Review

Computational Methodologies in the Exploration of Marine Natural Product Leads

Florbela Pereira * and Joao Aires-de-Sousa *

LAQV and REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

* Correspondence: florbela.pereira@fct.unl.pt (F.P.); joao@airesdesousa.com (J.A.-d.-S.);
Tel.: +351-21-2948300 (F.P.); +351-21-2948300 (J.A.-d.-S.)

check for
updates

Abstract: Computational methodologies are assisting the exploration of marine natural products (MNPs) to make the discovery of new leads more efficient, to repurpose known MNPs, to target new metabolites on the basis of genome analysis, to reveal mechanisms of action, and to optimize leads. In silico efforts in drug discovery of NPs have mainly focused on two tasks: dereplication and prediction of bioactivities. The exploration of new chemical spaces and the application of predicted spectral data must be included in new approaches to select species, extracts, and growth conditions with maximum probabilities of medicinal chemistry novelty. In this review, the most relevant current computational dereplication methodologies are highlighted. Structure-based (SB) and ligand-based (LB) chemoinformatics approaches have become essential tools for the virtual screening of NPs either in small datasets of isolated compounds or in large-scale databases. The most common LB techniques include Quantitative Structure–Activity Relationships (QSAR), estimation of drug likeness, prediction of adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties, similarity searching, and pharmacophore identification. Analogously, molecular dynamics, docking and binding cavity analysis have been used in SB approaches. Their significance and achievements are the main focus of this review.

Keywords: Computer-Aided Drug Design (CADD); drug discovery; chemoinformatics; bioinformatics; machine learning (ML); marine natural products (MNPs)

1. Introduction

Drug research and development (R&D) is comprehensive, complex, expensive, time-consuming, and full of risk. A 2016 study [1] reported a clinical success rate, i.e., the likelihood that a drug that enters clinical testing will eventually be approved, of approximately 12%. The development of a drug from concept to market currently takes 13–15 years and requires United States $2–3 billion on average [2]. Although such costs are going up, the number of drugs approved every year per billion dollars spent on R&D has remained flat or decreased for most of the past decade [3]. Several new methodologies have been developed and applied in drug R&D to shorten the research cycle and to reduce the costs. Computational methodologies have been instrumental at various stages of drug discovery [4,5] and continue to be indispensable in the incessant demand for life-saving drugs. Computer-Aided Drug Design (CADD) methods have emerged as a powerful tool in the development of therapeutically important small molecules for over three decades [6–8], enabling higher hit rates than experimental high-throughput screening (HTS) approaches alone [6]. For example, Mueller et al. [6] built a computational model using results from a previous HTS of metabotropic glutamate receptor 5 (mGlu5) activity [9], which was able to identify new lead-like mGlu5 modulators in a virtual screening experiment with a hit rate of 3.6% [6]: an enrichment factor of approximately 16 compared

with the original experimental HTS data (0.22%) [9]. Nowadays, CADD methodologies have been extended from their more conventional application of lead discovery and optimization toward new directions, e.g., target identification and validation and preclinical tests (prediction of adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties). They are generally classified in two categories, structure-based (SB) and ligand-based (LB), both of which have been used with marine natural products (MNPs). In this review, we highlight recent advances of CADD methodologies applied to NPs (particularly focusing on MNPs), as well as the importance of informatics in the analysis of marine extracts from an early stage of screening, to recognize and filter out known compounds (dereplication) [10,11]. Figure 1 illustrates the role of computational methodologies in a typical drug discovery pipeline and the importance of NPs and MNPs in this context.
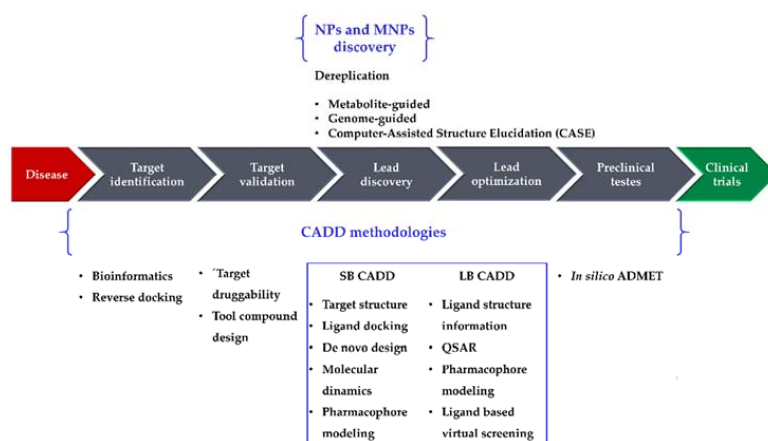

Figure 1. The drug discovery pipeline, computer-aided drug design (CADD), and natural product (NP)/marine natural product (MNP) discovery methodologies. SB, structure-based; LB, ligand-based; ADMET, adsorption, distribution, metabolism, excretion, and toxicity; QSAR, Quantitative Structure–Activity Relationship.

Statistics concerning novel drug approvals by the Food and Drug Administration (FDA) during 1969–2016 show a current downward trend since the high-point in 1996 (53 new molecular entities (NMEs)/year), and the minimum (after 1996) of 15 NMEs/year in 2010 and 2016. Figure 2 compares the global number of novel FDA approvals with the number of approvals of NP and derivatives between 1969 and 2016 and highlights the contribution of MNPs and CADD methodologies.
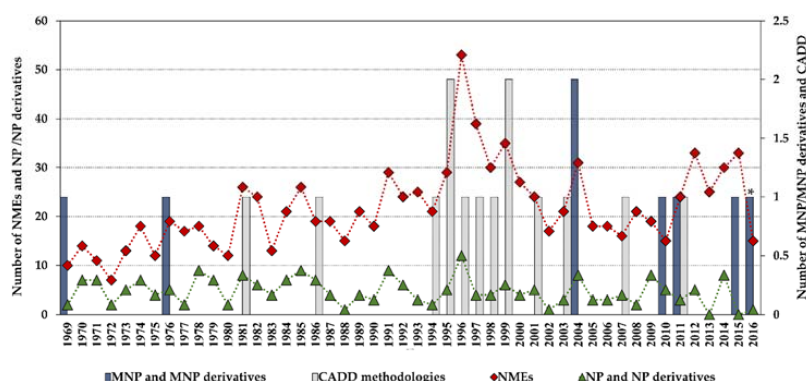

Figure 2. Novel Food and Drug Administration (FDA) approvals during 1969–2016, where new molecular entities (NMEs) are all approvals except biologics license applications; NP and derivatives are all non-mammalian NPs except MNPs; CADD methodologies are approvals that were developed using CADD; * an MNP that is an European Medicines Agency (EMEA)-approved drug. Data are from Drugs@FDA and the literature [7,12–17].

Although the high-point in the mid to late 1990s appears to be mainly due to regulatory factors [3] (i.e., clearing of a backlog at the FDA following the implementation of the 1992 Prescription Drug User Fee Act, and political lobbying for human immunodeficiency virus (HIV) drugs, which lowered the normal regulatory hurdles), major advances in many of the scientific and technological inputs into R&D had been accomplished during the 1980s and 1990s. For example, combinatorial chemistry increased the capability to produce drug-like molecules by approximately 800 fold, increasing the size of the known chemical space [3,18,19]; faster DNA sequencing allowed the identification of new drug targets [20]; advancements in the elucidation of three-dimensional protein structures via X-ray crystallography facilitated the identification of lead compounds through structure-guided strategies[21]; the advent of HTS led to an explosion in the rate of data generation [22]; and computational drug design and screening were implemented [7]. Interestingly, the high-point for NP and derivatives was also in 1996 (with 12 approved drugs) and the 1990s decade was also the most successful for CADD-driven drugs, with eight approved drugs (Figure 3).
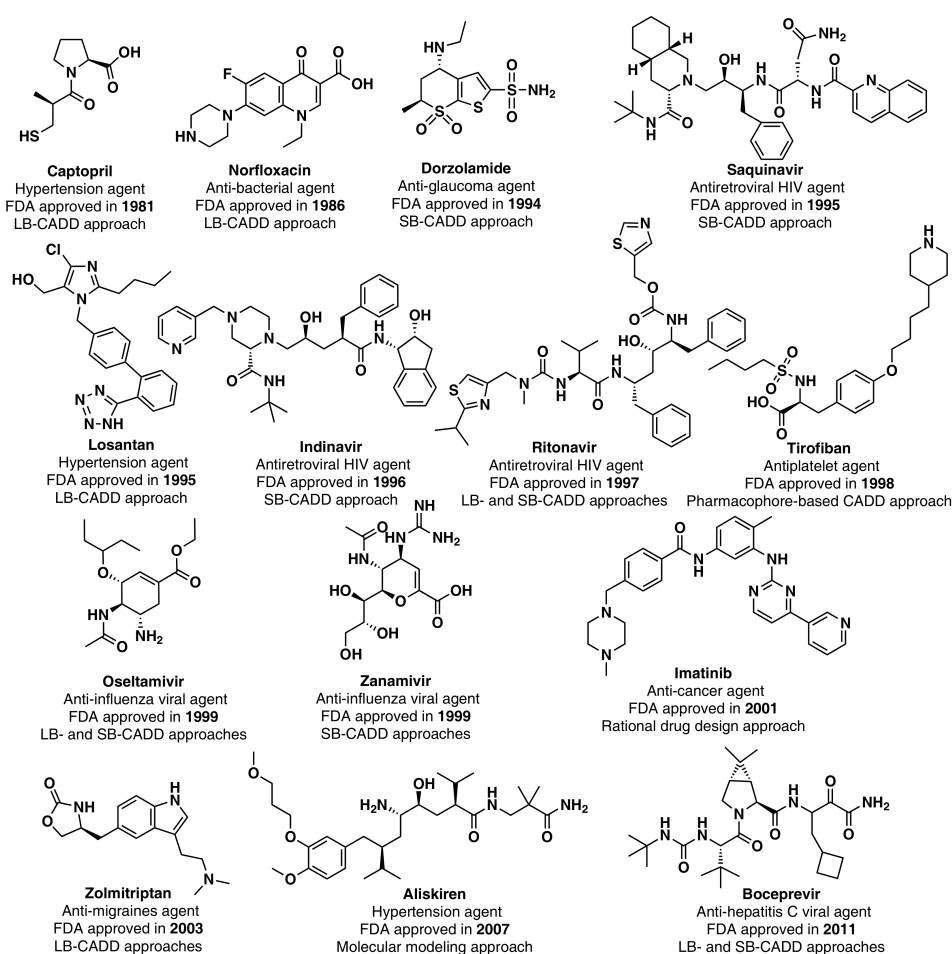

Figure 3. CADD-driven drugs and their chemical structures and clinical indication.

More than half of the total approvals of MNP and derivatives occurred in the 21st century (six out of eight approved drugs, Figure 4). The declining number of NMEs in development pipelines together with the higher success rate of marine compounds (1 in 3500 MNPs [13] against the industry average of 1 in 5000–10,000 compounds [23]) have led to the rekindling of interest in NP-like scaffolds [23,24]. More than 28,000 MNPs have been reported to date from a variety of marine sources (http://pubs.rsc.org/marinlit); in 2016, the literature reported 1277 new compounds [25] isolated from marine microorganisms and phytoplankton, green, brown, and red algae, sponges,

cnidarians, bryozoans, molluscs, tunicates, echinoderms, mangroves, and other intertidal plants and microorganisms. However, only eight MNPs have to date been approved as drugs (Figures 2 and 4), while 12 marine-derived metabolites are currently in different phases of clinical trials [25–29]. New approaches are needed to overcome the perceived disadvantages of NPs as compared with synthetic drugs, such as the difficulty in access and supply, the complexity of NP chemistry, and the inherent slowness of working with NPs [28,30]. Existing NP databases must be improved by filling missing activity records [31]. It shall be mentioned that the known biological activity space of MNPs has been biased due to funding sources, e.g., five out of the eight approved MNPs drugs are anti-cancer drugs (Figure 4). The emphasis on cancer is mainly due to the fact that the major funding agency in the U.S. for MNP and derivatives was for many years the National Institutes of Health (NIH)/National Cancer Institute (NCI), and this happened similarly in other countries [13].
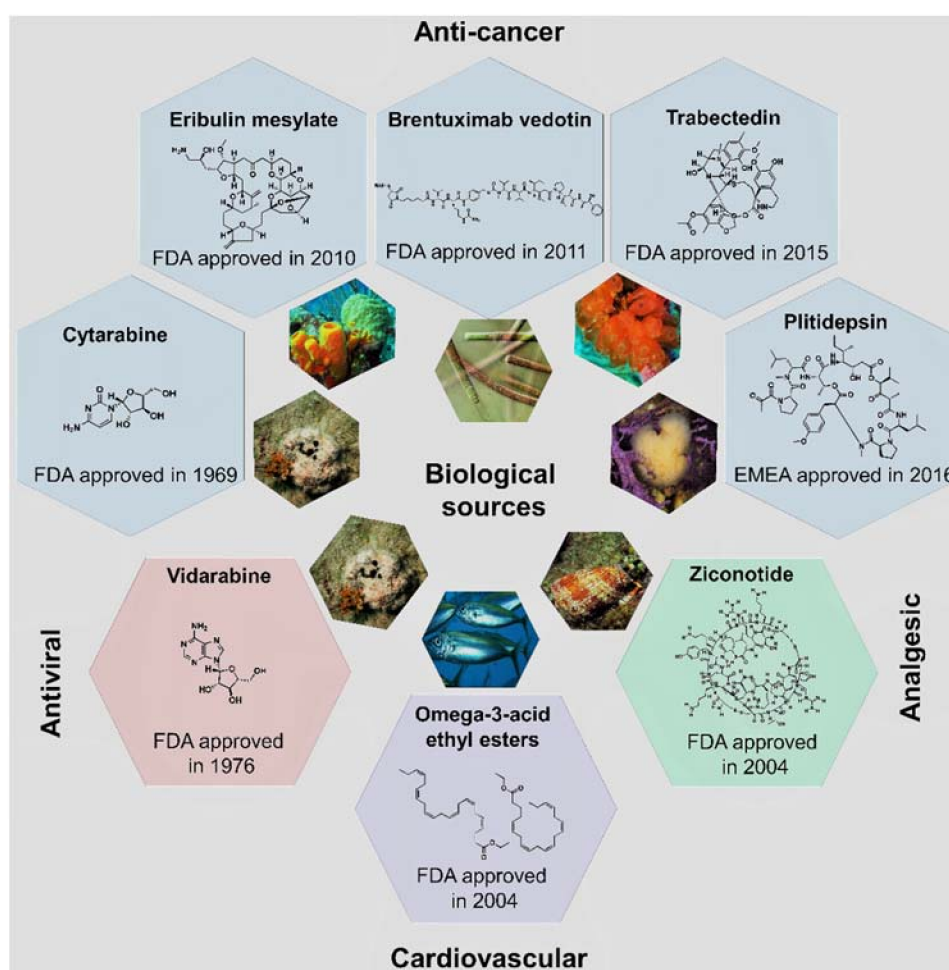

Figure 4. The eight approved MNP and derivative drugs and their biological sources, chemical structures, and clinical usage.

The vast majority of currently used antibiotics have been isolated from terrestrial microbes, accounting for more than 75% of all antibiotics discovered [32,33], but antimicrobial compounds from marine sources have not yet been developed into clinical testing phases [13,28]. Recently, the marine environment has been proposed as an untapped source of new bioactive molecules, and marine bacteria and fungi seem to be the most important sources for antibacterial discovery [28,34–36]. Computational methodologies are crucial in the systematic exploration of the biological activity of MNPs to improve the rate of drug discovery from marine sources. Their significance, achievements, and challenges are addressed in this review.

2. Databases

Specific databases of NPs and MNPs are available with physical, chemical, and biological properties. Furthermore, databases of larger scope also include compounds from marine sources, as well as similar compounds from other sources, and are useful resources for the development of MNP leads. The exploration of databases has become a well-established essential component of chemistry and biological research. Some of these databases are just collections of chemical structures, e.g., catalogues of commercially available samples for screening, while others provide additional data, such as measured bioactivities and protein targets as well as targeted diseases. Only a fraction of large general databases is directly related to NPs, but some exist that can assist in NP-based drug discovery and dereplication. To be useful for dereplication purposes, databases must cover extensively the chemical and biological space of the known NPs and must be searchable by several features, such as structure and substructure identity/similarity, spectroscopic identity/similarity, UV absorption maxima, accurate mass, physical properties, taxonomic identification of the producing macro- or micro-organism, biological activity, and biological targets. For CADD procedures, databases must provide compounds with their molecular structures in chemical file formats, bioactivity data (e.g., cell-based assays), and biomolecular targets. They contain advantageously medicinal chemistry data, NP data, approved drugs and failed drug candidates with data generated in the preclinical and clinical phases of drug discovery [37–39]. The most relevant databases for NPs as well as their searchable attributes are listed in Table 1 (the ReSpect and NaprAlert databases have not been updated since 2012 and 2016, respectively).

Table 1. Essential features of selected databases for NPs dereplication and CADD.

Database	Compounds [6]		Taxo. [7]	Bioact. [8]	Targets [9]	Spec. Data [10]
	Total	NPs				
CAS/SciFinder [1]	9.0×10^7	>283,000	+	+	-	-
ChemSpider [2]	5.9×10^7	>13,800	-	+	-	-
PubChem [2]	9.3×10^7	4.4×10^5	-	+	+	+ [10]
ChEMBL [2]	1.7×10^6	>75,000	-	+	+	-
REAXYS [1,2]	1.1×10^8	>215,000	+	+	-	-
ZINC [2,5]	1.2×10^8	>44,000	-	+	+	-
LOPAC [3,5]	1280	?	-	+	+	-
Prestwick [3,5]	1280	?	-	+	+	-
ACD/NMR DB [4]	>322,000	>50,000	-	-	-	+ [10.2]
NMRShiftDB [4]	43,440	?	-	-	-	+ [10.2]
Massbank [4]	>15,000	>2500	-	-	-	+ [10.3]
ReSpect [4]	-	>3595	-	-	-	+ [10.3]
METLIN [4]	-	75,000	-	-	-	+ [10.3]
GNPS [4]	22,644	>3000	+	-	-	+ [10.3]
NaprAlert [4]	-	>155,000 [12]	+	+	-	+ [10.1]
DNP [4]	-	>270,000	+	+	-	+ [10.1]
DMNP [4]	-	>30,000	+	+	-	+ [10.1]
MarinLit [4]	-	>29,000	+	+	-	+ [10]
AntiBase [4]	-	43,743	+	+	-	+ [10]
StreptomeDB [4]	-	3991	+	+	-	+ [11]
NPCARE [2,4]	-	6578 [12]	+	+	+	-

[1] Comprehensive compilation of information on NPs with no specific application in view; [2] Particularly suitable for CADD applications; [3] Chemogenomic libraries that were conceived for cell-based high-throughput screening (HTS) assays, but are also suitable for CADD applications; [4] Suitable for dereplication applications; [5] Commercially available compounds; [6] When possible an estimate number of NPs in the database is given; [7] Taxonomy; [8] Bioactivity; [9] Biological targets; [10] Spectral data comprising [10.1] UV, [10.2] NMR, and [10.3] MS data; [11] Predicted ^{1}H/^{13}C NMR and MS spectra; [12] Comprising extracts, NPCARE contains 2566 fractional extracts isolated from 1952 distinct biological species, including plants, marine organisms, fungi, and bacteria.

Substructure searching is available for all databases reported in Table 1 with the exception of the NaprAlert and NPCARE databases. CAS/SciFinder, available at Scientific and Technical Network (http://www.cas.org/products/scifinder) is a commercial database comprising one of the largest online repository of NPs structures, although it has several search limitations to be applied in dereplication

procedures (e.g., it does not allow one to search by spectral data or accurate mass). Other commercially available databases are: REAXYS, licensed by Elsevier B.V. (https://new.reaxys.com), which provides access to experimentally measured data (physical, chemical, and pharmacological data); ACD/NMR DB from ACD/Labs (http://www.acdlabs.com/products/dbs/nmr_db), which consists of experimental NMR spectra, currently including 210,000 ^{1}H, >200,000 ^{13}C, 16,780 ^{19}F, 9200 ^{15}N, and >27,000 ^{31}P NMR spectra; NaprAlert (http://www.napralert.org); and the Chapman & Hall/CRC Dictionary of NPs (http://dnp.chemnetbase.com).

More specific databases particularly focusing on MNPs are the Chapman & Hall/CRC Dictionary of MNPs (http://dmnp.chemnetbase.com), MarinLit (http://pubs.rsc.org/marinlit/), and AntiBase (http://application.wiley-vch.de/stmdata/antibase.php). AntiBase covers terrestrial and marine microbial NPs and includes predicted ^{13}C NMR spectra for compounds with no available experimental spectra.

The remaining databases listed in Table 1 are freely available. The StreptomeDB (http://www.pharmaceutical-bioinformatics.org/streptomedb/) is a versatile platform for the gathering of information concerning the genus *Streptomyces*, an actinobacteria that has stirred huge interest as a source of bioactive compounds over the last few decades; all molecular structures can be downloaded with metadata in the MDL SD file format [40] NPCARE (http://silver.sejong.ac.kr/npcare) is an online database of NPs and fractional extracts for anticancer activities, which were validated with 1107 cell lines for 34 cancer types [41]. Each record is annotated with the cancer type, the genus, and species names of the biological resource, the cell line used for demonstrating the anticancer activity, the PubChem ID, and information about the target gene or protein.

ChemSpider (http://www.chemspider.com) is a curated chemical database, which was made available from the Royal Society of Chemistry (RSC) and contains data for compounds gathered from over 500 different sources [42]. PubChem (http://pubchem.ncbi.nlm.nih.gov) is probably the largest freely available collection of chemical information and one of the largest repositories of NPs; it is organized as three interlinked databases (Substance, Compound, and BioAssay) [38] and includes more than 234 million depositor-provided chemical substance descriptions, 93 million unique chemical structures, and 1.2 million biological assay descriptions, covering about 10,300 and 22,000 unique protein target and gene target sequences, respectively. ChEMBL (http://www.ebi.ac.uk/chembl) is a large-scale curated bioactivity database with information on molecule–target interactions retrieved from the published literature; it has been expanded both in terms of data content (e.g., a neglected tropical disease archive including datasets from GlaxoSmithKline, Novartis, St. Jude Children´s Research Hospital , FDA-approved drugs, and drug candidates in clinical development) and annotation (e.g., properties and efficacy targets for FDA-approved drugs and drug candidates in clinical development) [37]. The ZINC (http://zinc15.docking.org/), LOPAC, and Prestwick databases comprise commercially available molecules, thus linking available collections of samples for experimental screening to known targets. LOPAC, available from Sigma-Aldrich, is a chemogenomic library that contains pharmacologically relevant small molecule agents and a complete list of compounds and their annotated targets (more than 450 targets); more than 50% of the compounds target G-protein-coupled receptor (GPCR), similarly to approved drugs, making it particularly well-suited to screen for GPCR-related phenotypic effects [43]. The Prestwick chemical library (http://www.prestwickchemical.com/prestwick-chemical-library.html) is also a chemogenomic library with mostly approved drugs that were selected for target diversity (more than 100 targets) and known safety and bioavailability [43]. ZINC is a free database designed to bring together biology and chemoinformatics; it is simultaneously easy to use by non-specialists and fully programmable for chemoinformaticians and computational biologists. The ZINC 15 version [44] was expanded from an exclusively molecule-centric database (mainly used for virtual screening, ligand discovery, pharmacophore screening, benchmarking, and force field development) to one that connects molecules to biological targets, processes, and other bioactive small molecules; the biological annotations, such as the identification of molecules as metabolites, drugs, and NPs and

the identification of molecules as ligands for particular proteins and processes, were derived from other databases and libraries, e.g., HMDB [45], ChEMBL [37], and DrugBank [46]. Moreover, several NPs databases have also been incorporated in the ZINC database, namely: AfroDb [47], a database of NPs from African sources; HIM (Herbal Ingredients In-Vivo Metabolism database) [48]; NPACT (naturally occurring plant-based anti-cancer compound-activity-target database) [49]; NuBBE [50], a NPs database from the biodiversity of Brazil; and TCM database@Taiwan [51] with traditional Chinese medicine compounds. The use of databases in dereplication and CADD procedures is further discussed in Section 3.1.1 (Secondary-metabolite-guided identification) and Section 4.1 (Ligand-based CADD), respectively.

3. Dereplication

Dereplication involves the comparison of experimental data from new extracts with those of known NPs, and therefore computational methodologies associated with databases are essential to increase the chance of isolating new molecules efficiently. For reviews of the NP dereplication literature in general, the reader is referred to Gaudêncio and Pereira [10], Pérez-Victoria et al. [11], and Zhang et al. [52]. Mohamed et al. [53] reviewed computational resources for NPs dereplication, and Hufsky et al. [54] is suggested for a review of informatics methods for NP discovery. Here, we highlight the most relevant recent advances in computational dereplication methodologies employing computational mass spectrometry or NMR spectroscopy (metabolite-guided and genome-guided approaches) and computer-assisted structure elucidation (CASE), in particular those concerning MNPs or likely to be applied to MNPs. Genome mining is a strategy to aim at the isolation of novel NPs [55] as the identification of genes encoding for the biosynthesis of secondary metabolites can guide the exploration of extracts to identify anticipated new molecules.

3.1. Computer-Assisted Identification of Compounds

3.1.1. Secondary Metabolite-Guided

Different analytical techniques, such as liquid chromatography-mass spectrometry (LC-MS) [56,57], liquid chromatography-high resolution mass spectrometry (LC-HRMS) [58–60], liquid chromatography time-of-flight MS (LC-TOF-MS) [61], high-resolution electrospray ionisation mass spectrometry (HRESI-MS) [62], and NMR spectroscopy [59,60,62], have been applied for fast dereplication followed by multivariate data analysis to minimize redundancy in the isolation steps.

Chanana et al. [56] developed an LC-MS-based principal component analysis (PCA) workflow, which comprises a new script written in R (PoPCAR, Planes of Principal Component Analysis in R), to distinguish unique versus common metabolites in ~50 marine actinomycete strains. PoPCAR allows researchers to identify masses or molecules unique to each strain by locating those in a bucket table with a peak list, which can be generated using commercial software, such as Bruker ProfileAnalysis or open source tools, e.g., MZmine [63] or XCMS [64]. The AntiBase database was also integrated into this workflow. With this strategy, the authors were able to pinpoint the skeleton of forazoline, one of three classes of novel compounds previously identified from an *Actinomadura* sp. (Figure 5).
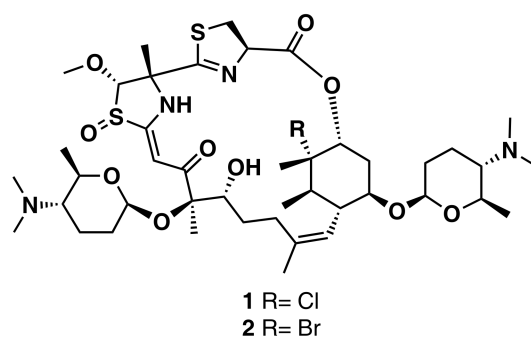
Figure 5. Chemical structures of the antifungal agents forazoline A (**1**) and B (**2**) isolated from an *Actinomadura* sp.

A similar approach was reported using PCA, hierarchical clustering (HCA), and orthogonal partial least square-discriminant analysis (OPLS-DA) to evaluate the high resolution Fourier transform mass spectrometry (HRFTMS) and NMR data of marine sponge-associated bacterium Actinokineospora sp. crude extracts, which were cultivated from the Red Sea sponge Spheciospongia vagabunda [60]. The differential analysis of sample populations was accomplished using the MZmine software; the MS and NMR records from the databases AntiBase and MarinLit were used to identify the known secondary metabolites. With this dereplication workflow, two new antiparasitic O-glycosylated angucyclines, actinosporins A and B, were identified.

Roullier et al. [65] highlighted the potential of marine-derived fungi for new bioactive metabolites and their under-investigated halogenated metabolome and focused on the detection of new halogenated compounds among a collection of marine-derived fungal strains. A new software tool, MeHaloCoA, was developed under R to automate the identification of halogenated compounds in HPLC-MS profiles and was demonstrated with the identification and isolation of two new MNPs from a Penicillium canescens strain, chlorogriseofulvine and griseophenone I, which exhibited antiproliferative activities.

The Global Natural Products Social Molecular Networking (GNPS; http://gnps.ucsd.edu) is an open-access knowledge base for community-wide organization and sharing of raw, processed, or identified tandem mass (MS/MS) spectrometry data. It provides access to spectral libraries, dereplication tools, and visualization of molecular networks based on spectral correlation [66]. Examples of a GNPS application with MNPs include the analysis of 146 marine Salinispora and Streptomyces strains [67] and the chemical profiling of the Alphaproteobacterium strain MOLA1416 associated with the marine lichen Lichina pygmaea [68].

Although the application of hyphenated analytical and statistical methods in metabolomics facilitates the discovery of potentially novel secondary metabolites from plant, animal, and microbial origin, there are still several challenges that have to be addressed in order to achieve a real leap forward in drug discovery from natural sources. For example, comprehensive MS and NMR databases are not available for small molecules; thus, compound deconvolution and identification often require extensive searching of individual databases. Most databases do not contain MS fragmentation spectra and two-dimensional (2D) NMR spectra, which are crucial for small molecule structure elucidation and unambiguous dereplication. Moreover, the NP drug discovery process can only be exponentially improved, in our opinion, with the inclusion of predicted spectral data using computational methods. In order to amplify the spectral data space, predicted spectra can be generated for known chemical spaces and for unknown chemical spaces exponentially amplified by automatic molecular structure generators [69].

The following examples illustrate these points. The PubChem database currently contains about 90 million compounds, while the two largest (commercial) MS spectral libraries, from the National Institute of Standards and Technology (version 17) and Wiley Registry (11th edition), enclose MS data

for 267,000 and 741,000 compounds, respectively; the largest (commercial) NMR spectral database, ACD/NMR DB, comprises NMR data for ~322,000 compounds. Kerber et al. [69] reported that among more than 109 million possible molecular structures with the formula $C_8H_6N_2O$ (mass 146 Da), only 1911 hits matched in PubChem database.

Several strategies have been devised to explore this huge searchable chemical space. For example, Jeffryes et al. [70] used the Biochemical Network Integrated Computational Explorer (BNICE) and expert-curated reaction rules based on the Enzyme Commission classification system to compute Metabolic in silico Network Expansions (MINEs). This is an extension of databases with known metabolites to include molecules that have not yet been observed, but are likely to occur based on known metabolites and biochemical reactions. These databases are freely available from http://minedatabase.mcs.anl.gov. Recently, Lai et al. [71] reported the integration of a metabolome database, BinBase (a large GC-MS-based untargeted metabolomics database covering various species, organs, and matrices), with the mass spectrometry chemoinformatics tools BinVestigate (http://binvestigate.fiehnlab.ucdavis.edu), MS-DIAL 2.0, and MS-FINDER 2.0 (http://prime.psc.riken.jp). The goal is to annotate unknown metabolites modified by enzymatic transformations that gain physiological functions in a given biological system (epimetabolites) [71]. This methodology revealed that N-methyl-uridine monophosphate was highly upregulated in cancer cells and cancer tissues compared with its levels in any other cell type or tissue [71]. Another example is LipidBlast (http://fiehnlab.ucdavis.edu/projects/LipidBlast), a simulated mass spectral library for 119,200 compounds automatically generated from typical structural motifs of lipids [72].

The combination of molecular structure generators and spectra prediction methods for augmented spectral data spaces has been very successful in proteomics for many years as the prediction of peptide fragmentation patterns is easier. Hufsky and Böcker [73] reviewed the literature and identified four main approaches to mine a database of metabolite structures beyond a straightforward comparison of experimental spectra: (1) rule-based fragmentation spectrum prediction; (2) combinatorial fragmentation; (3) competitive fragmentation modelling; and (4) molecular fingerprint prediction. Rules for fragmentation prediction can be automatically learned from experimental data using machine learning (ML) techniques [74,75]. Kangas et al. [74] reported an algorithm, the so-called "in silico identification software (ISIS)", which generates in silico spectra of lipids for the purpose of structural identification. This method uses artificial neural networks (ANN) to find accurate bond cleavage rates in a mass spectrometer employing collision-induced dissociation tandem mass spectrometry. Searching a database of 18,399 calculated spectra against the experimental spectra of 45 test lipids yielded the correct structure at the top position in 40 cases and at the second position in 5 cases.

In contrast to rule-based fragmentation, combinatorial fragmentation does not aim at predicting a mass spectrum but rather at explaining the peaks in the experimental fragmentation spectrum of a metabolite by matching against possible fragments enumerated with systematic bond dissociation, i.e., mapping fragmentation spectra to molecular structures. MetFrag [76] and MetFusion [77] are the most used tools for combinatorial fragmentation. MetFusion combines MetFrag results with a spectral library search in the MassBank database. More recently, the Metabolite Identification via Database Searching (MIDAS) algorithm was reported [78]. Similarly to MetFrag, MIDAS exhaustively enumerates possible fragments, but then calculates the plausibility of the fragments based on their fragmentation pathways, instead of bond dissociation energies, to evaluate a metabolite-spectrum match (MSM); the MSM score is calculated to reflect how well the metabolite explains the spectrum. MIDAS was designed to search high-resolution tandem mass spectra against a large metabolite database in an automated and high-throughput manner. It was tested with four standard ESI-MS/MS data sets from MassBank and revealed high accuracy in the identification of metabolites against the MetaCyc database, even outperforming MetFrag [78]. It was also demonstrated using a real-world LC-ESI-MS/MS measurement of a metabolome from Synechococcus sp. PCC 7002, a marine cyanobacterium: many metabolites previously found using spectral library searching, chemical formula matching, and manual interpretation were identified, but MIDAS additionally identified many other metabolites missed in

the previous study. In a further development, Ridder et al. reported a substructure-based annotation of high-resolution multistage MS^n spectral trees (MAGMa), which uses the hierarchical information available from this technique to explain the fragment peaks observed at consecutive levels of the MS^n spectral tree [79,80]. The MAGMa+ software is available that combines MIDAS and MAGMa and uses metabolite-dependent optimized parameters obtained with ML techniques [81].

The competitive fragmentation modelling (CFM) approach [75] predicts mass spectra using a probabilistic generative model for the MS/MS fragmentation process and an ML approach for learning model parameters from experimental data. The fragmentation process is modelled as a stochastic homogeneous Markov process. This model estimates the likelihood of any given fragmentation event and predicts those peaks that are most likely to be observed, thus improving precision. It was shown that CFM can be used to predict the MS/MS spectrum from a chemical structure and to rank possible structures for an observed spectrum.

The FingerID method of Heinonen et al. [82] uses an ML approach to predict structural properties (fingerprints) of unknown molecules from their MS spectra rather than predicting fragmentation MS spectra from chemical structures. Then, the predicted fingerprints can be used to search for the unknown molecule in a chemical structure database. In the training phase, each spectrum of the training set is transformed into a feature vector. For each structural property of the fingerprint, feature vectors are marked as possessing it or not. A support vector machine (SVM) ML technique is trained to predict which structural features of the fingerprint are present in a compound from its spectra. In a related work, Shen et al. [83] reported a kernel-based ML method to predict molecular fingerprints from MS data and fragmentation trees [84]. Fragmentation trees can be considered as an annotated representation of the original fragmentation mass spectrum. Experiments on two large reference datasets, METLIN and MassBank, have shown that the inclusion of fragmentation tree kernels significantly increases the molecular fingerprint prediction accuracy [83]. A further improvement was achieved by combining more kernels, more fingerprints, and a refined fingerprint similarity scoring (CSI:FingerID) [85].

3.1.2. Genome-Guided

The fast development of genome sequencing methods and the exponentially rising number of genome sequences available revolutionized almost every aspect of biology, including NP research. In spite of the large diversity of secondary metabolites, the structures of the involved enzymes are much conserved, making it possible to mine genomes for genes encoding biosynthetic enzymes [86]. The key feature of the renaissance of NP drug discovery would be to turn the ad-hoc process of discovering NPs into a high-throughput pipeline yielding many thousands of new small molecules from microbes [87]. However, more than 10 years after the first Streptomyces genomes were sequenced [88,89], this promise has not yet been realized. Indeed, over the last decade not more than a few hundred molecules have been discovered using genome mining, and many of those molecules were so challenging to discover that the process would be difficult to generalize and automate.

Ziemert et al. reviewed the evolution of genome mining in microbes and included an extensive list of examples where genome mining has directly led to the identification of metabolites [86]. For example, the discovery of the polyene macrolactam salinilactam A (**3**) (Figure 6) demonstrates the powerful interplay between genomic analysis and traditional studies of NP chemistry. The salinilactam gene cluster is the biggest gene cluster detected by bioinformatic analysis in the marine actinomycete Salinispora tropica CNB-440 genome [90]. The detection of the compound was possible based on putative structural features (characteristic UV chromophores) suggested by the initial inspection of the partial gene cluster. Then, the structural fragments and the molecular formula obtained by MS for an isolated product suggested a 10-module polyketide synthase (PKS) enzyme responsible for the biosynthesis of the compound, which facilitated assembly and therefore closure of the genome. Finally, further bioinformatic analysis of enzymatic domains refined the structure elucidation of the compound. A similar strategy was followed by Schulze et al. [91] and enabled the discovery of a

family of macrolactams from a marine actinomycete, Micromonospora sp. (lobosamide A (**4**), B (**5**), and C (**6**) are illustrated in Figure 6), as well as mirilactam A (**7**) and B (**8**) from a distantly related actinobacterium, Actinosynnema mirum. A genome mining study reported the identification of 31 cyanobactin gene clusters from 126 genomes of the marine cyanobacteria Microcystis aeruginosa PCC 9432 and Oscillatoria nigro-viridis PCC 7112 [92]. Cyanobactins are a growing family of cyclic ribosomal peptides produced by cyanobacteria, which have exhibited cytotoxic activity against cancer cell lines as well as antiviral, antimalarial, and allelopathic activities. Bioinformatic analysis of the genomes predicted that the strains produce cyanobactins with chain lengths of 3, 4, and 5 amino acids and containing thiazoles (the core encoded a cysteine and the gene cluster encoded heterocyclase and oxidase enzymes). Extensive chemical analyses demonstrated that some cyanobacteria produce short linear peptides with a chain length ranging from three to five amino acids. Three novel linear peptides, aeruginosamide B (**9**) and C (**10**) and viridisamide A (**11**) (Figure 6), were isolated, which were N-prenylated and O-methylated on the N and C termini, respectively.
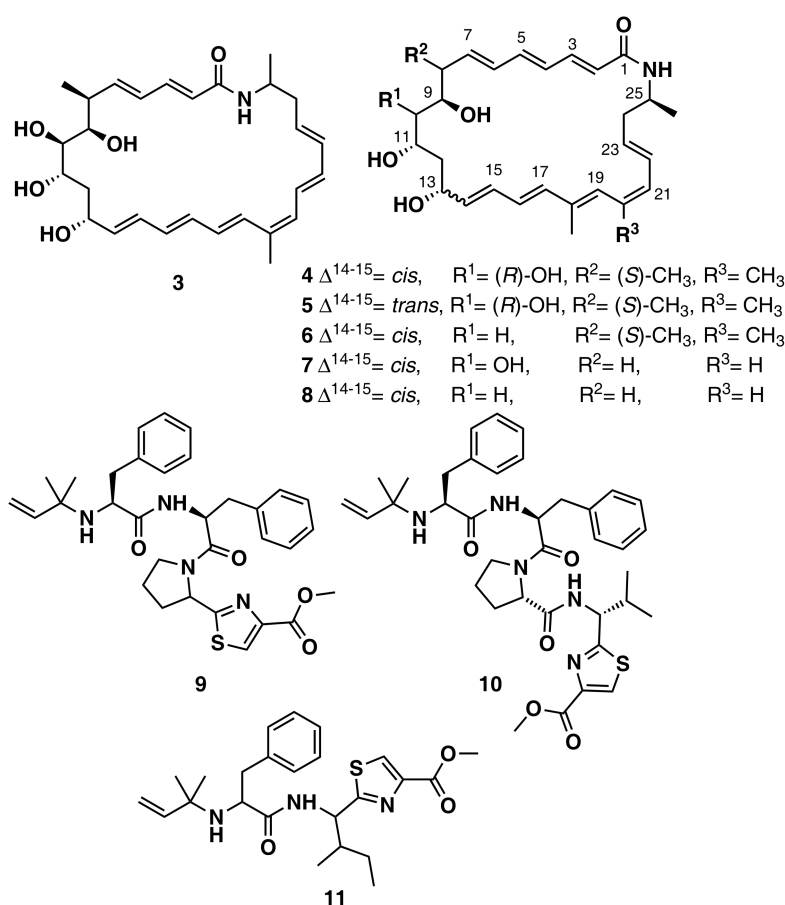

Figure 6. Chemical structures of MNPs identified using genome mining approaches.

Of particular relevance for the computational identification of genes encoding metabolic pathways is the fact that they are typically chromosomally adjacent, forming biosynthetic gene clusters (BGCs). These BGCs encode all the biosynthetic machinery to produce, process, and export a specialized metabolite (enzymes, regulatory proteins, and transporters) [87]. They are useful targets for mining genomes (to discover new metabolites) based on knowledge of homologous genes and rules/patterns extracted from them. Plenty of computational tools are available for researchers to mine genetic data and to connect them to known secondary metabolites. An overview of computational tools for genome mining is displayed in Figure 7. Reviews include references [86] and [87]. The Secondary Metabolite Bioinformatics Portal (SMBP) website at http://www.secondarymetabolites.org maintains

a catalogue of available software, databases, and hand-curated links to major resources used in the field [93]. We are currently far from the initial simple comparison techniques using manually constructed lists of genes as query sequences, such as the sequence-based comparison with BLAST [94] or profile-based tools, such as HMMer [95]. Nowadays, comprehensive software resources are available and typically classified into two categories: low-novelty methods using profiles of known and highly conserved biosynthetic machineries (e.g., polyketide synthases or non-ribosomal peptide synthetases domains) and high-novelty methods detecting new classes of gene clusters (Figure 7). Examples of software implementing low-novelty methods are ClustScan [96], SMURF [97], and antiSMASH [98]. The most comprehensive tool, antiSMASH, can detect more than 20 classes of pathways. High-novelty methods include pattern-based mining, phylogeny-based mining, comparative genomic alignment, resistance-based mining, and regulation-based mining.
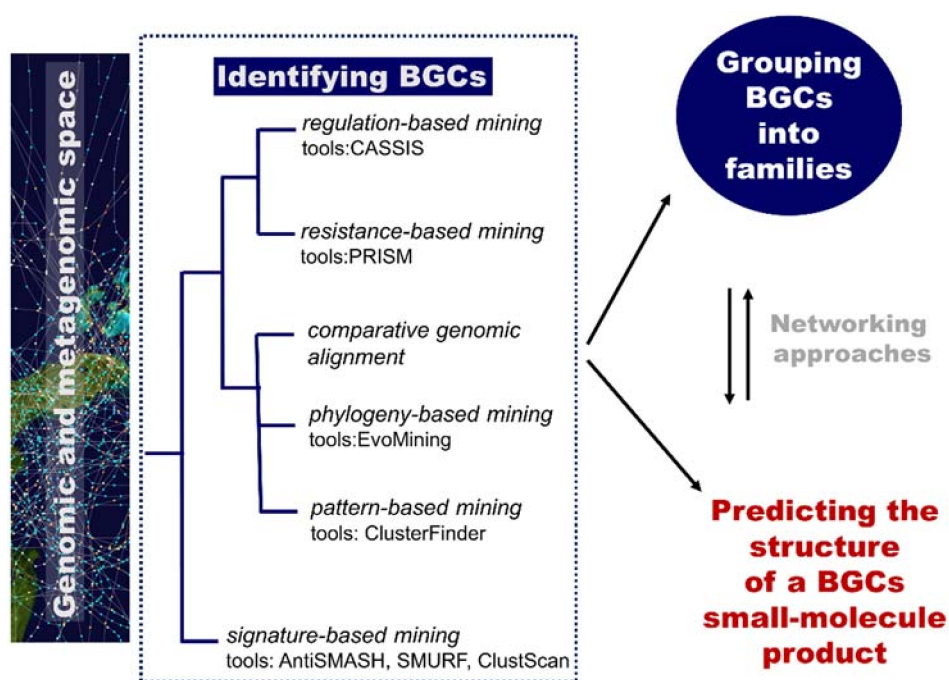

Figure 7. The role of computational methodologies in genome mining for natural product discovery. BGC, biosynthetic gene cluster.

The ClusterFinder software implements a pattern-based mining strategy (based on a hidden Markov model-based probabilistic algorithm) and aims to identify gene clusters of both known and unknown classes [99]. Instead of looking for specific individual signature genes, ClusterFinder recognizes patterns of broad gene functions encoded in a genomic region. In a study of secondary metabolites of proteobacteria, ClusterFinder enabled the identification of a large, previously unrecognized family of gene clusters that encode the biosynthesis of aryl polyenes [99].

Phylogeny-based mining incorporates evolutionary principles into gene mining: enzymes evolve in their substrate specificity and acquire new metabolic functions keeping detectable relationships with ancestral primary metabolic enzymes [100]. Cruz-Morales et al. [100] reported the use of EvoMining, a phylogeny-based mining approach, to discover a biosynthetic pathway for arseno-organic metabolites in Streptomyces coelicolor and Streptomyces lividans. The EvoMining method was implemented in a standalone tool distributed as a docker image developed by the EvoDivMet lab and has been made available at https://github.com/nselem/EvoMining.

Takeda et al. reported a comparative genomic alignment methodology based on the assumption that secondary metabolism genes are highly enriched in nonsyntenic blocks; a biosynthetic gene cluster can be detected by searching for a similar order of genes and their presence in nonsyntenic blocks.

This approach enabled the detection of biosynthetic gene clusters without core genes, e.g., the kojic acid biosynthesis gene cluster of Aspergillus oryzae [101].

PRISM (PRediction Informatics for Secondary Metabolomes) is an open-source web application for the genomic prediction and dereplication of nonribosomal peptide and type I and II polyketide chemical structures [102]. This software is based on hidden Markov models that can predict not only genes involved in NP biosynthesis but also in antibiotic resistance. Genes encoding resistance functions can lead to the identification of enzymes for the biosynthesis of new antibiotics [86] as bacteria producing antibiotics may have their own resistance mechanisms to avoid self-destruction. Such a resistance-based approach is illustrated with the work of Moore and co-workers [103] that screened the genomes of 86 marine Salinispora bacterial genomes and prioritized an orphan polyketide synthase–nonribosomal peptide synthetase hybrid BGC (tlm) with a putative fatty acid synthase resistance gene. The expression of the tlm and the related ttm BGCs in Streptomyces hosts led to the production of unusual thiotetronic acid antibiotics.

Finally, CASSIS is an example of a regulation-based mining tool that exploits the idea of co-regulation of the cluster genes and assumes the existence of common regulatory patterns in the cluster promoters; the method searches for "islands" of enriched cluster-specific motifs in the vicinity of anchor genes [104]. This strategy can be particularly useful in fungi as genes of the same BGC are highly co-regulated [86].

3.2. Computer-Assisted Structure Elucidation (CASE)

Fully automated structure elucidation from spectroscopy data has been achieved for small organic molecules, from 1D NMR data, or for complex NPs using 2D NMR data. CASE expert systems have been developed for over 40 years. Currently available packages include the open source Seneca platform [105,106], the commercial ACD/Structure Elucidator Suite [107–109], LSD [110], and CMC-se (http://www.bruker.com). Here, we review some recent achievements of CASE expert systems.

Troche-Pesqueira et al. reported enhanced CASE procedures for the determination of the relative configuration of NPs, which starts from the molecular formula and combines conventional one-dimensional (1D) and 2D NMR spectra with residual dipolar couplings (RDCs) and/or residual chemical shift anisotropy (RCSA) [111,112]. The employment of RDC data in conjunction with a CASE program automated the determination of relative configurations in molecules of medium complexity and a moderate degree of flexibility, such as naltrexone, 10-epi-8-deoxycumambrin, strychnine, eburnamorine, yohimbine, and N-methylcodeine. The pool of diastereoisomeric candidates was enumerated and the conformational space was explored for flexible molecules in the process of identifying the structure that best agrees with the RDC data. Moreover, the authors demonstrated that the assignment of absolute configurations can also be incorporated by comparison of experimental and density functional theory (DFT)-calculated vibrational or electronic circular dichroism (VCD or ECD) curves [111].

Liu et al. [112] proposed a protocol comprising the confluence of capabilities embodied by CASE methods, DFT calculations, and measurement of anisotropic NMR parameters (RDCs and RCSA) aiming at the growing general problem of structural mischaracterization. The authors demonstrated that the combination of RDCs and RCSAs provides a powerful orthogonal mean of confirming not only the relative configuration of a given stereocenter, but also the overall molecular structure and atomic connectivity of a molecule [112]. The protocol was applied to several examples of revised structures, including aquatolide, a sesquiterpene lactone isolated from the hexane extract of Asteriscus aquaticus. In 1989, a very rare ladderane moiety was proposed [113] for the aquatolide (12) (Figure 8). However, more recently, the proposed chemical structure of the aquatolide (**12**) was revised on the basis of quantum-chemical calculations and NMR experiments to the unusual core structure (**13**) (Figure 8) [114]. The revised structure of aquatolide was subsequently confirmed by X-ray crystallography [114] and by total synthesis [115]. Liu et al. compared the experimental

and back-calculated RDC/RCSA data for the model structures (**12**) and (**13**) (Figure 8) and readily established that the revised structure (**13**) is in best agreement with the data [112].
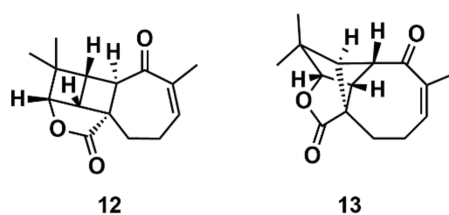

Figure 8. Proposed structure of aquatolide (**12**) and the corresponding revised structure (**13**).

Synergistic combinations of CASE algorithms and DFT calculations of chemical shifts have been reported that broaden the range of amenable structural problems to encompass proton-deficient molecules, molecules with heavy elements (e.g., halogens), conformationally flexible molecules, and configurational isomers [116–118]. Buevich and Elyashberg [118] illustrated this approach with previously established structures; one example is cycloshermilamine D (**14**) (Figure 9), a pyridoacridine alkaloid isolated from the marine tunicate Cystodytes violatinctus [119]. The ACD/Structure Elucidator system processed the experimental data, consisting of the molecular formula, 1D proton and carbon spectra, and 2D NMR data (COSY, HSQC, and HMBC), and yielded 263 candidate structures. The four top candidates included the structure of cycloshermilamine D at the first position, but the other three candidates had very similar sets of carbon chemical shift deviations. DFT calculations of carbon chemical shifts for the four structures were performed at the mPW1PW91/6-311 + G(2d,p) level of the theory, unequivocally showing that the first structure had the lowest root mean square deviation (RMSD) (^{13}C) and the smallest maximum chemical shift deviation, which convincingly supported the structure of cycloshermilamine D without any additional experimental data.
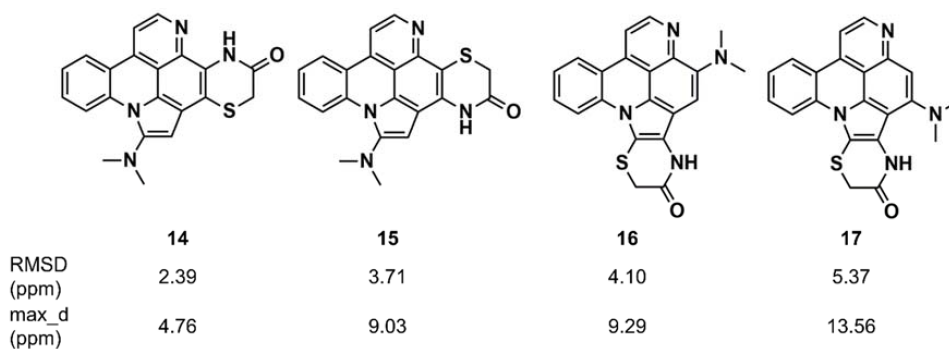

	14	**15**	**16**	**17**
RMSD (ppm)	2.39	3.71	4.10	5.37
max_d (ppm)	4.76	9.03	9.29	13.56

Figure 9. Root mean square deviation (RMSD) and maximum chemical shift deviation between experimental and density functional theory (DFT)-calculated carbon chemical shifts [118] for four isomers of cycloshermilamine D suggested by Computer-Assisted Structure Elucidation (CASE) analysis.

4. Computer-Aided Drug Design (CADD)

Computer prediction of biological activities of MNPs is required to guide decisions concerning the in vivo and in vitro testing of isolated NPs and extracts, to assist in the design of bioactive NP derivatives, and to virtually screen databases of known or proposed NPs. Additionally, the regions of the chemical space encompassing NPs are recognized as promising for the invention of new drug leads as they result from the evolution of chemical structures during millions of years for optimum performance of biochemical machineries [120]. Furthermore, advances have been reported on computational methodologies to explore global networks connecting active compounds and their

targets [121–124], to simulate interactions between ligands and binding sites [125–132], and to establish structure-activity relationships with NPs and MNPs [133–137]. Available ADMET predictors for several endpoints, e.g., human intestinal absorption, Caco2 (heterogeneous human epithelial colorectal adenocarcinoma), cell permeability, or blood brain barrier permeability, are often applied in screening procedures to filter out molecules with undesirable properties [132,134,138].

4.1. Ligand-Based (LB)

Ligand-based methodologies are useful to discover new lead compounds when sets of active molecules are known for specific targets. Developed strategies include similarity searches in databases of molecules, structure alignment for the identification of pharmacophores and virtual screening, and ML algorithms to establish Quantitative Structure-Activity Relationships (QSARs), predict properties of candidates, and guide the design of new molecules.

Dineshkumar et al. [139] performed target prediction for sporolides A and B using LB pharmacophore screening against known inhibitors and drugs. These NPs are polycyclic macrolides from the obligate marine actinomycete Salinispora tropica. Eight pharmacophore features were identified in sporolides A and B: six H-bond acceptors, one hydrophobic group, and one aromatic ring [139]. The three-dimensional (3D) models were generated and the pharmacophore pattern was used to screen the public Binding Database with 400,000 known ligands. A small group of targets was retrieved bearing similar pharmacophore features, and these were further explored with structure-based methods. HIV-1 reverse transcriptase chain A emerged as a predicted target. In vitro testing showed that sporolide B significantly reduced the activity of HIV-1 RT and could be a possible drug candidate for HIV and other retroviral viruses [139]. The same lab later reported a similar computational study for the MNPs salinosporamides A, B, and C from the same source and concluded that the glucocorticoid receptor and methionine aminopeptidase 2 could be new drug targets, suggesting possible antiinflammatory and anticancer activities of salinosporamides [140].

Waldmann and co-workers [141] suggested, from a statistical analysis of the structural classification of NPs, that more than half of all NPs have just the right size (i.e., a van der Waals volume between 300 and 800 Å^3) to serve as a starting point for hit and lead discovery. Indeed, Pereira et al. [142] have also observed, in a subset of PubChem and AntiMarin, a correlation between active compounds and three- or four-ring compounds with a van der Waals volume between 300 and 800 Å^3. Ertl et al. [120] developed a NP-likeness score to measure the similarity between a molecule and the structural space covered by NPs. A NP-likeness score was incorporated in SENECA, an open-source CASE platform, significantly improving the ranking of candidates in structure elucidation of metabolites [106]. Similar approaches can be used in virtual screening, in prioritization of compound libraries toward NP-likeness, and in the design of building blocks for the synthesis of NP-like libraries [120]. More recently, Shang et al. [143] analysed the differences between terrestrial and marine NPs using chemoinformatics methods on a data set with 32,937 MNPs and 132,071 terrestrial NPs. The authors observed a trend for MNPs to have lower solubility, longer chains and larger rings, more halogens (especially bromine), and nitrogen. MNP scaffolds are less represented in databases of known ligands, which agrees with the fact that MNPs have been less exploited in drug discovery projects and suggests their greater potential in developing new drugs.

Reymond and co-workers [144] enumerated possible organic saturated or aromatic ring systems with up to 4 cycles and 14 atoms to obtain the so-called GDB4c database containing 916,130 ring systems. This was further processed to generate all possible stereoisomers, yielding a GDB4c3D database with 6,555,929 compounds. Almost all of these ring systems are unknown and represent chiral 3D macrocycle structures; included are many polycyclic scaffolds reminiscent of NPs. The database is a useful resource for similarity and pharmacophore searching on the basis of known NPs. It is available for download at www.gdb.unibe.ch together with interactive tools for data mining. The authors illustrated the platform by searching for similar structures of the NPs hasubanonine (**18**) and vincadine

(**19**) (Figure 10). The results enabled the identification of similar 3D structures with new ring systems and led to the proposal of the six new analogs **20–22** and **23–25**.
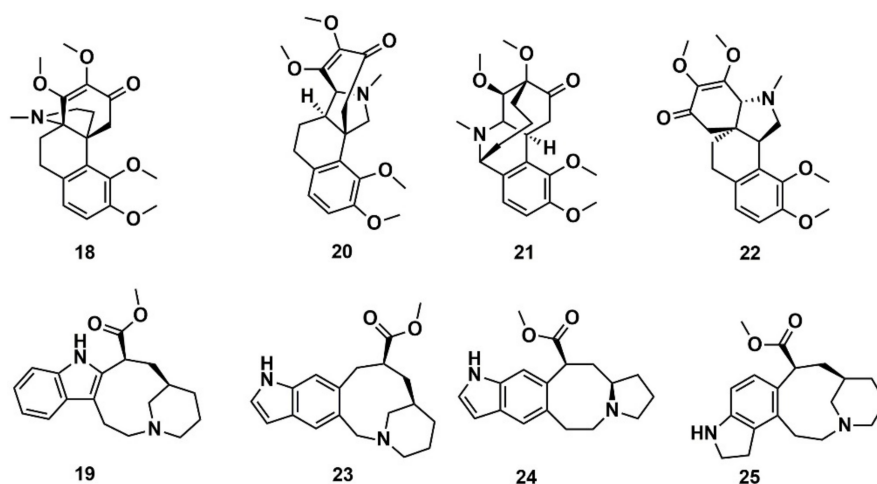

Figure 10. Chemical structure of hasubanonine (**18**) and vincadine (**19**) as well as their designed analogs, **20–22** and **23–25**, respectively.

NPs often contain macrocycles, which are problematic structures for CADD due to their size (generally >500 MW) and conformational complexity. Low-energy conformations must be identified to model conformation-dependent properties. Macrocyclic polyketides are medically and biologically important NPs characterized by structural and functional diversity [145]. Wang et al. proposed an improved dihedral angle-based macrocycle conformational sampling method and evaluated its performance with a data set of 37 polyketides with 9−22 rotatable bonds in the macrocyclic ring for which crystal structures were available [145]. The protocol was able to reproduce the crystal structure of polyketides' aglycone backbone within an RMSD of 0.50 Å for 31 out of 37 polyketides [145].

Drug interaction with multiple targets is a cause of drug side effects [146], but it can also be used to increase drug efficacy [147], repurposing [121,148], and design multitarget molecules [149]. Systematic experimental identification of drug targets for NPs or known drugs at the human proteome level is not feasible for the thousands of compounds currently available. Therefore, the development of computational tools to predict the targets of new or known molecules in a systematic way is of high interest [121]. It has been claimed [150] that ML models can point to potential target families and sometimes even to the target subtypes of approximately one-third of the NPs identified to date. Schneider et al. [150] computationally identified and biochemically confirmed an unknown, high-affinity macromolecular target of doliculide (**26**) (Figure 11), an MNP that is produced by the sea hare Dolabella auricularia. The authors performed automated target prediction with the SPiDER protocol for both doliculide, an NP with strong actin-polymerizing and anticancer activities, and 134 intermediates and precursors of a total synthesis. The SPiDER protocol performs a projection of query compounds, represented by pharmocophore topological descriptors, onto a self-organizing map (SOM) consisting of 120 receptive fields, which was previously trained with pharmacologically active reference compounds and their known targets [149]. The prostaglandin receptors (e.g., EP2, EFP3, and EP4) were predicted as targets not only for doliculide itself but also for most of the synthesis intermediates (100 out of the 134). Doliculide represented a novel chemotype among G-protein-coupled receptor ligands. A flexible three-dimensional pharmacophore alignment was also performed between doliculide (**26**) and three well-studied, non-selective prostanoid agonists (**27–29**) (Figure 11). The alignment revealed that the four compounds contain a total of five common pharmacophore points.
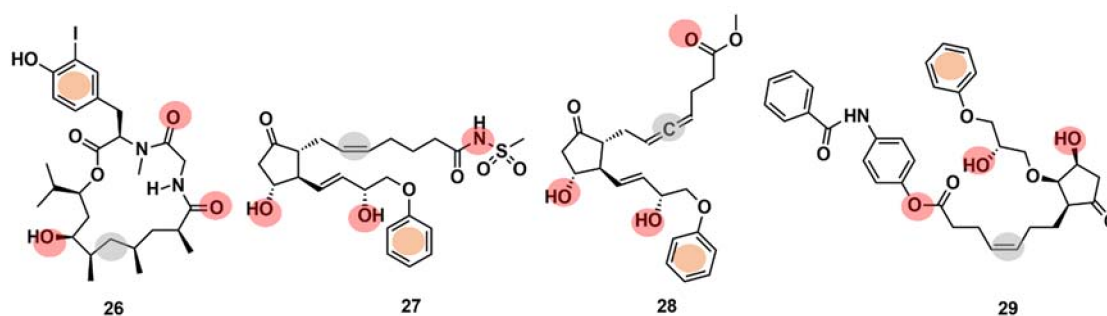
Figure 11. Chemical structures of doliculide (**26**) and three known prostanoid receptor ligands, sulprostone (**27**), enprostil (**28**), and GR63,799 (**29**). The pharmacophore features are indicated in the chemical structures by colored dots: red, hydrogen-bond donors; grey, lipophilic interaction centers; and orange, aromatic centers.

Network-based approaches have also been used for the systematic identification of drug−target interactions (DTIs) and assessment of drug safety profiles [121]. Fang et al. [121] proposed a statistical network model to predict new drug targets and anticancer indications of NPs. A global drug−target network was reconstructed that linked molecules, substructures, and targets and resulted in 7314 interactions connecting 751 targets and 2388 NPs. New interactions are predicted from the substructures of query compounds. The authors computationally identified multiple anticancer indications for several typical NPs with a new mechanism of action (MOA) across 13 cancer types. For example, naringenin (a flavanone mainly found in grapefruit, oranges, and tomatoes), disulfiram (an FDA-approved carbamate derivative for the treatment of chronic alcoholism), and metformin (a biguanide oral agent for treating type 2 diabetes) showed six (bladder, lung, uterine, colon, prostate, and breast), five (breast, colon, lung, thyroid, and uterine), and two (breast and ovarian) new MOAs, respectively [121].

Linear regressions and ML algorithms are well-known to establish QSARs, which are trained with available experimental data and molecular descriptors encoding structural features to make predictions for new molecules. Here, we describe recent examples of QSAR models used to estimate biological activities and ADMET properties of MNP. Davis and Vasanthi [134] retrieved 157 compounds from the Seaweed Metabolite Database of marine algal secondary metabolites (http://www.swmd.co.in) and developed a QSAR approach concerning anticancer activity against six different cancer cell lines: MCF-7 (human breast adenocarcinoma), A431 (human epithelial carcinoma), HeLa (human cervical adenocarcinoma), HT-29 (human colon adenocarcinoma grade II), P388 (murine leukemia), and A549 (human lung epithelial adenocarcinoma). The QSAR process was used to identify relevant structural features and to support the choice of protein kinase B (PKB) targets for further structure-based studies. ADMET predictions were later used to select a lead compound. A QSAR approach was also pursued by Knight et al. [135] using 43 synthetic derivatives of the marine alkaloid tambjamine to model transmembrane anion transport activity. The data set comprised bipyrrole core derivatives with three substitution patterns. A parabolic dependence of the anionophoric activity was observed with lipophilicity, which was quantified in two-, three-, and four-parameter linear model equations.

The quest for new antimalarial drugs has also led to the investigation of MNPs with QSAR methods [136,137]. Aswathy et al. [136] analyzed 42 analogs of the natural product thiaplakortone-A, which was found in the Australian marine sponge Plakortis lita and is active against chloroquine-sensitive and chloroquine-resistant Plasmodium falciparum. Several QSAR models, including both 2D and 3D QSAR, were developed, and the results were combined with simulated interactions with the P. falciparum calcium-dependent protein kinase 1 protein to design and screen new virtual molecules. Three new molecules were proposed as leads to potential anti-malarial drugs. In a different approach, quantitative relationships were established between thermodynamics/electronic properties calculated by DFT methods and antimalarial activity [137].

Linear regressions were performed with a data set of 14 sponge metabolites–bromopyrrole alkaloids. The best model (r^2 = 0.97, Q^2 = 0.86, F = 41.85) was obtained using the molecular descriptors entropy, dipole moment, molecular polarizability, energy of the highest occupied molecular orbital (HOMO), softness, and electrophilicity index [137]. The HOMO also performed remarkably well in discriminating overall biological activity of MNP and microbial NPs [151].

4.2. Structure-Based (SB)

Molecular docking has been the major SB methodology to predict affinities to macromolecular targets, to interpret binding modes, and to assist in the design of drug leads. Several recent publications illustrate the application of the method to MNPs [127–129,138,152–171], and some representative examples are here described.

Liu et al. [129] designed, synthesized, and evaluated 19 new derivatives of the MNP tasiamide B (**30**) (Figure 12) as inhibitors of BACE1, a potential therapeutic target for Alzheimer's disease. Tasiamide B is an acyclic peptide containing a statine-like unit and several aminoacid residues. The exploration of structure–activity relationship (SAR) with truncated derivatives identified a core structure as well as a free carboxylic acid group important for inhibitory activity. The conclusions were supported by a docking simulation.
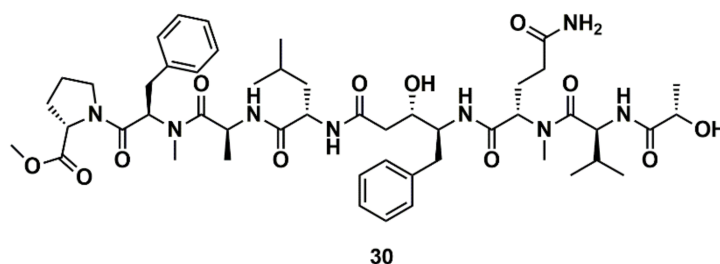

Figure 12. Chemical structures of tasiamide B (**30**).

SB computational studies and in vitro experimentation were combined to elucidate the molecular target of 13 low molecular weight MNPs from marine sponges and ascidians. Some are bioactive and the structural similarity to diverse cholinergic ligands anticipated their possible activity towards nicotinic acetylcholine receptors (nAChRs) [127]. In silico docking to the Lymnaea stagnalis acetylcholine-binding protein (AChBP), a model for the ligand-binding domains of nAChRs, was carried out. High affinity was predicted for some compounds, such as the polysulfide varacin (**31**) and the seven alkaloids pibocin (**32**), makaluvamines C and G (**33**, **34**), debromohymenialdesine (**35**), crambescidin 359 (**36**), aaptamine (**37**), and monanchocidin (**38**), while low efficiency of interaction was suggested for other compounds, such as the two sphingolipids rhizochalin (**39**) and its aglycone (**40**) as well as the three alkaloids 1,1′-dimethyl-[2,2′]-bipyridyldiium salt (**41**), 7,8-dihydroimidazo-[1,5-*c*]-pyrimidin-5(6*H*)-one (**42**), and 1,3-dimethylisoguaniniium hydrochloride (**43**) (Figure 13). The conclusions from computer modelling were verified by radioligand analysis. Nicotinic acetylcholine receptors exhibit multiple conformational states: resting (channel closed), active (channel open), and desensitized (channel closed). Homology modelling was used by Mallipeddi et al. [172] to generate structures of the Torpedo californica $\alpha_2\beta\delta\gamma$ nAChR that initially represent the resting state and the desensitized state. Molecular dynamics (MD) simulations were performed on the extracellular ligand binding domain on each nAChR conformational state with and without the agonist anabaseine present in each binding site. Anabaseine (a bipyridine derivative) is a marine alkaloid toxin that acts as an agonist on most nAChRs in the central nervous system. The MD simulations revealed that in the presence of agonist, loop C was drawn inward and attained a more stable conformation [172].
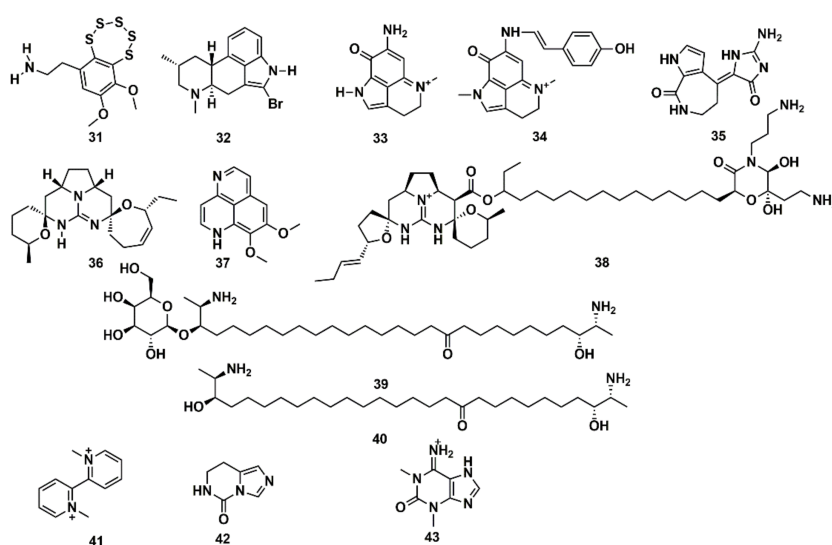
Figure 13. Chemical structures of MNPs predicted to have high (**31–38**) and low (**39–43**) affinity to *Lymnaea stagnalis* AChBP using three docking approaches.

Protein kinases and acetylcholinesterase (AChE) are potential targets for the treatment of Alzheimer's disease (AD). Llorach-Pares et al. reported a molecular docking investigation of meridianins A–G (a group of indole alkaloids isolated from the marine tunicate Aplidium) towards protein kinases in order to assist in the future development of anti-AD drugs [138]. Post-processing of docking results was performed with MD simulations. The results provided information concerning binding mode, strength, and selectivity and were complemented with ML predictions of ADMET properties. Botic et al. described four brominated pyrroloiminoquinone alkaloids (discorhabins) isolated from Latrunculia sp. sponges collected near the Antarctic Peninsula and their promising activity as reversible competitive inhibitors of cholinesterases. Docking calculations with different AChEs revealed the involved interactions in the active sites and provided further support for the experimental data [152].

Wang et al. [165] studied the antibacterial activity of a novel anthraquinone, 2-(dimethoxymethyl)-1-hydroxyanthracene-9,10-dione, together with nine known anthraquinone derivatives isolated from the marine-derived fungus Aspergillus versicolor. The novel molecule showed strong inhibitory activities against MRSA ATCC 43300, and MRSA CGMCC 1.12409 (with MIC values of 3.9 and 7.8 µg/mL, respectively). Molecular docking studies predicted that the new anthraquinone binds to the AmpC β-lactamase and topoisomerase IV enzymes, which could explain its antimicrobial properties. It bound to DNA topoisomerase IV receptor similarly to a co-crystallized ligand and with lower binding energy. The same was observed in the β-lactamase binding site.

Chen et al. [126] reported the synthesis of a series of novel 1,2-dithiolan-4-yl benzoate derivatives inspired by bruguiesulfurol, a marine cyclic disulphide, and their in vitro inhibitory activity against the enzyme protein tyrosine phosphatase 1B (PTP1B), a validated target for the treatment of diabetes and obesity. An SAR analysis assisted by molecular docking allowed the authors to reveal the derivative with a 2,5-dibromidebenzyloxy terminal moiety as the most potent PTP1B inhibitor among all 11 derivatives (IC$_{50}$ = 0.59 µM), with improved activity compared to the original hit [126]. Inhibitors of the same enzyme were isolated from the marine brown alga Sargassum serratifolium. Three plastoquinones (sargahydroquinoic acid, sargachromenol, and sargaquinoic acid) exhibited dose-dependent inhibitory activity against PTP1B (IC$_{50}$ range of 5.14–14.15 µM). In addition, sargachromenol and sargaquinoic acid also showed dose-dependent inhibitory activity against α-glucosidase (IC$_{50}$ 42.41 and 96.17 µM, respectively). The results of docking simulations indicated a high affinity and tight binding capacity towards the active site of the PTP1B and α-glucosidase

enzymes [157]. Docking was also used by Xu et al. [158] to understand the high activity against PTP1B (IC$_{50}$ 0.84 μM) of a marine-derived bromophenol compound isolated from the red alga Rhodomela confervoides.

Twelve pyrrole alkaloid derivatives, isolated from an Australian marine sponge, Ianthella sp., were evaluated as inhibitors of ATP binding cassette (ABC) transporters, a potentially useful activity to overcome multi-drug resistance of cancer cells [128]. One of them, lamellarin O, was found to be a potent selective inhibitor of the BCRP ABC transporter. An SAR analysis covering the 12 MNPs and 6 synthetic analogues was supported by in silico docking studies and identified structural elements of the inhibitory pharmacophore, including a methoxy-acetophenone, a carboxylic ester, and two phenolic residues.

Cen-Pacheco et al. [153] applied molecular docking to understand the different activity of two novel squalene derivatives, isolated from the red seaweed Laurencia viridis, as inhibitors of Ser-Thr protein phosphatase type 2A (PP2A). This enzyme has several functions in cells and is a tumour promoter and suppressor, making it a potential target for new anticancer drugs. The two novel squalene derivatives, (+)-longilene peroxide and (+)-prelongilene, were evaluated for their ability to inhibit PP2A. While (+)-longilene peroxide is an inhibitor (IC$_{50}$ 11.3 μM ±1.4), (+)-prelongilene is inactive at a concentration of 100 μM. Docking simulations onto the PP2A enzyme-binding region revealed that, although the two compounds have similar binding modes, the first establishes several favourable contacts that are not observed with the second, and the second has unfavourable contacts with several residues. The results indicated that the additional allylic hydroperoxide group at C-2 in (+)-longilene peroxide is responsible for key hydrogen bonds and appears to be the factor leading to the differences in bioactivity [153]. Similarly, Cruz et al. rationalized the different activity against protein phosphatase 1 and 2A of two new marine brominated bis(indole) alkaloids, dragmacidins I and J, with docking into the binding pocket of PP1 [154]. Structure-based virtual screening enabled Xin et al. to discover new DNA topoisomerase I (Topo I) inhibitors, which are potential antitumor agents. A collection of 138 structures from low-cytotoxic or non-cytotoxic coral-derived fungi and plants were docked to the central catalytic domain of the Topo I–DNA complex and the 27 molecules with the most favourable predicted interactions were evaluated in vitro. Among these, four compounds showed activity at 25 μM and two compounds were active at 5 μM [155].

The ability of reverse docking for target fishing of MNPs was evaluated by Chen et al. using 40 marine compounds with known antitumor activities and known target proteins but without their crystal structure determined [159]. A database of anti-tumor proteins was constructed with 470 crystal structures corresponding to 150 different target proteins. After docking the 40 MNPs to the proteins in the database, it was observed that, although the predicted binding energy for a given ligand to its known target is usually not the lowest, 55% of the compounds have their reported target ranked in the top 20, and 30% in the top 10. It is noted that the compounds may have multiple targets and some of them may have not been discovered and reported yet [159].

In general, the LB and SB methods are complementary and were used as such in several of the works here cited [134,136,139,140]. In a comparative study of docking and similarity searches (based on 2D and 3D fingerprints), Avram et al. concluded that fusing the results obtained by the two approaches can enhance the probability to find new chemotypes in virtual screening [173]. Ebrahim and Sayed [131] reported the exploration of a MNP-based mini-library comprising 71 molecules with diverse scaffolds (e.g., macrolides, sesquiterpenes, diterpenes, sesterterpenes, triterpenes, and alkaloids). They were submitted to the Lilly's Open Innovation for Phenotypic Drug Discovery (PD2-OIDD) program for biological screening after successfully passing the initial online bioinformatics screen (https://openinnovation.lilly.com/dd/). The bioinformatics filter calculates molecular descriptors and evaluates drug-like characteristics. Among the surviving 38 MNPs and semisynthetic derivatives, several compounds showed promising results in primary and secondary angiogenesis screening modules and minimal cytotoxicity at relevant doses. According to the authors,

molecular modelling and docking experiments aided in understanding molecular binding interactions, identifying pharmacophoric epitopes, and deriving structure-activity relationships of active hits.

Finally, Skariyachan et al. applied a computational workflow to identify possible lead molecules against the Ebola virus among compounds from microbial symbionts associated with marine sponges [132]. The procedure included the calculation of drug likeness and ADMET properties followed by docking of the selected molecules against the VP40 target of Ebola virus. Lead molecules, such as gymnastatin G (a sterol derivative with anti-leukemia activity), sorbicillactone A (an alkaloid derivative with anti-leukemia and anti-HVI-1 activities), marizomib (a β-lactone-γ-lactam derivative with anti-proteasome activity), and daryamide C (a polyketide derivative with anticancer activity against the human colon carcinoma cell line), were proposed as possible inhibitors against the VP40 matrix protein of the Ebola virus [132].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DiMasi, J.A.; Grabowski, H.G.; Hansen, R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33. [PubMed]
2. Nosengo, N. New tricks for old drugs. *Nature* **2016**, *534*, 314–316. [CrossRef] [PubMed]
3. Scannell, J.W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **2012**, *11*, 191–200. [PubMed]
4. Nantasenamat, C.; Prachayasittikul, V. Maximizing computational tools for successful drug discovery. *Expert Opin. Drug Discov.* **2015**, *10*, 321–329. [CrossRef] [PubMed]
5. Gasteiger, J. Chemoinformatics: Achievements and challenges, a personal view. *Molecules* **2016**, *21*, 151. [CrossRef] [PubMed]
6. Mueller, R.; Dawson, E.S.; Meiler, J.; Rodriguez, A.L.; Chauder, B.A.; Bates, B.S.; Felts, A.S.; Lamb, J.P.; Menon, U.N.; Jadhav, S.B.; et al. Discovery of 2-(2-benzoxazoyl amino)-4-aryl-5-cyanopyrimidine as negative allosteric modulators (NAMs) of metabotropic glutamate receptor 5 (mGlu5): From an artificial neural network virtual screen to an in vivo tool compound. *ChemMedChem* **2012**, *7*, 406–414. [CrossRef] [PubMed]
7. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W., Jr. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395. [CrossRef] [PubMed]
8. Katsila, T.; Spyroulias, G.A.; Patrinos, G.P.; Matsoukas, M.-T. Computational approaches in target identification and drug discovery. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 177–184. [CrossRef] [PubMed]
9. Rodriguez, A.L.; Grier, M.D.; Jones, C.K.; Herman, E.J.; Kane, A.S.; Smith, R.L.; Williams, R.; Zhou, Y.; Marlo, J.E.; Days, E.L.; et al. Discovery of novel allosteric modulators of metabotropic glutamate receptor subtype 5 reveals chemical and functional diversity and in vivo activity in rat behavioral models of anxiolytic and antipsychotic activity. *Mol. Pharmacol.* **2010**, *78*, 1105–1123. [CrossRef] [PubMed]
10. Gaudencio, S.P.; Pereira, F. Dereplication: Racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **2015**, *32*, 779–810. [CrossRef] [PubMed]
11. Perez-Victoria, I.; Martin, J.; Reyes, F. Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta Med.* **2016**, *82*, 857–871. [CrossRef] [PubMed]
12. Patridge, E.; Gareiss, P.; Kinch, M.S.; Hoyer, D. An analysis of FDA-approved drugs: Natural products and their derivatives. *Drug Discov. Today* **2016**, *21*, 204–207. [CrossRef] [PubMed]
13. Newman, D.J.; Cragg, G.M. Drugs and drug candidates from marine sources: An assessment of the current "State of play". *Planta Med.* **2016**, *82*, 775–789. [CrossRef] [PubMed]
14. Vijayakrishnan, R. Structure-based drug design and modern medicine. *J. Postgrad. Med.* **2009**, *55*, 301–304. [CrossRef] [PubMed]
15. Talele, T.T.; Khedkar, S.A.; Rigby, A.C. Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* **2010**, *10*, 127–141. [CrossRef] [PubMed]

16. Van Drie, J.H. Computer-aided drug design: The next 20 years. *J. Comput. Aided Mol. Des.* **2007**, *21*, 591–601. [CrossRef] [PubMed]

17. Clark, D.E. What has computer-aided molecular design ever done for drug discovery? *Expert Opin. Drug Discov.* **2006**, *1*, 103–110. [CrossRef] [PubMed]

18. Geysen, H.M.; Schoenen, F.; Wagner, D.; Wagner, R. Combinatorial compound libraries for drug discovery: An ongoing challenge. *Nat. Rev. Drug Discov.* **2003**, *2*, 222–230. [CrossRef] [PubMed]

19. Dolle, R.E. Historical overview of chemical library design. *Methods Mol. Biol.* **2011**, *685*, 3–25. [PubMed]

20. Sanger, F. Sequences, sequences, and sequences. *Annu. Rev. Biochem.* **1988**, *57*, 1–29. [CrossRef] [PubMed]

21. Joachimiak, A. High-throughput crystallography for structural genomics. *Curr. Opin. Struct. Biol.* **2009**, *19*, 573–584. [CrossRef] [PubMed]

22. Mayr, L.M.; Fuerst, P. The future of high-throughput screening. *J. Biomol. Screen.* **2008**, *13*, 443–448. [CrossRef] [PubMed]

23. Gerwick, W.H.; Moore, B.S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* **2012**, *19*, 85–98. [CrossRef] [PubMed]

24. Cragg, G.M.; Newman, D.J. Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta* **2013**, *1830*, 3670–3695. [CrossRef] [PubMed]

25. Blunt, J.W.; Carroll, A.R.; Copp, B.R.; Davis, R.A.; Keyzers, R.A.; Prinsep, M.R. Marine natural products. *Nat. Prod. Rep.* **2017**, *34*, 235–294. [CrossRef] [PubMed]

26. Mayer, A.M.S.; Rodriguez, A.D.; Taglialatela-Scafati, O.; Fusetani, N. Marine pharmacology in 2009–2011: Marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, and antiviral activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of action. *Mar. Drugs* **2013**, *11*, 2510–2573. [PubMed]

27. Ruiz-Torres, V.; Antonio Encinar, J.; Herranz-Lopez, M.; Perez-Sanchez, A.; Galiano, V.; Barrajon-Catalan, E.; Micol, V. An updated review on marine anticancer compounds: The use of virtual screening for the discovery of small-molecule cancer drugs. *Molecules* **2017**, *22*, 1037. [CrossRef] [PubMed]

28. Choudhary, A.; Naughton, L.M.; Montanchez, I.; Dobson, A.D.W.; Rai, D.K. Current status and future prospects of marine natural products (MNPs) as antimicrobials. *Mar. Drugs* **2017**, *15*, 272. [CrossRef] [PubMed]

29. Blunt, J.W.; Copp, B.R.; Keyzers, R.A.; Munro, M.H.G.; Prinsep, M.R. Marine natural products. *Nat. Prod. Rep.* **2016**, *33*, 382–431. [CrossRef] [PubMed]

30. Harvey, A.L. Natural products in drug discovery. *Drug Discov. Today* **2008**, *13*, 894–901. [CrossRef] [PubMed]

31. Koch, M.A.; Schuffenhauer, A.; Scheck, M.; Wetzel, 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] [PubMed]

32. Berdy, J. Bioactive microbial metabolites-A personal view. *J. Antibiot.* **2005**, *58*, 1–26. [CrossRef] [PubMed]

33. Wohlleben, W.; Mast, Y.; Stegmann, E.; Ziemert, N. Antibiotic drug discovery. *Microb. Biotechnol.* **2016**, *9*, 541–548. [CrossRef] [PubMed]

34. Kang, H.K.; Seo, C.H.; Park, Y. Marine peptides and their anti-infective activities. *Mar. Drugs* **2015**, *13*, 618–654. [CrossRef] [PubMed]

35. Valliappan, K.; Sun, W.; Li, Z. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7365–7377. [CrossRef] [PubMed]

36. Ng, T.B.; Cheung, R.C.F.; Wong, J.H.; Bekhit, A.A.; Bekhit, A.E.-D. Antibacterial products of marine organisms. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 4145–4173. [CrossRef] [PubMed]

37. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef] [PubMed]

38. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. Pubchem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef] [PubMed]

39. Chen, Y.; Kops, C.d.B.; Kirchmair, J. Data resources for the computer-guided discovery of bioactive natural products. *J. Chem. Inf. Model.* **2017**, *57*, 2099–2111. [CrossRef] [PubMed]

40. Klementz, D.; Doering, K.; Lucas, X.; Telukunta, K.K.; Erxleben, A.; Deubel, D.; Erber, A.; Santillana, I.; Thomas, O.S.; Bechthold, A.; et al. StreptomeDB 2.0-an extended resource of natural products produced by streptomycetes. *Nucleic Acids Res.* **2016**, *44*, D509–D514. [CrossRef] [PubMed]

41. Choi, H.; Cho, S.Y.; Pak, H.J.; Kim, Y.; Choi, J.-Y.; Lee, Y.J.; Gong, B.H.; Kang, Y.S.; Han, T.; Choi, G.; et al. NPCARE: Database of natural products and fractional extracts for cancer regulation. *J. Cheminformatics* **2017**, *9*, 2. [CrossRef] [PubMed]

42. Pence, H.E.; Williams, A. Chemspider: An online chemical information resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124. [CrossRef]

43. Jones, L.H.; Bunnage, M.E. Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* **2017**, *16*, 285–296. [CrossRef] [PubMed]

44. Sterling, T.; Irwin, J.J. Zinc 15-ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [CrossRef] [PubMed]

45. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. HMBD 3.0-the human metabolome database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801–D807. [CrossRef] [PubMed]

46. Barneh, F.; Jafari, M.; Mirzaie, M. Updates on drug-target network; facilitating polypharmacology and data integration by growth of drugbank database. *Brief. Bioinform.* **2016**, *17*, 1070–1080. [CrossRef] [PubMed]

47. Ntie-Kang, F.; Zofou, D.; Babiaka, S.B.; Meudom, R.; Scharfe, M.; Lifongo, L.L.; Mbah, J.A.; Mbaze, L.M.; Sippl, W.; Efange, S.M.N. AfroDB: A select highly potent and diverse natural product library from African medicinal plants. *PLoS ONE* **2013**, *8*, e78085. [CrossRef] [PubMed]

48. Kang, H.; Tang, K.; Liu, Q.; Sun, Y.; Huang, Q.; Zhu, R.; Gao, J.; Zhang, D.; Huang, C.; Cao, Z. HIM-herbal ingredients in-vivo metabolism database. *J. Cheminform.* **2013**, *5*, 28. [CrossRef] [PubMed]

49. Mangal, M.; Sagar, P.; Singh, H.; Raghava, G.P.S.; Agarwal, S.M. NPACT: Naturally occurring plant-based anti-cancer compound-activity-target database. *Nucleic Acids Res.* **2013**, *41*, D1124–D1129. [CrossRef] [PubMed]

50. Valli, M.; dos Santos, R.N.; Figueira, L.D.; Nakajima, C.H.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Development of a natural products database from the biodiversity of Brazil. *J. Nat. Prod.* **2013**, *76*, 439–444. [CrossRef] [PubMed]

51. Chen, C.Y.-C. TCM database@Taiwan: The world's largest traditional chinese medicine database for drug screening in silico. *PLoS ONE* **2011**, *6*, e15939. [CrossRef] [PubMed]

52. Zhang, G.; Li, J.; Zhu, T.; Gu, Q.; Li, D. Advanced tools in marine natural drug discovery. *Curr. Opin. Biotechnol.* **2016**, *42*, 13–23. [CrossRef] [PubMed]

53. Mohamed, A.; Canh Hao, N.; Mamitsuka, H. Current status and prospects of computational resources for natural product dereplication: A review. *Brief. Bioinform.* **2016**, *17*, 309–321. [CrossRef] [PubMed]

54. Hufsky, F.; Scheubert, K.; Bocker, S. New kids on the block: Novel informatics methods for natural product discovery. *Nat. Prod. Rep.* **2014**, *31*, 807–817. [CrossRef] [PubMed]

55. Zhang, M.M.; Qiao, Y.; Ang, E.L.; Zhao, H. Using natural products for drug discovery: The impact of the genomics era. *Expert Opin. Drug Discov.* **2017**, *12*, 475–487. [CrossRef] [PubMed]

56. Chanana, S.; Thomas, C.S.; Braun, D.R.; Hou, Y.; Wyche, T.P.; Bugni, T.S. Natural product discovery using planes of principal component analysis in R (POPCAR). *Metabolites* **2017**, *7*, 34. [CrossRef] [PubMed]

57. Ellis, G.A.; Hou, Y.; Braun, D.R.; Wyche, T.P.; Adnani, N.; Vazquez-Rivera, E.; Bugni, T.S. LC/MS untargeted metabolomics for prioritizing marine invertebrate-associated bacteria for discovery of natural products. *Planta Med.* **2013**, *79*, 844–844. [CrossRef]

58. Macintyre, L.; Zhang, T.; Viegelmann, C.; Martinez, I.J.; Cheng, C.; Dowdells, C.; Abdelmohsen, U.R.; Gernert, C.; Hentschel, U.; Edrada-Ebel, R. Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Mar. Drugs* **2014**, *12*, 3416–3448. [CrossRef] [PubMed]

59. Tawfike, A.F.; Viegelmann, C.; Edrada-Ebel, R. Metabolomics and dereplication strategies in natural products. *Methods Mol. Biol.* **2013**, *1055*, 227–244. [PubMed]

60. Abdelmohsen, U.R.; Cheng, C.; Viegelmann, C.; Zhang, T.; Grkovic, T.; Ahmed, S.; Quinn, R.J.; Hentschel, U.; Edrada-Ebel, R. Dereplication strategies for targeted isolation of new antitrypanosomal actinosporins A and B from a marine sponge associated-*Actinokineospora* sp. EG49. *Mar. Drugs* **2014**, *12*, 1220–1244. [CrossRef] [PubMed]

61. Sidebottom, A.M.; Johnson, A.R.; Karty, J.A.; Trader, D.J.; Carlson, E.E. Integrated metabolomics approach facilitates discovery of an unpredicted natural product suite from streptomyces coelicolor M145. *ACS Chem. Biol.* **2013**, *8*, 2009–2016. [CrossRef] [PubMed]

62. Tawfike, A.F.; Tate, R.; Abbott, G.; Young, L.; Viegelmann, C.; Schumacher, M.; Diederich, M.; Edrada-Ebel, R.A. Metabolomic tools to assess the chemistry and bioactivity of endophytic Aspergillus strain. *Chem. Biodivers.* **2017**, *14*, e1700040. [CrossRef] [PubMed]

63. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]

64. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, *78*, 779–787. [CrossRef] [PubMed]

65. Roullier, C.; Guitton, Y.; Valery, M.; Amand, S.; Prado, S.; du Pont, T.R.; Grovel, O.; Pouchus, Y.F. Automated detection of natural halogenated compounds from LC-MS profiles-application to the isolation of bioactive chlorinated compounds from marine-derived fungi. *Anal. Chem.* **2016**, *88*, 9143–9150. [CrossRef] [PubMed]

66. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Don Duy, N.; 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]

67. Crusemann, M.; O'Neill, E.C.; Larson, C.B.; Melnik, A.V.; Floros, D.J.; da Silva, R.R.; Jensen, P.R.; Dorrestein, P.C.; Moore, B.S. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J. Nat. Prod.* **2017**, *80*, 588–597. [CrossRef] [PubMed]

68. Parrot, D.; Intertaglia, L.; Jehan, P.; Grube, M.; Suzuki, M.T.; Tomasi, S. Chemical analysis of the alphaproteobacterium strain MOLA1416 associated with the marine lichen *Lichina pygmaea*. *Phytochemistry* **2018**, *145*, 57–67. [CrossRef] [PubMed]

69. Kerber, A.; Laue, R.; Meringer, M.; Rucker, C. Molecules in silico: The generation of structural formulae and its applications. *J. Comput. Chem. Jpn.* **2004**, *3*, 85–96. [CrossRef]

70. Jeffryes, J.G.; Colastani, R.L.; Elbadawi-Sidhu, M.; Kind, T.; Niehaus, T.D.; Broadbelt, L.J.; Hanson, A.D.; Fiehn, O.; Tyo, K.E.J.; Henry, C.S. Mines: Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *J. Cheminform.* **2015**, *7*, 44. [CrossRef] [PubMed]

71. Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **2018**, *15*, 53–56. [CrossRef] [PubMed]

72. Kind, T.; Liu, K.-H.; Lee, D.Y.; DeFelice, B.; Meissen, J.K.; Fiehn, O. Lipidblast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **2013**, *10*, 755–758. [CrossRef] [PubMed]

73. Hufsky, F.; Boecker, S. Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrom. Rev.* **2017**, *36*, 624–633. [CrossRef] [PubMed]

74. Kangas, L.J.; Metz, T.O.; Isaac, G.; Schrom, B.T.; Ginovska-Pangovska, B.; Wang, L.; Tan, L.; Lewis, R.R.; Miller, J.H. In silico identification software (ISIS): A machine learning approach to tandem mass spectral identification of lipids. *Bioinformatics* **2012**, *28*, 1705–1713. [CrossRef] [PubMed]

75. Allen, F.; Greiner, R.; Wishart, D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* **2015**, *11*, 98–110. [CrossRef]

76. Wolf, S.; Schmidt, S.; Mueller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]

77. Gerlich, M.; Neumann, S. Metfusion: Integration of compound identification strategies. *J. Mass Spectrom.* **2013**, *48*, 291–298. [CrossRef] [PubMed]

78. Wang, Y.; Kora, G.; Bowen, B.P.; Pan, C. Midas: A database-searching algorithm for metabolite identification in metabolomics. *Anal. Chem.* **2014**, *86*, 9496–9503. [CrossRef] [PubMed]

79. Ridder, L.; van der Hooft, J.J.J.; Verhoeven, S.; de Vos, R.C.H.; van Schaik, R.; Vervoort, J. Substructure-based annotation of high-resolution multistage MSn spectral trees. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2461–2471. [CrossRef] [PubMed]

80. Ridder, L.; van der Hooft, J.J.J.; Verhoeven, S. Automatic compound annotation from mass spectrometry data using MAGMa. *Mass Spectrom.* **2014**, *3*, S0033. [CrossRef] [PubMed]

81. Verdegem, D.; Lambrechts, D.; Carmeliet, P.; Ghesquiere, B. Improved metabolite identification with MIDAS and MAGMa through MS/MS spectral dataset-driven parameter optimization. *Metabolomics* **2016**, *12*, 98. [CrossRef]

82. Heinonen, M.; Shen, H.; Zamboni, N.; Rousu, J. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics* **2012**, *28*, 2333–2341. [CrossRef] [PubMed]

83. Shen, H.; Duhrkop, K.; Bocker, S.; Rousu, J. Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics* **2014**, *30*, 157–164. [CrossRef] [PubMed]

84. Boecker, S.; Rasche, F. Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics* **2008**, *24*, I49–I55. [CrossRef] [PubMed]

85. Duehrkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Boecker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12580–12585. [CrossRef] [PubMed]

86. Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes—A review. *Nat. Prod. Rep.* **2016**, *33*, 988–1005. [CrossRef] [PubMed]

87. Medema, M.H.; Fischbach, M.A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639–648. [CrossRef] [PubMed]

88. Bentley, S.D.; Chater, K.F.; Cerdeno-Tarraga, A.M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417*, 141–147. [CrossRef] [PubMed]

89. Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **2003**, *21*, 526–531. [CrossRef] [PubMed]

90. Udwary, D.W.; Zeigler, L.; Asolkar, R.N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P.R.; Moore, B.S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10376–10381. [CrossRef] [PubMed]

91. Schulze, C.J.; Donia, M.S.; Siqueira-Neto, J.L.; Ray, D.; Raskatov, J.A.; Green, R.E.; McKerrow, J.H.; Fischbach, M.A.; Linington, R.G. Genome-directed lead discovery: Biosynthesis, structure elucidation, and biological evaluation of two families of polyene macrolactams against *Trypanosoma brucei. ACS Chem. Biol.* **2015**, *10*, 2373–2381. [CrossRef] [PubMed]

92. Leikoski, N.; Liu, L.W.; Jokela, J.; Wahlsten, M.; Gugger, M.; Calteau, A.; Permi, P.; Kerfeld, C.A.; Sivonen, K.; Fewer, D.P. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chem. Biol.* **2013**, *20*, 1033–1043. [CrossRef] [PubMed]

93. Weber, T.; Kim, H.U. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Syst. Synth. Biotechnol.* **2016**, *1*, 69–79. [CrossRef] [PubMed]

94. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

95. Finn, R.D.; Clements, J.; Eddy, S.R. Hmmer web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef] [PubMed]

96. Starcevic, A.; Zucko, J.; Simunkovic, J.; Long, P.F.; Cullum, J.; Hranueli, D. Clustscan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* **2008**, *36*, 6882–6892. [CrossRef] [PubMed]

97. Khaldi, N.; Seifuddin, F.T.; Turner, G.; Haft, D.; Nierman, W.C.; Wolfe, K.H.; Fedorova, N.D. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **2010**, *47*, 736–741. [CrossRef] [PubMed]

98. Blin, K.; Wolf, T.; Chevrette, M.G.; Lu, X.W.; Schwalen, C.J.; Kautsar, S.A.; Duran, H.G.S.; Santos, E.; Kim, H.U.; Nave, M.; et al. AntiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45*, W36–W41. [CrossRef] [PubMed]

99. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Brown, L.C.W.; 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] [PubMed]

100. Cruz-Morales, P.; Kopp, J.F.; Martinez-Guerrero, C.; Alfonso Yanez-Guerra, L.; Selem-Mojica, N.; Ramos-Aboites, H.; Feldmann, J.; Barona-Gomez, F. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model Streptomycetes. *Genome Biol. Evol.* **2016**, *8*, 1906–1916. [CrossRef] [PubMed]

101. Takeda, I.; Umemura, M.; Koike, H.; Asai, K.; Machida, M. Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: Application to sequenced genomes of aspergillus and ten other filamentous fungal species. *DNA Res.* **2014**, *21*, 447–457. [CrossRef] [PubMed]

102. Skinnider, M.A.; Dejong, C.A.; Rees, P.N.; Johnston, C.W.; Li, H.; Webster, A.L.H.; Wyatt, M.A.; Magarvey, N.A. Genomes to natural products prediction informatics for secondary metabolomes (PRISM). *Nucleic Acids Res.* **2015**, *43*, 9645–9662. [CrossRef] [PubMed]

103. Tang, X.Y.; Li, J.; Millan-Aguinaga, N.; Zhang, J.J.; O'Neill, E.C.; Ugalde, J.A.; Jensen, P.R.; Mantovani, S.M.; Moore, B.S. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **2015**, *10*, 2841–2849. [CrossRef] [PubMed]

104. Wolf, T.; Shelest, V.; Nath, N.; Shelest, E. CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics* **2016**, *32*, 1138–1143. [CrossRef] [PubMed]

105. Steinbeck, C. Seneca: A platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1500–1507. [CrossRef] [PubMed]

106. Jayaseelan, K.V.; Steinbeck, C. Building blocks for automated elucidation of metabolites: Natural product-likeness for candidate ranking. *BMC Bioinform.* **2014**, *15*, 234. [CrossRef] [PubMed]

107. Blinov, K.A.; Carlson, D.; Elyashberg, M.E.; Martin, G.E.; Martirosian, E.R.; Molodtsov, S.; Williams, A.J. Computer-assisted structure elucidation of natural products with limited 2D NMR data: Application of the struceluc system. *Magn. Reson. Chem.* **2003**, *41*, 359–372. [CrossRef]

108. Elyashberg, M.E.; Blinov, K.A.; Williams, A.J.; Molodtsov, S.G.; Martin, G.E.; Martirosian, E.R. Structure elucidator: A versatile expert system for molecular structure elucidation from 1D and 2D NMR data and molecular fragments. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 771–792. [CrossRef] [PubMed]

109. Molodtsov, S.G.; Elyashberg, M.E.; Blinov, K.A.; Williams, A.J.; Martirosian, E.E.; Martin, G.E.; Lefebvre, B. Structure elucidation from 2D NMR spectra using the struceluc expert system: Detection and removal of contradictions in the data. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1737–1751. [CrossRef] [PubMed]

110. Plainchont, B.; de Paulo Emerenciano, V.; Nuzillard, J.-M. Recent advances in the structure elucidation of small organic molecules by the LSD software. *Magn. Reson. Chem.* **2013**, *51*, 447–453. [CrossRef] [PubMed]

111. Troche-Pesqueira, E.; Anklin, C.; Gil, R.R.; Navarro-Vazquez, A. Computer-assisted 3D structure elucidation of natural products using residual dipolar couplings. *Angew. Chem. Int. Ed.* **2017**, *56*, 3660–3664. [CrossRef] [PubMed]

112. Liu, Y.; Sauri, J.; Mevers, E.; Peczuh, M.W.; Hiemstra, H.; Clardy, J.; Martin, G.E.; Williamson, R.T. Unequivocal determination of complex molecular structures using anisotropic NMR measurements. *Science* **2017**, *356*, 43. [CrossRef] [PubMed]

113. Feliciano, A.S.; Medarde, M.; Delcorral, J.M.M.; Aramburu, A.; Gordaliza, M.; Barrero, A.F. Aquatolide—A new type of humulane-related sesquiterpene lactone. *Tetrahedron Lett.* **1989**, *30*, 2851–2854. [CrossRef]

114. Lodewyk, M.W.; Soldi, C.; Jones, P.B.; Olmstead, M.M.; Rita, J.; Shaw, J.T.; Tantillo, D.J. The correct structure of aquatolide-experimental validation of a theoretically-predicted structural revision. *J. Am. Chem. Soc.* **2012**, *134*, 18550–18553. [CrossRef] [PubMed]

115. Saya, J.M.; Vos, K.; Kleinnijenhuis, R.A.; van Maarseveen, J.H.; Ingemann, S.; Hiemstra, H. Total synthesis of aquatolide. *Org. Lett.* **2015**, *17*, 3892–3894. [CrossRef] [PubMed]

116. Buevich, A.V.; Elyashberg, M.E. Synergistic combination of CASE algorithms and DFT chemical shift predictions: A powerful approach for structure elucidation, verification, and revision. *J. Nat. Prod.* **2016**, *79*, 3105–3116. [CrossRef] [PubMed]

117. Saunders, C.M.; Tantillo, D.J. Application of computational chemical shift prediction techniques to the cereoanhydride structure problem-carboxylate complications. *Mar. Drugs* **2017**, *15*, 171. [CrossRef] [PubMed]

118. Buevich, A.V.; Elyashberg, M.E. Towards unbiased and more versatile NMR-based structure elucidation: A powerful combination of CASE algorithms and DFT calculations. *Magn. Reson. Chem.* **2017**, *56*, 493–504. [CrossRef] [PubMed]

119. Koren-Goldshlager, G.; Aknin, M.; Kashman, Y. Cycloshermilamine D, a new pyridoacridine from the marine tunicate *Cystodytes violatinctus*. *J. Nat. Prod.* **2000**, *63*, 830–831. [CrossRef] [PubMed]

120. 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] [PubMed]

121. 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]

122. 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] [PubMed]

123. Hsin, K.-Y.; Matsuoka, Y.; Asai, Y.; Kamiyoshi, K.; Watanabe, T.; Kawaoka, Y.; Kitano, H. Systemsdock: A web server for network pharmacology-based prediction and analysis. *Nucleic Acids Res.* **2016**, *44*, W507–W513. [CrossRef] [PubMed]

124. Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner, C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijer, M.B.; Matos, R.C.; Tran, T.B.; et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175. [CrossRef] [PubMed]

125. Espinoza-Moraga, M.; Njuguna, N.M.; Mugumbate, G.; Caballero, J.; Chibale, K. In silico comparison of antimycobacterial natural products with known antituberculosis drugs. *J. Chem. Inf. Model.* **2013**, *53*, 649–660. [CrossRef] [PubMed]

126. Chen, J.; Gao, L.-X.; Gong, J.-X.; Jiang, C.-S.; Yao, L.-G.; Li, J.-Y.; Li, J.; Xiao, W.; Guo, Y.-W. Design and synthesis of novel 1,2-dithiolan-4-yl benzoate derivatives as PTP1B inhibitors. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 2211–2216. [CrossRef] [PubMed]

127. Kudryavtsev, D.; Makarieva, T.; Utkina, N.; Santalova, E.; Kryukova, E.; Methfessel, C.; Tsetlin, V.; Stonik, V.; Kasheverov, I. Marine natural products acting on the acetylcholine-binding protein and nicotinic receptors: From computer modeling to binding studies and electrophysiology. *Mar. Drugs* **2014**, *12*, 1859–1875. [CrossRef] [PubMed]

128. Huang, X.-C.; Xiao, X.; Zhang, Y.-K.; Talele, T.T.; Salim, A.A.; Chen, Z.-S.; Capon, R.J. Lamellarin O, a pyrrole alkaloid from an Australian marine sponge, *Ianthella* sp., reverses BCRP mediated drug resistance in cancer cells. *Mar. Drugs* **2014**, *12*, 3818–3837. [CrossRef] [PubMed]

129. Liu, J.; Chen, W.; Xu, Y.; Ren, S.; Zhang, W.; Li, Y. Design, synthesis and biological evaluation of tasiamide B derivatives as BACE1 inhibitors. *Bioorg. Med. Chem.* **2015**, *23*, 1963–1974. [CrossRef] [PubMed]

130. Manglik, A.; Lin, H.; Aryal, D.K.; McCorvy, J.D.; Dengler, D.; Corder, G.; Levit, A.; Kling, R.C.; Bernat, V.; Huebner, H.; et al. Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **2016**, *537*, 185. [CrossRef] [PubMed]

131. Ebrahim, H.Y.; El Sayed, K.A. Discovery of novel antiangiogenic marine natural product scaffolds. *Mar. Drugs* **2016**, *14*, 57. [CrossRef] [PubMed]

132. Skariyachan, S.; Acharya, A.B.; Subramaniyan, S.; Babu, S.; Kulkarni, S.; Narayanappa, R. Secondary metabolites extracted from marine sponge associated Comamonas testosteroni and Citrobacter freundii as potential antimicrobials against MDR pathogens and hypothetical leads for VP40 matrix protein of Ebola virus: An in vitro and in silico investigation. *J. Biomol. Struct. Dyn.* **2016**, *34*, 1865–1883. [PubMed]

133. Fang, J.; Yang, R.; Gao, L.; Zhou, D.; Yang, S.; Liu, A.-L.; Du, G.-H. Predictions of BuChE inhibitors using support vector machine and naïve bayesian classification techniques in drug discovery. *J. Chem. Inf. Model.* **2013**, *53*, 3009–3020. [CrossRef] [PubMed]

134. Davis, G.D.J.; Vasanthi, A.H.R. QSAR based docking studies of marine algal anticancer compounds as inhibitors of protein kinase B (PKB). *Eur. J. Pharm. Sci.* **2015**, *76*, 110–118. [CrossRef] [PubMed]

135. Knight, N.J.; Hernando, E.; Haynes, C.J.E.; Busschaert, N.; Clarke, H.J.; Takimoto, K.; Garcia-Valverde, M.; Frey, J.G.; Quesada, R.; Gale, P.A. QSAR analysis of substituent effects on tambjamine anion transporters. *Chem. Sci.* **2016**, *7*, 1600–1608. [CrossRef] [PubMed]

136. Aswathy, L.; Jisha, R.S.; Masand, V.H.; Gajbhiye, J.M.; Shibi, I.G. Computational strategies to explore antimalarial thiazine alkaloid lead compounds based on an Australian marine sponge *Plakortis lita*. *J. Biomol. Struct. Dyn.* **2017**, *35*, 2407–2429. [CrossRef] [PubMed]

137. Flores, M.C.; Marquez, E.A.; Mora, J.R. Molecular modeling studies of bromopyrrole alkaloids as potential antimalarial compounds: A DFT approach. *Med. Chem. Res.* **2018**, *27*, 844–856. [CrossRef]

138. Llorach-Pares, L.; Nonell-Canals, A.; Sanchez-Martinez, M.; Avila, C. Computer-aided drug design applied to marine drug discovery: Meridianins as Alzheimer's disease therapeutic agents. *Mar. Drugs* **2017**, *15*, 366. [CrossRef] [PubMed]

139. Dineshkumar, K.; Aparna, V.; Madhuri, K.Z.; Hopper, W. Biological activity of sporolides A and B from salinispora tropica: In silico target prediction using ligand-based pharmacophore mapping and in vitro activity validation on HIV-1 reverse transcriptase. *Chem. Biol. Drug Des.* **2014**, *83*, 350–361. [CrossRef] [PubMed]

140. Dineshkumar, K.; Aparna, V.; Hopper, W. Ligand based-pharmacophore modeling and extended bioactivity prediction for salinosporamide A, B and C from marine actinomycetes *Salinispora tropica*. *Comb. Chem. High Throughput Screen.* **2017**, *20*, 3–19. [CrossRef] [PubMed]

141. Wetzel, S.; Bon, R.S.; Kumar, K.; Waldmann, H. Biology-oriented synthesis. *Angew. Chem. Int. Ed.* **2011**, *50*, 10800–10826. [CrossRef] [PubMed]

142. Pereira, F.; Latino, D.A.R.S.; Gaudencio, S.P. A chemoinformatics approach to the discovery of lead-like molecules from marine and microbial sources en route to antitumor and antibiotic drugs. *Mar. Drugs* **2014**, *12*, 757–778. [CrossRef] [PubMed]

143. Shang, J.; Hu, B.; Wang, J.; Zhu, F.; Kang, Y.; Li, D.; Sun, H.; Kong, D.-X.; Hou, T. A cheminformatic insight into the differences between terrestrial and marine originated natural products. *J. Chem. Inf. Model.* **2018**. [CrossRef] [PubMed]

144. Visini, R.; Arus-Pous, J.; Awale, M.; Reymond, J.-L. Virtual exploration of the ring systems chemical universe. *J. Chem. Inf. Model.* **2017**, *57*, 2707–2718. [CrossRef] [PubMed]

145. Wang, Q.; Sciabola, S.; Barreiro, G.; Hou, X.; Bai, G.; Shapiro, M.J.; Koehn, F.; Villalobos, A.; Jacobson, M.P. Dihedral angle-based sampling of natural product polyketide conformations: Application to permeability prediction. *J. Chem. Inf. Model.* **2016**, *56*, 2194–2206. [CrossRef] [PubMed]

146. Tatonetti, N.P.; Liu, T.; Altman, R.B. Predicting drug side-effects by chemical systems biology. *Genome Biol.* **2009**, *10*. [CrossRef] [PubMed]

147. Hopkins, A.L.; Mason, J.S.; Overington, J.P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127–136. [CrossRef] [PubMed]

148. Vanhaelen, Q.; Mamoshina, P.; Aliper, A.M.; Artemov, A.; Lezhnina, K.; Ozerov, I.; Labat, I.; Zhavoronkov, A. Design of efficient computational workflows for in silico drug repurposing. *Drug Discov. Today* **2017**, *22*, 210–222. [CrossRef] [PubMed]

149. Reker, D.; Rodrigues, T.; Schneider, P.; Schneider, G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4067–4072. [CrossRef] [PubMed]

150. Schneider, G.; Reker, D.; Chen, T.; Hauenstein, K.; Schneider, P.; Altmann, K.-H. Deorphaning the macromolecular targets of the natural anticancer compound doliculide. *Angew. Chem. Int. Ed.* **2016**, *55*, 12408–12411. [CrossRef] [PubMed]

151. Pereira, F.; Latino, D.A.R.S.; Gaudencio, S.P. QSAR-assisted virtual screening of lead-like molecules from marine and microbial natural sources for antitumor and antibiotic drug discovery. *Molecules* **2015**, *20*, 4848–4873. [CrossRef] [PubMed]

152. Botic, T.; Defant, A.; Zanini, P.; Zuzek, M.C.; Frangez, R.; Janussen, D.; Kersken, D.; Knez, Z.; Mancini, I.; Sepcic, K. Discorhabdin alkaloids from *Antarctic latrunculia* spp. Sponges as a new class of cholinesterase inhibitors. *Eur. J. Med. Chem.* **2017**, *136*, 294–304. [CrossRef] [PubMed]

153. Cen-Pacheco, F.; Perez Manriquez, C.; Souto, L.M.; Norte, M.; Javier Fernandez, J.; Hernandez Daranas, A. Marine longilenes, oxasqualenoids with Ser-Thr protein phosphatase 2A inhibition activity. *Mar. Drugs* **2018**, *16*, 131. [CrossRef] [PubMed]

154. Cruz, P.G.; Martinez Leal, J.F.; Hernandez Daranas, A.; Perez, M.; Cuevas, C. On the mechanism of action of dragmacidins I and J, two new representatives of a new class of protein phosphatase 1 and 2A inhibitors. *ACS Omega* **2018**, *3*, 3760–3767. [CrossRef]

155. Xin, L.-T.; Liu, L.; Shao, C.-L.; Yu, R.-L.; Chen, F.-L.; Yue, S.-J.; Wang, M.; Guo, Z.-L.; Fan, Y.-C.; Guan, H.-S.; et al. Discovery of DNA topoisomerase I inhibitors with low-cytotoxicity based on virtual screening from natural products. *Mar. Drugs* **2017**, *15*, 217. [CrossRef] [PubMed]

156. Wu, G.; Qi, X.; Mo, X.; Yu, G.; Wang, Q.; Zhu, T.; Gu, Q.; Liu, M.; Li, J.; Li, D. Structure-based discovery of cytotoxic dimeric tetrahydroxanthones as potential topoisomerase I inhibitors from a marine-derived fungus. *Eur. J. Med. Chem.* **2018**, *148*, 268–278. [CrossRef] [PubMed]

157. Ali, M.Y.; Kim, D.H.; Seong, S.H.; Kim, H.-R.; Jung, H.A.; Choi, J.S. Alpha-glucosidase and protein tyrosine phosphatase 1B inhibitory activity of plastoquinones from marine brown alga *Sargassum serratifolium*. *Mar. Drugs* **2017**, *15*, 368. [CrossRef] [PubMed]

158. Xu, Q.; Luo, J.; Wu, N.; Zhang, R.; Shi, D. BPN, a marine-derived PTP1B inhibitor, activates insulin signaling and improves insulin resistance in C2C12 myotubes. *Int. J. Biol. Macromol.* **2018**, *106*, 379–386. [CrossRef] [PubMed]

159. Chen, F.; Wang, Z.; Wang, C.; Xu, Q.; Liang, J.; Xu, X.; Yang, J.; Wang, C.; Jiang, T.; Yu, R. Application of reverse docking for target prediction of marine compounds with anti-tumor activity. *J. Mol. Graph. Model.* **2017**, *77*, 372–377. [CrossRef] [PubMed]

160. Chen, L.; Zhao, Y.-Y.; Lan, R.-F.; Du, L.; Wang, B.-S.; Zhou, T.; Li, Y.-P.; Zhang, Q.-Q.; Ying, M.-G.; Zheng, Q.-H.; et al. Dicitrinone D, an antimitotic polyketide isolated from the marine-derived fungus *Penicillium citrinum*. *Tetrahedron* **2017**, *73*, 5900–5911. [CrossRef]

161. Yu, G.; Wang, Y.; Yu, R.; Feng, Y.; Wang, L.; Che, Q.; Gu, Q.; Li, D.; Li, J.; Zhu, T. Chetracins E and F, cytotoxic epipolythiodioxopiperazines from the marine-derived fungus *Acrostalagmus luteoalbus* HDN13-530. *RSC Adv.* **2018**, *8*, 53–58. [CrossRef]

162. Sharma, K.; Sharma, D.; Sharma, M.; Sharma, N.; Bidve, P.; Prajapati, N.; Kalia, K.; Tiwari, V. Astaxanthin ameliorates behavioral and biochemical alterations in in-vitro and in-vivo model of neuropathic pain. *Neurosci. Lett.* **2018**, *674*, 162–170. [CrossRef] [PubMed]

163. Ko, S.-C.; Jang, J.; Ye, B.-R.; Kim, M.-S.; Choi, I.-W.; Park, W.-S.; Heo, S.-J.; Jung, W.-K. Purification and molecular docking study of angiotensin I-converting enzyme (ACE) inhibitory peptides from hydrolysates of marine sponge *Stylotella aurantium*. *Process Biochem.* **2017**, *54*, 180–187. [CrossRef]

164. Pereira, R.C.C.; Lourenco, A.L.; Terra, L.; Abreu, P.A.; Laneuville Teixeira, V.; Castro, H.C. Marine diterpenes: Molecular modeling of thrombin inhibitors with potential biotechnological application as an antithrombotic. *Mar. Drugs* **2017**, *15*, 79. [CrossRef] [PubMed]

165. Wang, W.; Chen, R.; Luo, Z.; Wang, W.; Chen, J. Antimicrobial activity and molecular docking studies of a novel anthraquinone from a marine-derived fungus *Aspergillus versicolor*. *Nat. Prod. Res.* **2018**, *32*, 558–563. [CrossRef] [PubMed]

166. Deplazes, E. Molecular simulations of disulfide-rich venom peptides with ion channels and membranes. *Molecules* **2017**, *22*, 362. [CrossRef] [PubMed]

167. Mohyeldin, M.M.; Akl, M.R.; Siddique, A.B.; Hassan, H.M.; El Sayed, K.A. The marine-derived pachycladin diterpenoids as novel inhibitors of wild-type and mutant EGFR. *Biochem. Pharmacol.* **2017**, *126*, 51–68. [CrossRef] [PubMed]

168. Sanchez-Murcia, P.A.; Cortes-Cabrera, A.; Gago, F. Structural rationale for the cross-resistance of tumor cells bearing the A399V variant of elongation factor eEF1A1 to the structurally unrelated didemnin B, ternatin, nannocystin A and ansatrienin B. *J. Comput. Aided Mol. Des.* **2017**, *31*, 915–928. [CrossRef] [PubMed]

169. Jung, H.A.; Roy, A.; Jung, J.H.; Choi, J.S. Evaluation of the inhibitory effects of eckol and dieckol isolated from edible brown alga Eisenia bicyclis on human monoamine oxidases A and B. *Arch. Pharm. Res.* **2017**, *40*, 480–491. [CrossRef] [PubMed]

170. Naine, S.J.; Devi, C.S.; Mohanasrinivasan, V.; Doss, C.G.P. Bioactivity of marine Streptomyces sp VITJS4: Interactions of cytotoxic phthalate derivatives with human topoisomerase II alpha: An in silico molecular docking analysis. *Interdiscip. Sci.* **2018**, *10*, 261–270. [CrossRef] [PubMed]

171. Sun, Y.; Ai, X.; Hou, J.; Ye, X.; Liu, R.; Shen, S.; Lia, Z.; Lu, S. Integrated discovery of FOXO1-DNA stabilizers from marine natural products to restore chemosensitivity to anti-EGFR-based therapy for metastatic lung cancer. *Mol. BioSyst.* **2017**, *13*, 330–337. [CrossRef] [PubMed]

172. Mallipeddi, P.L.; Pedersen, S.E.; Briggs, J.M. Interactions of acetylcholine binding site residues contributing to nicotinic acetylcholine receptor gating: Role of residues Y93, Y190, K145 and D200. *J. Mol. Graph. Model.* **2013**, *44*, 145–154. [CrossRef] [PubMed]

173. Avram, S.; Pacureanu, L.M.; Seclaman, E.; Bora, A.; Kurunczi, L. PLS-DA-docking optimized combined energetic terms (PLSDA-DOCET) protocol: A brief evaluation. *J. Chem. Inf. Model.* **2011**, *51*, 3169–3179. [CrossRef] [PubMed]