Metabolomics: a high-throughput screen for biochemical and bioactivity diversity in plants and crops

Alexandre Foito[*a], Derek Stewart[a,b,c]

[a]*Environmental and Biochemical Sciences Group, The James Hutton Institute, Invergowrie, Scotland, UK;* [b]*Institute of Mechanical Process and Energy Engineering, School of Engineering and Physical Sciences, Heriot Watt University, Edinburgh, Scotland, UK;* [b]*NIBIO, Norwegian Institute of Bioeconomy Research, Pb115, NO-1431 Ås, Norway.*

Abstract: Plants and crops contain a staggering diversity of compounds, many of which have pharmacological activity towards a variety of diseases. These properties have been exploited by traditional and modern medicine providing important sources of healthcare to this day. The contribution of natural products (such as plant-derived) to the modern pharmacopeia is indeed significant; however, the process of identifying novel bioactive compounds from biological sources has been a central challenge in the discovery of natural products. The resolution of these challenges relied extensively on the use of hyphenated mass spectrometry (MS) and nuclear magnetic resonance (NMR)-based analytical technologies for the structural elucidation and annotation of novel compounds. Technical developments in instrumentation and data processing have fostered the development of the field of metabolomics which provides a wealth of tools with huge potential for application in the process of drug/bioactive discovery from plant tissues. This manuscript provides an overview of the metabolomics toolbox available for the discovery of novel bioactive compounds and the integration of these tools in the bioprospection and drug discovery workflows.

Keywords: Metabolomics, Plants, Bioprospection, Drug Discovery, Bioactivity, Chemical Diversity

*Address correspondence to this author at the Environmental and Biochemical Sciences Group, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA Scotland, UK; Tel: ++ 44 1382 568893; E-mails: Alexandre.Foito@hutton.ac.uk

1. INTRODUCTION

The discovery of pharmacologically active compounds has been a long standing goal of medicine since its inception. Our hunter-gatherer ancestors sought to use the diverse products found in nature to provide relief and cure from their ailments and this approach remained unchanged for millennia. Indeed, early pharmacologists focused on natural substances, mainly extracts derived from plants. Interestingly, many pharmacologically active natural products (NPs) used in modern medicine began as complex mixtures of compounds similar to our ancestor's approaches. However, modern approaches aimed at identifying and purifying the active NP(s) present within these mixtures. This process has often resulted in the development of novel pharmacological ingredients from a wide variety of biological materials. An analysis of the rate of novel chemical entity (NCE) approvals from the period of 1981 to 2014 demonstrates that the field of NPs is still highly represented with a mean of 34±9% in the last 15 years [1]. From the number of NPs developed a significant proportion were derived from plant extracts [2]. Indeed, it has been repeatedly shown that plants contain a diverse set of compounds and that some of these have shown pharmacological activity towards a variety of diseases [2]. Finding a set of bioactive compounds among the sheer diversity of the compounds present in plant tissues, estimated at 200,000 different metabolites [3], comprises a significant challenge. Furthermore, purification may be hindered by the limitation of biological material available and/or synergistic/antagonistic effects that might be lost as extracts attain higher degrees of purity. Developments in the field of metabolomics in the past 20 years have facilitated improvements in the accuracy, sensitivity, throughput and data analysis capacity and have contributed significantly towards the field of discovery of new bioactive biomolecules from plants. This review addresses the developments in metabolomics in the context of bioactive discovery.

2. SOURCES OF NEW DRUGS AND BIOACTIVE COMPOUNDS

Novel bioactive compounds can be classified according to their origin. For example, novel compounds are often characterised as NPs, botanical drugs (defined mixture), NP derivatives, synthetic drugs and mimics of NPs [1]. The early 2000s saw the development of high-throughput screening (HTS) based on biochemical targets which allowed many pure compounds to be screened in a relatively short time. The combination of HTS with modern combinatorial chemistry opened up the possibility of generating libraries comprising millions of compounds and screening these for efficacy in a relatively short period of time. However,

the enormous scale of this approach did not translate into a significant increase in the number novel drugs developed. In fact, the number of *de novo* FDA approved compounds developed has been fairly modest, with Sorafenib, an anticancer drug, the most successful output [1]. The limited success has been attributed to the randomness of the approaches which may result in a lack of NCEs with high diversity [4].

This resulted in a shift toward smaller-scale compound collections based on NP scaffolds and more focused in chemical diversity [2]. With the growing realisation that the chemical diversity present in NPs has had enormous success with respect to the development of novel drugs, in comparison with high-throughput synthetic chemistry approaches, there has been resurgence in the application of natural chemical diversity to the field of drug discovery [5-7]. Indeed, NP molecules tend to have a higher number of chiral centers, oxygen atoms and a lower number of sulfur, nitrogen and halogen-containing groups [6] as well as a lower ratio of aromatic ring atoms to total heavy atoms in comparison with synthetic compounds [8]. Since its inception in the middle of the 19th century, the history of modern synthetic chemistry represents a small fraction of time compared with the millennia associated with natural chemical evolution and diversity. The production of some of these substances has often been a determinant factor whether organisms would survive in hostile environments often providing a significant advantage over competing organisms. This explains the relatively higher percentage of biologically active substances from natural sources when compared with substances obtained from purely synthetic sources [4]. Approximately 25% of about 1 million NPs are biologically active (i.e. show toxicity or a positive activity) with approximately 60% of these products derived from plants and the remaining by microbes and animals to lesser extent [9]. NPs to this day remain important sources of biologically active compounds and during the period of 1981-2014 approximately 60% of NCEs approved are represented by NPs, NP derivatives and mimics of NPs [1]. Indeed, NPs may be isolated from diverse biological sources such as, animal, fungi, bacteria, marine organisms and plants and there has been an increasing realisation of the contribution of global biodiversity towards affordable therapeutical solutions for the majority of the population in the world [7].

3. BIODIVERSITY AND NATURAL PRODUCT DISCOVERY

Global biodiversity has great potential for providing novel sources of NPs, and indeed it has provided a significant contribution to the repertoire of pharmaceutical products currently available. Despite many successful examples of NPs used for the development of novel pharmaceuticals, the enormous biodiversity present remains largely untapped. Many

successful pharmaceutical products such as penicillin and statins derive from Fungi but these comprise a total of approximately 100,000 known species although the total number is estimated to be ten-fold higher [10]. Prokarya and Animalia have also provided important novel antimicrobial agents such as chloramphenicol and streptomycin from prokaryotes [9] and Hadruirin from animal sources [11]. Animalia species numbers are estimated at 7,770,000 and 2 150,000 species present in terrestrial and marine environments, respectively, [12] whereas estimates for number of Prokaryote species remain unknowable in any scale in the environment [13]. Marine eukaryotic organisms are also being increasingly seen as a potential source of NPs with sponges, macroalgae and seaweeds reported to provide a wealth of NPs with anticancer and antiviral properties [14, 15]. The total number of marine species is estimated at 0.7-1.0 million species including 226,000 Eukariotic species which inhabit an extreme diversity of environments with respect to pressure, salinity and temperature [16].

Interestingly, the total number of described plant species pales in comparison with the aforementioned groups with the estimate of the total number of plant species at about 320,000, although some studies estimate this number to be as high as 450,000 [17]. Despite this, the distribution of drugs based on NPs across different stages of drug development and according to their biological origin (Plant, Bacterial, Fungal, Animal or semi-synthetic) highlights that plants are important sources of NPs for biomedical applications [2]. Indeed, out of the 225 drugs derived from NPs undergoing development in 2008, 108 were from plant origin alone which significantly outnumbers any of the other origin classes (25, 7, 24 and 61 for bacterial, fungal, animal and semisynthetic origin, respectively) [2]. Indeed there are many well-known examples of plant-derived drugs which include anticancer-drugs most notably paclitaxel (taxol), a diterpenoid compound isolated from the bark of *Taxus brevifolia* etoposide (vepesid) partly synthesized from the a lignin isolated from *Podophyllum peltatum* (podophyllotoxin) and irinotecan which is a derivative from an alkaloid extracted from *Camptotheca acuminata* [1, 18].

4. FROM TRADITIONAL MEDICINE TO THE DEVELOPMENT OF NOVEL DRUGS

Plants have been seen not only as a source of energy and nutrition but have been exploited in traditional medicine throughout humanity’s history. The oldest records of usage of medicinal plants dates back to 2400 B.C. from Mesopotamia but more recent records are also available from Ancient Egypt and China [7]. Indeed, NPs particularly from plants have been at the forefront of medicine since ancient times until the isolation of morphine by Sertürner around

1804 [7]. During that period, a large wealth of ethnobotanical knowledge was acquired over countless generations and constituted the foundation of traditional and indigenous peoples medicinal knowledge. Only with the advent of the scientific revolution in the 18th century has medicine shifted from the application of herbal crude preparations to the isolation and development of the active ingredients within the plant tissues. Despite unprecedented developments in modern medicine it is still worth noting that a large proportion of the world population still relies on traditional remedies (often plant-based) as the main source of healthcare [19, 20]. With the continuous growth in the use of medicinal herbal products, issues concerning their safe use are being increasingly highlighted [20] with the lack of pharmacological and clinical data of such products seen as a major impediment to their integration into conventional medical practices [21]. Therefore, a number of studies have been focused on providing such data such as that of the clinical studies of *Rhodiola rosea* as a therapeutical agent for depression related conditions [22]. These examples highlight not only the therapeutic potential of botanical products but also the need to investigate further the mechanisms of action, the bioactive ingredients and the safe use of such products. The discovery of novel ingredients in traditional medicine often involved direct experimentation of prospective medicines in humans utilising crude preparations of biological tissues with minimal considerations regarding scientific methodology. This approach raises unacceptable safety and ethical issues, and promises very little with respect to significant success rates by modern standards. However, there were no other alternatives available until the last 200 years.

5. BIOPROSPECTION OF NATURAL PRODUCTS

Several technological, medical and scientific advances eventually led to a shift towards isolation of pure compounds. Many such botanical products have been targeted with the goal of discovering commercially valuable bioactive NPs in a process known as natural product bioprospection [23].With an estimate of plant/crop diversity of 200,000 compounds [3], many of which possess bioactive properties, the teasing out of specific bioactivity becomes incredibly difficult. Traditionally, the discovery of bioactive NPs from biological material involved the collection and screening of plant material utilising biological assays. From this step onwards there are two leading approaches utilised in the discovery of bioactivity relying on bioassay-guided fractionation and pure compound screening [24, 25] (Figure 1). The bioassay-guided fractionation approach seeks to reduce the complexity of the extracts down to its single bioactive components by utilising repeated series of fractionation and bioassay

testing until the successful isolation of an active NP [24]. Several fractionation methodologies are available for separating crude extracts into fractions of reduced complexity. Solvent partitioning is a popular approach in which samples are extracted sequentially using solvents of differing polarity and, while it requires minimal capital investment, it provides limited separation capacity, is labour intensive and requires the extensive use of organic solvents [25] which may be problematic for bioassays. Liquid-chromatography (LC) separation methods which include LC, High performance LC (HPLC), Ultra performance LC (UPLC) and counter-current chromatography are popular methods of offline and online separation of complex mixtures [26]. Although challenging, it is possible to couple LC techniques to post-column biochemical assays and parallel chemical analysis which may aid the identification of bioactive compounds from crude extracts [27]. Gel-chromatography has also been modestly used in applications where targeted molecules include proteins and other biopolymers and while separation efficiency is modest, the aqueous composition is often compatible with bioassays [26]. Thin layer chromatography (TLC) is also a popular separation technique particularly when coupled to bioassays and allows the separation and bioassay to occur under the same experimental conditions within the same plate thereby allowing several samples to be analysed simultaneously utilising less solvent. On the other hand when compared with liquid chromatographic methods it does have lower separation efficiency [28]. Affinity chromatographic methods provide an additional targeted method which relies on the binding of biomolecules with artificially generated antibodies or synthetic ligands allowing the enrichment of biomolecules with targeted traits in the fractions [26]. Bioassays are often categorised as isolated molecular target assays, cell-free multicomponent assays or cell-based assays. The former is usually based on assessing specific interactions between purified enzymes (eg. proteases, kinases) and pure compounds whereas the second includes assays on activities derived from cell extracts, cell membranes or reconstituted signalling pathways [29]. The later utilises cell-models and is generally based on the assays which monitor reporter genes or phenotypic responses resulting from intact cellular processes [29]. Regardless of the bioassay strategy being utilised, the bioassay-guided “offline” fractionation process is an iterative process that ultimately depends on the availability of preparative-scale analytical methods that allow for the resolution of complex mixtures of primary and secondary metabolites that are typical from the source organisms and result in a purified natural compound ideally as a single chemical entity with a specific activity [30]. This approach allows fractions with no bioactivity to be excluded and will generate bioactive fractions of reduced chemical complexity, however it requires an abundant

source of biological material and may often lead to the isolation and identification of known bioactive compounds in a process known as replication [25]. A literature survey indicates that the average number of isolation steps involved in the purification of NPs is less than three [30] and the authors suggested that compounds that are often present in low concentrations are rarely pursued due to their arduous isolation process. Indeed, the effort required to purify a single chemical entity depends on the concentration of the NP, the physico-chemical properties (eg. solubility and crystalisation), the use of appropriate selectivity characteristic of the separation technique chosen for the target NP, and the nature of the matrix of the crude extract [30].

An alternative approach aims to isolate and elucidate the structure of the majority of the secondary metabolites present in the crude extract and subsequently test pure compounds (ie standards) for bioactivity (Figure 1). The nature of this approach allows the selection of extracts which contain compounds which are not already present in the libraries of pure compounds with recorded bioactivities [24]. Both the cited approaches should aim to identify as many NPs present in the extracts at the earliest opportunity in order to avoid redundant work in a process known dereplication [24, 25]. The most common methodologies used for this purpose often combine a separation step, usually a chromatography-based separation which attempts to resolve crude extracts into separate metabolites, linked to a structure elucidation step which would typically involve mass-spectrometry (MS) or nuclear magnetic resonance (NMR) technologies [24]. This type of approach is well aligned with the prevalent paradigm of drug development of the past few decades in which single proteins whose inhibition is likely to be involved in the treatment of a targeted disease are identified and subsequently are tested against large libraries of small-molecules in order to identify lead molecules which will be further tested in appropriate and complex model (*ex vivo* an *in vivo*) [31]. The discovery of novel drugs according to this approach often aims to identify compounds that would act/bind specific targets. However, advances in genomics and proteomics highlight natural variation present in individuals, which might be associated with lack of global drug effectiveness due to individual mutations/polymorphisms that are not present in the model target protein screened. In addition, individual variation in gut microbiome content and composition can modulate the response to drug therapies [32]. Furthermore, when addressing antimicrobial/antiviral drugs a single target approach is generally more vulnerable to the development of resistance to novel medicines. Indeed, a combination of antiretroviral drugs that target different viral proteins is often found to be

more effective [33]. As a result, novel paradigms are arising, which include system-level thinking approaches such as computational multitarget screening [31] and the development of high-throughput combination screening platforms [34]. The rekindled interest in system-level thinking has been aided by the development of high-throughput 'omics' platforms that are able to generate large amounts of reproducible molecular data and large curated databases integrating data from different disciplines and the large development of interdisciplinary research with the goal of integrating and mining these datasets [35]. As the paradigm in drug discovery shifts toward system-level approaches with respect to the screening of multiple targets and multiple effectors, so must the process of bioactive NP discovery from plant material. Indeed, in many cases, it is the combination of metabolites rather than a unique chemical entity that is responsible for the bioactive properties of crude preparations of biological material. For example, it has been observed that testing a mixture of the two major bioactive alkylphthalides (Z-ligustilide and n-butyldenephtalide) identified through bioassay-guided fractionation of *Angelica sinensis* roots yields lower anti-endothelial activities when compared to the crude volatile oil obtained which indicates potential synergistic effects with unknown compounds [36]. Furthermore, Skirycz et al (2016) cites the example of the use of flowers and fruits of *Psychotria colorata* in native American traditional medicine to treat pain. The analgesic effect was confirmed in rodents [37], and several alkaloids in combination were identified as the active ingredients [23]. Interestingly, the major compounds identified include hodgkinsine and psychotridine, which are opioid agonist and a N-methyl-D-aspartate receptor antagonist, respectively [38, 39]. The complex mixture of agonist-antagonist compounds in crude extracts may allow additive, synergistic and antagonistic interactions between metabolites that underpin the medicinal properties of plant extracts [23]. The complexity of these metabolite interactions and their contribution towards the bioactivity of plant extracts is in contrast to the standard aforementioned approach used in NP discovery (e.g. bioassay guided isolation of pure compounds). Indeed, it is not uncommon for the fractionation of bioactive crude extracts to generate non-bioactive extracts [36, 40]. Consequently, many advocate that a shift away from less reductionist approaches towards systems biology approaches is warranted [23, 41-43]. The fields of genomics, proteomics and metabolomics have experienced great advances in the last 20 years which have allowed system biology approaches to become a reality. While the application of genomics and proteomics to the field of drug discovery has been a relatively recent event, the use of some analytical tools utilised in metabolomics such as hyphenated MS and NMR have been central in the discovery of NPs since the 1970s. However, technical developments in instrumentation

and post-separation and acquisition data analysis in the past two decades have increased the throughput of such analytical tools and resulted in naming the field ‘metabolomics’.

6. METABOLOMICS: BRAVE NEW (DATA) WORLD

The term metabolomics was coined in 1998 by Oliver et al and refers to the study of the quantitative and qualitative collection of all the metabolites in a given cell or tissue [44]. Since then, there has been a growing interest in applying metabolomics methods particularly in the field of plant sciences [45-48]. Applications of metabolomics usually include NMR- and MS-based approaches often coupled to chromatographic separations and are diverse in scope. These include studies with the goals of understanding how metabolic networks respond in plants exposed to environmental stresses [49-51], monitoring unintended metabolic effects in genetically-modified crops [52-54], discovery of novel NPs [55, 56] amongst many others. Metabolomics approaches have been reviewed extensively elsewhere [23, 45, 46, 57] and the nature of these approaches has often been categorised either as a targeted, metabolite profiling or metabolite fingerprinting. The first aims to extract and quantify a limited predetermined set of metabolites by optimising linear ranges, efficiency and robustness of extraction, and stability of the target compounds in detriment of non-target compounds. Examples of this type of approach include the quantification of polyphenols in berries [58], identification and quantification of loline-type alkaloids in endophyte-infected grasses of the genera *Lolium* and *Festuca* [59] and quantification of glucosinolates from *Camelina sativa* seeds [60]. However, the main limitation of the targeted approach is that it provides no information regarding non-targeted metabolites which may be pathway related or relevant for the objective of the study [61].

Metabolite profiling, which can be defined as analysis of a group of selected pre-defined metabolites (e.g. polyphenols, carbohydrates or fatty acids), is also affected by the limitations cited above, however it differs from targeted analysis due to including a broader range of metabolites and may include known and unknown metabolites and often providing semi-quantitative rather than absolute quantitative information [23,62,63]. There is an extensive body of literature documenting the application of this type of approach but some examples include the monitoring of primary metabolites under environmental stress conditions [50,64,65] monitoring primary and secondary metabolites in genetically modified crops [54]. Conversely, truly untargeted approaches attempt to be as comprehensive as possible while minimising bias towards any classes of compounds. It is generally regarded that there is no “one method to rule them all” that allows a comprehensive coverage of the entire

metabolome, and that the choice of sample preparation, extraction and analytical tool inevitably introduces bias towards certain classes of metabolites [66]. This limitation is often overcome by combining multiple strategies for analysis that complement each other's weaknesses.

Metabolite fingerprinting is an untargeted approach that focuses in the recognition of patterns in spectra (usually NMR and MS) collected from the analysis of sample extracts without attempting to quantify or identify specific compounds [62]. This is often achieved by applying multivariate pattern recognition tools such as principal component analysis (PCA) and discriminant function analysis (DFA) in order to identify metabolic features that differentiate groups of samples [62]. Multivariate analysis has been associated with studies of biomarker discovery and disease diagnostic [67] and it has also been used to distinguish different plant species and ecotypes [68, 69]. Despite the global, unbiased nature of this approach (as discussed above) the initial workflows did not include automated metabolite annotation, and therefore, annotation of metabolites was usually performed manually in a time-intensive process, which was usually reserved to the compounds of interest such as potential biomarkers. This approach is still currently relevant; however, advances in metabolomics workflows are blurring the once distinct division between profiling and fingerprinting methods. Indeed the past 15 years have seen great advancements in pre-processing and processing of data particularly from MS-based technologies coupled to chromatographic separation. This can be illustrated by the wealth of pre-processing and processing tools available both from commercial and instrument vendors but also as open-source tools. Some examples of these tools include XCMS [70], XCMS online [71], TAGfinder [72], Sieve (Thermo Fisher Scientific Inc) [73], MassHunter Profinder (Agilent Technologies) MarkerLynx (Waters), Progenesis QI (Nonlinear Dynamics) [74], Marker View (AB Sciex), MZmine 2 [75], Metalign [76] amongst many others.

These tools are often incorporated in workflows which allow detected features to be processed, molecular formulas to be generated, and subsequently, queried against metabolite databases from various sources (e.g. online, commercial, in-house databases). Indeed the integration of automated annotation in untargeted analysis can be seen as a hybrid of metabolite fingerprinting and metabolite profiling particularly when tailored databases are utilised. For example, Skogerson et al (2011) developed an automated peak annotation and database system for the analysis of complex volatile mixtures by gas-chromatography coupled to mass spectrometry (GC-MS) allowing the annotation of large datasets comprising

of hundreds to thousands of samples [77]. The process of automated annotation is highly dependent on the quality of source database utilized, and although targeted biological databases often provide promising results, it is not unusual for compounds to be wrongly annotated. Therefore, the metabolomics community encourages studies to categorise the certainty level of compound identification according to a set of guidelines based on the degree of evidence provided [78, 79].

6.1. NMR-based metabolomics

NMR spectroscopy allows molecules containing one or more atoms with a non-zero magnetic field to be detected that include ^{1}H, ^{13}C, ^{14}N, ^{15}N and ^{31}P [80, 81]. This technique allows the characterisation of metabolites with at least one NMR signal with respect to their frequency (chemical shift), intensity and magnetic relaxation properties depending on the chemical environment that the nucleus occupies [80]. NMR spectra, via varied pulse techniques, often provide structural information which is essential for the identification of unknown compounds and it is used routinely in NP discovery. Additionally, it is possible to obtain quantitative data from NMR spectra which may be of interest to several types of metabolomics approaches. Indeed, NMR has been used in many metabolomics approaches which include the metabolite profiling of opium poppy [56] and metabolic flux analysis of linseed embryos [82]. NMR has been extensively used in fingerprinting approaches as spectra are often too convoluted [61]; however, many groups have addressed convolution by coupling liquid-chromatography separation prior NMR analysis [83]. Despite its applications, LC-NMR approaches have inherent limitations such as significant capital costs, solvent suppression [83, 84] and perhaps, more importantly, the relatively low sensitivity of NMR which often limits the analysis to the most abundant compounds [61, 80, 81, 83] (Figure 2).

6.2. Mass spectrometry-based metabolomics

Mass spectrometry has become the *de rigueur* approach used in metabolomics due to its high sensitivity and wide metabolite coverage. Overall, this technique relies on the ionisation of chemical compounds, most commonly utilising Electron Ionisation (EI), Electrospray Ionisation (ESI) and Atmospheric pressure chemical ionization (APCI), generating charged molecules or molecule fragments which are subsequently measured with respect to their mass-to-charge ratio (m/z) and abundance. Often crude extracts are analysed through direct infusion-mass spectrometry (DI-MS) generating a single mass spectrum representing the global metabolite profile of the sample. Utilising this approach it is possible to achieve a very

high-throughput metabolite fingerprinting approach (first pass) which can be applied in a variety of plant tissues [85, 86]. As there is no separation prior to analysis, this type of approach is usually carried out on high mass accuracy instruments as it allows the distinction between some compounds with the same nominal mass [66]. However, this technique does not allow isomers to be differentiated due to their identical molecular masses [87]. Furthermore, as multiple compounds are analysed simultaneously in the mass spectrometer this technique is particularly vulnerable to ion co-suppression effects in which non-volatile compounds influence the ionization and transfer of metabolites from liquid to gas phase in an ESI source leading to a misrepresentation of the signals present in the mass spectrum [66]. In order to prevent and minimize these effects, metabolomics approaches often resort to coupling mass spectrometers with separation techniques such as GC, LC or electrophoretic separation such as capillary electrophoresis [61] (Figure 2).

GC-MS has seen extensive use in plant metabolite profiling of primary metabolites ranging from studying plant metabolic responses to abiotic stress [50], analysis of the volatile profiles of *Allium* samples [88], to the mapping quantitative trait loci for metabolites in tomato fruits [89]. This approach requires that the target compounds are volatile in the operating range of the GC usually 100-320 °C which in many cases may require compounds to be extracted from biological material and subsequently derivatised to ensure their volatility. This requirement for volatility comes at a cost as only the compounds amenable to derivatisation may be analysed [90]. Despite this limitation, GC-MS applied in the context of metabolomics allows the simultaneous detection of several hundred different chemical compounds ranging from organic acids, sugars to amino acids fatty acids and aromatic amines [61]. Indeed, GC provides a high separation efficiency allowing compounds with mass spectral similarities such as isomers and enantiomers to be distinguished and produces reproducible retention times that allow comparison of results derived from different labs. Upon chromatographic separation, metabolites are usually ionised by electron ionisation (EI), which is regarded as a reproducible approach generally unaffected by ion co-supression [87]. The reproducible nature of this approach allows mass spectrum and kovat retention indexes to be used in querying readily available MS databases containing hundreds of thousands of compounds such as the NIST standard reference database 14 (https://www.nist.gov/srd/nist-standard-reference-database-1a-v14), Golm Metabolome database [91] and FiehnLib [92].

Liquid-chromatography coupled to mass spectrometry (LC-MS) takes advantage of the separating power of liquid chromatography techniques that allow the study of high molecular

weight compounds that cannot be analysed by GC-MS [87]. When coupled to high resolution MS instruments this constitutes a powerful technique for analysis of a broad range of secondary metabolites. It is used broadly not only in metabolomics studies but also in natural product chemistry studies. Indeed, LC does not require analytes to be volatile or derivatised prior to LC analysis, and allows for a wide range of analyte polarity and mass range [90], consequently, sample preparation is generally less complex. Following chromatographic separation the metabolites are ionised (generally by ESI or APCI) and mass spectrometers generally allow for positive and negative modes of operation. Metabolites often ionise more efficiently in one mode over another so the monitoring of data in both positive and negative ionisation modes allows a more comprehensive coverage of the metabolome [66]. Generally, ionisation techniques utilised in LC-MS are softer ionisation techniques when compared to EI, which is commonly utilised in GC-MS approaches, and provide spectra much less dominated by fragments. The limited number of fragments observed in LC-ESI-MS often limits the discrimination of isomeric compounds so mass spectrometrists often design tandem-MS (MSn) methods which aim to fragment selected ions utilising collision-induced dissociation [93]. The fragmentation patterns observed can then be compared against fragmentation databases of known compounds. However, instrument variability and matrix effects (e.g. presence of co-eluting compounds or ionic salts) can impact the ionisation efficiency of the instrument and may also interfere with the chromatography which ultimately leads to a lack of reproducibility which contributes to the great challenge of comparing data generated from difference sources.

6.3. Combining multiple analytical tools

Despite the high sensitivity of MS-based methods, they are ultimately limited with respect to metabolite structure elucidation and it is largely impossible to determine unambiguously the structure of a metabolite utilising MS-based methodologies alone regardless of the accuracy of the instrument [94]. Indeed, MS-based metabolomics methods provided relatively low discovery rates and often result in false identifications with a relatively low number of metabolites which can be assigned an identity with a high degree of confidence [95]. Ultimately it is the availability of authentic standards or the requirement to purify individual metabolites followed by NMR analysis for structure elucidation that remain the only strategies that allow high-confidence level in the identification of metabolites. Therefore it is not unusual for bioprospecting studies to incorporate results from these two analytical approaches in order to provide unambiguous identification of metabolites.

Interestingly, in face of the limitations associated with both MS and NMR-spectroscopy approaches there have been efforts to combine these techniques in an hyphenated approach. By interfacing LC with NMR and MS analysis (e.g. LC-NMR-MS) it is possible to capitalise on the strengths and mitigate the weaknesses of each of techniques providing comprehensive structural data that can aid in addressing the bottleneck of compound identification in extracts [96, 97]. This approach was applied to *E. coli* extracts and permitted the correct identification of a wide range of metabolites including amino acids, nucleic acids and carbohydrates [98]. Currently, there is no single procedure that allows the quantification and identification of the entire metabolite complement in any tissue. The choice of sampling and extraction procedure as well as the analytical tool applied will inadvertently introduce bias in the analysis, thus the methodologies should be adjusted in order to obtain the most relevant and comprehensive metabolite coverage possible for the study.

7. METABOLOMICS FOR SCREENING BIOCHEMICAL DIVERSITY AND CHEMOTAXONOMY

The biodiversity present in the plant kingdom is often reflected in large qualitative and quantitative differences in metabolite composition in plant extracts. Indeed a plant population may display not only presence-absence polymorphism with regards to secondary metabolite composition but also quantitative differences between the concentrations of secondary metabolites [99]. There is indeed enormous qualitative diversity present in plant tissues and a large fraction is accounted from differential modification of common backbone structures which may have resulted from the evolution of enzymes with the same product specificity but catalysing different reactions [100]. A notable example of this includes flavonoids such as quercetin and their respective different glycoside derivatives which number over 130 [101]. Conversely, there are enzymes with poor substrate specificity that catalyse a similar reaction in more than a single extract, for example recombinant *Vitis labrusca* flavonoid 3-O-glucosyltransferase was found to glycosylate both peonidin, malvidin and isorhamnetin [102]. Plant secondary metabolites are often involved in response to (a)biotic stresses and it has been recently demonstrated that plants can undergo rapid qualitative and quantitative evolution of secondary metabolites as a response to herbivore pressure [103, 104]. Interestingly, it can be observed that when plants are introduced into new habitats that it may favour the evolution of new secondary metabolite compositions compared with its native habitat [105, 106]. It has been suggested that synthesis of secondary plant metabolites are

among the most evolvable traits as unlike primary metabolites their presence is not essential for plant survival under favourable ecological conditions [99]. The general lack of pleiotropic effects of mutations in the genes underlying the biosynthesis of secondary metabolites may allow these biochemical systems to be relatively free to evolve [107]. Under unfavourable ecological conditions on the other hand secondary metabolites may play important roles in chemical defense against natural plant enemies. According to the screening hypothesis, the high degree of diversity and evolvability present in some classes of compounds is explained by the increased chance to evolve the production of novel active compounds if they maintain a diverse and rapidly mutating complex of compounds [108], thus explaining the presence of a large complex of compounds with no known activity which could act as precursors for novel metabolites [99]. The variation present in the metabolite complement of plant tissues is ideally suited for analysis using metabolomics approaches which facilitates the development of chemotaxonomic studies with potential applications in drug discovery processes.

The potential to categorise different accession/species according to chemical composition provides tools enabling the selection of plant raw materials, which display the broadest chemical diversity for bioprospection [109]. Additionally, the combination of chemotaxonomic studies with bioactivity data might aid the identification of chemical classes qualitatively and quantitatively overrepresented in plant species with medicinal properties [110]. A chemotaxonomic approach based on metabolite profiling of a collection of diverse indigenous Korean plant species showed clear separation according to family [111]. The authors subsequently combined the metabolite dataset with a bioactivity dataset to mine for correlations between individual metabolite levels and bioactivity which resulted in the identification of 5 metabolites with statistically significant correlations with bioactivity [111]. A similar strategy was employed to analyse nearly 90 species from the genus *Rhododendron* and correlate the metabolite profiles with plants extracts with antimicrobial and cytotoxity datasets resulting in the identification of seven metabolites with potential antimicrobial properties [112].

8. METABOLOMICS TOOLS FOR BIOPROSPECTING AND DEREPLICATION

The analytical chemistry toolbox available for metabolomics approaches have historically been used in traditional NP discovery approaches particularly with respect to targeted metabolite profiling, structural elucidation, qualitative and quantitative assessment of purified NPs. Indeed there are numerous examples of bioprospection studies of plant extracts using MS and/or NMR tools [113-117]. However, the discovery of novel bioactive NPs faces

challenges with respect to increasing rates of rediscovery of known compounds as active ingredients (replication) [118-120], high rates of attrition (lost to follow-up) in the translation of bioactivity due to unknown modes of action [121, 122] and difficulties in screening synergistic/antagonistic effects between multiple active compounds. As result "business as usual" in the field in bioprospecting is no longer an option if one is to address these challenges efficiently. Many are now turning to system-level approaches where the high-throughput chemical characterization of complex compounds plays a central role and as result metabolomics-approaches are increasingly being used in multidisciplinary approaches to address these challenges. Metabolomics approaches are particularly suited for addressing the challenge of increasing rates of rediscovery of known compounds in the process of dereplication, which is the process of identifying already known bioactive molecules in extracts allowing researchers to focus their purification efforts on novel lead molecules and avoid redundant work [24]. As untargeted metabolomics approaches have gained attention due to high-throughput methods for the chemical characterisation of bioactive extracts, there has been increased development of workflows that take advantage of MS^2 spectral information and high mass accuracy for rapid annotation of metabolites aligned with multivariate statistical analysis allowing for quick dereplication of complex extracts [24, 118, 119, 123].

When applied to the analysis of extracts from plants [124] or other organisms [120] these strategies can be successful and result in the identification of novel structures of the lead compounds thereby facilitating the pure compound screening approach (Figure 1) and avoid redundant work. The automated annotation of metabolomics datasets has been a central challenge in the field of metabolomics and has seen extensive development in the past decade. Therefore, continuing developments in this field and integration with known active ingredient libraries will ensure more efficient dereplication processes that can accelerate the process of bioprospection.

9. METABOLOMICS FOR MODE OF ACTION AND DRUG METABOLISM

The use of metabolomics tools has also seen application in the field of drug metabolism and toxicology and the elucidation the mode-of-action of lead compounds. The innovations observed in HTS have allowed a significant improvement in the level of automation of *in vitro* biochemical or cellular assays resulting in the systematic use in the early stages of drug discovery [96]. Isolated molecular target assays have been extensively employed in HTS approaches and often rely on the interaction between small molecules and protein targets.

NMR methods in particular have seen extensive application not only for small molecule structure determination but also analysis of target-ligand complexes [125, 126]. Indeed, the analysis of protein ligand-interactions is of central importance for understanding the regulation of biological functions of target proteins, cooperativity of ligand mechanisms and the development of novel drugs that modulate protein functions or inhibit protein interactions [126].

Alternatively, bioassays based on the use of microbial, animal and human cell culture metabolic models of human metabolism can provide an alternative source to animal experimentation in order to assess the effect drugs in putative human metabolism. In these instances metabolomics can be utilised as a tool to monitor cellular drug metabolism and assess the putative risks (or benefits) of drug metabolites in model organisms [96]. The use of metabolomics may be applied to rapidly evaluate the intracellular and extracellular metabolites and consequently evaluate cellular uptake, accumulation and metabolism of the tested drug/bioactive. The metabolomics approach when complemented with transcriptomic and proteomic data, can provide detailed mechanistic information of altered cellular processes [127, 128].

It was highlighted that the aforementioned reductionist approaches should be avoided when attempting to obtain evidence for the bioactivity of traditional medicines and that holistic approaches are advocated [41] which include *in vivo* approaches. Indeed, understanding the *in vivo* metabolic fate and mode of action of any potential drug candidate is of paramount importance for drug discovery and design [96]. The use of metabolomic-based approaches in the analysis of biofluids such as plasma or urine, has been explored widely within the context of the metabolic fate of potential drugs/bioactives [129, 130]. For example Sun et al (2009) investigated the excretion kinetic profile of acetaminophen in rats by profiling major drug metabolites (acetyl-L-cysteine acetaminophen) in urine utilising LC-MS and NMR metabolomics approaches. These results were correlated with endogenous oxidative stress-related metabolites leading the authors to postulate that the toxicity of the drug can be monitored by utilising metabolomics tools [131]. A combined LC-MS and NMR metabolomics approach has also been applied for profiling urine, plasma and tissue samples from penicillin-treated rats revealing a decrease in many host-gut microbiota metabolites and urinary conjugated metabolites (sulphate, glucuronide and glycine conjugates) suggesting a strong interaction between gut-microbiota and drug metabolism [132]. These approaches have strong potential for application in screening drug metabolism, toxicity and compliance

in the pre-clinical phases of the drug discovery process that are often associated with high levels of attrition.

Metabolomics approaches have been routinely applied in the field of biomarker discovery with particular application in the field of cancer research [133-135] with an extensive body of literature reporting the identification of candidate metabolic biomarkers [134,136,137]. Indeed during cancer initiation and progression the cellular metabolism is altered allowing cells to increase anabolic synthesis, avoid apoptosis and adapt to low nitrogen and oxygen availability [138]. These extensive metabolic changes allow the discovery of multiple cancer biomarkers which can be monitored utilising metabolomics-based approaches [137,138]. In contrast with preclinical cancer studies where tissue and tumour samples are routinely collected, the regular monitoring of biomarkers in clinical setting often requires the use of minimally invasive low-cost methods for sample collection and analysis as they provide minimal discomfort to the patient and can be collected from control individuals [138]. Metabolomics-based methodologies could therefore provide robust low-cost solutions for the monitoring of multiple biomarkers with in various tissues but also extracellular fluids leading to potential applications in the assessment of disease progression, prognosis, treatment efficacy and drug metabolic effects in clinical settings [134]. As result, there has been an increasing interest in the application of metabolomics in the realm of cancer research and treatment. For example, a global untargeted mass spectrometry-based metabolomics approach was used in a preoperative window clinical trial of metformin for the treatment of endometrial carcinoma with the goal of monitoring of the metabolic effects of the drug treatment in serum (pre- and post-treatment) and tumour tissue [139]. Samples were collected from twenty obese patients and 65% were found to respond to the treatment, reducing proliferation of tumours based on pre- and post-treatment evaluation. The application of metabolite profiling tools to the analysis of serum samples indicated that responders to metformin treatment experienced a greater induction of lipolysis compared to non-responders and this correlated with increased fatty acid oxidation and glycogen metabolism observed in the patient's tumour tissue [139]. This led the authors to suggest the potential use of metformin in the treatment of endometrial cancer in obese individuals. Subsequent work has been carried out in order to understand the mechanism of action behind the effect of metformin on cancer proliferation by applying a combined metabolomics and transcriptomics approach to study its effect over time in human-derived colon cancer LoVo cells. A total of 47, 45 and 66 metabolites, ranging from carbohydrates to amino acids and lipids, were found

to be differentially expressed at three time points. The majority of the metabolites experienced an upregulation 8h following exposure and were subsequently downregulated after 24h of exposure [140]. In combination with the transcriptome analysis the authors concluded that metformin is likely to modulate the cell energy metabolism at both metabolomics and transcriptomics level leading to the suppression of LoVo cell proliferation [140]. This illustrates the potential of metabolomics tools not only for providing insight into the drug mechanisms of action in models during the pre-clinical stages of drug development, but also for the monitoring of metabolic responses to the lead compound in patients during the clinical stages of drug development.

10. METABOLOMICS OF COMPLEX COMPOUND MIXTURES AND INTEGRATION OF BIOACTIVITY DATA

The use of complex mixtures of compounds such as typically found in extract libraries has traditionally been a challenge in NP discovery due to the diversity of small molecule present in varying levels some of which may interact and interfere in bioassays. As this is problematic for the identification of bioactive components in a single extract, the standard (reductionist) approach has been to generate fractions of reduced chemical complexity or testing pure compounds. However, with high-throughput screening and high-throughput chemical characterisation it is possible to generate a wealth of information regarding qualitative and quantitative differences in compounds mixtures for all the complex mixtures tested bioactivity. These multidisciplinary resources could then be data-mined in order to find correlations between chemical composition and bioactivity phenotype [141,142]. Compound activity mapping was developed by integrating high-content biological screening and untargeted metabolomics using a correlation of individual mass signals with specific phenotypes from a cell-based assay [141]. The authors combined 10 977 mass spectral features and 58 032 biological measurements from a library of 234 NP extracts which resulted in the identification of 13 clusters of fractions containing 11 known compound families and four new compounds. The authors found this methodology suitable for finding clusters containing single bioactives indicating a single mode of action. Furthermore clusters containing multiple bioactives were also reported which highlighted potential different modes of action although false positive compounds were also reported. Ultimately, in order to prevent false positives it is necessary that the extract libraries provide wide quantitative and

qualitative chemical differences and minimize the occurrence of compounds that consistently share quantitative trends. It is proposed that screening larger libraries of extracts of similar organisms could aid in solving this problem [141].

CONCLUSION

Metabolomics provides a wealth of tools with huge potential for aiding the process of drug/bioactive discovery from plant tissues (Figure 3). This process starts with the selection of target raw plant material and is often guided by ethnopharmacological leads. In addition, metabolomics can be used in the context of chemotaxonomy in order to identify sources of novel chemical entities, classes of compounds overrepresented in medicinal plants and novel plant sources for bioactive components. Furthermore, traditional bioprospection approaches rely on analytical tools integral to metabolomics such as NMR- and MS-based technologies. However, metabolomics aims to provide high-throughput chemical screening tools and these can be used for dereplication and identification of novel chemical entities with promising medicinal properties from complex plant extracts. When utilised as part of multidisciplinary approaches it provides the opportunity for mining the datasets in order to identify chemical diversity and potential novel leads for drug development as well as identifying the relevant biosynthetic pathways underlying the accumulation of bioactive components. Furthermore, metabolomics approaches can be applied to cellular, animal and human models in order to assess drug metabolism and mechanisms of action. Metabolomics can also be applied throughout the development of biotechnology-based approaches in order to identify promising metabolic engineering strategies and monitor unintended metabolic effects. Finally, analysis of biofluids utilising metabolomics-based tools has great application potential for monitoring the drug metabolism response in model species. In summary, the metabolomics toolbox holds great potential for application throughout the process of drug and bioactive discovery, ranging from the screening of promising plant raw material for NPs to the elucidation of drug metabolism through the analysis of biofluids.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

AF and DS acknowledge support from the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS) and EU-PF7 BachBerry (Project No. FP7-613793). DS also acknowledges support from DISCO (Project No. FP7-613513.

REFERENCES

[1] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 2016; 79:629-661

[2] Harvey AL. Natural products in drug discovery. Drug Discov Today 2008; 13:894-901.

[3] Weckwerth W. Metabolomics in systems biology. Annu Rev Plant Biol 2003; 54:669-689

[4] Tietze LF, Bell HP, Chandrasekhar S. Natural product hybrids as new leads for drug discovery. Angew Chem Int Ed 2003; 42:3996-4028

[5] Galm U, Shen B. Natural product discovery: the times have never been better. Chem Biol 2007; 14:1098-1104

[6] Yuliana ND, Jahangir M, Verpoorte R, Choi YH. Metabolomics for the rapid dereplication of bioactive compounds from natural sources. Phytochem Rev 2013; 12:292-304.

[7] David B, Wolfender JL, Dias DA. The pharmaceutical industry and natural products: historical status and new trends. Phytochem Rev 2015; 14:299-315.

[8] Feher M, Schimdt JM. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 2003; 43:218-227.

[9] Demain AL. Importance of microbial natural products and the need to revitalize their discovery. J Ind Microbiol Biotechnol 2014; 41:185.

[10] Schueffler A, Anke T. Fungal natural products in research and development. Nat Prod Rep 2014; 31:1425-1448.

[11] Torres-Larios A, Gurrola GB, Zamudio FZ, Possani LD. Hadrurin, a new antimicrobial peptide from the venom of the scorpion Hadrurus aztecus.Eur J Biochem 2000; 267:5023-5031.

[12] Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How many species are there on earth and in the ocean? PLOS Bio. 2011. 9:e1001127.

[13] Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci 2002; 99:10494-10499.

[14] Mayer AMS, Glaser KB, Cuevas C, *et al*. The odyssey of marine pharmaceuticals: a current pipeline perspective. Trends Pharmacol Sci 2010; 31:255-265.

[15] Newman DJ, Cragg GM. Drugs and drug candidates from marine sources: an assessment of the current "state of play". Planta Med 2016; 82:775-789.

[16] Appletans W, Ahyong ST, Anderson G, *et al*. The magnitude of global marine species diversity. Curr Biol 2012; 22:2189-2202.

[17] Pimm SL, Joppa LN. How many plant species are there, where are they, and at what rate are they going extinct. Ann Missouri Bot Gard 2015; 100:170-176.

[18] Proksch P, Edrada RA, Ebel R. Drugs from the seas – current status and microbiological implications. Appl Microbiol Biotechnol 2002; 59:125-134.

[19] Gilani AH, Rahman AU. Trends in ethnopharmacol. J Ethnopharmacol 2005; 100:43-49.

[20] Ekor M. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Front. Pharmacol 2014; 4:177.

[21] Mao JJ, Li QS, Soeller I, Xie SX, Amsterdam JD. Rhodiola rosea therapy for major depressive disorder: a study protocol for a randomized, double-blind, placebo- controlled trial. J Clin Trials 2014; 20:170.

[22] Amsterdam JD, Panossian AG. Rhodiola rosea L. as a putative botanical antidepressant. Phytomedicine 2016; 23:770-783.

[23] Skirycz A, Kierszniowska S, Meret M, Willmitzer L, Tzotzos G. Medicinal Bioprospecting of the Amazon rainforest: a modern Eldorado. Trends Biotechnol. 2016; 34:781-790.

[24] Wolf D, Siems K. Burning the Hay to find the needle – Data mining strategies in Natural product dereplication. Chimia 2007; 61:339-345.

[25] Duarte K, Rocha-Santos TAP. Analytical techniques for discovery of bioactive compounds from marine fungi. Trends Anal Chem 2012; 343:97-110.

[26] Weller MG. A unifying review of Bioassay-guided fractionation, effect-directed analysis and related techniques. Sensors 2012; 12:9181-9209.

[27] Shi SY, Zhang YP, Jiang XY, *et al*. Coupling HPLC to on-line, post-column (bio)chemical assays for high-resolution screening of bioactive compounds from complex mixtures. Trends Anal Chem 2009; 28:865-877.

[28] Ciesla LM, Waksmundzka-Hajnos M, Wojtunik KA, Haknos M. Thin-layer chromatography coupled with biological detection to screen natural mixtures for potential drug leads. Phytochem Lett 2015; 11:445-454.

[29] Inglese J, Shamu CE, Guy RK. Reporting data from High-throughput screening of small-molecule libraries. Nat Chem Biol 2007; 438-441.

[30] Pauli GF, Chen SN, Friesen JB, McAlpine JB, Jaki BU. Analysis and Purification of Bioactive Natural Products: The AnaPurNa Study. J Nat Prod 2012; 75:1243-1255.

[31] Jenwitheesuk E, Horst JA, Rivas KL, Van Voorhis WC, Samudrala R. Novel paradigms for drug discovery: computational multitarget screening. Trends Pharmacol Sci 2008; 29:62-71.

[32] Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut microbiota modulation of chemotherapy efficacy and toxicity. Nat Rev Gastroenterol Hepatol 2017; doi: 10.1038/nrgastro.2017.20. [Epub ahead of print]

[33] Metzner KJ, Allers K, Rauch P, Harrer T. Rapid selection of drug-resistant HIV-1 during the first months of suppressive ART in treatment-naïve patients. AIDS 2007; 21:703-711.

[34] Griner LAM, Guha R, Shinn P, *et al*. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. PNAS 2014; 111:2349-2354.

[35] Arrell DK, Terzic A. Network systems biology for drug discovery. Clin Pharmacol Ther 2010; 88:120-125.

[36] Yeh JC, Garrard IJ, Cho CWC, *et al*. Bioactivity-guided fractionation of the volatile oil of *Angelica sinensis radix* designed to preserve the synergistic effects of the mixture followed by identification of the active principles. J Chromatogr A 2012; 1236:132-138.

[37] Elisabetsky E, Amador TA, Albuquerque RR, Nunes DS, Carvalho ACT. Analgesic activity of Psychotria colorata (Willd. Ex R. & S.) Muell. Arg. Alkaloids. J Ethnopharmacol 1995; 48:77-83.

[38] Amador TA, Elisabetsky E, Souza DO. Effects of Psychotria colorata alkaloids in brain opioid system. Neurochem Res 1996; 21:97-102.

[39] Amador TA, Verotta L, Nunes DS, Elisabetsky E. Involvement of NMDA receptors in the analgesic properties of psychotridine. Phytomedicine 2001; 8:202-206.

[40] Byers JA. Optimal fractionation and bioassay plans for isolation of synergistic chemicals: the subtractive-combination method. J Chem Ecol 1992; 18:1603-1621.

[41] Verpoorte R, Choi YH, Kim HK. Ethnopharmacology and systems biology: a perfect holistic match. J Ethnopharmacol 2005; 100:53-56.

[42] Vicini P, van der Graaf PH. Systems pharmacology for drug discovery and development: Paradigm shift or flash in the Pan. Clin Pharmacol Ther 2013; 93:379-381.

[43] Wu C, Kim HK, van Wezel GP, Choi YH. Metabolomics in the natural products field – a gateway to novel antibiotics. Drug Discov Today: Technol 2015; 13:11-17.

[44] Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends Biotechnol 1998; 16:373-378.

[45] Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 2002; 48:155-171.

[46] Shulaev V, Cortes D, Miller G, Mittler R. Metabolomics for plant stress response. Physiol Plant 2008; 132:199-208.

[47] Wolfender JL, Rudaz S, Choi YH, Kim HK. Plant metabolomics: from holistic data to relevant biomarkers. Curr Med Chem 2013; 20:1056-1090.

[48] Hong J, Yang L, Zhang D, Shi J. Plant metabolomics: an indispensable system biology tool for plant science. Int J Mol Sci 2016; 17:767.

[49] Kaplan F, Kopka J, Haskell DW, *et al*. Exploring the temperature-stress metabolome of Arabidopsis. Plant Phys 2004; 136:4159-4168.

[50] Foito A, Byrne SL, Shepherd T, Stewart D, Barth S. Transcriptional and metabolic profiles of Lolium perenne L. genotypes in response to a PEG-induced water stress. Plant Biotechnol J 2009; 7:719-732.

[51] Moradi P, Ford-Lloyd B, Pritchard J. Metabolomic approach reveals the biochemical mechanisms underlying drought stress tolerance in thyme. Anal Biochem 2017; Doi: 10.1016/j.ab.2017.02.006.

[52] Barros E, Lezar S, Anttonen MJ, *et al*. Comparison of two GM maize varieties with near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnol J 2010; 8:436-451.

[53] Chang Y, Zhao C, Zhu Z, *et al*. Metabolic profiling based on LC/MS to evaluate unintended effects of transgenic rice with cry1Ac and sck genes. Plant Mol Biol 2012; 78:477-487.

[54] Shepherd LVT, Hackett CA, Alexander CJ, *et al*. Modifying glycoalkaloid content in transgenic potato – metabolome impacts. Food Chem 2015; 187:437-443.

[55] Joshi L, Van Eck JM, Mayo K, *et al*. Metabolomics of plant saponins: Bioprospecting triterpene glycoside diversity with respect to mammalian cell targets. OMICS 2002. 6:235-246.

[56] Hagel JM, Mandal R, Han B, *et al*. Metabolome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. BMC Plant Biol 2015; 15:220.

[57] Shepherd LVT, Fraser P, Stewart D. Metabolomics: a second-generation platform for crop and food analysis. Bioanalysis 2011; 3, 1143-1159.

[58] Mazur SP, Sonsteby A, Nes A, *et al*. Effects of post-flowering environmental variation along an altitudinal gradient on chemical composition of 'Glen Ample' red raspberry (Rubus idaeus L.). Eur J Hortic Sci 2014; 79:267-277.

[59] Adhikari KB, Boelt B, Fomsgaard IS. Identification and quantification of loline-type alkaloids in endophyte-infected grasses by LC-MS/MS. J Agric Food Chem 2016. 64:6212-6218.

[60] Berhow MA, Polat U, Glinski JA, et al. Optimized analysis and quantification of glucosinolates from Camelina sativa seeds by reverse-phase liquid chromatography. Ind Crops Prod 2013. 43:119-125.

[61] Shulaev V. Metabolomics technology and bioinformatics. Brief Bioinform 2006; 7:128-139.

[62] Nielsen J, Oliver SG. The next wave in metabolome analysis. Trends Biotechnol 2005; 23:544-546.

[63] Courant F, Antignac JP, Dervilly-Pinel G, Bizec BL. Basics of mass spectrometry based metabolomics. Proteomics 2014. 14:21-22.

[64] Byrne SL, Foito A, Hedley PE, Morris JA, Stewart D, Barth S. Early response mechanisms of perennial ryegrass (Lolium perenne) to phosphorus deficiency. Ann Bot 2011. 107:243-254.

[65] Obata T, Witt S, Lisec J, *et al*. Metabolite profiles of maize leaves in drought, heat and combined stress field trials reveal the relationship between metabolism and grain yield. Plant Phys 2015; Doi: 10.1104/pp.15.01164.

[66] Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. Trends Anal Chem 2005; 24:285-294.

[67] Hantao LW, Ribeiro FAL, Passador MM, *et al.* Metabolic profiling by ultra-performance liquid chromatography-mass spectrometry and parallel factor analysis for the determination of disease biomarkers in *Eucalyptus*. Metabolomics 2014. 10:1318-1325.

[68] Lesiak AD, Cody RB, Dane AJ, Musah RA. Plant seed species identification form chemical fingerprints: a high-throughput application of direct analysis in real time mass spectrometry. Anal Chem 2015; 87:8748-8757.

[69] Cottet K, Kouloura E, Kritsanida M, et al. Comparative metabolomic study between African and Amazonian *Symphonia globulifera* by tandem LC-HRMS. Phytochem Lett 2017. (*in press*)

[70] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak. XCMS: Processing Mass spectrometry data for metabolite profiling using non-linear peak alignment, matching and identification. Anal Chem 2006; 78:779-787.

[71] Tautenhahn R, Patti GJ, Rinehart D, Siuzdak. XCMS: Online: a web-based platform to process untargeted metabolomics data. Anal Chem 2012; 84:5035-5039.

[72] Luedemann A, Strassburg K, Erban A, Kopka J. TagFinder for the quantitative analysis of gas-chromatography-mass-spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics 2008; 24:732-737.

[73] Kamleh A, Barrett MP, Wildridge D, Burchmore RJS, Scheltema RA. Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. Rapid Commun Mass Spectrom 2008; 22:1912-1918.

[74] Burch TC, Isaac G, Booher CL, *et al*. Comparative metabolomics and lipidomic analysis of phenotype stratified prostate cells. PLOS One 2015; Doi: 10.1371/journal.pone.0134206.

[75] Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: Modular framework for processing, visualizing and analysing mass spectrometry-based molecular profile data. BMC Bioinformatics 2010; 11:395.

[76] Lommen A. MetAlign: Interface driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem 2009; 81:3079-3086.

[77] Skogerson K, Wohlgemuth G, Barupal DK, Fiehn O. The volatile compound BinBase mass spectral database. BMC Bioinformatics 2011; 12:321.

[78] Sumner LW, Amberg A, Barrett D *et al*. Proposed minimum reporting standards for chemical analysis. Metabolomics 2007; 3:211-221.

[79] Dunn WB, Erban A, Weber RJM, *et al*. Mass Appeal: metabolite identification in mass-spectrometry –focused untargeted metabolomics. Metabolomics 2013; 9:44-66.

[80] Krishnan P, Kruger NJ, Ratcliffe RG. Metabolite fingerprinting and profiling in plants using NMR. J Exp Bot 2005; 56:255-265.

[81] Verpoorte R, Choi YH, Kim HK. NMR-based metabolomics at work in phytochemistry. Phytochem Rev 2007; 6:3-14.

[82] Troufflard S, Roscher A, Thomasset B, Barbotin JN, Rawsthorne S, Portais JC. In vivo 13C NMR determines metabolic fluxes and steady state in linseed embryos. Phytochemistry 2007; 68:2341-2350.

[83] Wolfender JL, Ndjoko K, Hostettmann K. The potential of LC-NMR in phytochemical analysis. Phytochem Anal 2001; 12:2-22.

[84] Hiller W, Sinha P, Hehn M, Pasch H. Online LC-NMR – From an expensive toy to a powerful tool in polymer analysis. Prog Polym Sci 2013; 39:979-1016.

[85] Aharoni A, de Vos CHR, Verhoeven HA, *et al*. Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS 2004; 6:217-234.

[86] McDougall G, Martinussen I, Stewart D. Towards fruitful metabolomics: high throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry. J Chromatogr b 2008; 871:362-369.

[87] Kopka J, Fernie AR, Weckwerth W, Gibon Y, Stitt M. Metabolite profiling in plant biology: platforms and destinations. Genome Biol 2004; 5:109.

[88] Kusano M, Kobayashi M, Iizuka Y, Fukushima A, Saito K. Unbiased profiling of volatile organic compounds in the headspace of Allium plants using nd in-tube extraction device. BMC Res Notes 2016; 9:133.

[89] Schauer N, Semel Y, Roessner U, *et al*. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnol 2006; 24:447:454.

[90] Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. Chemical Derivatization and mass spectral libraries in metabolite profiling by GC/MS and LC/MS/MS. J Exp Bot 2005; 56:219-243.

[91] Kopka J, Schauer N, Krueger S, *et al*. GMD@CSB.DB: the Golm metabolome database. Bioinformatics 2005; 21:1635-1638.

[92] Kind T, Wohlgemuth G, Lee DY, *et al*. FiehnLib – mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal Chem 2009; 81:10038-10048.

[93] van der Hooft JJJ, Vervoot J, Bino RJ, Beekwilder J, de Vos RCH. Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. Anal Chem 2011; 83:409-413.

878 [94] Iijima Y, Nakamura Y, Ogata Y, *et al*. Metabolite annotations based on the integration
879 of mass spectral information. Plant J 2008; 54:949-962.

880 [95]. Matsuda F, Nakabayashi R, Sawada Y, *et al*. Mass spectra-based framework for
881 automated structural elucidation of metabolome data to explore phytochemical diversity.
882 Front Plant Sci 2011; doi. 10.3389/fpls.2011.00040.

883 [96] Corcoran O, Spraul M. LC-NMR-MS in drug discovery. Drug Discov Today 2003;
884 8:626-631.

885 [97] Bingol K, Brüschweiler R Knowns and unknowns in metabolomics identified by
886 multidimensional NMR and hybrid MS/NMR methods. Curr Opin Biotechnol 2017; 43:17-
887 24.

888 [98] Bingol K, Brüschweiler-Li L, Yu C, Somogyi A, Zhang F, Brüschweiler R.
889 Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid
890 identification of new metabolites in complex mixtures. Anal Chem 2015; 87:3864-3870.

891 [99] Moore BD, Andrew RL, Külheim C, Foley WJ. Explaining intraspecific diversity in
892 plant secondary metabolites in an ecological context. New Phytol 2014; 201:733-750.

893 [100] Schwab W. Metabolome diversity: too few genes, too many metabolites?
894 Phytochemistry 2003; 62:837-849.

895 [101] Harborne JB, The Flavonoids Advances in Research since 1980. In: Harborne JB, The
896 Flavonoids Advances in Research since 1980. 1st ed. Chapman and Hall Ltd 1988; pp. 576-
897 578.

898 [102] Hall D, Yuan XX, Murate J, De Luca V. Molecular cloning and biochemical
899 characterization of the UDP-glucose:Flavonoid 3-O-glucosyltransferase from Concord grape
900 (Vitis labrusca). Phytochemistry 2012; 74:90-99.

901 [103] Agrawal AA, Hastings AP, Johnson MTJ, Maron JJ, Salminen JP. Insect herbivores
902 drive real-time ecological and evolutionary change in plant populations. Science 2012;
903 338:113-116.

904 [104] Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA.
905 Natural enemies drive geographic variation in plant defences. Science 2012; 338:116-119.

906 [105] Alba C, Bowers MD, Blumenthal D, Hufbauer R. Evolution of growth but not
907 structural or chemical defence in Verbascum Thapsus (common mullein) following
908 introduction to North America. Biol Invasions 2011; 13:2379-2389.

[106] Franks SJ, Wheeler GS, Goodnight C. Genetic variation and evolution of secondary compounds in native and introduced populations of the invasive plant Melaleuca quinquenervia. Evolution 2012; 66:1398-1412.

[107] Carmona D, Lajeunesse MJ, Johnson MTJ. Plant traits that predict resistance to herbivores. Funct Ecol 2011; 25:358-367.

[108] Jones CG, Firn RD. On the evolution of plant secondary chemical diversity. Philos Trans R Soc Lond B Biol Sci 1991; 333:273-280.

[109] Singh R. Chemotaxonomy: A tool for plant classification. J Med Plants 2016; 4:90-93.

[110] Xie G, Plumb R, Su M, *et al*. Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomics research. J Sep Sci 2008; 31:1015-1026.

[111] Lee S, Oh D-G, Lee S, *et al*. Chemotaxonomic metabolite profiling of 62 indigenous plant species and its correlations with bioactivities. Molecules 2015; 20:19719-19734.

[112] Grimbs A, Shresta A, Resk ASD, *et al*. Bioactivity in Rhododendron: a systemic analysis of antimicrobial and cytotoxic activities and their phylogenetic and phytochemical origins. Front Plant Sci 2017; Doi: 10.3389/fpls.2017.00551.

[113] Freiburghaus F, Steck A, Pfander H, Brun R. Bioassay-guided isolation of a diastereoisomer of kolavenol from Entada abyssinica active on Trypanosoma brucei rhodesiense. J Ethnopharmacol. 1998; 61:179-183.

[114] Inayatullah S, Prenzler PD, Obied HK, Rehman A-U, Mirza B. Bioprospecting traditional Pakistani medicinal plants for potent antioxidants. Food Chem 2012; 132:222-229.

[115] Kainz KP, Krenn L, Erdem Z, *et al*. 2-Deprenyl-rheediaxanthone B isolated from Metaxya rostrate induces active cell death in colorectal tumor cells. PLOS One 2013; doi: 10.1371/journal.pone.0065745.

[116] Gogna N, Hamid N, Dorai K. Metabolomic profiling of the phytomedicinal constituents of Carica papaya L. leaves and seeds by 1H NMR spectroscopy and multivariate statistical analysis. J Pharm Biomed Anal 2015; 115:74-85.

[117] Geng C-A, Chen J-J. A fragmentation study on four unusual secoriridoid trimers, swerilactones H-K, by the electrospray tandem mass-spectrometry. Nat Prod Bioprospect 2016; 6:297-303.

[118] Nielsen KF, Mansson M, Rank C, Frisvad JC, Larsen TO. Dereplication of microbial natural products by LC-DAD-TOFMS. J Nat Prod 2011; 74:2338-2348.

[119] Yang JY, Sanchez LM, Rath CM, *et al*. Molecular Networking as a dereplication strategy. J Nat Prod 2013; 76:1686-1699.

[120] McIntyre L, Zhang T, Viegelmann C, *et al*. Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. Mar Drugs 2014; 12:3416-3448.

[121] Gerwick WH, Moore BS. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. Chem Biol 2012; 19:85-98.

[122] Kurita KL, Glassey E, Linington RG. Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. PNAS 2015; 112:11999-12004.

[123] Kind T, Fiehn O. Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. Phytochem Lett 2016; doi: 10.1016/j.phytol.2016.11.006.

[124] Jiang Y, Liu F-J, Wang Y-M, Li H-J. Dereplication-guided isolation of novel hepatoprotective triterpenoid saponins from Celosiae Semen by high-performance liquid chromatography coupled with electrospray ionization tandem quadrupole-time-of-flight mass spectrometry. J Pharm Biomed Anal 2017; 132:148-155.

[125] Moore JM. NMR screening in drug discovery. Curr Opin Biotechnol 1999; 10:54-58.

[126] Cala O, Guilliere F, Krimm I. NMR-based analysis of protein-ligand interactions. Anal Bioanal Chem 2014; 406:943-956.

[127] Wilme A, Limonciel A, Aschauer L, *et al*. Application of integrated transcriptomic, proteomic and metabolomics profiling for the delineation of mechanisms of drug induced cell stress. J Proteomics 2013; 79:180-194.

[128] Van den Hof WFPM, Ruiz-Aracama A, Summeren AV, *et al*. Integrating multiple omics to unravel mechanisms of Cyclosporin A induced hepatoxicity in vitro. Toxicol In Vitro 2015; 29:489-501.

[129] Pimpão, R.C., Dew, T., Figueira, M.E., *et al*. Urinary metabolite profiling identifies novel colonic metabolites and conjugates of phenolics in healthy volunteers. Mol Nutr Food Res 2014; 58:1414-1425.

[130] Zhang L, Hatzakis E, Patterson AD. NMR-based metabolomics and its application in drug metabolism and cancer research. Curr Pharm Rep 2016; 2:231-240.

[131] Sun J, Schnackenberg LK, Beger RD. Studies of acetaminophen and metabolites in urine and their correlations with toxicity using metabolomics. Drug Metab Lett 2009; 3:130-136.

[132] Sun J, Schnackenberg LK, Khare S, *et al*. Evaluating effects of penicillin treatment on the metabolome of rats. J Chromatogr B 2013; 932:134-143.

[133] Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. J Pharm Biomed Anal 2014. 87:1-11.

[134] Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. J Pharm Biomed Anal 2015. 107:63-74.

[135] Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: Beyond biomarkers and towards mechanisms. Nat Revs Mol Cell Biol 2016. 17:451-459.

[136] Jin X, Yun SJ, Jeong P, *et al.* Diagnosis of bladder cancer and prediction of survival by urinary metabolomics. Oncotarget 2014. 30:1635-1645.

[137] Ren S, Shao Y, Zhao X, *et al.* Integration of metabolomics and transcriptomics reveals major metabolic pathways and potential biomarker involved in prostate cancer. Mol Cell Proteomics 2016. 15:154-163.

[138] Armitage EG, Southam AF. Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. Metabolomics 2016. 12:146.

[139] Schuler KM, Rambally BS, DiFurio MJ, *et al*. Antiproliferative and metabolic effects of metformin in a preoperative window clinical trial for endometrial cancer. Cancer Med 2015. 4:161-173.

[140] He J, Wang K, Zheng N, et al. Metformin suppressed the proliferation of LoVo cells and induced a time-dependent metabolic and transcriptional alteration. Sci Rep 2015. 5:17423.

[141] Kurita KL, Glassey E, Linington RG. Integration of high-content screening of untargeted metabolomics for comprehensive functional annotation of natural product libraries. Proc Natl Acad Sci 2015; 112:11999-12004.

[142] Aliagiannis N, Halabalaki M, Chaita E, *et al*. Heterocovariance based metabolomics as a powerful tool accelerating bioactive natural product identification. ChemistrySelect 2016; 1:2531-2535.

FIGURES

Figure 1 – Diagram representing the two leading approaches utilised in bioprospection of bioactive compound: pure compound screening (left) and bioassay-guided fractionation (right).

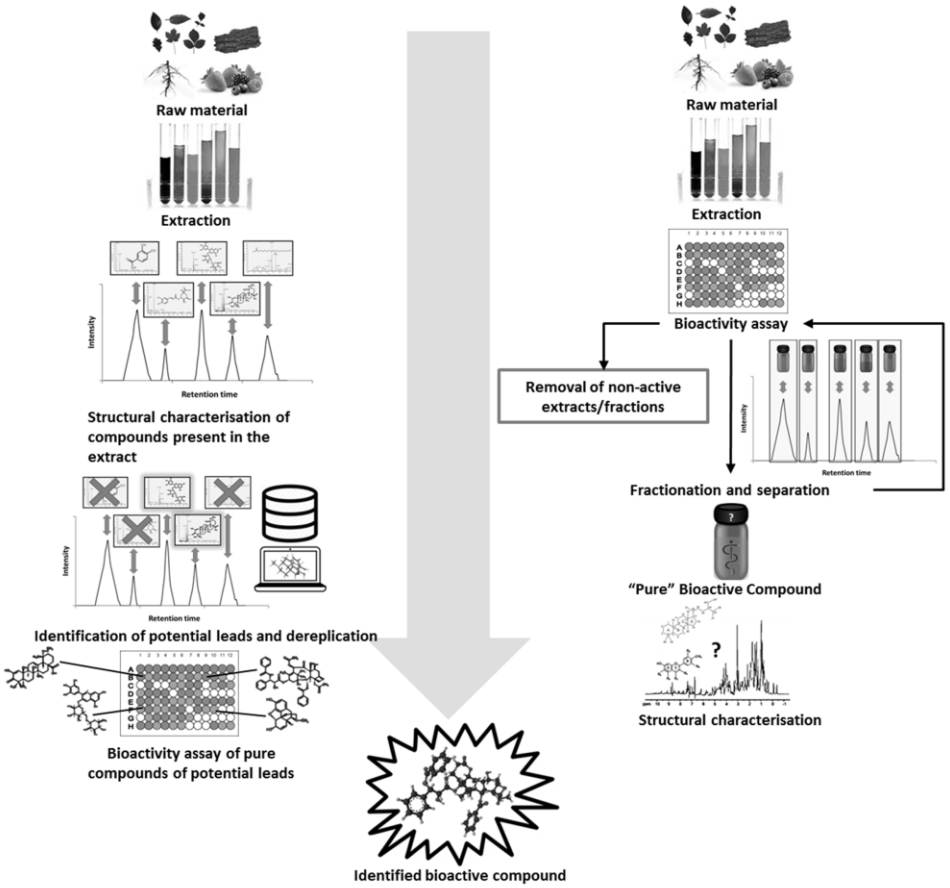
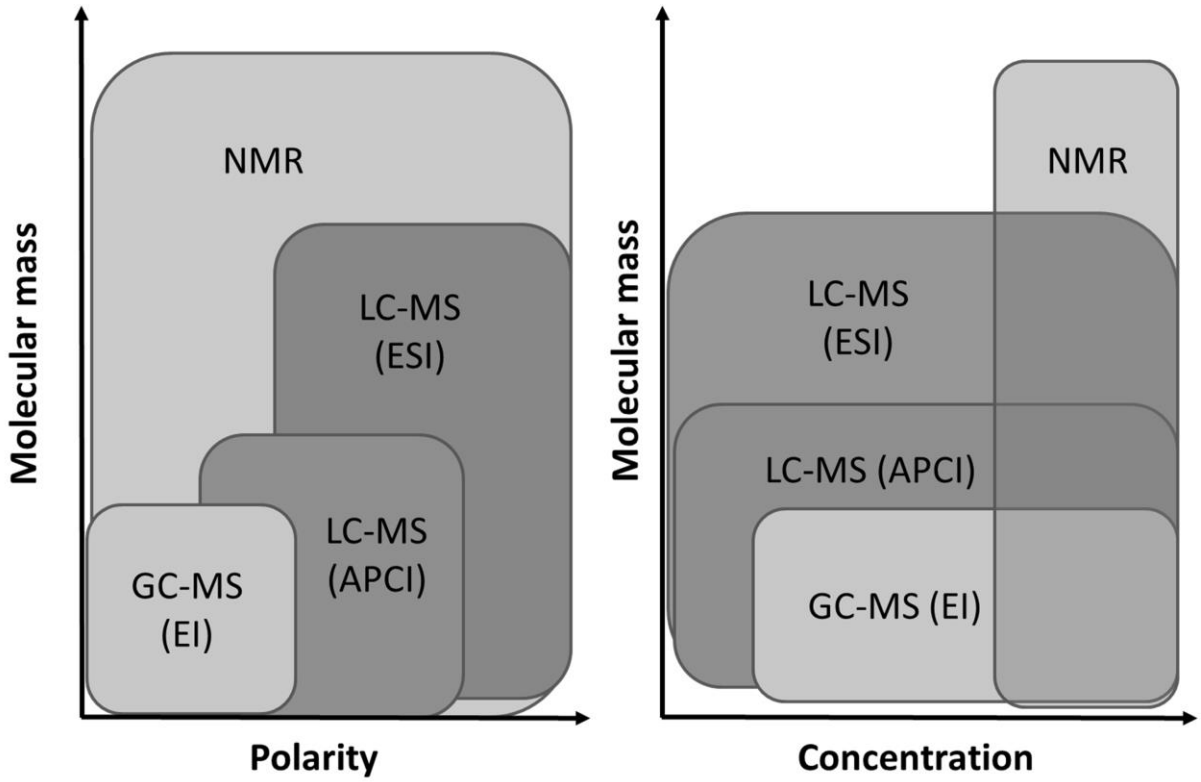
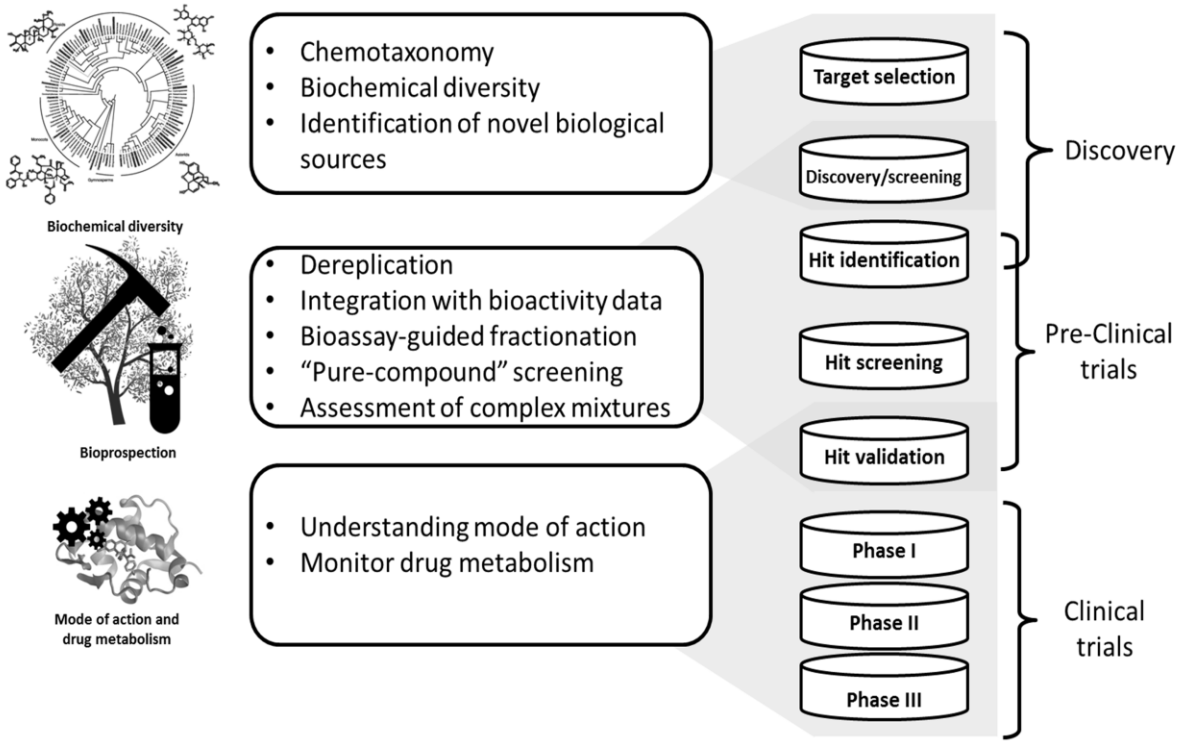


Figure 2 – Comparison of relative analyte coverage with respect to molecular mass, polarity and concentration ranges across different analytical platforms.



1011 **Figure 3** – Potential applications of metabolomics tools in the drug discovery workflow



1012