

Review

# Computational Approaches to Enzyme Inhibition by Marine Natural Products in the Search for New Drugs

Federico Gago 

Department of Biomedical Sciences & IQM-CSIC Associate Unit, School of Medicine and Health Sciences, University of Alcalá, E-28805 Madrid, Alcalá de Henares, Spain; federico.gago@uah.es; Tel.: +34-918854514

**Abstract:** The exploration of biologically relevant chemical space for the discovery of small bioactive molecules present in marine organisms has led not only to important advances in certain therapeutic areas, but also to a better understanding of many life processes. The still largely untapped reservoir of countless metabolites that play biological roles in marine invertebrates and microorganisms opens new avenues and poses new challenges for research. Computational technologies provide the means to (i) organize chemical and biological information in easily searchable and hyperlinked databases and knowledgebases; (ii) carry out cheminformatic analyses on natural products; (iii) mine microbial genomes for known and cryptic biosynthetic pathways; (iv) explore global networks that connect active compounds to their targets (often including enzymes); (v) solve structures of ligands, targets, and their respective complexes using X-ray crystallography and NMR techniques, thus enabling virtual screening and structure-based drug design; and (vi) build molecular models to simulate ligand binding and understand mechanisms of action in atomic detail. Marine natural products are viewed today not only as potential drugs, but also as an invaluable source of chemical inspiration for the development of novel chemotypes to be used in chemical biology and medicinal chemistry research.

**Keywords:** enzyme inhibitors; databases; cheminformatics



**Citation:** Gago, F. Computational Approaches to Enzyme Inhibition by Marine Natural Products in the Search for New Drugs. *Mar. Drugs* **2023**, *21*, 100. <https://doi.org/10.3390/md21020100>

Received: 21 December 2022

Revised: 26 January 2023

Accepted: 28 January 2023

Published: 30 January 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Overview

Both pharmacology and basic cell biology have traditionally benefited from the continuous identification and biochemical characterization of active principles obtained from natural sources. The scanty primitive chemical libraries of natural products (NPs), consisting mostly of the alkaloids and heterosides isolated from terrestrial plants that provided the foundations of modern pharmacology [1], were progressively enriched with a multitude of small- to medium-sized molecules present in numerous living creatures, both big and small, including those inhabiting seas and oceans, which together make up a huge water mass that covers >70% of Earth's total surface and hosts ~80% of all living species [2]. Nonetheless, and despite a notable renaissance in recent years [3,4], the list of marine natural products (MNPs) that have been approved or are currently found in the global marine pharmaceutical clinical pipeline (<https://www.midwestern.edu/departments/marinepharmacology/clinical-pipeline>, accessed on 20 December 2022) is still very limited, and only a few of these drugs actually target an enzyme.

The vastness of the largely unexplored chemical space [5] existing in marine environments poses daunting challenges in terms of (i) sample recollection, (ii) compound isolation, (iii) chemical characterization, (iv) evaluation in as many biochemical and/or biological assays as possible, preferably using validated targets and high-throughput state-of-the-art technologies [6], and (v) the identification and validation of pharmacologically relevant targets. Given the precedents of successful marine leads as a source of useful medicinal agents and biochemical probes, it can be argued that it makes sense to continue exploiting over four billion years of evolution in nature's combinatorial chemistry, often subjected

to unique ecological pressures and nutrient availability, that led to selective survival advantages in the producing organisms [7]. The best studied phylogenetically diverse living beings from marine habitats include green, brown, and red algae; sponges; coelenterates (i.e., jellyfishes, corals, and sea anemones); bryozoans (i.e., invertebrates known as moss animals); the Ascidiacea class (commonly known as the ascidians, tunicates, and sea squirts); mollusks; echinoderms; phytoplankton; and innumerable bacteria and fungi. Secondary metabolites are specialized organic compounds that are not considered essential for normal growth or reproduction (under laboratory culture conditions) but instead play roles in evolution, communication (as chemical cues), and competition, or else appear to be used as chemical weaponry against prey or natural enemies in their natural environments. MNPs often feature unique scaffolds and carbocyclic skeletons, and many have been discovered following their bioassay-guided isolation, although the paucity of material usually prevents the full profiling of bioactivity [8], which is often limited to some rudimentary tests (e.g., phenotype-oriented antimicrobial or cytotoxic assays [9], and/or inhibitory activity against one enzyme or a limited set of enzymes). In this respect, it has been pointed out that micromolar activities detected in extracts should be critically analyzed because of potential artefactual assay readouts due to unspecific aggregation [10], hence the recommendation to use  $\beta$ -lactamase and malate dehydrogenase as counter-screening enzymes [11], among other precautionary measures.

Recent progress in understanding the genetic basis of MNP biosynthesis and the ever-increasing availability of genomic information have created unique opportunities to develop sequence-based approaches for the discovery of novel bioactive molecular entities [12]. Polyketide synthases (PKSs) and multimodular nonribosomal peptide synthetases (NRPS) stand out among the enzymes that are ultimately responsible for the highly efficient synthesis of three large subclasses of important NPs (PKs, NRPs, and PK/RP or NRP/PK hybrids) [13] through the concerted assembly of relatively simple carboxylic acid and amino acid building blocks, respectively [14,15]. Type I PKSs consist of multiple modules, with each module minimally containing three core domains: acyltransferase (AT) domain, ketosynthase (KS) domain, and thiolation (T) domain [aka acyl carrier protein (ACP) domain] [16]. These (mega)enzymes are encoded in biosynthetic gene clusters (BGCs), which have been identified for hundreds of bacterial and fungal metabolites and are highly evolved for horizontal exchange [17]. Besides, attention continues to be drawn to two facts that have significantly expanded the area of MNP research, namely (i) that some isolated MNPs are bioaccumulated in the target organism from dietary sources, e.g., algae [18]; and (ii) that a significant number of MNPs are actually produced by microbes and/or microbial interactions with the “host from whence it was isolated” [8]. The growing emphasis on the study of compounds from microbial sources (both terrestrial and marine) has been fueled by interest in (i) the central role that microorganisms play in mediating both interspecies interactions and host-microbe relationships [19]; and (ii) their natural ability to produce ribosomally synthesized and post-translationally modified peptides (RiPPs), which often contain noncanonical amino acids and structural motifs that give rise to a currently under-represented class of biologically active molecules [20,21].

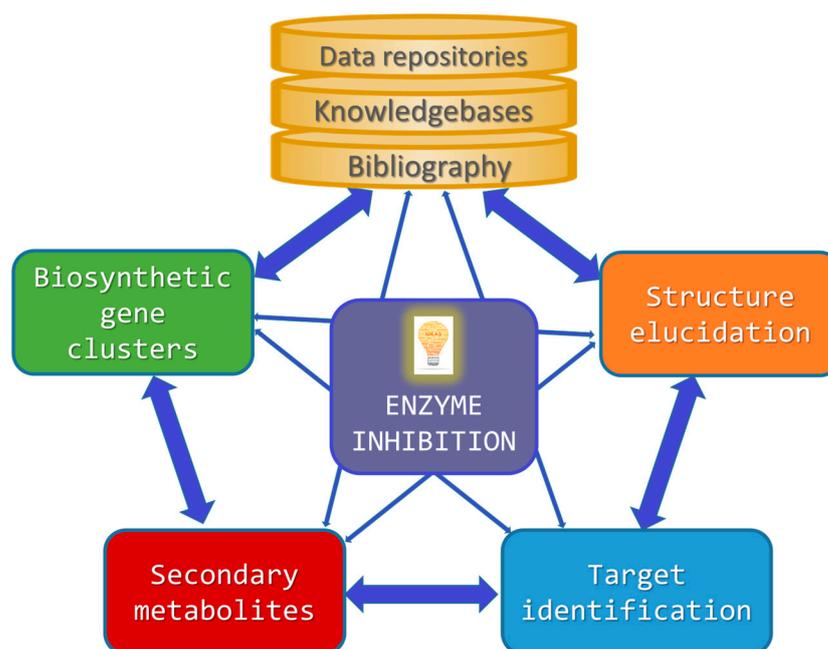
Modern science (and the world at large) is overly dependent on computer and internet technologies. Computers have a long history in data management, as well as in information storage, processing, retrieval, and dissemination, and for these purposes their use has expanded enormously in recent years and has contributed to shaping the current research landscapes in bioscience and biomedicine as we know them today. The World Wide Web has become a central source of (i) information on all possible subjects that is stored and (ideally) curated in extensively hyperlinked databases; (ii) educational and research tools; and (iii) services that are intended to make life easier not only for the general public, but also for scientists, including those devoted to chemical biology, medicinal chemistry, and drug discovery. Devices ranging from pocket computers, also known as mobile or cellular phones which have superseded earlier personal digital assistants (PDA), to tablets, laptops, desktops, mainframes, and supercomputers dominate many aspects of

our lives and complement human skills in numerous applications designed to utilize an ever-growing torrent of biological and chemical data in effective manners. While this is the driving force behind the increasing use of high-performance computing, machine learning and artificial intelligence for processing tons of data in a way that compensates for the inherent constraints of human cognition [22], better-informed decision making in drug discovery and development still largely relies (or so I like to believe) on the power of human judgement and life-long expertise.

The concise Guide to Pharmacology (<https://www.guidetopharmacology.org/>; latest release 13 October 2022, accessed on 20 December 2022) presented by the International Union of Basic and Clinical Pharmacology (IUPHAR) and the British Pharmacological Society (BPS) includes enzymes (Nature's catalysts essential to the chemistry of life) as one of the six major classes of pharmacological targets, the others being G protein-coupled receptors, ion channels, nuclear hormone receptors, catalytic receptors, and transporters (including the very large SLC superfamily of solute carriers) [23]. Over one thousand distinct human enzymes are described in the Universal Protein Knowledgebase (UniProtKB) [24], therefore representing almost half of all current human targets. Fortunately, the three-dimensional (3D) structures of many of these enzymes or closely related counterparts from other species—both in their apo forms and in complexes with ligands—have been solved and deposited in the Worldwide Protein Data Bank (wwPDB) [25], a continuously enlarging global repository established in 1971. These 3D structures facilitate the elucidation of functional mechanisms, aid in understanding the binding mode of inhibitors, and enable virtual screening (VS) and structure-based drug design (SBDD) technologies. For other enzymes of interest, we still depend on several homology modeling approaches [26], neural network-based models, such as those generated by AlphaFold [27], and artificial intelligence, which was recently employed to build the ESM Metagenomic Atlas (<https://esmatlas.com/>, accessed on 20 December 2022), with more than 617 million structures from all kingdoms of life [28].

Following the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB, <https://www.qmul.ac.uk/sbcs/iubmb/enzyme/>; accessed on 20 December 2022), the wwPDB assigns Enzyme Commission (EC) numbers to protein chains in macromolecular structures according to the type of chemical reaction that they catalyze. The main classes are oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5), ligases (EC 6), and translocases (EC 7), with subclasses (with up to 4 digits) being defined on the basis of the specific donors and receptors of chemical groups that participate in the reactions and additional considerations. The main collection of functional enzyme and metabolism data is possibly BRENDA (<https://www.brenda-enzymes.org/>, accessed on 20 December 2022), which was established in 1987 and selected as an ELIXIR Core Data Resource [29] in 2018 [30]. In addition, the merging of MACiE (Mechanism, Annotation and Classification in Enzymes), a database of enzyme mechanisms, and CSA (Catalytic Site Atlas), a database of catalytic sites of enzymes, has resulted in the M-CSA Mechanism and Catalytic Site Atlas (<http://www.ebi.ac.uk/thornton-srv/m-csa/browse/?sort=ec>, accessed on 20 December 2022) [31], which consolidates a body of knowledge on enzyme structures, gene sequences, reaction mechanisms, metabolic pathways, and kinetic data that any researcher working on enzyme inhibitors should be familiar with.

Building on this introductory background information, the following sections will separately cover each of the abovementioned aspects for which specialized computer technologies have been developed in the field of enzyme inhibition by MNPs (Figure 1).



**Figure 1.** Simplified interrelationship diagram illustrating how the design/identification of enzyme inhibitors from marine sources can benefit from the use of computer-aided methods.

## 2. Bibliographical Sources and Virtual NP Databases

Chemical libraries encompassing millions of compounds include the Chemical Abstracts Service (CAS) REGISTRY database (<http://www.cas.org/expertise/cascontent/registry/index.html>, accessed on 20 December 2022), which is updated on a daily basis and contains >250,000 NPs out of >150 million chemical substances, PubChem (including PC-Substance, PCCompound, and PCBioAssay) [32], ChEMBL (a manually curated database of >2,300,000 bioactive molecules with drug-like properties, last update July 2022) [33], and ChemSpider (with various levels of partial to complete stereochemistry) [34]. The free-to-access resource DrugBank is a web-enabled database (<https://go.drugbank.com/>, accessed on 20 December 2022) that incorporates comprehensive molecular information about drugs, their mechanisms, their interactions, and their targets. First described in 2006 as a knowledgebase for drugs, drug actions, and drug targets [35], DrugBank has evolved over time in response to improvements in web standards and changing needs for drug research and development. The latest update, DrugBank 5.0 [36], was expanded to cover not only drug binding data, numerous investigational drugs, drug-drug and drug-food interactions, and SNP-associated drug effects, but also information on the influence of hundreds of drugs on metabolite levels (pharmacometabolomics), gene expression levels (pharmacotranscriptomics), and protein expression levels (pharmacoproteomics). Enzyme inhibitors (DBCAT000003) are described as “compounds or agents that combine with an enzyme in such a manner as to prevent the normal substrate-enzyme combination and the catalytic reaction”.

Reviews on MNPs have been published on a regular basis in the scientific literature [37–39]. The renewed upsurge of interest in NPs, and MNPs in particular, over the last two decades has led to a rapid multiplication of databases in both the private sector and the public domain that compile general-purpose or thematic information on these naturally occurring compounds, often incorporating supplementary material published in scientific papers. A dedicated, searchable, and continuously updated database (MarinLit, <https://marinlit.rsc.org/>, accessed on 20 December 2022) that was established in the 1970s by Prof. John Blunt and Prof. Murray Munro (University of Canterbury, New Zealand) has been maintained by the Royal Society of Chemistry (UK) since 2014. MarinLit covers ~40,000 compounds from marine macro- and microorganisms and about

the same number of references to journal articles. Among the specialized MNP databases, the Dictionary of Marine Natural Products (DMNP) [40] appeared as the first of its kind in 2008 and encompassed a subset of data from the Dictionary of Natural Products (DNP, one of several Chapman & Hall chemical dictionaries) based on the biological source of the compounds. DMNP was marketed as a book together with a CD-ROM for a desktop version, and the searchable web-based version CHEMnetBASE (<https://dmnp.chemnetbase.com/>, accessed on 20 December 2022) is still available (v. 31.1; updated in 2022), but only to subscribing institutions.

Virtual chemical libraries of NPs can be categorized into (i) encyclopedic and general NP databases; (ii) special subsets within fully enumerated, ultra-large scale chemical libraries specifically built to facilitate VS campaigns, e.g., ZINC [41,42]; (iii) compound collections enriched with NPs used in traditional medicines; and (iv) specialized databases focused on specific habitats, geographical regions, organisms, biological activities, or even specific NP classes. Unfortunately, many NP databases belonging to the latter two categories are rather ephemeral or rapidly become either outdated or unavailable to the scientific community [43], and the same criticism applies to many bioinformatics web services related to NPs [44]. This is most likely due to (i) a lack of funds (and/or human resources) for their sustained management and continuous upgrading, and (ii) the current overwhelming “data deluge”. For these reasons, there is an urgent need for nonredundant, community-wide efforts that optimize the use of contemporary bioinformatic and chemoinformatic capabilities, as exemplified by the recently established open platform LOTUS (<https://lotus.naturalproducts.net>, accessed on 20 December 2022), a knowledgebase that is expected to have strong transformative potential for research on NPs and beyond [45]. In this praiseworthy initiative, data sharing within the Wikidata framework broadens interoperability and facilitates access to >750,000 referenced structure-organism pairs.

Another large and freely available NP database is Super Natural II ([https://bioinf-applied.charite.de/supernatural\\_new/index.php](https://bioinf-applied.charite.de/supernatural_new/index.php); last updated: October 2022, accessed on 20 December 2022), which provides two-dimensional (2D) structures and physicochemical properties for ~326,000 molecules, as well as information about the pathways associated to their synthesis, degradation, and mechanisms of action with respect to structurally similar drugs [46]. An additional recent compilation of 400,000 non-redundant NPs was made available in 2021 [47] as the open-access COLleCtion of Open NatUral producTs (COCONUT, <https://coconut.naturalproducts.net/>, accessed on 20 December 2022).

One important goal of these NP databases is to facilitate a quick assessment of novelty for any newly identified compound in a natural extract. To distinguish between known and unknown compounds, it is important to have rapid and trustworthy “dereplication” methods, which rely heavily on the interpretation of molecular mass and molecular formula, as well as UV and NMR spectral data [48]. Nevertheless, the dereplication process can be problematic sometimes because (i) the present validity and accuracy of the collected information is only as good as that of the original data source; and (ii) stereochemical information on NPs is often inaccurate or incomplete. In the field of MNPs alone, it was recently reported that more than 200 structures were misassigned in the last ten years only [49]. A comparative analysis of the original and the revised structures revealed that major pitfalls still plague the structural elucidation of small molecules and, consequently, that quite a few 3D molecular structures present in databases may be inaccurate. This finding emphasizes the roles of total synthesis, X-ray crystallography, as well as chemical and biosynthetic logic, to complement spectroscopic data. Nevertheless, it is noteworthy that a much lower incidence of “impossible” structures was found in MNPs compared to NPs of plant origin.

The utilization of computer-assisted structure elucidation (CASE) programs can minimize the risk of misassignment and help identify truly novel compounds (the “unknown unknowns”) [50] by generating all structures that are consistent with key data from 2D correlation spectroscopy (COSY), heteronuclear multiple bond correlation (HMBC), and 1,1-adequate sensitivity double-quantum spectroscopy (ADEQUATE) NMR experiments,

and by ranking the resulting structures in order of probability. The algorithms may additionally benefit from both stereospecific NMR data and use of optimized geometries and predicted chemical shifts provided by density functional theory (DFT) quantum mechanical calculations [51]. The absolute configuration of an MNP can be unequivocally confirmed by crystallographic analysis and, in the case of noncrystalline compounds containing a pseudo-meso core structure that results in a specific rotation ( $[\alpha]_D$ ) of almost zero (e.g., elatinyne), it may be necessary to absorb the compound into a porous coordination network (a “crystalline sponge”) [52].

The exploration of the identities and biological activities of metabolites present in complex mixtures has benefited enormously in recent years from scalable native and functional metabolomics approaches [53]. Novel techniques, such as affinity selection mass spectrometry (MS), complemented with pulsed ultrafiltration, size exclusion chromatography, and magnetic microbead affinity selection screening, now allow the separation of non-covalent ligand-receptor complexes from other nonbinding compounds [54].

Recognizing the need for community-wide platforms to effectively share and analyze raw, processed, or identified tandem MS (MS/MS or MS<sup>2</sup>) data of NPs, in an analogous fashion to what has been achieved in genomics and proteomics research with the GenBank<sup>®</sup> at the National Center for Biotechnology Information (NCBI) [55] and the UniProtKB [56], the open-access knowledgebase known as Global Natural Products Social Molecular Networking (GNPS, <http://gnps.ucsd.edu>, accessed on 20 December 2022) was presented in 2017 [57]. The spectral libraries enable unambiguous dereplication (by matching spectral features of the unknown compound(s) to curated spectral databases of reference compounds, i.e., identification of “known unknowns”) [50], variable dereplication (approximate matches to spectra of related molecules), and the identification of spectra in molecular networks. Importantly, GNPS allows for the community-driven, iterative re-annotation of reference MS/MS spectra in a wiki-like fashion, and therefore it will contribute to library improvements and eventual convergence of all curated MS/MS spectra. The visualization of molecular networks in GNPS represents each spectrum as a node, and spectrum-to-spectrum alignments as edges (connections) between nodes.

### 3. Linking Chemical Diversity of Secondary Metabolites to Biosynthetic Gene Clusters

Secondary metabolites can be considered genetically encoded small molecules that play a variety of roles in cell biology and therefore have the potential to become chemical probes or drug leads. Their identification and characterization can benefit from a growing number of databases and genomics-based computational tools that have been compiled and hyperlinked at the Secondary Metabolite Bioinformatics Portal (SMBP (<http://www.secondarymetabolites.org/>, accessed on 20 December 2022) website [58]. Inherent limitations related to their low production and difficult detection, and also high rediscovery rates, can be addressed, at least in part, by searching for BGCs in genomic data and unveiling their (sometimes cryptic) metabolic potential [59]. However, the highly repetitive nature of the associated genes creates major challenges for accurate sequence assembly and analysis, hence the need for new bioinformatic tools. An example is the Natural Product Domain Seeker (NaPDoS) web service (<https://npdomainseeker.sdsc.edu/napdos2/>, accessed on 20 December 2022), which provides an automated method to assess the secondary metabolite biosynthetic gene diversity and novelty of strains or environments. NaPDoS analyses are based on the phylogenetic relationships of sequence tags derived from genes encoding PKS and NRPS, respectively. The sequence tags correspond to PKS-derived KS domains and NRPS-derived condensation (C) domains and are compared to an internal database of experimentally characterized biosynthetic genes, so that genes associated with uncharacterized biochemistry can be identified [60]. The latest update (NaPDoS2) greatly expands the taxonomic and functional diversity represented in the webtool database and allows larger datasets to be analyzed. Importantly, NaPDoS2 can be used to detect genes involved in the biosynthesis of specific structural classes or new biosynthetic mechanisms, and also to predict biosynthetic potential [61].

The key role of marine microbial symbionts of invertebrates in MNP biosynthesis has been increasingly recognized [62] and “genome mining” (i.e., the exploitation of genomic information for the discovery of biosynthetic pathways) [63] provides unique opportunities for (i) the identification of yet undisclosed specialized metabolites [64] and their chemical variants [63]; (ii) the genetic engineering of BGCs to obtain novel “unnatural” NPs [65]; and (iii) the heterologous expression of secondary metabolic pathways that remain silent or are poorly expressed in the absence of a specific trigger or elicitor [66]. In fact, the results of a variety of genome sequencing projects have unveiled the metabolic diversity of microorganisms (which may be overlooked under standard fermentation and detection conditions) and their tremendous biosynthetic potential. Furthermore, studies on the evolutionary history of BGCs in relation to that of the bacteria harboring them (“comparative genomics”) beautifully illustrate the mechanisms by which chemical diversity is created in nature and how some NPs represent ecotype-defining traits while others appear selectively neutral [67].

Novel algorithms have been devised to systematically identify BGCs in microbial genomic sequences [12,63,68]. A network analysis of the predicted BGCs in Proteobacteria (aka Pseudomonadota, a major phylum of Gram-negative bacteria) has revealed large gene cluster families, and the experimental characterization of the most prominent one revealed two subfamilies consisting of hundreds of BGCs encoding the biochemical machinery for the synthesis of a series of remarkably conserved lipids with an aryl head group conjugated to a polyene tail (i.e., aryl polyenes) that are likely to play important roles in Gram-negative cell biology [17]. The systematic study of BGCs in Actinobacteria (actinomycetes mainly associated to sponges in marine habitats) is complicated by numerous repetitive motifs. By combining several metrics, a method for the global classification of these gene clusters into families (GCFs) has been developed, and the biosynthetic capacity of the resulting GCF network has been validated in hundreds of strains by correlating confident MS detection of known NPs with the presence or absence of their established BGCs [69].

The Minimum Information about a Biosynthetic Gene cluster (MIBiG, <https://mibig.secondarymetabolites.org/>, accessed on 20 December 2022) specification is a data standard that facilitates the consistent and systematic deposition and retrieval of metadata on BGCs and their molecular products [70]. MIBiG is a Genomic Standards Consortium project that builds on the Minimum Information about any Sequence (MIxS) framework to (i) identify which genes are responsible for the biosynthesis of which chemical moieties, thus systematically connecting genes and chemistry; (ii) understand the natural genetic diversity of BGCs within their environmental and ecological context; and (iii) develop an evidence-based parts registry for engineering biosynthetic pathways and gene clusters through synthetic biology. The MIBiG standard contains dedicated class-specific checklists for gene clusters encoding pathways to produce alkaloids, saccharides, terpenes, polyketides, NRPs, and RiPPs [20].

Natural antimicrobial peptides (AMPs) have been found not only in marine fish [71,72] but also in marine invertebrates [73,74] as major components of their innate host defense systems. The Antimicrobial Peptide Database (APD, <https://aps.unmc.edu/>, accessed on 20 December 2022), online since 2003 and last updated in June 2022 [75], defines four unified classes of AMPs on the basis of the polypeptide chain’s connection patterns: (I) linear polypeptide chains (e.g., cathelicidins) [76]; (II) sidechain-linked peptides, such as disulfide-containing defensins and lantibiotics (i.e., lanthionine-containing antibiotics, e.g., microbisporicin, produced by the soil actinomycete *Microbispora corallina* [77] and mathermycin from the marine actinomycete *Marinactinospora thermotolerans* [78]); (III) polypeptide chains with side chain to backbone connection (e.g., bacterial lassos and fusaricidins); and (IV) circular peptides with a seamless backbone, i.e., N- and C-termini linked by a peptide bond (e.g., plant cyclotides and animal  $\theta$ -defensins) [79]. The manually curated Database of Antimicrobial Activity and Structure of Peptides (DBAASP, <http://dbaasp.org>, accessed on 20 December 2022) provides detailed information (including chemical structure and activity against specific targets) on experimentally tested peptides (both natural and synthetic) that

have shown antimicrobial activity as monomers, multimers, or multi-peptides [80]. The Collection of Antimicrobial Peptides (CAMP), CAMPSign, and ClassAMP are open-access resources that have been developed to advance our current understanding of AMPs, from N- and C-terminal modifications and the presence of unusual amino acids to 3D structures thorough family-specific signatures that facilitate AMP identification and classification as antibacterial, antifungal, or antiviral [81,82]. Synthetic AMPs are substantially enriched in residues with physicochemical properties known to be critical for antimicrobial activity, such as high  $\alpha$ -helical propensity, positive charge, and hydrophobicity.

The Natural Products Atlas [83] was created as an open-access centralized knowledge-base encompassing ~25,000 microbially produced NPs using a combination of manual curation and automated data mining approaches, and was developed as a community-supported resource under findable, accessible, interoperable, and reusable (FAIR) [84] principles. It contains referenced data for molecular structure, source organism, isolation, total synthesis, and instances of structural reassignment for compounds of bacterial, fungal, and cyanobacterial origin. Its associated web interface (<https://www.npatlas.org>, v. 2.3.0, accessed on 20 December 2022) allows users to search by structure, substructure, and physical properties, as well as to explore the chemical space of these NPs from a variety of perspectives. The NP Atlas is integrated with other NP databases, including the MIBiG repository and the GNPS platform cited above. The NP Atlas was recently updated [19] and currently embodies (i) >32,000 compounds; (ii) a full RESTful (REST is an acronym for REpresentational State Transfer and an architectural style for distributed hypermedia systems) application programming interface (API); (iii) full taxonomic descriptions for all microbial taxa; (iv) integrated data from external resources, including CyanoMetDB (<https://www.eawag.ch/en/departement/uchem/projects/cyanometdb/>, accessed on 20 December 2022), a comprehensive public database of secondary metabolites from cyanobacteria (aka “blue-green algae”) [85]; and (v) chemical ontology terms from both ClassyFire [86] (see below) and NPClassifier (a deep-learning tool for the automated structural classification of NPs from their counted Morgan fingerprints) [87].

Finally, more than seven terabases of metagenomic data from samples collected in epipelagic and mesopelagic water locations across the globe by the *Tara* (<https://fondationtaraocean.org/en/foundation/>, accessed on 20 December 2022) Oceans project have been used to generate an ocean microbial reference gene catalog (<http://ocean-microbiome.embl.de/companion.html>, accessed on 20 December 2022) with >40 million nonredundant sequences from viruses, prokaryotes, and picoeukaryotes. Remarkably, almost three quarters of ocean microbial core functionality is shared with the human gut microbiome, and epipelagic community composition was found to be mostly driven by water temperature rather than geography or any other environmental factor [88]. A more recent analysis of 214 metagenome-assembled genomes (MAGs) recovered from the polar seawater microbiomes revealed strains that are prevalent in the polar regions while nearly undetectable in temperate seawater [89].

#### 4. Classification and Chemoinformatic Analyses of Natural Products

The long-established Gene Ontology (GO) resource [90,91] describes our knowledge of the “universe” of biology with respect to (i) molecular functions, (ii) cellular locations, and (iii) biological processes of gene products, in terms of a dynamic, controlled vocabulary that can be applied to prokaryotes and eukaryotes, as well as to single and multicellular organisms. Along the same vein, a standardized and purely structure-based chemical ontology (ChemOnt) was recently developed to automatically assign over 77 million compounds to a taxonomy consisting of >4800 different categories by means of a computer program named ClassyFire (<http://classyfire.wishartlab.com/>, accessed on 20 December 2022) that is freely accessible as a web server [86]. This new taxonomy for chemical substances consists of up to 11 different levels (kingdom, superclass, class, subclass, etc.), with each of the categories defined by unambiguous, computable structural rules.

As a follow-on, the Chemical Functional Ontology (ChemFOnT), another FAIR-compliant, web-enabled resource (<https://www.chemfont.ca>, accessed on 20 December 2022), describes the functions and actions of >341,000 biologically important chemical substances, including primary and secondary metabolites, as well as drugs and NPs. The functional hierarchy within ChemFOnT consists of four functional “aspects” (physiological effect; disposition; process; and role), which are subdivided into twelve functional categories (health effects and organoleptic effects; sources, biological locations, and routes of exposure; environmental, natural, and industrial processes; adverse biological roles, normal biological roles, environmental roles, and industrial applications) and a total of >170,000 functional terms. At the time of publishing, ChemFOnT contained almost four million protein-chemical relationships and more than ten million chemical-functional relationships that can be adopted by other databases and software tools and be of utility not only to general chemists but also to researchers involved in genomics, metagenomics, proteomics, and metabolomics [92].

NPs are the result of nature’s exploration of biologically relevant chemical space through eons of evolutionary time, hence their high diversity regarding atom connectivity and functional groups. Because they cover a broad range of sizes, 3D structures, and physicochemical properties that can be related to drug-likeness (including favorable ADME characteristics), NPs are considered not only as potential drugs, but also as an invaluable source of chemical inspiration for the development of new bioactive small molecules useful in chemical biology and medicinal chemistry research. The structural diversity of drugs was early assessed by making use of shape description methods and grouping the atoms of each drug molecule into ring, linker, framework (or scaffold) [93], and side chain [94]. A methodology that calculated the NP-likeness score—a Bayesian measure of similarity with respect to the structural space covered by NPs—proved capable of efficiently separating NPs from synthetic (i.e., man-made) molecules in a cross-validation experiment [95]. Nevertheless, rule-based procedures applied to the automated assignment of NPs to different classes, such as alkaloids, steroids, and flavonoids, have unveiled database-dependent differences in the coverage of chemical space [96]. Beyond that, several cheminformatics techniques have been used to analyze NPs and decompose them into fragments in the belief that their unique substructural features and chemical properties are likely to be optimized for protein recognition and enzyme inhibition. A recent cheminformatic analysis of the structural and physicochemical properties of NP-based drugs in comparison to top-selling brand-name synthetic drugs revealed that macrocycles occupied distinctive and relatively underpopulated regions of chemical space, while chemical probes largely overlapped with synthetic drugs [97].

Ideally, molecular diversity in drug discovery efforts should be focused on what is usually considered drug-like chemical space (aka “drug space”), which may (or may not) fully comply with Lipinski’s “rule of five” [98]. A pioneering initiative to map this space made use of 72 descriptors accounting for size, lipophilicity (calculated  $\log P_{o/w}$ ), polarizability, charge, flexibility (number of nonterminal rotatable bonds), rigidity (total number of rings and rigid bonds), and hydrogen bonding abilities for a set of ~400 compounds encompassing both representative drugs (“core structures”) and a number of “satellite molecules” intentionally placed outside of the drug space (i.e., possessing extreme values in one or several of the desired properties, while containing drug-like chemical fragments). By means of principal component analysis (PCA) and projections to latent structures (PLS) it was possible, after some iterations that involved the inclusion of additional randomly selected active molecules, to extract map coordinates in the form of *t*-score values and construct a chemical global positioning system (ChemGPS) [99]. The ChemGPS scores were found to describe well the latent structures extracted with PCA from a large set of compounds and appeared to be suitable for comparing multiple libraries and for keeping track of previously explored regions of chemical space. Later work (largely based on cyclooxygenase 1 and/or cyclooxygenase 2 (COX-1/2) inhibition) proposed an expansion of ChemGPS to better cover space for NPs, giving birth to ChemGPS-NP [100], which was further tuned

for the improved handling of the chemical diversity encountered in NP research with a view to increasing the probability of hit identification [101]. The public ChemGPS-NP Web tool (<http://chemgps.bmc.uu.se/>, accessed on 20 December 2022) was then developed to allow for the exploration of NPs by navigating in a consistent 8-dimensional global map of structural characteristics built by means of PCA [102].

Following a different philosophy to chart the known chemical space explored by nature, the structural classification of natural products (SCONP) was devised to accomplish a hierarchical grouping of the scaffolds present in ~170,000 entries from the DNP by establishing parent–child relationships between them and arranging the scaffolds in a tree-like fashion [103]. Some previous processing was necessary that included structure cleansing (i.e., separation from accompanying molecules) and deglycosylation (in the case of glycosides whose active component is the aglycon part). Unfortunately, stereochemistry could not be considered in this early cheminformatic analysis so that the different possible configurations of the NP scaffolds had to be treated as being equivalent. The conversion of the resulting NP scaffolds to SMILES (simplified molecular-input line-entry system) strings [104] allowed for the comparison with those of standard synthetic molecules represented by over 10 million drug-like commercially available samples from the ZINC database [41]. This analysis revealed interesting differences not only between natural and synthetic (i.e., man-made) molecules, but also between scaffolds originating from distinct classes of organisms, i.e., plants, bacteria, and fungi. Visual comparisons of the respective structural features were effectively displayed by plotting the scaffolds according to their frequency distributions [105]. Moreover, a flexible analytics framework named Scaffold Hunter (<https://scaffoldhunter.sourceforge.net/>, accessed on 20 December 2022) generates and enables the visualization of virtual scaffold trees in bioactive compound collections that easily allow for the identification of new starting points for the design and synthesis of biology-oriented small molecule libraries [106]. Interestingly, a recent cluster analysis of chemical fingerprints and molecular scaffolds of >55,000 compounds reportedly isolated from marine and terrestrial microorganisms showed that three quarters of the MNPs are closely related to compounds isolated from their terrestrial counterparts [107].

The cheminformatic deconstruction of hundreds of thousands of NPs has allowed for the definition of thousands of fragment groups that represent a large portion of the chemical space defined by NPs and may guide the synthesis of “non-natural” NPs or pseudo-NPs, that is, molecules made in the lab that contain at least some of the structural features present in NPs but have not yet been found in living organisms [108]. In this regard, we must bear in mind that the prototype “antimetabolite” 6-thioguanine, which was synthesized in 1955 by Nobel Prize winners Elion and Hitchings [109], was found in 2013 to be biosynthetically produced by *Erwinia amylovora*, the bacteria responsible for fire blight pathogenesis in apple and pear trees [110]. In fact, a recent cheminformatic analysis revealed that a significant portion of biologically active synthetic compounds can be regarded as pseudo-NPs and, as such, the result of human-directed “chemical evolution” of NP structure [111]. Once again, humans imitate nature by (i) performing atom/group replacement and/or decorating with novel fragments what are thought to be privileged scaffolds for bioactivity [112–114]; or (ii) combining fragment-sized NPs and/or NP fragments to provide “hybrid NPs” [115].

Historically, the total synthesis of NPs followed by derivative synthesis (“active analogue approach” or “analogue-oriented synthesis” [116,117]) and semisynthetic procedures aimed at modifying the chemical structure of complex fermentation products have enabled a deeper understanding of structure–activity relationships (SAR). In contrast, the de novo combination of NP fragments in unique arrangements, often by virtue of innovative strategies such as “diversity-oriented synthesis” [118,119], “target-oriented and diversity-oriented organic synthesis” [120], and “synthesis-informed design” [121], has been shown to generate focused NP-like libraries containing compounds endowed with bioactivities unrelated to those of the guiding NP(s) [122–124]. Examples of successful workflows of pseudo-NP design and development are “biology-oriented synthesis” [114,125] and “pharmacophore-directed retrosynthesis” [126]. In applying the latter approach, a key first

step is to elaborate a tentative pharmacophore, i.e., “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”, as defined by the International Union of Pure and Applied Chemistry (IUPAC) [127], and then devise a retrosynthetic procedure that ensures that the proposed pharmacophore is present in multiple intermediates of increasing complexity, ultimately leading to the NP. An important goal of these synthetic approaches is to find structurally simplified and optimized derivatives with lower molecular weights that can overcome commonly observed limitations, such as poor oral absorption, short half-life, and low blood–brain barrier permeability.

### 5. Linking NPs to Their Targets: Computational Methodologies for Building Global Networks

The popular term “druggable genome” [128] refers to the genes (or, more appropriately, gene products) that are known or predicted to interact with drugs, ideally resulting in a therapeutic benefit. Although drugs are intended to be selective (i.e., have high affinity for one single target), it is not uncommon for many molecules to bind to more than one protein, giving rise to polypharmacology and side effects. Due to the fact that many drug–target combinations are theoretically possible, the computational exploration of possible interactions can help identify potential targets.

Because the systematic identification of drug targets for NPs, regardless of their origin, using a battery of experimental binding or affinity assays, is both costly and time-consuming, a substantial amount of effort has gone into devising *in silico* tools that allow for the construction of global networks that connect active compounds to their cellular targets. It is expected that, by using these methods, the resulting system’s pharmacology infrastructure will help to predict new drug targets for pharmacologically uncharacterized NPs and identify secondary targets (off-targets) that can aid in the rationalization of side effects of known molecules [129]. The Drug-Gene Interaction Database (DGIdb 4.0, <https://www.dgidb.org/>, accessed on 20 December 2022) provides information on drug–gene interactions and druggable gene products collected from publications, databases, and other web sites [130]. The latest update mostly focused on (i) the integration with crowd-sourced efforts (e.g., Wikidata) to facilitate term normalization and with the open-data web platform Drug Target Commons (<https://dataverse.harvard.edu/dataverse/dtc2tdc>, accessed on 20 December 2022) [131] to enable the upload of community-contributed interaction data; and (ii) export to a Network Data Exchange (NDEx) infrastructure [132] for storing, sharing and publishing biological network knowledge. The tool named substructure–drug–target network-based inference (SDTNBI) was devised to prioritize potential targets for old drugs (“drug repositioning”), failed drugs, and new chemical entities by bridging the gap between new chemical entities and known drug–target interactions (DTIs) [133]. A later modification (wSDTNBI) [134] uses weighted DTI networks, whose edge weights are correlated with binding affinities, and network-based VS, which does not rely on the receptors’ 3D structures [135]. The publicly available SwissTargetPrediction web server (<http://www.swisstargetprediction.ch>, accessed on 20 December 2022) [136] also attempts to predict the most likely target(s) (in mice, rats, or human beings) for a SMILES-defined input molecule by using a computational method that combines different measures of similarities (both in 2D chemical structure and in 3D molecular shape) with known ligands [137]. All of these approaches, together with highly efficient receptor-based ligand docking [138], can be useful to narrow down the number of potential targets, but strict experimental confirmation and validation are needed [139,140].

The attention initially drawn [141] to certain synthetic molecules that were responsible for disproportionate percentages of hits in enzyme-based bioassays but, on closer inspection, turned out to be false actives and therefore nonprogressible hits, leading to the PAINS acronym (Pan Assay INterference compounds) [142], was later extended to NPs [143]. As a result, some NPs have been designated as “invalid metabolic panaceas” and the concept of “residual complexity” (<http://go.uic.edu/residualcomplexity>, accessed on 20

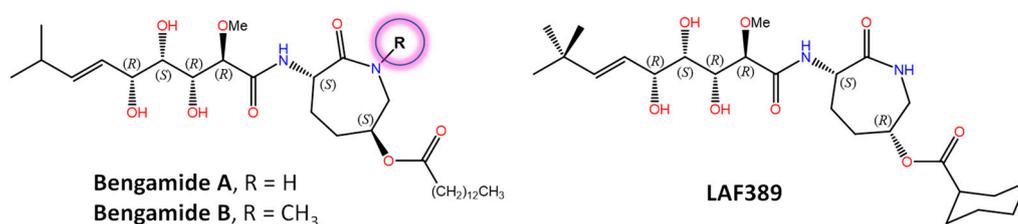
December 2022) has emerged [144]. Nowadays, compounds with a PAINS chemotype can be recognized and excluded from bioassays by the judicious use of electronic substructure filters [145] and machine learning approaches [146] (e.g., Hit Dexter, <https://nerdd.univie.ac.at/hitdexter3/>, accessed on 20 December 2022).

Because the best link connecting NPs to their targets is arguably the experimentally determined 3D structure of the respective complexes, in the following section, I will provide some examples of MNPs and synthetic analogues that were selected on the basis of chemical novelty and submicromolar inhibition data, preferably supported by structural evidence of complex formation with pharmacologically relevant enzyme targets.

## 6. Selected Examples of MNPs Acting as Enzyme Inhibitors

The road from the research laboratory to the drug pipeline is long and winding. Quite often, molecules originally assayed for one biological activity end up showing promise for another unintended indication, either fortuitously or by following one of the computational approaches outlined in the previous sections.

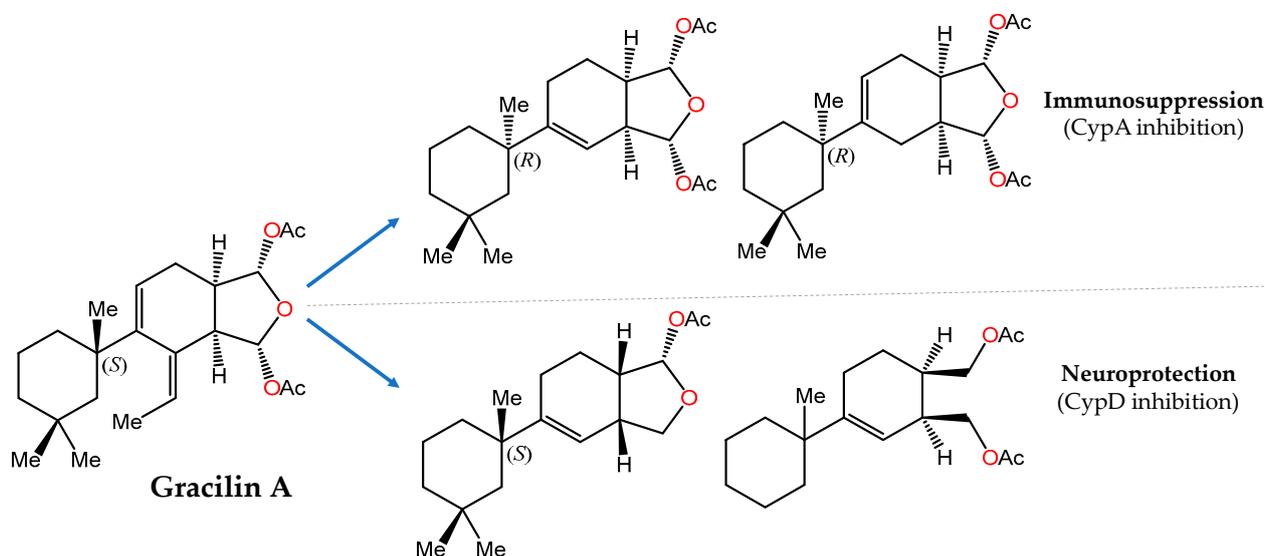
Bengamides A and B (Figure 2) were first described as heterocyclic anthelmintics naturally present in the sponge *Jaspis cf. coriacea* [147], and later on, not only in other sponges from many biogeographic sites, but also in the terrestrial Gram-negative bacterium *Mycococcus virescens*. Decades of further research have shown that methionine aminopeptidases MetAP1 and MetAP2 (essential metalloenzymes that remove the initiator amino-terminal methionine from nascent proteins) are molecular targets for bengamides, which also display notable antiproliferative and antiangiogenic properties [148,149]. In fact, a synthetic analogue of bengamide B, LAF389, was the subject of a phase I anticancer clinical trial that, unfortunately, demonstrated no objective responses and also the occurrence of unanticipated cardiovascular events. The high-resolution 3D structures of both human MetAP1 and MetAP2 enzymes in complex with bengamide derivatives, including LAF389 (PDB entry 1QZY) [150], have been solved [151] and show these compounds bound in a manner that mimics the binding of peptide substrates, with three key hydroxyl groups on the inhibitor coordinating the di-Co(II) center in the enzyme active site. Renewed interest in bengamides is currently focused on their antibacterial activities against various drug-resistant *Mycobacterium tuberculosis* [152] and *Staphylococcus aureus* strains [153]. Incidentally, the mycotoxin fumagillin, first isolated from *Aspergillus fumigatus* and originally studied also as an antiangiogenic agent and human MetAP2 inhibitor [154], has been widely used for more than 60 years in apiculture to control nosema disease in honey bees effectively [155] because the microsporidian *Nosema apis* lacks MetAP1 and targeting MetAP2 suppresses infection.



**Figure 2.** Chemical structures of naturally occurring bengamides A and B, which appear to be encoded by a mixed PKS/NRPS BGC, and the synthetic analogue LAF389.

Another showcase example is provided by gracilin A (Figure 3), a nor-diterpene metabolite originally isolated from the Mediterranean sponge *Spongionella gracilis* [156], that was initially reported as a potent phospholipase A<sub>2</sub> (PLA<sub>2</sub>) inhibitor [157] and later shown to mimic the immunosuppressive effects of cyclosporin A through interaction with cyclophilin A (CypA) [158]. In a recent pharmacophore-directed retrosynthesis application, a theoretically derived pharmacophore of gracilin A was chosen as an early synthetic target. Then, sequential increases in the complexity of this minimal structure enabled SAR

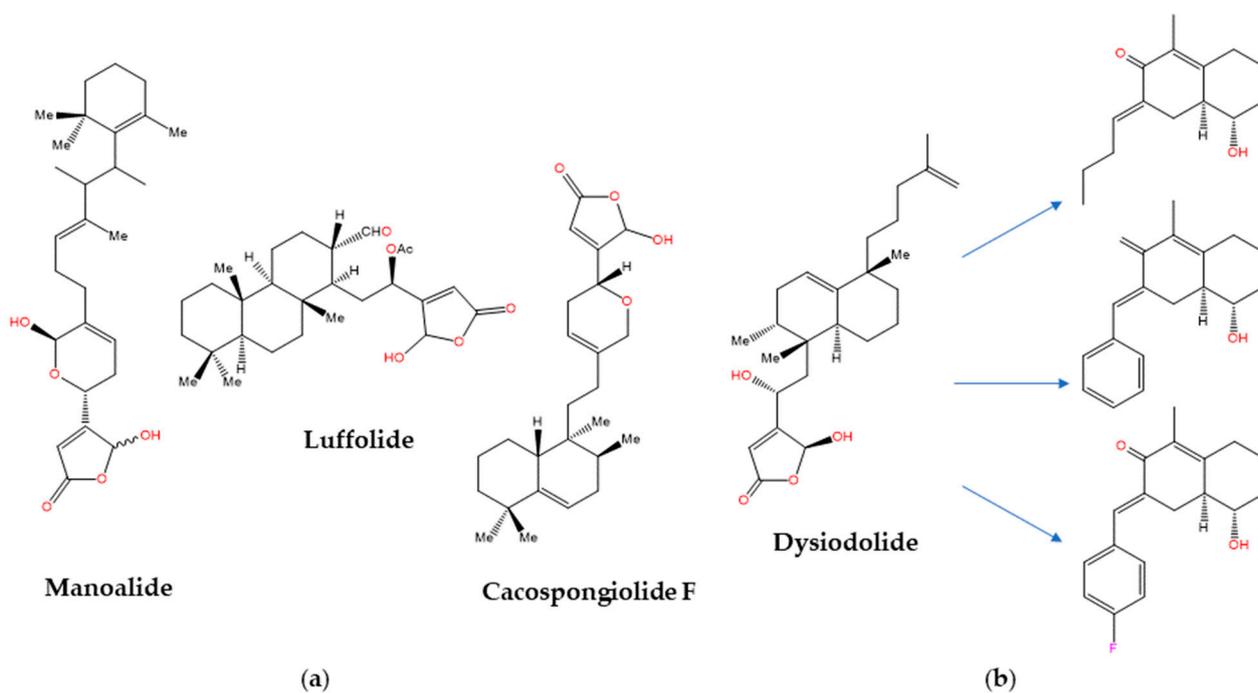
profiling and the identification of structurally less complex derivatives of gracilin A that displayed selectivity for mitochondrial CypD over CypA inhibition as well as significant neuroprotective and/or immunosuppressive activities [126].



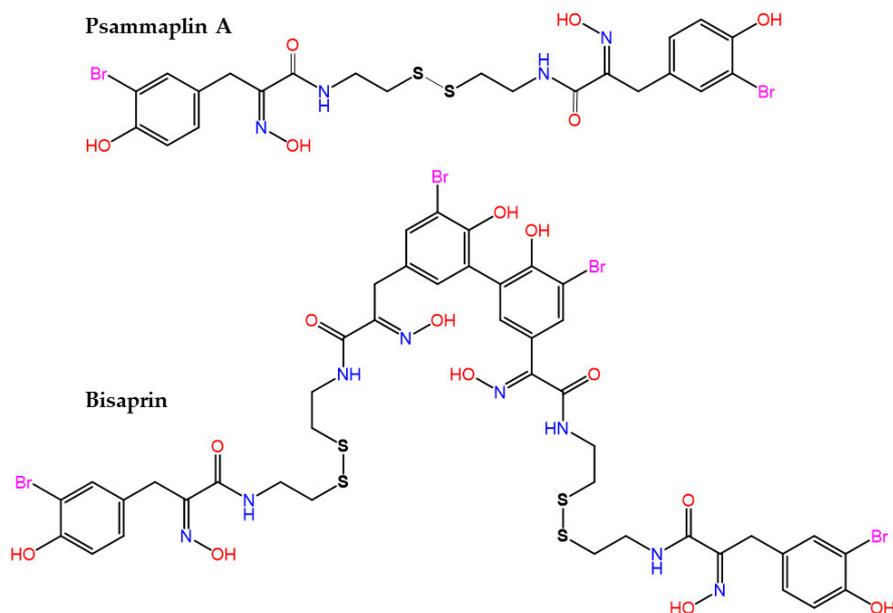
**Figure 3.** Pharmacophore-directed chemical modifications of gracilin A leading to novel compounds with distinct pharmacodynamic profiles by selective CypA vs. CypD inhibition [126].

The sesterterpenoids [159] are metabolites first isolated from marine sponges of the *Thorectidae* family, which includes the genera *Cacospongia*, *Fasciospongia*, *Luffariella*, and *Thorecta*, that often contain biologically active butenolide and hydroxybutenolide groups in their structures [160]. The anti-inflammatory activity of manoalide and luffolide (Figure 4a) was related to the inactivation of secretory PLA<sub>2</sub>, whereas for cacospongionolide F this biological effect was shown to involve the inhibition of the nuclear factor- $\kappa$ B (NF- $\kappa$ B) pathway as well [161]. In contrast, the related dysidiolide (Figure 4b) from the Caribbean sponge *Dysidea etherea* de Laubenfels (Dysideidae family) was the first known natural inhibitor of the cyclin-dependent kinase (CDK)-activating phosphatases cdc25A and cdc25B, with IC<sub>50</sub> values in the micromolar range [162]. Later on, dysidiolide and distinctly decorated analogues prepared from *ent*-halimic acid following a classical “active analogue approach” were shown to cause stage-specific arrest of proliferating cancer cells, but again at low micromolar concentrations [163]. Because cdc25A was found to be present in the same protein structure similarity cluster (PSSC) [164] as 11 $\beta$ -hydroxysteroid dehydrogenase type 1 (11 $\beta$ HSD1, an enzyme that catalyzes the conversion of cortisone to cortisol), the SCONP-guided [103] selection of the 1,2,3,4,4a,5,6,7-octahydronaphthalene scaffold present in dysidiolide led to a focused compound library (Figure 4b) that showed the submicromolar inhibition of 11 $\beta$ HSD1 and selectivity over 11 $\beta$ HSD2 [165].

The bisulfide bromotyrosine- and oxime-containing derivatives psammaplina A and bisaprasin (Figure 5) were originally characterized as nanomolar inhibitors of histone deacetylases and DNA methyltransferase enzymes [166,167], but it is known today that they are used by marine sponges, such as *Pseudoceratina purpurea* and *Aplysinella rhax*, in their chemical communication [168] and quorum sensing [169] systems to prevent biofilm formation and attenuate virulence factor expression by pathogenic microorganisms.



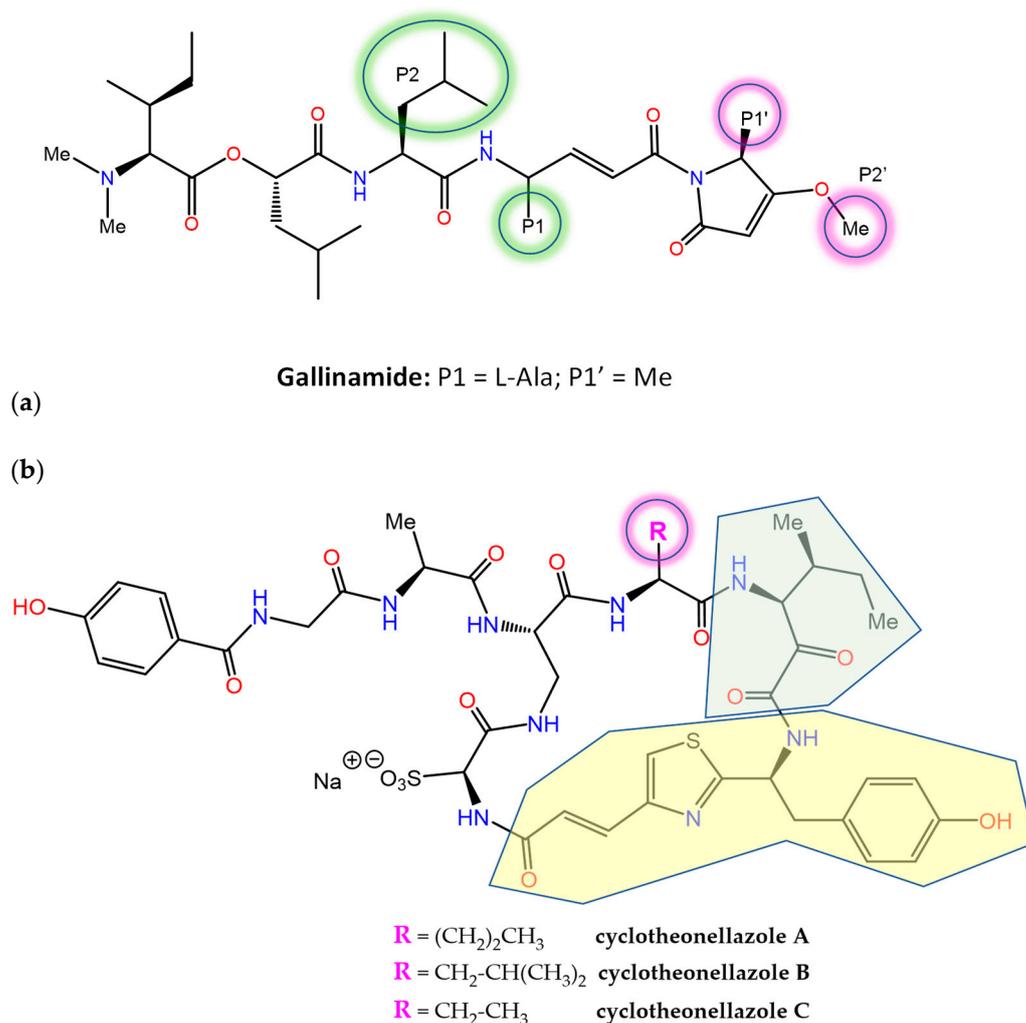
**Figure 4.** Examples of MNPs containing hydroxybutenolide moiety. (a) PLA<sub>2</sub>-inhibiting sesterterpenoids of marine origin; (b) SCONP-guided [103] evolution from dysiodolide to a focused library of submicromolar inhibitors of 11βHSD1 that showed some selectivity over 11βHSD2 [165].



**Figure 5.** Structures of psammaplin A and its biphenylic dimer bisaprasin.

A number of MNPs are potent inhibitors of proteases, an important drug target class in human diseases that is integrated in MEROPS (<http://www.ebi.ac.uk/merops/>, accessed on 20 December 2022), a database of proteolytic enzymes, their substrates and inhibitors [170]. Gallinamide A (Figure 6), a metabolite of the marine cyanobacterium *Schizothrix* sp. that originally displayed modest antimalarial activity, was subsequently reisolated and characterized as a potent and irreversible inhibitor of the human cysteine protease cathepsin L ( $k_i = 9000 \pm 260 \text{ M}^{-1} \text{ s}^{-1}$ ), with 8- to 320-fold greater selectivity over the closely related cathepsins V or B [171]. Docking-guided modifications to im-

prove the binding affinity resulted in notably enhanced potency against cathepsin L ( $K_i = 0.0937 \pm 0.01$  nM and  $k_{inact}/K_i = 8,730,000$ ). Gallinamide and its analogs also displayed the potent inhibition of the highly homologous cruzain, an essential *Trypanosoma cruzi* cysteine protease, as well as cytotoxic activity on intracellular *T. cruzi* amastigotes [172]. Importantly, the biochemical data indicated that inhibitor potency was driven by the rate of formation of the reversible enzyme:inhibitor complex, rather than by the rate of covalent modification. The resolution of the 3D co-crystal structure of the complex formed between cruzain and gallinamide A later confirmed the proposed binding pose and revealed the expected covalent bond formed between the drug's Michael acceptor enamide and the active site Cys25 thiol (PDB entry 7JUJ) [173].

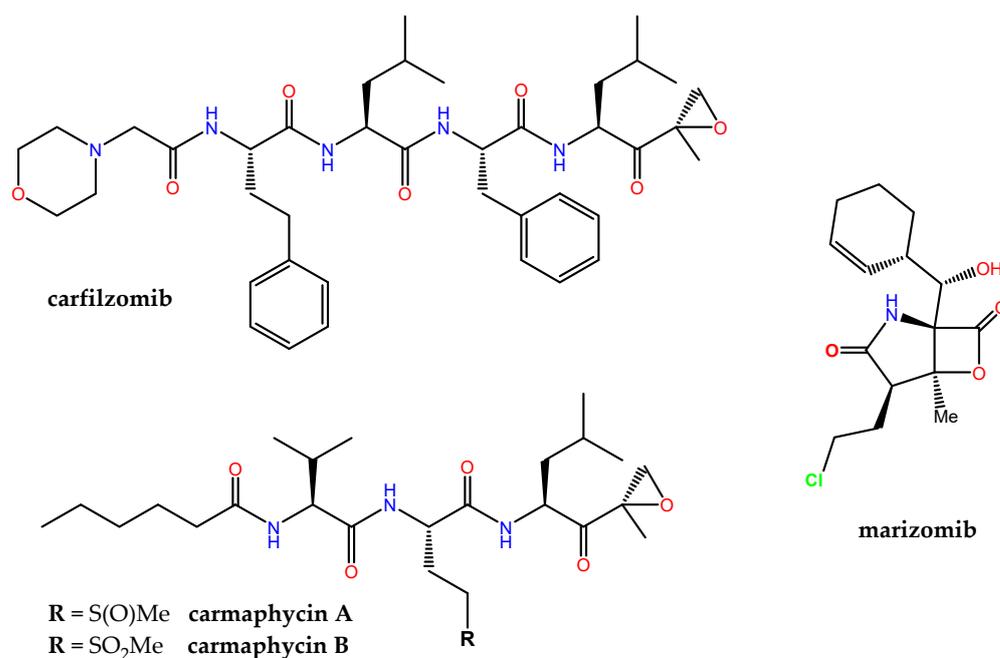


**Figure 6.** Chemical structures of protease inhibitors (a) gallinamide and (b) cyclotheonellazoles A–C. The 4-propenoyl-2-tyrosylthiazole (Ptt) and 3-amino-4-methyl-2-oxohexanoic acid (Amoha) subunits in the cyclotheonellazoles are highlighted in yellow and green, respectively.

The *in vitro* antiplasmodial activity of an extract of the sponge *Theonella aff. swinhoei* collected in Madagascar was ascribed, in part, to the previously known actin-binding metabolite swinholide A [174]. Further work disclosed the presence of three unusual cyclic peptides, cyclotheonellazoles A–C (Figure 6), containing six nonproteinogenic amino acids out of the eight composing units (of which the most novel were 4-propenoyl-2-tyrosylthiazole and 3-amino-4-methyl-2-oxohexanoic acid). These macrocyclic peptides are thought to be produced by hybrid PKS-NRPS enzymes from symbiotic bacteria and were found not to be active against *Plasmodium*, but instead displayed the nanomolar and sub-nanomolar inhibition of chymotrypsin and elastase, respectively [175]. This latter enzyme



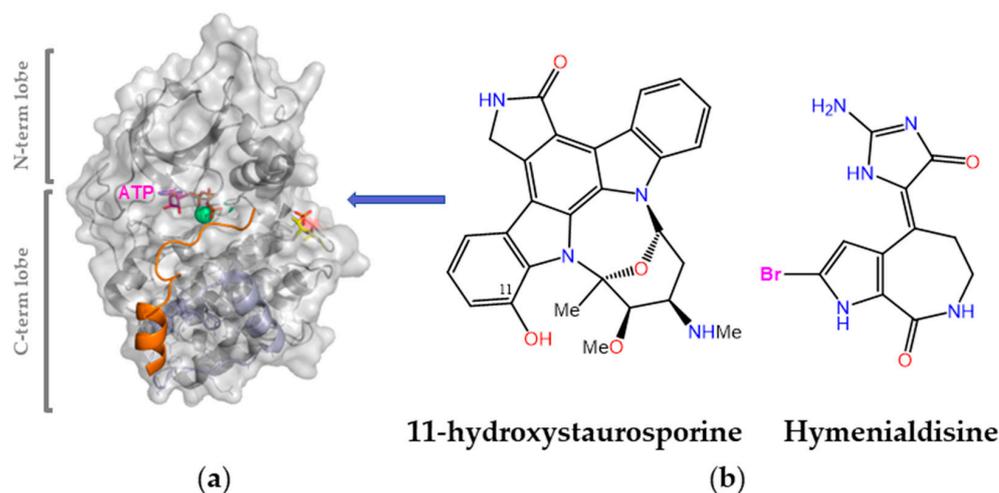
ical use in MM patients, in combination with lenalidomide and dexamethasone. Marizomib (aka salinosporamide A) is a structurally and pharmacologically unique MNP that contains a  $\beta$ -lactone- $\gamma$ -lactam (Figure 8) and is produced by the marine actinomycete *Salinispora tropica*. Marizomib not only inhibits the chymotrypsin-like activity of the proteasome (via a novel mechanism involving the acylation of the O $\gamma$  in the N-terminal catalytic Thr residue followed by the displacement of chloride) but also those of the caspase-like and trypsin-like subunits [186]. In addition to its “pan-proteasome” pharmacodynamic activity, marizomib crosses the blood-brain barrier, and for these reasons it has been extensively studied, first preclinically [187], and then in phase I–III clinical trials, both alone and in combination. Many other MNP scaffolds continue serving as inspiration for the design and synthesis of potent 20S human proteasome inhibitors, including carmaphycins A and B (Figure 8) from a marine cyanobacterium *Symploca* species and fellutamide B, originally isolated from *Penicillium fellutanium*, a fungus found in the gastrointestinal tract of the marine fish *Apogon endekataenia* [188]. In the development of potent covalent inhibitors of the proteasome, ligand docking and binding energy calculations have highlighted the importance of the optimization of the prior noncovalent binding mode, through conformational restraints, in a pose close to that found in the transition state [189].



**Figure 8.** Chemical structure of selected proteasome inhibitors: carfilzomib (clinically approved), marizomib (in clinical trials), and carmaphycins A and B (investigational).

Protein kinases are validated drug targets because (i) kinase deregulation plays an essential role in many disease states, and (ii) many inhibitors have already shown therapeutic benefit (almost one hundred are currently approved for clinical use). This bioactivity is of broad scope and has been reported for various MNPs obtained from different sources, including bacteria and cyanobacteria, fungi, algae, soft corals, sponges, and animals [6,190]. Of note, a significant number of them were originally isolated from terrestrial sources and subsequently found in marine organisms too, and vice versa. For example, the pan-kinase inhibitor staurosporine (a pentacyclic indolo(2,3-*a*)carbazole first discovered in 1977 from the bacterium *Streptomyces staurosporeus*) was one of the early tools used to probe the cellular effects of blocking the ATP-binding pocket in different protein kinases. In 2002, 11-hydroxystaurosporine (Figure 9) was reported to be present in an ascidian *Eudistoma* species collected in Micronesia and to be a more potent inhibitor of protein kinase C than staurosporine itself [191]. Another early potent pan-kinase inhibitor is (*Z*)-hymenialdisine, which owes its name to *Hymeniacidon aldis*, the sponge where it was originally found.

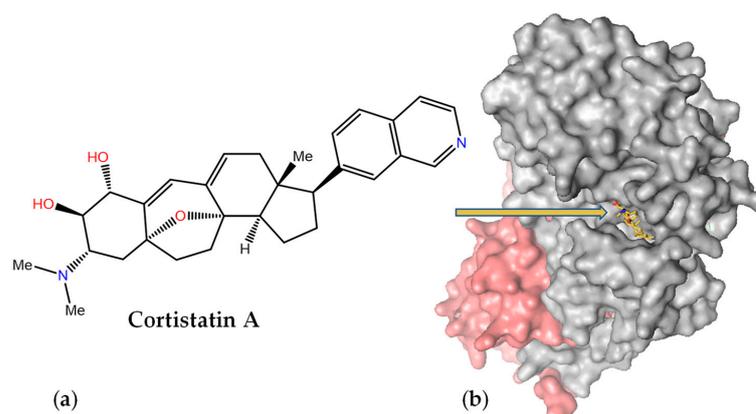
These and many other MNPs inspired synthetic work on analogues and novel scaffolds that paved the ground for the discovery of imatinib (Gleevec<sup>®</sup>), a landmark drug that has (i) significantly improved the outcomes of patients with chronic myelogenous leukemia by inhibiting the oncogenic BCR–ABL tyrosine kinase; (ii) shown remarkable clinical efficacy in the treatment of other malignancies; (iii) helped establish the concept of “targeted therapy” in the field of cancer research; (iv) fostered the concept of “precision medicine”, i.e., tailor the chemotherapeutic treatment to the unique genetic changes in an individual’s cancer cells; and (v) fuel the extremely rich and rewarding research on protein kinase inhibitors [192].



**Figure 9.** Early small-molecule protein kinase inhibitors of marine origin. (a) A representative protein kinase structure (PDB code 3FJQ) showing a bound peptide inhibitor (orange) and an ATP molecule (purple) plus two  $Mn^{2+}$  ions (green spheres) in the crevice that separates the N- and C-terminal lobes. The C-terminal lobe (bottom) contains the activation sites (T196 and phosphorylated T198 in yellow) and substrate-binding site (residues 230–260 in light blue). (b) Chemical structures of two early pan-kinase inhibitors.

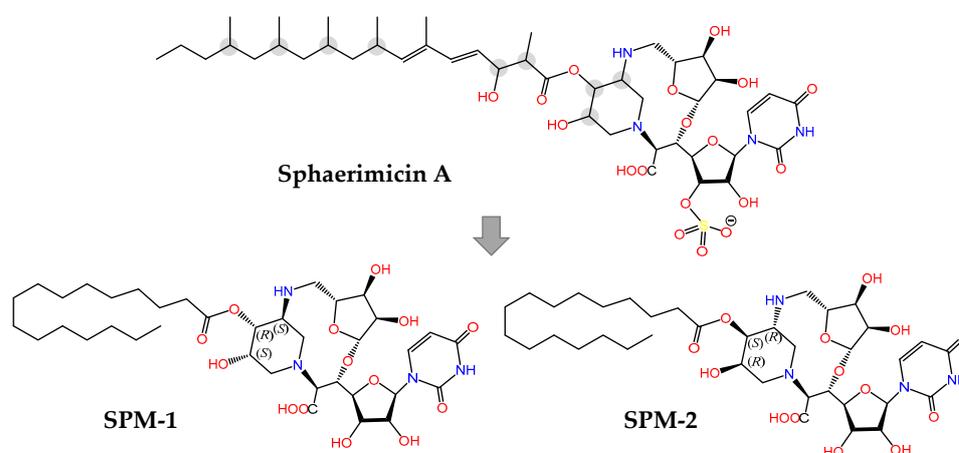
A top priority in the development of novel protein kinase inhibitors is to understand selectivity so that the tendency of one given drug to bind to other unintended kinases (off-targets) can be suppressed or attenuated. To this end, kinome-wide inhibitory selectivity profiling is necessary because small assay panels cannot provide a robust measure of selectivity [193]. Binding site similarity searches, as performed in the KinomeFEATURE (<https://simtk.org/projects/kdb>, accessed on 20 December 2022) [194] and KID [195] databases, along with machine learning models that map the activity profile of inhibitors across the entire human kinome [196], as exemplified by Drug Discovery Maps [197], can be of help not only to gain insight into the structural basis of kinase cross-inhibition, but also to predict the binding affinities of novel kinase inhibitors.

Cortistatin A (Figure 10a) was isolated as an antiangiogenic steroidal alkaloid from the marine sponge *Corticium simplex* and should not be confused with the somatostatin-like cortistatin neuropeptides. It consists of a 9(10→19)-abeo-androstane and isoquinoline skeleton and was originally shown to inhibit the proliferation of human umbilical vein endothelial cells at nanomolar concentrations [198]. It was later found that this MNP selectively inhibits the mediator-associated cyclin-dependent kinase CDK8 and disproportionately induces the upregulation of superenhancer-associated genes in acute myeloid leukemia cell lines [199]. The crystal structure of the ternary complex of CDK8 bound to cyclin C and cortistatin A (PDB code 4CRL) revealed exquisite shape complementarity between this alkaloid and the ATP-binding pocket of CDK8 (Figure 10b), with the crucial isoquinoline [200] making essential hydrogen bonding interactions with the peptide backbone.



**Figure 10.** Inhibition of mediator-associated cyclin-dependent kinase CDK8 by cortistatin A. (a) Chemical structure of cortistatin A. (b) Schematic view of the complex between cyclin C (pink) and CDK8 (grey) showing cortistatin A (yellow sticks) bound inside the ATP-binding pocket of CDK8 (PDB code 4CRL).

The last example in this review is provided by sphaerimicin A (Figure 11), a complex macrocyclic uridine nucleoside derivative isolated from *Sphaerisporangium* sp. SANK60911 using a genome mining approach focused on the enzyme uridine-5'-aldehyde transaldolase [201]. Even though this is a terrestrial actinomycete, it is closely related to other marine species [202] that contain similar BGCs, hence its inclusion in this section. Sphaerimicin A exhibits nanomolar inhibitory activity on bacterial *MraY*, an integral membrane enzyme that catalyzes the transfer of phospho-*N*-acetylmuramyl pentapeptide from UDP-*N*-acetylmuramyl pentapeptide (Park's nucleotide) to the phospholipid undecaprenyl phosphate during the lipid cycle of peptidoglycan biosynthesis. In an elegant example of molecular design assisted by theoretical conformational analysis and NMR data, the simplified analogues with defined stereochemistry SPM-1 and SPM-2 were synthesized [203]. The fact that SPM-1 turned out to be 54-fold more potent than SPM-2 against *MraY* from *Aquifex aeolicus* (*MraY*<sub>AA</sub>) revealed the importance of the conformationally restrained macrocycle for target binding, an aspect that was clarified even further when the 3D structure of the *MraY*<sub>AA</sub>:SPM-1 complex was solved by X-ray crystallography. Therefore structure-based optimization is now feasible in order to develop *MraY* inhibitors with the potential of becoming novel antibiotics against drug-resistant bacteria.



**Figure 11.** Chemical structures of the naturally occurring sphaerimicin A (containing the undetermined stereogenic centers circled in grey) and synthetic simplified analogues SPM-1 and SPM-2 with defined stereochemistry, which is critical for inhibitory potency. The X-ray crystal structure of the *MraY*:SPM-1 complex has recently been determined (PDB code 8CXR) [203].

## 7. Conclusions and Outlook

Computational methodologies play indispensable roles in the exploration of the vast chemical space covered by MNPs by helping, among many other tasks, to (i) elucidate their chemical composition and 3D structure; (ii) store, process, curate, and organize huge amounts of information related to source organisms, biosynthesis, and bioactivity; and (iii) connect biological activities with both molecular scaffolds and target binding sites [204]. The limits of biologically relevant chemical space for enzyme inhibitors are defined by the specific binding interactions taking place between small- and medium-sized molecules (e.g., terpenoids, alkaloids, polyketides, non-ribosomal peptides, and RiPPs) [20] and a number of selected orthosteric and allosteric pockets in macromolecular catalysts that have evolved over billions of years [205].

Some of the molecular entities recently found in marine microbiota can easily defy and outperform a chemist's imagination and ingenuity, and also be endowed with unexpected, and even unprecedented, bioactivities that may inspire more synthetic creativity. A historical example is the clinically used cytarabine (aka cytosine arabinoside or arabinosyl cytosine, Ara-C), a synthetic pyrimidine nucleoside that was developed in the imitation of spongothymidine, a nucleoside originally isolated from the Caribbean sponge *Tethya crypta*. Many other analogues, however, did not follow the same fate and, in fact, the potential of MNPs as therapeutic agents for human diseases has been realized only in a few cases, which attests to the enormous difficulties of progressing many of these compounds through the drug pipeline with the final goal of demonstrating an acceptable benefit-risk balance in clinical trials and thereafter. We must be confident that the new generations of cross-disciplinary trained scientists working in community-wide networks (e.g., Ocean Medicines, <https://cordis.europa.eu/project/id/690944>, accessed on 20 December 2022) will overcome existing hurdles to find valuable new medicines inspired by, or based on, MNPs.

It seems clear that the integration of information from various sources, including high-throughput phenotypic screening and BGC engineering, using computational methods has revolutionized NP research and can speed up the process of discovering new biologically active molecules from marine and terrestrial sources. A major bottleneck in these efforts is to identify the macromolecular target that is responsible for the observed (or assigned) mechanism of action, a problem that is usually aggravated when dealing with complex mixtures of MNPs. The recent success in the functional characterization of several NPs and the identification of bioactive metabolites upon integrating results from untargeted metabolomics, high-content image analysis of perturbation-treated cells, and gene expression signatures [206] on a data-driven multi-platform raises the hope for the accelerated discovery of novel pharmacologically active MNPs in the near future.

**Funding:** This research was funded by the Spanish MICINN (Project PID2019-104070RB-C22).

**Data Availability Statement:** Data sharing is not applicable.

**Acknowledgments:** I am grateful to Xavier Avilés (Universidad de Barcelona, Spain) for his kind invitation to contribute to this special issue, and to Schrodinger LLC for the provision of the educational version of PyMOL v. 1.7.4.5 (2022) that was used for 3D figure composition.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Funayama, S.; Cordell, G.A. *Alkaloids: A Treasury of Poisons and Medicines*; Elsevier: Amsterdam, The Netherlands, 2014.
2. Mora, C.; Tittensor, D.P.; Adl, S.; Simpson, A.G.; Worm, B. How many species are there on Earth and in the ocean? *PLoS Biol.* **2011**, *9*, e1001127. [CrossRef] [PubMed]
3. Mayer, A.M.; Glaser, K.B.; Cuevas, C.; Jacobs, R.S.; Kem, W.; Little, R.D.; McIntosh, J.M.; Newman, D.J.; Potts, B.C.; Shuster, D.E. The odyssey of marine pharmaceuticals: A current pipeline perspective. *Trends Pharmacol. Sci.* **2010**, *31*, 255–265. [CrossRef] [PubMed]
4. Glaser, K.B.; Mayer, A.M. A renaissance in marine pharmacology: From preclinical curiosity to clinical reality. *Biochem. Pharmacol.* **2009**, *78*, 440–448. [CrossRef] [PubMed]

5. Reymond, J.L. The chemical space project. *Acc. Chem. Res.* **2015**, *48*, 722–730. [[CrossRef](#)] [[PubMed](#)]
6. Bharate, S.B.; Sawant, S.D.; Singh, P.P.; Vishwakarma, R.A. Kinase inhibitors of marine origin. *Chem. Rev.* **2013**, *113*, 6761–6815. [[CrossRef](#)] [[PubMed](#)]
7. Newman, D.J.; Cragg, G.M.; Battershill, C.N. Therapeutic agents from the sea: Biodiversity, chemo-evolutionary insight and advances to the end of Darwin's 200th year. *Diving Hyperb. Med.* **2009**, *39*, 216–225.
8. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803. [[CrossRef](#)]
9. Nakao, Y.; Fusetani, N. Enzyme Inhibitors from Marine Invertebrates. In *Handbook of Marine Natural Products*; Fattorusso, E., Gerwick, W.H., Tagliatalata-Scafati, O., Eds.; Springer Science+Business Media B.V.: Dordrecht, The Netherlands, 2012; pp. 1145–1229.
10. Duan, D.; Doak, A.K.; Nedyalkova, L.; Shoichet, B.K. Colloidal aggregation and the in vitro activity of traditional Chinese medicines. *ACS Chem. Biol.* **2015**, *10*, 978–988. [[CrossRef](#)]
11. Seidler, J.; McGovern, S.L.; Doman, T.N.; Shoichet, B.K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **2003**, *46*, 4477–4486. [[CrossRef](#)]
12. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [[CrossRef](#)]
13. Weissman, K.J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **2015**, *11*, 660–670. [[CrossRef](#)]
14. Fischbach, M.A.; Walsh, C.T. Biochemistry. Directing biosynthesis. *Science* **2006**, *314*, 603–605. [[CrossRef](#)] [[PubMed](#)]
15. Jurjens, G.; Kirschning, A.; Candito, D.A. Lessons from the synthetic chemist nature. *Nat. Prod. Rep.* **2015**, *32*, 723–737. [[CrossRef](#)] [[PubMed](#)]
16. Sherman, D.H.; Rath, C.M.; Mortison, J.; Scaglione, J.B.; Kittendorf, J.D. Biosynthetic Principles in Marine Natural Product Systems. In *Handbook of Marine Natural Products*; Fattorusso, E., Gerwick, W.H., Tagliatalata-Scafati, O., Eds.; Springer Nature Switzerland AG: Dordrecht, The Netherlands, 2012; pp. 947–976.
17. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Wieland Brown, L.C.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421. [[CrossRef](#)]
18. Harizani, M.; Ioannou, E.; Roussis, V. The *Laurencia* paradox: An endless source of chemodiversity. *Prog. Chem. Org. Nat. Prod.* **2016**, *102*, 91–252. [[CrossRef](#)]
19. van Santen, J.A.; Poynton, E.F.; Iskakova, D.; McMann, E.; Alsup, T.A.; Clark, T.N.; Fergusson, C.H.; Fewer, D.P.; Hughes, A.H.; McCadden, C.A.; et al. The Natural Products Atlas 2.0: A database of microbially-derived natural products. *Nucleic Acids Res.* **2022**, *50*, D1317–D1323. [[CrossRef](#)] [[PubMed](#)]
20. Arnison, P.G.; Bibb, M.J.; Bierbaum, G.; Bowers, A.A.; Bugni, T.S.; Bulaj, G.; Camarero, J.A.; Campopiano, D.J.; Challis, G.L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108–160. [[CrossRef](#)] [[PubMed](#)]
21. Dang, T.; Sussmuth, R.D. Bioactive peptide natural products as lead structures for medicinal use. *Acc. Chem. Res.* **2017**, *50*, 1566–1576. [[CrossRef](#)]
22. Korteling, J.E.; van de Boer-Visschedijk, G.C.; Blankendaal, R.A.M.; Boonekamp, R.C.; Eikelboom, A.R. Human-versus Artificial Intelligence. *Front. Artif. Intell.* **2021**, *4*, 622364. [[CrossRef](#)]
23. Alexander, S.P.H.; Fabbro, D.; Kelly, E.; Mathie, A.; Peters, J.A.; Veale, E.L.; Armstrong, J.F.; Faccenda, E.; Harding, S.D.; Pawson, A.J.; et al. The concise guide to pharmacology 2019/20: Enzymes. *Br. J. Pharmacol.* **2019**, *176* (Suppl. S1), S297–S396. [[CrossRef](#)]
24. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [[CrossRef](#)]
25. Burley, S.K.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods Mol. Biol.* **2017**, *1607*, 627–641. [[CrossRef](#)] [[PubMed](#)]
26. Hameduh, T.; Haddad, Y.; Adam, V.; Heger, Z. Homology modeling in the time of collective and artificial intelligence. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3494–3506. [[CrossRef](#)] [[PubMed](#)]
27. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
28. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic level protein structure with a language model. *bioRxiv* **2022**. [[CrossRef](#)]
29. Drysdale, R.; Cook, C.E.; Petryszak, R.; Baillie-Gerritsen, V.; Barlow, M.; Gasteiger, E.; Gruhl, F.; Haas, J.; Lanfear, J.; Lopez, R.; et al. The ELIXIR Core Data Resources: Fundamental infrastructure for the life sciences. *Bioinformatics* **2020**, *36*, 2636–2642. [[CrossRef](#)]
30. Chang, A.; Jeske, L.; Ulbrich, S.; Hofmann, J.; Koblit, J.; Schomburg, I.; Neumann-Schaal, M.; Jahn, D.; Schomburg, D. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **2021**, *49*, D498–D508. [[CrossRef](#)]
31. Ribeiro, A.J.M.; Holliday, G.L.; Furnham, N.; Tyzack, J.D.; Ferris, K.; Thornton, J.M. Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **2018**, *46*, D618–D623. [[CrossRef](#)]
32. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [[CrossRef](#)]
33. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magarinos, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]

34. Pence, H.E.; Williams, A. Chemspider: An online chemical information resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124. [[CrossRef](#)]
35. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [[CrossRef](#)] [[PubMed](#)]
36. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)] [[PubMed](#)]
37. Faulkner, D.J. Marine natural products. *Nat. Prod. Rep.* **2002**, *19*, 1R–49R. [[CrossRef](#)]
38. Blunt, J.W.; Carroll, A.R.; Copp, B.R.; Davis, R.A.; Keyzers, R.A.; Prinsep, M.R. Marine natural products. *Nat. Prod. Rep.* **2018**, *35*, 8–53. [[CrossRef](#)] [[PubMed](#)]
39. Carroll, A.R.; Copp, B.R.; Davis, R.A.; Keyzers, R.A.; Prinsep, M.R. Marine natural products. *Nat. Prod. Rep.* **2022**, *39*, 1122–1171. [[CrossRef](#)]
40. Blunt, J.; Munro, M.H.G. *Dictionary of Marine Natural Products, with CD-ROM*; Chapman & Hall/CRC: Boca Raton, FA, USA, 2008.
41. Sterling, T.; Irwin, J.J. ZINC 15—Ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [[CrossRef](#)]
42. Irwin, J.J.; Tang, K.G.; Young, J.; Dandarchuluun, C.; Wong, B.R.; Khurelbaatar, M.; Moroz, Y.S.; Mayfield, J.; Sayle, R.A. ZINC20-A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073. [[CrossRef](#)]
43. Sorokina, M.; Steinbeck, C. Review on natural products databases: Where to find data in 2020. *J. Cheminform.* **2020**, *12*, 20. [[CrossRef](#)]
44. Kern, F.; Fehlmann, T.; Keller, A. On the lifetime of bioinformatics web services. *Nucleic Acids Res.* **2020**, *48*, 12523–12533. [[CrossRef](#)]
45. Rutz, A.; Sorokina, M.; Galgonek, J.; Mietchen, D.; Willighagen, E.; Gaudry, A.; Graham, J.G.; Stephan, R.; Page, R.; Vondrasek, J.; et al. The LOTUS initiative for open knowledge management in natural products research. *Elife* **2022**, *11*, e70780. [[CrossRef](#)] [[PubMed](#)]
46. Banerjee, P.; Ereman, J.; Gohlke, B.O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II—a database of natural products. *Nucleic Acids Res.* **2015**, *43*, D935–D939. [[CrossRef](#)] [[PubMed](#)]
47. Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M.A.; Steinbeck, C. COCONUT online: Collection of Open Natural Products database. *J. Cheminform.* **2021**, *13*, 2. [[CrossRef](#)] [[PubMed](#)]
48. Blunt, J.; Munro, M.; Upjohn, M. The Role of Databases in Marine Natural Products Research. In *Handbook of Marine Natural Products*; Fattorusso, E., Gerwick, W.H., Tagliatela-Scafati, O., Eds.; Springer Science+Business Media B.V.: Dordrecht, The Netherlands, 2012; pp. 389–421.
49. Shen, S.M.; Appendino, G.; Guo, Y.W. Pitfalls in the structural elucidation of small molecules. A critical analysis of a decade of structural misassignments of marine natural products. *Nat. Prod. Rep.* **2022**, *39*, 1803–1832. [[CrossRef](#)]
50. Wishart, D.S. Computational strategies for metabolite identification in metabolomics. *Bioanalysis* **2009**, *1*, 1579–1596. [[CrossRef](#)]
51. Burns, D.C.; Mazzola, E.P.; Reynolds, W.F. The role of computer-assisted structure elucidation (CASE) programs in the structure elucidation of complex natural products. *Nat. Prod. Rep.* **2019**, *36*, 919–933. [[CrossRef](#)]
52. Urban, S.; Brkljaca, R.; Hoshino, M.; Lee, S.; Fujita, M. Determination of the absolute configuration of the pseudo-symmetric natural product elatenyne by the crystalline sponge method. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 2678–2682. [[CrossRef](#)]
53. Rinschen, M.M.; Ivanisevic, J.; Giera, M.; Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 353–367. [[CrossRef](#)]
54. Muchiri, R.N.; van Breemen, R.B. Affinity selection-mass spectrometry for the discovery of pharmacologically active compounds from combinatorial libraries and natural products. *J. Mass Spectrom.* **2021**, *56*, e4647. [[CrossRef](#)]
55. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *41*, D36–D42. [[CrossRef](#)]
56. Magrane, M.; UniProt, C. UniProt Knowledgebase: A hub of integrated protein data. *Database* **2011**, *2011*, bar009. [[CrossRef](#)] [[PubMed](#)]
57. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)] [[PubMed](#)]
58. Weber, T.; Kim, H.U. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.* **2016**, *1*, 69–79. [[CrossRef](#)]
59. Scherlach, K.; Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* **2021**, *12*, 3864. [[CrossRef](#)] [[PubMed](#)]
60. Ziemert, N.; Podell, S.; Penn, K.; Badger, J.H.; Allen, E.; Jensen, P.R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **2012**, *7*, e34064. [[CrossRef](#)]
61. Klau, L.J.; Podell, S.; Creamer, K.E.; Demko, A.M.; Singh, H.W.; Allen, E.E.; Moore, B.S.; Ziemert, N.; Letzel, A.C.; Jensen, P.R. The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *J. Biol. Chem.* **2022**, *298*, 102480. [[CrossRef](#)]
62. Albarano, L.; Esposito, R.; Ruocco, N.; Costantini, M. Genome mining as new challenge in natural products discovery. *Mar. Drugs* **2020**, *18*, 199. [[CrossRef](#)]
63. Medema, M.H.; Fischbach, M.A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639–648. [[CrossRef](#)]

64. Winter, J.M.; Behnken, S.; Hertweck, C. Genomics-inspired discovery of natural products. *Curr. Opin. Chem. Biol.* **2011**, *15*, 22–31. [[CrossRef](#)]
65. Lane, A.L.; Moore, B.S. A sea of biosynthesis: Marine natural products meet the molecular age. *Nat. Prod. Rep.* **2011**, *28*, 411–428. [[CrossRef](#)]
66. Bonet, B.; Teufel, R.; Crusemann, M.; Ziemert, N.; Moore, B.S. Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J. Nat. Prod.* **2015**, *78*, 539–542. [[CrossRef](#)] [[PubMed](#)]
67. Jensen, P.R. Natural products and the gene cluster revolution. *Trends Microbiol.* **2016**, *24*, 968–977. [[CrossRef](#)] [[PubMed](#)]
68. Medema, M.H. The year 2020 in natural product bioinformatics: An overview of the latest tools and databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306. [[CrossRef](#)] [[PubMed](#)]
69. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968. [[CrossRef](#)] [[PubMed](#)]
70. Medema, M.H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J.B.; Blin, K.; de Bruijn, I.; Chooi, Y.H.; Claesen, J.; Coates, R.C.; et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625–631. [[CrossRef](#)] [[PubMed](#)]
71. Masso-Silva, J.A.; Diamond, G. Antimicrobial peptides from fish. *Pharmaceuticals* **2014**, *7*, 265–310. [[CrossRef](#)] [[PubMed](#)]
72. Barroso, C.; Carvalho, P.; Goncalves, J.F.M.; Rodrigues, P.N.S.; Neves, J.V. Antimicrobial peptides: Identification of two b-defensins in a teleost fish, the european sea bass (*Dicentrarchus labrax*). *Pharmaceuticals* **2021**, *14*, 566. [[CrossRef](#)]
73. Tincu, J.A.; Taylor, S.W. Antimicrobial peptides from marine invertebrates. *Antimicrob. Agents Chemother.* **2004**, *48*, 3645–3654. [[CrossRef](#)]
74. Sychev, S.V.; Sukhanov, S.V.; Pantelev, P.V.; Shenkarev, Z.O.; Ovchinnikova, T.V. Marine antimicrobial peptide arenicin adopts a monomeric twisted beta-hairpin structure and forms low conductivity pores in zwitterionic lipid bilayers. *Pept. Sci.* **2017**, *110*, e23093. [[CrossRef](#)]
75. Wang, G.; Li, X.; Wang, Z. APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **2016**, *44*, D1087–D1093. [[CrossRef](#)] [[PubMed](#)]
76. Broekman, D.C.; Zenz, A.; Gudmundsdottir, B.K.; Lohner, K.; Maier, V.H.; Gudmundsson, G.H. Functional characterization of codCath, the mature cathelicidin antimicrobial peptide from Atlantic cod (*Gadus morhua*). *Peptides* **2011**, *32*, 2044–2051. [[CrossRef](#)] [[PubMed](#)]
77. Castiglione, F.; Lazzarini, A.; Carrano, L.; Corti, E.; Ciciliato, I.; Gastaldo, L.; Candiani, P.; Losi, D.; Marinelli, F.; Selva, E.; et al. Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. *Chem. Biol.* **2008**, *15*, 22–31. [[CrossRef](#)] [[PubMed](#)]
78. Chen, E.; Chen, Q.; Chen, S.; Xu, B.; Ju, J.; Wang, H. Mathermycin, a lantibiotic from the marine actinomycete *Marinactinospora thermotolerans* SCSIO 00652. *Appl. Environ. Microbiol.* **2017**, *83*, e00926–17. [[CrossRef](#)] [[PubMed](#)]
79. Wang, G. The antimicrobial peptide database provides a platform for decoding the design principles of naturally occurring antimicrobial peptides. *Protein Sci.* **2020**, *29*, 8–18. [[CrossRef](#)]
80. Pirtskhalava, M.; Amstrong, A.A.; Grigolava, M.; Chubinidze, M.; Alimbarashvili, E.; Vishnepolsky, B.; Gabrielian, A.; Rosenthal, A.; Hurt, D.E.; Tartakovsky, M. DBAASP v3: Database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res.* **2021**, *49*, D288–D297. [[CrossRef](#)]
81. Waghu, F.H.; Idicula-Thomas, S. Collection of antimicrobial peptides database and its derivatives: Applications and beyond. *Protein Sci.* **2020**, *29*, 36–42. [[CrossRef](#)]
82. Gawde, U.; Chakraborty, S.; Waghu, F.H.; Barai, R.S.; Khanderkar, A.; Indraguru, R.; Shirsat, T.; Idicula-Thomas, S. CAMPR4: A database of natural and synthetic antimicrobial peptides. *Nucleic Acids Res.* **2022**, *51*, D377–D383. [[CrossRef](#)]
83. van Santen, J.A.; Jacob, G.; Singh, A.L.; Aniebok, V.; Balunas, M.J.; Bunsko, D.; Neto, F.C.; Castano-Espriu, L.; Chang, C.; Clark, T.N.; et al. The Natural Products Atlas: An open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.* **2019**, *5*, 1824–1833. [[CrossRef](#)]
84. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
85. Jones, M.R.; Pinto, E.; Torres, M.A.; Dorr, F.; Mazur-Marzec, H.; Szubert, K.; Tartaglione, L.; Dell’Aversano, C.; Miles, C.O.; Beach, D.G.; et al. CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. *Water Res.* **2021**, *196*, 117017. [[CrossRef](#)]
86. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **2016**, *8*, 61. [[CrossRef](#)]
87. Kim, H.W.; Wang, M.; Leber, C.A.; Nothias, L.F.; Reher, R.; Kang, K.B.; van der Hooft, J.J.J.; Dorrestein, P.C.; Gerwick, W.H.; Cottrell, G.W. NPClassifier: A deep neural network-based structural classification tool for natural products. *J. Nat. Prod.* **2021**, *84*, 2795–2807. [[CrossRef](#)]
88. Sunagawa, S.; Coelho, L.P.; Chaffron, S.; Kultima, J.R.; Labadie, K.; Salazar, G.; Djahanschiri, B.; Zeller, G.; Mende, D.R.; Alberti, A.; et al. Ocean plankton. Structure and function of the global ocean microbiome. *Science* **2015**, *348*, 1261359. [[CrossRef](#)]

89. Cao, S.; Zhang, W.; Ding, W.; Wang, M.; Fan, S.; Yang, B.; McMinn, A.; Wang, M.; Xie, B.B.; Qin, Q.L.; et al. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome* **2020**, *8*, 47. [[CrossRef](#)]
90. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
91. Gene Ontology, C. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334. [[CrossRef](#)]
92. Wishart, D.S.; Girod, S.; Peters, H.; Oler, E.; Jovel, J.; Budinski, Z.; Milford, R.; Lui, V.W.; Sayeeda, Z.; Mah, R.; et al. ChemFOnT: The chemical functional ontology resource. *Nucleic Acids Res.* **2022**, *51*, D1220–D1229. [[CrossRef](#)]
93. Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893. [[CrossRef](#)]
94. Bemis, G.W.; Murcko, M.A. Properties of known drugs. 2. Side chains. *J. Med. Chem.* **1999**, *42*, 5095–5099. [[CrossRef](#)]
95. Ertl, P.; Roggo, S.; Schuffenhauer, A. Natural product-likeness score and its application for prioritization of compound libraries. *J. Chem. Inf. Model.* **2008**, *48*, 68–74. [[CrossRef](#)]
96. Chen, Y.; García de Lomana, M.; Friedrich, N.O.; Kirchmair, J. Characterization of the chemical space of known and readily obtainable natural products. *J. Chem. Inf. Model.* **2018**, *58*, 1518–1532. [[CrossRef](#)]
97. Stone, S.; Newman, D.J.; Colletti, S.L.; Tan, D.S. Cheminformatic analysis of natural product-based drugs and chemical probes. *Nat. Prod. Rep.* **2022**, *39*, 20–32. [[CrossRef](#)]
98. Zhang, M.Q.; Wilkinson, B. Drug discovery beyond the ‘rule-of-five’. *Curr. Opin. Biotechnol.* **2007**, *18*, 478–488. [[CrossRef](#)]
99. Oprea, T.I.; Gottfries, J. Chemography: The art of navigating in chemical space. *J. Comb. Chem.* **2001**, *3*, 157–166. [[CrossRef](#)]
100. Larsson, J.; Gottfries, J.; Bohlin, L.; Backlund, A. Expanding the ChemGPS chemical space with natural products. *J. Nat. Prod.* **2005**, *68*, 985–991. [[CrossRef](#)]
101. Larsson, J.; Gottfries, J.; Muresan, S.; Backlund, A. ChemGPS-NP: Tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **2007**, *70*, 789–794. [[CrossRef](#)]
102. Rosén, J.; Lövgren, A.; Kogej, T.; Muresan, S.; Gottfries, J.; Backlund, A. ChemGPS-NP(Web): Chemical space navigation online. *J. Comput. Aided Mol. Des.* **2009**, *23*, 253–259. [[CrossRef](#)]
103. Koch, M.A.; Schuffenhauer, A.; Scheck, M.; Wetzels, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272–17277. [[CrossRef](#)]
104. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [[CrossRef](#)]
105. Ertl, P.; Schuhmann, T. Cheminformatics analysis of natural product scaffolds: Comparison of scaffolds produced by animals, plants, fungi and bacteria. *Mol. Inform.* **2020**, *39*, e2000017. [[CrossRef](#)]
106. Schafer, T.; Kriege, N.; Humbeck, L.; Klein, K.; Koch, O.; Mutzel, P. Scaffold Hunter: A comprehensive visual analytics framework for drug discovery. *J. Cheminform.* **2017**, *9*, 28. [[CrossRef](#)]
107. Voser, T.M.; Campbell, M.D.; Carroll, A.R. How different are marine microbial natural products compared to their terrestrial counterparts? *Nat. Prod. Rep.* **2022**, *39*, 7–19. [[CrossRef](#)]
108. Over, B.; Wetzels, S.; Grutter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* **2013**, *5*, 21–28. [[CrossRef](#)]
109. Elion, G.B.; Hitchings, G.H. The synthesis of 6-thioguanine. *J. Am. Chem. Soc.* **2002**, *77*, 1676. [[CrossRef](#)]
110. Coyne, S.; Chizzali, C.; Khalil, M.N.; Litomska, A.; Richter, K.; Beerhues, L.; Hertweck, C. Biosynthesis of the antimetabolite 6-thioguanine in *Erwinia amylovora* plays a key role in fire blight pathogenesis. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 10564–10568. [[CrossRef](#)]
111. Grigalunas, M.; Brakmann, S.; Waldmann, H. Chemical evolution of natural product structure. *J. Am. Chem. Soc.* **2022**, *144*, 3314–3329. [[CrossRef](#)]
112. Müller, G. Medicinal chemistry of target family-directed masterkeys. *Drug Discov. Today* **2003**, *8*, 681–691. [[CrossRef](#)]
113. Bon, R.S.; Waldmann, H. Bioactivity-guided navigation of chemical space. *Acc. Chem. Res.* **2010**, *43*, 1103–1114. [[CrossRef](#)]
114. Wetzels, S.; Bon, R.S.; Kumar, K.; Waldmann, H. Biology-oriented synthesis. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10800–10826. [[CrossRef](#)]
115. Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on natural products for drug design. *Nat. Chem.* **2016**, *8*, 531–541. [[CrossRef](#)]
116. Seiple, I.B.; Zhang, Z.; Jakubec, P.; Langlois-Mercier, A.; Wright, P.M.; Hog, D.T.; Yabu, K.; Allu, S.R.; Fukuzaki, T.; Carlsen, P.N.; et al. A platform for the discovery of new macrolide antibiotics. *Nature* **2016**, *533*, 338–345. [[CrossRef](#)] [[PubMed](#)]
117. Köst, Z.A.; Szklarski, A.R.; Pellegrino, S.; Michalak, S.E.; Meyer, M.; Zanette, C.; Cencic, R.; Nam, S.; Voora, V.K.; Horne, D.A.; et al. Synthesis facilitates an understanding of the structural basis for translation inhibition by the lissoclimides. *Nat. Chem.* **2017**, *9*, 1140–1149. [[CrossRef](#)]
118. Tan, D.S.; Foley, M.A.; Shair, M.D.; Schreiber, S.L. Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays. *J. Am. Chem. Soc.* **1998**, *120*, 8565–8566. [[CrossRef](#)]
119. Galloway, W.R.J.D.; Isidro-Llobet, A.; Spring, D.R. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nat. Commun.* **2010**, *1*, 80. [[CrossRef](#)]

120. Schreiber, S.L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964–1969. [[CrossRef](#)]
121. Wender, P.A. Toward the ideal synthesis and molecular function through synthesis-informed design. *Nat. Prod. Rep.* **2014**, *31*, 433–440. [[CrossRef](#)] [[PubMed](#)]
122. Cremosnik, G.S.; Liu, J.; Waldmann, H. Guided by evolution: From biology oriented synthesis to pseudo natural products. *Nat. Prod. Rep.* **2020**, *37*, 1497–1510. [[CrossRef](#)] [[PubMed](#)]
123. Karageorgis, G.; Foley, D.J.; Laraia, L.; Waldmann, H. Principle and design of pseudo-natural products. *Nat. Chem.* **2020**, *12*, 227–235. [[CrossRef](#)]
124. Karageorgis, G.; Foley, D.J.; Laraia, L.; Brakmann, S.; Waldmann, H. Pseudo natural products-chemical evolution of natural product structure. *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 15705–15723. [[CrossRef](#)]
125. van Hattum, H.; Waldmann, H. Biology-oriented synthesis: Harnessing the power of evolution. *J. Am. Chem. Soc.* **2014**, *136*, 11853–11859. [[CrossRef](#)]
126. Abbasov, M.E.; Alvariño, R.; Chaheine, C.M.; Alonso, E.; Sánchez, J.A.; Conner, M.L.; Alfonso, A.; Jaspars, M.; Botana, L.M.; Romo, D. Simplified immunosuppressive and neuroprotective agents based on gracilin A. *Nat. Chem.* **2019**, *11*, 342–350. [[CrossRef](#)] [[PubMed](#)]
127. Wermuth, C.G.; Ganellin, C.R.; Lindberg, P.; Mitscher, L.A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70*, 1129–1143. [[CrossRef](#)]
128. Hopkins, A.L.; Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730. [[CrossRef](#)] [[PubMed](#)]
129. Fang, J.; Wu, Z.; Cai, C.; Wang, Q.; Tang, Y.; Cheng, F. Quantitative and systems pharmacology. 1. In silico prediction of drug-target interactions of natural products enables new targeted cancer therapy. *J. Chem. Inf. Model.* **2017**, *57*, 2657–2671. [[CrossRef](#)] [[PubMed](#)]
130. Freshour, S.L.; Kiwala, S.; Cotto, K.C.; Coffman, A.C.; McMichael, J.F.; Song, J.J.; Griffith, M.; Griffith, O.L.; Wagner, A.H. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* **2021**, *49*, D1144–D1151. [[CrossRef](#)]
131. Tang, J.; Tanoli, Z.U.; Ravikumar, B.; Alam, Z.; Rebane, A.; Vaha-Koskela, M.; Peddinti, G.; van Adrichem, A.J.; Wakkinen, J.; Jaiswal, A.; et al. Drug Target Commons: A community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem. Biol.* **2018**, *25*, 224–229.e2. [[CrossRef](#)]
132. Pillich, R.T.; Chen, J.; Churas, C.; Liu, S.; Ono, K.; Otasek, D.; Pratt, D. NDEX: Accessing network models and streamlining network biology workflows. *Curr. Protoc.* **2021**, *1*, e258. [[CrossRef](#)]
133. Wu, Z.; Cheng, F.; Li, J.; Li, W.; Liu, G.; Tang, Y. SDTNBI: An integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief. Bioinform.* **2017**, *18*, 333–347. [[CrossRef](#)]
134. Wu, Z.; Ma, H.; Liu, Z.; Zheng, L.; Yu, Z.; Cao, S.; Fang, W.; Wu, L.; Li, W.; Liu, G.; et al. wSDTNBI: A novel network-based inference method for virtual screening. *Chem. Sci.* **2022**, *13*, 1060–1079. [[CrossRef](#)]
135. Wu, Z.; Li, W.; Liu, G.; Tang, Y. Network-based methods for prediction of drug-target interactions. *Front. Pharmacol.* **2018**, *9*, 1134. [[CrossRef](#)]
136. Gfeller, D.; Michielin, O.; Zoete, V. Shaping the interaction landscape of bioactive molecules. *Bioinformatics* **2013**, *29*, 3073–3079. [[CrossRef](#)] [[PubMed](#)]
137. Gfeller, D.; Grosdidier, A.; Wirth, M.; Daina, A.; Michielin, O.; Zoete, V. SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* **2014**, *42*, W32–W38. [[CrossRef](#)] [[PubMed](#)]
138. Gentile, F.; Yaacoub, J.C.; Gleave, J.; Fernandez, M.; Ton, A.T.; Ban, F.; Stern, A.; Cherkasov, A. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **2022**, *17*, 672–697. [[CrossRef](#)] [[PubMed](#)]
139. Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner, C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijjer, M.B.; Matos, R.C.; Tran, T.B.; et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181. [[CrossRef](#)] [[PubMed](#)]
140. Lounkine, E.; Keiser, M.J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J.L.; Lavan, P.; Weber, E.; Doak, A.K.; Cote, S.; et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **2012**, *486*, 361–367. [[CrossRef](#)]
141. McGovern, S.L.; Caselli, E.; Grigorieff, N.; Shoichet, B.K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002**, *45*, 1712–1722. [[CrossRef](#)]
142. Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [[CrossRef](#)]
143. Baell, J.B. Feeling Nature’s PAINS: Natural products, natural product drugs, and pan assay interference compounds (PAINS). *J. Nat. Prod.* **2016**, *79*, 616–628. [[CrossRef](#)]
144. Bisson, J.; McAlpine, J.B.; Friesen, J.B.; Chen, S.N.; Graham, J.; Pauli, G.F. Can invalid bioactives undermine natural product-based drug discovery? *J. Med. Chem.* **2016**, *59*, 1671–1690. [[CrossRef](#)]
145. Baell, J.B.; Nissink, J.W.M. Seven year itch: Pan-assay interference compounds (PAINS) in 2017-utility and limitations. *ACS Chem. Biol.* **2018**, *13*, 36–44. [[CrossRef](#)]
146. Stork, C.; Chen, Y.; Sicho, M.; Kirchmair, J. Hit Dexter 2.0: Machine-learning models for the prediction of frequent hitters. *J. Chem. Inf. Model.* **2019**, *59*, 1030–1043. [[CrossRef](#)] [[PubMed](#)]
147. Quiñoà, E.; Adamczeski, M.; Crews, P.; Bakus, G.J. Bengamides, heterocyclic anthelmintics from a Jaspidae marine sponge. *J. Org. Chem.* **1986**, *51*, 4494–4497. [[CrossRef](#)]
148. Wang, Y.Q.; Miao, Z.H. Marine-derived angiogenesis inhibitors for cancer therapy. *Mar. Drugs* **2013**, *11*, 903–933. [[CrossRef](#)]

149. White, K.N.; Tenney, K.; Crews, P. The bengamides: A mini-review of natural sources, analogues, biological properties, biosynthetic origins, and future prospects. *J. Nat. Prod.* **2017**, *80*, 740–755. [CrossRef]
150. Towbin, H.; Bair, K.W.; DeCaprio, J.A.; Eck, M.J.; Kim, S.; Kinder, F.R.; Morollo, A.; Mueller, D.R.; Schindler, P.; Song, H.K.; et al. Proteomics-based target identification: Bengamides as a new class of methionine aminopeptidase inhibitors. *J. Biol. Chem.* **2003**, *278*, 52964–52971. [CrossRef]
151. Xu, W.; Lu, J.P.; Ye, Q.Z. Structural analysis of bengamide derivatives as inhibitors of methionine aminopeptidases. *J. Med. Chem.* **2012**, *55*, 8021–8027. [CrossRef] [PubMed]
152. Lu, J.P.; Yuan, X.H.; Yuan, H.; Wang, W.L.; Wan, B.; Franzblau, S.G.; Ye, Q.Z. Inhibition of *Mycobacterium tuberculosis* methionine aminopeptidases by bengamide derivatives. *ChemMedChem* **2011**, *6*, 1041–1048. [CrossRef] [PubMed]
153. Porrás-Alcalá, C.; Moya-Utrera, F.; García-Castro, M.; Sánchez-Ruiz, A.; López-Romero, J.M.; Pino-González, M.S.; Díaz-Morilla, A.; Kitamura, S.; Wolan, D.W.; Prados, J.; et al. The development of the bengamides as new antibiotics against drug-resistant bacteria. *Mar. Drugs* **2022**, *20*, 373. [CrossRef]
154. Liu, S.; Widom, J.; Kemp, C.W.; Crews, C.M.; Clardy, J. Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science* **1998**, *282*, 1324–1327. [CrossRef]
155. Bailey, L. Effect of fumagillin upon *Nosema apis* (Zander). *Nature* **1953**, *171*, 212–213. [CrossRef]
156. Rateb, M.E.; Houssen, W.E.; Schumacher, M.; Harrison, W.T.; Diederich, M.; Ebel, R.; Jaspars, M. Bioactive diterpene derivatives from the marine sponge *Spongionella* sp. *J. Nat. Prod.* **2009**, *72*, 1471–1476. [CrossRef] [PubMed]
157. Potts, B.C.; Faulkner, D.J.; Jacobs, R.S. Phospholipase A<sub>2</sub> inhibitors from marine organisms. *J. Nat. Prod.* **1992**, *55*, 1701–1717. [CrossRef] [PubMed]
158. Sánchez, J.A.; Alfonso, A.; Leirós, M.; Alonso, E.; Rateb, M.E.; Jaspars, M.; Houssen, W.E.; Ebel, R.; Tabudravu, J.; Botana, L.M. Identification of *Spongionella* compounds as cyclosporine A mimics. *Pharmacol. Res.* **2016**, *107*, 407–414. [CrossRef]
159. Li, K.; Gustafson, K.R. Sesterterpenoids: Chemistry, biology, and biosynthesis. *Nat. Prod. Rep.* **2021**, *38*, 1251–1281. [CrossRef] [PubMed]
160. Ebada, S.S.; Lin, W.; Proksch, P. Bioactive sesterterpenes and triterpenes from marine sponges: Occurrence and pharmacological significance. *Mar. Drugs* **2010**, *8*, 313–346. [CrossRef] [PubMed]
161. Posadas, I.; De Rosa, S.; Terencio, M.C.; Paya, M.; Alcaraz, M.J. Cacospongionolide B suppresses the expression of inflammatory enzymes and tumour necrosis factor- $\alpha$  by inhibiting nuclear factor- $\kappa$ B activation. *Br. J. Pharmacol.* **2003**, *138*, 1571–1579. [CrossRef]
162. Gunasekera, S.P.; McCarthy, P.J.; Kelly-Borges, M.; Lobkovsky, E.; Clardy, J. Dysidiolide: A novel protein phosphatase inhibitor from the Caribbean sponge *Dysidea etheria* de Laubenfels. *J. Am. Chem. Soc.* **1996**, *118*, 8759–8760. [CrossRef]
163. Marcos, I.S.; Escola, M.A.; Moro, R.F.; Basabe, P.; Diez, D.; Sanz, F.; Mollinedo, F.; de la Iglesia-Vicente, J.; Sierra, B.G.; Urones, J.G. Synthesis of novel antitumoural analogues of dysidiolide from *ent*-halimic acid. *Bioorg. Med. Chem.* **2007**, *15*, 5719–5737. [CrossRef]
164. Dekker, F.J.; Koch, M.A.; Waldmann, H. Protein structure similarity clustering (PSSC) and natural product structure as inspiration sources for drug development and chemical genomics. *Curr. Opin. Chem. Biol.* **2005**, *9*, 232–239. [CrossRef]
165. Koch, M.A.; Wittenberg, L.O.; Basu, S.; Jeyaraj, D.A.; Gourzoulidou, E.; Reinecke, K.; Odermatt, A.; Waldmann, H. Compound library development guided by protein structure similarity clustering and natural product structure. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16721–16726. [CrossRef]
166. Pina, I.C.; Gautschi, J.T.; Wang, G.Y.; Sanders, M.L.; Schmitz, F.J.; France, D.; Cornell-Kennon, S.; Sambucetti, L.C.; Remiszewski, S.W.; Perez, L.B.; et al. Psammaplins from the sponge *Pseudoceratina purpurea*: Inhibition of both histone deacetylase and DNA methyltransferase. *J. Org. Chem.* **2003**, *68*, 3866–3873. [CrossRef] [PubMed]
167. Kim, D.H.; Shin, J.; Kwon, H.J. Psammaplin A is a natural prodrug that inhibits class I histone deacetylase. *Exp. Mol. Med.* **2007**, *39*, 47–55. [CrossRef] [PubMed]
168. El-Sayed, A.M. The Pherobase: Database of Pheromones and Semiochemicals; 2022. Available online: <https://pherolist.org/> (accessed on 20 December 2022).
169. Oluwabusola, E.T.; Katermeran, N.P.; Poh, W.H.; Goh, T.M.B.; Tan, L.T.; Diyaolu, O.; Tabudravu, J.; Ebel, R.; Rice, S.A.; Jaspars, M. Inhibition of the quorum sensing system, elastase production and biofilm formation in *Pseudomonas aeruginosa* by psammaplin A and bisaprasin. *Molecules* **2022**, *27*, 1721. [CrossRef]
170. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef] [PubMed]
171. Miller, B.; Friedman, A.J.; Choi, H.; Hogan, J.; McCammon, J.A.; Hook, V.; Gerwick, W.H. The marine cyanobacterial metabolite gallinamide A is a potent and selective inhibitor of human cathepsin L. *J. Nat. Prod.* **2014**, *77*, 92–99. [CrossRef] [PubMed]
172. Boudreau, P.D.; Miller, B.W.; McCall, L.I.; Almaliti, J.; Reher, R.; Hirata, K.; Le, T.; Siqueira-Neto, J.L.; Hook, V.; Gerwick, W.H. Design of gallinamide A analogs as potent inhibitors of the cysteine proteases human cathepsin L and *Trypanosoma cruzi*. *J. Med. Chem.* **2019**, *62*, 9026–9044. [CrossRef]
173. Barbosa Da Silva, E.; Sharma, V.; Hernandez-Alvarez, L.; Tang, A.H.; Stoye, A.; O'Donoghue, A.J.; Gerwick, W.H.; Payne, R.J.; McKerrow, J.H.; Podust, L.M. Intramolecular interactions enhance the potency of gallinamide A analogues against *Trypanosoma cruzi*. *J. Med. Chem.* **2022**, *65*, 4255–4269. [CrossRef]

174. Klenchin, V.A.; King, R.; Tanaka, J.; Marriott, G.; Rayment, I. Structural basis of swinholide A binding to actin. *Chem. Biol.* **2005**, *12*, 287–291. [[CrossRef](#)]
175. Issac, M.; Aknin, M.; Gauvin-Bialecki, A.; De Voogd, N.; Ledoux, A.; Frederich, M.; Kashman, Y.; Carmeli, S. Cyclotheonellazoles A-C, potent protease inhibitors from the marine sponge *Theonella aff. swinhoei*. *J. Nat. Prod.* **2017**, *80*, 1110–1116. [[CrossRef](#)]
176. Cui, Y.; Zhang, M.; Xu, H.; Zhang, T.; Zhang, S.; Zhao, X.; Jiang, P.; Li, J.; Ye, B.; Sun, Y.; et al. Elastase inhibitor cyclotheonellazole A: Total synthesis and in vivo biological evaluation for acute lung injury. *J. Med. Chem.* **2022**, *65*, 2971–2987. [[CrossRef](#)]
177. Köcher, S.; Resch, S.; Kessenbrock, T.; Schräpp, L.; Ehrmann, M.; Kaiser, M. From dolastatin 13 to cyanopeptolins, micropeptides, and lymgbyastatins: The chemical biology of Ahp-cyclodepsipeptides. *Nat. Prod. Rep.* **2020**, *37*, 163–174. [[CrossRef](#)] [[PubMed](#)]
178. Chen, Q.Y.; Luo, D.; Seabra, G.M.; Luesch, H. Ahp-Cyclodepsipeptides as tunable inhibitors of human neutrophil elastase and kallikrein 7: Total synthesis of tutuilamide A, serine protease selectivity profile and comparison with lymgbyastatin 7. *Bioorg. Med. Chem.* **2020**, *28*, 115756. [[CrossRef](#)] [[PubMed](#)]
179. Pettit, G.R.; Kamano, Y.; Herald, C.L.; Dufresne, C.; Cerny, R.L.; Herald, D.L.; Schmidt, J.M.; Kizu, H. Antineoplastic agent. 174. Isolation and structure of the cytostatic depsipeptide dolastatin 13 from the sea hare *Dolabella auricularia*. *J. Am. Chem. Soc.* **1989**, *111*, 5015–5017. [[CrossRef](#)]
180. Gunasekera, S.P.; Miller, M.W.; Kwan, J.C.; Luesch, H.; Paul, V.J. Molassamide, a depsipeptide serine protease inhibitor from the marine cyanobacterium *Dichothrix utahensis*. *J. Nat. Prod.* **2010**, *73*, 459–462. [[CrossRef](#)] [[PubMed](#)]
181. Salvador, L.A.; Taori, K.; Biggs, J.S.; Jakoncic, J.; Ostrov, D.A.; Paul, V.J.; Luesch, H. Potent elastase inhibitors from cyanobacteria: Structural basis and mechanisms mediating cytoprotective and anti-inflammatory effects in bronchial epithelial cells. *J. Med. Chem.* **2013**, *56*, 1276–1290. [[CrossRef](#)]
182. Yu, M.; Dolios, G.; Petrick, L. Reproducible untargeted metabolomics workflow for exhaustive MS2 data acquisition of MS1 features. *J. Cheminform.* **2022**, *14*, 6. [[CrossRef](#)]
183. Keller, L.; Canuto, K.M.; Liu, C.; Suzuki, B.M.; Almaliti, J.; Sikandar, A.; Naman, C.B.; Glukhov, E.; Luo, D.; Duggan, B.M.; et al. Tutuilamides A-C: Vinyl-chloride-containing cyclodepsipeptides from marine cyanobacteria with potent elastase inhibitory properties. *ACS Chem. Biol.* **2020**, *15*, 751–757. [[CrossRef](#)]
184. Reher, R.; Aron, A.T.; Fajtova, P.; Stincone, P.; Wagner, B.; Perez-Lorente, A.I.; Liu, C.; Shalom, I.Y.B.; Bittremieux, W.; Wang, M.; et al. Native metabolomics identifies the rivulariapeptolide family of protease inhibitors. *Nat. Commun.* **2022**, *13*, 4619. [[CrossRef](#)]
185. Hanada, M.; Sugawara, K.; Kaneta, K.; Toda, S.; Nishiyama, Y.; Tomita, K.; Yamamoto, H.; Konishi, M.; Oki, T. Epoxomicin, a new antitumor agent of microbial origin. *J. Antibiot.* **1992**, *45*, 1746–1752. [[CrossRef](#)]
186. Levin, N.; Spencer, A.; Harrison, S.J.; Chauhan, D.; Burrows, F.J.; Anderson, K.C.; Reich, S.D.; Richardson, P.G.; Trikha, M. Marizomib irreversibly inhibits proteasome to overcome compensatory hyperactivation in multiple myeloma and solid tumour patients. *Br. J. Haematol.* **2016**, *174*, 711–720. [[CrossRef](#)]
187. Potts, B.C.; Albitar, M.X.; Anderson, K.C.; Baritaki, S.; Berkers, C.; Bonavida, B.; Chandra, J.; Chauhan, D.; Cusack, J.C., Jr.; Fenical, W.; et al. Marizomib, a proteasome inhibitor for all seasons: Preclinical profile and a framework for clinical trials. *Curr. Cancer Drug Targets* **2011**, *11*, 254–284. [[CrossRef](#)] [[PubMed](#)]
188. Hubbell, G.E.; Tepe, J.J. Natural product scaffolds as inspiration for the design and synthesis of 20S human proteasome inhibitors. *RSC Chem. Biol.* **2020**, *1*, 305–332. [[CrossRef](#)] [[PubMed](#)]
189. Kawamura, S.; Unno, Y.; Tanaka, M.; Sasaki, T.; Yamano, A.; Hirokawa, T.; Kameda, T.; Asai, A.; Arisawa, M.; Shuto, S. Investigation of the noncovalent binding mode of covalent proteasome inhibitors around the transition state by combined use of cyclopropylic strain-based conformational restriction and computational modeling. *J. Med. Chem.* **2013**, *56*, 5829–5842. [[CrossRef](#)] [[PubMed](#)]
190. Li, T.; Wang, N.; Zhang, T.; Zhang, B.; Sajeevan, T.P.; Joseph, V.; Armstrong, L.; He, S.; Yan, X.; Naman, C.B. A systematic review of recently reported marine derived natural product kinase inhibitors. *Mar. Drugs* **2019**, *17*, 493. [[CrossRef](#)] [[PubMed](#)]
191. Kinnel, R.B.; Scheuer, P.J. 11-Hydroxystaurosporine: A highly cytotoxic, powerful protein kinase C inhibitor from a tunicate. *J. Org. Chem.* **1992**, *57*, 6327–6329. [[CrossRef](#)]
192. Cohen, P.; Cross, D.; Janne, P.A. Kinase drug discovery 20 years after imatinib: Progress and future directions. *Nat. Rev. Drug Discov.* **2021**, *20*, 551–569. [[CrossRef](#)]
193. Karaman, M.W.; Herrgard, S.; Treiber, D.K.; Gallant, P.; Atteridge, C.E.; Campbell, B.T.; Chan, K.W.; Ciceri, P.; Davis, M.I.; Edeen, P.T.; et al. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132. [[CrossRef](#)]
194. Lo, Y.C.; Liu, T.; Morrissey, K.M.; Kakiuchi-Kiyota, S.; Johnson, A.R.; Broccatelli, F.; Zhong, Y.; Joshi, A.; Altman, R.B. Computational analysis of kinase inhibitor selectivity using structural knowledge. *Bioinformatics* **2019**, *35*, 235–242. [[CrossRef](#)]
195. Dai, X.; Xu, Y.; Qiu, H.; Qian, X.; Lin, M.; Luo, L.; Zhao, Y.; Huang, D.; Zhang, Y.; Chen, Y.; et al. KID: A kinase-focused interaction database and its application in the construction of kinase-focused molecule databases. *J. Chem. Inf. Model.* **2022**, *62*, 6022–6034. [[CrossRef](#)]
196. Manning, G.; Whyte, D.B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912–1934. [[CrossRef](#)]
197. Janssen, A.P.A.; Grimm, S.H.; Wijdeven, R.H.M.; Lenselink, E.B.; Neefjes, J.; van Boeckel, C.A.A.; van Westen, G.J.P.; van der Stelt, M. Drug discovery maps, a machine learning model that visualizes and predicts kinome-inhibitor interaction landscapes. *J. Chem. Inf. Model.* **2019**, *59*, 1221–1229. [[CrossRef](#)]

198. Aoki, S.; Watanabe, Y.; Sanagawa, M.; Setiawan, A.; Kotoku, N.; Kobayashi, M. Cortistatins A, B, C, and D, anti-angiogenic steroidal alkaloids, from the marine sponge *Corticium simplex*. *J. Am. Chem. Soc.* **2006**, *128*, 3148–3149. [[CrossRef](#)] [[PubMed](#)]
199. Pelish, H.E.; Liao, B.B.; Nitulescu, II; Tangpeerachaikul, A.; Poss, Z.C.; Da Silva, D.H.; Caruso, B.T.; Arefolov, A.; Fadeyi, O.; Christie, A.L.; et al. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **2015**, *526*, 273–276. [[CrossRef](#)] [[PubMed](#)]
200. Aoki, S.; Watanabe, Y.; Tanabe, D.; Arai, M.; Suna, H.; Miyamoto, K.; Tsujibo, H.; Tsujikawa, K.; Yamamoto, H.; Kobayashi, M. Structure-activity relationship and biological property of cortistatins, anti-angiogenic spongean steroidal alkaloids. *Bioorg. Med. Chem.* **2007**, *15*, 6758–6762. [[CrossRef](#)]
201. Funabashi, M.; Baba, S.; Takatsu, T.; Kizuka, M.; Ohata, Y.; Tanaka, M.; Nonaka, K.; Spork, A.P.; Ducho, C.; Chen, W.C.; et al. Structure-based gene targeting discovery of sphaerimicin, a bacterial translocase I inhibitor. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 11607–11611. [[CrossRef](#)] [[PubMed](#)]
202. Li, L.; Gui, Y.H.; Xu, Q.H.; Lin, H.W.; Lu, Y.H. *Spongiactinospora rosea* gen. nov., sp. nov., a new member of the family Streptosporangiaceae. *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 427–433. [[CrossRef](#)] [[PubMed](#)]
203. Nakaya, T.; Yabe, M.; Mashalidis, E.H.; Sato, T.; Yamamoto, K.; Hikiji, Y.; Katsuyama, A.; Shinohara, M.; Minato, Y.; Takahashi, S.; et al. Synthesis of macrocyclic nucleoside antibacterials and their interactions with MraY. *Nat. Commun.* **2022**, *13*, 7575. [[CrossRef](#)]
204. Pereira, R.B.; Evdokimov, N.M.; Lefranc, F.; Valentao, P.; Kornienko, A.; Pereira, D.M.; Andrade, P.B.; Gomes, N.G.M. Marine-derived anticancer agents: Clinical benefits, innovative mechanisms, and new targets. *Mar. Drugs* **2019**, *17*, 329. [[CrossRef](#)]
205. Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855–861. [[CrossRef](#)]
206. Hight, S.K.; Clark, T.N.; Kurita, K.L.; McMillan, E.A.; Bray, W.; Shaikh, A.F.; Khadilkar, A.; Haeckl, F.P.J.; Carnevale-Neto, F.; La, S.; et al. High-throughput functional annotation of natural products by integrated activity profiling. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2208458119. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.