Chemical structure metagenomics of microbial natural products: surveying nonribosomal peptides and beyond

Thomas Ma and John Chu[*]

Abstract

Bioactivity-guided fractionation (BGF) has historically been a fruitful natural product discovery workflow. However, it is plagued by increasing rediscovery rates in recent years and new methods capable of exploring the natural product chemical space more broadly and more efficiently is in urgent need. Chemical structure metagenomics as one such method is the theme of this Perspective. It emphasizes a chemical-structure-centered viewpoint toward natural product research. Key to chemical structure metagenomics is the ability to predict the structure of a natural product based on its biosynthetic gene sequences, which facilitated the discovery of numerous new bioactive molecules and helped uncover oversampled/underexplored niches of decades of BGF based discovery. While microbial nonribosomal peptides have been the focus of chemical structure metagenomics efforts thus far, it is in principle applicable to other natural product families. The future outlook of this new approach will also be discussed.

Introduction

Natural products present diverse structures to elicit a wide range of biological activities and are of paramount importance to both translational science and basic research. Despite not knowing the underlying active ingredient(s) and their mechanisms of action, humans have taken advantage of the therapeutic effects of herbal and microbial extracts for thousands of years [1]. We now know that natural products, genetically encoded secondary metabolites, are in most cases responsible for the desirable effects in these traditional medicine formulations [2]. Modern science has found that natural products are uniquely suited as the starting point for the development of new drugs [3,4]. They

may serve as chemical probes and help confirm laboratory findings [5]; furthermore, entirely new physical and chemical principles are sometimes uncovered as a result of natural product research [6-9].

Two milestone achievements ushered in the modern era of natural product research. The first was the serendipitous discovery of penicillin in 1929 by Alexander Fleming, which went on to save millions of lives [10]. The second was the invention of the bioactivity-guided fractionation (BGF) workflow by Selman Waksman to search systematically for new bio-

active small molecules (Figure 1) [11]. He believed that the fungus *Penicillium rubens* produces penicillin to suppress the growth of other microorganisms, especially those it competes with for nutrient and space, and that "chemical weapons" of this sort are widespread in nature. Waksman used the BGF workflow to find more than 20 new antibiotics throughout his career, wherein microbial culture extracts were fractionated and then tested for antimicrobial activity. The active fractions would be fractionated further, tested again, and this process would be performed iteratively until a pure compound is obtained. Notably, BGF is not limited to the identification of metabolites with antimicrobial activity or those of bacterial origin. It is applicable to the screening of natural products of almost any bioactivity of interest that originate from plants, fungi, and even animals (such as sponges and corals).
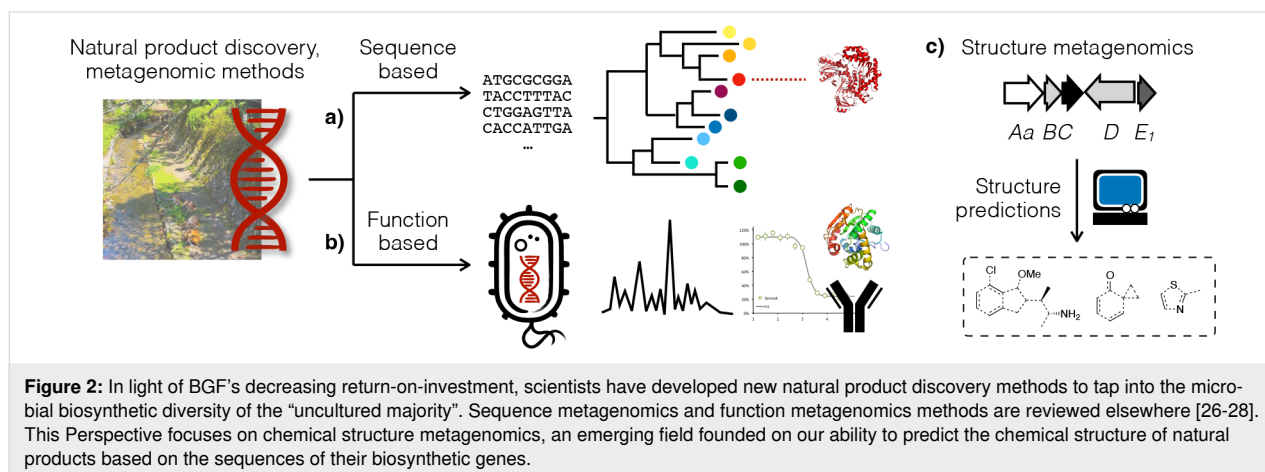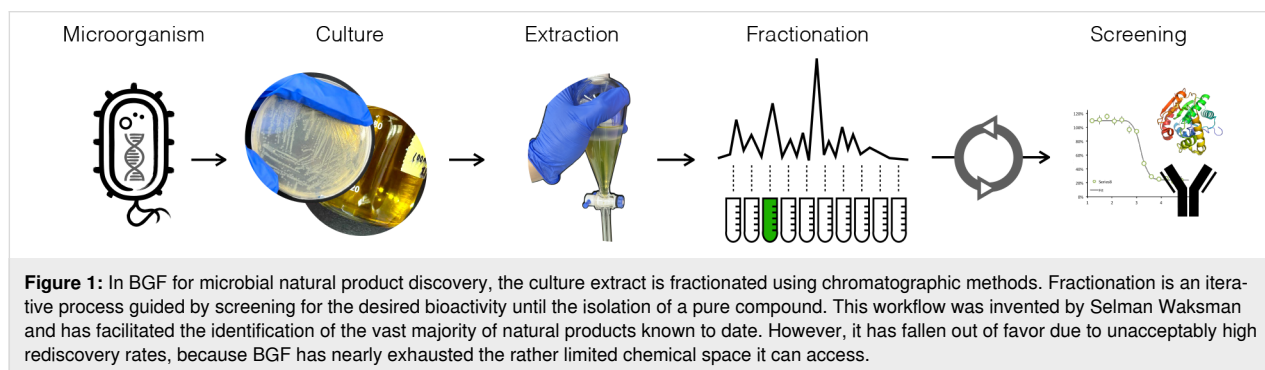
Perspective
BGF can access only a fraction of the natural product chemical space

For nearly a century, BGF has been the method of choice and identified the vast majority of natural products known to date. However, not long after the extensive application of BGF, scientists noticed an unwanted trend, i.e., it became more and

more likely to isolate a natural product that has already been characterized, also known as "rediscovery" [4]. As rediscovery amounts to nothing but wasted time and money, BGF faced a diminished return-on-investment problem and gradually fell out of favor. This problem can be attributed to the fact that BGF nearly exhausted the rather limited natural product chemical space it can access [12,13]. Indeed, it has long been known that only a fraction of microorganisms is readily cultured in the laboratory [14-16]. More recently, large-scale sequencing studies and bioinformatic analyses estimated that BGF-based discovery covered only ≈1% of the biosynthetic diversity nature has to offer [17]. This is because there are two pre-requisites for a natural product to be amenable to the BGF workflow. Not only the microorganism of interest needs to be readily cultured, it must actively express its natural product biosynthetic gene cluster(s) (BGC) under laboratory culture conditions [18,19]. It turns out these pre-requisites are rarely met [16].

New approaches to enable broader exploration

New approaches have been developed to circumvent the aforementioned constraints and enable the exploration of a broader natural product chemical space. Existing approaches fall into two broad categories – sequence metagenomics (Figure 2a) and

Figure 1: In BGF for microbial natural product discovery, the culture extract is fractionated using chromatographic methods. Fractionation is an iterative process guided by screening for the desired bioactivity until the isolation of a pure compound. This workflow was invented by Selman Waksman and has facilitated the identification of the vast majority of natural products known to date. However, it has fallen out of favor due to unacceptably high rediscovery rates, because BGF has nearly exhausted the rather limited chemical space it can access.

Figure 2: In light of BGF's decreasing return-on-investment, scientists have developed new natural product discovery methods to tap into the microbial biosynthetic diversity of the "uncultured majority". Sequence metagenomics and function metagenomics methods are reviewed elsewhere [26-28]. This Perspective focuses on chemical structure metagenomics, an emerging field founded on our ability to predict the chemical structure of natural products based on the sequences of their biosynthetic genes.

function metagenomics (Figure 2b). The former analyzes nucleic acid sequences to prioritize BGCs that are worth exploring [20]. For example, it has been shown that tracking characteristic biosynthetic or self-resistance gene(s) can facilitate the discovery of new congeners of a natural product family [21,22]. Function metagenomics uses genetically tractable model organisms, such as *Escherichia coli* or *Streptomyces albus*, to express DNA extracted from the environment and then screen for the phenotype of interest [23,24]. This approach was used to identify BGCs that produce antibiotics, pigments, compounds that alter the morphology of the host, etc. [25]. Sequence and function metagenomic approaches have been reviewed elsewhere and will not be covered herein [26-28].

This Perspective focuses on chemical structure metagenomics (Figure 2c), an emerging field that integrates bioinformatics, chemical synthesis, molecular biology, etc., and emphasizes a chemical-structure-centered viewpoint toward natural product research. It should be noted that these metagenomic approaches are not mutually exclusive; rather, they complement each other and together provide a more complete picture. Chemical structure metagenomics approaches have already facilitated the discovery of numerous new bioactive molecules [29-36] and shed light on the scope of past discovery efforts by uncovering select oversampled/underexplored niches in the natural product chemical spaces [37]. While nonribosomal peptides (NRP) from microorganisms have been the focus of chemical structure metagenomics approaches thus far, it is in principle applicable to other natural product families as well, and the possibility of extending it beyond NRPs will be discussed at the end.

Chemical structure is the universal language of nature

The chemical structure of a molecule is defined as the arrangement of its atoms and bonds, which describes not just the size and shape of a molecule, but also encompass its stereochemistry, charges, polarity, functional groups, as well as the way these elements are spatially oriented. The chemical structure of a molecule defines its properties and reactivity, and therefore dictates how it interacts with other molecules. Microorganisms need to communicate with each other and the environment, secreting signals of friendship, disdain, and many other sentiments in between. Natural products are the vocabulary for this molecular language, and knowing their chemical structures is key to understanding this form of information exchange [38].
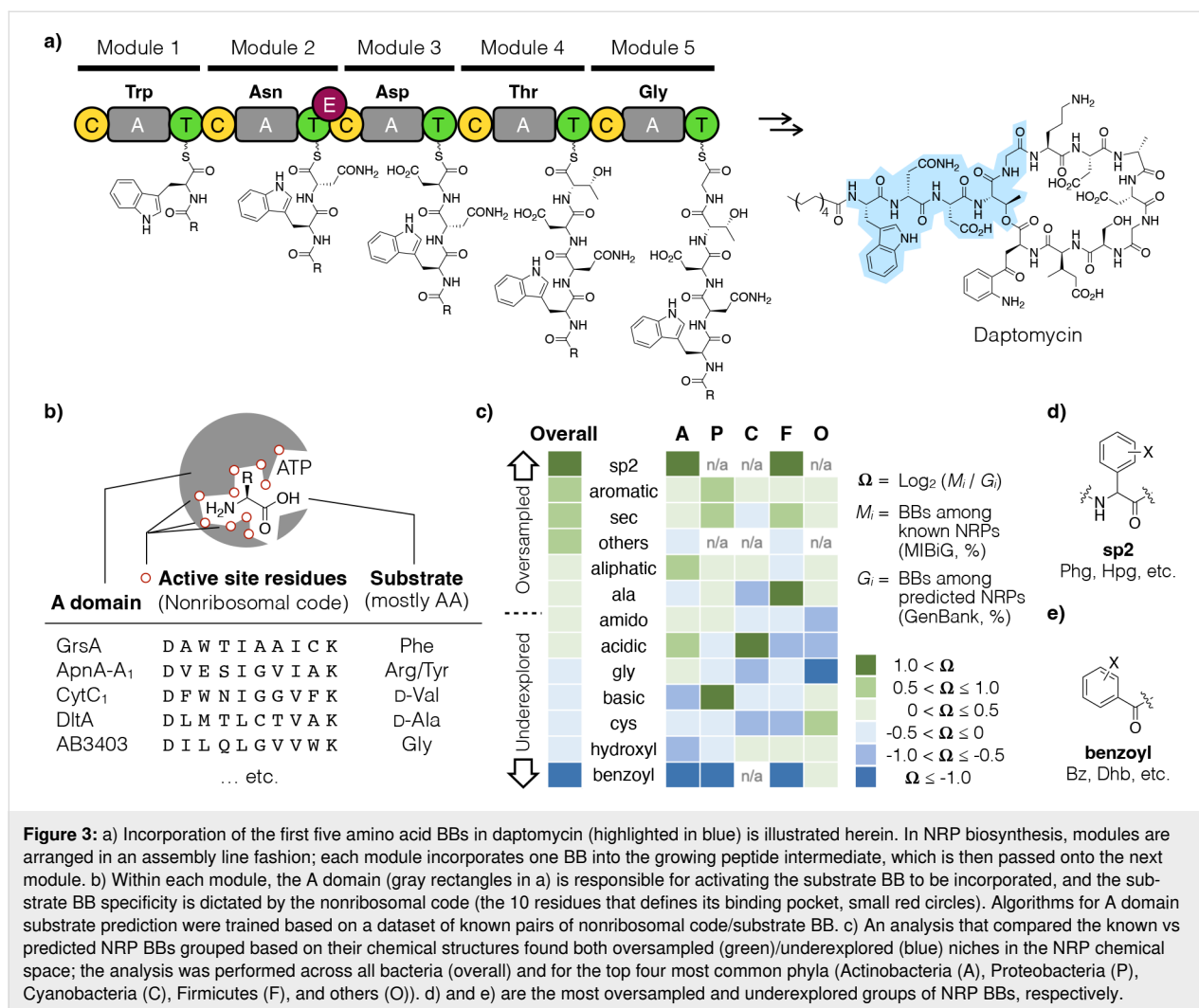
It is worth pointing out that chemical structure, but not bioactivity, is the unique descriptor of a molecule. Molecules with the same chemical structures are (by definition) identical; they must act on the same cellular target and display the same bioactivity. It is not necessarily the case the other way around for

bioactivity. However, bioactivity has long been used as a proxy to help find new natural products, whose chemical structure is typically determined after it has been isolated via the lengthy iterative purification process guided by bioactivity. Rediscovery is inevitable based on such a workflow, and this is essentially due to our inability to predict the chemical structure ahead of time. Does this have to be the case?

The instructions for natural product biosynthesis, along with all the information an organism needs, are encoded in its genome. Genetic information is transcribed and translated into proteins that carry out chemical reactions that sustain life, which include both primary metabolism and the biosynthesis of natural products. Direct protein sequencing by Edman degradation used to be standard practice in studying proteins but nowadays is rarely performed [39]. This is because the rules of transcription and translation are understood well enough for the sequence of a protein (its primary structure) to be predicted based on the corresponding nucleic acid sequence. Proteinaceous enzymes then go on to catalyze biosynthetic reactions that put together small molecule building blocks (BB) to generate natural products with extremely diverse chemical structures. Because the intricacy of this process is not fully understood, scientists still need to wait for enzymes to complete the entire course of biosynthesis, and then characterize chemical structure of the final natural product. While a generalized algorithm is not yet available, scientists in recent years have made some headways toward predicting the chemical structure of a natural product. Specifically, it is now possible to predict the order and identity of the BBs in NRPs based solely on the nucleic acid sequences of its BGC [40-49]. These algorithms not only obviated the need for culture and gene expression, dereplication can now be done in silico to avoid rediscovery. They are the cornerstone of chemical structure metagenomics; a few examples in this area of research are described below.

NRP biosynthesis and structure prediction

NRPs are biosynthesized by either type I or II nonribosomal peptide synthetase (NRPS) [50,51]. Type I NRPS is a megaenzyme machinery that contains multiple modules arranged in an assembly line fashion, each of which is responsible for incorporating a single BB into the growing peptide intermediate (Figure 3a). One module typically contains one adenylation (A) domain that folds and operates semi-autonomously, which recognizes and activates a specific substrate BB. Notably, the BBs are in most cases amino acids, wherein a much broader variety are used in NRP biosynthesis than the 20 canonical amino acids used in protein biosynthesis [52]. The modules are usually arranged co-linearly to the BGC sequence, which makes bioinformatic analysis much more straightforward, and for this reason, NRPs had been the target of early efforts aimed at

Figure 3: a) Incorporation of the first five amino acid BBs in daptomycin (highlighted in blue) is illustrated herein. In NRP biosynthesis, modules are arranged in an assembly line fashion; each module incorporates one BB into the growing peptide intermediate, which is then passed onto the next module. b) Within each module, the A domain (gray rectangles in a) is responsible for activating the substrate BB to be incorporated, and the substrate BB specificity is dictated by the nonribosomal code (the 10 residues that defines its binding pocket, small red circles). Algorithms for A domain substrate prediction were trained based on a dataset of known pairs of nonribosomal code/substrate BB. c) An analysis that compared the known vs predicted NRP BBs grouped based on their chemical structures found both oversampled (green)/underexplored (blue) niches in the NRP chemical space; the analysis was performed across all bacteria (overall) and for the top four most common phyla (Actinobacteria (A), Proteobacteria (P), Cyanobacteria (C), Firmicutes (F), and others (O)). d) and e) are the most oversampled and underexplored groups of NRP BBs, respectively.

predicting the chemical structure of natural products based on biosynthetic gene sequences.

In 1999, Stachelhaus et al. reported that the substrate specificity of an A domain, i.e., the identity of the amino acid BB it recognizes and activates, can be predicted based on its gene sequence alone [40]. Numerous prediction algorithms of this kind have been reported since [40-49]. The A domains are highly conserved in terms of both structure and sequence [41,53-57]. The exception is the 10 residues that constitute the A domain active site, whose high variability creates binding pockets of varying shapes and sizes. These residues therefore dictate substrate BB specificity of an A domain and are referred to as the nonribosomal code (in analogy to the genetic code) [58]. Hundreds of known nonribosomal codes and their corresponding BBs can be extracted from natural products that have been characterized over the past several decades, generating a dataset to train NRP prediction algorithms (Figure 3b) [49,59,60]. A software suite called antibiotics and secondary metabolite anal-

ysis shell (antiSMASH) further automated all of the following steps: take (meta-)genomic sequences as the input, identify BGCs for NRPs (and other natural products as well), parse out modules and domains, and finally outputs the order and identity of BBs of a predicted NRP. AntiSMASH is freely available for the research community worldwide [61].

Underexplored and oversampled NRP building blocks

With these prediction tools at hand, it became possible to survey the biosynthetic diversity of NRPs from a new perspective. Specifically, Jian et al. compiled data in GenBank into a custom microbial genome collection, termed GB1 [37]. Very little is known about most GB1 microorganisms aside from their genome sequences, much less their potential ability to produce natural products. On the other hand, MIBiG is the most comprehensive collection of known natural products whose BGCs have been sequenced and annotated [59]. As mentioned above, even though BGF found the vast majority of natural

products we know thus far, they collectively account for a meager ≈1% of the chemical space. It begs the question of whether BGF has been sampling evenly or was there any bias. In contrast to BGF only being amenable to actively expressed BGCs in readily cultured microorganisms, there are very few pre-requisites to DNA sequencing, and therefore GB1 represents a (nearly) even sampling of the universe of microbial biosynthetic diversity.

Jian et al. applied A domain prediction algorithms to genome sequences in MIBiG and GB1. They estimated the relative abundance of predicted BBs among known NRPs (MIBiG) versus the predicted NRPs (GB1). The Ω parameter they presented was based on the log2 scale, wherein Ω = +1 and −1 means BGF-based natural product research has oversampled and underexplored by two-fold a particular BB (or a group of BBs with similar chemical structures), respectively (Figure 3c). Phenylglycine and its derivatives (the **sp2** group of BBs, Figure 3d) turned out to be the most oversampled group of BBs. This group of BBs included noncanonical amino acids characteristic of the glycopeptide antibiotic family [62,63]. Their aromatic rings undergo oxidative coupling to form biaryl moieties that restrict atropisomerism; the resulting rigid structure is key to their high affinity binding to peptidoglycan intermediates [64,65]. Glycopeptide antibiotics include vancomycin, teicoplanin, ramoplanin, etc., and were once an intense research focus in both academia and industry [66]. Such a historical backdrop may explain the apparent oversampling of the **sp2** BBs. In contrast, hydroxylated benzoic acids turned out to be the most underexplored group of BBs (**benzoyl**, Figure 3e). In NRPs, they serve as ligands for chelating iron in siderophores. Because ferric cations (Fe^{3+}) mostly exist as insoluble solids in the Earth crust, most microorganisms produce and secrete siderophores to scavenge this scarce resource from their surroundings [67]. Whereas siderophores do not always contain hydroxylated benzoyl BBs, all known NRPs that do contain the benzoyl BB are siderophores. One plausible interpretation for this group of BBs being the most underexplored is that there are as yet unknown benzoyl BB containing NRPs that play other biological roles; we may have thus far completely missed them. Together, this chemical structure metagenomic analysis showed that BGF has not examined the natural product chemical space evenly, as there are niches that have been examined more frequently than random sampling, and there appear to be many stones left unturned as well.

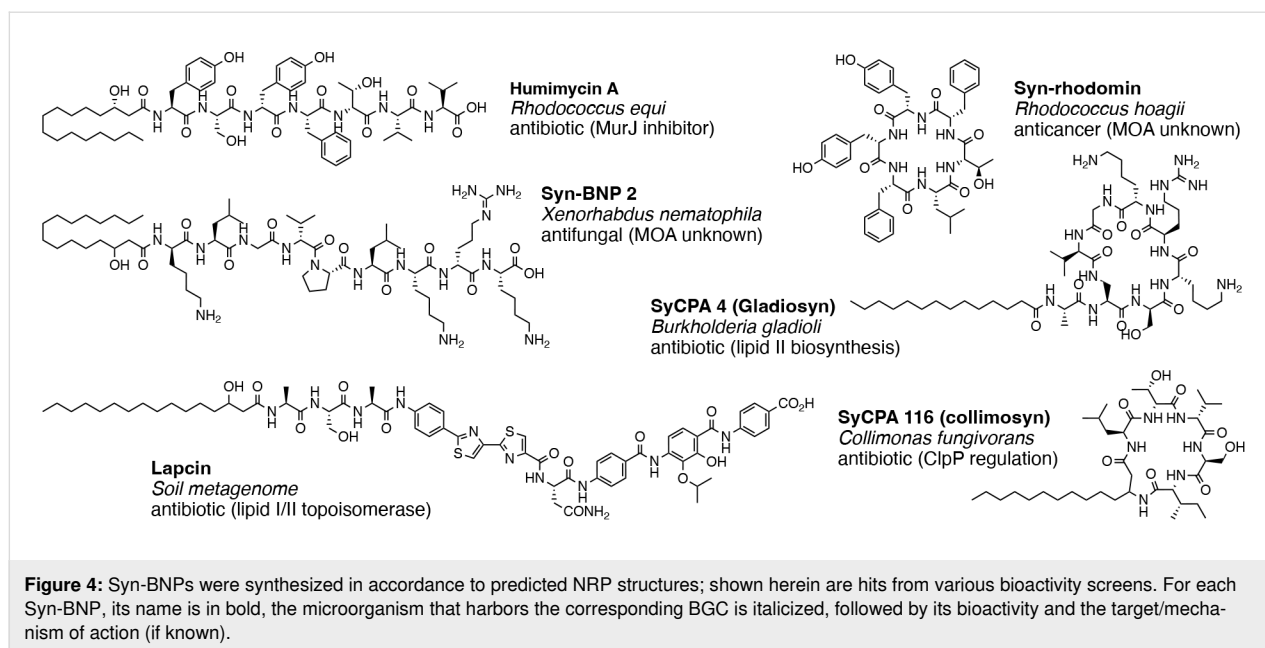## Converting predicted NRPs into real molecules

Aside from an analysis of the BB usage pattern, real NRP-like molecules can be constructed in accordance to the predicted identity and order of the BBs. Brady and co-workers used solid-phase peptide synthesis to convert the predicted NRPs from virtual into reality; they called these molecules synthetic-bioinformatic natural products (Syn-BNPs) [29-36]. By excluding Syn-NRPs that resemble known NRPs, they were able to focus on chemical spaces that have not been explored yet. Because many NRPs are produced and isolated as a collection of structurally similar congeners, it was believed that this approach shall be viable as long as a Syn-BNP bears enough resemblance to the natural counterparts to recapitulate their biological functions. More than 500 Syn-BNP NRPs were designed and synthesized based on the analysis of bacterial genome and metagenome sequences. These molecules were screened for various bioactivities and led to the discovery of new antibiotics, antifungals, as well as anticancer compounds (Figure 4).

Importantly, the underlying mechanism of bacterial growth suppression has been identified for several Syn-BNP antibiotics, which includes both general mode of action (MOA) (e.g., membrane lysis and depolarization) [30,34] and specific MOA (e.g., dysregulation of ClpP protease [33], inhibition of topoisomerase I/II [36,68], blocking lipid II transport by flippase [29], sequestration of cell wall biosynthetic intermediate C55-(di)phosphate, etc.) [35]. It is unlikely that a Syn-BNP is able to target a specific protein or pathway, unless it in fact recapitulated key structural feature(s) of a NRP. These observations are a testament to the feasibility of this approach. This approach has also been applied to focus on NRPs with particular physical properties. Specifically, Qian and co-workers examined 7395 bacterial genomes and identified more than ten thousand potential cationic NRPs. They focused on a few promising candidates after bioinformatic-based dereplication and found two NRPs, brevicidine and laterocidine, that worked in an animal thigh infection model [69].

Future directions 1: Improve A domain prediction algorithms

NRP prediction algorithms are central to chemical structure metagenomics. Several mechanistically novel Syn-BNP antibiotics were discovered, and a systematic BB usage pattern analysis provided insights into oversampled/underexplored niches in NRP chemical space (see above). These success stories suggest that the existing algorithms for A domain substrate prediction were reasonably accurate. However, only about 6 out of 10 A domains were amenable to the current prediction algorithms. Jian et al. reported that ≈40% of A domains failed to match a nonribosomal code in the algorithm training dataset and were deemed "unpredictable" [37]. It is also possible that the A domain in question aligned so poorly to prototypical A domains that prevented the proper identification of the nonribosomal code itself [70]. Regardless of the scenario, these bioinformatically intractable A domains are distinct from known ones and

Figure 4: Syn-BNPs were synthesized in accordance to predicted NRP structures; shown herein are hits from various bioactivity screens. For each Syn-BNP, its name is in bold, the microorganism that harbors the corresponding BGC is italicized, followed by its bioactivity and the target/mechanism of action (if known).

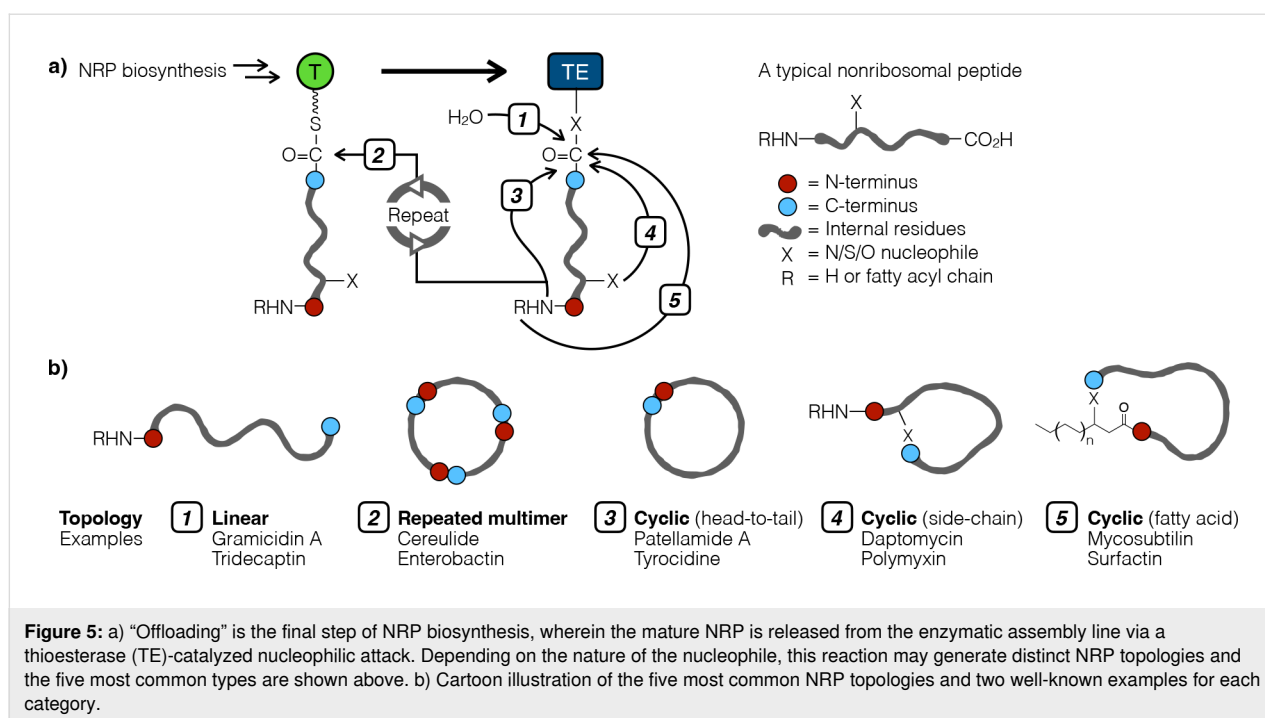point to enormous biosynthetic novelty that still awaits our exploration.

Compiling a dataset for training A domain substrate prediction algorithms has never been the objective for natural product research in the past. The current dataset has been the byproduct of cumulative NRP discovery, and its rate of expansion has been disappointingly slow. This is because new NRPs do not necessarily contain new A domains or new nonribosomal codes, so that the effective size of the training dataset does not always benefit from the discovery of a new NRP. Conversely, every "unpredictable" A domain, if its substrate specificity were to be experimentally determined, is guaranteed to be a new datapoint. In fact, various in vitro substrate characterization assays that studied A domains as stand-alone enzymes have been reported [53,57,71-77]. Investigating strategically these bioinformatically intractable A domains is a much more efficient way to acquire new datapoints, and the performance of A domain prediction algorithms should improve simply after re-training on an expanded dataset. These algorithms may also benefit from bioinformatic studies of adenylating enzymes in general; the discovery of a novel β-lactone biosynthesis pathway in *Nocardia* species is a good case in point [78].

Future directions 2: Understanding thioesterase function

No algorithm is currently capable of predicting the topology of an NRP despite the fact that this feature is known to be important for bioactivity [79,80]. Typically, the C-terminus of the NRP intermediate is covalently linked via a thioester bond to the phosphopantetheine prosthetic arm of the peptide carrier protein (also known as the thiolation (T) domain) throughout biosynthesis (Figure 5a). The NRP intermediate is passed from one module to the next as BBs are incorporated one at a time into this growing peptide chain. The very last step in NRP biosynthesis – offloading the final product from the enzymatic assembly line – is usually catalyzed by the thioesterase (TE) domain and determines the topology of the final NRP product. The TE may release the NRP as a linear peptide or cyclic peptide, and the latter further manifests many possible topologies (Figure 5b). In a nutshell, offloading is a TE-catalyzed nucleophilic attack to release an NRP from the megaenzyme machinery. Water can act as the nucleophile during the offloading step, which amounts to hydrolysis and results in a linear peptide. On the other hand, a cyclic peptide forms when an intramolecular functional group acts as the nucleophile in this step. Regardless of the resulting topology, the NRP offloading step always entails the same chemical reaction, wherein nucleophilic attack is promoted by the catalytic triad of a TE via general base catalysis. This is likely why traditional mechanistic studies that focused on the enzyme active site failed to work out how TEs control NRP topology.

A priori, two factors should be important in the offloading step. First, the biosynthetic command is undoubtedly encoded somewhere within the TE protein sequence. The second factor, one that has been overlooked, is the chemical structure of the NRP. Both of these factors may need to be included in an analysis to fully understand the determinant(s) of NRP topology. There are many ways to classify NRPs when topology and chemical structural features are considered simultaneously. For example, a macrocycle may be a lactam or a lactone depending on whether

Figure 5: a) "Offloading" is the final step of NRP biosynthesis, wherein the mature NRP is released from the enzymatic assembly line via a thioesterase (TE)-catalyzed nucleophilic attack. Depending on the nature of the nucleophile, this reaction may generate distinct NRP topologies and the five most common types are shown above. b) Cartoon illustration of the five most common NRP topologies and two well-known examples for each category.

the internal nucleophile is an amine or an alcohol, respectively. Based on the position of the nucleophile, a NRP can be cyclized head-to-tail, via an amino acid side-chain, a nucleophilic heteroatom on the N-terminal fatty acyl chain, or as a multimer of repeating sub-structures (Figure 5b). The ring size, ratio of L- and D-amino acids, etc. may also be the basis of classification. TE sequence alignment guided by NRPs grouped based on one or more of the above topological features may yield key insights.

Future directions 3: Chemical structure metagenomics of type I polyketides

The two largest natural product families are NRP and polyketides (PKs) [81]. Type I PKs are also constructed in a modular assembly line biosynthetic logic and may be amenable to chemical structure metagenomic approaches described herein [82]. In fact, the software suite antiSMASH can already predict the substrate specificity of individual polyketide synthase (PKS) modules, i.e., the type of alkylmalonate it incorporates into a growing PK intermediate [61]. Furthermore, Xiang et al. reported recently that the stereochemistry of each new chiral center resulting from each alkylmalonate BB incorporation can be predicted based on the corresponding ketoreductase domain sequence [83]. Last but not least, predicted polyketide structures can be synthesized in a streamlined and automated fashion by using a microfluidic device reported by Burke and co-workers [84,85]. As such, the necessary computational and synthetic tools are all in place to examine the PK chemical space and to support a Syn-BNP based PK discovery campaign.

Future directions 4: Natural products with more complex biosynthetic logic

Type I NRPs and PKs, as well as many sub-families of ribosomally synthesized and post-translationally modified peptides (RiPPs), are readily amenable to chemical structure metagenomics studies, because translating their modular biosynthetic logic into a chemical synthetic scheme is rather straightforward [86]. Aside from constructing the core molecular scaffold of these natural products, further modifications may be installed either before or after the assembly of the amino acid or alkylmalonate BBs. While predicting most modifications de novo is not yet built into existing algorithms, it is often possible to work them out based on chemical structure context, background knowledge, and educated guesses. For example, cyclodehydration, a feature frequently seen in both NRPs and RiPPs [87], requires the presence of a nucleophile at the β-position of the amino acid and occurs exclusively on select amino acids. Specifically, cyclodehydration of a serine or threonine (followed by oxidation or reduction) generates an oxazole, oxazoline, or oxazolidine moiety in the NRP backbone, and the analogous thiazo moieties come from cysteines. The same requirement applies to the formation of dehydroalanine and dehydrobutyrine moieties. These chemical principles hold true even though the tailoring enzymes for these modifications in NRPs and RiPPs are nonhomologous [88,89].

Some types of modifications are biased towards (or against) certain amino acids; while these trends are statistically valid,

whether there is an underlying chemical principle that governs the observed selectivity remains unclear [52]. For example, tailoring enzymes for β-hydroxylation most often act on aspartate and asparagine. In contrast, despite being one of the most common BBs in NRPs, no ornithine is β-hydroxylated to the best of our knowledge.

Last but not least, natural product research has forayed into the use of artificial intelligence (AI) tools. For example, Magarvey and co-workers used machine learning to help improve bioinformatic analysis. Their updated PRISM 4 pipeline outperformed antiSMASH5 and was able to predict the chemical structures of a wide variety of secondary metabolites [90]. Notably, the repertoire of PRISM 4 spans beyond NRPs and PKs and includes alkaloids, terpenoids, aminoglycosides, nucleosides, etc. Structure prediction of natural products other than NRPs and PKs has historically progressed slower. However, it was not due to a lack of data. Indeed, terpenoids and alkaloids are the two largest families of natural products in plants and also represent a sizable minority in microorganisms. The lagging progress is in fact due to our inability to interpret the corresponding genetically encoded biosynthetic instructions. While scientists will undoubtedly find more applications of AI, its strongest suit is arguably to help humans find hidden patterns in bulk data, such as the many bits and pieces of information that are available for terpenoids and alkaloids biosynthesis. Data standardization (albeit a tedious and labor-intensive task) may be the final roadblock between AI and generalizable chemical structure metagenomics [91].

Conclusion and Outlook
Microorganisms produce specialized metabolites to communicate with each other and to interact with the environment. The chemical structure is the unique descriptor of these molecules and dictates the way these molecules behave and interact. However, the research and discovery of natural products have historically been guided by bioactivity and not chemical structure. We argue that, whether it is for the purpose of finding new lead compounds for drug development or gaining a deeper understanding in life science, the field of natural product research may benefit from placing chemical structure front and center. A few examples are described herein to illustrate the power of this viewpoint. It complements existing approaches to facilitate a broader and more efficient survey of the vast natural product chemical space that awaits our exploration.

Funding

Conflict of Interest
None to declare.

Author Contributions
Thomas Ma: writing – original draft; writing – review & editing. John Chu: conceptualization; supervision; writing – original draft; writing – review & editing.

ORCID® iDs
Thomas Ma - https://orcid.org/0009-0004-7298-3859
John Chu - https://orcid.org/0000-0002-7033-7229

Data Availability Statement
Data sharing is not applicable as no new data was generated or analyzed in this study.

References
1. Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. *Nat. Prod. Res.* **2018,** *32,* 1926–1950. doi:10.1080/14786419.2017.1356838
2. Yuan, H.; Ma, Q.; Ye, L.; Piao, G. *Molecules* **2016,** *21,* 559. doi:10.3390/molecules21050559
3. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2020,** *83,* 770–803. doi:10.1021/acs.jnatprod.9b01285
4. Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; The International Natural Product Sciences Taskforce; Supuran, C. T. *Nat. Rev. Drug Discovery* **2021,** *20,* 200–216. doi:10.1038/s41573-020-00114-z
5. Carlson, E. E. *ACS Chem. Biol.* **2010,** *5,* 639–653. doi:10.1021/cb100105c
6. Bandaranayake, W. M.; Banfield, J. E.; Black, D. S. C. *J. Chem. Soc., Chem. Commun.* **1980,** 902–903. doi:10.1039/c39800000902
7. Bentley, R. *J. Chem. Educ.* **2004,** *81,* 1462. doi:10.1021/ed081p1462
8. Mak, J. Y. W.; Pouwer, R. H.; Williams, C. M. *Angew. Chem., Int. Ed.* **2014,** *53,* 13664–13688. doi:10.1002/anie.201400932
9. "Wooddward molecules: Reserpine, lysergic acid, cephalosporin C a colchicine" American Chemical Society. https://www.acs.org/molecule-of-the-week/archive/w/woodward-molecules.html (accessed July 10, 2024).
10. Bennett, J. W.; Chung, K.-T. *Adv. Appl. Microbiol.* **2001,** *49,* 163–184. doi:10.1016/s0065-2164(01)49013-7
11. Kresge, N.; Simoni, R. D.; Hill, R. L. *J. Biol. Chem.* **2004,** *279,* e7–e8. doi:10.1016/s0021-9258(20)67861-9
12. Milshteyn, A.; Schneider, J. S.; Brady, S. F. *Chem. Biol.* **2014,** *21,* 1211–1223. doi:10.1016/j.chembiol.2014.08.006
13. Lewis, K. *Cell* **2020,** *181,* 29–45. doi:10.1016/j.cell.2020.02.056
14. Amann, J. *Zentralbl. Bakteriol., Parasitenkd., Infektionskrankh. Hyg., Abt. 1 Orig.* **1911,** *29,* 381–384.
15. Muller, P. T. *Arch. Hyg.* **1912,** *75,* 189–223.
16. Staley, J. T.; Konopka, A. *Annu. Rev. Microbiol.* **1985,** *39,* 321–346. doi:10.1146/annurev.mi.39.100185.001541
17. Charlop-Powers, Z.; Owen, J. G.; Reddy, B. V. B.; Ternei, M. A.; Guimarães, D. O.; de Frias, U. A.; Pupo, M. T.; Seepe, P.; Feng, Z.; Brady, S. F. *eLife* **2015,** *4,* 10.7554/eLife.05048. doi:10.7554/elife.05048

18. Xu, F.; Nazari, B.; Moon, K.; Bushin, L. B.; Seyedsayamdost, M. R. *J. Am. Chem. Soc.* **2017,** *139,* 9203–9212. doi:10.1021/jacs.7b02716

19. Hoskisson, P. A.; Seipke, R. F. *mBio* **2020,** *11,* e02642–20. doi:10.1128/mbio.02642-20

20. Medema, M. H.; de Rond, T.; Moore, B. S. *Nat. Rev. Genet.* **2021,** *22,* 553–571. doi:10.1038/s41576-021-00363-7

21. Mungan, M. D.; Alanjary, M.; Blin, K.; Weber, T.; Medema, M. H.; Ziemert, N. *Nucleic Acids Res.* **2020,** *48,* W546–W552. doi:10.1093/nar/gkaa374

22. Hobson, C.; Chan, A. N.; Wright, G. D. *Chem. Rev.* **2021,** *121,* 3464–3494. doi:10.1021/acs.chemrev.0c01214

23. Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998,** *5,* R245–R249. doi:10.1016/s1074-5521(98)90108-9

24. Brady, S. F. *Nat. Protoc.* **2007,** *2,* 1297–1305. doi:10.1038/nprot.2007.195

25. Craig, J. W.; Chang, F.-Y.; Kim, J. H.; Obiajulu, S. C.; Brady, S. F. *Appl. Environ. Microbiol.* **2010,** *76,* 1633–1641. doi:10.1128/aem.02169-09

26. Charlop-Powers, Z.; Milshteyn, A.; Brady, S. F. *Curr. Opin. Microbiol.* **2014,** *19,* 70–75. doi:10.1016/j.mib.2014.05.021

27. Rosenzweig, A. F.; Burian, J.; Brady, S. F. *Curr. Opin. Microbiol.* **2023,** *75,* 102335. doi:10.1016/j.mib.2023.102335

28. Hemmerling, F.; Piel, J. *Nat. Rev. Drug Discovery* **2022,** *21,* 359–378. doi:10.1038/s41573-022-00414-6

29. Chu, J.; Vila-Farres, X.; Inoyama, D.; Ternei, M.; Cohen, L. J.; Gordon, E. A.; Reddy, B. V. B.; Charlop-Powers, Z.; Zebroski, H. A.; Gallardo-Macias, R.; Jaskowski, M.; Satish, S.; Park, S.; Perlin, D. S.; Freundlich, J. S.; Brady, S. F. *Nat. Chem. Biol.* **2016,** *12,* 1004–1006. doi:10.1038/nchembio.2207

30. Vila-Farres, X.; Chu, J.; Inoyama, D.; Ternei, M. A.; Lemetre, C.; Cohen, L. J.; Cho, W.; Reddy, B. V. B.; Zebroski, H. A.; Freundlich, J. S.; Perlin, D. S.; Brady, S. F. *J. Am. Chem. Soc.* **2017,** *139,* 1404–1407. doi:10.1021/jacs.6b11861

31. Chu, J.; Vila-Farres, X.; Inoyama, D.; Gallardo-Macias, R.; Jaskowski, M.; Satish, S.; Freundlich, J. S.; Brady, S. F. *ACS Infect. Dis.* **2018,** *4,* 33–38. doi:10.1021/acsinfecdis.7b00056

32. Chu, J.; Vila-Farres, X.; Brady, S. F. *J. Am. Chem. Soc.* **2019,** *141,* 15737–15741. doi:10.1021/jacs.9b07317

33. Chu, J.; Koirala, B.; Forelli, N.; Vila-Farres, X.; Ternei, M. A.; Ali, T.; Colosimo, D. A.; Brady, S. F. *J. Am. Chem. Soc.* **2020,** *142,* 14158–14168. doi:10.1021/jacs.0c04376

34. Wang, Z.; Koirala, B.; Hernandez, Y.; Zimmerman, M.; Park, S.; Perlin, D. S.; Brady, S. F. *Nature* **2022,** *601,* 606–611. doi:10.1038/s41586-021-04264-x

35. Wang, Z.; Koirala, B.; Hernandez, Y.; Zimmerman, M.; Brady, S. F. *Science* **2022,** *376,* 991–996. doi:10.1126/science.abn4213

36. Wang, Z.; Forelli, N.; Hernandez, Y.; Ternei, M.; Brady, S. F. *Nat. Commun.* **2022,** *13,* 842. doi:10.1038/s41467-022-28292-x

37. Jian, B.-S.; Chiou, S.-L.; Hsu, C.-C.; Ho, J.; Wu, Y.-W.; Chu, J. *ACS Chem. Biol.* **2023,** *18,* 476–483. doi:10.1021/acschembio.2c00761

38. Yim, G.; Huimi Wang, H.; Davies, J. *Philos. Trans. R. Soc., B* **2007,** *362,* 1195–1200. doi:10.1098/rstb.2007.2044

39. Edman, P.; Begg, G. A Protein Sequenator. *Eur. J. Biochem.;* Springer: Berlin, Heidelberg, 1967; pp 80–91. doi:10.1007/978-3-662-25813-2_14

40. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999,** *6,* 493–505. doi:10.1016/s1074-5521(99)80082-9

41. Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000,** *7,* 211–224. doi:10.1016/s1074-5521(00)00091-0

42. Minowa, Y.; Araki, M.; Kanehisa, M. *J. Mol. Biol.* **2007,** *368,* 1500–1517. doi:10.1016/j.jmb.2007.02.099

43. Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005,** *33,* 5799–5808. doi:10.1093/nar/gki885

44. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011,** *39* (Suppl. 2), W362–W367. doi:10.1093/nar/gkr323

45. Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C. *PLoS One* **2013,** *8,* e62136. doi:10.1371/journal.pone.0062136

46. Baranašić, D.; Zucko, J.; Diminic, J.; Gacesa, R.; Long, P. F.; Cullum, J.; Hranueli, D.; Starcevic, A. *J. Ind. Microbiol. Biotechnol.* **2014,** *41,* 461–467. doi:10.1007/s10295-013-1322-2

47. Lee, T. V.; Johnson, R. D.; Arcus, V. L.; Lott, J. S. *Proteins: Struct., Funct., Bioinf.* **2015,** *83,* 2052–2066. doi:10.1002/prot.24922

48. Knudsen, M.; Søndergaard, D.; Tofting-Olesen, C.; Hansen, F. T.; Brodersen, D. E.; Pedersen, C. N. S. *Bioinformatics* **2016,** *32,* 325–329. doi:10.1093/bioinformatics/btv600

49. Chevrette, M. G.; Aicheler, F.; Kohlbacher, O.; Currie, C. R.; Medema, M. H. *Bioinformatics* **2017,** *33,* 3202–3210. doi:10.1093/bioinformatics/btx400

50. Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. *Nat. Prod. Rep.* **2020,** *37,* 355–379. doi:10.1039/c9np00047j

51. Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006,** *106,* 3468–3496. doi:10.1021/cr0503097

52. Walsh, C. T.; O'Brien, R. V.; Khosla, C. *Angew. Chem., Int. Ed.* **2013,** *52,* 7098–7124. doi:10.1002/anie.201208344

53. Lee, T. V.; Johnson, L. J.; Johnson, R. D.; Koulman, A.; Lane, G. A.; Lott, J. S.; Arcus, V. L. *J. Biol. Chem.* **2010,** *285,* 2415–2427. doi:10.1074/jbc.m109.071324

54. Sundlov, J. A.; Shi, C.; Wilson, D. J.; Aldrich, C. C.; Gulick, A. M. *Chem. Biol.* **2012,** *19,* 188–198. doi:10.1016/j.chembiol.2011.11.013

55. Kudo, F.; Miyanaga, A.; Eguchi, T. *J. Ind. Microbiol. Biotechnol.* **2019,** *46,* 515–536. doi:10.1007/s10295-018-2084-7

56. Tan, K.; Zhou, M.; Jedrzejczak, R. P.; Wu, R.; Higuera, R. A.; Borek, D.; Babnigg, G.; Joachimiak, A. *Curr. Res. Struct. Biol.* **2020,** *2,* 14–24. doi:10.1016/j.crstbi.2020.01.002

57. Chen, I.-H.; Cheng, T.; Wang, Y.-L.; Huang, S.-J.; Hsiao, Y.-H.; Lai, Y.-T.; Toh, S.-I.; Chu, J.; Rudolf, J. D.; Chang, C.-Y. *ChemBioChem* **2022,** *23,* e202200563. doi:10.1002/cbic.202200563

58. von Döhren, H.; Dieckmann, R.; Pavela-Vrancic, M. *Chem. Biol.* **1999,** *6,* R273–R279. doi:10.1016/s1074-5521(00)80014-9

59. Kautsar, S. A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J. C.; Terlouw, B. R.; van der Hooft, J. J. J.; van Santen, J. A.; Tracanna, V.; Suarez Duran, H. G.; Pascal Andreu, V.; Selem-Mojica, N.; Alanjary, M.; Robinson, S. L.; Lund, G.; Epstein, S. C.; Sisto, A. C.; Charkoudian, L. K.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2020,** *48,* D454–D458. doi:10.1093/nar/gkz882

60. Flissi, A.; Ricart, E.; Campart, C.; Chevalier, M.; Dufresne, Y.; Michalik, J.; Jacques, P.; Flahaut, C.; Lisacek, F.; Leclère, V.; Pupin, M. *Nucleic Acids Res.* **2020,** *48,* D465–D469. doi:10.1093/nar/gkz1000

61. Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B. R.; Metcalf, W. W.; Helfrich, E. J. N.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2023,** *51,* W46–W50. doi:10.1093/nar/gkad344

62. James, R. C.; Pierce, J. G.; Okano, A.; Xie, J.; Boger, D. L. *ACS Chem. Biol.* **2012,** *7,* 797–804. doi:10.1021/cb300007j

63. Sarkar, P.; Yarlagadda, V.; Ghosh, C.; Haldar, J. *MedChemComm* **2017,** *8,* 516–533. doi:10.1039/c6md00585c

64. Smyth, J. E.; Butler, N. M.; Keller, P. A. *Nat. Prod. Rep.* **2015,** *32,* 1562–1583. doi:10.1039/c4np00121d

65. Wang, Z.; Meng, L.; Liu, X.; Zhang, L.; Yu, Z.; Wu, G. *Eur. J. Med. Chem.* **2022,** *243,* 114700. doi:10.1016/j.ejmech.2022.114700

66. Butler, M. S.; Hansford, K. A.; Blaskovich, M. A. T.; Halai, R.; Cooper, M. A. *J. Antibiot.* **2014,** *67,* 631–644. doi:10.1038/ja.2014.111

67. Kramer, J.; Özkaya, Ö.; Kümmerli, R. *Nat. Rev. Microbiol.* **2020,** *18,* 152–163. doi:10.1038/s41579-019-0284-4

68. Wang, Z.; Kasper, A.; Takahashi, M.; Morales Amador, A.; Bhattacharjee, A.; Kan, J.; Hernandez, Y.; Ternei, M.; Brady, S. F. *Angew. Chem., Int. Ed.* **2024,** *63,* e202317187. doi:10.1002/anie.202317187

69. Li, Y.-X.; Zhong, Z.; Zhang, W.-P.; Qian, P.-Y. *Nat. Commun.* **2018,** *9,* 3273. doi:10.1038/s41467-018-05781-6

70. Miyanaga, A.; Kudo, F.; Eguchi, T. *Curr. Opin. Chem. Biol.* **2022,** *71,* 102212. doi:10.1016/j.cbpa.2022.102212

71. Wilson, D. J.; Aldrich, C. C. *Anal. Biochem.* **2010,** *404,* 56–63. doi:10.1016/j.ab.2010.04.033

72. Xia, S.; Ma, Y.; Zhang, W.; Yang, Y.; Wu, S.; Zhu, M.; Deng, L.; Li, B.; Liu, Z.; Qi, C. *PLoS One* **2012,** *7,* e37487. doi:10.1371/journal.pone.0037487

73. Hara, R.; Suzuki, R.; Kino, K. *Anal. Biochem.* **2015,** *477,* 89–91. doi:10.1016/j.ab.2015.01.006

74. Kasai, S.; Konno, S.; Ishikawa, F.; Kakeya, H. *Chem. Commun.* **2015,** *51,* 15764–15767. doi:10.1039/c5cc04953a

75. Duckworth, B. P.; Wilson, D. J.; Aldrich, C. C. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2016,** 53–61. doi:10.1007/978-1-4939-3375-4_3

76. Kittilä, T.; Schoppet, M.; Cryle, M. J. *ChemBioChem* **2016,** *17,* 576–584. doi:10.1002/cbic.201500555

77. Lundy, T. A.; Mori, S.; Thamban Chandrika, N.; Garneau-Tsodikova, S. *ACS Chem. Biol.* **2020,** *15,* 282–289. doi:10.1021/acschembio.9b00929

78. Robinson, S. L.; Terlouw, B. R.; Smith, M. D.; Pidot, S. J.; Stinear, T. P.; Medema, M. H.; Wackett, L. P. *J. Biol. Chem.* **2020,** *295,* 14826–14839. doi:10.1074/jbc.ra120.013528

79. Lange, A.; Sun, H.; Pilger, J.; Reinscheid, U. M.; Gross, H. *ChemBioChem* **2012,** *13,* 2671–2675. doi:10.1002/cbic.201200532

80. Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016,** *33,* 183–202. doi:10.1039/c4np00148f

81. Blin, K.; Medema, M. H.; Kottmann, R.; Lee, S. Y.; Weber, T. *Nucleic Acids Res.* **2017,** *45,* D555–D559. doi:10.1093/nar/gkw960

82. Grininger, M. *Nat. Chem. Biol.* **2023,** *19,* 401–415. doi:10.1038/s41589-023-01277-7

83. Xiang, C.; Yao, S.; Wang, R.; Zhang, L. *Beilstein J. Org. Chem.* **2024,** *20,* 1476–1485. doi:10.3762/bjoc.20.131

84. Woerly, E. M.; Roy, J.; Burke, M. D. *Nat. Chem.* **2014,** *6,* 484–491. doi:10.1038/nchem.1947

85. Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M. E.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. *Science* **2015,** *347,* 1221–1226. doi:10.1126/science.aaa5414

86. Ortega, M. A.; van der Donk, W. A. *Cell Chem. Biol.* **2016,** *23,* 31–44. doi:10.1016/j.chembiol.2015.11.012

87. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013,** *30,* 108–160. doi:10.1039/c2np20085f

88. McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009,** *26,* 537–559. doi:10.1039/b714132g

89. Cheng, Z.; He, B.-B.; Lei, K.; Gao, Y.; Shi, Y.; Zhong, Z.; Liu, H.; Liu, R.; Zhang, H.; Wu, S.; Zhang, W.; Tang, X.; Li, Y.-X. *Nat. Commun.* **2024,** *15,* 4901. doi:10.1038/s41467-024-49215-y

90. Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A. *Nat. Commun.* **2020,** *11,* 6058. doi:10.1038/s41467-020-19986-1

91. Mullowney, M. W.; Duncan, K. R.; Elsayed, S. S.; Garg, N.; van der Hooft, J. J. J.; Martin, N. I.; Meijer, D.; Terlouw, B. R.; Biermann, F.; Blin, K.; Durairaj, J.; Gorostiola González, M.; Helfrich, E. J. N.; Huber, F.; Leopold-Messer, S.; Rajan, K.; de Rond, T.; van Santen, J. A.; Sorokina, M.; Balunas, M. J.; Beniddir, M. A.; van Bergeijk, D. A.; Carroll, L. M.; Clark, C. M.; Clevert, D.-A.; Dejong, C. A.; Du, C.; Ferrinho, S.; Grisoni, F.; Hofstetter, A.; Jespers, W.; Kalinina, O. V.; Kautsar, S. A.; Kim, H.; Leao, T. F.; Masschelein, J.; Rees, E. R.; Reher, R.; Reker, D.; Schwaller, P.; Segler, M.; Skinnider, M. A.; Walker, A. S.; Willighagen, E. L.; Zdrazil, B.; Ziemert, N.; Goss, R. J. M.; Guyomard, P.; Volkamer, A.; Gerwick, W. H.; Kim, H. U.; Müller, R.; van Wezel, G. P.; van Westen, G. J. P.; Hirsch, A. K. H.; Linington, R. G.; Robinson, S. L.; Medema, M. H. *Nat. Rev. Drug Discovery* **2023,** *22,* 895–916. doi:10.1038/s41573-023-00774-7