Giant barrel sponges in diverse habitats: a story about the metabolome

Bayona Maldonado, L.M.

Chapter 2

Recent progress in marine organism metabolomics

Lina M. Bayona [1], Nicole J. de Voogd [2,3], Young Hae Choi [1,*]

[1] Natural Products Laboratory, Institute of Biology, Leiden University, 2333 BE Leiden, The Netherlands.

[2] Naturalis Biodiversity Center, Marine Biodiversity, 2333 CR Leiden, The Netherlands.

[3] Institute of Environmental Sciences, Leiden University, 2333 CC Leiden, The Netherlands.

[*] Corresponding author e-mail: y.choi@chem.leidenuniv.nl

Manuscript submitted to Metabolomics

Abstract

Marine ecosystems are hosts to a vast array of organisms, being among the most richly biodiverse locations on the planet. The study of these ecosystems is very important, as they are not only a significant source of food for the world but have also, in recent years, become a prolific source of compounds with therapeutic potential. Studies of aspects of marine life have involved diverse fields of marine science, and the use of metabolomics as an experimental approach has increased in recent years. As part of the "omics" technologies, metabolomics has been used to deepen the understanding of interactions between marine organisms and their environment at a metabolic level and to discover new metabolites produced by these organisms.

This review provides an overview of the developments in the field of marine organism metabolomics, and applications of metabolomics common to other fields that could potentially contribute to its use in marine organism studies. It deals with the entire process of a metabolomic study, from sample collection considerations, metabolite extraction, analytical techniques and data analysis. It also includes an overview of recent applications of metabolomics in fields such as marine ecology and drug discovery and future perspectives of its use in the study of marine organisms.

Keywords: Marine organisms, metabolomics, drug discovery, marine chemical ecology.

1. Introduction

Marine ecosystems show remarkable differences with their terrestrial counterparts. The first and most evident is that they develop in an aqueous media, with significant implications on the dynamics of the ecosystem. One of the most influential factors is the distribution of species and chemical nutrients in three dimensions, as depth is an additional factor to be considered in the study of marine organisms. Another key difference is the distribution of chemical nutrients and other compounds in a medium which is in continual movement due to waves and currents, resulting in very dynamic systems (Carr et al. 2003). Possibly as a result of this, marine organisms are more sensitive to environmental changes, and react relatively fast. Thus, while changes in terrestrial ecosystems caused by large-scale disruptive events (e.g., global warming) can take centuries, changes in marine ecosystems can be observed within a few decades.

Metabolism is the first to response to environmental changes experienced by a marine organism (Viant 2007). In this context, a comprehensive analysis of the metabolites produced by these organisms is crucial to understand the dynamics of marine ecosystems. Such a comprehensive analysis can be achieved through metabolomics, an "omics" technology defined as the study of all metabolites or small molecules present in an organism, cell or tissue under certain conditions (Bundy et al. 2008). Metabolomics has an incomparable edge over conventional analysis, offering several tools for chemical profiling-related studies such as metabolic fingerprinting and footprinting (Dunn and Ellis 2005). Using a systems biology approach, metabolomics studies have played an essential role in the reconstruction of molecular networks, providing renewed insights into the relationship between genes and the final products of metabolism (Kell 2004; Weckwerth 2003).

In addition to offering a full picture of molecular networks, one of the most important fields in which metabolomics could be useful is in marine chemical ecology, the discipline that investigates the individual biological/physiological roles of marine natural products. This field focuses on the interactions between marine organisms mediated by chemical compounds (Hay 2009). Despite its rapid growth over the past years, the comprehension of the possible roles of metabolites as well as the environmental, genetic and phenotypic factors affecting the biosynthesis of metabolites has remained marginal considering the vast number of metabolites that have been isolated (Paul et al. 2006; Paul and Ritson-Williams 2008). It is expected that metabolomics will provide useful information by identifying the metabolites that vary during the interaction between organisms and assigning a putative function to these metabolites.

Apart from in-depth information in advances in physiological and ecological issues, metabolomics could potentially give some key indications on the way to go for an efficient exploitation of marine resources. Unbiased metabolomic analysis allows the identification of compounds that can be used as biomarkers for toxicity, diseases or the detection of novel molecules with biological activity in big sample sets (Harvey et al. 2015; Viant 2007). The sea offers a vast array of secondary metabolites produced by marine organisms, an unexplored source of potential new drug candidates. These molecules introduce a new set of molecular features that greatly differ from compounds produced by terrestrial organisms and have thus been used as leads to develop new anticancer or antiretroviral drugs and painkillers (Gerwick and Moore 2012). For this reason, an unbiased metabolomics approach is essential to provide an extensive overview of the metabolism of marine organisms facilitating the process of discovery of new compounds that can be used for the development of novel drugs.

Despite the success of the application of metabolomics to marine organisms, its implementation has suffered a number of practical setbacks that have thwarted the development of its full potential. The collection of significant amounts of sample is often complicated by an extremely low availability, not only due to scarce natural abundance but to the difficulties encountered when accessing the collection sites and/or in their transportation in appropriate conditions from these remote collection locations to laboratories. Another challenging aspect for the experimental design is marine organism chemistry itself. Marine organism metabolomes are characterized by the massive predominance of salt and lipids, requiring specific sample pre-treatment to remove these and obtain extracts that comply with analytical requirements and can provide data sets suitable for metabolomic analysis.

Marine organism metabolomics can lead to the discovery of many bioactive compounds and an increased understanding of the interaction between certain organisms and their environment. In the last few years, a number of technological advances in existing analytical tools or the development of new related ones have contributed to pave the way for exciting applications for metabolomics in this field. This paper reviews all the required steps in the workflow of marine organism metabolomics based on recent studies: sample preparation including extraction, data analysis, and identification of metabolites. It also provides examples of the application of metabolomics to a large diversity of organisms, and the discussion of scientific and technological advances.

2. Sample preparation and extraction: how to maximize the number of intact metabolites in a simple manner

The complexity of marine metabolomics is intrinsic to both the organisms and the environment that they inhabit. Marine organisms establish tight relationships in order to survive the harsh conditions of the marine environment. Thus, when designing an experiment, it is important to bear in mind that this complexity implies that access to marine environments is much harder, requiring more complicated logistics to collect reliable sample sets. Marine ecosystems also differ from terrestrial ones in key features and salinity, water currents, depths, etc. must all be considered apart from common ones such as temperature, light, oxygen concentration, for example, among the significant environmental conditions.

Once samples are collected, i.e., separated from their natural state, it is generally highly recommended to quench the metabolism immediately, as the collection or harvesting process can *per se* rapidly produce changes in their metabolism. Many protocols, designed in general for plant or animal tissues, indicate flash freezing the samples using liquid nitrogen (Kim et al. 2010). This, however, is often impossible in the case of marine organisms due to practical constraints related to the installations and available technical equipment at collection locations which are often very distant from a laboratory or similar facilities. Existing alternatives include freezing at -20° using dry ice or adding organic solvents that can stop or reduce metabolic reactions. Representative quenching techniques that have been used in marine organism metabolomic studies are listed in Table 2.1.

Sample treatment of the collected samples involves the extraction of metabolites and eventually the removal of impurities that may be incompatible with the analytical method to be used. The selection of the extraction method, includes the choice of a solvent and conditions, is driven by two basic aims: the extraction of as many metabolites as possible and ensuring their integrity by the elimination or at least reduction of time-consuming steps and manipulation that can lead to the degradation or modifications in the chemical structure of the metabolites. Similarly to common protocols used in plant natural products studies, organic solvents such as methanol and ethanol or their aqueous mixtures are used, but the selection of solvents varies according to the predominant or targeted metabolites in the studied organism (Ebada et al. 2008). The efficiency of some extraction systems or related technology and conditions (e.g. temperature and pressure), have been evaluated using algae (Heavisides et al. 2018) and sponges (Bayona et al. 2018) as examples. In these cases, extraction conditions for automatic pressure-assisted extractions systems were evaluated to optimize yields of certain groups of metabolites or the chemical diversity of the extracts. Factors such as

temperature, solvent polarity and the number of cycles were found to greatly influence the diversity of chemical composition of the extracts. However, it is still too early to propose general guidelines for optimum conditions for the extraction of a wider range of marine organism metabolites.

Table 2.1: Examples of quenching methods applied in marine organism metabolomics.

Quenching method	Organisms
Centrifugation of the liquid culture media and filtration	Bacteria (Forner et al. 2013; Hou et al. 2012; Viegelmann et al. 2014).
Addition of solvent and freezing at -20 °C	Cyanobacteria (Luzzatto-Knaan et al. 2017; Winnikoff et al. 2014).
Freeze-drying after harvesting	Cyanobacteria (Kleigrewe et al. 2015)
Addition of solvent and centrifugation	Bacteria (Bose et al. 2015)
Frozen and stored at -20 °C	Sponge (Ali et al. 2013; Olsen et al. 2016; Reverter et al. 2018) Corals (He et al. 2014)
Snap frozen using liquid nitrogen	Coral (Farag et al. 2016; Sogin et al. 2014, 2016a) Bacteria (Boroujerdi et al. 2009)

As mentioned, marine organism samples usually contain high amounts of salts and lipids. The removal of salt from their extracts is essential since it interferes with most analytical procedures used for data acquisition. For example, samples with high salt concentration are incompatible with common analytical methods such as LC-MS, HPLC or HPTLC. One of the most common sample pre-treatment methods for desalting marine organism extracts is solid phase purification with Diaion HP-20 resins preequilibrated with methanol (Houssen and Jaspars 2012), C18 and PS-DVB SPE cartridges (Cutignano et al. 2015; Ivanišević et al. 2011). Unless specifically targeted, lipids must also be removed since especially in chromatographic analysis, they build up in lipophilic stationary phases used most commonly, causing serious reproducibility problems. This can be avoided by defatting the extract on Sephadex LH-20 with a mobile phase of methanol and dichloromethane (1:1) or C18 SPE cartridges that retain these highly lipophilic compounds (Houssen and Jaspars 2012).

3. Analytical Methods: to detect a variety of intact metabolites in a simple and robust manner

Metabolomics studies aim to detect all the metabolites present in an organism under a given set of conditions. Nevertheless, due to the limitations of each analytical platform, it is now accepted that no single method can provide a complete picture of the metabolome. In the case of marine organisms, the limitation is more critical given the vast diversity of the metabolites. Furthermore, there are relatively fewer identified metabolites to date. This difficulty has been quite acceptably solved with the increase in available single or combined spectroscopic or spectrometric and chromatographic techniques. In this section, the potential and limitations of each analytical method are discussed with various examples.

3.1 Nuclear Magnetic Resonance spectroscopy: a way of visualizing the overall metabolome

NMR spectroscopy is a widely used analytical technique, mainly for structure elucidation. Since the emergence of metabolomics, NMR, mainly ^{1}H-NMR, has been one of the most common analytical platforms used for data acquisition together with MS-based methods. Several advantages are hard-wired into NMR-based metabolomics: its high signal robustness, ease of quantification of detected metabolites and most importantly, the structural information provided directly by the spectra themselves. In the past, the evident edge of ^{1}H-NMR over other analytical platforms for metabolomics resulted in its application to the study of a wide range of organisms such as mussels (Wu and Wang 2010), shrimps (Schock et al. 2013), sponges (Ali et al. 2013), corals (Sogin et al. 2014) and various fish (Cappello et al. 2018; Melis et al. 2017). In these applications, NMR-based methods showed strong advantages over other methods, particularly in monitoring the levels of primary metabolites submitted to changes in environmental, biological or ecological conditions. Overall, NMR allowed the identification of metabolites present in high concentrations such as amino acids, ketone bodies, organic acids and the energy-related compounds ATP and carbohydrates, which could piece together a metabolic picture of primary metabolism. Notwithstanding its strength in profiling highly concentrated metabolites, ^{1}H-NMR is severely limited when it comes to the detection and identification of the less abundant secondary metabolites. This occurs due to its extremely low sensitivity, the congestion of signals in the spectra (Kim et al. 2010; Viant 2007), the difficulty in the identification of minor compounds in databases or the concurrence of one or more of these situations.

There have been many attempts to compensate for the weak points of ^{1}H-NMR. The information lost due to densely overlapping peaks in ^{1}H-NMR spectra has been partially compensated using the ^{13}C- and 2D NMR spectra. Up to now, J-resolved (Ludwig and Viant

2010), HSQC (Bingol et al. 2014) and ^{13}C-NMR (Clendinen et al. 2014) have been used in metabolomics studies. Conventionally, 2D NMR has been used to obtain additional information on the structural connectivity of molecules which is crucial for their structural elucidation. When applied to mixture analysis for profiling, 2D NMR experiments such as heteronuclear single quantum coherence spectroscopy (HSQC) and J-resolved have proved to be valuable tools for metabolomics applications. Additionally, the projection of some 2D NMR spectra such as J-resolved can be useful to reduce the signal congestion observed in ^{1}H-NMR as the multiplicity of the proton signals is removed (Verpoorte et al. 2007). However, in these experiments one of the main advantages of ^{1}H-NMR, which is the proportionality of the signals to the molar concentration of compounds, is lost, thus resulting in an information vs quantification trade-off.

Despite the limitations of the multidimensional NMR methods, such as longer acquisition times, larger file size and problems with accurate quantification, there have been some interesting advances in the use of these platforms in metabolomics analysis. Recently, improvements in J-resolved pulse experiments have allowed a sensitivity similar to ^{1}H-NMR in close to 20 minutes, making it possible to use 2D J-resolved experiments for metabolomics analysis (Ludwig and Viant 2010). A J-resolved experiment follows a workflow similar to that of ^{1}H-NMR experiments. It includes the removal of regions of the spectra that are not useful (e.g., a solvent signal), as well as normalization, scaling and bucketing to obtain the matrix used for multivariate data analysis. For the scaling step a generalized logarithm (glog) transformation has been specifically developed for the processing of J-resolved spectra. The efficacy of this method for metabolomic studies was confirmed by its use for the classification of mussel samples according to the environmental conditions to which they were exposed (Parsons et al. 2007).

As the backbone of all organic compounds is composed of carbon atoms, ^{13}C-NMR spectra also contain detailed structural information of molecules. However, due to the low natural abundance of ^{13}C and long T2 relaxation times as well as non-quantitative features in the common broadband decoupling method, it has been less used in metabolomics. In the past years, a workflow using ^{13}C-NMR spectra at natural abundance was developed by Clendinen and co-workers (Clendinen et al. 2014). With the use of an optimized ^{13}C probe, the acquisition times for mixtures were reduced to just 2 hours. With this methodology, it was possible to improve the differentiation of groups and increase the number of metabolites identified from the mixture. To date however, this method has not been applied to the metabolomics of

marine organisms studies but could nevertheless be considered as a complementary tool for metabolic studies.

3.2 Liquid chromatography-based methods with mass spectrometer detectors: A closer look at minor metabolites

Liquid chromatography hyphenated to mass spectrometry is the most commonly used technique in marine organism metabolomics. Among its several strong points is the separation provided by the chromatographical step, which also provides additional chemical information as metabolites can be distinguished by their retention time as well as by their m/z value. Moreover, the separation step avoids saturation of the ionization system, allowing more metabolites to be ionized and detected. In LC-MS analysis, hundreds of thousands of metabolites can be detected including some that are present at trace levels (Goulitquer et al. 2012). This higher sensitivity is ideal for the detection of bioactive and ecologically relevant compounds, since in many cases these are present in very small quantities (Belarbi et al. 2003).

The term LC-MS refers generically to all hyphenated systems consisting of a liquid chromatograph connected to a mass spectrometer detector. However, there are diverse types of liquid chromatographs and mass detectors. Currently, the chromatographic instrument in most LC-MS setups is an ultrahigh performance liquid chromatography system (UHPLC). This system reduces analysis time without losing resolution as compared to conventional high-performance liquid chromatography (HPLC) systems. The most common stationary phase used in LC is C18, but in recent times new normal phases such as HILIC have been developed, extending the range of metabolites that can be efficiently analyzed to include very hydrophilic and ionic compound mixtures which cannot be resolved using C18 (Belghit et al. 2017; Borras et al. 2017). Different types of mass detectors have been coupled to chromatography systems for marine metabolomics studies. Among the most commonly used are triple quadrupole (QQQ), quadrupole/ time of flight (Q-TOF), ion trap (IT) and Orbitrap. The potential and drawbacks of each system have been widely discussed in the literature (Allwood and Goodacre 2010; Forcisi et al. 2013; Lei et al. 2011). In general, Orbitrap has a higher resolution, but when it comes to quantification, QQQ and Q-TOF systems are expected to provide higher accuracy. All these MS instruments can be used for MS^n experiments, but ion trap is considered to be the most powerful, generating robust fragmentation patterns which are essential for structure elucidation.

Despite its popularity in the field of metabolomics, one of the main drawbacks of LC-MS-based studies is the complicated data mining process given the size of the generated data sets. In

practice, molecules cannot be identified simply on the basis of their molecular mass. Even with high-resolution molecular mass data, reliable fragmentation patterns are needed to confirm the molecular structure. Thus, target compounds must be submitted to tandem MS, i.e., MS^n experiments, to generate fragmentation patterns which can then be matched with a reference spectrum, much like a fingerprint. Furthermore, recent development of the GNPS (Global Natural Products Social network) platform which uses MS/MS data has enabled the deduction of structural similarities between metabolites, allowing them to be grouped accordingly (Wang et al. 2016). The detailed procedure and potential of this platform are discussed in the data analysis section.

3.3 Gas chromatography-based-mass detector hyphenated methods: Detailed metabolic analysis for primary metabolites.

Gas chromatography hyphenated to mass spectrometry is one of the earliest analytical platforms used in metabolomics studies. Even today, among all the MS-based techniques, GC-MS is still considered to offer the highest resolution, reproducibility and robustness. These two last features of the fragmentation patterns of compounds generated with electronic impact ionization (EI), the most common ionizer in GC-MS, makes their identification by matching their spectra to data libraries relatively easy (Kopka 2006). In marine metabolomics however, GC-MS has been used less frequently, mainly because GC analysis requires the metabolites to be thermostable and volatile, limiting the number and kind of compounds that can be analyzed directly using this method. This is overcome by the derivatization of samples (Sumner et al. 2003), but even with the improvements made in derivatization methods, this is still a limitation to the application of GC-MS in metabolomics and it has thus been mainly used for the detailed analysis of primary metabolism, especially for amino acids and organic acids.

Although the number of reports of metabolomics studies using GC-MS is notably lower than NMR or LC-MS, some examples show the potential of this technique for marine organism metabolomics. The GC-MS analysis of scallops (*Chlamys farreri*), an important organism in the food industry, was used to evaluate modifications in metabolic pathways related to their survival rates in different preservation conditions. The presence of metabolites indicative of a trade-off between aerobic and anaerobic metabolism detected in live scallops preserved in semi-anhydrous conditions provided essential information for the marketing of scallops, due to its direct incidence on the final value of the product (Chen et al. 2015). The metabolomes of other organisms, such as starfish and marine polychaeta have also been investigated. These studies focused on establishing the best extraction and analytical conditions for the optimized detection of metabolites. In both examples, GC-MS analysis provided detailed information of

the primary metabolites such as amino acids, lipids, and carbohydrates in the samples (Fernández-Varela et al. 2015; Pereira et al. 2012).

Along with primary metabolites, volatiles have historically been the most frequent targets of GC-MS analysis. In the past, volatiles were suspected to play various key physiological roles. In recent decades, they have also come to be considered interesting vectors in the interactions between organisms given their unique physical characteristics and high diffusion rate. Recently, *in situ* extraction methods such as headspace extraction and solid phase microextraction (SPME) have been applied to the analysis of volatile compounds present in marine organisms, such as algae (de Alencar et al. 2017; Jerković et al. 2018) as well as in some marine microorganisms (Barra et al. 2017; Salvatore et al. 2018). The use of *in situ* methodologies for the extraction and further analysis of this kind of compounds can provide additional information in metabolomics studies.

3.4 Direct infusion mass spectrometry: Getting a picture of the metabolome

Although less popular, MS-based metabolomics can be performed with data generated directly by the MS-spectrometer. A wide range of direct infusion methods are used in this case, all of which have their own application fields in metabolomics.

In studies involving marine organisms, the amount of available sample tends to be extremely low and this is aggravated by the fact that they require some preparation, starting with extraction. Ultimately, no extraction would be the ideal method for metabolomics studies. Nowadays, this is possible if using direct infusion MS techniques such as matrix-assisted laser desorption/ionization (MALDI) imaging, direct analysis in real time (DART) and desorption electrospray ionization (DESI) all of which offer the possibility of analyzing samples without an extraction step (Esquenazi et al. 2009; Parrot et al. 2018). In marine organisms, MALDI-TOF imaging has been used to detect metabolites such as viridamides, jamaicamides and curacin in filaments of marine cyanobacteria and to determine their location within sponge cells (Esquenazi et al. 2008). This study is of particular interest because while MALDI-TOF has been generally applied to analyze macromolecules, this application proves that MALDI-TOF imaging is capable of detecting low molecular mass molecules despite the interference of the matrix, eventually providing valuable information on the spatial distribution of secondary metabolites in the tissues of marine organisms.

Notwithstanding their evident value, the widespread implementation of mass imaging techniques has had its setbacks: the lack of reproducibility between samples, the difficulty in sample preparation, the low resolution of the mass measurements and complexity of data

processing. These limitations, however, have not diminished its potential as a powerful approach for studying ecological interactions involving for example surface-related phenomena and spatial location of metabolites. Recently, the combination of LC-MS based metabolomics with mass imaging has been used to study the chemical defense mechanism of eelgrass against microbial foulers (Papazian et al. 2019). In this study, untargeted metabolomics of the marine plant *Zostera marina* revealed different concentrations of a group of fatty acids and phenolic compounds in a leaf surface extract and the whole leaf. These compounds were targeted in a DESI-MS imaging analysis, showing that some fatty acids were more abundant on the surface of the leaf while phenolic compounds were more abundant within the leaf. With this information a defense mechanism of *Zostera marina* against microbial foulers was proposed. This is an example of the potential of direct infusion-mass imaging methods as an approach for marine organism metabolomics.

3.5 Other analytical platforms: To complete the metabolome puzzle

There is no single method that can cover the full metabolome of an organism, despite the increasing analytical range achieved over the past years. It has therefore become increasingly important to develop novel analytical tools to complement the information obtained by current techniques. In this context, SFC coupled to mass spectrometry has been used in metabolomic studies, aiming to increase the metabolic coverage and decrease the time of analysis (Shulaev and Isaac 2018). The particular physicochemical characteristics of the mobile phase, usually supercritical CO_2, which features the advantages of a gas and a liquid, gives SFC a niche in the field of metabolomics. Using a simple gaseous-like mobile phase, SFC fills the gap between LC- and GC-based methods allowing the analysis of compounds of a wider range of polarity, with shorter run times (mostly within 10 min), and with very high resolution and reproducibility between analyses without the need to volatilize the samples. Despite its potential, one of the main snags in the use of SFC has been the availability of a stationary phase that could cover a wide range of metabolites in a single analysis. Recent studies using plant extracts have shown that stationary phases such a DIOL and C18 are suitable for the analysis of crude extracts, resolving 88% of the compounds present in the mixture (Grand-Guillaume Perrenoud et al. 2016). Another recently used chromatographic technique in the field of metabolomics is capillary electrophoresis (CE) coupled to mass spectrometry. The developments in CE have allowed the detection of up to 900 molecular features in a single run, making this technique ideal for metabolomics studies (Zhang W. et al. 2017). However, despite the remarkable sensitivity of CE for specific metabolic groups, variation in the migration time has been one of the main obstacles to its development as a robust

metabolomics analysis tool. There have been successful efforts to overcome this issue (Hirayama et al. 2014) and though there are no reports to our knowledge of marine metabolomics studies using CE-MS, it is likely that in the foreseeable future, both SFC-MS and CE-MS may be used as supplementary metabolomics tools, increasing the chemical space.

Infrared spectroscopy has been considered a very versatile, inexpensive and robust tool for the analysis of samples. However, the use of FT-IR in metabolomics has been confined to fingerprinting studies due to its limitations in the identification or dereplication of metabolites. This can be appreciated in the study of the effect of the exposure of the green algae *Ulva lactuca* to different concentrations of oil fuels (Pilatti et al. 2017). This showed FT-IR to be a reliable tool for fingerprinting or footprinting applied to the detection of changes induced by marine organism exposure to fossil fuels, even at low concentrations. Another example of the use of FT-IR for fingerprinting is a study performed with the mussels *Mytilus edulis* (Gidman et al. 2007). In this case, FT-IR was used to observe chemical changes in the hemolymph caused by environmental stress.

Thin layer chromatography (TLC) was among the first separation techniques used in analytical chemistry and it is still being applied for the rapid evaluation of chemical profiles. Recent improvements of this technique aimed at increasing its relatively low resolution, have resulted in the development of high-performance thin-layer chromatography (HPTLC). This renewed technique has recently been applied to metabolic fingerprinting, mainly in plant metabolomic studies (Bayona et al. 2019). Several other studies had revealed the potential of HPTLC for data acquisition in marine organism metabolomics, e.g. sponges and algae. In these cases, the possibility of performing *in situ* bioactivity tests provided by HPTLC separation allowed the rapid recognition of bioactive metabolites. Moreover, the integration of HPTLC with a MS-detector can facilitate the identification of compounds, providing, eventually, guidance for preparative separation work.

4. Data analysis and multivariate data analysis: to extract hidden information from a raw data matrix

The data acquired using any of the described analytical platforms is generally massive and must be profiled in an untargeted manner. This is a key step in any metabolomics study. Data processing involves the conversion of raw spectroscopic and chromatographic data to a numerical matrix which can be used for statistical or multivariate data analysis. It is an essential step for the extraction of hidden valuable information from the bulky data sets and normally requires several steps depending on the source (analytical method) of the data (Figure 2.1 and

2.2). The preprocessing of NMR data is somewhat simpler than that required for other analytical tools and consists of phase correction, baseline adjustment, shift adjustment, binning (also

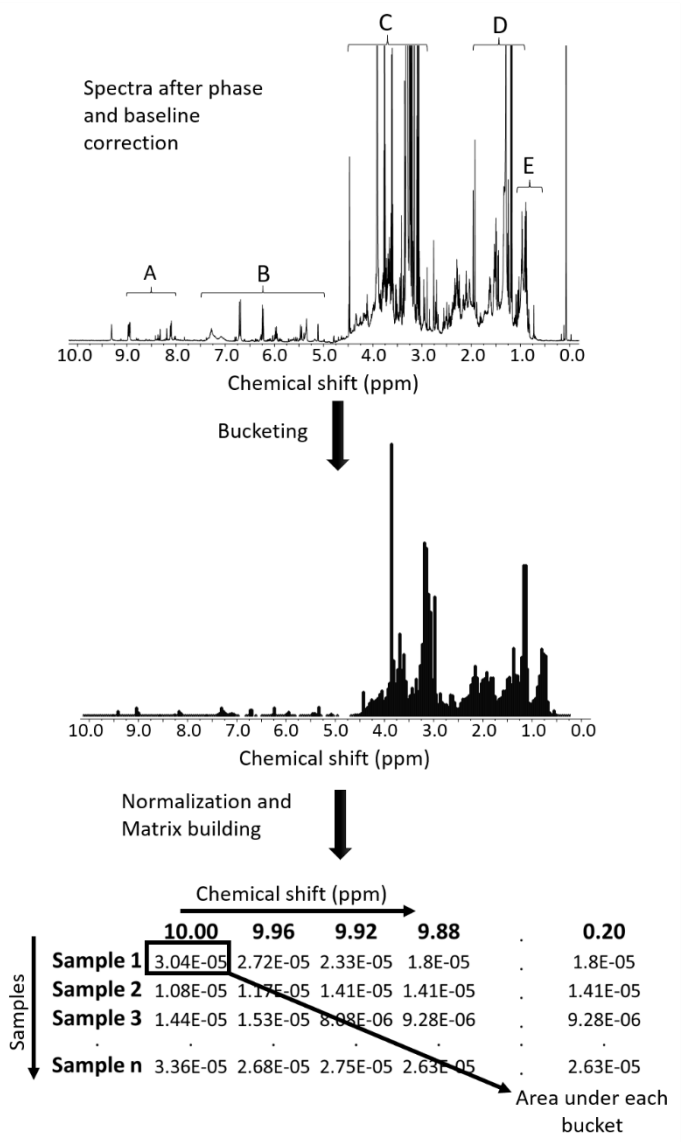

Figure 2.1: Schematic diagram of the workflow to build a data matrix of ^{1}H-NMR spectra using an example of the marine sponge *Xestospongia* sp. in CH_3OH-d_4. The regions pointed in the NMR spectra correspond to A) aromatic quinones, B) unsaturated brominated fatty acids, C) hydroxylated fatty acids and glycerol derivates D) methylene groups E) Sterols

and the flexibility of the binning range have a substantial impact on the final results. Ideally, a smaller binning size should result in a more precise representation of the original spectrum, but this can cause the shifting of signals which eventually may result in a loss of important information. In the past, the smallest binning size was suggested to be between 0.01-0.04 ppm for 400 - 600 MHz NMR as long as no further alignment was applied to. However, the signal robustness gained from binning is a trade-off for resolution that is decreased compared to the raw data (Figure 2.1). To overcome this dilemma, many alternative algorithms have been developed to improve the binning procedure or the alignment of the spectra, e.g. dynamic adaptative binning (Anderson et al. 2011), fuzzy warping (Wu et al. 2006) and peak alignment using a genetic algorithm (Forshed et al. 2003).

Compared to the simplicity of the NMR dataset, LC-MS datasets are inherently more complex due to the multi-dimensional features provided by the response of analytes to both the chromatographic separation and mass spectrometry. Moreover, the higher sensitivity of this method results in large data sets that can be more challenging to work with. Several software programs are available for data processing, some of which are open access, such as MZMine 2 (Pluskal et al. 2010), OpenMS (Röst et al. 2016), XCMS (Smith et al. 2006), and MS-DIAL (Tsugawa et al. 2015). Although each software uses its own algorithms with different steps, they share the goal of creating a matrix that contains the intensity of all molecular features in each sample and removes possible technical variations (Figure 2.2). Alignment of the molecular features in both retention time and mass values is seen as the most essential part of LC-MS data processing, and variability in retention times, mass values and isotopic patterns must all be considered.

Regardless of the analytical methods used for metabolomics analyses, a huge number of variables will always be generated. This can range from several hundred up to tens of thousands of variables. Therefore, the first important step in data mining is the reduction of data dimensionality to extract hidden information from the raw data matrix. Multivariate data analysis (MVDA) is usually divided into two categories: unsupervised and supervised methods. For unsupervised MVDA, many approaches have been developed to reduce the dimensionality of the data, out of which principal component analysis (PCA) is currently the most commonly applied in the field of metabolomics. The aim of PCA is to collect most of the variance present in a data set using new orthogonal variables, known as principal components (Worley and Powers 2013). By analyzing the first few components of the PCA, it is possible to observe

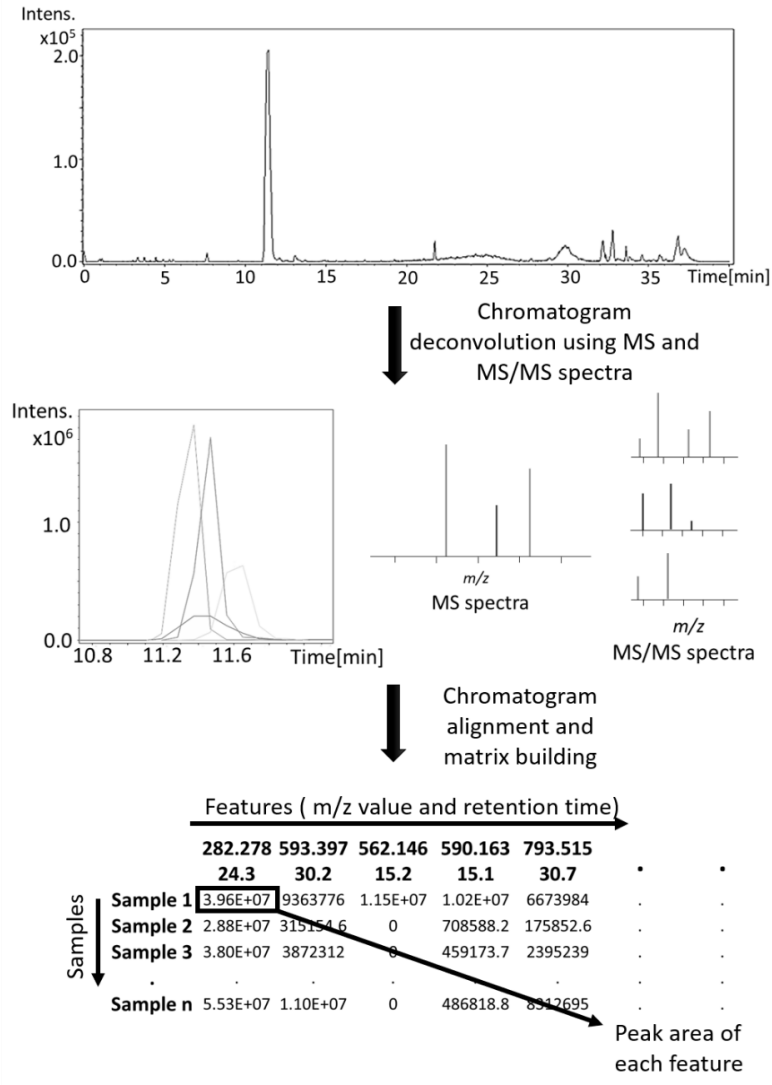

Figure 2.2: Schematic diagram of the workflow to build a matrix of LC-MS data, as an example the chromatogram of a sample from the marine sponge *Xestospongia* sp. The chromatograms deconvolution is done using both MS and MS/MS data. Followed by the alignment of the features present in each sample using m/z values and retention time criteria.

whether samples that belong to a specific class are grouped together. The metabolites associated with the classes or factors of interest can then be singled out using the loading plots. There are a number of examples of marine organism studies in which PCA was successfully employed: the metabolic discrimination of three species of crabs (*Callinectes sapidus*, *Eriphia verrucosa* and *Cancer pagurus*) (Zotti et al. 2016), the differentiation of

bacterial strains that produce new compounds (Macintyre et al. 2014), and the detection of metabolic changes caused by exposure to steroids in mussels (Cappello et al. 2017).

Similarly to unsupervised methods, there are numerous methods that can be applied for supervised approaches in many biological systems in marine organisms. Typical examples of the applications of various unsupervised methods are shown in Table 2.2 and 2.3. Although there is no universal way to perform an supervised MVDA, currently, partial least square (PLS) modeling, or its variations such as PLS-discriminant analysis (PLS-DA), orthogonal partial least square (OPLS) modeling and OPLS-discriminant analysis (OPLS-DA) are the most common methods applied after completing a PCA. In PLS (for continuous Y-variables) or PLS-DA (for discrete Y variables) the new variable will be calculated to maximize the covariance between the X- data matrix (e.g., NMR, LC-MS, GC-Ms data) and Y-data matrix (e.g., age, distance, specie, and biological activity) (Liland 2011). However, even after the use of PLS modeling, it may be difficult to find patterns in the data set or in the validation of the model due to prevailing variation caused by factors that are not included in the study. To solve this problem, OPLS and OPLS-DA have been developed. These analyses are a rotation of the original solution of the PLS and PLS-DA analyses. Although these models do not represent an improvement in the predictive capacity of the model, the interpretation of the data is more straightforward (Liland 2011). Especially OPLS-DA is being applied to discriminate groups such as active and non-active samples in a data set, after which the metabolites likely to be responsible for the activity are selected using the samples as a guideline.

5. Databases: both for the dereplication of known metabolites and for a repository for future identification

The most efficient way to annotate the metabolites presumably associated with certain treatments is to dereplicate the compounds using available data bases (Blunt et al. 2012). There are two types of databases used in metabolomics studies; general chemical entities and spectral data libraries. The general chemical databases such as PubChem (PubChem, NIH), ChemSpider (ChemSpider, RCS), the CAS registry through SciFinder (SciFinder, CAS), Dictionary of Natural Products (CRC Press), and AntiBase (AntiBase, Wiley) are available online. These databases provide information on nomenclature, molecular formula and mass, source and physicochemical properties of the compounds. Databases such as PubChem and ChemSpider, are non-specialized and cover all kinds of chemical entities regardless of their origin. On the other hand, databases like Dictionary of Natural Products and Antibase focus on compounds isolated from specific natural sources such as plants, animals or microorganisms. In the field of marine natural products, the unique chemical features that characterize compounds

isolated from marine organisms, particularly marine invertebrates, together with the wide chemical space that is covered by marine natural products, demands the use of specialized databases. MarinLit is perhaps the most important database in the marine natural products field containing more than 27500 compounds reported from marine sources (MarinLit, RSC). Another of its advantages is the possibility of including a wide range of features in its search criteria. In particular, compounds can be searched using spectroscopic data that include UV maximum, NMR chemical shifts and the exact mass.

The second type of databases focus on spectral information and are very useful for dereplication in the field of metabolomics. These databases are usually specialized in one analytical technique. NMR databases such as SpecInfo-Consortium Member-NMR, IR and MS (SpecInfo-Consortium Member-NMR, IR and MS, Wiley) and Chenomx contain information on the ^{1}H-NMR and/or ^{13}C-NMR spectra of several thousand compounds. Particularly for ^{1}H-NMR spectra, complete databases are required for full identification, as signals in a specific region of the spectra can be assigned to different compounds or kinds of compounds and detailed information on specific compounds is very helpful. In the case of GC-MS data, the widespread use of electronic impact (EI) as an ionization method generates highly reproducible fragmentation patterns that allow reliable identification using databases. In contrast, LC-MS analysis has produced less reproducible fragmentation data. In the past, analyses focused on mass of adduct ions because of their soft-ionization modes. However, recently the necessity of tandem mass information (MS/MS and MS^{n}) for identification has become evident. Databases such as NIST 20 (NIST 20), MassBank (Horai et al. 2010), and METLIN (Guijas et al. 2018) have available MS/MS spectra that can be used to perform dereplication of compounds, which include different ionization methods such as EI, ESI and MALDI coupled to various mass analyzers. These databases enable the matching of putative compounds using both their molecular mass and the MS/MS spectra. Additional search features, such as fragment losses in the MS/MS spectra can help to find analogs of unknown compounds.

Marine metabolomic studies, as any other field in metabolomics, greatly rely on the information available in databases for the annotation of compounds. The availability of specialized databases such as MarinLit and Antibase is particularly useful for the identification of secondary metabolites. However, the use of generic databases should not be dismissed, as several metabolomics studies of marine organisms have revealed changes in primary metabolism that may not be available yet in specialized libraries. Mass spectrometry databases are continually updated with additions of new information / molecules, making them a powerful tool for the dereplication of compounds, even if the standardization of

experimental conditions is still a distant concept. In the case of NMR metabolomics, it is important to note that there is a need for the inclusion of more NMR spectroscopic data of marine natural products into databases to improve the annotation rate of compounds. Moreover, the *in silico* predictions of NMR spectra, already available in software such as ACD/Labs, could benefit from the addition of more experimental data into the databases. Lastly, with the increase in public spectroscopic data that is now required in metabolomics studies and the inclusion of this information into databases, an increase in the annotation of compounds in marine metabolomics studies can be expected.

6. Statistical total correlation (STOCSY), small molecule accurate recognition technology (SMART) and molecular networking (MN): *in silico* identification of metabolites from complex mixtures

The final aim of any metabolomics study is the identification of metabolites from the signals selected from the MVDA. However, the process between the selection of a signal and the identification of the corresponding metabolites is arduous and time-consuming and usually considered to be the bottleneck in metabolomics studies. A reasonable approach for identification is to use currently available databases as shown in the previous section. However, the limited number of entities in databases (e.g. NMR and MS spectra) was an incentive for the development of *in silico* identification strategies using both NMR and MS based methods. The different nature of the features obtained from NMR and MS data have led to the development of diverse approaches.

One of the NMR-based computational tools used in some marine organism metabolomic studies is statistical total correlation spectroscopy (STOCSY). This method uses the correlation between the intensity of the signals in the ^{1}H-NMR spectra to distinguish which signals belong to the same compound or family of compounds (Figure 2.3a), facilitating the identification of some molecules (Lindon and Nicholson 2008). STOCSY was used to identify compounds such as alanine, trigonelline, threonine and lactate, present in coral extracts, in a study that aimed to distinguish the chemical profiles of reef building corals (Sogin et al. 2014). It was also used to identify the changes in the metabolic profile of breast cancer cells treated with candidate anticancer marine natural products (Bayet-Robert et al. 2010). It was possible to determine that when treated with ascididemin, a compound isolated from the marine tunicate *didemnum* sp., the cells accumulated unusual amounts of gluconic acid.

Other computational tools based on NMR data, combine HSQC spectra with deep convolutional neural networks in a new platform called Small Molecule Accurate Recognition

Technology (SMART) for the dereplication of natural products (Zhang et al. 2017). This platform enables rapid dereplication of molecules into a family of compounds. In combination with metabolomics platforms based on MS data, such as molecular networking, it can lead to the fast and efficient discovery of new molecules. This was the case for viqueamide C, a new cyclopeptide isolated from marine cyanobacteria *Rivularia* sp. and *Moorea producens* (Gerwick 2017). This molecule was highlighted by both molecular networking and SMART approaches as a new compound from the family of viqueamides, leading to its isolation and confirmation of its structure.

Molecular networking (MN) is a metabolomics workflow based on MS-data sets, which has emerged in the last few years (Yang et al. 2013). When it was first introduced, it was used as a dereplication methodology, but this was later extended to the visualization of complex connections between all metabolites, which could reveal metabolic correlations in biosynthesis. The construction of molecular networks is based on the analysis of MS/MS spectra of compounds, assuming that molecules with similar chemical structures should display similar fragmentation patterns. The basic idea of MN data calculation is to give cosine scores to MS/MS spectral similarity: The closer the score is to 1 the higher the similarity between spectra with a threshold usually set at 0.7 (Figure 2.3b) (Quinn et al. 2017). Recently, a new workflow called feature-based molecular networking was developed (Nothias et al. 2020). This network has some advantages compared to traditional Molecular Networking as isomers can be distinguished based on differences in retention times and the quantification of the features in the samples is more accurate, as it uses peak area instead of ion intensity.

The power of MN lies in grouping the metabolites with structural similarities. This has proved to be very useful for metabolomics and particularly in marine organism metabolomics, such as the study of three marine cyanobacteria strains *Trichodesmium erythraeum*, *Okeania* sp, and *Oscillatoria* sp. that allowed the identification of the aplysiatoxins family from the metabolomic pools (Ding et al. 2018). The *T. erythraeum* strains collected in different locations displayed highly characteristic clusters depending on the location, while those from *Okeania* sp. and *Oscillatoria* sp. appeared to be similar even at a distance of 100 m. Ecologically, this investigation is an example of the strong influence of environmental factors on the metabolic production of these cyanobacteria. From a chemical perspective, it provided an example of the possibility of identifying compounds within clusters using the structural similarity to known compounds present in this same cluster (Yang et al. 2013).

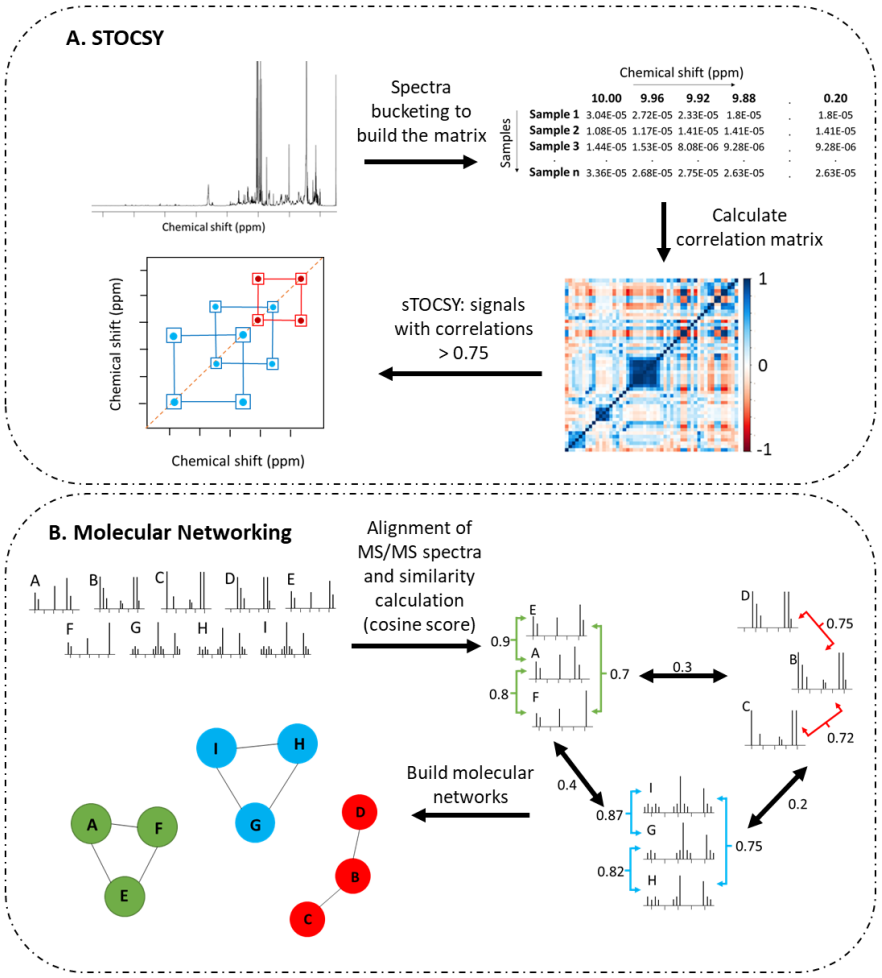

Figure 2.3: Comparison of the workflows of A) STOCSY (statistical total correlation spectroscopy) and B) Molecular networking. A: the correlation matrix shows the correlation values. Those with correlation values above 0.75 are plotted in the spectra. In red are the correlations that are also present in COSY spectra as a result of chemical bonds, and in blue are the correlations between signals that are not chemically bonded B: shows the alignment of MS/MS spectra and calculation of the similarity cosine score, features with scores above 0.7 are connected adapted from (Watrous et al. 2012).

The use of computational tools has increased the amount of information that can be extracted from NMR spectra and consequently the identification of compounds in a mixture. One of the

Even with existing databases, the chance of success in the identification of metabolites would still be low, mainly due to the limitation of data entities, which require time-consuming measurements. To circumvent this limitation, *in silico* dereplication tools, especially *in silico*

MS fragmentation, appeared a few years ago as a solution to the lack of experimental information. Following the proposal of the Wolfender group to use *in silico* dereplication together with molecular networks (Allard et al. 2016), many researchers have followed suit and developed platforms with this purpose. Platforms such as network annotation propagator NAP (da Silva et al. 2018), MS2LDA (van der Hooft et al. 2016) and MolNEtEnhancer (Ernst et al. 2019) have all been implemented in the GNPS webpage, to be used together with molecular networks for the identification of molecules or families of molecules.

7. Metabolomics applications to marine organisms: biological, ecological, pharmaceutical and physiological studies

7.1 Marine natural products and discovery of new bioactive compounds

The structural novelty of marine natural products shows great potential, as they can provide a multitude of leads for the development of novel drugs. There are, however, a few critical limitations for the isolation of natural products from marine organisms: the structural complexity of their metabolites, the extremely low concentration of active molecules and the limited amount of source materials. Considering that these organisms cannot be easily obtained simply by cultivating or farming, the only point that can be optimized is the deconvolution of chemical structures from small amounts of complex mixtures. Therefore, any successful project for drug discovery from marine sources must adopt an approach that can speed up identification of bioactive molecules in these relatively adverse circumstances.

Metabolomics has been increasingly used in recent years for the study of marine natural products as a new tool for metabolic mining. Metabolomics platforms have improved the ability to prioritize which samples within a sample set have either novel chemical structures or possible bioactive compounds. Moreover, it is even possible to target putative active compounds before starting the isolation process. This pre-profiling procedure based on metabolomics saves time and effort, highlighting valuable sources of new structures and activities. In general, the procedure for the discovery of bioactive molecules using metabolomics in marine organisms consists of metabolic profiling and bioactivity testing, correlation of features by MVDA, selection of signals responsible for positive correlations and identification of the metabolites corresponding to the signal. Table 2.2 lists applications of metabolomics platforms for the discovery of new molecules from various marine organisms.

Table 2.2: Examples of application of metabolomics for the discovery of new bioactive compounds

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Bacteria *Vibrio* sp. QWI-06	UHPLC-ESI-Orbitrap-MS	Molecular networking	no	Discovery of vitroprocines A-J active against *Acinetobacter baumannii.*	(Liaw et al. 2015)
Cyanobacteria *Moorea bouillonii* and *Moorea producens*	HPLC-ESI-Q-TOF-MS	Molecular networking	no	Discovery of three new metabolites columbamides A-C.	(Kleigrewe et al. 2015)
Cyanobacteria *Moorea producens* JHB	LC-ESI-LTQ-FTICR-MS	Molecular networking	no	Discovery of new compounds of the Jamaicamide and Hectochlorins family.	(Boudreau et al. 2015)
Bacteria *Salinispora arenícola*	UHPLC-ESI-Q-TOF-MS	PCA	OPLS-DA	The incubation time and salinity of the culture media of *Salinispora arenicola* change the quantity and type of Rifamycins produced.	(Bose et al. 2015)
Fungi *Aspergillus terreus*	LC-ESI-TOF-MS	PCA	no	Using diverse cultivation media lead to the discovery of the new compound 7-desmethylcitreoviridin.	(Adpressa and Loesgen 2016)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Bacteria 1000 marine microorganisms	UHPLC-ESI-Q-TOF-MS	Molecular networking and PCoA[a]	no	Increment in the chemical space by using diverse extraction methods and the discovery of two compounds: maridric acids A and B.	(Floros et al. 2016)
Sponge *Geodia macandrewii* and *Geodia baretti*	UHPLC-ESI-TOF-MS	PCA	OPLS-DA	The differentiation of two sponges of the genera *Geodia* and the isolation of the new compound geodiataurine.	(Olsen et al. 2016)
Bacteria *Micromonospora* sp and *Rhodococcus* sp	UHPLC-ESI-Q-TOF-MS	PCA	no	Discovery of keycin, a new antibiotic from the co-culture of the bacteria.	(Adnani et al. 2017)
317 marine cyanobacteria and benthic algae	UHPLC-ESI-Q-TOF-MS	Molecular networking and PCoA[a]	no	Differences in the chemical composition of cyanobacteria collected in a different location giving and the discovery of yuvalamine A, a new compound.	(Luzzatto-Knaan et al. 2017)
Bacteria *Streptomyces* sp. WU20	HPLC-ESI-QQQ-MS	no	PLS	The cultivation of *Streptomyces* sp WU 20 under nickel stress lead to the discovery of a new compound with antibacterial activity.	(Shi et al. 2017)
Bacteria 24 *actinobacteria* like strains	HPLC-ESI-IT-MS	HCA	OPLS-DA	24 strains were group depending on their metabolic production. Putative active compounds in quorum sensing assays were selected and dereplicated.	(Betancur et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Sponge *Spongia officinalis*	HPLC-ESI-Q-TOF-MS	PCA and molecular networking	PLS-DA	Composition of furanoterpenes in *S. officinalis* is changed depending on the geographical location and season. Three new compounds: Furofficin and Spongialactam A and B were isolated.	(Bauvais et al. 2017)
Fungi 21 isolated fungi	UHPLC-ESI-Q-TOF-MS	Molecular networking	PLS-DA	The co-culture of isolated fungi with phytopathogenic bacteria and fungi trigger the production putative active secondary metabolites. One new putative peptide from the emerimicin family was annotated.	(Oppong-Danquah et al. 2018)
Alga *Fucus vesiculosus*	UHPLC-ESI-Q-TOF-MS	Molecular networking	no	Seasonal changes in the metabolome the algae were observed and the variation in the concentration of some metabolites was related to changes in the bioactivity of the alga extracts.	(Heavisides et al. 2018)

[a] Principal coordinate analysis

7.2 Biological and Ecological Applications

The discovery of new molecules from marine organisms focusses mainly on sedentary organisms such as sponges, corals, algae, and microorganisms. In contrast, biological or ecological studies pose a greater variety of research questions and include wide variety of animals such as fish, mussels, and turtles. The diversity of organisms in these studies results in a great heterogenicity in the type of samples. The preanalytical and analytical protocols confined to sponges, corals, and other sessile organisms might be similar to the ones used in general drug discovery studies. However, for other organisms such as, fish, mussels and marine mammals, samples from organs, tissues and biofluids require preanalytical treatment similar to the ones used in human metabolomics studies.

The investigation of the effect of environmental changes in the metabolome of marine organisms resulting from human activities has been growing in recent years. Currently, many industries produce residues such as pesticides, antibiotics, hormones, and heavy metals that end up polluting the oceans. In the past few years, a number of studies concerning the effect of artificial chemical residues on the metabolome of a wide range of marine organisms have been actively carried out. Studies on the connection between environmental changes and metabolomes have been well reviewed (Bundy et al. 2008). Additionally, reports of studies that apply metabolomics to explain physiological processes of marine organisms such as mating, predator defense, and growth development, all of which are likely mediated by chemical compounds, have been growing gradually, especially related to sessile animals. A summary of recent applications of metabolomics to biological and environmental sciences are shown in Table 2.3.

Lastly, one of the central and most substantial issues in marine biology is taxonomical classification. Many of the traditionally used taxonomic characteristics are not useful for the classification of marine organisms and the available knowledge on genetic markers is still scarce. Classical characterization has in many cases resulted in misclassifications of many species, generating a constant creation and reclassification of species. The use of metabolomics in this field could provide an additional taxonomical marker or at least a holistic overview of the metabolome, which is unlikely to be achieved with conventional methods. For example, metabolic fingerprinting was implemented to distinguish two morphotypes of the zoanthid, *Parazoanthus axinellae,* collected in different locations of the Mediterranean Sea. This resulted in the detection of taxonomic marker type metabolites: ecdysteroids, zoanthoxanthins, and parazoanthines (Cachet et al. 2015). Parazoanthines were found to be

present only in the "slender" morphotype during the whole year but not in the other morphotype. This result endorsed the revision of the classification of *Parazoanthus axinellae.*

8. Summary and perspectives

Although knowledge of marine organism and their ecosystems has been rapidly increasing over the past decades, with more species and thousands of new molecules reported every year, the scale of unexplored organisms is thought to be astounding. Moreover, even with the developments in technologies such as SCUBA diving, ROV's (Remotely operated vehicles) and research submersibles, which allow a study of the environment that could not even be imagined in the past, the study of marine organisms remains an arduous task. The application of metabolomics in marine organism studies provides a new approach for the discovery of compounds that can be used for the benefit of humans in different ways. Perhaps even more importantly, it has increased the understanding of the function that all these metabolites have within the producing organism as well as the interactions of these organisms with their environment.

Recent studies have revealed the association of microorganisms to be the real synthesizers of many of the secondary metabolites isolated from marine invertebrates. This has prompted the study of marine associated microorganisms, where new active novel compounds can be discovered using metabolomics as a criterion to select strains for the study. The main focus of marine chemical ecology and environmental metabolomics is the study of interactions of marine organisms with their environment and with other organisms. There is particular interest in the study of the effect of modifications in environmental conditions resulting from human activities. Within this field, metabolomics-based analyses could be expected to lead to increased understanding of the interactions between some organisms and their symbionts, or biosynthetic pathways of compounds and how these are transferred between organisms. Another interesting topic to study is the biotransformation of secondary metabolites along the trophic chain, since some metabolites are known to originate through this complex mechanism. In this case, metabolomics studies together with biosynthetic pathway studies can lead to the discovery of how compounds are transformed.

The ultimate goal of metabolomics, that is, acquiring a picture of the whole metabolome of an organism is still a dream, as none of the existing analytical platforms can detect all metabolites in a single analysis. The addition of new analytical tools, such as CE-MS, SFC-MS, HPTLC and IR- based metabolomics, is a step in this direction but even then, the use of more analytical platforms does not necessarily guarantee more useful information. Rather, it is

Table 2.3: Examples of the application of metabolomics to environmental and biological studies of marine organisms

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Coral *Pocillopora damicornis*	UHPLC-ESI-Q-TOF-MS and GC-EI-TOF-MS	no	OPLS-DA	Exposure to different conditions of temperature and pCO_2 shift metabolic pathways including carbohydrate metabolism, cell structural maintenance, defense mechanisms among others.	(Sogin et al. 2016)
Coral *Sarcophyton* spp., *Lobophytum pauciflorum*, and *Sinularia polydactyla*	UHPLC-ESI-LCQ-MS and ^{1}H-NMR	PCA	OPLS	Coral growth in the wild exhibit higher levels of cembranoids, the most common group of diterpenes reported for soft corals, while corals growing in an aquarium have a higher content of oxylipins.	(Farag et al. 2016)
Bacteria *Persicivirga* (Nonlabens) *mediterranea* TC4 and TC7 *Pseudoalteromonas lipolytica* TC8 and *Shewanella* sp. TC11	UHPLC–ESI-Q-TOF-MS	PCA and molecular networking	PLS-DA	Different culture conditions are reflected in the metabolome of the four bacteria studied. Compounds of the family of hydroxylated ornithine lipids, diamine lipids and glycine lipids are putative biomarkers.	(Favre et al. 2017)
Alga *Ulva mutabilis*	GC-EI-TOF-MS and UHPLC-ESI-TOF-MS	PCoA[a]	CAP[b]	The exo-metabolome and development of *Ulva mutabilis* is changed by the presences of symbionts.	(Alsufyani et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Clams *Ruditapes philippinarum*	^{1}H-NMR	no	PLS-DA	The analysis of *R. philippinarum* gills after been exposed to hypoxia showed changes in the concentration of some amino acids and energy related metabolites.	(Y. Zhang et al. 2017)
Sea snail *Haliotis diversicolor*	^{1}H-NMR	PCA	OPLS-DA	The exposure of *H. diversicolor* to organotin contaminant cause changes in the energy metabolism, osmotic balance oxidative stress. Moreover, the metabolic response is different depending on the sex and the tissue analyzed.	(Lu et al. 2017)
Mussels *Mytilus galloprovincialis*	^{1}H-NMR	PCA	no	The exposure of *M. galloprovincialis* to drospirenone has no effect in their sexual development but disrupt energy, amino acids, and glycerophospholipid metabolism.	(Cappello et al. 2017)
Macroalga *Asparagopsis taxiformis* Coral *Astroides calycularis*	UHPLC–ESI-QTOF-MS	PCA	no	The interaction between a coral and an invasive alga revealed no changes in the metabolome of the coral while the metabolome of the alga was changed when in contact with the coral.	(Greff et al. 2017)
Fish *Salvelinus alpinus*	^{1}H-NMR	PCA	PLS-DA	The metabolic changes caused by a test diet were reflected in the metabolome of the plasma, liver and muscle. Based on this, improvements in the diet were proposed.	(Cheng et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Sponge *Haliclona mucosa* and *Haliclona fulva*	UHPLC-ESI-Q-TOF	PCA	no	Differences in the metabolome between the two species were observed. Additionally, a decrease in the diversity of the metabolome during the period between April and May was observed and variation due to the location was detected over a 200 km ratio in the Mediterranean Sea.	(Reverter et al. 2018)
Fish *Thunnus thynnus*	^{1}H-NMR	PCA	no	Study of the liver tissue of bluefin tuna showed differences in the metabolic changes according to the gender, caused by the accumulation of environmental contaminant. Energy- related metabolites, amino acids and lipids were identified as the most affected metabolites.	(Cappello et al. 2018)

[a] Principal coordinate analysis

[b] Canonical analysis of principle coordinates

increasingly clear that the solution to this problem may come via the combination of available metabolomics tools in a way that can provide the most useful information about the particular organism under study using bioinformatic tools.

Identification of the metabolites of interest continues to be the major restriction in metabolomics studies. Although dereplication can be performed in the initial steps, on many occasions the compounds cannot be detected within the mixture and have to be isolated for their identification. In the isolation process, the limited amount of sample is always an obstacle and this has been circumvented in some cases by the implementation of microscale isolation protocols for low amounts of crude extracts. Additionally, other dereplication tools such as *in silico* MS/MS comparison can be useful for the identification of compounds, as reference compounds needed for identity confirmation are generally very difficult, if not impossible to acquire. The use of these strategies can facilitate metabolomics studies and speed up the process of discovering new compounds and thereby reveal the function of these compounds in the organisms.

Until now studies have focused on the metabolites present inside the organisms, ignoring those that are released into the ocean. To achieve a deeper understanding of the interactions occurring in marine organisms, it is crucial to undertake the study of metabolites that are released into the water, as it is highly likely that these metabolites are possible mediators in these interactions. However, the study of exuded metabolites represents a major experimental challenge, as they are rapidly diluted in the ocean making their analysis very difficult. As a first approach, studies could be performed in the laboratory, in a controlled environment that reproduces the conditions is in the field. In addition, water filtering systems could allow to concentrate the metabolites using methods such as solid-phase extraction. The use of this kind of experiment enables the collection of metabolites in a sufficient quantity to perform further metabolomics analyses. These kinds of studies will become an indispensable tool to understand the transfer of metabolites between organisms, providing valuable information for fields such as aquaculture, marine chemical ecology, and environmental metabolomics.

Acknowledgements

The authors greatly appreciate the contribution of Dr. Erika G. Wilson in the scientific discussion of this paper. This work was supported by the COLCIENCIAS (#728, Science Technology and Innovation Ministry, Colombian Government)