

# Marine Biodiscovery in a Changing World



Maggie M. Reddy, Laurence Jennings, and Olivier P. Thomas

## Contents

1	Introduction	2
2	Exploration of Marine Biodiversity and Establishing Biomaterial Repositories	3
2.1	Definitions and Examples	4
2.2	Access and Benefit Sharing	5
2.3	Field Identification and Taxonomy	6
2.4	The Irish Marine Biomaterial Repository	8
2.5	Challenges and Opportunities for Biorepositories	10
3	New Chemical Entities from Marine Bioresources	12
3.1	Chemical Screening: Dereplication	13
3.2	Biological Screening	19
3.3	Ireland's National Marine Biodiscovery Approach	20
3.4	Prioritization Strategies	21
3.5	Natural Product Chemistry	23
4	Concluding Remarks	27
	References	29

---

M. M. Reddy · L. Jennings · O. P. Thomas (✉)  
Marine Biodiscovery, School of Chemistry and Ryan Institute, NUI Galway, University Road,  
Galway H91TK33, Ireland  
e-mail: [olivier.thomas@nuigalway.ie](mailto:olivier.thomas@nuigalway.ie)

M. M. Reddy  
e-mail: [mageshneem.reddy@nuigalway.ie](mailto:mageshneem.reddy@nuigalway.ie)

L. Jennings  
e-mail: [laurence.jennings@nuigalway.ie](mailto:laurence.jennings@nuigalway.ie)

## 1 Introduction

The marine environment represents an untapped reservoir of bioresources that play an increasing role in the wellbeing of our societies as stated in the emerging field of Blue Growth [1]. Several maritime countries have identified the potential of marine bioresources for the development of their societies. Together with the exploration of our oceans, the most lucrative application was soon identified as drug discovery and the term “marine pharmacology” was used in the late 1960s [2]. At around the same time, the term “bioprospecting,” also known as “biodiversity prospecting,” was used to describe the exploration of natural resources in the search for biomolecules of commercial value.

The term “biodiscovery” was only defined later in the early 2000s and used specifically in Australia and New Zealand, where it replaced “bioprospecting,” as this term was often associated with biopiracy. The creation of a biodiscovery act in 2004 in Australia, Queensland, set the scene for a broader use of this term, especially, for marine exploration. Investigations in this area quickly extended to other maritime countries including but not limited to South Korea, Japan, Italy, Norway, Germany, and the United Kingdom (Scotland). Significant academic national efforts in China, the USA, and Europe, coupled with private initiatives like Pharmamar in Spain have contributed to the collection of marine macro- and microorganisms from several regions around the world. The main objective was to discover new and high-value drugs from these organisms, with a special focus on anticancer agents [3–5]. Marine biodiscovery has focused largely on marine invertebrates, but this has recently shifted toward microorganisms that represent a much more diverse group, and thus opening new avenues for biotechnology and an efficient producer of bioactive metabolites [6].

Some lessons have been learnt from the early development of this field, and some recent changes should prompt us to propose new directions in the field of marine biodiscovery:

- the decreasing rate of discovery of new chemical entities among marine natural products;
- the decreasing number of taxonomists specializing in marine organisms;
- climate change especially affecting the oceans;
- accelerated loss in biodiversity;
- increasing inequalities, especially for developing countries, which are often the richest in terms of marine biodiversity; and
- the emergence of new pandemics in the context of public health.

In this contribution, we will detail the construction of a National Marine Biodiscovery Laboratory in Ireland that has been developed, by considering some or all of these changes. Key modifications have been applied to our first initiative to take into account most of the abovementioned changes [7, 8]. First, the construction of regional marine biomaterial repositories can mitigate some of these recent changes. Then, the development of new technologies for the screening and dereplication of

the biomaterials will also contribute to address some of the other issues. Finally, we will propose some important changes that could make the term marine biodiscovery more inclusive and central for the development of the three pillars of sustainability: the blue economy, environmental challenges, and also social impacts.

## **2 Exploration of Marine Biodiversity and Establishing Biomaterial Repositories**

All key indicators point to an unequivocal global decline in biodiversity across most terrestrial [9] and marine [10, 11] organisms. Anthropogenic activities, such as over-population and habitat destruction, are the leading factors responsible for this unprecedented loss of biodiversity. This is heightened further by unsustainable resource use and climate change [12]. The overall negative trend in biodiversity has gained momentum over the past few decades and is projected to continue in a downward trajectory [13] in the absence of global intervention. Drastic changes in current agricultural practices and fishing methods or ambitious conservation efforts are urgently needed if humanity is to reverse current biodiversity trends by 2050 [9]. However, further mitigation strategies are also needed and will require innovative solutions and synergistic efforts from industries, governments, and scientists.

Safeguarding against biodiversity loss goes beyond immediate conservation goals and human benefit, but may also be important for tackling future climate-related challenges [14–16]. Biodiversity consistently has represented a source of food, fertilizer, medicine, cosmetics, and textiles for human society. For example, traditional medicine was integral to ancient societies and was used to treat various ailments and disease [17, 18]. Interestingly, naturally inspired drugs remain important today and have greatly benefitted from advances in natural product chemistry and drug discovery, including from the marine environment [5].

Bioactive compounds originating from the ocean are opening new avenues for the bioeconomy due to their unique structures and bioactivities. Similar to terrestrial plants, marine macro- and especially microorganisms, have been shown to produce highly diverse biomolecules usually absent in the terrestrial environment [19, 20]. However, despite several decades of exploration in marine biodiscovery, only a few groups of interest are well studied, such as bacteria, fungi, sponges, tunicates, and octocorals, and even fewer (Porifera, Chordata, Mollusca) account for marine-derived or marine-inspired drugs available in the market today [21]. Interestingly, marine-derived drugs have exponentially grown in the last 15 years, showing potential for future research in this field. Although it may be possible to discover new compounds with new applications from the same organism [22], or through the application of new methods, exploration of novel biodiversity ecoregions or targeting non-model organisms will more likely result in the discovery of new natural products.

Expanding chemical and biological screening efforts beyond model organisms, however, will depend on access to a wide range of biodiversity samples that can be

difficult, costly, and time consuming, especially for marine organisms [6]. Research expeditions typically focus on specific groups of interest [23], and those investing in wider sampling efforts are understandably confined to specific geographic regions. The wealth of biodiversity samples collected during these costly research expeditions are then retained in university-based or private biodiversity collections and are only used to address specific research questions locally. Another major drawback of one-off research expeditions is the possibility of overlooking transient or rare species with potentially new chemical compounds. Taken together, and on considering additional limitations, such as stricter permit regulations, limited funding for exploratory research missions, and a lack of research visibility, it is imperative that we rethink our current approach for the collection, storage, and utilization of biodiversity samples for marine biodiscovery.

## ***2.1 Definitions and Examples***

Biodiversity collections broadly encompass preserved, living, or digital biodiversity collections and may focus on specific taxa or specific geographic regions and represent publicly accessible or private collections. Natural History Collections (NHC), Biorepositories or Biobanks, and university-based or private collections exist within the broad concept of biodiversity collections. Natural History Collections are probably the best-known example of a biodiversity collection and comprise preserved collections such as those housed in herbariums and museums, and living collections such as botanical gardens, zoos, aquaria, and tissue/culture collections [24]. The recent emergence of digital repositories also adds to the broader concept of NHCs and may be associated with physical specimens, e.g., NMNH Biorepository, GBIF [25], and IDigBio [26], or based on observational data (e.g., iNaturalist), and may account for global inventories, e.g., GBIF, inventories for specific groups, e.g., AlgaeBase or The World Porifera Database or regional inventories (e.g., NIST). In general, biodiversity collections have gained far more visibility from digitization, reaching a much wider user base than previously imagined [15].

Historically, natural history collections, such as herbariums and museums, primarily have housed “type specimens” and served as a record of both extant and extinct biodiversity. However, associated specimen details have since been used to address a broad range of biological, ecological, biogeography, and evolutionary questions [24]. It has also been used to assess changes in biodiversity patterns over time or to detect and track invasive species [27]. Natural History Collections contain a wealth of knowledge that has more recently been disseminated through public outreach and educational programs [28]. However, the main purpose of NHCs remains the collection and storage of biodiversity specimens. Although voucher specimens may be available on loan, access to such specimens is usually difficult and often require non-invasive utilization of samples, which usually are not suitably preserved for applied research such as genomic, biotechnological/bioprospecting, or biomedical research.

Biorepositories essentially extend the function of NHCs and are involved mainly in the collection, processing, storage, and distribution of biospecimens or derived biomaterials for scientific research, with a large focus on applied research. Biorepositories were developed over the last few decades in response to a rising demand for high-quality samples for “-omic” sciences [29, 30]. Although the terms “biobanks” and “biorepositories” are often used interchangeably, it is important to note that a biobank is a type of biorepository. Biobanks were developed with a central focus on human-derived samples and associated information for biomedical research [31]. There are numerous different types of biobanks, each designed for a specific application, e.g., cancer or HIV/AIDS research and generally covering a broad geographic range. Even though the term biobank is largely reserved for human-derived samples, it also has been adopted for plant (seed banks), animal, microbial, and environmental collections.

More commonly, however, plant, animal (excluding humans), or microbial biospecimens and derived biomaterials are typically held in biorepositories and may comprise preserved (e.g., NIST) or living collections (Roscoff Culture Collection). Biorepositories either target a broad range of taxa from a specific geographic region (e.g., GEC BioRepository) or habitat (e.g., MarBank), or focus on a particular group (e.g., microorganisms: Roscoff Culture Collection). More recently, there has been a major focus on building marine genomic biorepositories, mostly in North and Central America to ensure the preservation of biodiversity in the face of a rapidly changing world. Examples of these include the GEC BioRepository in Guam, the Ocean Genome Legacy Center, the NIST Biorepository, Genomic Biorepository of Coastal Marine Species in Estero Padre Ramos and Estero Real, Nicaragua [32], the Smithsonian, MarBank in Norway [33], and the NEON Biorepository at Arizona State University. In all cases, biorepositories operate at a national level and cater to specific research needs, e.g., genomic studies. Despite the huge benefits associated with these biorepositories, in some cases, a lack of funding and support, especially in developing countries, has limited their continuation [31].

## ***2.2 Access and Benefit Sharing***

Biodiversity is overwhelmingly concentrated in tropical regions of the world, which also corresponds with the locations of many developing nations. Furthermore, the marine environment of several maritime developing nations remains relatively unexplored. There is of course enormous potential for marine biodiversity and biodiversity research in these regions, as they often host a wide range of undocumented taxonomic diversity, and hence an untapped source of potential new natural products [6].

All coastal countries have sovereignty and exclusive access to marine resources within 200 nautical miles of their coastline, regions referred to as their “Exclusive

Economic Zone.” Access and utilization of samples, particularly in coastal developing nations, raises an ethical concern for access and benefit sharing. The “Convention on Biological Diversity” later supplemented by the “Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization” were drafted to address concerns regarding the misappropriation of marine and other genetic resources. The resultant Nagoya Protocol was implemented in 2014 and is a legally binding agreement signed by 128 member states as of 2020. In cases where requirements for the “United Nations Convention on the Law of the Sea” (350 nautical miles) and the Nagoya Protocol (250 nautical miles) overlap, both conventions come into effect.

In summary, access and benefit sharing broadly encourage fair access and utilization of resources between nations, with a strong focus on shared benefits among developing and developed nations. In addition to monetary benefits, non-monetary benefits which may include traditional knowledge, research outputs, or building research capacity and transferring technology within providing countries are also considered [15]. Interestingly, Schindel and Cook [15] also discuss shared risks in addition to shared benefits. There are two main but different aspects of the Nagoya Protocol, which permits for (1) access and (2) utilization of samples [34]. Collection and access to samples typically involve local partners and traditional knowledge and begin with “Mutually Agreed Terms” negotiated between “providers” and “users.” It outlines detailed access and benefit sharing, “Material Transfer Agreements,” and third-party access granted for interested parties not originally involved in the mutually agreed terms agreement. Material Transfer Agreements (MTAs) should be approached in two phases, first, by agreement on research terms, and then, subsequent agreement terms for commercialization. Material Transfer Agreements are required before users request a “Prior Informed Consent” which is typically granted by competent authorities of individual member states, for example, the Ministry of Fisheries. It is the joint responsibility of “providers” and “users” to ensure that all necessary domestic permits are obtained, and national regulations respected. Additionally, coastal countries need to be informed of planned marine research activities 6 months in advance. The resultant “Prior Informed Consent” is proof that resources were obtained legally [35]. Although not yet widely implemented, correctly performed “Prior Informed Consent” may become a requirement for scientific publication in the future [34]. Non-compliance with access and benefit sharing may result in fines or interruption and even cessation of research and commercialization.

### ***2.3 Field Identification and Taxonomy***

A major challenge in collecting, characterizing, and utilizing biodiversity samples is the identification of specimens. Species names are tags and as such contain important biological, ecological, and evolutionary information. The accurate identification of specimens is therefore fundamental to subsequent research but depends on available taxonomic information and expertise. New species are defined (species delimitation)

based on the available information and the application of available methods or tools [36–38], and assigned names (nomenclature) following various naming conventions. As a result, taxa are commonly re-assigned to more appropriate groups as new information becomes available. Species names and their classification are therefore in a constant flux and continue to evolve, as do the species themselves.

It is important to remember that species identification and species characterization are fundamentally different. Species are characterized based on shared traits or similarity and identifiable by a set of diagnostic traits using different species concepts as working hypotheses. The Morphological Species Concept, Reproductive Species Concept, and more recently, the Evolutionary or Phylogenetic Species Concept are among the most widely applied, each relying on different types of data and subject to various advantages and limitations [39]. Most taxonomists today generally apply a unified species concept. Different species concepts are still widely debated and this has been referred to as “the species problem” by Ernest Mayr. This long-standing debate reflects the notorious difficulty associated with certain species.

A broad spectrum of diagnostic traits is found commonly among individuals of a single species and therefore species boundaries (species delimitation) needs to be carefully considered. New species are discovered against this background and are described when sufficient evidence supports their placement as distinct entities. Although there are no rules in taxonomy, there are rules when naming (nomenclature) a species. In recent years, there has been a strong move toward integrative taxonomy, which is intended to reduce the subjectivity often involved with species delimitation. For this approach, multiple lines of evidence, such as morphology/anatomy, ecological/biological, DNA, and DNA-based algorithms, commonly are used more than a single method. More recently, metabolomics, which provides a unique metabolic fingerprint, has also been used as an additional line of evidence in integrative taxonomy and has been successful in identifying taxonomic ambiguities [40, 41] and defining species [42].

Taxonomy is fundamental in terms of subsequent species identification. However, this may be challenging for a large majority of marine organisms as they often represent understudied groups. Species identification may be even more challenging when blurred species boundaries, such as overlapping diagnostic traits or cryptic taxa, are encountered. The problem of proper species identification is hampered further when field identifications are confused with species identification. Field names are commonly assigned to taxa based on visual inspection of the overall gross morphology and distinctiveness. However, field identifications based on morphology can be misleading, and typically require further verification using microscopy or DNA [43]. This is particularly true for sister species of structurally less complex organisms, such as seaweeds or sponges, which superficially resemble one another.

Deoxyribonucleic acid barcoding was developed as a fast, cost-effective, and standardized method for species identification [44]. The cytochrome oxidase subunit I is utilized commonly as the marker of choice for barcoding, but different DNA barcodes or a set of DNA barcodes may be preferred for different groups. Reliable identification based on DNA barcoding depends on comprehensive reference libraries against which unknown specimens are identified based on similarity [45]. Although not

intended for this purpose, DNA barcoding has also been used for species discovery. It is important to remember that even though DNA barcoding may expedite the identification of new species, it does not replace alpha taxonomy and taxonomic expertise that are essential for species descriptions.

Accurate identification of taxa is fundamental to marine biodiscovery/natural product chemistry for several reasons [6]. First, it may be informative for prioritizing taxa for marine biodiscovery, as similar taxa are known to produce similar chemical analogs [46]. Accurate species identification is also critical at the early stages of drug discovery, as recollections of species may be required for further testing. Although chemical synthesis of products may limit the need for recollections thereafter, this procedure cannot be used routinely to synthesize complex chemical structures [47] of alginates for example. Last, misapplied names and hidden diversity could lead to misleading species distributions and mask potentially interesting regions for marine biodiscovery exploration.

In addition to accurate identification of taxa, fully understanding the biology and ecology of species is also important for marine biodiscovery and natural product chemistry. While shared common metabolic pathways can be a more significant predictor at higher order phylogenetic ranks, habitat or ecological diversity are believed to evoke certain secondary metabolic pathways and account for similar bioactivity among similar genera or species [19].

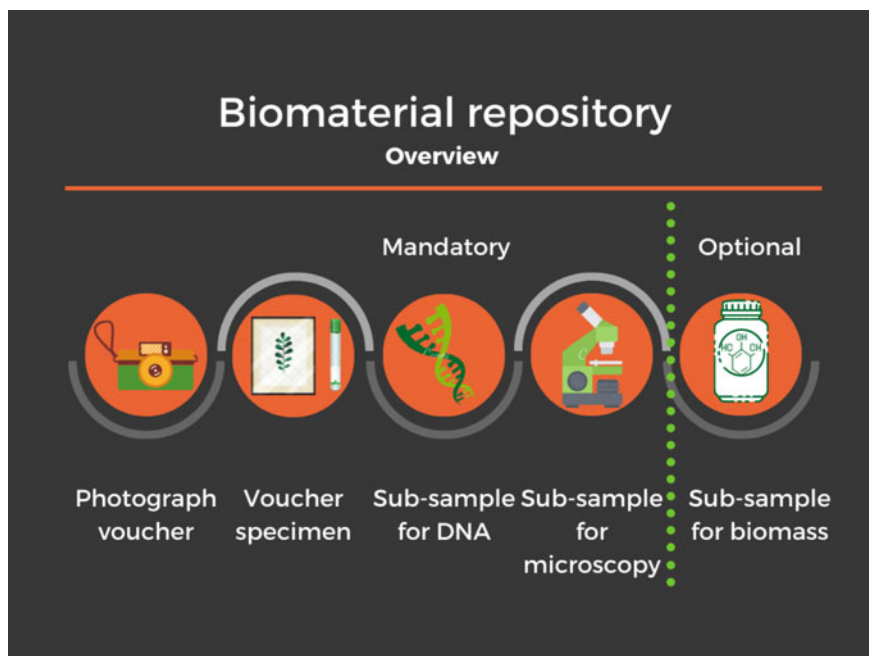
## ***2.4 The Irish Marine Biomaterial Repository***

The Irish coast extends for roughly 3,000 km, bounded by the Irish Sea to the east and the north-east Atlantic Ocean to the west. This dynamic coastline is richly diverse in marine life and has a long history of marine exploitation with seaweed utilization dating back to the Middle Ages. However, the exploitation and use of marine resources have remained largely unchanged since then, despite enormous potential applications.

The Irish Marine Biomaterial Repository (IMBR) has been developed recently, with the aim of bridging the gap between research, industry, and conservation. The central tenet of the biorepository is to provide a focal point for shared bioresources. The IMBR focuses on, but is not limited to, the collection, processing, storage, and distribution of coastal macro-marine organisms with a main priority on the preparation of chromatographic fractions ready to be used for biological and chemical screening. It is organized according to a data management plan centered around a web database that allow the recording and tracking of all the data present in the IMBR [48]. Voucher specimens associated with subsamples for DNA, microscopy, and biomass are prepared following best practices [31] (Fig. 1). Careful characterization of morphological, genetic, and chemical profiles are generated as a baseline for future comparison purposes.

The IMBR endeavors to provide accurate species identification from a combined morphological and DNA approach by working with taxonomic experts, especially





**Fig. 1** The Irish Marine Biomaterial Repository (IMBR)

when cryptic taxa are encountered. Proper taxonomic identification of species may also reveal related organisms of interest and will aid in prioritizing taxa for marine biodiscovery. The need for accurate species identification is crucial to all research and is particularly important for marine biodiscovery research, as a growing body of evidence suggests that different species may produce structurally different bioactive compounds with different applications [22]. However, as mentioned earlier, accurate identification depends on the available taxonomic information, which is often poorly documented for the large majority of marine organisms, particularly in understudied regions. For this reason, voucher specimens will remain permanently in the collection and will be revised continually to reflect updated taxonomic changes. All specimens will also be associated with a DNA barcode that will facilitate the process and may be used as a unique identifier until taxonomic research on specific groups is carried out. This will avoid the waiting time for species to be described as new, which may take a few decades in some cases [49]. Taxonomic changes will be made and communicated with end-users of the IMBR. All available specimens and associated metadata, including in situ and field photographs, are available online in a standardized and open access format. Future developments such as providing access of subsamples to interested users (researchers or the industry) for further study remain to be developed, as does testing the viability of chromatographic fractions or extracts of organisms for long-term storage. Although some challenges and additional future work remain,

this pilot-scale project offers enormous opportunities for future development when applied to different regions of the world.

## ***2.5 Challenges and Opportunities for Biorepositories***

In terms of the cost and benefit of biorepositories, it is recognized that establishing and maintaining a biorepository will require initial capital investment (facilities), capacity (staff), and time (curation). However, investing in these costs today may yield better returns for the future and could mean a trade-off between costly research expeditions and maintaining redundant collections. Below, are discussed important benefits that the authors believe marine biorepositories provide, and hence may further justify the costs and risks associated with maintaining such biodiversity collections.

Marine biorepositories, such as the IMBR initiative, are intended to foster collaboration between research sectors (academia, industries) that would otherwise work in isolation or in parallel. For academic researchers, access to a broad range of samples may be highly beneficial for students and postdoctoral fellows working on time-limited projects, which require extensive spatial and temporal comparisons. As a result, it may also lead to achieving broader study aims or guide a better study design. Furthermore, linking data (DNA, morphology, chemistry) originating from a single specimen will promote robust and accurate results, especially considering the rampant misidentification of specimens and the misapplication of species names. Another major benefit of marine biorepositories is extending the longevity and value of samples beyond immediate study goals. Specimen donations from previous projects or from private or university-based biodiversity collections are at risk of being lost when researchers relocate or retire. These collections could add to a growing and comprehensive knowledge base and promote easy access to samples and wider availability of associated information if donated to open access biorepositories.

### **2.5.1 Working with Natural History Collections**

Biorepositories should be viewed as being auxiliary rather than a substitute for natural history collections, as each biodiversity collection functions with different but complementary goals. For example, long-term or baseline data from natural history collections may be supplemented with new collections from biorepositories. In some instances, natural history collections have recognized the changing needs of modern-day research, including host museum-based biorepositories, e.g., the Smithsonian [30]. Clearly, similar efforts will benefit greatly natural history collections at large, but these are unlikely to be achieved for many natural history collections due to escalating challenges related to limited funding and lack of personnel. Given the complementary functions of natural history collections and biorepositories, it is

believed that it could be mutually beneficial for these facilities to work together for shared funding, expertise, and facilities.

The value of natural history collections and biorepositories has received considerable attention recently in the light of emerging infectious diseases. Holistic sampling is the collection of either host and parasite collections, or extended specimen details, which are among the eminent changes needed for the next generation of biodiversity collections [15]. This, complemented with a range of chromatographic fractions ready-to-be tested against new infectious diseases, could prevent or alleviate the damaging social and economic losses incurred by infectious diseases such as the current COVID-19 pandemic. However, these chemical and biological libraries can also be used to resolve other biodiversity-related issues. The impact of plastic pollution is well documented in the literature, but far fewer studies propose solutions to this growing problem. The utilization and convenience of single-use products is so pervasive today that changing the mindset of society may take time. However, finding alternative natural sources to plastics, such as biopolymers, could be an effective intermediate solution until drastic paradigm shifts in societal attitudes occur.

### **2.5.2 Biodiscovery and Conservation**

If biodiversity is required as a source of biomolecules to cope with future challenges, then intuitively our future initiatives should be developed with sustainability in mind. In the short term, this could be achieved by targeting species with high biomasses and not at risk of extinction or species amenable to aquaculture (see the species prioritization list below). Wild-harvested species with commercial applications should also be coupled with monitoring and environmental impact assessments. Identifying regions of high biodiversity and potential application could also be used to advise in marine-protected areas for planning purposes and to encourage the integration of such data from biodiscovery. Planning in marine-protected areas will therefore not only consider environmental and social sustainability goals but also include economic sustainability.

### **2.5.3 Addressing Social Challenges**

Bioproperty and shared derived benefits from bioresources remain a major concern today. At present, marine biodiscovery may be carried out with little to no involvement by participants from the providing member states where the samples are collected, typically in developing nations. This is understandable as the availability of facilities and expertise are largely concentrated in the developed world. However, more rudimentary stages of marine biodiscovery that involve less technical equipment and requirements should focus on building local capacities. For example, establishing and maintaining marine biorepositories could have a strong social impact on developing nations. This has the potential to build and support research capacity and add value to resources. Other future directions may include developing “green” and

sustainable extraction processes in-state using renewable and accessible resources, thereby diversifying the use of resources rather than depending on a single resource. It is also needed to focus on more diverse applications of the biomolecules, such as exploring alternatives for current animal feed that is often protein fit for human consumption, or tackling other food security issues even though this may be less potentially profitable than drug discovery.

In a changing world, scientific research may be propelled into new and unpredictable directions. As biorepositories extend the function of natural history collections, new research methods may require new preservation methods or sample requirements. Biodiversity collections therefore need to be dynamic and able to meet the ever-changing needs of research and development. The combination of this approach with new tools in drug discovery described below has the potential to address many of the new challenges of our present century.

### **3 New Chemical Entities from Marine Bioresources**

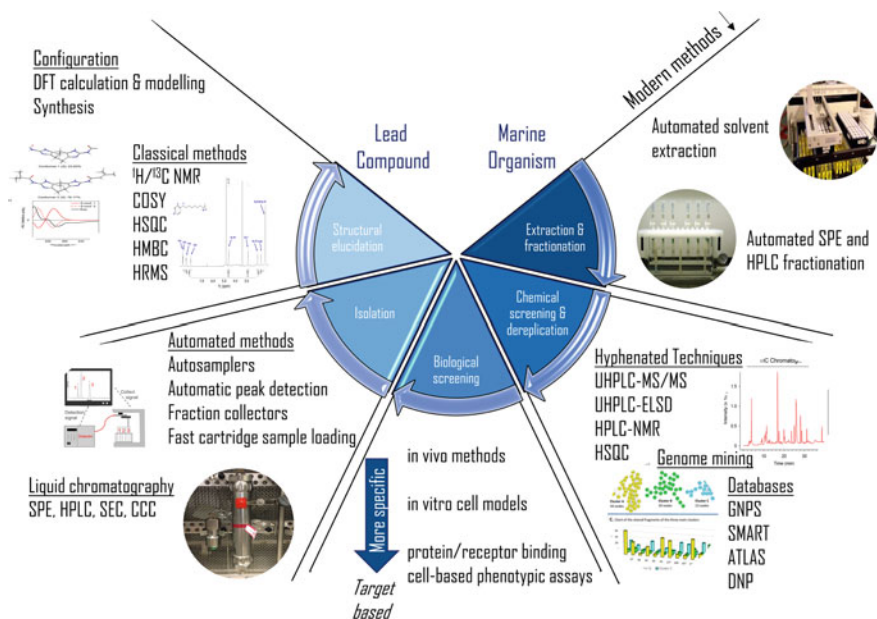
The identification of new bioactive chemicals from marine organisms traditionally has been conducted through the use of conventional biological screening, liquid chromatographic separation, and nuclear magnetic resonance (NMR) structural elucidation. However, over the last two decades, the development of highly sensitive robotic hyphenated methods has led to alternate approaches to marine natural product isolation. This has increased significantly the ability to produce large libraries of extracts and chromatographic fractions, as well as chemically and biologically screen a large number of samples so as to focus on the directed isolation of new natural products. Therefore, the contemporary marine biodiscovery workflow typically now includes an initial chemical analysis and dereplication stage utilizing these modern techniques. These methods are usually coupled with updated natural product databases and calculated chemical data, so that specimens can be prioritized rapidly for the isolation of new natural products.

These modern methods can be further utilized for the fast and semi-automated isolation and structural elucidation of compounds. Additionally, the significant increase in resolution and the sensitivity of traditional methods has led to a greater ability to characterize new molecules using extremely low nanomolar quantities. These modern procedures have had a significant impact on the field of marine natural products research over the last 20 years, leading to a substantial increase in the number of marine natural products reported in the scientific literature each year.

This section will cover the typical contemporary workflow employed by marine natural product chemists, starting from the use of samples present in a marine biorepository to the identification of a bioactive molecule, focusing mainly on new developments. This includes chemical screening and dereplication stages, biological screening, and the isolation and identification of natural products.

### 3.1 Chemical Screening: Dereplication

With over 35,000 marine natural products isolated, it has now become standard practice in marine natural products chemistry to use high-throughput chemical screening and dereplication methods to prioritize isolation from unknown specimens. Normally, it is preferable to screen chemically and biologically fractionated extracts, as more complex crude extracts may confound screening results [50]. While the use of fractions can provide better screening efficiency, this greatly increases the quantity of samples that need to be evaluated. However, more recent high-throughput technology has allowed for the development of methods that can be used to screen thousands of samples.



**Fig. 2** The marine biodiscovery workflow from a biomaterial repository to the isolation of bioactive compounds. This process includes five major steps including extraction and fractionation, chemical screening and dereplication, biological screening, purification, and finally, structure elucidation. This schematic highlights the most important modern methods that increase the efficiency and outcomes of this process. Abbreviations: DFT (Density Functional Theory); NMR (Nuclear Magnetic Resonance), COSY (CORrelation SpectroscopY), HSQC (Heteronuclear single quantum coherence), HMBC (Heteronuclear Multiple Bond Correlation); HRMS (High Resolution Mass Spectroscopy); SPE (Solid Phase Extraction); HPLC (High Performance Liquid Chromatography); SEC (Size Exclusion Chromatography); CCC (CounterCurrent Chromatography); UHPLC (Ultra High Performance Liquid Chromatography); MS (Mass Spectroscopy); GNPS (Global Natural Product Social Molecular Networking); SMART (Small Molecule Accurate Recognition Technology); DNP (Dictionary of Natural Products).

Over the last two decades, a number of new hyphenated methods have become highly popular for the dereplication of known natural products. The standard combination is an analytical chromatographic separation procedure coupled with an analytical detector. These methods include high-performance liquid chromatography-MS, high-performance liquid chromatography-evaporative light scattering detection, and high-performance liquid chromatography-NMR spectroscopy. These results then are compared to dedicated databases containing chemical data from known natural products (Fig. 2).

### 3.1.1 Chemical Structure Databases

The process of dereplication can be defined as the identification of already known natural products through the comparison of experimental data with pre-existing analytical information. Therefore, over the last 20 years, publicly available databases have been developed that contain continuously updated lists of known natural products as well as analytical data related to each compound. Many of these databases are now being designed specifically for dereplication purposes to contain appropriate search tools including structural and analytical data searches, as well as bioactivity and taxonomy searches. The latter is extremely important in marine natural product chemistry as the efficiency of dereplication can be improved substantially if the taxonomy of a given marine organism is known. While a number of general chemical databases are available, more recently, more specialized marine natural product databases have been developed.

A number of common chemical and natural product databases are typically used in marine natural products for dereplication. General chemical databases (synthetic and natural molecules) typically used for this purpose include SciFinder [51], PubChem [52], ChemSpider [53], ChEBI [54], ChEMBL [55], and REAXYS [56]. These databases all contain comprehensive records of new chemical entities that can be inspected readily using structural data and search terms. While these databases all contain broad levels of data, the search results obtained can often be confusing due to the number of synthetic products also included. As such, a number of databases specific for natural products have been developed, including the “Dictionary of Natural Products” (DNP) [57], among others [58]. The “Dictionary of Natural Products,” in particular, is regarded widely as the most comprehensive database of natural products, and in addition to structural and keyword based searches, it also offers researchers an ability to search for marine natural products based on biological activity and taxonomy.

However, with over 300,000 reported natural products thus far, more specialized databases are becoming ever more common in marine natural product chemistry [59]. These databases include Antimarin, MarinLit [60], ATLAS [61], CyanoMet [62], and StreptomeDB [62]. The latter three databases have been designed specifically for microbial natural products and contain information regarding the gene sequences of some species. However, for the last two decades, the database that has become most popular within marine natural products chemistry is the MarinLit

database. MarinLit was established in the 1970s by Professors John Blunt and Murray Munro to specifically record marine-derived natural products. Today, it is hosted by the Royal Society of Chemistry and contains >35,000 compounds. This database has become very popular among marine natural product chemists due to its unique tailored searches. These are separated into searches for compounds (structure, NMR chemical shift data, exact mass, molecular formula, or UV), taxonomy (phyla through to species), prior literature, and geographic location. A comprehensive compilation of all these data make this an important tool for modern marine natural product dereplication methods.

### 3.1.2 Analytical Separation Methods

Over the last 20 years, liquid chromatography-based separation techniques have become the most frequently used for the screening of marine extracts and fractions. Typically, these methods include the use of high-performance liquid chromatography, though this has begun to change in recent years with the introduction of the more efficient ultra-high-performance liquid chromatography (UHPLC). This method utilizes higher pressure systems (up to 1200 bar) with lower diameter, smaller particle size columns to allow separations with rapid run times and high resolution [64]. This is performed typically using reversed-phase columns, which are more appropriate for dealing with the polarity of most drug-like bioactive marine natural products [65]. Currently, it is becoming common for laboratories to include in-house dereplication databases that can also incorporate retention times of metabolites for the specific systems used.

Other common separation methods for the isolation of marine products include size-exclusion chromatography, countercurrent chromatography, and capillary electrophoresis. These additional methods are useful typically when separating highly polar materials due the high quantity of ocean salts in marine extracts that can hinder chemical analysis.

### 3.1.3 Mass Spectrometric Screening

Undoubtedly, mass spectroscopy (MS) has become the most highly utilized detection method in marine biodiscovery over the last two decades, due to the significant improvement in both resolution and sensitivity of modern time-of-flight spectrometers (ToF) and the development of newer mass spectrometers (e.g., Orbitrap™) [66]. Mass spectrometric data allow researchers to determine unequivocally the molecular formula of natural products even at nanomolar concentration levels. Furthermore, the ability to fragment metabolites through tandem mass spectroscopy (MS/MS) with no additional time involvement has made this method of chemical screening vastly popular. Additionally, with a wide variety of commercial ionization methods, including ESI (Electrospray Ionization), FAB (Fast Atom Bombardment), APCI (Atmospheric-Pressure Chemical Ionization), DESI (Desorption Electrospray

Ionization), and MALDI (Matrix-Assisted Laser Desorption Ionization), users now have the ability to obtain greater sensitivity for a wider range of different biological samples. Ionization methods such as DESI and MALDI allow the direct MS detection on solid samples of marine organisms without the need for extraction and other preparation processes [67, 68].

Mass spectrometric data usually can then be evaluated with the molecular databases described above to identify rapidly known compounds and can provide the exact mass and molecular formula. However, more recently, new resources have become available to allow for the automated dereplication of MS/MS data. Most notably, for marine biodiscovery, is the recent development of the Global Natural Products Social molecular networking resource (GNPS) [69]. This resource contains MS/MS databases populated and curated by the natural products community and allows for the visualization of clusters of structurally similar natural products. The software compares the experimental data to the database and returns fragmentation similarities between two molecules. Furthermore, it is becoming more common in marine biodiscovery to use the feature-based molecular networking tool for dereplicated metabolites and to identify potential new analogs [70]. While this is a useful tool for dereplication, some aspects of this approach can be highly laborious if conducting large-scale chemical screening. However, the GNPS feature-based molecular networking procedure has been widely adopted by the marine natural products community, with there being a rising number of marine natural products reported each year utilizing this tool [71]. Some recent examples of new natural products identified through molecular networking include lamellarin sulfates from the ascidian *Didemnum ternerratum* [72], microcolin lipopeptides from the cyanobacterium *Moorea producens* [73], and new chlorinated polyketides from a *Smenospongia* sp. sponge [74].

Other new MS/MS dereplication methods include the use of in silico fragmentation algorithms so that experimental fragmentation can be compared to computationally calculated fragmentation patterns of molecules matching the exact mass. These in silico packages include but are not limited to CFM-ID [75], MetFrag [76], and CSI:FingerID [77]. These can be highly useful resources, as a majority of published natural products do not have MS/MS data reported. These packages can be utilized to produce in silico fragmented assemblies of compounds and marine natural products to be compared to experimental data. Additionally, the use of moiety and pharmacophore in silico fragmentation has been employed for identified substructures within a molecule to target subsequently bioactive metabolites using MS/MS [78].

A primary limitation of MS methods is that the data obtained are not able to be used to confidently assign the definite structure of a natural product, especially for regio- and stereoisomers. Another limitation of MS is the ionization efficiency of natural products for detection. While a number of useful ionization methods are commercially available, issues can still arise from molecules that cannot be easily ionized or that form a plethora of different adduct ions that may complicate the analysis. Problems can also arise from highly instable compounds that can readily fragment with low ionization energies, hindering the identification of the exact mass of a given molecule.



### 3.1.4 Nuclear Magnetic Resonance Spectroscopic Screening

While mass spectroscopy has become highly popular due to its easy use in hyphenated setups, Nuclear Magnetic Resonance (NMR) spectroscopy provides far superior data for organic compound structural determination. For example, where MS cannot, NMR spectroscopy can be used to identify the different isomers from a change in the chemical shifts. Furthermore, the non-destructive nature of NMR spectroscopy is a major benefit over other dereplication methods, especially due to the low quantities of material typically available in marine natural product studies. Nuclear magnetic resonance spectroscopy has only recently become a more commonly used dereplication method, as a result of the recent developments in NMR spectroscopy automation including auto-sampling and auto-acquisition.  $^1\text{H}$  NMR experiments are rapid experiments and can be used to quickly obtain proton chemical shift data. However, due to the complex mixtures of compounds in marine natural product screening, overlapping signals with just a one-dimensional NMR experiment can hinder a dereplication process. With the increase in resolution and sensitivity of modern high-field NMR instruments, it has become more common to use rapid two-dimensional experiments including COSY (Correlation Spectroscopy), DOSY (Diffusion Ordered Spectroscopy), HSQC (Heteronuclear Single Quantum Correlation), and TOCSY (Total Correlation Spectroscopy) for chemical screening [79, 80]. This allows overlapping signals in one-dimensional NMR data to be identified and assigned. More recently, other NMR experiments have been designed to address these issues including the use of hyphenated NMR procedures. Modern hyphenated NMR techniques that have been described for the identification of marine natural products include high-performance liquid chromatography-NMR [81] and high-performance liquid chromatography-solid-phase-extraction (SPE)-NMR [82]. These methods can be used for the acquisition of NMR spectra using small amounts of crude material. Such hyphenated methods have been used previously for the identification of a number of known natural products from marine organisms including sponges, brown alga, and red alga [81, 83, 84]. An example of the importance of HPLC-NMR is the isolation of plocamenone and iso-plocamenone from the red alga, *Plocamium angustum* [83]. The NMR detection method was highly important as the use of other methods would have not allowed the discrimination of these isomers.

As NMR data has been an essential requirement for the publication of new molecules, since the late 1980s, a considerable amount of NMR data is available for known natural products. Therefore, a direct comparison of published and experimental data can be used to confidently identify known structures. A number of databases including MarinLit and DEREPA-NP [85] are utilized in marine biodiscovery to allow searches of chemical shifts and NMR features, respectively, for rapid dereplication work. The most recent NMR dereplication resource is the SMART NMR database [80]. This resource offers an automated dereplication through comparison of an experimental HSQC spectrum with the SMART database. The first compound to be reported following the use of the SMART dereplication was symplocolide A, a maritime natural product from the marine cyanobacteria *Symploca* sp. [86].

The current limitations of NMR spectroscopy are due primarily to the cost of buying and running this equipment. Costs associated with the preparation of samples for analysis are also typically expensive in requiring special glassware and deuterated solvents. Also, for an exact match between two sets of experimental data, NMR experiments must sometimes be conducted in the same solvent and in some cases at the same pH. This can make the selection of a solvent for screening mixtures of natural products complicated when weighing up the solubility and data availability.

### 3.1.5 Molecular Genomics Screening

With the development of new cheaper and faster DNA sequencing methods, it has become much more efficient to sequence and identify biosynthetic gene clusters. This, therefore, has become a useful tool for the identification of biosynthetic genes encoding for the production of natural products. The sequencing and analysis of these “biosynthetic gene clusters” can reveal a range of information including the precursor metabolites, the biosynthetic intermediates, and also the structure of the metabolites present. For this reason, researchers have begun to investigate this approach as a potential dereplication tool. The genomic investigation of organisms for the proposition of new natural products has been coined “genome mining” [87, 88]. This method is now becoming a more standard approach among groups with a strong focus on microbial-derived products. With the significant rise in marine microorganism natural products being discovered, this has become more popular in marine biodiscovery in recent years.

A number of new databases currently are used in marine biodiscovery for the rapid identification of biosynthetic gene cluster sequences. These include the MIBiG repository [89], the taxon-specific Prospect (fungal genes) [90], and ActDES (Actinomycetes genes) [91] databases as well as the computationally predicted gene databases antiSMASH [92] and IMG-ABC [93]. After identification of biosynthetic genes, it is then necessary to identify the structures or structural features for which these genes encode, to dereplicate natural products. With recent developments in machine learning, tools have been developed to use the identified genes to predict chemical structures. These tools include PRISM [94], SBSPKS v3 [95], and Adenyl-Pred [96] for the chemical structure prediction of predominately polyketides and nonribosomal peptides. Currently, a number of these tools have been tailored towards bacterial metabolites and gene sequences, although with the recent focus in marine biodiscovery on the isolation of fungal natural products, it is highly likely that there will be a shift towards fungal biosynthetic genes in the coming years [21].

There are some pitfalls to the use of genome mining that need to be addressed. The first issue is that this method is, currently, mainly applicable for the identification of polyketide and nonribosomal peptide sequences. This limits the use of genome mining for marine natural products due to the significant number of unique marine-derived terpene, alkaloid, and polyaromatic metabolites. Additionally, the structures of these natural products cannot be identified based solely on the genomic data. Unknown genetic transformations or spontaneous reactions of a natural product

may not be predicted confidently as the molecule encoded by the genes may not match the final product. Another limitation is that a number of natural products are produced by multiple organisms and that some moieties may be incorporated into a molecule through an organism's diet [97, 98]. While this method will likely improve in the future, currently, there is still a need for this genomic approach to be used in conjunction with chemical methods to confirm molecular structures.

### **3.2 *Biological Screening***

Over the past 60 years, marine organisms have been a rich source of unique bioactive compounds with activities including antitumor, antibacterial, antifungal, antiinflammatory, antiviral, antiparasitic, and antioxidant activities, useful for a broad range of applications, but mainly for drug discovery. High-throughput screening methods of the twenty-first century have transformed the field of drug discovery from marine sources. The use of methods utilizing high-throughput automated liquid handlers, plate readers, and 384/1,534-well assay plates have increased throughput to a point where thousands of marine samples can be screened over a few days. Traditional assay methods utilize *in vivo* whole organisms and *in vitro* organism-derived cell lines. The screening of natural product extracts in cell line assays can often lead to non-specific hits and false positives that typically are not advantageous for development as lead compounds. However, more recently, bioassays have shifted to more specific phenotypic cellular function assays and protein/receptor binding assays [99]. These highly specialized assays can be useful for the identification of target specific compounds that may then be classified as lead compounds.

Cell line models are still the most common biological assays used for screening marine natural product extracts. This is due to the primary targets for drug discovery being traditionally antitumor and antimicrobial compounds. Cell-based assays have been used significantly in cancer research due to the availability of tumor cell lines. Similarly, microbial cell cultures are very common and can be screened with relative ease. Furthermore, a number of other more specific phenotypic genetically modified cell lines have become common when screening for antiinflammatory, antiviral, and antiparasitic activities. Cellular assays are very common due to their development as high-throughput assays in microwell plates with robotic liquid handlers. A major problem with cell-based assays is that they often use established and immortalized cell lines that are vastly different from cells of *in vivo* systems [100]. However, typically this is a trade-off that is made to achieve higher throughput results.

Biological assays in marine biodiscovery are shifting slowly from these *in vitro* cell assays to more specific cell-free protein biochemical assays [99]. With a goal of obtaining natural products with more specific activity, target-based receptor binding assays have become more common primarily due to their increased specificity. Additionally, these offer the ability for multiple different methods of analysis so that not only drug targets can be identified but also diagnostic tools. An enzyme substrate can be used to identify if the activity is inhibited, and fluorescent dyes can be used

to bind to the protein–ligand complex, and more recently, in marine biodiscovery, native MS can be used to identify protein–ligand complexes [101, 102]. In particular, the protein–ligand native MS assay has been of interest in marine biodiscovery due to its ability to obtain the MS data for unknown active compounds allowing a mass-guided purification as opposed to a bioassay-guided study. However, these assays further trade cellular mechanisms and similarities to in vivo systems, to achieve higher throughput, higher specificity, and often false results.

Bioassay-guided purification is probably the most widely utilized method when conducting natural product drug discovery projects. This process is very appealing as the bioactivity at each purification step can target only bioactive components. However, difficulties can include inactivity due to false positives [103]. This is a common occurrence, especially when starting with crude extracts instead of less complex fractions. The use of more modern receptor binding assays led to some success for complex mixtures of natural products in overcoming this issue [100]. As such, receptor binding assays have become more common for bioassay-guided purification of active samples [101]. These methods are also advantageous due to the microscale of the assays in requiring smaller quantities of material. However, due to the common possibility of false positives when screening marine organism extracts, it is becoming more common to rely less on the biological screening results and more on chemical data. Furthermore, methods that integrate the chemical screening with biological screening can help in the identification of false results and also point to potential molecules within fractions that may be responsible for activity [104, 105]. We suggest that the use of both target-based receptor binding assays and chemical screening on marine fractions followed by cell-based and in vivo models would be the most beneficial process.

### ***3.3 Ireland's National Marine Biodiscovery Approach***

In Ireland, a major focus of the National Marine Biodiscovery Laboratory is the large-scale chemical screening and dereplication of known metabolites, and targeting unknown metabolite identification. This dereplication method utilizes a library of fractions produced through the organic extraction and fractionation using reversed-phase solid-phase extraction and storage in the biomaterial repository described above. These fractions are then subjected to chemical screening using two hyphenated methods; UHPLC-MS/MS for dereplication using the MarinLit database and the GNPS molecular networking tool as discussed above, as well as UHPLC-ELSD (evaporative light-scattering detection) to identify relative quantities of compounds within a sample for prioritization. This screening is performed utilizing the same ultra-high-performance liquid chromatography methods and columns so that the MS data and ELSD data can be compared. This ELSD dataset is important, as a large proportion of peaks that are identified using MS data are typically compounds that are strongly ionized and are present at low nano- and picogram quantities. This can

be problematic if samples containing new natural products obtained in quantities too low for isolation are prioritized. The ELSD data allows researchers to identify the relative quantity of material available in an extract so that time is not wasted purifying samples that will not yield compounds that can be analyzed by NMR spectroscopy. Additionally, if only a small amount of biomass is available, this can be prioritized for recollection so that minor compounds of interest can be isolated. On the other hand, compounds that display high quantities of natural products in the ELSD but do not display ions in the MS/MS can be prioritized for alternative methods of screening such as NMR analysis.

This dereplication method was used successfully to identify known bisindole alkaloids from the sponge *Spongosorites calcicola* and also identify two minor unknown brominated metabolites, the calcicamides, for isolation [106]. The feature-based molecular networking using GNPS was also used for the identification of a new family of highly branched thiolane derivatives, the nebulosins, from the marine annelid *Eupolymnia nebulosa*. The screening displayed three clusters of unknown natural products that constituted the major metabolites of this annelid [107].

While both types of analysis use high-throughput autosamplers, this method can be improved in the future through the addition of a splitter unit between the UHPLC and detectors to obtain results in a single run. It would also be advantageous to integrate the ELSD data with feature-based molecular networking to better understand clusters and the production of natural products.

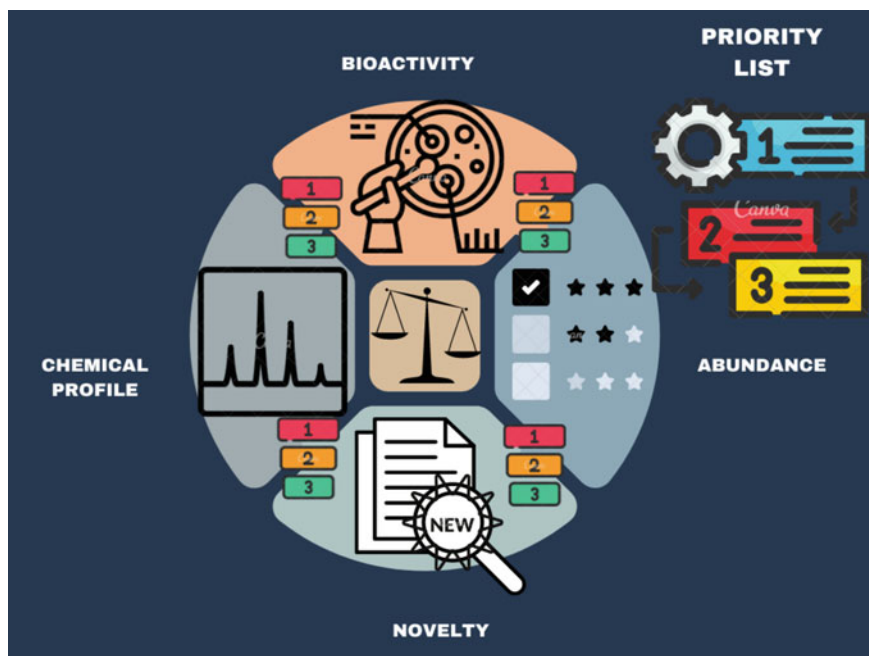
In terms of bioassays, the choice was made in Ireland to outsource the necessary biological screening, for several reasons:

- expertise in pharmacology is usually required to develop robust bioassays;
- biological screening is more efficient when using expensive automated robots;
- automatic robots require maintenance and dedicated space as usually available in companies; and
- as new targets appear, it would not be possible to embrace the full range of potential bioactivities targeted.

Our group therefore established a collaboration with Fundacion Medina in Spain, leaders in Europe in the biological screening of metabolites of natural origin, especially for rapid antimicrobial and cytotoxicity assessments.

### 3.4 Prioritization Strategies

Due to the high number of samples present in collections or biorepositories, a prioritization strategy is a key component of a successful process in marine biodiscovery. When thousands of samples are present in collections, a decision has to be made about where to start, and the biological assay should not be the only criterion taken into account. We decided to develop a prioritization strategy on each targeted bioassay (application) based on weights and scores assigned for (Fig. 3):



**Fig. 3** Prioritization strategy for marine biodiscovery

- the availability of the sample (biomass in Nature);
- new metabolites following chemical screening (dereplication) but also new species following taxonomy;
- ease of purification using chemical profiling; and
- biological activities evident.

The use of prioritization is very flexible, and the weights applied will depend on the main question addressed. Importantly, this strategy can only be based on the discovery of new species and chemical entities for the construction of repositories and advances of knowledge. In this case, the weight of novelty and chemical profile should be the highest. For more commercial applications, several options can be envisaged. Indeed, for a company working on the production of low-value products, the most important will be the availability of the resource and therefore the abundance (wild or cultured) will have the highest weight. Conversely, for a pharmaceutical company, the novelty and the bioactivity will have the highest weights. In the case of the IMBR, we gave very similar weights to the four different factors, as our main goal was to find bioactive natural products from a diverse range of species and that were easy to purify for metabolites (chemical profile).

At this point, it is important to extend the concept of marine biodiscovery beyond drug discovery. This high risk and long-term research question should not only correspond to one aspect of marine biodiscovery. Short-term applications of the extracts or

fractions as fertilizers or other agricultural application would contribute significantly to the blue economy. Applications requiring large biomass of the marine resource will be assigned the highest weights for prioritization.

This strategy has streamlined the selection of promising species for the identification of new bioactive marine natural products.

### **3.5 Natural Product Chemistry**

The purification and structural elucidation of marine natural products is the most time-consuming and expensive process of the marine biodiscovery pipeline. A large biomass is needed for isolation of active metabolite at a level that can facilitate the subsequent structure elucidation procedure. The presence of a highly potent minor metabolite in trace amounts can further complicate the isolation and structure determination process. However, recent technological advancements and the availability of higher field NMR spectrometers now mean that even natural products available in only sub-milligram quantities can be isolated and identified. This also means that smaller quantities of biomass (<10 g) can also undergo purification. These are especially important developments for marine biodiscovery for species obtained with a natural low biomass. Typically, the isolation and structure elucidation of marine natural products utilizes similar methodologies to terrestrial compounds, but with a fractionation process using a reversed-phase material such as vacuum-liquid chromatography.

#### **3.5.1 Purification of Marine Natural Products**

The complex and unique nature of marine organism extracts can make the isolation of the associated natural products difficult and highly time consuming. Moreover, most chemists consider the purification of natural products to be the most arduous part of the marine natural products pipeline. Therefore, samples are prioritized using their chemical and biological screening results, so that only the specimens that contain unknown metabolites and exhibit bioactivity undergo this laborious process. As unknown compounds are the target of most natural product isolations, there is no set procedure for purification and often multiple different subsequent methods must be used. However, the chemical profile and dereplication data obtained should aid in the selection of isolation methods. A usual purification process follows a process of large-scale, low-pressure fractionation, followed by subsequent small-scale, high-pressure HPLC separations of these fractions.

These large-scale fractionation methods include solid-phase extraction, vacuum-liquid chromatography, column chromatography, and liquid-liquid partitioning. These methods can utilize a variety of solid phases depending on the targeted molecules including a reversed phase (polar to non-polar molecules), normal phase (non-polar molecules), ion-exchange resins (ionic molecules), and size-exclusion

gels (polymers and water-soluble molecules). The primary aim of this first large-scale step in marine biodiscovery is to separate the secondary metabolites of interest from the bulk primary metabolites and ocean salts.

Subsequent smaller scale steps usually include higher resolution separation methods including column chromatography, thin-layer chromatography, high-performance liquid chromatography (HPLC), size-exclusion chromatography, and countercurrent chromatography. The primary aim of the high-resolution purification step is to obtain pure natural products of interest for chemical characterization. In marine natural product chemistry, preparative high-performance liquid chromatography utilizing a reversed phase (typically C<sub>18</sub> or phenyl bonded silica) is the most common and possibly the most useful method for isolating secondary metabolites.

Most modern preparative high-performance liquid chromatography systems are connected in series with a UV or a diode array detector to identify compounds eluting from the column and also a fraction collector for automated sample collection. Additionally, a number of new technologies for an increase in speed and automation of purification are now being incorporated into marine biodiscovery workflows. These include active high-performance liquid chromatography splitters in conjunction with other detectors (MS, ELSD) that can be used to purify compounds without a UV chromophore. Stop-flow high-performance liquid chromatography-NMR can be used to acquire sets of NMR data on compounds throughout a high-performance liquid chromatography purification [81], while peak-detecting fraction collectors that automatically collect compounds as they are detected eluting from the column are available.

While the processes used in marine biodiscovery closely match terrestrial methods, the difficulties inherent can differ slightly. Two main difficulties that may occur during the purification of marine extracts are the isolation of water-soluble small molecules, as the separation of these molecules from salts can often be difficult [108], and that marine natural products can be unstable and contain extremely labile functional groups (sulfates, polyenes).

### 3.5.2 Structure Elucidation

The most significant technological improvement in natural products chemistry has been the development of two-dimensional NMR spectroscopic experiments from the mid-1970s to mid-1980s. With the invention of 2D-NMR experiments, the time to elucidate structures of compounds decreased from years to days, and the quantity of product required decreased from hundreds of milligrams to just a few milligrams. Now, with modern NMR spectrometers, 2D-NMR data can be acquired on sub-milligram quantities of molecules. However, since the development of 2D-NMR pulse sequences, little has changed in the approach to determining the planar structure of marine natural products. Typically, planar structural elucidation is still performed through manual interpretation of <sup>1</sup>H NMR, <sup>13</sup>C NMR, COSY, HSQC, and HMBC data as well as the exact mass of a molecule. While NMR spectroscopy made the assignment of the planar structure “relatively” straight forward, the complete stereo



chemical assignments of marine natural products is still a challenging task. Traditionally, relative configurational assignments have been made through the use of coupling constants and nuclear Overhauser effect NMR experiments. In turn, absolute configurational assignments have been conducted using degradation and derivatization procedures, and X-ray crystallography. With the exception of NMR experiments, these methods often require quite large quantities of isolated compounds and can often be destructive to the sample.

More recently, a number of new computational tools have become more widely utilized for the structural assignment of marine natural products. These include computer-assisted structure elucidation and quantum chemical calculation of chiroptical properties and NMR spectra. In a large number of circumstances, a compound structural assignment using NMR spectroscopy is relatively straight forward, but there are instances when NMR assignments can be complicated, usually when a molecule has a low H:C ratio. In these situations, either the use of computer-assisted structure elucidation to generate a structure or the calculation of NMR spectrometric parameters to support structures are useful methods. Computer-assisted structure elucidation tools typically generate a number of different structures and give each structure a relative probability [109]. Common computer-assisted structure elucidation tools include ACD Structure Elucidator, Bruker CMC-se, and MNova Structure Elucidation. Currently, only a limited number of published marine natural products report the use of computer-assisted structure elucidation steps in their structural determination. However, assignments made using Gauge independent atomic orbital calculated NMR has become a popular method in marine natural products to support structural assignments [110].

Over the last 10 years, there has been a considerable shift in the methods used for configurational assignments in marine natural products. More recently, the use of time-dependent density functional theory for predicted electronic circular dichroism and Gauge independent atomic orbital NMR has dominated such assignments [111]. These time-dependent density functional theory methods have become popular in marine biodiscovery as a result of advancements in high performance computation early in the twenty-first century. These methods facilitate the reliable and fast prediction of both chiroptical properties (electronic circular dichroism, vibrational circular dichroism, and optical rotation) for absolute configuration and NMR for relative configuration [112]. The predicted data can be compared to experimental data on sub-milligrams of compounds to allow structural assignments. Statistical approaches including DP4 and its upgrades have become commonly used in marine biodiscovery for the calculation of a relative configuration with calculated NMR shifts or coupling constants [113–115]. A new tool, DP4-AI, recently has become available and combines both the automated computer-assisted structure elucidation 2D structure elucidation and NMR predicted relative configuration [116]. This method merging computer-assisted structure elucidation and DP4 is a significant improvement that is likely to make automated marine natural products assignment to the point of relative configuration more common.

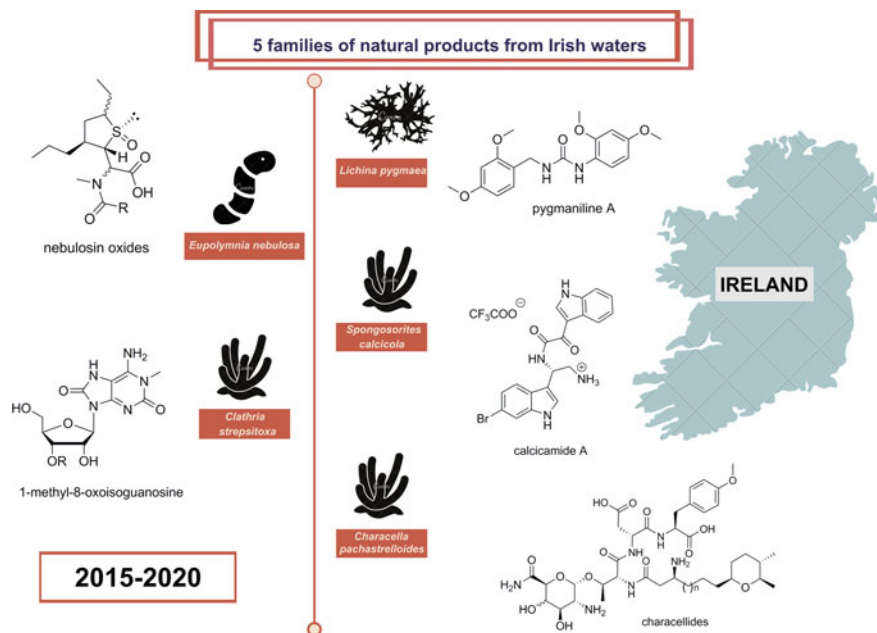
One other technique that has recently been developed and has so far been exclusively used for marine natural products structural elucidation is the crystalline sponge

X-ray crystallographic method [117]. This method allows X-ray crystallography to be performed on non-crystalline compounds by allowing them to form a complex with a host crystal. This method so far has been used to identify the structure of nanogram amounts of several red algal terpenoids. The limitations of this method are that only non-polar compounds can be absorbed by current crystals in non-polar solvents such as hexane. However, this procedure offers great promise for a new and robust method of absolute structural elucidation on sub-milligrams of sample.

### 3.5.3 Bioactive Natural Products from Irish Waters

During the last 5 years, the application of the workflow for marine biodiscovery in Ireland has led to a number of interesting compounds and five main publications (Fig. 4).

The first original aromatic urea derivatives were isolated from the intertidal lichen *Lichina pygmaea* [118]. Still in the intertidal area of the western coast of Ireland, the chemical study of the marine worm *Eupolyornia nebulosa* led to the discovery of an entire family of sulfur-containing metabolites [107]. These discoveries demonstrate the potential of overlooked marine organisms for marine biodiscovery that has been often restricted to sponges or tunicates. Then, the chemical studies of the subtidal sponges *Clathria strepsitoxa* and *Spongosorites calcicola* provided some interesting



**Fig. 4** Main families of natural products identified from Irish waters using the newly developed marine biodiscovery workflow

bioactive alkaloids [105, 119]. Finally, in the context of a program dedicated to the exploration of a deep-sea hotspot of bio- and chemodiversity, a new family of characellides was identified from the sponge *Chraccella pachastrelloides* [120].

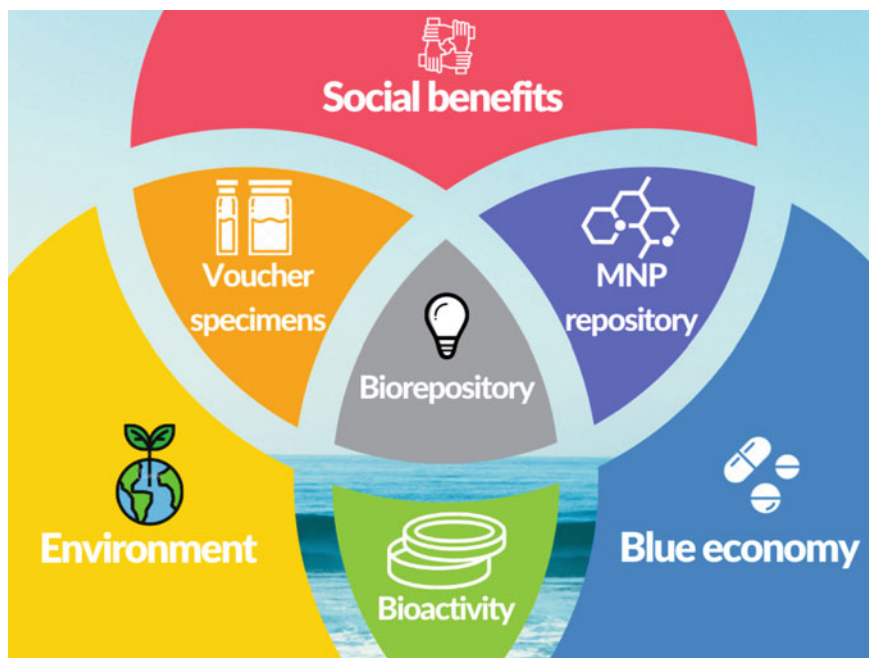
## 4 Concluding Remarks

Throughout this chapter, we have highlighted the importance of the construction of marine biomaterial repositories, especially in regions of the world with high marine biodiscovery potential, i.e., maritime developing nations. A number of advantages associated with the establishment and maintenance of such facilities have been detailed above. We believe that the Irish Marine Biomaterial Repository (MBR), which recently has been developed, can serve as an example for developing nations to follow, using best practices and data management systems already implemented over the past 5 years at the IMBR. A marine biomaterial repository requires a minimum investment from local authorities and can provide the following benefits:

- a marine biorepository represents the biological and chemical patrimony present in local species over time, and may be used to track changes in species distributions due to global change;
- it will contribute to building capacity in developing countries, especially in research areas such as taxonomy, where specialists are lacking;
- actions are available to quickly assess the potential of local marine bioresources against emerging pathogens or targets (e.g., SARS-COVID-2);
- it will provide new opportunities in the field of the blue economy for local communities through intellectual property associated with the genetic material stored; and
- it may also provide job opportunities and capacity building through training and technology transfer in member states or developing maritime nations.

Overall, the approach developed above contributes to the three pillars of sustainability as outlined below (Fig. 5).

The environmental impact of developing local marine biorepositories may be envisioned to be positive overall, as vouchers will be stored and processed at local facilities. Furthermore, local scientists will be consulted when access to samples is required and will be kept up-to-date on research and development being performed on local marine bioresources. Changes in the distribution of some species could lead to their protection by local and also international agencies. As only small quantities of samples are required for chemical and biological screening, local populations are not likely to be affected by initial bioprospecting collections. Subsequent collections of larger biomasses, however, will need to be carefully monitored by local authorities and communities so as to prevent population declines and possible extinction. It is important to include all the marine diversity present in the different coastal habitats, and not only focus on common groups such as sponges or tunicates. Some annelids or mollusks can also be promising in the marine biodiscovery workflow.



**Fig. 5** A sustainable marine biodiscovery approach

Finally, the possible economic value associated with some bioactive species has the potential to influence decision makers with regard to marine spatial planning and the establishment of special areas of conservation.

The presence of voucher specimens hosted by local entities ensures not only a positive social impact through capacity building and training of young scientists, but also interactions with worldwide experts and developing taxonomic skills in different groups of marine organisms. Members of the younger generation will have a better understanding of their rich marine biodiversity. On the other hand, the presence of chemical ensembles and regulations on access and benefit sharing could provide some positive economic inputs to the communities. Indeed, economic benefits will be shared between the industrial companies and the local communities through royalties on the licenses of the new biomolecules developed.

Finally, this type of marine organism organization will clearly contribute positively to the blue economy. The biomaterials produced are easily available for screening and the discovery of new applications is clearly accelerated. We recommend that marine biodiscovery should be more inclusive and cover both biological and biochemical discovery. The repositories should also integrate the large array of species available in an ecoregion and not only the most emblematic species like corals, sponges, and seaweeds. The future of marine biodiscovery will be bright if researchers in the field start building capacities in all regions of the world. The global trends that led us

from invertebrates to microbes and now a large investment in the seaweeds can be dangerous if a strong foundation is not built for the next generation of researchers.

**Acknowledgments** This project is financed by the Irish Government and the European Maritime and Fisheries Fund (EMFF) 2014–2020, Blue Growth & Marine Spatial Planning Scheme (SERV-20-MEFS-044), and also by the Marine Institute (Grant PDOC/19/02/01). We also acknowledge some fruitful discussions with colleagues of the Interreg project EBB.

## References

1. Eikeset AM, Mazzarella AB, Davíðsdóttir B, Klinger DH, Levin SA, Rovenskaya E, Stenseth NC (2018) What is blue growth? The semantics of “Sustainable Development” of marine environments. *Mar Policy* 87:177
2. Baslow MH (1969) Marine pharmacology. A study of toxins and other biologically active substances of marine origin. The Williams and Wilkins Company, Baltimore, USA
3. Molinski TF, Dalisay DS, Lievens SL, Saludes JP (2009) Drug development from marine natural products. *Nat Rev Drug Discov* 8:69
4. Gerwick WH, Moore BS (2012) Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem Biol* 19:1631
5. Jaspars M, De Pascale D, Andersen JH, Reyes F, Crawford AD, Ianora A (2016) The marine biodiscovery pipeline and ocean medicines of tomorrow. *J Mar Biol Assoc UK* 96:151
6. Sigwart JD, Blasiak R, Jaspars M, Jouffray J-B, Tasdemir D (2021) Unlocking the potential of marine biodiscovery. *Nat Prod Rep*. <https://doi.org/10.1039/D1030NP00067A>
7. Rae M, Folch H, Moniz MJB, Wolff CW, McCormack GP, Rindi F, Johnson MP (2013) Marine bioactivity in Irish waters. *Phytochem Rev* 12:555
8. Firsova D, Mahajan N, Solanki H, Morrow C, Thomas OP (2017) Current status and perspectives in marine biodiscovery. In: Paterson R, Lima N (eds) *Bioprospecting: success, potential and constraints*, vol 16. Topics in biodiversity and conservation. Springer, Cham, Switzerland, p 29
9. Leclère D, Obersteiner M, Barrett M, Butchart SHM, Chaudhary A, De Palma A, DeClerck FAJ, Di Marco M, Doelman JC, Dürrauer M, Freeman R, Harfoot M, Hasegawa T, Hellweg S, Hilbers JP, Hill SLL, Humpenöder F, Jennings N, Krisztin T, Mace GM, Ohashi H, Popp A, Purvis A, Schipper AM, Tabeau A, Valin H, van Meijl H, van Zeist WJ, Visconti P, Alkemade R, Almond R, Bunting G, Burgess ND, Cornell SE, Di Fulvio F, Ferrier S, Fritz S, Fujimori S, Grooten M, Harwood T, Havlík P, Herrero M, Hoskins AJ, Jung M, Kram T, Lotze-Campen H, Matsui T, Meyer C, Nel D, Newbold T, Schmidt-Traub G, Stehfest E, Strassburg BBN, van Vuuren DP, Ware C, Watson JEM, Wu W, Young L (2020) Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* 585:551
10. Blasiak R, Wynberg R, Grorud-Colvert K, Thambisetty S (2020) The ocean genome: conservation and the fair, equitable and sustainable use of marine genetic resources. World Resources Institute, Washington, DC, USA
11. Hillebrand H, Brey T, Gutt J, Hagen W, Metfies K, Meyer B, Lewandowska A (2018) Climate change: warming impacts on marine biodiversity. *Handbook on marine environment protection*, p 353
12. Di Marco M, Harwood TD, Hoskins AJ, Ware C, Hill SLL, Ferrier S (2019) Projecting impacts of global climate and land-use scenarios on plant biodiversity using compositional-turnover modelling. *Glob Change Biol* 25:2763
13. Ledger S (2020) Living planet report 2020 - bending the curve of biodiversity loss. WWF, Gland, Switzerland

14. Colella, JP, Agwanda BR, Khan FAA, Bates J, Carrión Bonilla CA, de la Sancha N, Dunnum JL, Ferguson AW, Greiman SW, Kiswele PK, Lessa EP, Soltis P, Thompson CW, Vanhove MPM, Webala PW, Weksler M, Joseph A (2020) Build international biorepository capacity. *Science* 370:773
15. Schindel DE, Cook JA (2018) The next generation of natural history collections. *PLoS Biol* 16:1
16. Thompson CW, Phelps KL, Allard MW, Cook JA, Dunnum JL, Ferguson AW, Gelang M, Khan FAA, Paul DL, Reeder DM, Simmons NB, Vanhove MPM, Webala PW, Weksler M, Kilpatrick CW (2021) Preserve a voucher specimen! The critical need for integrating natural history collections in infectious disease studies. *mBio* 12:1
17. Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. *Metabolites* 2:303
18. Pérez-Lloréns JL, Mouritsen OG, Rhatigan P, Cornish ML, Critchley AT (2020) Seaweeds in mythology, folklore, poetry, and life. *J Appl Phycol* 32:3157
19. Evans-Illidge EA, Logan M, Doyle J, Fromont J, Battershill CN, Ericson G, Wolff CW, Muirhead A, Kearns P, Abdo D, Kininmonth S, Llewellyn L (2013) Phylogeny drives large scale patterns in Australian marine bioactivity and provides a new chemical ecology rationale for future biodiscovery. *PLoS One* 8:1
20. Mühling M, Joint I, Willetts AJ (2013) The biodiscovery potential of marine bacteria: an investigation of phylogeny and function. *Microb Biotechnol* 6:361
21. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR (2021) Marine natural products. *Nat Prod Rep* 38:362
22. Bergmann W, Feeney RJ (1951) Contributions to the study of marine products. XXXII. The nucleosides of sponges. I. *J Org Chem* 16:981
23. Planes S, Allemand D, Agostini S, Banaigs B, Boissin E, Boss E, Bourdin G, Bowler C, Douville E, Flores JM, Forcioli D, Furla P, Galand PE, Ghiglione J-F, Gilson E, Lombard F, Moulin C, Pesant S, Poulain J, Reynaud S, Romac S, Sullivan MB, Sunagawa S, Thomas OP, Troublé R, de Vargas C, Vega Thurber R, Voolstra CR, Wincker P, Zoccola D (2019) The Tara Pacific expedition—a pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLoS Biol* 17:e3000483
24. Miller SE, Barrow LN, Ehlman SM, Goodheart JA, Greiman SE, Lutz HL, Misiewicz TM, Smith SM, Tan M, Thawley CJ, Cook JA, Light JE (2020) Building natural history collections for the twenty-first century and beyond. *Bioscience* 70:674
25. Edwards JL, Lane MA, Nielsen ES (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289:2312
26. Page LM, Macfadden BJ, Fortes JA, Soltis PS, Riccardi G (2015) Digitization of biodiversity collections reveals biggest data on biodiversity. *Bioscience* 65:841
27. Lavoie C (2013) Biological collections in an ever changing world: herbaria as tools for biogeographical and environmental studies. *Perspect Plant Ecol Evol Syst* 15:68
28. Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC, Edwards SV, Davis CC (2020) Digitization and the future of natural history collections. *Bioscience* 70:243
29. Day JG, Stacey GN (2008) Biobanking. *Mol Biotechnol* 40:202
30. Funk VA, Gostel M, Devine A, Kelloff CL, Wurdack K, Tuccinardi C, Radosavljevic A, Peters M, Coddington J (2017) Guidelines for collecting vouchers and tissues intended for genomic work (Smithsonian Institution): botany best practices. *Biodivers Data J* 5:e11625
31. De Souza YG, Greenspan JS (2013) Biobanking past, present and future: responsibilities and benefits. *AIDS* 27:303
32. Hueté-Pérez JA, Mendoza-Ramírez EN (2012) Genomic biorepository of coastal marine species in Estero Padre Ramos and Estero Real Nicaragua. *Encuentro* 93:6
33. Gabrielsen KL (2012) Marbank – a biobank of arctic marine organisms. *Planta Med* 78:PL18
34. Lallier LE, McMeel O, Greiber T, Vanagt T, Dobson ADW, Jaspars M (2014) Access to and use of marine genetic resources: understanding the legal framework. *Nat Prod Rep* 31:612

35. Morgera E, Tsioumani E, Buck M (2014) Unraveling the Nagoya Protocol: a commentary on the Nagoya Protocol on access and benefit-sharing to the Convention on Biological Diversity. Brill & Nijhoff, Leiden, The Netherlands. <https://doi.org/10.1163/9789004217188>
36. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* 55:595
37. Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic barcode gap discovery for primary species delimitation. *Mol Ecol* 21:1864
38. Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869
39. De Queiroz K (2007) Species concepts and species delimitation. *Syst Biol* 56:879
40. Cachet N, Genta-Jouve G, Ivanisevic J, Chevaldonné P, Sinniger F, Culioli G, Pérez T, Thomas OP (2015) Metabolomic profiling reveals deep chemical divergence between two morphotypes of the zoanthid *Parazoanthus axinellae*. *Sci Rep* 5:1
41. Villamor A, Signorini LF, Costantini F, Terzin M, Abbiati M (2020) Evidence of genetic isolation between two Mediterranean morphotypes of *Parazoanthus axinellae*. *Sci Rep* 10:1
42. Ruiz C, Ivanišević J, Chevaldonné P, Ereskovsky AV, Boury-Esnault N, Vacelet J, Thomas OP, Pérez T (2015) Integrative taxonomic description of *Plakina kanaky*, a new polychromatic sponge species from New Caledonia (Porifera: Homoscleromorpha). *Mar Ecol* 36:1129
43. Reddy MM, De Clerck O, Leliaert F, Anderson RJ, Bolton JJ (2020) An appraisal of the genus *Pyropia* (Bangiales, Rhodophyta) from southern Africa based on a multi-gene phylogeny, morphology and ecology, including the description of *Pyropia meridionalis* sp. nov. *SAfr J Bot* 131:18
44. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA* 101:14812
45. Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Syst Biol* 54:852
46. Rutz A, Dounoue-Kubo M, Ollivier S, Bisson J, Bagheri M, Saesong T, Ebrahimi SN, Ingkaninan K, Wolfender JL, Allard PM (2019) Taxonomically informed scoring enhances confidence in natural products annotation. *Front Plant Sci* 10:1
47. Hirata UD (1986) Halichondrins—antitumor polyether macrolides from a marine sponge. *Pure Appl Chem* 58:701
48. Marine Biodiscovery Ireland (2018) <http://www.imbd.ie/>. Accessed 14 March 2021
49. D'Archino R, Nelson WA, Sutherland JE (2020) Unnamed for over 30 years: *Gigartina falshawiae* sp. nov. (Gigartinales, Rhodophyta) and its confusion with *Iridaea tuberculosa* in New Zealand. *Phycologia* 59:45
50. Camp D, Davis RA, Campitelli M, Ebdon J, Quinn RJ (2012) Drug-like properties: guiding principles for the design of natural product libraries. *J Nat Prod* 75:72
51. SciFinder American Chemical Society. <https://scifinder.cas.org/scifinder/view/scifinder/scifinderExplore.jsf>
52. PubChem National Library of Medicine, National Institutes of Health. <https://pubchem.ncbi.nlm.nih.gov/>
53. ChemSpider Royal Society of Chemistry. <https://www.chemspider.com/>
54. Chemical Entities of Biological Interest (ChEBI) European Molecular Biology Laboratory. <https://www.ebi.ac.uk/chebi/>
55. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, Bellis L, Overington JP (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucl Acids Res* 43:W612
56. Reaxys Elsevier Life Sciences IP Limited. <https://www.reaxys.com>
57. Dictionary of Natural Products 29.2. CRC Press, Taylor & Francis Group. <http://dnp.chemnetbase.com>



58. Sorokina M, Steinbeck C (2020) Review on natural products databases: where to find data in 2020. *J Cheminform* 12:20
59. Blunt J, Munro M, Upjohn M (2012) The role of databases in marine natural products research. In: Fattorusso E, Gerwick WH, Tagliatalata-Scafati O (eds) *Handbook of marine natural products*. Springer, Netherlands, Dordrecht, p 389
60. MarinLit. Royal Society of Chemistry. <http://pubs.rsc.org/marinlit/>
61. van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, Neto FC, Castaño-Espriu L, Chang C, Clark TN, Cleary Little JL, Delgadillo DA, Dorrestein PC, Duncan KR, Egan JM, Galey MM, Haeckl FPJ, Hua A, Hughes AH, Iskakova D, Khadilkar A, Lee J-H, Lee S, LeGrow N, Liu DY, Macho JM, McCaughey CS, Medema MH, Neupane RP, O'Donnell TJ, Paula JS, Sanchez LM, Shaikh AF, Soldatou S, Terlouw BR, Tran TA, Valentine M, van der Hoof JJJ, Vo DA, Wang M, Wilson D, Zink KE, Linington RG (2019) The Natural Products Atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent Sci* 5:1824
62. Jones MR, Pinto E, Torres MA, Dörr F, Mazur-Marzec H, Szubert K, Tartaglione L, Dell'Aversano C, Miles CO, Beach DG, McCarron P, Sivonen K, Fewer DP, Jokela J, Janssen EM-L (2020) Comprehensive database of secondary metabolites from cyanobacteria. *BioRxiv:2020.2004.2016.038703*
63. Moubock AFA, Gao M, Qaseem A, Li J, Kirchner Pascal A, Ndingkokhar B, Bekono BD, Simoben CV, Babiaka Smith B, Malange YI, Sauter F, Zierep P, Ntie-Kang F, Günther S (2020) StreptomeDB 3.0: an updated compendium of Streptomycetes natural products. *Nucl Acids Res* 49:D600
64. Fekete S, Schappler J, Veuthey J-L, Guillaume D (2014) Current and future trends in UHPLC. *TrAC, Trends Anal Chem* 63:2
65. Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR (2012) Marine natural products. *Nat Prod Rep* 29:144
66. Seger C, Sturm S, Stuppner H (2013) 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* 30:970
67. Esquenazi E, Dorrestein PC, Gerwick WH (2009) Probing marine natural product defenses with DESI-imaging mass spectrometry. 106:7269
68. Yarnold JE, Hamilton BR, Welsh DT, Pool GF, Venter DJ, Carroll AR (2012) High resolution spatial mapping of brominated pyrrole-2-aminoimidazole alkaloids distributions in the marine sponge *Stylissa flabellata* via MALDI-mass spectrometry imaging. *Mol Biosyst* 8:2249
69. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya PCA, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJM, Granatosky EA, Glukhov E, Ryyffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BØ, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 34:828



70. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC (2013) Molecular networking as a dereplication strategy. *J Nat Prod* 76:1686
71. Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard P-M, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kamenfk Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin HC, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweiger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hoof JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC (2020) Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* 17:905
72. Bracegirdle J, Robertson LP, Hume PA, Page MJ, Sharrock AV, Ackerley DF, Carroll AR, Keyzers RA (2019) Lamellarin sulfates from the Pacific tunicate *Didemnum ternerratum*. *J Nat Prod* 82:2000
73. Yu H-B, Glukhov E, Li Y, Iwasaki A, Gerwick L, Dorrestein PC, Jiao B-H, Gerwick WH (2019) Cytotoxic microcolin lipopeptides from the marine cyanobacterium *Moorea producens*. *J Nat Prod* 82:2608
74. Teta R, Della Sala G, Esposito G, Via CW, Mazzoccoli C, Piccoli C, Bertin MJ, Costantino V, Mangoni A (2019) A joint molecular networking study of a *Smenospongia* sponge and a cyanobacterial bloom revealed new antiproliferative chlorinated polyketides. *Org Chem Front* 6:1762
75. Djoumbou-Feunang Y, Pon A, Karu N, Zheng J, Li C, Arndt D, Gautam M, Allen F, Wishart DS (2019) CFM-ID 3.0: Significantly improved ESI-MS/MS prediction and compound identification. *Metabolites* 9:72
76. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S (2010) In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform* 11:148
77. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci USA* 112:12580
78. Deyon-Jung L, Morice C, Chéry F, Gay J, Langer T, Frantz M-C, Rozot R, Dalko-Csiba M (2016) Fragment pharmacophore-based in silico screening: a powerful approach for efficient lead discovery. *MedChemComm* 7:506
79. Buedenbender L, Habener LJ, Grkovic T, Kurtböke DÍ, Duffy S, Avery VM, Carroll AR (2018) HSQC-TOCSY Fingerprinting for prioritization of polyketide- and peptide-producing microbial isolates. *J Nat Prod* 81:957
80. Zhang C, Idelbayev Y, Roberts N, Tao Y, Nannapaneni Y, Duggan BM, Min J, Lin EC, Gerwick EC, Cottrell GW, Gerwick WH (2017) Small molecule accurate recognition technology (SMART) to enhance natural products research. *Sci Rep* 7:14243
81. Dias DA, Urban SJJoss (2009) Application of HPLC-NMR for the rapid chemical profiling of a southern Australian sponge, *Dactylospongia* sp. 32:542
82. Jaroszewski JW (2005) Hyphenated NMR methods in natural products research, 2: HPLC-SPE-NMR and other new trends in NMR hyphenation. *Planta Med* 71:795
83. Timmers MA, Dias DA, Urban S (2012) Application of HPLC-NMR in the identification of plocamenone and isoplocamenone from the marine red alga *Plocamium angustum*. *Mar Drugs* 10:2089
84. Urban S, Timmers M (2013) HPLC-NMR Chemical profiling and dereplication studies of the marine brown alga, *Cystophora torulosa*. *Nat Prod Commun* 8:715
85. Zani CL, Carroll AR (2017) Database for rapid dereplication of known natural products using data from MS and fast NMR experiments. *J Nat Prod* 80:1758

86. Reher R, Kim HW, Zhang C, Mao HH, Wang M, Nothias L-F, Caraballo-Rodriguez AM, Glukhov E, Teke B, Leao T, Alexander KL, Duggan BM, Van Everbroeck EL, Dorrestein PC, Cottrell GW, Gerwick WH (2020) A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. *J Am Chem Soc* 142:4114
87. Mohimani H, Pevzner PA (2016) Dereplication, sequencing and identification of peptidic natural products: from genome mining to peptidogenomics to spectral networks. *Nat Prod Rep* 33:73
88. Udworthy DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* 104:10376
89. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hoof JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH (2020) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucl Acids Res* 48:D454
90. Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL (2020) An interpreted atlas of biosynthetic gene clusters from 1000 fungal genomes. *BioRxiv* 2020.2009.2021.307157
91. Schniete JK, Selem-Mojica N, Birke AS, Cruz-Morales P, Hunter IS, Barona-Gomez F, Hoskisson PA (2021) ActDES – a curated actinobacterial database for evolutionary studies. *Microb Genom* 7:498
92. Blin K, Shaw S, Kautsar SA, Medema MH, Weber T (2020) The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucl Acids Res* 49:D639
93. Palaniappan K, Chen I-MA, Chu K, Ratner A, Seshadri R, Kyrpides NC, Ivanova NN, Mouncey NJ (2019) IMG-ABC v.5.0: an update to the IMG/atlas of biosynthetic gene clusters knowledgebase. *Nucl Acids Res* 48:D422
94. Skinnider MA, Johnston CW, Gunabalasingam M, Merwin NJ, Kieliszek AM, MacLellan RJ, Li H, Ranieri MRM, Webster ALH, Cao MPT, Pfeifle A, Spencer N, To QH, Wallace DP, Dejong CA, Magarvey NA (2020) Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat Commun* 11:6058
95. Agrawal P, Mohanty D (2020) A machine learning-based method for prediction of macrocyclization patterns of polyketides and non-ribosomal peptides. *Bioinformatics* Oct3, btaa851
96. Robinson SL, Terlouw BR, Smith MD, Pidot SJ, Stinear TP, Medema MH, Wackett LP (2020) Global analysis of adenylate-forming enzymes reveals lactone biosynthesis pathway in pathogenic *Nocardia*. *J Biol Chem* 295:14826
97. Saporito RA, Donnelly MA, Norton RA, Garraffo HM, Spande TF, Daly JW (2007) Oribatid mites as a major dietary source for alkaloids in poison frogs. *Proc Natl Acad Sci USA* 104:8885
98. Fahey SJ, Garson MJ (2002) Geographic variation of natural products of tropical nudibranch *Asteronotus cespitosus*. *J Chem Ecol* 28:1773
99. Swinney DC, Anthony J (2011) How were new medicines discovered? *Nat Rev Drug Disc* 10:507
100. Wilson BAP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR (2020) Creating and screening natural product libraries. *Nat Prod Rep* 37:893
101. Futamura Y, Muroi M, Osada H (2013) Target identification of small molecules based on chemical biology approaches. *Mol Biosyst* 9:897
102. Liu M, Van Voorhis WC, Quinn RJ (2021) Development of a target identification approach using native mass spectrometry. *Sci Rep* 11:2387
103. Strömstedt AA, Felth J, Bohlin L (2014) Bioassays in natural product research – strategies and methods in the search for anti-inflammatory and antimicrobial activity. *Phytochem Anal* 25:13
104. Nothias L-F, Nothias-Esposito M, da Silva R, Wang M, Protsyuk I, Zhang Z, Sarvepalli A, Leyssen P, Touboul D, Costa J, Paolini J, Alexandrov T, Litaudon M, Dorrestein PC (2018) Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. *J Nat Prod* 81:758

105. Wolfender J-L, Litaudon M, Touboul D, Queiroz EF (2019) Innovative omics-based approaches for prioritisation and targeted isolation of natural products – new strategies for drug discovery. *Nat Prod Rep* 36:855
106. Jennings LK, Khan NMD, Kaur N, Rodrigues D, Morrow C, Boyd A, Thomas OP (2019) Brominated bisindole alkaloids from the Celtic sea sponge *Spongosorites calcicola*. *Molecules* 24:3890
107. Calabro K, Jennings LK, Lasserre P, Doohan R, Rodrigues D, Reyes F, Ramos C, Thomas OP (2020) Nebulosins: trisubstituted thiolane natural products from the northeastern Atlantic annelid *Eupolyornia nebulosa*. *J Org Chem* 85:14026
108. Sarker SD, Latif Z, Gray AI (eds) (2006) Natural products isolation, 2 edn. (Methods in Biotechnology, vol 20). Humana Press Inc., Totowa, NJ, USA
109. Burns DC, Mazzola EP, Reynolds WF (2019) The role of computer-assisted structure elucidation (CASE) programs in the structure elucidation of complex natural products. *Nat Prod Rep* 36:919
110. Guillen PO, Jaramillo KB, Genta-Jouve G, Sinniger F, Rodriguez J, Thomas OP (2017) Terrazoanthines, 2-aminoimidazole alkaloids from the tropical eastern Pacific zoantharian *Terrazoanthus onoi*. *Org Lett* 19:1558
111. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR (2020) Marine natural products. *Nat Prod Rep* 37:175
112. Mándi A, Kurtán T (2019) Applications of OR/ECD/VCD to the structure elucidation of natural products. *Nat Prod Rep* 36:889
113. Smith SG, Goodman JM (2010) Assigning stereochemistry to single diastereoisomers by GIAO NMR calculation: the DP4 probability. *J Am Chem Soc* 132:12946
114. Grimblat N, Zanardi MM, Sarotti AM (2015) Beyond DP4: an improved probability for the stereochemical assignment of isomeric compounds using quantum chemical calculations of NMR shifts. *J Org Chem* 80:12526
115. Grimblat N, Gavín JA, Hernández Daranas A, Sarotti AM (2019) Combining the power of *J* coupling and DP4 analysis on stereochemical assignments: the *J*-DP4 Methods. *Org Lett* 21:4003
116. Howarth A, Ermanis K, Goodman JM (2020) DP4-AI automated NMR data analysis: straight from spectrometer to structure. *Chem Sci* 11:4351
117. Kersten RD, Lee S, Fujita D, Pluskal T, Kram S, Smith JE, Iwai T, Noel JP, Fujita M, Weng J-K (2017) A red algal bourbonane sesquiterpene synthase defined by microgram-scale NMR-coupled crystalline sponge X-ray diffraction analysis. *J Am Chem Soc* 139:16838
118. Mahajan N, Chadda R, Calabro K, Solanki H, O'Connell E, Murphy PV, Thomas OP (2017) Isolation and synthesis of pygmanilines, phenylurea derivatives from the northeastern Atlantic lichen *Lichina pygmaea*. *Tetrahedron Lett* 58:1237
119. Firsova D, Calabro K, Lasserre P, Reyes F, Thomas OP (2017) Isoguanosine derivatives from the northeastern Atlantic sponge *Clathria (Microciona) strepsitoxa*. *Tetrahedron Lett* 58:4652
120. Afoullouss S, Calabro K, Genta-Jouve G, Gegunde S, Alfonso A, Nesbitt R, Morrow C, Alonso E, Botana LM, Allcock AL, Thomas OP (2019) Treasures from the deep: Characellides as anti-inflammatory lipoglycotripeptides from the sponge *Characella pachastrelloides*. *Org Lett* 21:246



**Maggie M. Reddy** is a postdoctoral researcher at the National University of Galway, Ireland. Her passion for the sea began as a little girl exploring the seashores of southern KwaZulu-Natal. She graduated with a Ph.D. degree in Marine Biology from the University of Cape Town in South Africa working in collaboration with the University of Ghent in Belgium. Her doctoral research revealed extensive species diversity including new and cryptic species in the Bangiales, an economically valuable group of seaweeds. Her main research focuses on marine biodiversity, taxonomy and molecular systematics. She has extensive experience in marine biodiversity collections having sampled much of the southern African coast from southern Mozambique to southern Angola.



**Laurence Jennings** obtained a Bachelor of Marine Science degree from Griffith University, Gold Coast, Australia. He then continued his studies completing a Ph.D. degree in marine natural products chemistry under the supervision of Prof. Anthony R. Carroll. Since 2018 he has held a postdoctoral research position at the National University of Ireland Galway within the group of Prof Olivier P. Thomas. This postdoctoral position, funded by the Marine Institute, has focused on the development of a biodiscovery laboratory and the investigation of secondary metabolites with pharmacological properties from Irish coastal organisms. His main research interests are the isolation and identification of bioactive marine natural products. He has a particular interest in the structural elucidation of complex natural products using NMR spectroscopic and computational methods.



**Olivier P. Thomas** has an Established Chair in Marine Biodiscovery at the National University of Ireland, Galway (NUIG). In 2002, he obtained a Ph.D. degree in medicinal chemistry from the University of René Descartes Paris. After two years as a post-doctoral researcher in biochemistry at the CEA Saclay in France, he was recruited as a lecturer in marine natural product chemistry at the University of Côte d'Azur, France. Following his passion for the oceans and their hidden treasures, he moved to the island of Ireland where he is leading the marine biodiversity group at NUIG. The research focus of the laboratory is the chemistry of marine metabolites with an emphasis on drug discovery but also biotoxins and chemical ecology through the use of environmental metabolomics.