplet probe to herbarium voucher specimens of *Garcinia mangostana* L. to profile multiple cytotoxic prenylated xanthones. The droplet probe has several advantages over other mass spec techniques, including the benefit of being a nondestructive imaging methodology (Oberlies et al., 2019). Put together, single-cell (or low-cell) level metabolomics will continue to advance and represent a potentially innovative way to explore how metabolite profiles vary in response to biotic and abiotic stressors with high spatial resolution, as well as impacting our understanding of the cooperative and antagonistic effects among metabolites.

**Annotation/identification**

Untargeted metabolomic approaches to bioactive discovery rely on the spectral data generated by the instrumentation for the independent variables. In MS-based metabolomics, these data are represented by the \(m/z\) value of the feature, which can be further specified by pairing it with retention time if a chromatography separation was used beforehand (e.g., LC-MS or GC-MS). For NMR metabolomics, the resulting peaks are then binned into narrow windows, which depends on the resolution of the instrument (Clark et al., 2016; Clendinnen et al., 2015; Robinette, Brüscheiler, Schroeder, & Edison, 2012). However, neither \(m/z\) values nor binning provides any significant structural information about the detected features, especially if they are unknown. Because of this, and due to the fact that metabolomic profiling results in large datasets, annotating or identifying metabolites often faces challenges, and thus it is estimated that only 2%—10% of the features from untargeted metabolomics LC-MS studies can be annotated directly from the experiment (Aksenov, Da Silva, Knight, Lopes, & Dorrestein, 2017). However, annotation and structural information can be improved through the addition of other analyses. One of the most significant improvements in compound annotation and identification for Ms-based metabolomics analysis was improvements in mass accuracy (measured as parts per million, p.p.m.) and the corresponding resolving power (\(m/\Delta m\)) of the instruments, which can reach a resolution of more than 1,000,000 (at full-width half-maximum) (Kueger, Steinhauser, Willmitzer, & Giavalisco, 2012). With the advancements in relative isotopic abundance, in combination with sub-p.p.m. mass accuracies, these data help to distinguish between overlapping mass signals and aid in predicting the elemental composition of measured metabolic signals. In addition, one approach for accurate annotation of elemental compositions that is independent of mass spectrometer platform is the use
of isotopically labeled compounds. Feeding a culture or organism with either a single labeled compound and follow its intercellular metabolism or isotopically label whole carbon, nitrogen, and/or sulfur input renders the whole metabolome isotopically labeled. These labeled metabolites are apparent through shifts in the composition-specific m/z and can be used to drastically improve precision in the elemental composition and reduce false positives, as well as differentiate between system compounds and contaminants and noise (Bueschl, Kraska, Kluger, & Schuhmacher, 2013; Giavalisco, Köhl, Hummel, Seiwert, & Willmitzer, 2009).

As noted above, molecular networking clusters MS/MS data based upon similarity of fragmentation pattern and has become a dominant analytical method in metabolomics data analysis and annotation. The resulting molecular networks can be seeded with spectra of standards to provide enhanced annotation of structural features or the network can serve as a search against community-curated databases of MS/MS spectra, such as the GNPS platform (https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp) (Nguyen et al., 2013; Watrous et al., 2012; Yang et al., 2013). Other spectral databases, including MassBank (https://mona.fiehnlab.ucdavis.edu/, http://www.massbank.jp, https://massbank.eu/MassBank/) and the National Institute of Standards and Technology Mass Spectral Library (https://chemdata.nist.gov/), are additional reference sources. Beyond the ability to search through the publicly available libraries or seeded networks with reference MS/MS spectra, the molecular network can also provide an opportunity to extend an annotation to an unknown molecule (node) via an “analogues search” option, which expands and accelerates the capacity to ascertain similar structures (Quinn et al., 2017; Watrous et al., 2012; Yang et al., 2013). The annotation of the unknown can be propagated using differences in masses that correspond to known augmentations or features of a molecule; for example, a difference of 14 Da between two nodes would suggest a presumed methylene (CH₂) addition or deletion, a difference of 16 Da an oxygenation, and a difference of 34 Da could represent the substitution of proton by a chlorine atom. It is possible that, through thorough interpretation of these differences in masses coupled with diagnostic fragmentation patterns observed, one might be able to discern which part of the molecule is modified.

For unknown metabolites highlighted during an untargeted metabolomics study, with no commercially available standard, MS/MS databases or other mass spectral metrics only provide at best a tentative identification of the molecular structure. To definitively determine their structure, it is necessary to isolate these compounds for analysis. There are several approaches that can be employed to separate out individual compounds from a more complex extract (Agatonovic-Kustrin, Morton, & Yusof, 2015; Bucar, Wube, & Schmid, 2013; Kim et al., 2001; Michalkiewicz, Biesaga, & Pyrzynska, 2008; Pfoze, Myrboh, Kumar, & Rohman, 2014). Typically, compounds are isolated via a combination of solvent extraction, liquid–liquid partitioning, and column chromatography (Fig. 10.8). The extraction solvent, chromatographic phases (both stationary and mobile) employed for the various stages of fractionation and separation can be augmented to optimize the isolation of the bioactive compound(s) of interest. Once isolated (and purity is checked by a combination of spectroscopic means), structural elucidation can proceed via a combination of NMR and MS techniques; these have been extensively reviewed previously. (Bouslimani,
FIGURE 10.8 Generic schema for extraction and fractionation of a botanical natural product. Initial biomass is dried, ground, and mixed with an organic solvent, after which it is subjected to multiple liquid–liquid partitions to yield an organic extract. This extract is chromatographically separated using normal-phase and reverse-phase liquid chromatography systems to generate simpler fractions, and ultimately purified compounds. Source: From Kellogg, J.J., Paine, M.F., McCune, J.S., Oberlies, N.H., & Cech, N.B. (2019). Selection and characterization of botanical natural products for research studies: A NaPDI center recommended approach. Natural Product Reports, 36, 1196–1221.
Synergy

Often in botanical samples, mixtures of compounds are more effective than their individual constituents in isolation due to additive or synergistic interactions among compounds. This is a potential advantage of plant natural products combating infectious diseases (Patwardhan & Mashelkar, 2009), which may lead to increased efficacy and a diminished tendency for the evolution of resistance (Wagner & Ulrich-Merzenich, 2009). One recent study revealed that the antituberculosis effect of *Artemisia annua* L. was substantially stronger than equivalent concentrations of the putative bioactive artemisinin, suggesting that *A. annua* extracts exert their antituberculosis effects through a combination of additional compounds (Martini et al., 2020).

However, it remains difficult to ascertain whether compounds are in fact working in concert to produce a biological effect. Increasingly, metabolomics approaches have been suggested as a way to identify multiple components that could work in concert due to their de-emphasis of separation and isolation (Caesar & Cech, 2019). *Terminalia sericea* Burch. ex DC., a popular remedy in South Africa for the treatment of infectious diseases, was evaluated for antibacterial compounds. The resulting biochemometric analysis highlighted no consistent association between the phytochemical levels and the activity of the active or nonactive extracts. The authors deduced that multiple constituents of *T. sericea* root bark contributed to the observed activity; however, further investigation of the interactions of compounds present in the root bark was warranted (Anokwuru et al., 2020). Another study combined biochemometrics with synergy-directed fractionation to identify active compounds and/or synergists from the botanical *Hydrastis canadensis* L. (Britton et al., 2018). The analysis revealed a new synergistic flavonoid, 3,3′-dihydroxy-5,7,4′-trimethoxy-6,8-C-dimethylflavone which, when tested in combination with berberine, lowered berberine’s IC50 from 132.2 ± 1.1 to 91.5 ± 1.1 μM, yet in isolation the flavonoid did not demonstrate antimicrobial activity (Britton et al., 2018). However, this approach suggested that the complexity of botanical extracts may necessitate several iterations of fractionation to simplify the mixture before biochemometric analysis yields useful results.

Conclusions

It should be noted that despite the great promise of untargeted metabolomics for bioactive molecule discovery, metabolomics is not a panacea for all botanical investigations. The data analysis of untargeted metabolomics datasets relies on statistical inference, and thus can only ascribe putative bioactivity to metabolites. Any such predictions require confirmation, which can be achieved via an isolation/structural isolation/bioassay approach (Fig. 10.8), genetic analysis of associated biosynthetic pathways followed by knockout assays resulting in defective mutants, and heterologous expression of the candidate biosynthetic pathway in a different organism to demonstrate a concomitant gain of antiinfective activity. Especially when considering the function and modes of action of
bioactive metabolites, metabolomics alone is not sufficient for answering all of these queries, and it is here where other -omics tools and data-driven approaches should be employed to complement metabolomics analyses.

The developments over the last two decades have led metabolomics to emerge as a key means for analyzing complex mixtures of compounds, and within the last decade metabolomics has evolved into an instrument for the discovery of bioactive compounds. Given the vast untapped botanical diversity across the globe, there is tremendous potential for uncovering novel phytochemicals that could play crucial roles in combatting infectious diseases, and the metabolomics tools and approaches that have been described above, as well as the new tools that will continue to be developed, will increase the efficiency and robustness of the discovery effort. While this review noted the existing technical limitations that, for now, hamper certain metabolomics-driven inquiries, continued innovation in metabolomics data acquisition, data analysis, and statistical inference will propel the discipline forward. And, as public data repositories and molecular networking and chemical identification tools become more widespread and advanced, the challenges of annotating metabolomics datasets and structural elucidation of highlighted bioactive candidates will become more streamlined and robust. Furthermore, the evolution of new analytical technologies such as MSI, droplet probe, and single-cell or small culture analysis will help determine the spatial distribution of various antinfecive metabolites and provide clues to their biosynthesis and storage. Thus, as both the analytical and statistical capabilities of untargeted metabolomics continue to improve, these techniques will drive antinfecive natural product discovery from plants, annotating the responsible metabolites, and identifying biological roles of these small molecules.

References


