Metabolomic tools used in marine natural product drug discovery

Kevin Andrew Stuart[a], Keira Welsh[a] Molly Clare Walker[a], and RuAngelie Edrada-Ebel[a]*,

[a]*Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, The John Arbuthnott Building, 161 Cathedral Street, Glasgow G4 0RE, UK*

*corresponding author: ruangelie.edrada-ebel@strath.ac.uk

Introduction: The marine environment is a very promising resource for natural product research, with many of these reaching the market as new drugs, especially in the field of cancer therapy as well as the drug discovery pipeline for new antimicrobials. Exploitation for bioactive marine compounds with unique structures and novel bioactivity such as the isoquinoline alkaloid; trabectedin, the polyether macrolide; halichondrin B, and the peptide; dolastatin 10, requires the use of analytical techniques, which can generate unbiased, quantitative and qualitative data to benefit the biodiscovery process. Metabolomics has shown to bridge this understanding and facilitate the development of new potential drugs from marine sources and particularly their microbial symbionts.

Areas covered: In this review, articles on applied secondary metabolomics ranging from 1990-2018 as well as to the last quarter of 2019 were probed to investigate the impact of metabolomics on drug discovery for new antibiotics and cancer treatment.

Expert opinion: The current literature review highlighted the effectiveness of metabolomics in the study of targeting biologically active secondary metabolites from marine sources for optimized discovery of potential new natural products to be made accessible to a R&D pipeline.

Keywords: secondary metabolomics; marine natural products; antibiotics, anti-cancer drugs, NMR, MS, drug discovery, molecular networking, multivariate analysis

Article Highlights

- Metabolomics is an effective analytical tool to aid in drug discovery. A metabolomics approach utilizes different high-resolution analytical techniques and combinations thereof, that is crucial in the discovery of novel drugs overcoming flaws in targeting the new bioactive molecules earlier in the workflow thereby reducing the cost of development failure at a later stage.
- The employment of both multivariate analysis (MVA) and molecular networking (MNW) in a metabolomics-based dereplication of metabolites effectively avoids the bioassay steps used in bioactivity-guided fractionation and prevents a tedious isolation work on known compounds.
- Metabolomics affords a "snap shot" of the metabolome and the biosynthetic potential of an organism while genomics complementarily expands on this. Genomic screening of biosynthetic gene clusters of marine microorganisms that are difficult to mass cultivate prior to metabolomics analysis would allow prediction of the occurrence of novel bioactive secondary metabolites and eliminate strains with low biosynthetic potential from further scale-up to save time and resources being wasted.
- A metabolomics-targeted approach has the ability to pinpoint the bioactive metabolite on the first fractionation stage even at low mg levels while rendering subsequent bioassays unnecessary and the chance of isolating the target natural product at higher yield is increased.

1 Introduction

1.1 Marine drugs on the market

Around 70% of the world's surface is covered by ocean mass. Between 1985 and 2012, 1241 new natural compounds from marine sources were identified and elucidated [1,2]. This represents a vast untapped resource for the isolation of potentially new natural products (NPs) with therapeutic applications. In 2014, there were eight drugs of marine origin approved by the Food and Drug Administration (FDA) and European Medicines Agency (EMA), while twelve NPs (or derivatives thereof) are currently in various phases of clinical trials [2].

The first marine-derived drug to be approved was Cytosar-U® (cytarabine) in 1969 for the treatment of non-Hodgkin's lymphoma and meningeal leukemia; followed by Vira-A® (vidarabine) in 1976 as an antiviral, prescribed against herpes, pox, and some rhabdoviruses [2]. Cytarabine and vidarabine are synthetic nucleosides developed from isolates of the sponge *Tethya crypta* [3]. The mechanism of action of cytarabine is not yet fully understood, though it appears to be intracellularly converted to its active form cytarabine triphosphate, which competes with deoxycytidine triphosphate as a substrate for DNA polymerase, inhibiting DNA synthesis [2]. Vidarabine has a higher toxicity and is less stable than more popular antiviral agents, and exhibits a low oral bioavailability and poor solubility [4], which contributed to its withdrawal from use in the US in 2001 [2]. Cytarabine remains a staple in cancer therapy today [2].

Prialt® (ziconotide) was granted approval by the FDA and EMA, in 2004 and 2005 respectively, as an intrathecal anesthetic. Ziconotide is a synthetic analogue of a peptide isolated from *Conus magus* [2,3]. It acts by blockading the N-type voltage gated calcium channel in neuronal cells, preventing pain neurotransmitter release [5].

Other currently approved marine drugs include:

- Yondelis® (trabectedin®), a novel alkaloid with a unique mechanism of action as it acts to bind the minor groove of the DNA supercoil and inhibit DNA repair machinery [2,6]. It was isolated from the tunicate *Ecteinascidia turbinata* [3].
- Halaven® (erbulin mesylate), used in metastatic breast cancer, is a synthetic analogue of the cytotoxic halichondrin B, isolated from the sponge *Halichondria okadai* [3].

Having established that marine organisms are genetically and chemically unique [7], it follows that the continued study of these organisms will lead to the discovery of a plethora of NPs with biologically active properties [8,9]. Referred to as secondary metabolites [2], these chemicals are not known to serve any primary function related to growth, development, or propagation of an organism [7]. Rather these NPs are more commonly associated with defense and survival [10], competition, and communication [2]. In its infancy, the study of marine natural products (MNPs) was restricted to easy-to-collect organisms such as soft corals and sponges; leading to the isolation of chemicals such as the toxin tedanolide from the fire sponge, *Tedania ignis*, and prostaglandins from the Caribbean sea whip, *Plexaura homomalla* [11]. It has been shown that vast number of secondary metabolites possess a range of biologically active properties which are unrelated to their *in vivo* role to aid marine organisms to adapt to and survive in their respective ecosystem [12]. These include cytotoxic, neurotoxic, anti-inflammatory, and anti-infective properties [13], all of which, after further study and development may have pharmaceutical applications. A major obstacle that has arisen, however, is the largely uncultivable nature of marine invertebrates. This paired

with an unsustainable collection of insufficient samples for the extraction of an adequate yield of NPs to be used for further R&D has caused a bottleneck in marine drug development [14].

Hypothesized to stabilize the sponge matrix, provide waste metabolism, and chemical defenses, symbiotic bacteria can constitute between 50 and 60% of a sponge's dry weight. They serve these functions by production of an immense range of secondary metabolites possessing unique chemical structures [15-17] that included alkaloids [18], peptides [19], and polyketides [20], as well as terpenoids, polyhydroxylated lipids [21], and halogenated compounds [22] with novel pharmacological activities or mechanism of action [14]. Marine invertebrates such as sponges and soft corals represent an excellent source of marine microorganisms for drug discovery and development [23]. Of particular interest is the drug Adcretis® (brentuximab vetodin) used in the treatment of Hodgkin lymphoma [2]. Adcretis® is a chimeric anti-CD30 antibody, conjugated to a synthetic analogue of the peptide, dolastatin 10 [11,24]. Dolastatin 10 was initially isolated from *Dolabella auricularia*, but later it was discovered that it is acrtaullay produced by the cyanobacteria, *Symploca hydnoides* and *Lyngbya majuscula*, that constitute part of *D. auricularia*'s diet [2]. Dolastatin 10 was primarily described as an antitumour and antimitotic agent by inhibiting tubulin polymerisation [25-27]. It has since been indicated that many NPs isolated from marine organisms may in fact be synthesized by microorganisms [12].

1.2 A brief history of metabolomics

Metabolomics is the identification and quantification of metabolites that are end products of a biological process in tissues, cells, or an organism [28]. The methodology involves high analytical precision and comprehensiveness allowing an unbiased quantitative and qualitative evaluation of biochemical precursors, intermediates, and

end products in a sample [28]. Metabolomics has been a growing field that entails the measurement of small molecules in a system that comprises of metabolite profiling and metabolic flux analysis [29] to contour changes occurring in living systems in response to numerous genetic and environmental factors. A metabolomics-based approach attempts to express the global and physiological changes to reveal information on the metabolic pathway involved in a cellular process at a molecular level [30]. In pharmacology, the understanding of the down and up regulation of certain metabolites allows the disclosure of disease mechanisms, recognition of new diagnostic and prognostic markers while increasing our comprehension on drug response phenotypes [31]. Metabolites do play essential roles in the cell and membrane structure (lipids), signaling (lipid mediators, neurotransmitters), and building blocks (amino acids). They are also involved in many processes affected by exogenous influences; such as response to oxidative stress (redox potential), inflammatory response, and energy metabolism [29]. Detecting and validating changes in metabolite levels in diseased individuals had become feasible with the rapid progress in instrumentation technology, and therefore evaluation of metabolomics data could essentially play a more important role in clinical practice [30].

The process of metabolomics has been around for many years. The first records were documented in ancient Chinese cultures (200-1500 BC) with the use of body fluids to establish the occurrence of a biological disorder. Chinese physicians determined the concentration of glucose in the urine of diabetic patients by using ants as detectors [31]. In the 16th and 17th century, influential discoveries of Santorio Sancatorius (1561-1636) on his studies on perspiration contributed to the basic foundation of metabolomics. Sancatorius created a transportable platform that was attached to a steelyard scale to calculate the changes in body weight of his subjects who participated on the platform

when studying the effects of perspiration [32]. Later, the German chemist Eduard Buchner (1860-1917) opened the field of biochemistry when he demonstrated the importance of enzymes in intracellular chemical reactions and biochemical pathways [33]. An early case of metabolite profiling was reported in 1948 when differences in alcoholic and schizophrenic patients were investigated by analyzing patterns of respective metabolic components found in their urine and saliva by paper chromatography [34]. It was not after several decades that technological advances in both chromatography and spectrometry were later able to quantify metabolites in biological samples. In the early 20th century, new technologies in mass spectrometry (MS) were developed to quantify molecules in biochemical samples, which then were able to examine potential factors that causes the release of metabolites in disease states [35]. In 1971, compounds present in body tissues could already be quantified by gas chromatography-mass spectrometry (GC-MS) [36]. Meanwhile, nuclear magnetic resonance (NMR) spectroscopy was also introduced for quantification [37,38], mixture analysis [39,40], and to detect metabolites in raw biological samples [41]. A promising breakthrough was perceived in 2007, when the Human Metabolome Project completed its first draft. Today, metabolomics research is not only being used in human studies but also in yeast, fungi, insects and plant models [31].

In recent years, there has been an increasing interest on employing a metabolomics approach in data analysis and interpretation. The use of genomics, transcriptomics, and proteomics has allowed understanding biology at a molecular level, and now there is a demand to widen this knowledge base to further explore the processes within a biological system [42]. Prior to the advent of metabolomics, there has been a primary focus on genomics and proteomics. Although these fields are useful, they can only offer a depiction of the potential outcomes of a biological system.

Metabolomics affords evidential information of the end products of a cellular process, and thus, providing a direct correlation between the phenotype of a biological system and the occurring metabolic processes within. As this applies to both healthy and diseased systems, it follows that metabolomics will not only provide information pertaining to disease markers, but also possible targets to control the aberrant system itself [43]. In medicine, the use of metabolomics was initially restricted to diagnosis, particularly in metabolic disorders [44,45]. However, the ability of metabolomics to profile a significantly larger number of molecules than other standardized techniques promises to present it as an invaluable tool in the future of molecular medicine, drug development, and drug target discovery [44]. This is even more evident when we consider that the number of known, expected and predicted human metabolites continues to grow, that is currently documented at 114,100 in the year 2018, the most recent update on the Human Metabolome Database [46]. As the field has continued to grow and evolve, applications have been discovered spanning the length of R&D and the drug discovery pipeline [47].

The concentrations of different metabolites within a biological system are often considered as a metabolic "fingerprint". A metabolic fingerprint is defined as being the "representative of the state of the organism" or cell at the time of study. Metabolomics utilizes the techniques discussed previously to quantify these metabolites, and build a temporal picture of the responses to stimuli displayed by cells and organisms[48]. This enables the identification of secondary metabolites which may be potential targets for drug therapy or, indeed, precursor molecules for drug discovery and development.

A more recently developed approach involves the study of metabolic "footprints". This approach allows the understanding of the effects of various external/internal signals on a biological system. Contrary to fingerprinting, where the

internal metabolome of a cell is studied and defined, footprinting defines the metabolites secreted, excreted, and consumed by a cell or organism [49]. This approach eliminates the need for lengthy extraction techniques or rapid metabolism quenching. Footprinting is of particular interest in the study of pathogenic microorganisms as much of their pathogenesis is dependent on environmental factors e.g. availability of nutrients and the presence of other microbes [49] or simply the ability of an organism to respond to its environment.

This paper aims to provide an unbiased review on how metabolomics is being increasingly utilized in the field of NP drug discovery, to expand the readers' knowledge of metabolomics as a viable analytical option. Moreover, to address the limitations of a metabolomics approach in terms of data interpretation in the translation process of bringing NPs to a drug development pipeline. The main objective is to construct a clear image of the direct applications of metabolomics in NPs research by critically analyzing sample studies where a metabolomics approach was utilized to exploit the discovery and development of new potential drugs. This study will address advantages of this approach in terms of feasibility and time required for a newly discovered natural product to come into the R&D pipeline. This will be achieved by looking into the different analytical techniques and multivariate analysis methods used in metabolomics for profiling, targeting, isolation and elucidation of the bioactive NPs. In addition, these aims will be achieved by evaluating the advantages and disadvantages of different metabolomics tools that include molecular networking and metabolomic-guided isolation of bioactive metabolites. Examples of the direct applications of metabolomics in NPs research will be presented by scrutinizing sample studies on the discovery of new potential antibiotics and anti-cancer drugs as well as illustrating the

advantages of a metabolomics-based approach in increasing the efficiency of NPs research for drug discovery.

2 Analytical techniques used in metabolomics

The study of small molecules, coined as metabolomics, is a relatively new scientific field, which focuses on their characterization using a number of techniques such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled with chromatography [47]. It enables a unique "top-down approach" for the study of complex biological systems [48]. The variety and wide range of analytical techniques that includes mass spectrometry [50-56], NMR spectroscopy [57-63], and infrared spectrophotometry [64-69], used within metabolomics has contributed to its rapid growth. The current tools made it possible to study thousands of metabolites in complex biological samples [70]. Due to the intricacy of a biological system and the chemical diversity of the metabolome, there is not one single analytical program or technique available to recognize all different metabolites in a biological model. To increase coverage of detected metabolites, new methods are being combined to increase the efficiency of detection and dereplication [71]. To date, NMR and MS have been the two main analytical tools used in generating metabolomics data. However, other analytical hyphenated-platforms like gas-chromatography mass spectrometry (GC-MS), high-performance liquid-chromatography (HPLC), ultra-performance liquid chromatography (UPLC) and combinations thereof are being utilized more effectively for metabolite separation, quantification, and elucidation [71]. While NMR can rapidly produce accurate and precise information, appropriate for the examination of bulk materials, the high sensitivity of mass spectrometry allows for the qualitative and quantitative analysis of many metabolites.

2.1 Mass Spectrometry

Mass spectrometry (MS) plays a massive role in metabolomics and it is the analytical method of choice when investigating small molecules. In metabolomics studies, MS has the capability of profiling the effect of time as well as other environmental factors on metabolite regulation. To date, MS has become a highly sensitive analytical tool used not only for quantification, but as well as for identification of hundreds more metabolites at lower concentrations [72]. However, often metabolites require to be extracted for MS analysis prior to acquisition [73]. The use of mass spectrometry (MS) has been gaining increasing interest for metabolomics applications [72] . One limitation of MS is the required separation step to reduce sample complexity and minimize ionization suppression effects. MS has a significantly higher sensitivity and resolution than NMR, allowing it to detect a huge number of metabolites [48]. However, detection is dependent on the metabolites' capability to ionize in a particular analyzer or ionization method and mode. A number of MS detectors are utilized for analysis subsequent to chromatographic separation. Applications and limitations of various types of mass analysers used as MS detectors are presented in Table 1 [74]. MS techniques are able to quantitatively profile the metabolites in samples depending on the ionization capability of the congeners.

****Please add Table 1 here*****

It is common for MS to be used in a hyphenated separation technology such as with liquid and gas chromatography [72]. Such hyphenated separation techniques commonly used includes GC-MS, HPLC or UPLC [75,76]. Hyphenated or coupled systems combine both the advantages of chromatographic and spectral methods to

increase the efficiency in the identification of metabolites with diverse chemistry. Chromatography affords unadulterated segments of biological compounds, whereas, spectrometry and/or spectroscopy produces selective information for the identification of the metabolites by employing reference standards or library spectra [76]. The importance of linking both separation and spectroscopic techniques ensures that the identification and quantification of the metabolites in an unknown sample can be properly dereplicated and evaluated. The multi-step nature of this type of analysis makes it slower than NMR [48]. To obtain structural information for the respective metabolites, a chromatographic system, like an HPLC or GC, could be equipped with different spectrometric or spectroscopic detectors, such as photo-diode array (PDA), UV, MS, NMR and Fourier-transform infrared (FTIR) [76]. This has resulted in a variety of coupled-systems being set-up e.g. LC-MS, GC-MS, CE-MS and HPLC-MS/NMR (Table 2) [76-78]. GC-MS was the first hyphenated system used in analytical research and drug development. Advantages and disadvantages of various types of chromatographic separation are presented in Table 3 [74,77,79].

*****Please add table 2 here******

*****Please add table 3 here******

Gas chromatography coupled with mass spectrometer (GC-MS) was also the first analytical technique used in metabolite profiling. This technique separates metabolites from a complex mixture according to the compound's volatility and molecular weight. Clinical research studies in the early 1970s used GC-MS techniques to analyze steroids, acids and drug metabolites from both human and rat urine [36]. It was not until the 1980s that the diagnostic potential of GC-MS was utilized to analyze

metabolic disorders as in testing for inborn errors of metabolism [80]. Advancements in computer-operated systems as well as decreased costs permitted GC-MS to be readily employed in more laboratories in the 1990s [81]. For MS coupled to GC, with electron impact (EI) as the ionization method usually at 70 eV, this results to a higher incidence of fragmentation that the spectra could be compared with an online library. However, GC-MS could only analyze small volatile molecules that are thermally stable, particularly during the fragmentation process [81].

Liquid chromatography coupled with mass spectrometer (LC-MS) was used to separate a complex mixture of compounds according to their polarity. During the early development of LC-MS, there was an incompatibility issue in combining LC and MS. This issue was resolved in the 1980s with the discovery of electrospray ionization (ESI), which is compatible to both the LC and MS process [73]. LC-MS or HPLC-MS combines the separating ability of LC with the selective proficiency of an ESI-MS. This technique determines not only the molecular weight of a molecule but also the fragmentation pattern of respective molecules in a sample. Soft ionization techniques are generally used in LC-MS which is commonly restricted to displaying only the molecular ion species of a few fragment ions resulting to a less complex set of spectra.

Mass spectral data can be processed with a quantitative differential expression profiling platform software for metabolite profiling preceding statistical or multivariate (MVA) analysis. The steps entail deconvolution, nonlinear peak alignment, peak normalization, peak matching, and identification that may include molecular formula prediction from high resolution datasets. A number of software are available online *via* open access. These include Mzmine 2.0 [82-84] and mzMatch [85], both of which are modularized tools coded in JAVA, a very flexible software that is easy to couple to a variety of databases [14]; XCMS is coupled to the METLN database [86]; Galaxy-M for

processing and analyzing both direct infusion and liquid chromatography mass spectrometry-based metabolomics data [87]; MetSign for high-resolution MS data[88]; MAIT (Metabolite Automatic Identification Toolkit) in R package[89]; MET-COFEA for **MET**abolite **co**mpound **f**eature **e**xtraction; and annotation [90] and Mass-Up is an open software application for MALDI-TOF mass spectrometry [91]. For all the above-mentioned software, it is necessary to convert raw data files to mzXML format [92,93].

2.2 Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is another technique commonly employed in metabolomics [71,73]. NMR is rapid and does not require complex sample preparation therefore; the sample's biological integrity is maintained [48]. It is straight-forward to use and non-destructive to the samples that they can be reused for further analysis. Several classes of metabolites can be analyzed simultaneously using NMR to gain an insight into the purity and molecular structure of metabolites in a sample under investigation. NMR can be used to study metabolites from both biological fluids such as urine, saliva, solid tissue samples from biopsy as well as extracts from other natural sources [94]. It has been widely used for metabolite fingerprinting, profiling and metabolite flux analysis [71]. The obtained NMR spectral data is highly reproducible and quantitative offering advantages for compounds that are difficult to ionize or would require to be derivatized for MS analysis [95]. However, the key limitation of NMR compared to MS, is its low sensitivity and does not allow for the investigation of a large proportion of metabolites at low abundance [42]. Hence, biomarkers may remain undetected if their concentrations are too low requiring increased number of scans and therefore longer measurement time or could also be resolved by employing higher magnetic fields [71]. NMR can exclusively identify and concurrently quantify a wide range of organic compounds but only above micro-molar concentrations [71].

Natural products research has become unimaginable without the use of high resolution chromatographic separation techniques coupled with UV and/or IR detectors then later combined with mass spectrometry (MS). But more recently, nuclear magnetic resonance (NMR) has become a valuable tool for natural product dereplication for metabolic profiling and metabolic fingerprinting applications [96-98]. Hyphenated HPLC-NMR is a fast growing technology, allowing rapid and detailed structural characterization of unknown mixtures [99]. Pioneering works on related flow probe NMR technologies has seen evolved to well established analytical platforms around the turn of the century [97,98]. Various applications of HPLC-NMR and integrated HPLC-NMR-MS in drug discovery, especially in the separation and structure elucidation of drug impurities, reaction mixtures, degradation products, in vitro and in vivo metabolites, and combinatorial library samples have been illustrated in the literature [100]. However, analysis of mixtures or crude extracts remains to be a challenge to interpret depending on the number of overlapping signals present particularly when secondary metabolites with complex spin patterns are investigated. The complexity of a secondary metabolomic profile of a natural product derived extract could make the analytical task arduous and toilsome in terms of requiring a highly specialized staff expertise as well as higher overhead costs.

2.3 Multivariate analysis (MVA)

Multivariate analysis (MVA) presents visualization plots to gain an understanding of the dataset trends produced from tested samples. There are many techniques employed that allow the conversion of highly multivariate data into compatible and interpretable sized data.

2.3.1 Principle component analysis (PCA)

Principle component analysis (PCA) is an unsupervised method, most commonly used to analyze multivariate data. PCA is mainly used for a hypothesis free explorative analysis [101]. The aim of a PCA is to define datasets by converting large number of correlated variables into smaller linear datasets [102]. The result is generated into a scores scatter plot, which express patterns that reflect fundamental arrangements present in the dataset divided into separate clusters by grouping the samples into specific metabolic phenotypes [103]. PCA permits the identification of patterns in the data matrix by producing clusters of columns (samples) or rows (features) that have comparable designs [104]. These clusters facilitate the detection of "outliers", samples that are numerically distant from other data points [105]. Often outliers are caused by a measurement, experimental, or sampling error and can be removed from the analysis. However, the outlier should not be removed if it is caused by a natural process of the construct that is being measured [105].

Advantages of PCA include low noise sensitivity, a decreased need for capacity and memory, and higher efficiency for processes occurring in a smaller dimension. PCA has a smaller database representation through trainee images stored and predicted on a reduced base making it easier to inter-operate a metabolomics dataset. PCA shows differences between groups by identifying respective distinguishing metabolic features [106]. However, PCA has its limitations requiring further validation when handling huge datasets. PCA may significantly fail to disclose underlying groups of unidentified variables resulting to a false image of the original data structure. Barnes *et al*. suggested the need to move the metabolomics software onto the cloud to save money on standalone software and computer hardware when handling large datasets [107]. PCA had also its limitation on second order statistical dependencies between variables.

Further pitfalls partially remained in the interpretation of results on a reduction analysis dimension, when the data is normalized. Further, simple variance could not be captured by PCA unless a trained dataset explicitly provides this information [108]. PCA only operates on the matrix data and cannot take into account the additional information that might be associated with the datasets. Hence, orthogonal projections to latent structures discriminant analysis (OPLS-DA) is used to solve this problem.

2.3.2 Orthogonal projections to latent structures discriminant analysis (OPLS-DA)

OPLS-DA is a supervised method, which separates Y-predictive from Y-orthogonal variation, facilitating the grouping of unrestrained data by extracting information on changing molecular composition of a sample [103]. OPLS-DA was designed for modelling two or more classes of data to increase and simplify class separation. A response value is openly assigned to each measurement. It is a predictive regression method that finds information in the data relating to known information and identifies discrete variables. These discrete variables can include factors such as environmental, temporal, dietary, physiological, drug treatments, sex, genetic, age and bioactivity are among those that can affect the clustering and considered as grouping or class factors in OPLS-DA [109]. OPLS-DA is primarily useful in classification studies and biomarker identification by pinpointing discriminating features for each of the respective variables. The pinpointed feature can then be preliminarily dereplicated through a quantitative differential expression profiling platform software such as mzMine 2.0. OPLS-DA has the unique advantage of interpreting data by showing which feature(s) is/are responsible for the class discrimination. It requires a second matrix of information linking to each sample to reduce the impact of unrelated data points on the data matrixes. This approach allows for the better representation of different defining variables as it is expressed on a continuum rather than binary scale. In comparison to

PCA, OPLS-DA is able to separate experimental groups within samples. It can do this due to its integrated orthogonal signal correction (OSC) filter that eliminates variation within the systematic spectrum that does not belong to the allocated groups [110]. However, OPLS-DA can often yield false positive results and require rigorous cross-validation to safeguard reliability [110]. Disadvantages of OPLS-DA includes the risk of overfitting and incorporating noise into the statistical model. Cross validation techniques like permutation tests could overcome this problem [111].

To increase specificity and reliability, both PCA and OPLS-DA should be used. As PCA does not have group specificity information, the plots generated on a scatter graph are essentially the least biased method of assessing trends and discrimination capacity. However, PCA is not always able to expose group separations and require the use of OPLS-DA to identify certain unnoticed groups or features. The use of both PCA and OPLS-DA on spectral dataset results provides important insights on both the general spectral trends (PCA) and group-predictive spectral features (OPLS-DA). For investigative studies, where the metabolic groups are unidentified or random, PCA is a good starting point to deliver an initial overview into the dataset arrangement and associations between groups. It provides unbiased information on trends on the occurrence of specific metabolites prior to the use of supervised methods (OPLS-DA). The results formed from PCA should then be used to conclude results which can then be confirmed using OPLS-DA and tested in more detail [110].

Studies have demonstrated the importance of PCA and OPLS-DA to analyze metabolomics data collected by NMR and MS [112,113]. However, without validation tests between PCA and OPLS-DA models, data can lead to inconclusive results about the underlying chemistry [114]. Cheng *et al* used multivariate analysis (MVA) for chemical dereplication to profile actinomycetes isolated from Mediterranean sponges,

which were a great source of novel biologically active compounds for drug discovery [113]. Conjoining both chemical dereplication and MVA could be an established method in NP drug discovery [115,116]. PCA and OPLS-DA identified chemically significant strains that could produce novel bioactive secondary metabolites [14]. PCA proved the occurrence of metabolites unique to the outlier strain by verifying the dereplication results. The PCA loadings plot indicated target metabolites that were validated by MS fragmentation and NMR spectral data. PCA plots organized over 1000 featured metabolites, while notable metabolites defined phenotypes of unique strains. MVA aided in the prioritization process to find new bioactive metabolites. The study by Cheng *et al* demonstrated the limitations of PCA, therefore, it has become a common practice to carry out OLPS-DA as a complementary method [113].

For NPs research, MVA is a highly effective tool in metabolomics profiling for discovering new drugs. LC-MS based metabolomics study done by Bose *et al* analyzed marine bacterial secondary metabolites and the production of rifamycin in *Salinispora arenicola* species. Rifamycin is a well-known antibiotic for the treatment of tuberculosis. PCA identified outliers and assessed groupings or trends. PCA and OPLS-DA analyzed the LCMS datasets of samples with various salt concentrations obtained at 7-day intervals to assess the optimum conditions for rifamycin production [117]. Earlier studies by HPLC-UV, which has become the basis of the latter investigation, monitored the production of rifamycin W, S and B as species-specific marker over a 43-day time period [118]. Albeit, the PCA scores plot showed groupings based on the different salt concentrations used. This was more clearly illustrated using the OPLS-DA plot. The use of MVA this study gave an insight on the optimum conditions for the production of the target antibiotics in *Salinispora* species [117].

2.4 Molecular networking (MNW)

Molecular networking (MNW) is a dereplication strategy, which aids the identification of new chemical entities against known compounds and thereby, reducing time and cost for further research [73]. By employing the MS/MS (MS^n) fragmentation data, MNW is also able to match molecular families with gene clusters by comparing their MS^n datasets with specific clusters found in the database network and eventually establish links with hundreds of other sequences. MNW utilized MS^n data based on the assumption that molecules belonging to the same structural family will have similar fragmentation patterns [73]. The MS^n spectral data are aligned and scored against other similar molecular families. The molecular network is put together using "cosine scores", which indicate the similarities of the MS^n data between metabolites. The processed data assembles a visual map indicating correlations between ion peaks in a chemical space as provided by the mass spectral data. The molecular network allowed a mapped comparison of molecular clusters with spectrally similar or identical structures. When a cluster is constructed, a node represents the link to similar MS^n spectrum, a "mathematical merging of spectra with nearly identical precursor mass and peak patterns" [73], and an edge (lines) joins the similar MS^n ion peaks. Therefore, a cluster with numerous nodes connected by several edges indicates the occurrence of related structures. Whereas, those with fewer nodes and connected by a single edge implicates that the spectral data is unlinked to a molecular family and may indeed be unique or a novel compound. The fragmentation data is dereplicated employing databases such as Metabolights [119] or the Global NP Social Molecular Networking (GNPS) web-platform containing thousands of both full MS scans and MS^n spectral data of various metabolites. GNPS has been the web platform database coupled to MNW for

dereplicating structural identities of a chemical family network [120]. The MNW plot is generated by MATLAB or Cytoscape software [121].

With increased DNA sequencing data access, MNW studies had exhibited a large number of undiscovered gene clusters still yet to be explored for biologically valuable properties. One such example was the isolation of a cytotoxic cyclic octapeptide from an American Samoan marine cyanobacterium as an encouraging resource for the identification of new anticancer treatments [122].

MNW and MVA strategies are complementary in the discovery of new drugs from NPs [73,101]. Like MVA, MNW is also a visualization tool but employs an organizational approach based on tandem MS^n data. While MNW can interpret several variables simultaneously, MVA can only assess two or three variables, as it is difficult to extrapolate a three-dimensional dataset expressed on a two-dimensional context, [73,101]. Both MVA and MNW could be coupled to a database to target metabolites for their structure novelty. A dereplication step prior to commencing an isolation work thrusts the early recognition of known molecules while reducing the cost and time invested in the detection of novel pharmacologically active composites. NMR, MS, and MS^n fragmentation data are the most widely used dereplication tools used to highlight compounds of particular interest. By identifying potential analogues of interest, both MVA and MNW are able to enable the user to identify unknown fragments and substructures, thereby decreasing the decision time to prioritize and plan the isolation protocol for novel bioactive NPs. Additionally, both MNW and MVA effectively avoid the bioassay steps used in bioactivity-guided fractionation and prevent a tedious isolation work on known compounds. The disadvantages of using MNW depended entirely on the limitations of mass spectrometry. Sometimes the results of the MS^n are varied from a narrow number of fragmented ions due to the diverse ionizing capability

of molecules in different types of instruments. Currently available MS^n databases have only been set-up for specific MS analyzers, which is not able to tackle the non-reproducible fragmentation pathways between the type of instruments employed for the analysis. There is still no program available that could accomplish data conversion for the different vendor-specific MS^n data. Small molecular weight compounds produce a small number of MS fragments, giving rise to similar MS^n spectra in spite of significant differences in structures. MS is unable to reveal the stereochemistry of particular molecules. Moreover, the time taken for a single MS analysis to be uploaded to a database network can take from 10 minutes to several hours depending on the class and structure complexity of the samples. This has a large impact on the quality of the results and the ability to gain global data sharing [123]. However, the use of a combined MNW and MVA approach appears advantageous. One study looked at the use of molecular networking in isolating secondary metabolite production in both Scottish and Antarctic locations [124]. The results revealed that with the use of PCA in combination with molecular networking, metabolomics-based dereplication were shown to be very effective tools for the identification of the metabolites. This also allowed other metabolite spectral patterns to be prioritized for further isolation, structural analysis, and bioassay work [124].

MNW has been utilized in many different areas of research. MS-based molecular networking has been applied in the dereplication of marine and terrestrial microbial samples. MNW retained similar analogues that were difficult to detect in other dereplication approaches and provided a context of information obtained from MVA. MNW illustrated the dereplication of 58 abyssomicin analogues [73]. Abyssomicin C (Figure 1) was a new potential antibiotic against methicillin-resistant

Staphylococcus aureus (MRSA). People with MRSA were estimated 64% more likely to die from an infection compared to people that have the non-resistant form [125].

Figure 1. Potential antibiotics against MRSA

The incorporation of MNW into NPs drug discovery required minimum changes in the workflow and complemented other dereplication methods as well as bioactivity fingerprints, like cytological profiling or BioMAP for phenotypic profiling. MNW is able to facilitate a range of ionization platforms allowing cross correlation of MS^n spectral data that could be obtained by direct infusion, ambient ionization and LC-based methods. However, since MNW depends on the existence of related fragmentation patterns of NPs, the number of possible fragments observed decreases with low intensity molecular ion peaks. Development and improvement of new dereplication algorithms is still essential for the study and elucidation of novel structures, which may stay elusive due to their absence in the database.

The application of MNW was also exemplified to gain knowledge on secondary metabolites of a relatively unknown species of *Peitigera* lichen [126]. By using MNW and MALDI-MS complementarily, the study proved that the microbes within the lichen community work together as a single unit to survive. Small molecules were used as a defense mechanism through the production of fungal pyridine alkaloid PF1140 (Figure 1), which exhibited both antifungal and antibacterial properties. A limitation of MNW in this particular study was not all molecules at this point could be annotated. Like OPLS-DA, MNW can incorporate noise into the signaling dataset, however in both cases this could be overcome [111,126].

Studies using MNW have facilitated the increase in production of target metabolites and thereby increasing their antibiotic activity. One study involved the marine-derived *Streptomyces* species strain PTY08712 co-cultured with human pathogens such as methicillin-sensitive *Staphylococcus aureus* (MSSA), *Bacillus subtilis* and MRSA [127]. By using MNW, it was proven that there was an increased production of three antibiotics granatomycin C, granatomycin D, and dihydrogranaticin B (Figure 2), which became more evident when PTY08712 was co-cultured with MRSA.

Figure 2. Antibiotic compounds obtained from marine-derived *Streptomyces* species strain PTY08712

3 Application of metabolomics in NPs research

The application of metabolomics is being applied throughout NPs Research, especially in the investigations of new antibacterial and anticancer drugs. Several NPs have entered clinical trial between 1981 and 2010, 34% of approved small molecules were NPs or direct derivatives thereof [128] while 74% of the currently available antibiotics and 59% of new anticancer chemical entities originated or were derived from NPs [129,130]. Despite huge numbers of NPs being approved by the FDA, the pharmaceutical industry has reduced their efforts in utilizing NPs as a source for drug discovery. This can be attributed to the high cost of research and high incidence of rediscovery of known compound during the late stages of the research. Metabolomics affords the opportunity for a plethora of advancements, particularly in the field of drug discovery from NPs. Recent improvements reveal the indisputable value of the analytical techniques used within metabolomics in the discovery of NPs, gene-function

analysis and diagnostic platforms [131]. However, metabolomics could be plagued with experimental difficulty due to an incomplete catalogue or library of existing metabolomes available. Consequently, it is common to find unknown metabolites, thereby, making the interpretation and analysis of the data difficult, albeit a motivating prospect for novelty. Due to the vast range in both physical and chemical properties of the studied metabolites, simultaneous quantification and identification of compounds can be impossible to achieve using a single analytical technique and methodology.

In order to bring a potential NP-based drug to the global market, it must be presented into the drug discovery and development pipeline. This is a multi-step process which includes target validation, lead optimization, drug formulation and clinical trials. Drug development takes on average 10 to 15 years and requires between 800 million and 1.8 billion dollars to bring a new drug to market [132]. Of 5000 possible new drug candidates, on average only five would be tested in humans and only one would be approved for therapeutic use. Therefore, the need to utilize different analytical techniques and combinations thereof, is crucial in the discovery of novel drugs. The ability to identify flaws in targeting the new bioactive molecules earlier in the workflow helps reduce financial loss caused by development failure at a later stage [133]. Metabolomics is effective as an aid to drug discovery, as it has a universal application with large amounts of information being generated from the results and datasets. A metabolomics based approach has also been successful in confirming a desirable *in vivo* mechanism of action for high throughput screening chemical leads, and validating the maintenance of mechanism of action as the binding affinity is iteratively increased. Metabolomics can verify a chemical lead selective to an *in vivo* inhibitor along with its possible side effects. Metabolomics has become increasingly important

within the drug development process with its aforementioned strengths as an analytical tool [132].

3.1 Search for new potential anticancer drugs

Within the field of NPs research, metabolomics has made an impact in drug discovery, particularly in the study of cancer chemotherapeutic agents. Due to the difficulties of getting access to the desired pharmacological potency and selectivity for new compounds, no new approved drugs have been produced and released onto the market in recent years. Few studies originating from NPs have been approved by the EMA and the FDA. However, the growing interest in metabolomics allowed for the discovery of new potential drugs, in particular in the treatment of cancer. Metabolomics has aimed to tackle this problem by making the process of discovering new active compounds easier. A study performed by the Université d'Auvergne looked at the biochemical disorders induced by the use of cytotoxic MNPs in the against MCF7 breast cancer cell lines. The study used high resolution magic angle spinning (HRMAS) proton nuclear magnetic resonance (NMR) spectroscopy-based metabolomics. The three antineoplastic MNPs used in the study were ascididemin (Asc), lamellarin-D (Lam-D) and kahalalide F (KF) [134] (Figure 3). The study included the identification of the response of MCF7 cells to Asc which involved the obstruction of enzymes in citrate metabolism, mitochondrial-cytosolic carrier disorders in response to Lam-D, and enhanced lipid membrane breakdown as a result to exposure to KF [134]. The aim of the study by Bayet-Robert *et al* was to identify metabolic targets and the cytotoxic properties of candidate MNPs using NMR metabolomics.

Figure 3. Marine-derived anticancer drugs currently in clinical trials.

The metabolomics profiling workflow managed a batch of 1500 samples in replicates for NMR spectroscopy analysis. Partial least squares-discriminant analysis (PLS-DA) was used to help compare the different NMR spectrum signals from the treated and control groups. This included enzyme blockade in citrate metabolism to Asc, mitochondrial-cytosolic carrier disorders with regards to Lam-D, and an increased lipid membrane catabolism in reaction to KF. These metabolomics study provided clear evidence to the response of MNPs to breast cancer cells. The metabolome information in the involved pathways has aided in the design of potential new medicines from MNPs prior to their clinical trials in breast cancer.

3.2 Search for new potential antibiotics

Between 1942 and 1972, antibiotics increased life expectancy by eight years [132]. The effectiveness of antibiotic therapy is threatened by the emergence of antimicrobial resistance. Diseases more common in 20th century, including pneumonia and tuberculosis, are causing prolonged illnesses, increased health care cost and increasing mortality [135]. Despite the clear need for new antibiotics, this has been neglected by the pharmaceutical industry in order to concentrate their efforts on other new "blockbuster" drugs. Antibiotics often fail in the drug development process providing less financial returns, also contributing to the lack of input from the pharmaceutical industry [132]. Most antibiotics in the clinical pipeline are modifications of known antibiotics, creating a temporary solution to a more permanent problem. Only 51 antibiotics have been further developed since the discovery of penicillin [125]. The advent of antimicrobial resistance has become a huge burden on healthcare worldwide. We know that stagnation in the discovery of new antibiotics has occurred, as evidenced in Figure 4 [136]. In fact, a gap of 32 years occurred with the introduction of no new

antimicrobial agents to clinical practice (1968-2000). Based on this evidence, there is a pressing need to utilize all means of drug discovery, including NPs and metabolomics to address this global crisis.

Figure 4. Timeline of the introduction of new antibiotic classes from 1935 – 2003 [136].

The metabolomics approach is effective in secondary metabolite profiling as it is fast, reproducible and requires a relatively simple sample preparation step. In terms of feasibility, the process undertaken to identify a producer of a potentially novel antibiotic could be vast [137]. However, the process would still be cost-feasible due to the market size for a drug that may show a broad-spectrum of activity against multidrug-resistant bacterial pathogens.

3.3 Marine microbes as source of new potential drugs

3.3.1 Marine fungi

NPs have been investigated to aid the discovery of new drugs, including tubulin-binding anticancer medications, immunosuppressants, antibiotics, and antiparasitic treatments. NPs research has also revealed new classes of antifungals, such as the echinocandins (Figure 5), a polypeptide-antibiotic from *Aspergillus nidulans var. echinulatus* [138]. Research in the field of marine-derived fungi aided the discovery of several metabolites. Metabolic fingerprints of a collection of French Atlantic marine-derived fungal extracts afforded the identification of new bioactive halogenated compounds [139]. An untargeted time-scaled metabolomics-based approach was also used on strains of marine-derived *Penicillium* to study biosynthetic pathway regulation and their effect on metabolome variation [140]. Integrated metabolomics and imaging techniques has been also used to map the epiphytic fungi on seaweeds *Fucus*

vesiculosus [141] while metabolomics tools were also utilized to study the anti-quorum sensing activity of metabolomes from fungi derived from Baltic seagrass *Zostera marina* [142]. MNW-based metabolomics has been used to analyze the bioactivity of marine adapted fungi co-cultivated with phytopathogens to exploit the secondary metabolites for agrochemical applications [143]. Metabolomics has aided the discovery of many new metabolites including the cytotoxic halimide from the marine-derived fungus *Aspergillus ustus*, which was a lead compound in the development of plinablin that entered phase II clinical trials [144], and the antimicrotubule agent KPU-300 [145] (Figure 6).

Figure 5. Some bioactive compounds from marine-derived microbes

Figure 6. Plinablin and KPU-300 from the marine-derived fungal metabolite halimide

NPs discovery from marine-derived fungi has not been fully utilized since the discovery of cephalosporin from the marine fungus *Acremonium* in the 1940s [144,146]. Studies on marine-derived fungi were rare until the 1990s, and only started to grow at end of this decade [146]. The objectives of discovering new chemistry from marine sources has neglected the exploration of obligate marine fungi for drug discovery when compared to marine-derived fungi [144]. A range of different analytical techniques has been recently being employed to isolate and discriminate between fungi actively growing in the substratum and spurious propagule to determine if a fungus is facultative of marine origin. The long standing definition of marine fungi by Kohlmeyer has been continuously challenged, therefore fungi isolated from marine sediments have largely been ignored or labelled as terrestrial [147]. Metabolomics tools have also been

exploited to scrutinize evidential hypothesis and such included measuring the concentration of ergosterol, a secondary metabolite unique to fungal cell membranes, which can be used as a biomarker for fungal growth [148,149]. Another study that employed the metabolomics approach used LC-MS to quantify the concentration of 20-residue peptaibols longibrachins and trichokonins in *Trichoderma* strains to determine their active growth [150,151]. The review by Overy et al intensively analyzed their stated objectives and used 150 references but neglected several other groups of fungi such as specific yeast, chytrids, and basal group fungi thereby narrowing the reader's overview of marine fungi research as a whole. The article tried to address several hypotheses that were frequently left unanswered due to the lack of comparative data within this field [144].

3.3.2 Marine bacteria

It is not just fungi that are paving the road to drug discovery from NPs; bacteria such as actinobacteria harbors biosynthetic properties to produce antimicrobial secondary metabolites for new antibiotics [137]. There is an increasing need for the discovery of new classes of antibiotics as well as utilizing innovative analytical platforms such as metabolomics to achieve this goal. Around 70% of NPs currently in clinical use are isolated from actinobacteria; this phylum represents a potential source for a plethora of NPs with pharmaceutical applications [48]. Phylum *Actinobacteria* is one of the largest within the bacteria domain. It encompasses 6 orders and 14 suborders [152], and 408 species included in the List of Prokaryotic Names with Standing in Nomenclature [153]. The sampling of marine sediment for new bacterial sources of novel NPs has led to the discovery of *Salinispora*, a new genus of *Actinomycetes* [154]. Utilization of both genomics and a chemical profiling workflow resulted in the structural elucidation of

salinilactam A in *Salinispora tropica* [155] (Figure 5).

Streptomycetes species are known source of NPs [152]. Many antimicrobial agents have been described to be biosynthesized by *Streptomyces* [156,157]. Since 2000, 22 new potential antibiotics have been reported from *Streptomyces*, while five represented new compound classes and three of these new classes are NPs derived [158]. Three of these new classes originate from NPs: the lipopeptide daptomycin (Figure 5), the pleuromutilin retapamulin™, and the tiacumicin fidaxomicin™ [128]. In fact, the genus *Streptomyces* has been the main source organism for the semi-synthesis of antimicrobial products utilized by the pharmaceutical industry [159]. Due to the rapid evolution of antimicrobial resistance in pathogenic microorganisms, the discovery of new antimicrobial drugs is vitally important [160]. Historically, most research into the secondary metabolites of *Streptomycetes* has concentrated on terrestrial bacteria but the last decade has seen the focus shift to the untapped ecological niches of the marine environment [160,161]. Sponges, being sessile by nature, rely on chemical means of defense and provide a habitat for a plethora of microbial life. It has been reported that 37% of described MNPs were isolated from sponges. It has, however, been found that a number of earlier reported NPs attributed to sponges are in fact biosynthesized by their endosymbionts [162], many of which are Actinomycetes. Through the utilization of metabolomics along with other omics tools, marine Actinomycetes have been explored and profiled [163-167] while many new potential drugs [112,113,117,160,162,168-174] and agrochemicals [175] have been recently described. For the purpose of this literature appraisal, three research papers were selected and analyzed to assess the application of metabolomics to current drug discovery programs, with a particular focus on novel antibiotics isolated from marine *Streptomyces*. These papers present three different combinations of bioassay and omics-guided approaches to drug discovery.

3.4. Isolation strategies for biologically active natural products

3.4.1 Bioassay- and phylogenetic-guided isolation

From British Columbian waters, 186 *Streptomyces* isolates were obtained from 49 sediment samples collected from various depths (20-200 mm) and locations over a 2-year period as well as inoculated in different media [176]. The isolates were tolerant of salt concentrations of up to 3.5% but did not require salt to proliferate, indicating the strains are not obligate marine species but could have been terrestrial species which have adapted to the marine environment by being washed down as dormant spores. The collected isolates of *Streptomyces* were screened for activity against six microorganisms (methicillin-resistant *Staphylococcus aureus* strain ATCC 33591, *Bacillus subtilis* strain H344, *Escherichia coli* strain UBC 8161, *Pseudomonas aeruginosa* strain ATCC 27853, *Mycobacterium fortuitum* strain ATCC 6842 and *Candida albicans* strain ATCC 90028) and 47 isolates showed bioactivity. Bio-assay guided selection and phylogenetic study of the isolated strains facilitated further isolation work on prioritized isolates but did not reveal detailed information on the biosynthetic potential of these strains to produce unique NPs. Phylogenetic analyses of the isolates allowed rapid comparison and identification of four most bioactive isolates with unique phenotypic characteristics, which were selected for further chemical analyses and drug discovery work. A bioassay-guided isolation and preparative HPLC workflow successfully afforded two known (novobiocin and its desmethyldescarbamoyl congener) and four new anti-MRSA novobiocin analogues: desmethylnovobiocin, 5-hydroxynovobiocin, desmethyldescarbamoyl-5-hydroxynovobiocin, and desmethyl-5-hydroxynovobiocin [176] (Figure 7). The isolated compounds were elucidated by high resolution ESIQIT-MS (Electrospray Ionization Quadrupole Ion Trap-Mass Spectrometry) and NMR while the known metabolites were dereplicated from Antibase, Sci-finder and MarinLit.

Novobiocin was shown to target bacterial gyrase by inhibiting ATP hydrolysis, and interacts with heat shock protein Hsp90, destabilizing Hsp90 chaperone proteins. It showed strong activity against Gram positive bacteria and MRSA in combination with rifampicin, but was withdrawn from the market after its licensing in the 1960's due to a poor toxicity profile, aqueous insolubility, and poor activity against multi-resistant Gram negative infection [177].

Figure 7. Anti-MRSA active novobiocins

The dereplication study was only performed to a small subset of the *Streptomyces* isolates at the end of the study allowing only a snapshot of a portion of the genus. It is also important to account for the impact of horizontal gene transfer between microorganisms. It is not uncommon for two microbes with 100% rRNA similarity to synthesize different secondary metabolites, likewise it is known that microbes with very different sequences can synthesize the same secondary metabolites [131]. Although phylogenetics is a useful tool in organism identification and classification, the application of metabolomics tools resolves these issues and allows differentiation between microbes with a high level of genetic similarity, showing the complementary nature of the two approaches. Although this study efficiently identified and isolated four novel novobiocin compounds, the manner in which the strains were selected was a simple process of elimination based on the activity of the strains sampled against a small subset of microorganisms. The ability to predict the secondary metabolites that will be synthesized by a microbe prior to large scale culture and bioassays may prove invaluable in its ability to expedite the process of NP discovery.

3.4.2 Metabolomics and genomics: the perfect pairing

Utilizing a genomics platform with a metabolomics approach not only revealed the biosynthetic gene clusters coding for unidentified NPs but as well as validated their production by the respective strains [160]. This combinatory approach was further developed by inactivation of a biosynthetic gene cluster for a known antimicrobial agent, allowing the identification of a NP which had previously been masked by the effects of the inactivated gene cluster [162].

The genome of a strain of *Streptomyces* isolated from marine sediment from the Trondheim Fjord in Norway has been sequenced and established to contain coding for multiple of both known and undescribed NPs [160]. The initial genomic analysis identified the strains' biosynthetic potential to be comparable to other species within the genus, having similar average number of biosynthetic gene clusters. The genomic information predicted the production of a vast array of secondary metabolites in the strain. The metabolomics tools utilized in this study perfectly complemented the genomics-based approach that facilitated the isolation and identification of several of the predicted secondary metabolites from the strain. The sediment sample was treated with extremely high frequency radiation (EHF), which has been shown to promote the growth of a number of rare actinomycetes. The antiSMASH search tool revealed 36 potential secondary metabolite biosynthetic gene clusters, accounting for 8.4% of the chromosome, generally occurring in regions with low G&C content. Cluster 2 was noted as similar to the cluster coding for coelibactin, a putative peptide implicated in antibiotic regulation in *Streptomyces coelicolor* [178]. Five gene clusters contained genes encoding for terpene biosynthesis, while another (cluster 23) encoded for a type I polyketide synthase and a terpene synthase/cyclase. A number of unconfirmed production of other secondary metabolites were also detected, one of which was

presumed to govern the biosynthesis of a pyrrolopyrimidine nucleoside antibiotic. The antimicrobial activities of the extracts were assessed against *Bacillus subtilis* ATCC 6633 and *Pseudomonas putida* KT 2440 using a disc diffusion assay. They were found to be active against *B. subtilis* but not *P. putida*. This activity was then attributed to the production of bisindole pyrrole antibiotics that included spiroindimicins B and C and lynamicins A-E (Figure 8). A compound was identified with the expected absorption spectra of a bisindole pyrrole, but with a mass not corresponding to any of the known derivatives available from the database. This compound was found to be a mixture of two previously undescribed compounds, which were then isolated and structurally elucidated using NMR as spiroindimicins E and F. The mass disparity of 35Da from spiroindimicin B was attributed to the absence of a chlorine atom.

Figure 8. Bisindole pyrrole antibiotics

Both genomics and metabolomics were utilized to provide information on both the phenotype and genotype of a *Streptomyces* species (SM8) isolated from the marine sponge, *Haliclona simulans*, off the coast of Ireland [162]. Partial 16s rRNA sequencing showed 100% similarity between this strain and *Streptomyces violascens* strain XSD-115, other strains of which have been reported to biosynthesize antibiotics including actinomycin X_2 and actinomycin D [179], as well as a number of other *Streptomyces* species. Extracts were screened for antimicrobial activity against *Saccharomyces cerevisidae*, *Kluyveromyces marxianus*, *Aspergillus fumigatus*, *Candida albicans*, *Candida glabrita*, *Bacillus subtilis*, *Escherchia coli*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*. SM8 was found to have activity against all but *S. aureus* and *E. coli* while greatest activity was observed against *C. albicans* and *B. subtilis*. The

initial screening of metabolites produced by SM8 was performed using metabolomics tools. NMR and liquid chromatography-high resolution mass spectroscopy (LC-HRMS) facilitated swift identification of the classes of compounds present in the metabolome. 2D NMR revealed information on the major components of SM8, while LC-HRMS, which is more sensitive, detected metabolites in at ng to μg level concentrations depending on the ability of the compounds to ionize in the respective modes. A search on the Dictionary of NPs database provisionally identified metabolites already reported as *Streptomyces* metabolites. Metabolomic techniques were also applied to compare extracts from SM8 to those of *H. simulans* to establish the presence of similar secondary metabolites of interest between the host and the symbiont. Polyhydroxylated fatty acids were dereplicated by MVA to be bioactive against fungi and gram-negative bacteria. The majority of bioactive compounds identified by MVA were members of the antimycin family (Figure 9). SM8 underwent genomic analysis resulting in a draft genome sequence. Biosynthetic gene clusters for antimycin and candicidin (Figure 9) were identified and confirmed by comparison to known polyene PKS clusters, and the published gene cluster from *Streptomyces albus* sp. S4, respectively. This gene cluster was inactivated in the mutant strain by deletion of *ant*C, the integral gene for the biosynthetic pathway of antimycin, to establish its contribution to the antimicrobial activity of SM8. A 32-fold decrease in antimicrobial activity was observed for the mutant strain compared to the wild type. LC-MSn analysis of extracts on the Orbitrap revealed the presence of antimycins A1-a in the wild type but no traces of any antimycin congener in the mutant. Further compounds were isolated from a large scale culture of SM8 in oatmeal media. Three butenolides were found in the antifungal fractions, which may also have a regulatory role in antibiotic production [180] as earlier described for avenolide (Figure 9), a butenolide biosynthesized by *Streptomyces*

avermitilis [181].

Figure 9. Examples of antibiotic compounds from sponge-derived *Streptomyces*.

The combined approach employing both genomics and metabolomics approach allowed rapid identification of antimicrobial compounds in extracts of *Streptomyces* species SM8 [162]. Metabolomic tools complemented the application of genomics through identification of similar gene clusters for known antimicrobial agents. Inactivation of the antimycin coding gene cluster quantified the contribution of the respective secondary metabolites to the bioactivity of the extract and in parallel identified the occurrence of the low-yielding antifungal butenolides, the bioactivity of which may have been masked by the activity of the antimycins. Furthermore, combining genomics with metabolomics served to effectively "shortlist" the compounds responsible for a specific bioactivity, which warranted the importance of the strain for scale-up.

Metabolomics has developed rapidly in terms of its application in natural products research. Metabolomics afforded a "snap shot" of the metabolome and the biosynthetic potential of an organism while genomics complementarily expands on this. High quality data can be obtained and interpreted rapidly utilizing genome and metabolome databases such as antiSMASH 3.0, which however, like any databases, will have its limitation for dereplicating yet undescribed biosynthetic pathways in the occurrence of secondary metabolites with novel chemistry. The ability to employ data from one analytical techniques to interrogate and validate another investigative approach serves to expand our understanding of a cellular process through the detected metabolites as their end products. The ability of a metabolomics approach to model the

metabolic phenotype of an organism is of great use in the discovery of novel pharmaceutical agents. Its integration with other omics-based approaches permits a more accurate model of a complex biological system and could more completely depict the biosynthetic potential of an organism. Genomic screening of biosynthetic gene clusters of marine microorganisms that are difficult to mass cultivate prior to metabolomics analysis would allow prediction of the occurrence of novel bioactive secondary metabolites and eliminate strains with low biosynthetic potential from further scale-up to save time and resources being wasted.

4 Conclusion

The fundamental concept of metabolite profiling that has been around for thousands of years led to the founding of metabolomics, but its potential as an analytical technique had only recently began to be acknowledged within the field of drug discovery [35]. Each analytical technique has its limitations, however; it is common for the most sensitive methods like mass spectrometry to be used as a hyphenated technique to overcome such issues [72]. The use of NPs for the discovery of new potential antibiotics and anticancer drugs has become increasingly evident in recent years due to the discovery of several compounds with novel mechanisms of action and potent bioactivity. However, to date metabolomics is still within its infancy as a means of targeting NPs for drug discovery and is still often plagued with experimental difficulties due to the handling of large datasets and the incomplete nature of metabolomic databases. MVA such as PCA and OPLS-DA methods provide an essential means of quick interpretation of complex dataset, therefore, creating significant implications behind the metabolomics data. However, misunderstanding such methods can result in misleading or false biological conclusions [182]. MNW is also an effective visualization

and dereplication tool, allowing fast identification of known analogues and has allowed simultaneous interpretation of several variables, although it does not directly provide a chemical structure for respective analogues [73,101]. MVA and molecular networking are complementary methods and aid the discovery of new antibiotics [73,117]. Metabolomics has been shown to be an effective means of discovering novel antibiotics from NPs. Nevertheless, metabolomics has yet to achieve a firm position in the drug development process, in part due to the lack of interest of the pharmaceutical industry in antibiotics discovery [132]. Bioassay-guided isolation of bioactive metabolites is often perceived as an inefficient conventional method to the more targeted metabolomics approach.

The application of metabolomics to drug discovery can be biased for older methodologies. For instance, because of the past success of the "golden era" of antibiotic discovery. Unfortunately, the effectivity of antibiotics and anticancer drugs had dwindled due to the ever-emerging multi-drug resistance. Bacterial infections are not a third world or 20th century issue and require an urgent response. However, this will not aid the imminent crisis we are facing, requiring the discovery of new and innovative drugs as opposed to applying structural changes to current drugs. NPs may be the means of achieving this goal. Metabolomics has shown great promise as an efficient drug discovery protocol; however, it must be allowed to have a more prominent position in this process to prove its effectiveness.

Metabolomics, and other "-omics" approaches, afford a holistic view in systems biology. As metabolomics progresses as a scientific field in its own right, more and more applications are coming to light within the drug discovery pipeline. The application of untargeted metabolomics has been utilized to determine mechanisms of action of antibiotics [183] and potential anticancer drugs [134] by targeting metabolic

pathways. A metabolomics based approach has also found its relevance in probing the evolution of multi-drug resistance [184]. As presented in this review, metabolomics has been effective in the study and elucidation of disease models and phenotypes. Further study into the metabolic impact of a diseased state may prove vital in target validation processes. The efficient nature of a metabolomics-based approach, along with its high information output could steer towards a plethora of potential candidates for drug development and expedite the drug discovery process.

Obtaining information from a metabolomics data can be achieved by a range of approaches, tools and methods. The main issue regarding data mining is that extraction from datasets using various methods may alter data points, resulting in a different statistical model. For different approaches, there could be large changes in the occurrence of respective sets of the metabolome, necessitating a validation step prior to conclusive interpretation. Data extraction errors have a massive impact on the outcome of investigations. Such errors in the process increase the financial and temporal cost of a drug discovery program [185].

Metabolomics tools have been applied, not only in targeted isolation work but, in studying drug toxicology and disease pathology, which are employed in drug discovery. The pharmaceutical industry has been pushing to broaden its application by bringing new medicines to the market in a quicker and more cost-effective way. Pharmaceutical companies aim to discover drugs which are efficacious and successful but most importantly are going to save them money when the drug enters the clinical trial phases. Currently, drug discovery and development also consists of gene analysis and sequencing to identify genes correlated to a disease [186]. When a potential drug lead is discovered, it has to pass through the R&D pipeline to reach final development. Some drug leads successfully complete the final development pipeline only to fail in the

concluding clinical trial due to unwanted toxicity. Metabolomics aims to bridge this problem by offering a more cost-effective and practical way to discover potential new drugs through the study of metabolic pathways. Target analysis quantitatively evaluates the regulation of particular metabolites in a specific pathway [109]. Additionally, it holds many advantages by considering the affected chemical changes at the cellular basis and molecular level over the more conventional ways of drug discovery of utilizing only the pharmacological data. There are numerous analytical techniques and technologies employed in metabolomics which have aided in the breakthrough of discovering new or existing drugs and NPs with new mechanism of action [134,187-193]. A metabolomics based approach could amplify the presence of explicit metabolites allowing targeting of pathways specific to the disease under probe [109]. Information on the target pathway could lessen toxicity failures occurring during late clinical trials while more time and money could be spent wisely on creating a drug that was specific to the disease. Likewise, by using metabolic fingerprinting, large numbers of intracellular metabolites and new targets could be scanned that would otherwise have been unidentified. Metabolome information could afford an overall picture of physiological processes when a certain disease is active. All cellular processes, including anabolism and catabolism, are accounted for which would allow for an informative image to illustrate the particular metabolites and pathways targeted. Another major advantage of metabolomics in drug discovery is the vast number of metabolic samples and replicates that could be handled from many different sources or batches.

There could be more than one preferred method used for metabolite dereplication as well. The significant amount of data generated and to be interpreted has been a challenge. Within the process, there are numerous analytical techniques

accessible and often the choice relies on the main objective of the study. The effectivity to quantify and identify the metabolites with adequate sensitivity and precision is highly dependent on the choice of analytical method and instrumentation. High-end instrumentation cost is the disadvantage of a metabolomics approach. The high equipment and maintenance costs are a problem and often require specialized technical skills to operate. However, metabolomics had demonstrated an increased efficiency of NPs research for drug discovery. With important advances in analytical technologies, and target validation, the metabolomics approach can identify and quantify larger amounts of metabolites that can produce faster, more sensitive and reliable data. Unrelenting advancements in metabolomics technologies will undoubtedly achieve greater results both in NPs research but also in the drug discovery.

5 Expert Opinion

Metabolomics has become an indispensable analytical tool to aid the drug discovery process although it remains to be utilised effectively by NP drug discovery programs. Due to high instrumentation cost of metabolomics, it is still perceived as a relatively new field of research being effectively only used as an analytical approach since the 1970s [35,80]. The discovery of penicillin by Sir Alexander Fleming in 1928 ushered in a "golden era" of antibiotic discovery between the years of 1942 and 1972. Most of these were discovered, not with the aid of metabolomics [194] but by serendipity and bioassay-guided isolation work. In bioassay-guided fractionation, crude organic extracts were then subjected to repeated chromatographic isolation with each purification step being supported by a bioassay to follow the bioactivity of the target metabolite until its purest form is isolated [195]. The structural composition of the target metabolite was then determined with high-resolution MS and NMR. However, if no dereplication study

was done prior to the isolation work, the isolated purified compound could be a known metabolite. In the presence of low yielding potent secondary metabolites, the isolated compound could also be either the active or an inactive constituent particularly when monitoring was only followed by single detection by UV [116]. Nonetheless, bioassay-guided fractionation has brought about the discovery of significantly vital pharmaceuticals, such as taxol (paclitaxel), camptothecin, and vinblastine [196] (Figure 10).

Figure 10. Plant-derived anticancer drugs isolated bioassay-guided fractionation

The marine-derived drug called trabectedin® (yondelis®), also known as ecteinascidin-743 (ET-743) was developed by PharmaMar by semi-synthesis from safracin B (Figure 11), an antibiotic obtained by fermentation of the bacterium *Pseudomonas fluorescens* [197]. Trabectedin® was granted orphan designation (EU/3/01/039) by the European Commission on 30 May 2001 for the treatment of soft tissue sarcoma [198,199]. It was first isolated from the marine tunicate *Ecteinascidia turbinata* which belonged to a family of a Caribbean Sea squirt. The initial accounts of its antitumor effects on P388 murine leukemia cells were reported in 1969 against various cancer cell lines. It was only in 1990 that Kenneth L. Rinehart isolated and elucidated six further compounds called ecteinascidins (ET-729, ET-729A, ET-743, ET-745, ET-759B AND ET-770) [200]. Through bioassay-guided fractionation, Rinehart was able to discern the most abundant constituent ET-743 and was later licensed to PharmaMar for further research and development. ET-743, now marketed as yondelis®, was the first anticancer marine-derived drug that has been authorized by the European Commission for the treatment of patients with advanced soft tissue sarcoma in 2007, then approved by the EMA in 2008

and later by the FDA in 2015 for the treatment of advanced or metastatic soft tissue sarcoma. Bioassay-guided isolation in drug discovery did not require any expensive high-resolution equipment in the initial stages of the isolation work or any complex statistical algorithm to process generated spectral data. However, it took 20 years to isolate and elucidate the compound after the first positive results on anticancer activity was reported. The question still remains as to whether this 20-year time-frame could have been decreased with the use of a metabolomics approach?

Figure 11. Structures of safacin B and ecteinascidin 743 also known as yondelis® or trabectedin®

Bioassays determine the potency of NPs by assessing and comparing the level of response of test samples with known standards under specific conditions either *in vivo* or *in vitro*. Bioassays can be biochemical or cell-based. A biochemical assays examines a particular bioactivity without any interference from other cellular processes. Cell-based assays are dependent on the ability of the sample constituents to either passively or actively pass through a permeable cell membrane while any interactions with other cellular constituents were taken into account [201]. Bioassay results anticipated the discovery of bioactive metabolites for specific pharmacological applications. Clinical trials are largely based on bioassays of drugs compared to their clinical effects given at a range of different doses [202]. By taking advantage of new bioassays along with new chromatographic separation and elucidation methods (NMR and MS), bioassay-guided isolation has been an effective tool in drug discovery programs allowing greater access on chemical diversity of the NPs. Bioassay-guided fractionation workflow tolerated handling large quantities of an entire series of structurally derived compounds to

establish structural activity relationships [203]. A major issue regarding the use of bioassay-guided isolation is when testing a sample to determine its minimum inhibition concentration (MIC) or IC_{50} or EC_{50}. Quantifying the bioactivity is a process that is long and often requiring over three weeks to validate the results. This may add problems for a traditional bioassay-guided fractionation work program as it is common for such a project to be ran for only two to three months during an intensive screening campaign. The purification of identified compounds of interest may not be possible within this period. Therefore, bioassay-guided isolation is often regarded as too slow for a two-year funded project [203].

Despite the fact that many drugs have been discovered through bioassay-guided fractionation, there are several disadvantages to the process. The procedure usually favors the purification of more dominant peaks in a fraction detected by a particular spectroscopic technique, often resulting to missing the low yielding potent metabolites. A metabolomics approach has the ability to pinpoint the bioactive metabolite even at µg levels in the extract or in a fraction [116]. Additionally, during fractionation, as the volume of fractions decreases, this increases the risk of not isolating the bioactive component(s) as it becomes more difficult to work with mg levels of fractions [196]. In summary, there is the need to utilize both bioassay-guided isolation techniques as well as employing a targeted metabolomics approach. The important goal of identifying bioactive compounds without multiple bioassay-guided isolation steps is still very apparent. By doing this, the effectiveness and revenue of a NP discovery program can be improved and more drugs could be brought to the R&D pipeline [196]. By using a targeted metabolomics approach, multivariate statistical modelling coupled with bioassay tests can efficiently reveal the bioactive compounds from a complex chemical profile and accelerate the process of drug development.

Compared to a metabolomics-based approach, bioassay-guided isolation of the bioactive metabolites involves more assay steps, requiring higher gram levels of fraction material to end up with sufficient yield of purified compounds that could be subjected for structural elucidation work and toxicity assays. With the metabolomics approach, the target bioactive metabolite is already pinpointed during the bioassay on the first fractionation stage rendering subsequent bioassays unnecessary [112-116,137,162,204-206]. Due to the ability of a metabolomics approach to categorize and rule out flawed drug candidates earlier in the workflow, this reduces the time taken to identify new natural products with the potential to be submitted to the R&D pipeline [133]. A metabolomics-guided isolation process requires less steps, as only the fractions containing the active metabolites would move on to further isolation work, reducing expenditure on solvents when compared to repeated untargeted chromatographic separation, which also requires the testing of larger numbers of samples. However, inactive fractions could still be revisited for structure-activity relationship studies.

Funding:

This manuscript was not funded.

Declaration of Interest:

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer Disclosures:

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

1. Blunt JW, Copp BR, Keyzers RA, et al. Marine natural products. Nat Prod Rep. 2014; 31(2): 160-258.

2. Martins A, Vieira H, Gaspar H, et al. Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. Mar Drugs. 2014; 12(2): 1066-1101.

   *Outlines the paths of marine natural products discovery and development focusing on compounds that successfully reached the market and presenting the approaches tackled by the pharmaceutical and cosmetic companies.

3. Montaser R, Luesch H. Marine natural products: a new wave of drugs? Future Med Chem. 2011; 3(12): 1475-1489.

4. Shen W, Kim JS, Kish PE, et al. Design and synthesis of vidarabine prodrugs as antiviral agents. Bioorg Med Chem Lett. 2009; 19(3): 792-796.

5. Schmidtko A, Lötsch J, Freynhagen R, et al. Ziconotide for treatment of severe chronic pain. Lancet. 2010; 375(9725): 1569-1577.

6. D'Incalci M, Galmarini CM. A review of trabectedin (ET-743): a unique mechanism of action. Mol Cancer Ther. 2010; 9(8): 2157-2163.

7. Malve H. Exploring the ocean for new drug developments: Marine pharmacology. J Pharm Bioallied Sci. 2016; 8(2): 83-91.

8. Khalifa SAM, Elias N, Farag MA, et al. Marine natural products: A source of novel anticancer drugs. Mar Drugs. 2019; 17(9): 491-522.

   **Highlights the impact of marine organisms on the search for new anti-cancer drugs.

9. Carroll AR, Copp BR, Davis RA, et al. Marine natural products. Nat Prod Rep. 2019; 36(1): 122-173.

10. Demain AL, Fang A. The natural functions of secondary metabolites. Adv Biochem Eng Biotechnol. 2000; 69: 1-39.

11. 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(1): 85-98.

12. Rinehart KL. Secondary metabolites from marine organisms. Ciba Found Symp. 1992; 171: 236-49; discussion 249-54.

13. Nunnery JK, Mevers E, Gerwick WH. Biologically active secondary metabolites from marine cyanobacteria. Curr Opin Biotechnol. 2010; 21(6): 787-793.

14. Macintyre L, Zhang T, Viegelmann C, et al. Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. Mar Drugs. 2014; 12(6): 3416-3448.

*Highlights a strategic workflow to prioritise extracts for further isolation work.

15. Sun W, Wu W, Liu X, et al. Bioactive compounds isolated from marine-derived microbes in China: 2009-2018. Mar Drugs. 2019; 17(6): 339-365.

16. Zhang H, Zhao Z, Wang H. Cytotoxic natural products from marine sponge-derived microorganisms. Mar Drugs. 2017; 15(3): 68-81.

*Comprehensive review on sponge-associated microorganisms.

17. Gemperlein K, Zaburannyi N, Garcia R, et al. Metabolic and biosynthetic diversity in marine Myxobacteria. Mar Drugs. 2018; 16(9): 314-330.

18. Watters DJ. Ascidian toxins with potential for drug development. Mar Drugs. 2018; 16(5): 162-195.

19. Giordano D, Costantini M, Coppola D, et al. Biotechnological applications of bioactive peptides from marine sources. Adv Microb Physiol. 2018; 73: 171-220.

*A current comprehensive review on marine bioactive peptides.

20. Dou X, Dong B. Origins and bioactivities of natural compounds derived from marine ascidians and their symbionts. Mar Drugs. 2019; 17(12): 670-694.

**An updated comprehensive review of marine ascidian natural products particularly higlighthing those from their microbial symbionts equally producing diverse natural products.

21. Stonik VA, Stonik IV. Sterol and sphingoid glycoconjugates from microalgae. Mar Drugs. 2018; 16(12): 514-534.

22. Morita M, Schmidt EW. Parallel lives of symbionts and hosts: chemical mutualism in marine animals. Nat Prod Rep. 2018; 35(4): 357-378.

**Highlights the chemistry involve in the symbiosis between associated microrganisms and their respective marine hosts.

23. Jimenez C. Marine natural products in medicinal chemistry. ACS Med Chem Lett. 2018; 9(10): 959-961.

24. Katz J, Janik JE, Younes A. Brentuximab Vedotin (SGN-35). Clin Cancer Res. 2011; 17(20): 6428-6436.

25. Yamamoto N, Andoh M, Kawahara M, et al. Phase I study of TZT-1027, a novel synthetic dolastatin 10 derivative and inhibitor of tubulin polymerization, given weekly to advanced solid tumor patients for 3 weeks. Cancer Sci. 2009; 100(2): 316-321.

26. Bai R, Pettit GR, Hamel E. Dolastatin 10, a powerful cytostatic peptide derived from a marine animal. Inhibition of tubulin polymerization mediated through the vinca alkaloid binding domain. Biochem Pharmacol. 1990; 39(12): 1941-1949.

27. Mooberry SL, Leal RM, Tinley TL, et al. The molecular pharmacology of symplostatin 1: a new antimitotic dolastatin 10 analog. Int J Cancer. 2003; 104(4): 512-521.

28. Weckwerth W. Metabolomics: Methods and Protocols: Totowa, NJ : Humana Press; 2007.

**An excellent textbook for handling, processing, and interpreting metabolomics data

29. Shaffer M, Armstrong AJS, Phelan VV, et al. Microbiome and metabolome data integration provides insight into health and disease. Transl Res. 2017; 189: 51-64.

30. Wang XJ, Yan GL, Zhang AH, et al. Metabolomics and proteomics approaches to characterize and assess proteins of bear bile powder for hepatitis C virus. Chin J Nat Med. 2013; 11(6): 653-665.

31. Kusonmano K, Vongsangnak W, Chumnanpuen P. Informatics for metabolomics. Adv Exp Med Biol. 2016; 939: 91-115.

32. Eknoyan G. Santorio Sanctorius (1561-1636) - founding father of metabolic balance studies. Am J Nephrol. 1999; 19(2): 226-233.

33. Kohler R. The background to Eduard Buchner's discovery of cell-free fermentation. J Hist Biol. 1971; 4(1): 35-61.

34. Gates SC, Sweeley CC. Quantitative metabolic profiling based on gas chromatography. Clin Chem. 1978; 24(10): 1663-1973.

35. van der Greef J, van Wietmarschen H, van Ommen B, et al. Looking back into the future: 30 years of metabolomics at TNO. Mass Spectrom Rev. 2013; 32(5): 399-415.

36. Horning EC, Horning MG. Metabolic profiles: gas-phase methods for analysis of metabolites. Clin Chem. 1971; 17(8): 802-809.

37. Pauli GF, Godecke T, Jaki BU, et al. Quantitative 1H NMR. Development and potential of an analytical method: an update. J Nat Prod. 2012; 75(4): 834-851.

38. Pauli GF, Chen SN, Simmler C, et al. Importance of purity evaluation and the potential of quantitative (1)H NMR as a purity assay. J Med Chem. 2014; 57(22): 9220−9231.

39. Correa J, Pinto LF, Zhao L, et al. Filtering the NMR spectra of complex mixtures through polymer-mediated paramagnetic spin relaxation. Chemistry. 2018; 24(72): 19236-19242.

40. Novoa-Carballal R, Fernandez-Megia E, Jimenez C, et al. NMR methods for unravelling the spectra of complex mixtures. Nat Prod Rep. 2011; 28(1): 78-98.

**An excellent guide on handling and interpresting the NMR data for complex mixtures.

41. Smith Y. History of Metabolomics: @NewsMedical; 2017 [updated 2017-03-08; access date 28/08/2019]. Available from: https://www.news-medical.net/life-sciences/History-of-Metabolomics.aspx

42. Klassen A, Faccio AT, Canuto GA, et al. Metabolomics: Definitions and significance in systems biology. Adv Exp Med Biol. 2017; 965: 3-17.

43. Cuperlovic-Culf M, Culf A, Morin P, et al. Application of metabolomics in drug discovery, development and theranostics. . Curr Metabolomics. 2013; 1(1): 41-57.

44. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud. 2015; 1(1): a000588.

45. Rabinowitz JD, Purdy JG, Vastag L, et al. Metabolomics in drug target discovery. Cold Spring Harb Symp Quant Biol. 2011; 76: 235-246.

46. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018; 46(D1): D608-D617.

47. Wishart DS. Applications of metabolomics in drug discovery and development. Drugs R D. 2008; 9(5): 307-322.

*An excellent introduction on the application of metabolomics in drug disccovery.

48. Kosmides AK, Kamisoglu K, Calvano SE, et al. Metabolomic fingerprinting: challenges and opportunities. Crit Rev Biomed Eng. 2013; 41(3): 205-221.

49. Schlegel V. Metabolomic Foot Printing and Microbial Pathogens: Metabolomics: Open Access; 2015 [access date 16/12/2017]. Available from: https://www.omicsonline.org/open-access/metabolomic-foot-printing-and-microbial-pathogens-2153-0769-1000e132.php?aid=52991

50. Chiu HH, Kuo CH. Gas chromatography-mass spectrometry-based analytical strategies for fatty acid analysis in biological samples. J Food Drug Anal. 2020; 28(1): 60-73.

51. Rodrigues AM, Miguel C, Chaves I, et al. Mass spectrometry-based forest tree metabolomics. Mass Spectrom Rev. 2020; 39(1): 1- 32.

52. Wang XJ, Ren JL, Zhang AH, et al. Novel applications of mass spectrometry-based metabolomics in herbal medicines and its active ingredients: Current evidence. Mass Spectrom Rev. 2019; 38(4-5): 380-402.

53. Wang S, Blair IA, Mesaros C. Analytical methods for mass spectrometry-based metabolomics studies. Adv Exp Med Biol. 2019; 1140: 635-647.

54. Wang R, Yin Y, Zhu ZJ. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. Anal Bioanal Chem. 2019; 411(19): 4349-4357.

**Highlights liquid chromatography-mass spectrometry (LC-MS) with data-independent acquisition (DIA) as a powerful technology for untargeted metabolomics

55. Duncan KD, Fyrestam J, Lanekoff I. Advances in mass spectrometry based single-cell metabolomics. Analyst. 2019; 144(3): 782-793.

56. Diez-Simon C, Mumm R, Hall RD. Mass spectrometry-based metabolomics of volatiles as a new tool for understanding aroma and flavour chemistry in processed food products. Metabolomics. 2019; 15(3): 41-61.

57. Emwas AH, Roy R, McKay RT, et al. NMR spectroscopy for metabolomics research. Metabolites. 2019; 9(7): 123-162.

** Explored and highlighted emerging NMR techniques and technologies that are used to strengthen its employment while overcoming its inherent limitations in metabolomic applications.

58. Vignoli A, Ghini V, Meoni G, et al. High-throughput metabolomics by 1D NMR. Angew Chem Int Ed Engl. 2019; 58(4): 968-994.

59. Ranjan R, Sinha N. Nuclear magnetic resonance (NMR)-based metabolomics for cancer research. NMR Biomed. 2019; 32(10): e3916.

60. Lucas-Torres C, Wong A. Current developments in microMAS NMR analysis for metabolomics. Metabolites. 2019; 9(2): 29-41.

61. Consonni R, Cagliani LR. The potentiality of NMR-based metabolomics in food science and food authentication assessment. Magn Reson Chem. 2019; 57(9): 558-578.

62. Song Z, Wang H, Yin X, et al. Application of NMR metabolomics to search for human disease biomarkers in blood. Clin Chem Lab Med. 2019; 57(4): 417-441.

63. De Castro F, Benedetti M, Del Coco L, et al. NMR-based metabolomics in metal-based drug research. Molecules. 2019; 24(12): 2240-2254.

64. Volpe MG, Costantini S, Coccia E, et al. Evaluation of metabolic changes induced by polyphenols in the crayfish *Astacus leptodactylus* by metabolomics using Fourier transformed infrared spectroscopy. J Biosci. 2018; 43(4): 585-596.

65. Dai S, Lin Z, Xu B, et al. Metabolomics data fusion between near infrared spectroscopy and high-resolution mass spectrometry: A synergetic approach to boost performance or induce confusion. Talanta. 2018; 189: 641-648.

*A rare study that utilized the use of concatenated datasets from two analytical methods featuring IR and HRMS.

66. Pilatti FK, Ramlov F, Schmidt EC, et al. Metabolomics of *Ulva lactuca* Linnaeus (Chlorophyta) exposed to oil fuels: Fourier transform infrared spectroscopy and multivariate analysis as tools for metabolic fingerprint. Mar Pollut Bull. 2017; 114(2): 831-836.

67. Martens J, Berden G, van Outersterp RE, et al. Molecular identification in metabolomics using infrared ion spectroscopy. Sci Rep. 2017; 7(1): 3363-3368.

68. Yusof NA, Isha A, Ismail IS, et al. Infrared-metabolomics approach in detecting changes in *Andrographis paniculata* metabolites due to different harvesting ages and times. J Sci Food Agric. 2015; 95(12): 2533-2543.

69. Kuligowski J, Perez-Guaita D, Escobar J, et al. Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. Talanta. 2014; 127: 181-190.

70. Alonso A, Marsal S, Julia A. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol. 2015; 3(5): fbioe.2015.00023.

71. Zhang A, Sun H, Wang P, et al. Modern analytical techniques in metabolomics analysis. Analyst. 2012; 137(2): 293-300.

72. Gowda GA, Djukovic D. Overview of mass spectrometry-based metabolomics: opportunities and challenges. Methods Mol Biol. 2014; 1198: 3-12.

73. Yang JY, Sanchez LM, Rath CM, et al. Molecular networking as a dereplication strategy. J Nat Prod. 2013; 76(9): 1686-1699.

74. Demetrowitsch T, Schwarz K. Metabolomics: new analytical methods for metabolome research. Ernaehrungs-Umschau. 2014; 61(7): 102-111.

75. Sussulini A. Metabolomics : from fundamentals to clinical applications. Cham, Switzerland: Springer; 2017. (Sussulini A, editor.).

76. Patel KN, Patel JK, Patel MP, et al. Introduction to hyphenated techniques and their applications in pharmacy. Pharm Methods. 2010; 1(1): 2-13.

77. Wolfender JL, Marti G, Thomas A, et al. Current approaches and challenges for the metabolite profiling of complex natural extracts. J Chromatogr A. 2015; 1382: 136-64.

**An excellent review highlighting both the advanatges and disadvantages of the most common used hyphenated sytstems

78. Wolfender JL, Nuzillard JM, van der Hooft JJJ, et al. Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics. Anal Chem. 2019; 91(1): 704-742.

**A comprehensive review utilising MS and NMR datasets for chemometrics and dereplication study.

79. Nováková L, Matysová L, Solich P. Advantages of application of UPLC in pharmaceutical analysis. Talanta. 2006; 68(3): 908-918.

80. Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics. 2001; 2(3): 155-168.

81. Looser R, Krotzky AJ, Trethewey RN. Metabolite profiling with GC-MS and LC-MS. In: Vaidyanathan S, Harrigan GG, Goodacre R, editors. Metabolome Analyses: Strategies for Systems Biology. Boston, MA: Springer US; 2005. p. 103-118.

82. Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010; 11: 395-406.

83. Myers OD, Sumner SJ, Li S, et al. Detailed investigation and comparison of the XCMS and MZmine 2 Chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data. Anal Chem. 2017; 89(17): 8689-8695.

*An excellent critical review of two commonly use pre-processing platform for MS datasets.

84. Hu M, Krauss M, Brack W, et al. Optimization of LC-Orbitrap-HRMS acquisition and MZmine 2 data processing for nontarget screening of environmental samples using design of experiments. Anal Bioanal Chem. 2016; 408(28): 7905-7915.

85. Scheltema RA, Jankevics A, Jansen RC, et al. PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. Anal Chem. 2011; 83(7): 2786-93.

86. Smith CA, Want EJ, O'Maille G, et al. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006; 78(3): 779-787.

87. Davidson RL, Weber RJ, Liu H, et al. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. Gigascience. 2016; 5(1): 10-19.

88. Wei X, Sun W, Shi X, et al. MetSign: A computational platform for high-resolution mass spectrometry-based metabolomics. Anal Chem. 2011; 83(20): 7668-7675.

89. Fernández-Albert F, Llorach R, Andrés-Lacueva C, et al. An R package to analyse LC/MS metabolomic data: MAIT (Metabolite Automatic Identification Toolkit). Bioinformatics. 2014; 30(13): 1937-1939.

90. Zhang W, Chang J, Lei Z, et al. MET-COFEA: A liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation. Anal Chem. 2014; 86(13): 6245-6253.

91. Lopez-Fernandez H, Santos HM, Capelo JL, et al. Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. BMC Bioinformatics. 2015; 16(1): 318-330.

92. Pedrioli PG, Eng JK, Hubley R, et al. A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol. 2004; 22(11): 1459-66.

93. Adusumilli R, Mallick P. Data conversion with ProteoWizard msConvert. Methods Mol Biol. 2017; 1550: 339-368.

94. Kaiser KA, Merrywell CE, Fang F, et al. Chapter 5 - Metabolic Profiling. In: Holzgrabe U, Wawer I, Diehl B, editors. NMR Spectroscopy in Pharmaceutical Analysis. Amsterdam: Elsevier; 2008. p. 233-267.

95. Markley JL, Bruschweiler R, Edison AS, et al. The future of NMR-based metabolomics. Curr Opin Biotechnol. 2017; 43: 34-40.

96. Seger C, Sturm S, Stuppner H. Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques--state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. Nat Prod Rep. 2013; 30(7): 970-987.

97. Jaroszewski JW. Hyphenated NMR methods in natural products research, part 1: direct hyphenation. Planta Med. 2005; 71(8): 691-700.

98. Jaroszewski JW. Hyphenated NMR methods in natural products research, Part 2: HPLC-SPE-NMR and other new trends in NMR hyphenation. Planta Med. 2005; 71(9): 795-802.

99. Peng SX. Hyphenated HPLC-NMR and its applications in drug discovery. Biomed Chromatogr. 2000; 14(6): 430-441.

100. Lindon JC, Nicholson JK, Wilson ID. The development and application of coupled HPLC-NMR spectroscopy. Adv Chromatogr. 1996; 36: 315-382.

101. Ramadan Z, Jacobs D, Grigorov M, et al. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. Talanta. 2006; 68(5): 1683-1691.

102. Kemsley EK, Le Gall G, Dainty JR, et al. Multivariate techniques and their application in nutrition: a metabolomics case study. Br J Nutr. 2007; 98(1): 1-14.

103. Madala NE, Piater LA, Steenkamp PA, et al. Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and *Sorghum bicolor* cells. Springerplus. 2014; 3(1): 254-264.

104. PennState ECoS. Lesson 16 - Multivariate Statistics and Dimension Reduction | STAT 555 2017 [cited 2017; access date 14/02/2018]. Available from: https://onlinecourses.science.psu.edu/stat555/node/15

105. Williams Y. Outlier in Statistics: Definition & Explanation - Video & Lesson Transcript | Study.com 2017 [cited 2017; access date 14/02/2018]. Available from: http://study.com/academy/lesson/outlier-in-statistics-definition-lesson-quiz.html

106. Chanana S, Thomas CS, Braun DR, et al. Natural product discovery using planes of principal component analysis in R (PoPCAR). Metabolites. 2017; 7(3): 34-46.

107. Barnes S, Benton HP, Casazza K, et al. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. J Mass Spectrom. 2016; 51(8): 535-548.

* Comprehensively described the methods of metabolomics on data analysis,

108. Karamizadeh S, Abdullah SM, Manaf AA, et al. An overview of principal component analysis. J Signal Inf Process. 2013; 4(3): 173-175.

109. Claudino WM, Quattrone A, Biganzoli L, et al. Metabolomics: Available results, current research projects in breast cancer, and future applications. J Clin Oncol. 2007; 25(19): 2840-2846.

110. Worley B, Powers R. PCA as a practical indicator of OPLS-DA model reliability. Curr Metabolomics. 2016; 4(2): 97-103.

** An excellent guide in utising multivariate analysis in metabolomics profiling.

111. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. Comput Struct Biotechnol J. 2013; 4: e201301009.

112. Abdelmohsen UR, Cheng C, Viegelmann C, et al. Dereplication strategies for targeted isolation of new antitrypanosomal actinosporins A and B from a marine sponge associated-*Actinokineospora* sp. EG49. Mar Drugs. 2014; 12(3): 1220-1244.

113. Cheng C, MacIntyre L, Abdelmohsen UR, et al. Biodiversity, anti-trypanosomal activity screening, and metabolomic profiling of Actinomycetes isolated from Mediterranean sponges. PLoS One. 2015; 10(9): e0138528.

114. Raheem DJ, Tawfike AF, Abdelmohsen UR, et al. Application of metabolomics and molecular networking in investigating the chemical profile and antitrypanosomal activity of British bluebells (*Hyacinthoides non-scripta*). Sci Rep. 2019; 9(1): 2547-2760.

115. Kamal N, Viegelmann CV, Clements CJ, et al. Metabolomics-guided isolation of anti-trypanosomal metabolites from the endophytic fungus *Lasiodiplodia theobromae*. Planta Med. 2017; 83(6): 565-573.

116. Mazlan NW, Tate R, Yusoff YM, et al. Metabolomics-guided isolation of anti-trypanosomal compounds from endophytic fungi of the mangrove plant *Avicennia lanata*. Curr Med Chem. 2019; 26: 1-20.

117. Bose U, Hewavitharana AK, Ng YK, et al. LC-MS-based metabolomics study of marine bacterial secondary metabolite and antibiotic production in *Salinispora arenicola*. Mar Drugs. 2015; 13(1): 249-266.

118. Ng YK, Hodson MP, Hewavitharana AK, et al. Effects of salinity on antibiotic production in sponge-derived *Salinispora* actinobacteria. J Appl Microbiol. 2014; 117(1): 109-125.

119. Kale NS, Haug K, Conesa P, et al. MetaboLights: An open-access database repository for metabolomics data. Curr Protoc Bioinformatics. 2016; 53(1): 14.13.1-14.13.18.

120. Quinn RA, Nothias LF, Vining O, et al. Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. Trends Pharmacol Sci. 2017; 38(2): 143-154.

*An excellent and comprehensive review on molecular networking.

121. Wallace IM, Bader GD, Giaever G, et al. Displaying chemical information on a biological network using Cytoscape. Methods Mol Biol. 2011; 781: 363-376.

122. Naman CB, Rattan R, Nikoulina SE, et al. Integrating molecular networking and biological assays to target the isolation of a cytotoxic cyclic octapeptide, samoamide A, from an American Samoan marine cyanobacterium. J Nat Prod. 2017; 80(3): 625-633.

123. Watrous J, Roach P, Alexandrov T, et al. Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A. 2012; 109(26): E1743-E1752.

124. Purves K, Macintyre L, Brennan D, et al. Using molecular networking for microbial secondary metabolite bioprospecting. Metabolites. 2016; 6(1): 2-20.

125. World Health Organisation. Antimicrobial resistance 2017 [access date 16/12/2017]. Available from: http://www.who.int/mediacentre/factsheets/fs194/en/

126. Garg N, Zeng Y, Edlund A, et al. Spatial molecular architecture of the microbial community of a *Peltigera* lichen. mSystems. 2016; 1(6): e00139-16.

127. Sung AA, Gromek SM, Balunas MJ. Upregulation and identification of antibiotic activity of a marine-derived *Streptomyces* sp. via co-cultures with human pathogens. Mar Drugs. 2017; 15(8): 250-262.

128. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov. 2015; 14(2): 111-129.

129. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod. 2012; 75(3): 311-35.

130. Newman DJ, Cragg GM. Natural products as sources of new Drugs from 1981 to 2014. J Nat Prod. 2016; 79(3): 629-661.

131. Cox DG, Oh J, Keasling A, et al. The utility of metabolomics in natural product and biomarker characterization. Biochim Biophys Acta. 2014; 1840(12): 3460-3474.

132. Powers R. The current state of drug discovery and a potential role for NMR metabolomics. J Med Chem. 2014; 57(14): 5860-5870.

133. Kraljevic S, Stambrook PJ, Pavelic K. Accelerating drug discovery. EMBO Rep. 2004; 5(9): 837-842.

134. Bayet-Robert M, Lim S, Barthomeuf C, et al. Biochemical disorders induced by cytotoxic marine natural products in breast cancer cells as revealed by proton NMR spectroscopy-based metabolomics. Biochem Pharmacol. 2010; 80(8): 1170-1179.

135. Wu C, Kim HK, van Wezel GP, et al. Metabolomics in the natural products field-a gateway to novel antibiotics. Drug Discov Today Technol. 2015; 13: 11-17.

136. Conly J, Johnston B. Where are all the new antibiotics? The new antibiotic paradox. Can J Infect Dis Med Microbiol. 2005; 16(3): 159-60.

137. Wu C, Du C, Gubbens J, et al. Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. J Nat Prod. 2015; 78(10): 2355-2363.

138. Nyfeler R, Keller-Schierlein W. Metabolites of microorganisms. 143. Echinocandin B, a novel polypeptide-antibiotic from *Aspergillus nidulans* var. *echinulatus*: isolation and structural components. Helv Chim Acta. 1974; 57(8): 2459-2477.

139. Roullier C, Guitton Y, Valery M, et al. Automated detection of natural halogenated compounds from LC-MS profiles-application to the isolation of bioactive chlorinated compounds from marine-derived fungi. Anal Chem. 2016; 88(18): 9143-9150.

140. Roullier C, Bertrand S, Blanchet E, et al. Time dependency of chemodiversity and biosynthetic pathways: An LC-MS metabolomic study of marine-sourced *Penicillium*. Mar Drugs. 2016; 14(5): 103-118.

141. Parrot D, Blumel M, Utermann C, et al. Mapping the surface microbiome and metabolome of brown seaweed *Fucus vesiculosus* by Amplicon sequencing, integrated metabolomics and imaging techniques. Sci Rep. 2019; 9(1): 1061-1078.

142. Petersen LE, Marner M, Labes A, et al. Rapid metabolome and bioactivity profiling of fungi associated with the leaf and rhizosphere of the Baltic seagrass *Zostera marina*. Mar Drugs. 2019; 17(7): 419-443.

143. Oppong-Danquah E, Parrot D, Blumel M, et al. Molecular networking-based metabolome and bioactivity analyses of marine-adapted fungi co-cultivated with phytopathogens. Front Microbiol. 2018; 9: 2072-2107.

144. Overy DP, Bayman P, Kerr RG, et al. An assessment of natural product discovery from marine (sensu strictu) and marine-derived fungi. Mycology. 2014; 5(3): 145-167.

145. Hayashi Y, Takeno H, Chinen T, et al. Development of a new benzophenone-diketopiperazine-type potent antimicrotubule agent possessing a 2-pyridine structure. ACS Med Chem Lett. 2014; 5(10): 1094-1098.

146. Bugni TS, Ireland CM. Marine-derived fungi: a chemically and biologically diverse group of microorganisms. Nat Prod Rep. 2004; 21(1): 143-163.

147. Jones EBG, Sakayaroj J, Seutrong S, et al. Classification of marine ascomycota, anamorphic taxa and basidiomycota. Fungal Diversity. 2009; 35: 1-187.

148. Lau APS, Lee AKY, Chan CK, et al. Ergosterol as a biomarker for the quantification of the fungal biomass in atmospheric aerosols. Atmos Environ. 2006; 40(2): 249-259.

149. Pasanen AL, Yli-Pietila K, Pasanen P, et al. Ergosterol content in various fungal species and biocontaminated building materials. Appl Environ Microbiol. 1999; 65(1): 138-142.

150. Poirier L, Montagu M, Landreau A, et al. Peptaibols: stable markers of fungal development in the marine environment. Chem Biodivers. 2007; 4(6): 1116-1128.

151. Poirier L, Amiard JC, Mondeguer F, et al. Determination of peptaibol trace amounts in marine sediments by liquid chromatography/electrospray ionization-ion trap-mass spectrometry. J Chromatogr A. 2007; 1160(1-2): 106-113.

152. Barka EA, Vatsa P, Sanchez L, et al. Taxonomy, physiology, and natural products of Actinobacteria. Microbiol Mol Biol Rev. 2016; 80(1): 1-43.

153. Euzéby JP. List of bacterial names with standing in nomenclature: a folder available on the internet. Int J Syst Bacteriol. 1997; 47(2): 590-592.

154. Mincer TJ, Fenical W, Jensen PR. Culture-dependent and culture-independent diversity within the obligate marine actinomycete genus *Salinispora*. Appl Environ Microbiol. 2005; 71(11): 7019-7028.

155. Udwary DW, Zeigler L, Asolkar RN, et al. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proc Natl Acad Sci U S A. 2007; 104(25): 10376-10381.

156. Chater KF. Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. Philos Trans R Soc Lond B Biol Sci. 2006; 361(1469): 761-768.

157. Kemung HM, Tan LT, Khan TM, et al. Streptomyces as a prominent resource of future anti-MRSA drugs. Front Microbiol. 2018; 9: 2221-2247.

158. Butler MS, Blaskovich MA, Cooper MA. Antibiotics in the clinical pipeline in 2013. . J Antibiot (Tokyo). 2013; 66(10): 571-591.

159. Ventura M, Canchaya C, Tauch A, et al. Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev. 2007; 71(3): 495-548.

160. Paulus C, Rebets Y, Tokovenko B, et al. New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18. Sci Rep. 2017; 7: 42382-42393.

161. Tortorella E, Tedesco P, Palma Esposito F, et al. Antibiotics from deep-sea microorganisms: Current discoveries and perspectives. Mar Drugs. 2018; 16(10): 355-371.

162. Viegelmann C, Margassery LM, Kennedy J, et al. Metabolomic profiling and genomic study of a marine sponge-associated *Streptomyces* sp. Mar Drugs. 2014; 12(6): 3323-3351.

163. Forner D, Berrue F, Correa H, et al. Chemical dereplication of marine actinomycetes by liquid chromatography-high resolution mass spectrometry profiling and statistical analysis. Anal Chim Acta. 2013; 805: 70-79.

164. Adnani N, Vazquez-Rivera E, Adibhatla SN, et al. Investigation of interspecies interactions within marine *Micromonosporaceae* using an improved co-culture approach. Mar Drugs. 2015; 13(10): 6082-6098.

165. Bose U, Hewavitharana AK, Vidgen ME, et al. Discovering the recondite secondary metabolome spectrum of *Salinispora* species: a study of inter-species diversity. PLoS One. 2014; 9(3): e91488.

166. Duncan KR, Haltli B, Gill KA, et al. Exploring the diversity and metabolic potential of actinomycetes from temperate marine sediments from Newfoundland, Canada. J Ind Microbiol Biotechnol. 2015; 42(1): 57-72.

167. Tian J, Chen H, Guo Z, et al. Discovery of pentangular polyphenols hexaricins A-C from marine *Streptosporangium* sp. CGMCC 4.7309 by genome mining. Appl Microbiol Biotechnol. 2016; 100(9): 4189-4199.

168. Sidebottom AM, Johnson AR, Karty JA, et al. Integrated metabolomics approach facilitates discovery of an unpredicted natural product suite from *Streptomyces coelicolor* M145. ACS Chem Biol. 2013; 8(9): 2009-2016.

169. Adnani N, Chevrette MG, Adibhatla SN, et al. Coculture of marine invertebrate-associated bacteria and interdisciplinary technologies enable biosynthesis and discovery of a new antibiotic, keyicin. ACS Chem Biol. 2017; 12(12): 3093-3102.

170. Bose U, Ortori CA, Sarmad S, et al. Production of N-acyl homoserine lactones by the sponge-associated marine actinobacteria *Salinispora arenicola* and *Salinispora pacifica*. FEMS Microbiol Lett. 2017; 364(2): fnx002.

171. Cheng C, Othman EM, Stopper H, et al. Isolation of petrocidin A, a new cytotoxic cyclic dipeptide from the marine sponge-derived bacterium *Streptomyces* sp. SBT348. Mar Drugs. 2017; 15(12): 383-391.

172. Shi Y, Pan C, Auckloo BN, et al. Stress-driven discovery of a cryptic antibiotic produced by *Streptomyces* sp. WU20 from Kueishantao hydrothermal vent with an integrated metabolomics strategy. Appl Microbiol Biotechnol. 2017; 101(4): 1395-1408.

173. Takehana Y, Umekita M, Hatano M, et al. Fradiamine A, a new siderophore from the deep-sea actinomycete Streptomyces fradiae MM456M-mF7. J Antibiot (Tokyo). 2017; 70(5): 611-615.

174. Villegas-Plazas M, Wos-Oxley ML, Sanchez JA, et al. Variations in microbial diversity and metabolite profiles of the tropical marine sponge *Xestospongia muta* with season and depth. Microb Ecol. 2019; 78(1): 243-256.

175. Betancur LA, Naranjo-Gaybor SJ, Vinchira-Villarraga DM, et al. Marine actinobacteria as a source of compounds for phytopathogen control: An integrative metabolic-profiling / bioactivity and taxonomical approach. PLoS One. 2017; 12(2): e0170148.

176. Dalisay DS, Williams DE, Wang XL, et al. Marine sediment-derived *Streptomyces* bacteria from British Columbia, Canada are a promising microbiota resource for the discovery of antimicrobial natural products. PLoS One. 2013; 8(10): e77078.

177. Bisacchi GS, Manchester JI. A new-class antibacterial-Almost. Lessons in drug discovery and development: A critical analysis of more than 50 Years of effort toward ATPase inhibitors of DNA gyrase and topoisomerase IV. ACS Infect Dis. 2015; 1(1): 4-41.

178. Zhao B, Moody SC, Hider RC, et al. Structural analysis of cytochrome P450 105N1 involved in the biosynthesis of the zincophore, coelibactin. Int J Mol Sci. 2012; 13(7): 8500-13.

179. Jonsson AG, Torstensson NT. Protease inhibitors from *Streptomyces violascens*. II. Production of the inhibitors. Arch Mikrobiol. 1972; 83(1): 71-77.

180. Wu H, Song Z, Hentzer M, et al. Synthetic furanones inhibit quorum-sensing and enhance bacterial clearance in *Pseudomonas aeruginosa* lung infection in mice. J Antimicrob Chemother. 2004; 53(6): 1054-1061.

181. Kitani S, Miyamoto KT, Takamatsu S, et al. Avenolide, a Streptomyces hormone controlling antibiotic production in *Streptomyces avermitilis*. Proc Natl Acad Sci U S A. 2011; 108(39): 16410-16415.

182. Worley B, Powers R. Multivariate analysis in metabolomics. Curr Metabolomics. 2013; 1(1): 92-107.

183. Vincent IM, Ehmann DE, Mills SD, et al. Untargeted metabolomics to ascertain antibiotic modes of action. Antimicrob Agents Chemother. 2016; 60(4): 2281-2291.

184. Rao E, Kumavath R. Bacterial metabolomics reveals the evolution of antibiotic resistance. J Syst Biol Proteome Res. 2017; 1(1): 1-2.

185. Scalbert A, Brennan L, Fiehn O, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009; 5(4): 435–458.

186. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016; 15(7): 473-484.

187. Bafor EE, Rowan EG, Edrada-Ebel R. Toward understanding myometrial regulation: Metabolomic investigation reveals new pathways of oxytocin and ritodrine activity on the myometrium. Reprod Sci. 2017; 24(5): 691-705.

188. Bafor EE, Rowan EG, Edrada-Ebel R. Metabolomics-coupled functional pharmacology of chlorophyll compounds isolated from the leaves of *Ficus exasperata* Vahl (Moraceae) provides novel pathways on myometrial activity. Reprod Sci. 2018; 25(6): 923-937.

189. Gao D, Wang Y, Xie W, et al. Metabolomics study on the antitumor effect of marine natural compound flexibilide in HCT-116 colon cancer cell line. J Chromatogr B Analyt Technol Biomed Life Sci. 2016; 1014: 17-23.

190. Wyche TP, Piotrowski JS, Hou Y, et al. Forazoline A: marine-derived polyketide with antifungal in vivo efficacy. Angew Chem Int Ed Engl. 2014; 53(43): 11583-11586.

191. Morvan D. Functional metabolomics uncovers metabolic alterations associated to severe oxidative stress in MCF7 breast cancer cells exposed to ascididemin. Mar Drugs. 2013; 11(10): 3846-3860.

192. Peeters L, Beirnaert C, Van der Auwera A, et al. Revelation of the metabolic pathway of hederacoside C using an innovative data analysis strategy for dynamic multiclass biotransformation experiments. J Chromatogr A. 2019; 1595: 240-247.

193. Bottcher T, Pitscheider M, Sieber SA. Natural products and their biological targets: proteomic and metabolomic labeling strategies. Angew Chem Int Ed Engl. 2010; 49(15): 2680-2698.

194. Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. ChemBioChem. 2009; 10(4): 625-633.

195. Fu W, Nelson DR, Yi Z, et al. Chapter 6 - Bioactive Compounds From Microalgae: Current Development and Prospects. In: Atta ur R, editor. Studies in Natural Products Chemistry. Vol. 54: Elsevier; 2017. p. 199-225.

196. Kellogg JJ, Todd DA, Egan JM, et al. Biochemometrics for natural products research: Comparison of data analysis approaches and application to identification of bioactive compounds. J Nat Prod. 2016; 79(2): 376-386.

197. Cuevas C, Francesch A. Development of Yondelis (trabectedin, ET-743). A semisynthetic process solves the supply problem. Nat Prod Rep. 2009; 26(3): 322-337.

198. Kim S-K. Handbook of Anticancer Drugs from Marine Origin. Cham, Switzerland: Springer 2018. (Kim S-K, editor.).

199. Assessment report for Yondelis. Procedure No.EMEA/H/C/000773/II/0008 European Medicines Agency 2009 [access date 16/12/2017]. Available from: https://www.ema.europa.eu/en/documents/variation-report/yondelis-h-c-773-ii-0008-epar-assessment-report-variation_en.pdf

200. Rinehart KL, Holt TG, Fregeau NL, et al. Ecteinascidins 729, 743, 745, 759A, 759B, and 770: potent antitumor agents from the Caribbean tunicate *Ecteinascidia turbinata*. J Org Chem. 1990; 55(15): 4512-4515.

201. Olsen EK, Andersen JH. Bioprospecting of Arctic marine organisms: Employing bioassay-, chemistry-, and metabolomics-guided isolation [Doktorgradsavhandling]: UiT Norges arktiske universitet; 2015.

202. Panuganti SJ. Principles involved in bioassay by different methods: A mini-review. Res Rev: Res J Biol. 2015; 3(2): 1-18.

203. Nguta JM, Appiah-Opong R, Nyarko AK, et al. Current perspectives in drug discovery against tuberculosis from natural products. Int J Mycobacteriol. 2015; 4(3): 165-83.

204. Siheri W, Ebiloma GU, Igoli JO, et al. Isolation of a novel flavanonol and an alkylresorcinol with highly potent anti-trypanosomal activity from Libyan propolis. Molecules. 2019; 24(6): 1041-1052.

205. Chagas-Paula DA, Zhang T, Da Costa FB, et al. A metabolomic approach to target compounds from the Asteraceae family for dual COX and LOX inhibition. Metabolites. 2015; 5(3): 404-430.

206. Wu C, Du C, Ichinose K, et al. Discovery of C-glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76 by a Combined NMR-Based metabolomics and bioinformatics workflow. J Nat Prod. 2017; 80(2): 269-277.

Table 1. Application and limitations of different types of mass analysers commonly used as metabolomics tools. [74]

Type of Analyser	Applications and Advantages	Limitations
Quadrupole	• Offers higher sensitivity for lower limits of detection and quantification • Provides good fragmentation of larger molecules like peptides and polysaccharides	• Lower and poorer resolution attained • Fragmentation dependent on collision gas and energy
Ion trap	• Quite compact in construction • With high sensitivity • Suitable for structure clarification, target molecules can be fragmented several times in tandem or multi-staged mass spectrometry (MSn) • Excellent in reaction monitoring and target screening	• Poor quantification capability • Very poor dynamic range • Not well-defined collision energy
Time of flight	• Wide spectrum of applications that can be coupled with various ionization techniques • Excellent high resolution data achieved for small to medium-sized molecules • Separates ions according to their masses, lower masses exhibit shorter flight time • Method of choice for Matrix-Assisted Laser Desorption Ionization (MALDI), which is widely used for imaging	• Flight time is longer for higher masses causing lower resolution • Limited precursor-ion selectivity
Fourier transform ion cyclotron resonance	• High resolution mass spectral data is exceptionally precise achieving < 5ppm accuracy • Excellent tandem MS capability • Usually coupled to electrospray ionization (ESI)	• Limited dynamic range • Presence of artifacts • Expensive • Demanding facility requirements

Table 2. Applications and limitations of most commonly used hyphenated techniques [76-78].

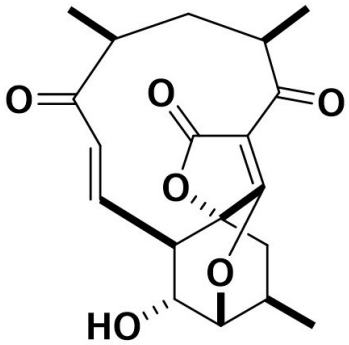
Hyphenated Techniques	Application	Limitations
CE-MS	• The versatility of the CE can be used for the separation of almost all molecules particularly for biofluids; • Used for separating ionised compounds that would move at distinct speeds at a certain applied voltage depending on the size and charge of the respective molecules • Used for purity determination, assays, and trace level determinations. • Hyphenation of CE with MS is suitable because of the low flow rates (10–100 μL/min) required in CE.	• Poor and inadequate reproducibility when tackling large amounts of samples, which makes it unsuitable as a metabolomics tool. • Highly dependent on ionisation capability of the analytes
GC-MS-EI	More structural information based on the interpretation of reproducible fragmentations that can be compared with library spectra.	• Only for compounds that are adequately volatile, small, and stable in high temperature • Can be useful for larger molecules through chemical derivatization but can lead to increase complexity from incomplete reactions
GC-MS-TOF	Confirmation of purity and identity of the components by measuring exact mass and establishing elemental composition	Only for compounds that are adequately volatile, small, and stable in high temperature
LC-MS-ESI/APCI LC-MS-MS/MS	One of the most sensitive and highly selective methods of molecular analysis, and provides information on the molecular weight as well as the fragmentation pattern of the analyte molecule for tandem MS/MS	• Soft ionization techniques that mainly display the molecular ion species with only a few fragment ions • Detection is highly dependent on the ionisation capability of the respective analytes on different mass analysers that offers varying degree of mass accuracy and resolution
LC-IR	Identification of organic compounds by utilizing the many absorption bands at the mid-IR region that are characteristic of particular functionalities, e.g., –OH, –COOH	• IR is much less sensitive compared to various other detection methods such as UV and MS • Extremely slow because of the huge mid-IR region of the 237 absorption bands of the mobile phase solvent often obscure the small signal generated by the sample components
LC-NMR	Provides the most useful structural information toward the structure elucidation of natural products, which is of great value in the analysis of complex mixtures of all types, not only in the analysis of natural products but as well as drug-related metabolites in biofluids.	NMR is probably the least sensitive. The NMR spectrometer is not able to handle the intense solvent signals and the weak signals of the analyte at the same time
LC-PDA-MS	Have proved to be extremely useful in combination with biological screening for a rapid survey of natural products	Detection of the analytes under UV and UV-vis range is dependent on the presence of chromophores

Table 3. Advantages and disadvantages of various chromatographic separation techniques grouped according to their separation principles. [74,77-79]

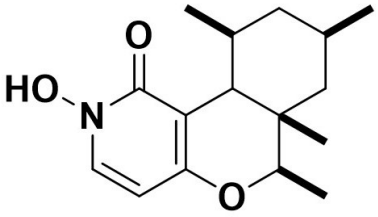
Type of separation principle	Chromatographic method	Advantages and uses	Disadvantages and limitations
A) Separation is based on the volatility of the various analytes according to a temperature gradient ranging usually between 120 and 360 degrees Celsius.	Gas chromatography (GC)	• Precise, robust and widely applicable • Analyte does not have to contain a chromophore • Environmentally friendly – low levels of waste produced • Automation • Usually coupled to Electron Impact Mass spectrometry (EIMS) that offers a huge library of mass spectral data for a broad spectrum of compounds	• Only thermally stable and volatile compounds can be analyzed • More suitable for non-polar compounds • Samples sometimes require chemical derivatisation prior to analysis • Requires an internal standard due to variable injection volume
B) By capillary action based on the affinity of the solute to the stationary phase and its partition coefficient with the mobile phase	Thin layer chromatography (TLC)	• Rapid, easy to follow, requiring minimum training, cheap and uncomplicated instrumentation • Qualitative visualization of separated metabolites • Quick dereplication of metabolites in comparison to a reference standard • Reaction monitoring in synthetic chemistry	• Quantification needs expensive instrumentation (e.g. densitometer), which depends on the absorbance capability of an analyte that is usually more suitable for chromophores • Use of a universal chemical spray reagent for detection is destructive
C) Adsorption and partition based on the difference in affinity of various metabolites or analytes between stationary phase and the mobile phase	Flash and/or Medium pressure liquid chromatography (MPLC)	• High-throughput separation of complex mixtures at higher loading gram level concentrations • Non-destructive for metabolites	• Uses large volumes of solvents with flow rates between 10 to 100ml/min
	High performance liquid chromatography (HPLC)	• Non-destructive • Low sample volume required • Quantitative analysis	• Expensive set-up • Solvent waste • Sensitivity highly dependent on the type of detector used. • Irreversible solute adsorption
	Ultra performance liquid chromatography (UPLC)	• Shorter time of analysis, higher sensitivity, higher resolution and separation efficiency, and lower solvent volumes • Quantitative analysis of low volume of samples (e.g. biofluids)	• Expensive set-up • Higher risk of irreversible solute adsorption

		• Highly suitable for mass spectrometry hyphenated systems	
	Hydrophilic interaction chromatography (HILIC)	• Separation of polar components (e.g. amino sugars, nucleotides)	• Poor reproducibility due to slow equilibration
	Supercritical fluid chromatography (SFC)	• Separation of complex mixtures or extracts • Separation of compounds of low to medium molecular weights or small molecules • No solvent waste because it utilizes CO_2 as mobile phase under pressurized condition • Recommended for high-throughput industrial scale-up isolation work	• Metabolites must be thermally stable • Not recommended for higher molecular weight compounds > 1000 Da
D) Liquid-Liquid partitioning of analytes between two immiscible solvents. The stationary phase is retained by centrifugal force.	Countercurrent chromatography (CCC) High-speed countercurrent chromatography (HSCCC)	• Good separation between lipophilic and hydrophilic analytes • Used in medium scale-up isolation work due to high loading capacity with lesser solvent waste. • 100% recovery of analytes • pH adjustment is not required • Shorter purification time • Does not require expensive columns	• Choice of appropriate suitable solvent-system is decisive, not straight-forward and can be laborious • Emulsification can be a problem • Not for thermo-labile compounds
E) Adsorption based on size exclusion or molecular sizes	Gel permeation chromatography (usually known as LH20 or SEPHADEX according to the type of resin used)	• Set-up is quick, cheap, and simple • 100% recovery of metabolites • Reusable resin solid phases • Separation of primary metabolites from large biomolecules like proteins and polysaccharides	• Very expensive resin solid phases • Cannot be used for purification of closely related structural and stereoisomers
F) Based on different rate of migration through an electric field	Capillary electrophoresis (CE) Capillary isoelectric focusing (CIEF)	• Widely used in chiral separation and amphoteric compounds such as proteins. Low volume of samples required therefore can be used for trace analysis of toxins in biofluids and immunoassays • Low volume of solvents used	• Sensitivity and resolution limit • Metabolites or analytes can be easily degraded • Applicable only to compounds that amphoteric or are easy to ionize • Not useful for chiral compounds

Figure 1. Potential antibiotics against MRSA



abyssomicin C

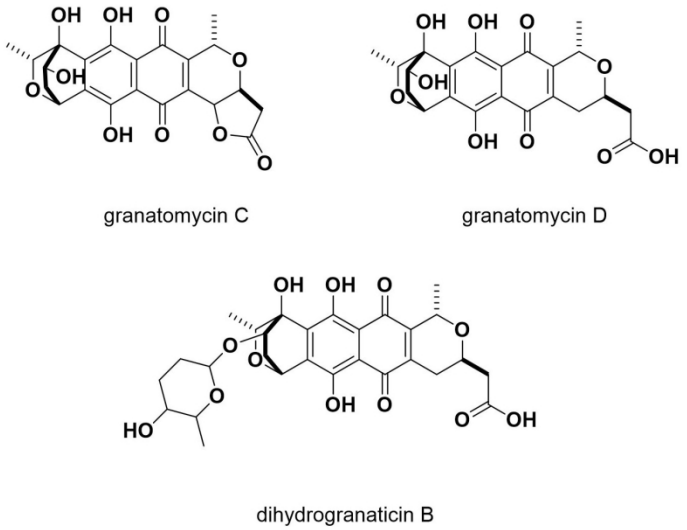


PF1140

Potential antibiotics against MRSA

328x314mm (96 x 96 DPI)

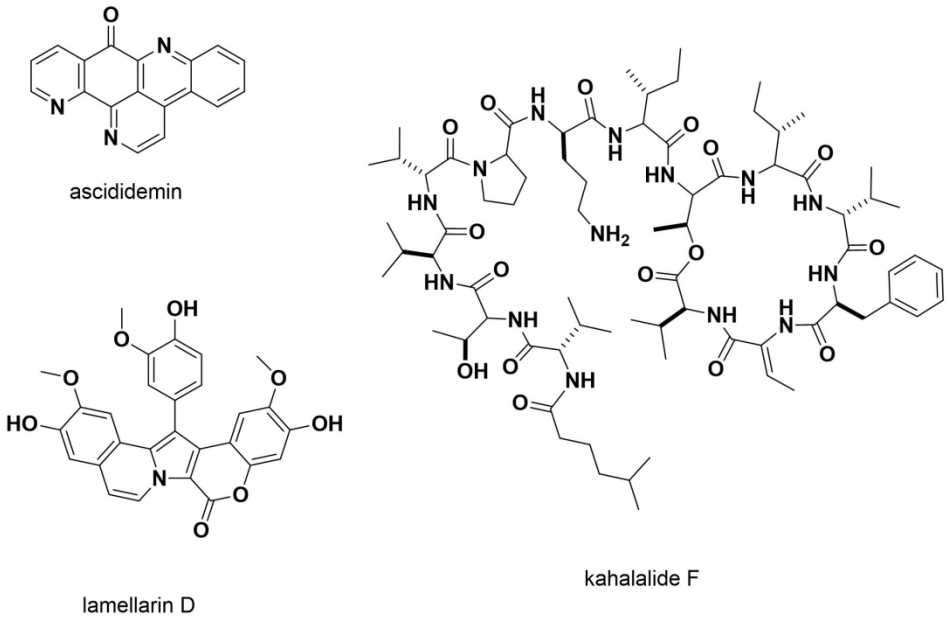
Figure 2. Antibiotic compounds obtained from marine-derived Streptomyces species strain PTY08712



Antibiotic compounds obtained from marine-derived Streptomyces species strain PTY08712

525x379mm (96 x 96 DPI)

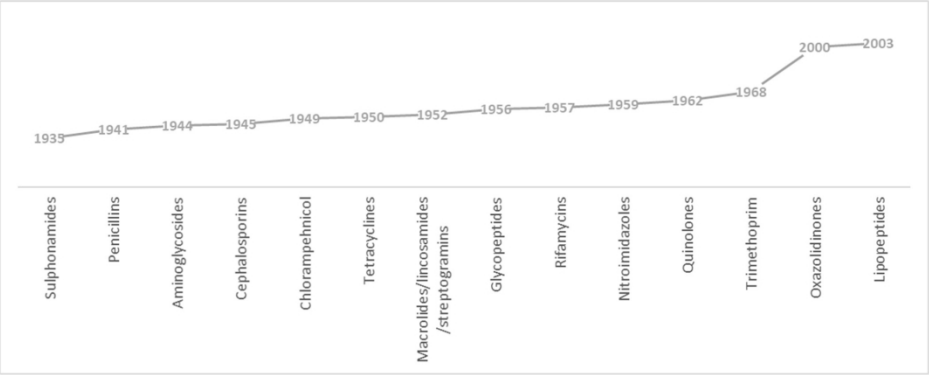
Figure 3. Marine-derived anticancer drugs currently in clinical trials.



Marine-derived anticancer drugs currently in clinical trials

539x442mm (96 x 96 DPI)

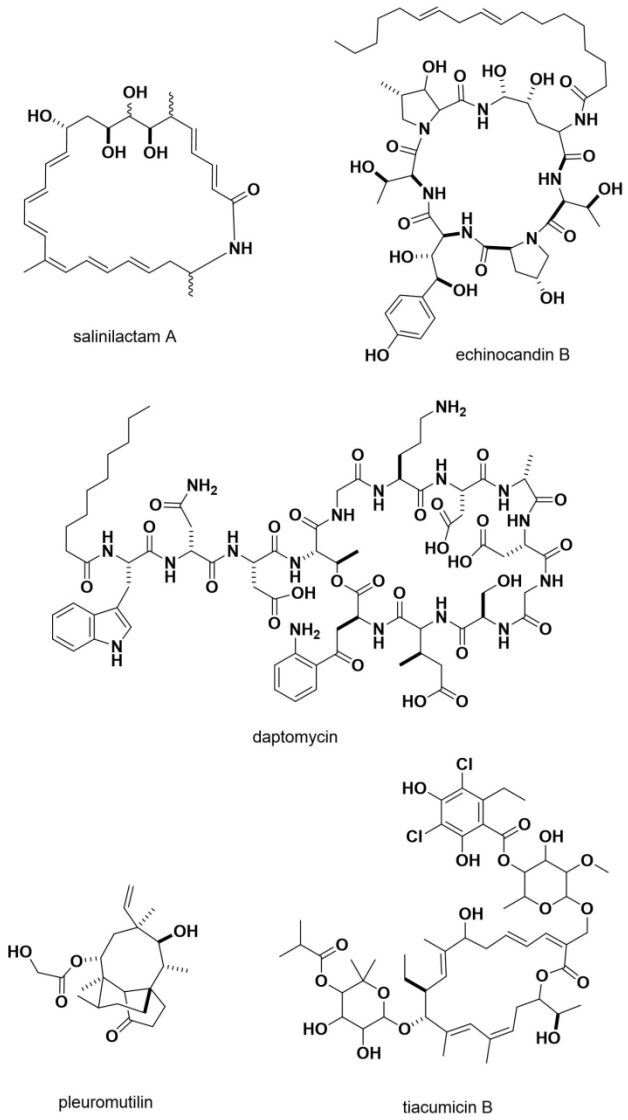
Figure 4. Timeline of the introduction of new antibiotic classes from 1935 – 2003 [136].



Timeline of the introduction of new antibiotic classes from 1935 – 2003

595x303mm (96 x 96 DPI)

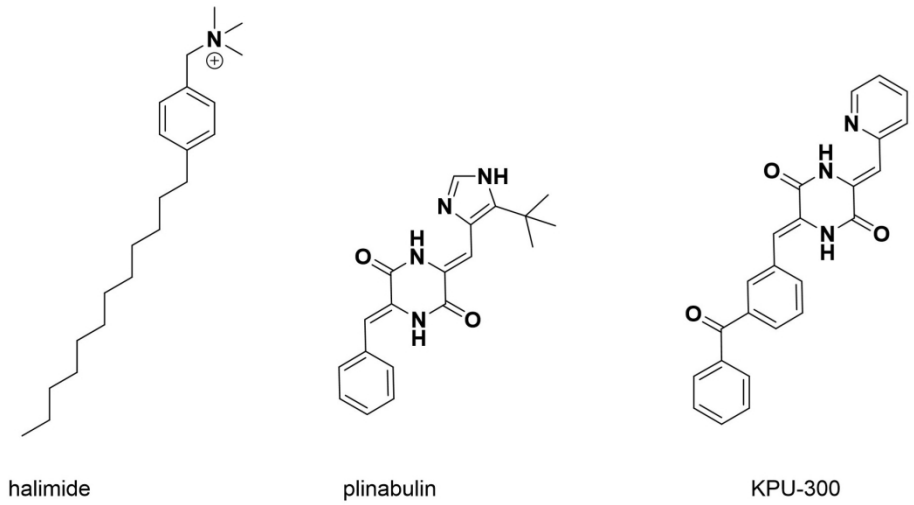
Figure 5. Bioactive compounds from marine-derived microbes



Bioactive compounds from marine-derived microbes

433x731mm (96 x 96 DPI)

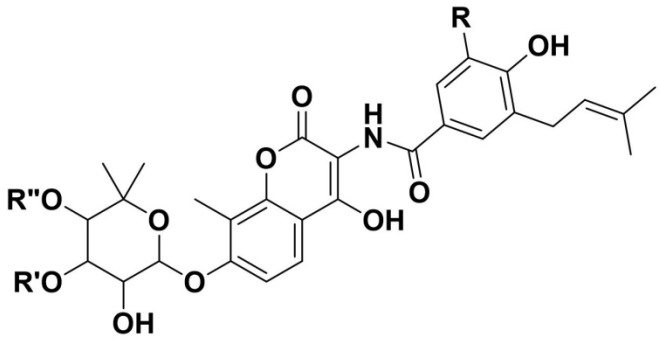
Figure 6. Plinablin and KPU-300 from the marine-derived fungal metabolite halimide



Plinablin and KPU-300 from the marine-derived fungal metabolite halimide

532x342mm (96 x 96 DPI)

Figure 7. Anti-MRSA active novobiocins

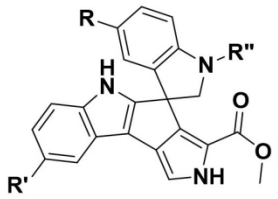


novobiocin: R=H, R'= -C(O)NH_2, R"= Me
desmethyldescarbamoylnovobiocin: R=H, R'= H, R"= H
desmethylnovobiocin: R=H, R'= -C(O)NH_2, R"= H
5-hydroxynovobiocin: R=OH, R'= -C(O)NH_2, R"= Me
desmethyldescarbamoyl-5-hydroxynovobiocin: R=OH, R'= H, R"= H
desmethyl-5-hydroxynovobiocin : R=OH, R'= -C(O)NH_2, R"= H

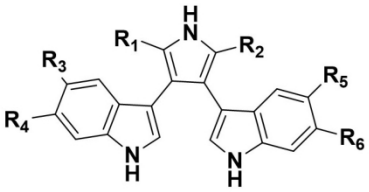
Anti-MRSA active novobiocins

515x407mm (96 x 96 DPI)

Figure 8. Bisindole pyrrole antibiotics



spiroindimicin B: R= Cl, R'= Cl, R"= Me
spiroindimicin C: R= Cl, R'= Cl, R"= H
spiroindimicin E: R= Cl, R'= H, R"= Me
spiroindimicin F: R= H, R'= Cl, R"= Me

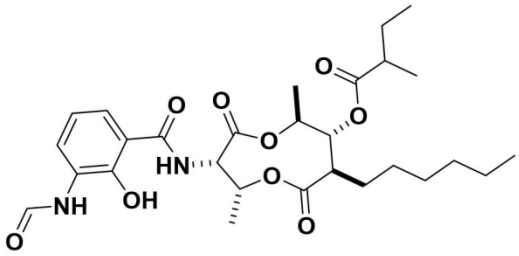


lynamicin A: R_1, R_4, R_6 = H; R_2 = COOMe; R_3, R_5 = Cl
lynamicin B: R_1, R_6 = H; R_2 = COOMe; R_3, R_4, R_5 = Cl
lynamicin C: R_1, R_2 = H; R_3, R_4, R_5, R_6 = Cl
lynamicin D: R_4, R_6 = H; R_1, R_2 = COOMe; R_3, R_5 = Cl
lynamicin E: R_3, R_4, R_6 = H; R_1, R_2 = COOMe; R_5 = Cl

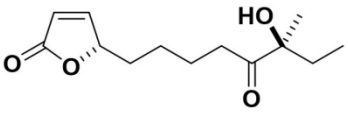
Bisindole pyrrole antibiotics

490x307mm (96 x 96 DPI)

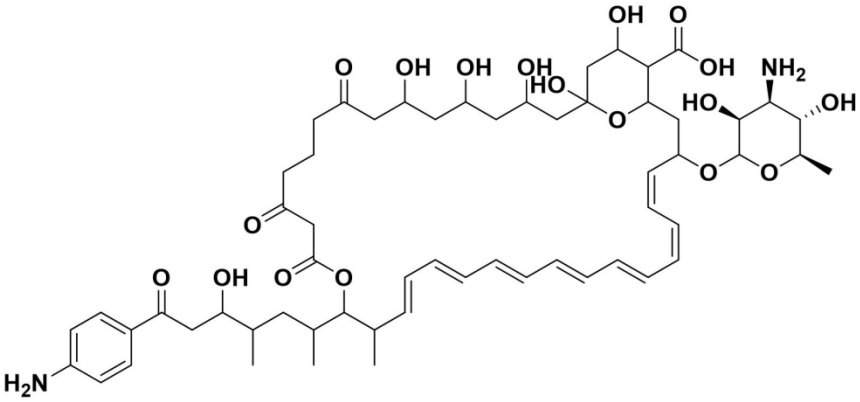
Figure 9. Examples of antibiotic compounds from sponge-derived Streptomyces.



antimycins A1-a



avenolide

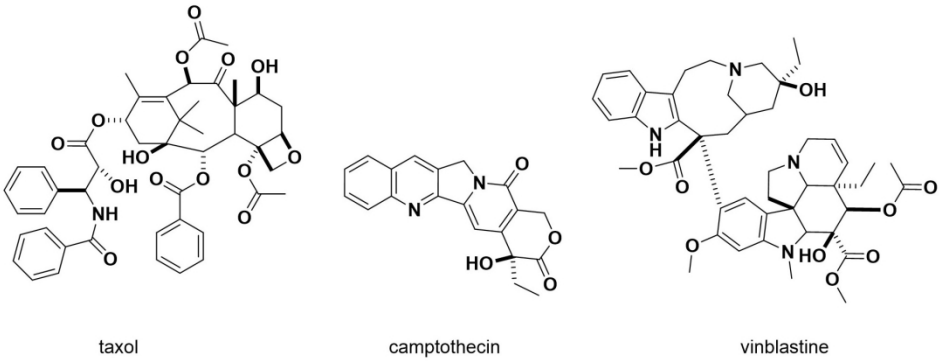


candicidin

Examples of antibiotic compounds from sponge-derived Streptomyces

491x497mm (96 x 96 DPI)

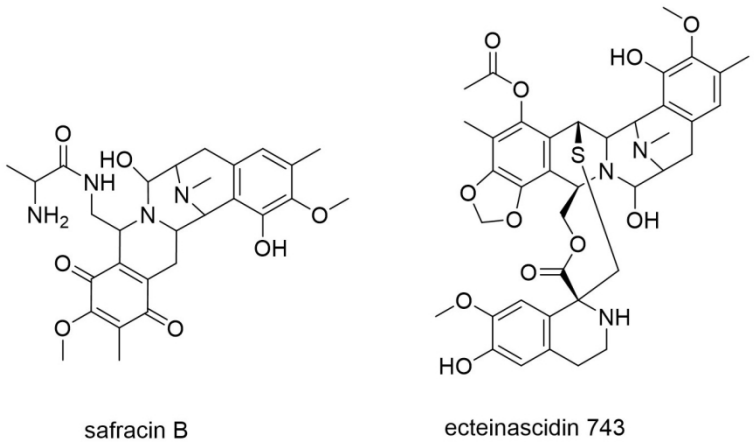
Figure 10. Plant-derived anticancer drugs isolated bioassay-guided fractionation



Plant-derived anticancer drugs isolated bioassay-guided fractionation

550x279mm (96 x 96 DPI)

Figure 11. Structures of safracin B and ecteinascidin 743 also known as yondelis or trabectedin



Structures of safracin B and ecteinascidin 743

478x258mm (96 x 96 DPI)