Cheminformatics-Guided Cell-Free Exploration of Peptide Natural Products

Jarrett M. Pelton[#,1], **Joshua E. Hochuli**[#,2], **Patric W. Sadecki**[3], **Takayuki Katoh**[5], **Hiroaki Suga**[5], **Leslie M. Hicks**[3,4], **Eugene N. Muratov**[1], **Alexander Tropsha**[1,*], **Albert A. Bowers**[1,3,4,*]

[1]Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[2]Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[3]Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[4]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC, 27599, USA.

[5]Department of Chemistry, Graduate School of Science, The University of Tokyo, Tokyo, 113-0033, Japan.

Abstract

There have been significant advances in the flexibility and power of *in vitro* cell-free translation systems. The increasing ability to incorporate non-canonical amino acids and complement

*Corresponding Authors: Authors to whom correspondence should be addressed: **Albert A. Bowers** - Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA; Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA; abower2@email.unc.edu, **Alexander Tropsha** - Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA; tropsha@email.unc.edu. Author to whom correspondence should be addressed: alex_tropsha@unc.edu.
#J.M.P. and J.E.H. contributed equally.

Jarrett M. Pelton - Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA;

Joshua E. Hochuli - Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA; Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, North Carolina, 27599, USA

Patric W. Sadecki - Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA;

Takayuki Katoh - Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Hiroaki Suga - Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Leslie M. Hicks - Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA; Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA;

Eugene N. Muratov - Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA;

The supporting information is available free of charge at hyperlink here Experimental methods and supporting figures (SI PDF) Supplemental Tables S1–7 (240103_CFB_SI Tables Excel Doc)

Source code for the peptide parsing and retrobiosynthetic analysis can be found at http://github.com/molecularmodelinglab/peptide_parser.

translation with recombinant enzymes has enabled cell-free production of peptide-based natural products (NPs) and NP-like molecules. We anticipate that many more such compounds and analogs might be accessed in this way. To assess the peptide NP space that is directly accessible to current cell-free technologies, we developed a peptide parsing algorithm that breaks down peptide NPs into building blocks based on ribosomal translation logic. Using the resultant dataset, we broadly analyze the biophysical properties of these privileged compounds and perform a retrobiosynthetic analysis to predict which peptide NPs could be directly synthesized in augmented cell-free translation reactions. We then test these predictions by preparing a library of highly modified peptide NPs. Two macrocyclases, PatG and PCY1, were used to effect head-to-tail macrocyclization of candidate NPs. This retrobiosynthetic analysis identified a collection of high-priority building blocks that are enriched throughout peptide NPs, yet had not previously been tested in cell-free translation. To expand the cell-free toolbox into this space, we established, optimized, and characterized the flexizyme-enabled, ribosomal incorporation of piperazic acids. Overall, these results demonstrate the feasibility of cell-free translation for peptide NP total synthesis, while expanding the limits of the technology. This work provides a novel computational tool for exploration of peptide NP chemical space, that could be expanded in future to allow design of ribosomal biosynthetic pathways for NPs and NP-like molecules.

Graphical Abstract
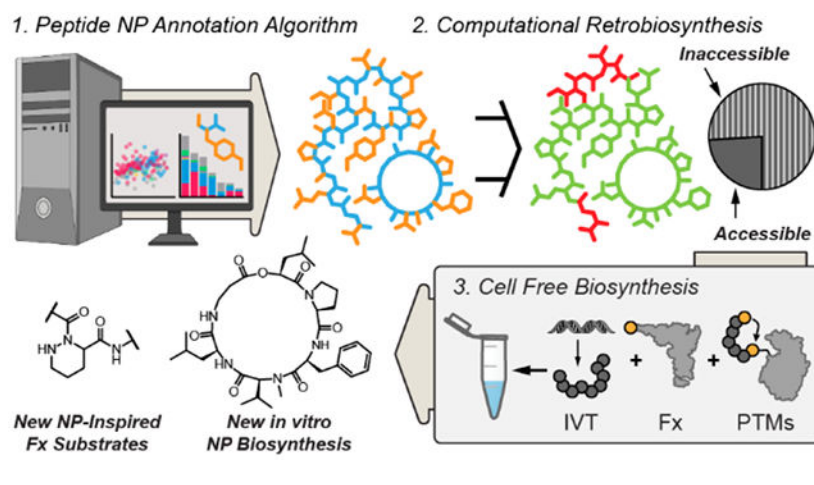
Introduction

Macrocyclic peptide-based natural products (NPs) are an important class of therapeutic agents comprised of tightly constrained and often heavily modified peptide backbones, as exemplified by bacitracin, vancomycin, and FK228 among others.[1] Additional members of this broad class, such as the recently discovered teixobactin and darobactin represent promising new drug leads.[2,3] The group, as a whole, can be distinguished based on the biosynthetic origins of their amide bond constituents, some of which are ribosomally synthesized and post-translationally modified peptides (RiPPs – e.g., darobactin and others) and others are non-ribosomal peptides (NRPs – e.g., bacitracin and others), made by non-ribosomal peptide synthetases (NRPSs).[4,5] In general, NRPs exhibit much broader chemical and structural diversity than RiPPs, although the number of new and novel RiPPs being

discovered is still growing. Consequentially, these two classes have begun to overlap.[6] For example, both NRPs and RiPPs exhibit numerous *N*-methyl and D-amino acids, as well as heavily oxidized and at times cross-linked amino acid side chains, such as those seen in the NRP vancomycin or the RiPP darobactin.[7,8] Importantly, while RiPPs tend to exhibit multiple iterations of a single type of modified amino acid, NRPs will often have multiple, differently modified amino acids embedded within the same scaffold. Together these modifications are essential for the novel bioactivities of both classes of macrocyclic peptide natural products.

While there is some overlap between structures of RiPPs and NRPs, their biosynthetic machinery could not be more different. RiPP biosynthesis exploits RNA-templated, ribosomal translation with amino acids supplied by the tRNA pool to construct peptide scaffolds that are then modified by suites of leader peptide-guided enzymes.[9,10] In contrast, NRPs are made by large, enzymatic assembly line complexes that recognize and charge specific amino acids through the action of dedicated, in-line adenylation domains (or A-domains).[5] Thus, while RiPP biosynthetic pathways have proven remarkably promiscuous, NRPs are decidedly less so. Indeed, RiPPs have proven amenable to the preparation of many hundreds, even thousands of analogs of peptide natural products, simply by altering the precursor peptide at the gene level, while NRPS engineering has had some profound but still limited successes.[11–16] Both biosynthetic machineries have been instrumental in the discovery of new compounds through genome mining efforts and for the preparation of new analogs to improve access to the natural products and explore structure-activity relationships.[17–19]

Recently, augmentations to cell-free biosynthesis (CFB) have significantly expanded the repertoire of ribosomally accessible peptide sequences, making it increasingly possible to obtain NP-like macrocyclic peptides by direct ribosomal translation. In particular, flexizymes, which are ribozymes that effect the transesterification of select, activated esters with the 3'-hydroxyl of tRNAs to allow tRNA loading of a diverse array of amino acid analogues and other carboxylic acids, have transformed the technology.[20] Importantly, the use of flexizyme-loaded tRNAs has exposed the broad promiscuity of ribosomal peptide synthesis and allowed the incorporation of many building blocks, previously thought to be only accessible to non-ribosomal assembly-line biosynthesis.[21] For example, ribosomal translation of *N*-methyl amino acids, beta-amino acids, gamma-amino acids, hydroxy acids, thioacids, D-amino acids, and long-chain carboxylates, such as those found in lipopeptides, have all been achieved by means of flexizyme pre-loaded tRNA translations.[22–28] In addition to these submonomers, flexizyme-based chemistries have allowed co-translational macrocyclization of ribosomal peptides to give highly natural product-like molecules.[29,30] Thus, by bypassing the need for the bulky biosynthetic machinery of NRPSs, flexizyme has significantly simplified CFB approaches to NRP and NRP-like molecules. Beyond the co-translational incorporation of noncanonical building blocks, the post-translational integration of various chemistries and recombinant enzymes has begun to further diversify the chemical space accessible by cell free approaches. For example, recent work from our lab combined flexizyme, with both a chemical oxidation step and three enzymatic transformations for the CFB of thiopeptides, a class of RiPP NPs.[31] Importantly, this work demonstrated that integrating flexizyme and chemical modification simplified accessibility to RiPP NPs by

reducing the number of enzyme-catalyzed steps. Combined, these augmentations further supplement the high throughput capabilities of CFB to rapidly and simply prepare libraries of both NRPs and RiPPs.

Given the increasing scope of ribosomal translation, we sought to assess the direct ribosomal accessibility of the vast array of peptide NPs under the current state-of-the-art in cell-free translation. To this end, we herein develop an algorithm that can cull NP databases for peptide NPs, parse their structures for canonical and non-canonical building blocks, regardless of cyclization state, and then identify cyclization junctures and corresponding linear precursors for macrocyclic peptide NPs (Figure 1a). The resultant data set enables several follow-on analyses and experiments, including: (1) a comprehensive analysis of the biophysical characteristics of peptide NPs and recurrent building blocks that contribute to these features (2) predictions of peptide NPs that should be accessible with current cell-free technology, and (3) identification of current gaps in the cell-free technology that limit access to peptide NPs. There have been a limited number of studies of the bulk biophysical composition of peptide NPs, but, to the best of our knowledge, no such study has sought to correlate those characteristics with the building block composition of the compounds.[32] Based on this data set and attendant analysis, we validate the cell-free ribosomal accessibility of a series of linear, as well as cyclic, peptide NPs; for the latter compounds, we harness the chemistry of two macrocyclase enzymes, PatG and PCY1. Finally, we identify piperazic acid as a non-canonical amino acid (ncAA), that is highly enriched in peptide NPs and has yet to be incorporated into peptides using cell-free translation. To address this gap, we establish, optimize, and characterize the flexizyme-based incorporation of piperazic acid in cell-free translation, opening a new pathway to this privileged chemical space. We anticipate that the retrobiosynthetic analysis and subsequent compound preparation will aid in prioritizing and benchmarking future improvements of flexizyme and other cell-free technologies towards NP or NP-like structures.

Results

Development of Peptide Natural Product Annotation Algorithm

We first developed an algorithm to rapidly analyze peptide NPs according to a ribosomal retrobiosynthetic logic. Development of this algorithm began with a collection of around 500,000 NP structures, in the form of SMILES strings, from three online datasets: Lotus, Supernatural II, and MiBiG (Figure 1b–1).[33–35] NP structures were then rendered as molecular graphs, with atoms as nodes and bonds as edges. These graphs were searched for patterns that aligned with those of amino acids or other NP building blocks (Figure 1b–2). Importantly, the search algorithm allows for deviations from conventional amino acid structures that are common in NPs: extended monomers, such as β- and γ-amino acids, alternative backbone heteroatoms, such as hydroxyacids which form esters instead of amides, and backbone dehydrations, as in azol(in)es. Once all possible building blocks were detected, the structures were rendered as graphs again, this time with full building blocks represented as unidirectional nodes based on N- vs C-termini and edges defined as peptidic connections (e.g. amides, esters, azol(in)es) between building blocks (Figure 1b–3). These peptide connectivity-based graphs were then parsed to identify consecutive stretches of at

least four building blocks deemed the peptide backbone. In many cases, NP diversity makes defining the backbone challenging because multiple potential backbone patterns can co-exist depending on the monomers that are considered acceptable building blocks (Figure S1). Therefore, we implemented a scoring function based on presumptive ribosomal translation to detect the optimal (typically longest) peptide backbone (see Supporting Information for full description of scoring rules). Identification of backbones condensed the dataset from roughly 500,000 NPs to 5,291 peptide NPs.

With backbones identified for each NP, the algorithm next gathers information on side chains by tracing the graphs of the non-backbone "R-group" of each building block (Figure 1b–4). Again, this graph-tracing can be complicated by the complex structures of NPs. For example, tracing along the side chain of cyclic monomers, such as Pro, will encounter the backbone again. Similarly, a side chain trace may encounter atoms from the backbone of a different building block as a result of macrocyclization. Therefore, encountering a backbone atom during a side chain trace terminates the tracing and the information is stored for future use. All other side chain traces are terminated by reaching the end of the side chain. The identities of recorded side chain graphs are stored as unique hashes along with various properties of each building block, which include stereochemistry, N-modifications (such as N-methylation), amide vs ester, $\alpha/\beta/\gamma$ amino acid, and others. In a final step, the algorithm analyzes peptide NPs for the presence of macrocyclization by searching for connections between different areas of the detected peptide backbone (Figure 1b–5). This last step resulted in identification of 1,441 linear peptides with no macrocyclic linkage and 3,850 macrocyclic peptide NPs that were further subdivided based on the type of cyclization linkage. Although this is an accurate description of the algorithm, the outputs were further refined, during its development, through an iterative design and review process. A more detailed description is provided in the Supplementary Information and the code is publicly available at [http://github.com/molecularmodelinglab/peptide_parser]. Below we present insights from this analysis of the macrocyclic peptide NPs.

Classes of Macrocyclic Peptide Natural Products

The macrocyclic NPs were split into non-overlapping classes based on five major types of cyclization connection (Figure 1c). The largest group, Class 1, consists of macrocycles where the detected backbone in the connectivity map is a continuous loop, resulting from head-to-tail cyclization between a terminal amine and carboxylic acid. These backbone macrocycles, such as cyclosporin and tyrocidine, consist of 1,843 NPs. Class 2 is the second largest class, consisting of 1,439 macrocycles where cyclization occurs between a side chain (such as Thr, Lys, Glu) and the backbone (either the N- or C-term). These macrocycles were detected when side chain tracing directly encounters either the N-terminal amine or C-terminal carboxylate backbone of another building block. Due to this method of cyclization, a portion of the peptide backbone extends outside of the macrocyclic portion of the molecule. Colistin and daptomycin are both examples of important antibiotics that fall within this class. Class 3 consists of 71 macrocycles that are formed by cyclization between two side chains, in which the backbone can extend from the macrocycle in both the N- and C-terminal directions. Detection of these macrocycles occurs when side chain tracing runs into the backbone of an alternative building block. However, unlike Class 2 macrocycles, the

connections in Class 3 occur at either the α-, β-, or γ-carbon. Detection of these cyclization sites is redundant due to detection occurring during side chain tracing from either involved building block. Streptide, arylomycin, and oxytocin are examples of NPs from this class. Class 4 is specifically made up of 153 thiopeptides, such as thiostrepton and thiocillin.[36] Thiopeptides, characterized by a side chain-to-side chain, formal cycloaddition technically also belong to Class 3. However, due to the uniqueness of their ring juncture and with an eye toward retrobiosynthetic assessment, we chose to separately classify the thiopeptides. The unique, tri-substituted, nitrogenous, heterocyclic ring juncture was used as the graph pattern for specific detection of thiopeptides. Finally, Class 5 includes 344 multicyclic NPs. These multicycles were grouped based on the presence of multiple sites of connection between distant sections of the backbone and can contain different combinations of the other cyclization connectivities. For example, microviridins and lanthipeptides consist of multiple side chain-to-side chain linkages, meanwhile cyclotides are backbone macrocycles with multiple disulfide crosslinks that form a knot-like motif. Using the NPs collected from MiBiG, we compared the biosynthetic origin of the NPs across these classes, finding that Classes 1, 2 and the linear peptides were more enriched toward non-ribosomal peptides, meanwhile Classes 3 – 5 generally had higher numbers of RiPPs (Figure S2).[35]

Biophysical Analysis of Peptide Natural Products

We next sought to examine the bulk biophysical characteristics of these peptide NPs. To visualize the chemical space of peptide NPs, we generated a tree map based on the chemical similarity of NPs to one another, color-coded by macrocycle Class (Figure 2a).[37] This tree map aligns well with macrocycle class assignments as members of the same class generally cluster along the same branch. In order to further characterize each class of peptide NPs an expanded set of biophysical properties were examined (Figure S3). Specifically, size (represented as volume in Å^3) and lipophilicity (represented as a calculated partition coefficient, clogP) have been closely correlated with the bioavailability and "drug-likeness" of peptide NPs and we, therefore, analyzed the distribution of these metrics across the database of peptide NPs (Figure 2b).[38] A vast majority of the NPs fall within a cluster that is generally considered to have a higher degree of intrinsic cell permeability, specifically in the range of higher lipophilicity (−2 – 10) and a volume between 500 and 1000 Å^3. Within this overall NP cluster, a degree of sub-clustering by classification can also be observed. For example, the linear peptides and Class 1 backbone macrocycles are generally smaller than the other classes, while Class 2 macrocycles contain a large cluster with higher volume and cLogP. Notably, the data set appears to fall off precipitously in size at approximately 1000 Å^3. This cutoff likely results from the low frequency of larger RiPPs and other bacteriocins in the data set; we now know from bioinformatic efforts that there are many more RiPP gene clusters in bacterial genomes than RiPP natural products that have been isolated and reported in the literature.[39,40] Furthermore, a general trend of increasing cLogP is observed as volume increases.

Canonical and non-Canonical Amino Acids in Peptide Natural Products

Given the bulk biophysical properties of the peptide NPs, we sought to assess what mixtures of canonical and non-canonical amino acids are used to achieve these characteristics. We first analyzed the frequency of the occurrence of each of the 20 canonical amino acids

in these peptide NPs (Figure 2c). Complementing the generally high lipophilicity of the data set, hydrophobic and aromatic amino acids are heavily over-represented, while charged residues are under-represented. Proline (Pro) is by far the most common building block in Class 1 macrocycles, while being less common in linear peptides, perhaps implying its importance in promoting conformations that facilitate backbone macrocyclization. Somewhat surprisingly, Arg is broadly present in the Class 2 macrocycles, suggesting a possible role for specific charge in the activity of many of these NPs. The low frequency of Cys observed is due to the frequency of modifications to this residue (often used in cyclization) that leave very few free Cys side chains.

ncAAs, including both modifications to canonical amino acids, such as *N*-methylation or epimerization, and entirely unique side chains, make up just over half of the building blocks in the peptide NP data set (20,416, 51.4%). To validate our coverage of ncAAs in the NPs, we compared our collection to the named amino acid building blocks provided by MiBiG, where we observed 129 out of 167 (77%) building blocks throughout our collection of NPs.[35] Figure 2d presents a heatmap that compares the frequency of ncAAs to the total number of AA building blocks in a peptide NP. Most of the NPs range in size from four to ten residues with three to five often being non-canonical building blocks. Linear peptides cluster more towards the smaller size (4 AAs) and generally lower ncAA frequency, while the macrocycles favor larger sizes and slightly higher ncAA frequency (Figure S4). Comparing the macrocycle classes, the total number of ncAAs (Figure 2e) and the longest run of consecutive ncAAs (Figure 2f) were generally very similar except for the Class 4 thiopeptides, which typically have both more ncAAs and longer stretches of them. The ncAAs could further be sub-categorized based on common features (Figure 2g). Variations in the side chains of α-amino acids are by far the most common category of ncAA, followed by significant numbers of D-AAs and *N*-methylated AAs. Interestingly, β-hydroxyacids, which are thought to be able to participate in a hydrogen bonding interaction with the peptide backbone and induce secondary structure, are also abundantly present.[41] Azol(in)e residues stand out most frequently in the Class 1 and 4 macrocycles, aligning with their established presence in RiPPs. Modifications to canonical amino acids, including *N*-methylated versions and D-stereoisomers, account for just under half of all ncAAs (8,487, 41.6%), with many of these most-frequent modifications occurring on hydrophobic and aromatic AAs (Figure 2h). To further breakdown the specific non-canonical side chains that make up the rest of the ncAAs, the most common individual building blocks (ignoring modifications to canonical amino acids) were identified (Figure 2i). The α-disubstituted building block α-aminoisobutyric acid (Aib) is the highest frequency ncAA due to its high abundance in linear peptides.[42] Additionally, dehydroamino acids (dehydrobutyrine – Dhb, dehydroalanine – Dha) are highly abundant across peptide NPs and are unsurprisingly enriched in the RiPPs in Classes 4 and 5. A series of building blocks from the microcystin family of NPs are abundant in the dataset due to the large number of analogs found within this family (at least 279 variations isolated to date).[43] These building blocks include the PKS-derived Adda ((2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-(4E,6E)-dienoic acid) unit, along with ^{Me}Dha and β-Me-D-Asp where the side chain carboxyl is utilized in the peptide backbone. Furthermore, the unique

cyclic 3-amino-6-hydroxy-2-piperidone (Ahp) residue is observed frequently, but strictly within Class 2 macrocycles.[44]

Cyclization Motifs in Peptide Natural Products

Lastly, we assessed the use of side chains for cyclization in the Class 2, 3, and 5 macrocycles. Within Class 2, the natural amino acid, Thr is by far the most heavily used for cyclization, followed by Glu, Lys, and Ser (Figure 2j). A number of ncAAs are also employed in side chain-to-backbone cyclic junctures, such as variable lengths of alkyl amine alternatives to lysine (i.e. diaminopropionic acid – Dap, and diaminobutyric acid – Dab). Nearly half of all β-hydroxyacids in the NP database are involved in depsipeptide linkages through the hydroxy group and macrocyclization using the side chain hydroxyl of β-hydroxyleucine is particularly heavily utilized in this manner. The algorithm can also be used to distinguish common side chain-to-side chain cyclization motifs in Classes 3 and 5 (Figure 2k). Here, disulfides between Cys residues are the most frequently observed cyclization motif, along with Cys often showing up in the lanthionine and (methyl)lanthionine bridges arising from cyclization with dehydroalanine or dehydrobutyrine in lanthipeptide biosynthesis. All iterations of lactone and lactam bridges between Asp/Glu and Thr/Ser/Lys residues can be found in the data set, with the Asp to Thr linkage being the most common. Furthermore, C-C crosslinks between hydroxyphenyl glycines (Hpg), characteristic of glycopeptide NPs like vancomycin, are also frequently observed. Because of this vast diversity of cyclization motifs, significant future method development will likely be required to access all structures in the database.

Ribosomal Accessibility Analysis of Peptide Natural Products

We next sought to probe this catalog of NP building blocks to predict NPs that might already be directly accessible to ribosomal translation, as well as opportunities for advancement of the CFB technology. NP accessibility was determined based on reported compatibility of individual amino acid building blocks with ribosomal translation, as well as select post-translational modifications that have been shown to expand CFB capabilities. We began by compiling a list of CFB-compatible building blocks from literature reports – the complete set together with references is provided in Tables S3 and S4.[22,24–28,31,45–84] In the subsequent NP analysis, any peptide that contains a single 'inaccessible' building block, such as a D-enantiomer that the ribosome does not translate or an ncAA that has not yet been proven in translation, is itself, considered 'inaccessible'. For example, flexizyme has been used to show that the D-enantiomers of sixteen of the canonical amino acids can be elongated by *E. coli* ribosomes; meanwhile, the remaining three (I, R, W) cannot.[27,45,48] Therefore, a NP containing one of the latter three D-amino acids is here considered 'inaccessible'. Additionally, many NP-derived building blocks have never been tested for ribosomal compatibility and despite the possibility that a portion of these might be incorporated upon testing, remain inaccessible in our analysis. Furthermore, ribosomal translation is known to have difficulty with 'challenging' sequences, incorporating many low yielding ncAAs across a peptide sequence and/or tracts of consecutive, low yielding ncAAs.[22,27] To account for this in the analysis, we established a three-tiered ranking system for accessible NPs, where High confidence represents the most accessible (0 – 2 ncAAs, nonconsecutive), Medium is moderately challenging (3 – 5 ncAAa, 3 consecutive), and

Low confidence is the most challenging (remaining accessible NPs). Importantly, as a result of the variability in the ncAA incorporation data (*in vitro* translation methods, optimizations, yield reporting), this accessibility assignment and ranking system remains a crude qualitative assessment of the ribosomal compatibility of NP peptides.

We then applied these rules to our full database to identify peptide NP sequences that are predicted to be ribosomally tractable; here we define ribosomally tractable as being comprised only of building blocks previously reported to be permissible to ribosomal translation, without considering macrocyclization. This retrobiosynthetic analysis predicted 1,270 (24%) NPs as ribosomally tractable with the largest number of accessible NPs arising from Class 1 and linear peptides (Figure 3a). Classes 3 and 5 had comparatively higher proportions of accessible macrocycles out of their total, likely due to the high abundance of RiPPs in these classes. Meanwhile, the large set of Class 2 macrocycles contained just 165 NPs (12.9%) predicted to be ribosomally tractable. Compared to the theoretical probability of accessibility based on the proportion of building blocks that are accessible, the NPs followed a similar trend with generally slightly higher than expected accessibility perhaps hinting that many of the inaccessible NPs contained multiple inaccessible building blocks (Figure S5). The confidence of accessibility across each class was also compared using the three-tiered ranking system based on total and consecutive ncAAs (Figure S6). Notably, more than half of all accessible NPs (867, 68.3%) fall in the High confidence tier that are very likely to be accessed by the ribosome. In general, if a Class 1 or 2 sequence was predicted to be accessible, it more often fell within the High confidence tier, whereas a greater percentage of the class 3 – 5 fell in the Medium and Low confidence tiers, even though there were higher proportions of accessible macrocycles in these classes. Overall, the frequency of medium and low confidence accessible NPs suggests that incorporation of multiple ncAAs, often consecutively, would be required for a large portion of NP accessibility. To further visualize CFB accessibility across the peptide NP space, we highlighted the predicted accessible NPs within our TreeMap representation (Figure S7). In line with the TreeMap having similar NPs in close proximity, predicted accessibility generally clustered together within branches of NPs within each class, leaving large gaps in NP structures that are currently likely inaccessible to ribosome-based synthesis.

We could further use this analysis to identify building blocks that currently prevent access to NPs, either due to difficulty of incorporation or else lack of incorporation data. We began by categorizing the top ncAAs "blocking" translation accessibility (Figure 3b), where untested or incompatible non-canonical α-side chains are the most common inaccessible category, followed by building blocks with multiple modifications beyond their side chain. For example, both ^{Me}Phe and D-Phe are compatible with the ribosome, however, D-^{Me}Phe and all other multiply modified building blocks (D-hydroxy acids, β/γ-^{Me}AAs) have rarely been tested and are therefore leading examples of 'inaccessibility'. Additionally, the abundance of D-stereoisomers, and in particular those that are known to not be tolerated in ribosomal elongation (D-Ile, D-Trp, D-Arg, and non-canonical side chains), are among the leading inaccessible NP building blocks. To further explore the individual side chains obstructing accessibility, we investigated the specific cases that most commonly limited retrobiosynthetic accessibility (Figure 3c). The previously described microcystin-derived building blocks and the heterocyclic Ahp are the most common inaccessible side

chains and are almost entirely found within Class 2. Interestingly, we observed a high abundance of a cyclic-hydrazine-containing building block called piperazic acid (Piz) that had never been tested for translation. This unique building block is associated with strong conformational effects and NPs containing this building block demonstrate an array of potent biological activities.[85,86] Lastly, (Me)oxazoles, Ome-dehydrothreonine (DhT(Ome)), and the β-hydroxy fatty acid ((R)-β-OH-Dec) are all consistently blocking access and the development of methods for their incorporation would rapidly expand ribosomal access to additional NPs.

As a final prelude to using this analysis to direct the *in vitro* biosynthesis of peptide NPs, we sought to assess compatibility of the compound database with known cyclization methods. While previous work on lasso peptides, lanthipeptides, and thiopeptides has shown that these groups can be amenable to enzymatic cyclization in a CFB context, there are relatively few robust, biocompatible macrocylization methods that might be used to make either Class 1 or Class 2-like macrocyclic linkages.[31,87–89] We therefore focused in on the RiPP macrocyclases PatG and PCY1, from Patellamide and Orbitide biosynthesis, respectively.[90–94] Both enzymes follow a conserved, two-step mechanism of site-selective proteolysis followed directly by head-to-tail macrocyclization (Figure S8). Both enzymes have well-documented substrate tolerance profiles, however, neither enzyme has been significantly tested with diverse sets of ncAA-containing substrates.[95] For PatG, the C-terminus must be a turn inducing residue, such as proline or azolines, meanwhile PCY1 can also tolerate some small side chains at this position, such as Gly, Ala, and Ser. Additionally, neither enzyme appears to tolerate D-isomers at the C- or N-terminal cyclization positions (Complete set of cyclization limitations is provided in the SI – Figure S8). Thus, we applied these cyclization requirements to our retrobiosynthetic analysis of the Class I macrocycles and found that 329 (69.9%) of the 471 building block 'accessible' NPs are predicted cyclization substrates of either PatG or PCY1 (Figure 3d). Finally, we generated a new TreeMap containing only class 1 backbone macrocycles and highlighted both building block and macrocyclization accessibility, again predicting that accessibility generally clustered within specific branches (Figure S9). Overall, this macrocyclase analysis highlights several candidate macrocycles for direct ribosomal translation and *in vitro* macrocyclization.

CFB for Natural Product Preparation

We sought to test the algorithm's accessibility predictions by *in vitro* expression of a series of candidate peptide NPs. To this end, we selected a focused panel of 43 peptide NPs, including 12 linear and 31 Class I macrocyclic peptides. The majority of the panel was focused on high to medium confidence predictions but a select number of low confidence and therefore, high complexity "challenge" substrates were also included (Table 1). At the gene level, sequences were fitted with an N-terminal HA tag to enable purification from the translation machinery and a Factor Xa protease site to allow site specific removal of this tag, as well as the requisite formyl-methionine (fMet) initiator residue. In addition, genes for cyclic peptide NPs were prepared as two versions, with either C-terminal PatG or PCY1 recognition sequences for cyclization. Lastly, translations were carried out with either a standard (NEB PURExpress product #E6840S) or a customized PURE kit (PUREfrex systems product #PFC-Z9901), which allowed for optimized concentrations

of translation factors, EF-P, and reservation of up to three select aaRSs (HisRS, AsnRS, and GlnRS), in line with the current best reported conditions for Flexizyme-based codon reprogramming.[96,97] All additional proteins required for CFB reactions were individually expressed and purified (Figure S10).

Using this general design, genes were translated with the relevant flexizyme pre-charged tRNAs for ncAA incorporation (Table S5), incubated with Factor Xa, and then analyzed by mass spectrometry.[98] As shown in Table 1, translation of nearly all linear NPs that were predicted to be high confidence for accessibility and contained up to two ncAAs were readily produced and detected (entries **1** – **11**, Figure S11, Table S6). Compound **12**, was a particularly challenging, low confidence substrate containing eleven α-aminoisobutyric acid (Aib) residues and two β-glycine residues and did not appear to translate in our hands. Several of these linear NPs exhibit N-terminal fatty acyl modifications, which we also sought to incorporate *via* late-stage chemical acylation. In this case, after release of the core by Factor Xa, the peptides were subjected to acylation with either acetyl (**6**) or hexanoyl chloride (**7** - **12**).[99,100] Chemical acylation proved highly efficient for all but one of the translated peptides, compound **11**, which suffered from a combination of a low yielding translation and poor Factor Xa cleavage due to the N-terminal D-AA (Figure S11).

A similar pattern of success was observed in the group of macrocycle test substrates. Following translation and Factor Xa treatment, we could detect accurate translation of all 14 high (**13** – **26**) and 6 medium (**27** – **32**) confidence macrocyclase substrates tested, as well as 2 out of 5 low confidence (**35** – **36**), challenge substrates, demonstrating both the accuracy of the predictions and the power of flexizyme-augmented *in vitro* translation for NP synthesis (Tables 1 and S7, Figures S12–S13, S16–S17). Failure in translation of several of the medium and low confidence prediction substrates highlights the clash between the features of many of the more complex NPs and the current limits of ribosomal translation of ncAAs: extended tracts of multiple D- and *N*-Methyl amino acids caused severe drop-offs in translation efficiency in our hands. For example, lack of translation of **33** was likely due to the poor incorporation, even in presence of EF-P, of the already low yielding ^{Me}Ile consecutively with ^{Me}Leu. Furthermore, selection of lower confidence NPs that were predicted to be cyclizable required slightly adjusting our accessibility requirements. Only a limited set of analogs from a few families of NPs fit within our original requirements. Therefore, we selected two medium (**33** – **34**) and three low (**37** – **39**) confidence NPs that were outside of our original predicted accessibility by single amino acids that were likely to be tolerated due to similarity to 'accessible' building blocks (Val-OH, D-Lys-Az, D-Leu-OH, ^{Me}Ile). Despite confirming these ncAAs could be incorporated individually, translation in the context of the NPs was unsuccessful (Figure S14).

Cyclization with PatG or PCY1, while robust with substrates comprised of only canonical amino acids, proved variable with more complex substrates (Tables 1 and S9, Figures S12–S13, S16–S17). These cyclase enzymes can yield three different species: the starting linear substrate, the linear hydrolysis product (less the C-terminal recognition sequence), and the intended macrocycle. Of the four substrates that consisted only of canonical amino acids, we could observe efficient cyclization of each (**13** – **16**). Notably, it proved beneficial to have multiple enzymes with complementary substrate tolerance profiles; using PatG,

cyclization of **16** gave mostly proteolysis product, but PCY1 gave mainly macrocycle. In the series of high confidence, ncAA-containing substrates (**17 - 26**), three (**17 – 19**) were able to be macrocyclized, but most were poor substrates for the enzymes, resulting in predominantly hydrolysis product (**20 – 26**). Among the lower confidence NPs, several contained C-terminal cyclization residues that not been reported, but were very similar to known tolerated residues. Thus, we confirmed PCY1 tolerance of both a C-terminal *trans*-4-hydroxyproline (Hyp) and ^{Me}Ala on macrocyclization of a test substrate (Figure S15), however, poor translation yields of the NPs these residues were observed in prevented us from showing macrocyclization of these residues in a NP context (**34** and **38**). In contrast, we were able to translate eight analogs within and similar to the Isaridin family of macrocyclic NPs (**27 – 32, 35 – 36**), with five of these substrates undergoing cyclization (**27 – 31**).[101,102] The Isaridin series highlights the potential for using Flexizyme-augmented CFB to coarsely traverse structure-activity-relationships (SAR) within NP macrocycles, even from the medium and lower confidence tiers. Alternative macrocyclases with expanded substrate scope could significantly enhance such efforts.

In a final set of substrates for CFB, we sought to express a series of thiazole containing macrocycles. Azoline-forming enzymes from RiPP biosynthetic pathways are well tolerated in CFB, but here, the incorporation of multiple unnatural amino acids, could pose a severe test of their permissivity. In the event, we introduced the recognition sequence for the promiscuous cyclodehydratase LynD (from Aesturamide biosynthesis) in place of the HA tag in these gene designs and, following translation, peptides were incubated with LynD and one of the cyanobactin azoline dehydrogenases ThcOxi, from *Cyanothece* PCC 7425, or ApOxi, from *A. platensis*, followed by Factor Xa and macrocyclization by PatG or PCY1.[31,103] This strategy successfully provided access to three high confidence substrates (**41 – 43**), although only **41** and **42** could be converted to the macrocycles, while **42** strongly favored hydrolysis. An additional thiazole-containing NP from the medium confidence accessibility tier (**43**) could not be observed by mass spectrometry, in either linear or cyclic forms, or before LynD-mediated thiazoline formation suggesting ribosomal incompatibility of this substrate, rather than cyclodehydration.

Directed Expansion of the CFB Toolbox for Piperazate Incorporation

We last turned our attention to the incorporation of new ncAAs that had not yet been shown compatible with ribosomal translation. or this effort, we chose to focus on piperazate (Piz) as both a structurally and chemically distinguished ncAA, present in a large number of structurally diverse NPs with wide ranging biological activities.[85] Notably, Piz NPs exhibit acylation and therefore backbone continuation at either the α/N2 or β/N1 position, although the α/N2 is significantly more common (Figure 4a).[85] This natural substitution pattern stands in stark contrast to the chemical reactivity of the two hydrazine nitrogens; under conventional coupling conditions, the β/N1 is considerably more reactive due to steric and electronic effects and α/N2 has proven the more synthetically challenging.[104] Based on recent reports of ribosomal translation of α-aminoxy and substituted α-hydrazino acids, we hypothesized that an augmented *in vitro* translation might accept piperazates when pre-charged onto a tRNA and thus allow ribosomal access to this abundant motif.[54] Therefore, we first tested and then optimized the flexizyme (dFx) catalyzed charging of L-

and D-Piz-DBE by the μ-helix gel shift assay (Figure S18).[20] Subsequent translation into test sequences containing either a single L-/D-Piz or four consecutive residues established the feasibility of ribosomal incorporation as confirmed by mass spectrometry (Figure 4b/ Figure S19a). Inclusion of elongation factor, EF-P, proved essential for robust incorporation of D-Piz and consecutive incorporations of either enantiomer (Figure S19b).[97] With these optimized conditions, we measured the incorporation efficiency of piperazate translation using a NanoBiT luciferase assay (Figure S20).[48] Compared to a fully canonical translation, L- and D-Piz were incorporated with a 68% and 23% yield, respectively, with both being higher than flexizyme-based Methionine incorporation (21%), further highlighting the utility of this method for Piz-containing peptide synthesis.

We next sought to assess the regioselectivity (α/N2 vs. β/N1) of elongation from a ribosomally incorporated piperazate. In order to deconvolute the outcome of ribosomal elongation, we chemically synthesized a small series of peptide standards that contain either an N1 or N2 acyl-Piz subunit (Figure S21).[105,106] Importantly, the two versions of each standard peptide exhibit different retention times when analyzed by LCMS (Figure 4c). We translated these same Piz-containing sequences, treated them with peptide deformylase to remove the formyl group on the initiating Met, and compared retention times to the standards.[29] In all cases, the translated peptides aligned with the retention times of the N2 standards, suggesting that the ribosome prefers the N2 position for peptide bond formation. Interestingly, the ribosomal selectivity for the α(N2)-position is opposite to what would be expected based solely on chemical reactivity. However, this regioselectivity is beneficial, since it provides access to the more common and more challenging NP linkage. In an attempt to exploit these capabilities, we designed and translated four genes for macrocyclic Piz-containing NP analogs (Figure 4d–g). Gratifyingly, all four linear substrates could be readily detected by mass spectrometry. Despite demonstrating PCY1 tolerated a C-terminal Piz building block as the cyclization residue, efforts at macrocyclization of NP substrates with PatG or PCY1 failed to yield the respective NP macrocycles (Figure S15). Nevertheless, these results suggest that a number of Piz-containing NPs might be accessible with the availability of better cyclization enzymes or conditions.

Discussion

Ribosome-based cell free biosynthesis presents a rapid method for the templated synthesis of libraries of diverse peptides. Despite recent augmentations expanding the reaches of this technology to large numbers of non-proteinogenic, ribosome-compatible building blocks, the applicability of this technology towards NP synthesis has not been thoroughly defined. Therefore, to facilitate the ribosomal translation of NP structures, we assembled a database of NP structures and carried out an exhaustive cheminformatic characterization of this database. We then conducted a CFB-oriented, computational retrobiosynthetic analysis of these NPs with the goal of both measuring current NP accessibility and providing guidance for the development of new methods to expand accessible space. This analysis allowed for classification of NPs by both the ability to incorporate the building blocks within a NP, mainly by ribosomal translation, and the ability to macrocyclize the NP post-translationally, if applicable. In line with our retrobiosynthetic predictions, we harnessed CFB to prepare a library of 24 NPs, while the ribosomal translation machinery alone accessed 36 NP

precursors. Importantly, these findings validated our retrobiosynthetic predictions and further support CFB technology as a tool for the rapid preparation of peptide NP libraries.

One potential limitation of our analysis is the reliance on established NP datasets for the NP structures. These databases are inherently biased toward NP structures originating from prioritized genera, along with the nature of the discovery of NPs often occurring in clusters of highly similar analogs. To assess such bias, we plotted the occurrence frequency of NPs from the most common genera (Figure S22). This graph shows the distribution of genera present and indicates that there is no obvious trend in the translation accessibility across genera.

Although many ncAAs present in the macrocyclic peptide NPs already have demonstrated compatibility with ribosomal translation, our retrobiosynthetic analysis identifies many that are incompatible or remain untested. Identification of these building blocks provides guidance on the most crucial ncAAs for broadening NP accessibility via direct *in vitro* translation. Several building blocks, including select D- and β-amino acids, have previously been reported as incompatible with ribosomes. These challenging building blocks will likely require dramatic engineering efforts or else alternative approaches for optimal incorporation. Other building blocks, however, have never been tested for ribosome compatibility, and therefore, present a rapid avenue for expanding NP accessibility. Piz proved an opportune example of this latter group that is highly enriched in NP structures. In our hands, both L- and D-Piz could be readily incorporated in augmented CFB reactions via Flexizyme enabled pre-loading of tRNAs. Additionally, this work showed that ribosomal elongation is able to selectively access the otherwise synthetically challenging α/N2 linkage. Similar translational bias over inherent chemical reactivity may aid in accessing other synthetically challenging linkages.

Despite successful translation of nearly all predicted medium and high confidence linear and Class 1 NP substrates, ribosome-based peptide synthesis still presents limitations for complex stretches containing several, consecutive, challenging building blocks, such as *N*-methyl, D- and β-amino acids. Under heavily optimized conditions, Flexizymes have been used to incorporate up to 10 challenging ncAAs in a row, but this kind of consecutive incorporation works best when a single codon is reprogrammed and typically requires a minimal translation system to prevent background incorporation of canonical amino acids.[22,23,96] Select ncAAs have lower incorporation efficiencies and thus tracts of high complexity within a peptide, requiring incorporation of multiple of these low efficiency residues will invariably lead to lower translation yields. However, Flexizyme certainly does highlight what the ribosome is capable of and it is possible that aaRS or even ribosome engineering could provide significant improvements in incorporation efficiency. Alternatively, new post-translational modifications could potentially help access these more complex substrates by reducing the ncAA load on the ribosome. In recent years, RiPP pathways have yielded *N*-methyl transferases and epimerases that could synergize with Flexizyme-incorporated ncAAs.[107,108]

In addition to building block-based accessibility, our retrobiosynthetic analysis and subsequent CFB experiments also examined macrocyclization capabilities. Although our

preliminary retrobiosynthetic analysis predicted that a substantial number of Class 1 macrocycles might be accessible with established RiPP macrocyclases, PatG and PCY1, the availability of CFB-compatible cyclization methods remains a major limitation in the accessibility of macrocyclic NPs. We observed successful cyclization of 14 NPs, suggesting that these enzymes are applicable within our predictions. However, competitive hydrolysis of moderate to weak substrates is a significant problem, preventing cyclization of 11 NPs predicted to be substrates for these enzymes. In particular, smaller substrates containing five or six building blocks were typically more challenging, perhaps due to increased strain required to adopt the pre-cyclization transition state conformation. Elsewhere, computational modeling of peptide conformation has been used to predict the proclivity of sequences to undergo PatG cyclization with some success.[109] It is possible that integration of such tools with our retrobiosynthetic analysis could further improve the accuracy of our predictions. Chemical methods, such as the spontaneous cyclization of C-terminal translated thioesters, which has been used to access **25** (Scleramide), might enable more cyclic NPs, however, these methods require additional ncAA incorporation.[29] Even taking these methods into account, it is clear that deep characterization of additional enzymatic cyclases or the development of new chemical cyclization methods will be essential to access many of these macrocyclic NP structures. Importantly, we expect our retrobiosynthetic predictions for ribosomal translation to apply across all of the Classes of NPs as long as the component building blocks are able to be accessed. This is further supported by Figure 2e–f, which highlights that there is no significant difference in the number and the length of consecutive runs of ncAAs, both of which could make translation more challenging beyond the identities of the building blocks. Therefore, the development of macrocyclization methods beyond Class 1 would likely rapidly provide access to these NPs. In particular, the Class 2 side chain-to-backbone macrocycles present a similarly large set of NPs and although there has been some limited success at making depsipeptides in CFB, there are as yet, no known CFB-compatible cyclization methods that might be able to access the vast majority of these Class 2 linkages.[108] The macrocyclizing thioesterase domains from NRPS assembly lines, themselves, which regularly carry out such side chain-to-backbone, as well as the more standard, head-to-tail macrocyclizations, might be able to be recruited to expand accessibility in Classes 1 and 2.[109]

In conclusion, many peptide NPs are already accessible through augmented CFB but there are still significant gaps to be filled. By cataloging the constituents of these NPs according to a ribosomal retrobiosynthetic logic, we have more clearly defined these gaps and quantified the number of NPs that might be enabled by each new co-translational or post-translational modification added to the CFB toolkit. Clearly, NP mimicry is not the only endpoint for CFB technology, but we anticipate that using NP scaffolds as challenge molecules might focus and direct biosynthetic methods development, as it has in organic synthesis and lead to insights into their bioactivity. In the near term, this scaffold analysis can be used to help prioritize building blocks for ribosomal incorporation and biosynthetic transformations for enzyme discovery efforts; in the long term, the analysis may also enable analog and library design for compound screening efforts (e.g., mRNA display), as well as predictive modeling of biological activity by incorporating published NP activities into our analysis. The annotation and retrobiosynthesis algorithms developed here can be readily updated to

accommodate new CFB technologies, along with updates to NP databases and predicted structures from genome mining efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

(1). Luther A; Bisang C; Obrecht D Advances in Macrocyclic Peptide-Based Antibiotics. Bioorg. Med. Chem 2018, 26 (10), 2850–2858. DOI: 10.1016/j.bmc.2017.08.006. [PubMed: 28886999]

(2). Ling LL; Schneider T; Peoples AJ; Spoering AL; Engels I; Conlon BP; Mueller A; Schäberle TF; Hughes DE; Epstein S; Jones M; Lazarides L; Steadman VA; Cohen DR; Felix CR; Fetterman KA; Millett WP; Nitti AG; Zullo AM; Chen C; Lewis K A New Antibiotic Kills Pathogens without Detectable Resistance. Nature 2015, 517 (7535), 455–459. DOI: 10.1038/nature14098. [PubMed: 25561178]

(3). Imai Y; Meyer KJ; Iinishi A; Favre-Godal Q; Green R; Manuse S; Caboni M; Mori M; Niles S; Ghiglieri M; Honrao C; Ma X; Guo JJ; Makriyannis A; Linares-Otoya L; Böhringer N; Wuisan ZG; Kaur H; Wu R; Mateus A; Typas A; Savitski MM; Espinoza JL; O'Rourke A; Nelson KE; Hiller S; Noinaj N; Schäberle TF; D'Onofrio A; Lewis K A New Antibiotic Selectively Kills Gram-Negative Pathogens. Nature 2019, 576 (7787), 459–464. DOI: 10.1038/s41586-019-1791-1. [PubMed: 31747680]

(4). Arnison PG; Bibb MJ; Bierbaum G; Bowers AA; Bugni TS; Bulaj G; Camarero JA; Campopiano DJ; Challis GL; Clardy J; Cotter PD; Craik DJ; Dawson M; Dittmann E; Donadio S; Dorrestein PC; Entian K-D; Fischbach MA; Garavelli JS; Göransson U; Gruber CW; Haft DH; Hemscheidt TK; Hertweck C; Hill C; Horswill AR; Jaspars M; Kelly WL; Klinman JP; Kuipers OP; Link AJ; Liu W; Marahiel MA; Mitchell DA; Moll GN; Moore BS; Müller R; Nair SK; Nes IF; Norris GE; Olivera BM; Onaka H; Patchett ML; Piel J; Reaney MJT; Rebuffat S; Ross RP; Sahl H-G; Schmidt EW; Selsted ME; Severinov K; Shen B; Sivonen K; Smith L; Stein T; Süssmuth RD; Tagg JR; Tang G-L; Truman AW; Vederas JC; Walsh CT; Walton JD; Wenzel SC; Willey JM; van der Donk WA Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. Nat. Prod. Rep 2013, 30 (1), 108–160. DOI: 10.1039/c2np20085f. [PubMed: 23165928]

(5). Süssmuth RD; Mainz A Nonribosomal Peptide Synthesis-Principles and Prospects. Angew. Chem. Int. Ed 2017, 56 (14), 3770–3821. DOI: 10.1002/anie.201609079.

(6). Walsh CT Blurring the Lines between Ribosomal and Nonribosomal Peptide Scaffolds. ACS Chem. Biol 2014, 9 (8), 1653–1661. DOI: 10.1021/cb5003587. [PubMed: 24883916]

(7). Yim G; Thaker MN; Koteva K; Wright G Glycopeptide Antibiotic Biosynthesis. J. Antibiot 2014, 67 (1), 31–41. DOI: 10.1038/ja.2013.117.

(8). Guo S; Wang S; Ma S; Deng Z; Ding W; Zhang Q Radical SAM-Dependent Ether Crosslink in Daropeptide Biosynthesis. Nat. Commun 2022, 13 (1), 2361. DOI: 10.1038/s41467-022-30084-2. [PubMed: 35487921]

(9). Burkhart BJ; Hudson GA; Dunbar KL; Mitchell DA A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. Nat. Chem. Biol 2015, 11 (8), 564–570. DOI: 10.1038/nchembio.1856. [PubMed: 26167873]

(10). Shemesh R; Novik A; Cohen Y Follow the Leader: Preference for Specific Amino Acids Directly Following the Initial Methionine in Proteins of Different Organisms. Genomics Proteomics Bioinformatics 2010, 8 (3), 180–189. DOI: 10.1016/S1672-0229(10)60020-4. [PubMed: 20970746]

(11). Montalbán-López M; Scott TA; Ramesh S; Rahman IR; van Heel AJ; Viel JH; Bandarian V; Dittmann E; Genilloud O; Goto Y; Grande Burgos MJ; Hill C; Kim S; Koehnke J; Latham JA; Link AJ; Martínez B; Nair SK; Nicolet Y; Rebuffat S; Sahl H-G; Sareen D; Schmidt EW; Schmitt L; Severinov K; Süssmuth RD; Truman AW; Wang H; Weng J-K; van Wezel GP; Zhang Q; Zhong J; Piel J; Mitchell DA; Kuipers OP; van der Donk WA New Developments in RiPP Discovery, Enzymology and Engineering. Nat. Prod. Rep 2021, 38 (1), 130–239. DOI: 10.1039/d0np00027b. [PubMed: 32935693]

(12). Vinogradov AA; Chang JS; Onaka H; Goto Y; Suga H Accurate Models of Substrate Preferences of Post-Translational Modification Enzymes from a Combination of MRNA Display and Deep Learning. ACS Cent. Sci 2022, 8 (6), 814–824. DOI: 10.1021/acscentsci.2c00223. [PubMed: 35756369]

(13). Vinogradov AA; Nagai E; Chang JS; Narumi K; Onaka H; Goto Y; Suga H Accurate Broadcasting of Substrate Fitness for Lactazole Biosynthetic Pathway from Reactivity-Profiling MRNA Display. J. Am. Chem. Soc 2020. DOI: 10.1021/jacs.0c10374.

(14). Kries H Biosynthetic Engineering of Nonribosomal Peptide Synthetases. J. Pept. Sci 2016, 22 (9), 564–570. DOI: 10.1002/psc.2907. [PubMed: 27465074]

(15). Winn M; Fyans JK; Zhuo Y; Micklefield J Recent Advances in Engineering Nonribosomal Peptide Assembly Lines. Nat. Prod. Rep 2016, 33 (2), 317–347. DOI: 10.1039/c5np00099h. [PubMed: 26699732]

(16). Ruijne F; Kuipers OP Combinatorial Biosynthesis for the Generation of New-to-Nature Peptide Antimicrobials. Biochem. Soc. Trans 2021, 49 (1), 203–215. DOI: 10.1042/BST20200425. [PubMed: 33439248]

(17). Bauman KD; Butler KS; Moore BS; Chekan JR Genome Mining Methods to Discover Bioactive Natural Products. Nat. Prod. Rep 2021, 38 (11), 2100–2129. DOI: 10.1039/d1np00032b. [PubMed: 34734626]

(18). Mordhorst S; Ruijne F; Vagstad AL; Kuipers OP; Piel J Emulating Nonribosomal Peptides with Ribosomal Biosynthetic Strategies. RSC Chem. Biol 2023, 4 (1), 7–36. DOI: 10.1039/d2cb00169a. [PubMed: 36685251]

(19). Chevrette MG; Gavrilidou A; Mantri S; Selem-Mojica N; Ziemert N; Barona-Gómez F The Confluence of Big Data and Evolutionary Genome Mining for the Discovery of Natural Products. Nat. Prod. Rep 2021, 38 (11), 2024–2040. DOI: 10.1039/d1np00013f. [PubMed: 34787598]

(20). Goto Y; Katoh T; Suga H Flexizymes for Genetic Code Reprogramming. Nat. Protoc 2011, 6 (6), 779–790. DOI: 10.1038/nprot.2011.331. [PubMed: 21637198]

(21). Iskandar SE; Haberman VA; Bowers AA Expanding the Chemical Diversity of Genetically Encoded Libraries. ACS Comb. Sci 2020, 22 (12), 712–733. DOI: 10.1021/acscombsci.0c00179. [PubMed: 33167616]

(22). Kawakami T; Murakami H; Suga H Messenger RNA-Programmed Incorporation of Multiple N-Methyl-Amino Acids into Linear and Cyclic Peptides. Chem. Biol 2008, 15 (1), 32–42. DOI: 10.1016/j.chembiol.2007.12.008. [PubMed: 18215771]

(23). Katoh T; Suga H Ribosomal Incorporation of Consecutive β-Amino Acids. J. Am. Chem. Soc 2018, 140 (38), 12159–12167. DOI: 10.1021/jacs.8b07247. [PubMed: 30221942]

(24). Adaligil E; Song A; Cunningham CN; Fairbrother WJ Ribosomal Synthesis of Macrocyclic Peptides with Linear Γ4- and β-Hydroxy-Γ4-Amino Acids. ACS Chem. Biol 2021, 16 (8), 1325–1331. DOI: 10.1021/acschembio.1c00292. [PubMed: 34270222]

(25). Ohta A; Murakami H; Higashimura E; Suga H Synthesis of Polyester by Means of Genetic Code Reprogramming. Chem. Biol 2007, 14 (12), 1315–1322. DOI: 10.1016/j.chembiol.2007.10.015. [PubMed: 18096500]

(26). Takatsuji R; Shinbara K; Katoh T; Goto Y; Passioura T; Yajima R; Komatsu Y; Suga H Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with In Vitro Display. J. Am. Chem. Soc 2019, 141 (6), 2279–2287. DOI: 10.1021/jacs.8b05327. [PubMed: 30648857]

(27). Fujino T; Goto Y; Suga H; Murakami H Reevaluation of the D-Amino Acid Compatibility with the Elongation Event in Translation. J. Am. Chem. Soc 2013, 135 (5), 1830–1837. DOI: 10.1021/ja309570x. [PubMed: 23301668]

(28). Passioura T; Liu W; Dunkelmann D; Higuchi T; Suga H Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code. J. Am. Chem. Soc 2018, 140 (37), 11551–11555. DOI: 10.1021/jacs.8b03367. [PubMed: 30157372]

(29). Kawakami T; Ohta A; Ohuchi M; Ashigai H; Murakami H; Suga H Diverse Backbone-Cyclized Peptides via Codon Reprogramming. Nat. Chem. Biol 2009, 5 (12), 888–890. DOI: 10.1038/nchembio.259. [PubMed: 19915537]

(30). Goto Y; Suga H The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. Acc. Chem. Res 2021, 54 (18), 3604–3617. DOI: 10.1021/acs.accounts.1c00391. [PubMed: 34505781]

(31). Fleming SR; Bartges TE; Vinogradov AA; Kirkpatrick CL; Goto Y; Suga H; Hicks LM; Bowers AA Flexizyme-Enabled Benchtop Biosynthesis of Thiopeptides. J. Am. Chem. Soc 2019, 141 (2), 758–762. DOI: 10.1021/jacs.8b11521. [PubMed: 30602112]

(32). Ricart E; Leclère V; Flissi A; Mueller M; Pupin M; Lisacek F RBAN: Retro-Biosynthetic Analysis of Nonribosomal Peptides. J. Cheminform 2019, 11 (1), 13. DOI: 10.1186/s13321-019-0335-x. [PubMed: 30737579]

(33). Rutz A; Sorokina M; Galgonek J; Mietchen D; Willighagen E; Gaudry A; Graham JG; Stephan R; Page R; Vondrášek J; Steinbeck C; Pauli GF; Wolfender J-L; Bisson J; Allard P-M The LOTUS Initiative for Open Knowledge Management in Natural Products Research. eLife 2022, 11. DOI: 10.7554/eLife.70780.

(34). Banerjee P; Erehman J; Gohlke B-O; Wilhelm T; Preissner R; Dunkel M Super Natural II--a Database of Natural Products. Nucleic Acids Res 2015, 43 (Database issue), D935–9. DOI: 10.1093/nar/gku886. [PubMed: 25300487]

(35). Terlouw BR; Blin K; Navarro-Muñoz JC; Avalon NE; Chevrette MG; Egbert S; Lee S; Meijer D; Recchia MJJ; Reitz ZL; van Santen JA; Selem-Mojica N; Tørring T; Zaroubi L; Alanjary M; Aleti G; Aguilar C; Al-Salihi SAA; Augustijn HE; Avelar-Rivas JA; Avitia-Domínguez LA; Barona-Gómez F; Bernaldo-Agüero J; Bielinski VA; Biermann F; Booth TJ; Carrion Bravo VJ; Castelo-Branco R; Chagas FO; Cruz-Morales P; Du C; Duncan KR; Gavriilidou A; Gayrard D; Gutiérrez-García K; Haslinger K; Helfrich EJN; van der Hooft JJJ; Jati AP; Kalkreuter E; Kalyvas N; Kang KB; Kautsar S; Kim W; Kunjapur AM; Li Y-X; Lin G-M; Loureiro C; Louwen JJR; Louwen NLL; Lund G; Parra J; Philmus B; Pourmohsenin B; Pronk LJU; Rego A; Rex DAB; Robinson S; Rosas-Becerra LR; Roxborough ET; Schorn MA; Scobie DJ; Singh KS; Sokolova N; Tang X; Udwary D; Vigneshwari A; Vind K; Vromans SPJM; Waschulin V; Williams SE; Winter JM; Witte TE; Xie H; Yang D; Yu J; Zdouc M; Zhong Z; Collemare J; Linington RG; Weber T; Medema MH MIBiG 3.0: A Community-Driven Effort to Annotate Experimentally Validated Biosynthetic Gene Clusters. Nucleic Acids Res 2023, 51 (D1), D603–D610. DOI: 10.1093/nar/gkac1049. [PubMed: 36399496]

(36). Vinogradov AA; Suga H Introduction to Thiopeptides: Biological Activity, Biosynthesis, and Strategies for Functional Reprogramming. Cell Chem. Biol 2020, 27 (8), 1032–1051. DOI: 10.1016/j.chembiol.2020.07.003. [PubMed: 32698017]

(37). Probst D; Reymond J-L Visualization of Very Large High-Dimensional Data Sets as Minimum Spanning Trees. J. Cheminform 2020, 12 (1), 12. DOI: 10.1186/s13321-020-0416-x. [PubMed: 33431043]

(38). Pye CR; Hewitt WM; Schwochert J; Haddad TD; Townsend CE; Etienne L; Lao Y; Limberakis C; Furukawa A; Mathiowetz AM; Price DA; Liras S; Lokey RS Nonclassical Size Dependence of Permeation Defines Bounds for Passive Adsorption of Large Drug Molecules. J. Med. Chem 2017, 60 (5), 1665–1672. DOI: 10.1021/acs.jmedchem.6b01483. [PubMed: 28059508]

(39). Ayikpoe RS; Shi C; Battiste AJ; Eslami SM; Ramesh S; Simon MA; Bothwell IR; Lee H; Rice AJ; Ren H; Tian Q; Harris LA; Sarksian R; Zhu L; Frerk AM; Precord TW; van der Donk WA; Mitchell DA; Zhao H A Scalable Platform to Discover Antimicrobials of Ribosomal Origin. Nat. Commun 2022, 13 (1), 6135. DOI: 10.1038/s41467-022-33890-w. [PubMed: 36253467]

(40). Hu YL; Yin FZ; Shi J; Ma SY; Wang ZR; Tan RX; Jiao RH; Ge HM P450-Modified Ribosomally Synthesized Peptides with Aromatic Cross-Links. J. Am. Chem. Soc 2023. DOI: 10.1021/jacs.3c07416.

(41). Dougherty PG; Sahni A; Pei D Understanding Cell Penetration of Cyclic Peptides. Chem. Rev 2019, 119 (17), 10241–10287. DOI: 10.1021/acs.chemrev.9b00008. [PubMed: 31083977]

(42). Stoppacher N; Neumann NKN; Burgstaller L; Zeilinger S; Degenkolb T; Brückner H; Schuhmacher R The Comprehensive Peptaibiotics Database. Chem. Biodivers 2013, 10 (5), 734–743. DOI: 10.1002/cbdv.201200427. [PubMed: 23681723]

(43). Bouaïcha N; Miles CO; Beach DG; Labidi Z; Djabri A; Benayache NY; Nguyen-Quang T Structural Diversity, Characterization and Toxicology of Microcystins. Toxins (Basel) 2019, 11 (12). DOI: 10.3390/toxins11120714.

(44). Köcher S; Resch S; Kessenbrock T; Schrapp L; Ehrmann M; Kaiser M From Dolastatin 13 to Cyanopeptolins, Micropeptins, and Lyngbyastatins: The Chemical Biology of Ahp-Cyclodepsipeptides. Nat. Prod. Rep 2020, 37 (2), 163–174. DOI: 10.1039/c9np00033j. [PubMed: 31451830]

(45). Katoh T; Suga H Ribosomal Incorporation of Negatively Charged D-α- and N-Methyl-l-α-Amino Acids Enhanced by EF-Sep. Philos. Trans. R. Soc. Lond. B Biol. Sci 2023, 378 (1871), 20220038. DOI: 10.1098/rstb.2022.0038.

(46). Fujino T; Goto Y; Suga H; Murakami H Ribosomal Synthesis of Peptides with Multiple β-Amino Acids. J. Am. Chem. Soc 2016, 138 (6), 1962–1969. DOI: 10.1021/jacs.5b12482. [PubMed: 26807980]

(47). Iwane Y; Hitomi A; Murakami H; Katoh T; Goto Y; Suga H Expanding the Amino Acid Repertoire of Ribosomal Polypeptide Synthesis via the Artificial Division of Codon Boxes. Nat. Chem 2016, 8 (4), 317–325. DOI: 10.1038/nchem.2446. [PubMed: 27001726]

(48). Chan AI; Sawant MS; Burdick DJ; Tom J; Song A; Cunningham CN Evaluating Translational Efficiency of Noncanonical Amino Acids to Inform the Design of Druglike Peptide Libraries. ACS Chem. Biol 2023. DOI: 10.1021/acschembio.2c00712.

(49). Goto Y; Murakami H; Suga H Initiating Translation with D-Amino Acids. RNA 2008, 14 (7), 1390–1398. DOI: 10.1261/rna.1020708. [PubMed: 18515548]

(50). Kawakami T; Murakami H; Suga H Ribosomal Synthesis of Polypeptoids and Peptoid-Peptide Hybrids. J. Am. Chem. Soc 2008, 130 (50), 16861–16863. DOI: 10.1021/ja806998v. [PubMed: 19053417]

(51). Kawakami T; Sasaki T; Reid PC; Murakami H Incorporation of Electrically Charged N-Alkyl Amino Acids into Ribosomally Synthesized Peptides via Post-Translational Conversion. Chem. Sci 2014, 5 (3), 887. DOI: 10.1039/c3sc52744a.

(52). Adaligil E; Song A; Hallenbeck KK; Cunningham CN; Fairbrother WJ Ribosomal Synthesis of Macrocyclic Peptides with B2- and B2,3-Homo-Amino Acids for the Development of Natural Product-Like Combinatorial Libraries. ACS Chem. Biol 2021, 16 (6), 1011–1018. DOI: 10.1021/acschembio.1c00062. [PubMed: 34008946]

(53). Katoh T; Suga H Ribosomal Elongation of Aminobenzoic Acid Derivatives. J. Am. Chem. Soc 2020, 142 (39), 16518–16522. DOI: 10.1021/jacs.0c05765. [PubMed: 32946698]

(54). Katoh T; Suga H Consecutive Ribosomal Incorporation of α-Aminoxy/α-Hydrazino Acids with l/d-Configurations into Nascent Peptide Chains. J. Am. Chem. Soc 2021, 143 (45), 18844–18848. DOI: 10.1021/jacs.1c09270. [PubMed: 34731572]

(55). Katoh T; Sengoku T; Hirata K; Ogata K; Suga H Ribosomal Synthesis and de Novo Discovery of Bioactive Foldamer Peptides Containing Cyclic β-Amino Acids. Nat. Chem 2020, 12 (11), 1081–1088. DOI: 10.1038/s41557-020-0525-1. [PubMed: 32839601]

(56). Kuroda T; Huang Y; Nishio S; Goto Y; Suga H Post-Translational Backbone-Acyl Shift Yields Natural Product-like Peptides Bearing Hydroxyhydrocarbon Units. Nat. Chem 2022, 14 (12), 1413–1420. DOI: 10.1038/s41557-022-01065-1. [PubMed: 36329180]

(57). Katoh T; Suga H Ribosomal Elongation of Cyclic γ-Amino Acids Using a Reprogrammed Genetic Code. J. Am. Chem. Soc 2020, 142 (11), 4965–4969. DOI: 10.1021/jacs.9b12280. [PubMed: 32129615]

(58). Iwane Y; Kimura H; Katoh T; Suga H Uniform Affinity-Tuning of N-Methyl-Aminoacyl-TRNAs to EF-Tu Enhances Their Multiple Incorporation. Nucleic Acids Res 2021, 49 (19), 10807–10817. DOI: 10.1093/nar/gkab288. [PubMed: 33997906]

(59). Lee J; Schwieter KE; Watkins AM; Kim DS; Yu H; Schwarz KJ; Lim J; Coronado J; Byrom M; Anslyn EV; Ellington AD; Moore JS; Jewett MC Expanding the Limits of the Second Genetic Code with Ribozymes. Nat. Commun 2019, 10 (1), 5097. DOI: 10.1038/s41467-019-12916-w. [PubMed: 31704912]

(60). Goto Y; Ohta A; Sako Y; Yamagishi Y; Murakami H; Suga H Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. ACS Chem. Biol 2008, 3 (2), 120–129. DOI: 10.1021/cb700233t. [PubMed: 18215017]

(61). Ohshiro Y; Nakajima E; Goto Y; Fuse S; Takahashi T; Doi T; Suga H Ribosomal Synthesis of Backbone-Macrocyclic Peptides Containing γ-Amino Acids. Chembiochem 2011, 12 (8), 1183–1187. DOI: 10.1002/cbic.201100104. [PubMed: 21506233]

(62). Torikai K; Suga H Ribosomal Synthesis of an Amphotericin-B Inspired Macrocycle. J. Am. Chem. Soc 2014, 136 (50), 17359–17361. DOI: 10.1021/ja508648s. [PubMed: 25454254]

(63). Rogers JM; Passioura T; Suga H Nonproteinogenic Deep Mutational Scanning of Linear and Cyclic Peptides. Proc Natl Acad Sci USA 2018, 115 (43), 10959–10964. DOI: 10.1073/pnas.1809901115. [PubMed: 30301798]

(64). Tsutsumi H; Kuroda T; Kimura H; Goto Y; Suga H Posttranslational Chemical Installation of Azoles into Translated Peptides. Nat. Commun 2021, 12 (1), 696. DOI: 10.1038/s41467-021-20992-0. [PubMed: 33514734]

(65). Maini R; Kimura H; Takatsuji R; Katoh T; Goto Y; Suga H Ribosomal Formation of Thioamide Bonds in Polypeptide Synthesis. J. Am. Chem. Soc 2019, 141 (51), 20004–20008. DOI: 10.1021/jacs.9b11097. [PubMed: 31815469]

(66). Goto Y; Ito Y; Kato Y; Tsunoda S; Suga H One-Pot Synthesis of Azoline-Containing Peptides in a Cell-Free Translation System Integrated with a Posttranslational Cyclodehydratase. Chem. Biol 2014, 21 (6), 766–774. DOI: 10.1016/j.chembiol.2014.04.008. [PubMed: 24856821]

(67). Goto Y; Iwasaki K; Torikai K; Murakami H; Suga H Ribosomal Synthesis of Dehydrobutyrine- and Methyllanthionine-Containing Peptides. Chem. Commun 2009, No. 23, 3419–3421. DOI: 10.1039/b904314d.

(68). Goto Y; Suga H In Vitro Biosynthesis of Peptides Containing Exotic Azoline Analogues. Chembiochem 2020, 21 (1–2), 84–87. DOI: 10.1002/cbic.201900521. [PubMed: 31523895]

(69). Yin Y; Ochi N; Craven TW; Baker D; Takigawa N; Suga H De Novo Carborane-Containing Macrocyclic Peptides Targeting Human Epidermal Growth Factor Receptor. J. Am. Chem. Soc 2019, 141 (49), 19193–19197. DOI: 10.1021/jacs.9b09106. [PubMed: 31752491]

(70). Kawakami T; Ogawa K; Hatta T; Goshima N; Natsume T Directed Evolution of a Cyclized Peptoid-Peptide Chimera against a Cell-Free Expressed Protein and Proteomic Profiling of the Interacting Proteins to Create a Protein-Protein Interaction Inhibitor. ACS Chem. Biol 2016, 11 (6), 1569–1577. DOI: 10.1021/acschembio.5b01014. [PubMed: 27010125]

(71). Ad O; Hoffman KS; Cairns AG; Featherston AL; Miller SJ; Söll D; Schepartz A Translation of Diverse Aramid- and 1,3-Dicarbonyl-Peptides by Wild Type Ribosomes in Vitro. ACS Cent. Sci 2019, 5 (7), 1289–1294. DOI: 10.1021/acscentsci.9b00460. [PubMed: 31403077]

(72). Sako Y; Goto Y; Murakami H; Suga H Ribosomal Synthesis of Peptidase-Resistant Peptides Closed by a Nonreducible Inter-Side-Chain Bond. ACS Chem. Biol 2008, 3 (4), 241–249. DOI: 10.1021/cb800010p. [PubMed: 18338852]

(73). Nakajima E; Goto Y; Sako Y; Murakami H; Suga H Ribosomal Synthesis of Peptides with C-Terminal Lactams, Thiolactones, and Alkylamides. Chembiochem 2009, 10 (7), 1186–1192. DOI: 10.1002/cbic.200900058. [PubMed: 19370739]

(74). Yamagishi Y; Ashigai H; Goto Y; Murakami H; Suga H Ribosomal Synthesis of Cyclic Peptides with a Fluorogenic Oxidative Coupling Reaction. Chembiochem 2009, 10 (9), 1469–1472. DOI: 10.1002/cbic.200900021. [PubMed: 19472249]

(75). Kawakami T; Ishizawa T; Murakami H Extensive Reprogramming of the Genetic Code for Genetically Encoded Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics. J. Am. Chem. Soc 2013, 135 (33), 12297–12304. DOI: 10.1021/ja405044k. [PubMed: 23899321]

(76). Lee J; Torres R; Kim DS; Byrom M; Ellington AD; Jewett MC Ribosomal Incorporation of Cyclic β-Amino Acids into Peptides Using in Vitro Translation. Chem. Commun 2020, 56 (42), 5597–5600. DOI: 10.1039/d0cc02121k.

(77). Lee J; Schwarz KJ; Kim DS; Moore JS; Jewett MC Ribosome-Mediated Polymerization of Long Chain Carbon and Cyclic Amino Acids into Peptides in Vitro. Nat. Commun 2020, 11 (1), 4304. DOI: 10.1038/s41467-020-18001-x. [PubMed: 32855412]

(78). Jongkees SAK; Caner S; Tysoe C; Brayer GD; Withers SG; Suga H Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. Cell Chem. Biol 2017, 24 (3), 381–390. DOI: 10.1016/j.chembiol.2017.02.001. [PubMed: 28262556]

(79). Murakami H; Ohta A; Ashigai H; Suga H A Highly Flexible TRNA Acylation Method for Non-Natural Polypeptide Synthesis. Nat. Methods 2006, 3 (5), 357–359. DOI: 10.1038/nmeth877. [PubMed: 16628205]

(80). Xiong H; Reynolds NM; Fan C; Englert M; Hoyer D; Miller SJ; Söll D Dual Genetic Encoding of Acetyl-Lysine and Non-Deacetylatable Thioacetyl-Lysine Mediated by Flexizyme. Angew. Chem. Int. Ed 2016, 55 (12), 4083–4086. DOI: 10.1002/anie.201511750.

(81). Murakami H; Kourouklis D; Suga H Using a Solid-Phase Ribozyme Aminoacylation System to Reprogram the Genetic Code. Chem. Biol 2003, 10 (11), 1077–1084. DOI: 10.1016/j.chembiol.2003.10.010. [PubMed: 14652075]

(82). Morimoto J; Hayashi Y; Suga H Discovery of Macrocyclic Peptides Armed with a Mechanism-Based Warhead: Isoform-Selective Inhibition of Human Deacetylase SIRT2. Angew. Chem. Int. Ed 2012, 51 (14), 3423–3427. DOI: 10.1002/anie.201108118.

(83). Hartman MCT Non-Canonical Amino Acid Substrates of E. Coli Aminoacyl-TRNA Synthetases. Chembiochem 2022, 23 (1), e202100299. DOI: 10.1002/cbic.202100299. [PubMed: 34416067]

(84). Sako Y; Morimoto J; Murakami H; Suga H Ribosomal Synthesis of Bicyclic Peptides via Two Orthogonal Inter-Side-Chain Reactions. J. Am. Chem. Soc 2008, 130 (23), 7232–7234. DOI: 10.1021/ja800953c. [PubMed: 18479111]

(85). Morgan KD; Andersen RJ; Ryan KS Piperazic Acid-Containing Natural Products: Structures and Biosynthesis. Nat. Prod. Rep 2019. DOI: 10.1039/c8np00076j.

(86). Xi N; Alemany LB; Ciufolini MA Elevated Conformational Rigidity in Dipeptides Incorporating Piperazic Acid Derivatives. J. Am. Chem. Soc 1998, 120 (1), 80–86. DOI: 10.1021/ja9729903.

(87). Si Y; Kretsch AM; Daigh LM; Burk MJ; Mitchell DA Cell-Free Biosynthesis to Evaluate Lasso Peptide Formation and Enzyme-Substrate Tolerance. J. Am. Chem. Soc 2021, 143 (15), 5917–5927. DOI: 10.1021/jacs.1c01452. [PubMed: 33823110]

(88). Liu R; Zhang Y; Zhai G; Fu S; Xia Y; Hu B; Cai X; Zhang Y; Li Y; Deng Z; Liu T A Cell-Free Platform Based on Nisin Biosynthesis for Discovering Novel Lanthipeptides and Guiding Their Overproduction In Vivo. Adv Sci (Weinh) 2020, 7 (17), 2001616. DOI: 10.1002/advs.202001616.

(89). Vinogradov AA; Shimomura M; Kano N; Goto Y; Onaka H; Suga H Promiscuous Enzymes Cooperate at the Substrate Level En Route to Lactazole A. J. Am. Chem. Soc 2020, 142 (32), 13886–13897. DOI: 10.1021/jacs.0c05541. [PubMed: 32664727]

(90). Lee J; McIntosh J; Hathaway BJ; Schmidt EW Using Marine Natural Products to Discover a Protease That Catalyzes Peptide Macrocyclization of Diverse Substrates. J. Am. Chem. Soc 2009, 131 (6), 2122–2124. DOI: 10.1021/ja8092168. [PubMed: 19166292]

(91). McIntosh JA; Robertson CR; Agarwal V; Nair SK; Bulaj GW; Schmidt EW Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst. J. Am. Chem. Soc 2010, 132 (44), 15499–15501. DOI: 10.1021/ja1067806. [PubMed: 20961047]

(92). Oueis E; Stevenson H; Jaspars M; Westwood NJ; Naismith JH Bypassing the Proline/Thiazoline Requirement of the Macrocyclase PatG. Chem. Commun 2017, 53 (91), 12274–12277. DOI: 10.1039/c7cc06550g.

(93). Ludewig H; Czekster CM; Oueis E; Munday ES; Arshad M; Synowsky SA; Bent AF; Naismith JH Characterization of the Fast and Promiscuous Macrocyclase from Plant PCY1 Enables the Use of Simple Substrates. ACS Chem. Biol 2018, 13 (3), 801–811. DOI: 10.1021/acschembio.8b00050. [PubMed: 29377663]

(94). Ongpipattanakul C; Nair SK Biosynthetic Proteases That Catalyze the Macrocyclization of Ribosomally Synthesized Linear Peptides. Biochemistry 2018, 57 (23), 3201–3209. DOI: 10.1021/acs.biochem.8b00114. [PubMed: 29553721]

(95). Sarkar S; Gu W; Schmidt EW Applying Promiscuous RiPP Enzymes to Peptide Backbone N-Methylation Chemistry. ACS Chem. Biol 2022. DOI: 10.1021/acschembio.2c00293.

(96). Katoh T; Tajima K; Suga H Consecutive Elongation of D-Amino Acids in Translation. Cell Chem. Biol 2017, 24 (1), 46–54. DOI: 10.1016/j.chembiol.2016.11.012. [PubMed: 28042044]

(97). Katoh T; Iwane Y; Suga H Logical Engineering of D-Arm and T-Stem of TRNA That Enhances d-Amino Acid Incorporation. Nucleic Acids Res 2017, 45 (22), 12601–12610. DOI: 10.1093/nar/gkx1129. [PubMed: 29155943]

(98). Coronado JN; Ngo P; Anslyn EV; Ellington AD Chemical Insights into Flexizyme-Mediated TRNA Acylation. Cell Chem. Biol 2022. DOI: 10.1016/j.chembiol.2022.03.012.

(99). Mikami T; Takao T; Yanagi K; Nakazawa HN (α) Selective Acetylation of Peptides. Mass Spectrom (Tokyo) 2012, 1 (2), A0010. DOI: 10.5702/massspectrometry.A0010. [PubMed: 24349911]

(100). Hakogi T; Taichi M; Katsumura S Synthesis of a Nitrogen Analogue of Sphingomyelin as a Sphingomyelinase Inhibitor. Org. Lett 2003, 5 (16), 2801–2804. DOI: 10.1021/ol034771u. [PubMed: 12889878]

(101). Chung Y-M; El-Shazly M; Chuang D-W; Hwang T-L; Asai T; Oshima Y; Ashour ML; Wu Y-C; Chang F-R Suberoylanilide Hydroxamic Acid, a Histone Deacetylase Inhibitor, Induces the Production of Anti-Inflammatory Cyclodepsipeptides from Beauveria Felina. J. Nat. Prod 2013, 76 (7), 1260–1266. DOI: 10.1021/np400143j. [PubMed: 23822585]

(102). Du F-Y; Zhang P; Li X-M; Li C-S; Cui C-M; Wang B-G Cyclohexadepsipeptides of the Isaridin Class from the Marine-Derived Fungus Beauveria Felina EN-135. J. Nat. Prod 2014, 77 (5), 1164–1169. DOI: 10.1021/np4011037. [PubMed: 24742254]

(103). Houssen WE; Bent AF; McEwan AR; Pieiller N; Tabudravu J; Koehnke J; Mann G; Adaba RI; Thomas L; Hawas UW; Liu H; Schwarz-Linek U; Smith MCM; Naismith JH; Jaspars M An Efficient Method for the in Vitro Production of Azol(in)e-Based Cyclic Peptides. Angew. Chem. Int. Ed 2014, 53 (51), 14171–14174. DOI: 10.1002/anie.201408082.

(104). Ciufolini MA; Xi N Synthesis, Chemistry and Conformational Properties of Piperazic Acids. Chem. Soc. Rev 1998, 27 (6), 437. DOI: 10.1039/a827437z.

(105). Nicolaou KC; Murphy F; Barluenga S; Ohshima T; Wei H; Xu J; Gray DLF; Baudoin O Total Synthesis of the Novel Immunosuppressant Sanglifehrin A. J. Am. Chem. Soc 2000, 122 (16), 3830–3838. DOI: 10.1021/ja994285v.

(106). Yoshida M; Sekioka N; Izumikawa M; Kozone I; Takagi M; Shin-Ya K; Doi T Total Synthesis and Structure Elucidation of JBIR-39: A Linear Hexapeptide Possessing Piperazic Acid and γ-Hydroxypiperazic Acid Residues. Chem. Eur. J 2015, 21 (7), 3031–3041. DOI: 10.1002/chem.201406020. [PubMed: 25524716]

(107). van der Velden NS; Kälin N; Helf MJ; Piel J; Freeman MF; Künzler M Autocatalytic Backbone N-Methylation in a Family of Ribosomal Peptide Natural Products. Nat. Chem. Biol 2017, 13 (8), 833–835. DOI: 10.1038/nchembio.2393. [PubMed: 28581484]

(108). Freeman MF; Helf MJ; Bhushan A; Morinaka BI; Piel J Seven Enzymes Create Extraordinary Molecular Complexity in an Uncultivated Bacterium. Nat. Chem 2017, 9 (4), 387–395. DOI: 10.1038/nchem.2666. [PubMed: 28338684]

(109). Booth J; Alexandru-Crivac C-N; Rickaby KA; Nneoyiegbe AF; Umeobika U; McEwan AR; Trembleau L; Jaspars M; Houssen WE; Shalashilin DV A Blind Test of Computational Technique for Predicting the Likelihood of Peptide Sequences to Cyclize. J. Phys. Chem. Lett 2017, 8 (10), 2310–2315. DOI: 10.1021/acs.jpclett.7b00848. [PubMed: 28475844]

(110). Nagano M; Huang Y; Obexer R; Suga H One-Pot In Vitro Ribosomal Synthesis of Macrocyclic Depsipeptides. J. Am. Chem. Soc 2021. DOI: 10.1021/jacs.1c00466.

(111). Kopp F; Marahiel MA Macrocyclization Strategies in Polyketide and Nonribosomal Peptide Biosynthesis. Nat. Prod. Rep 2007, 24 (4), 735–749. DOI: 10.1039/b613652b. [PubMed: 17653357]
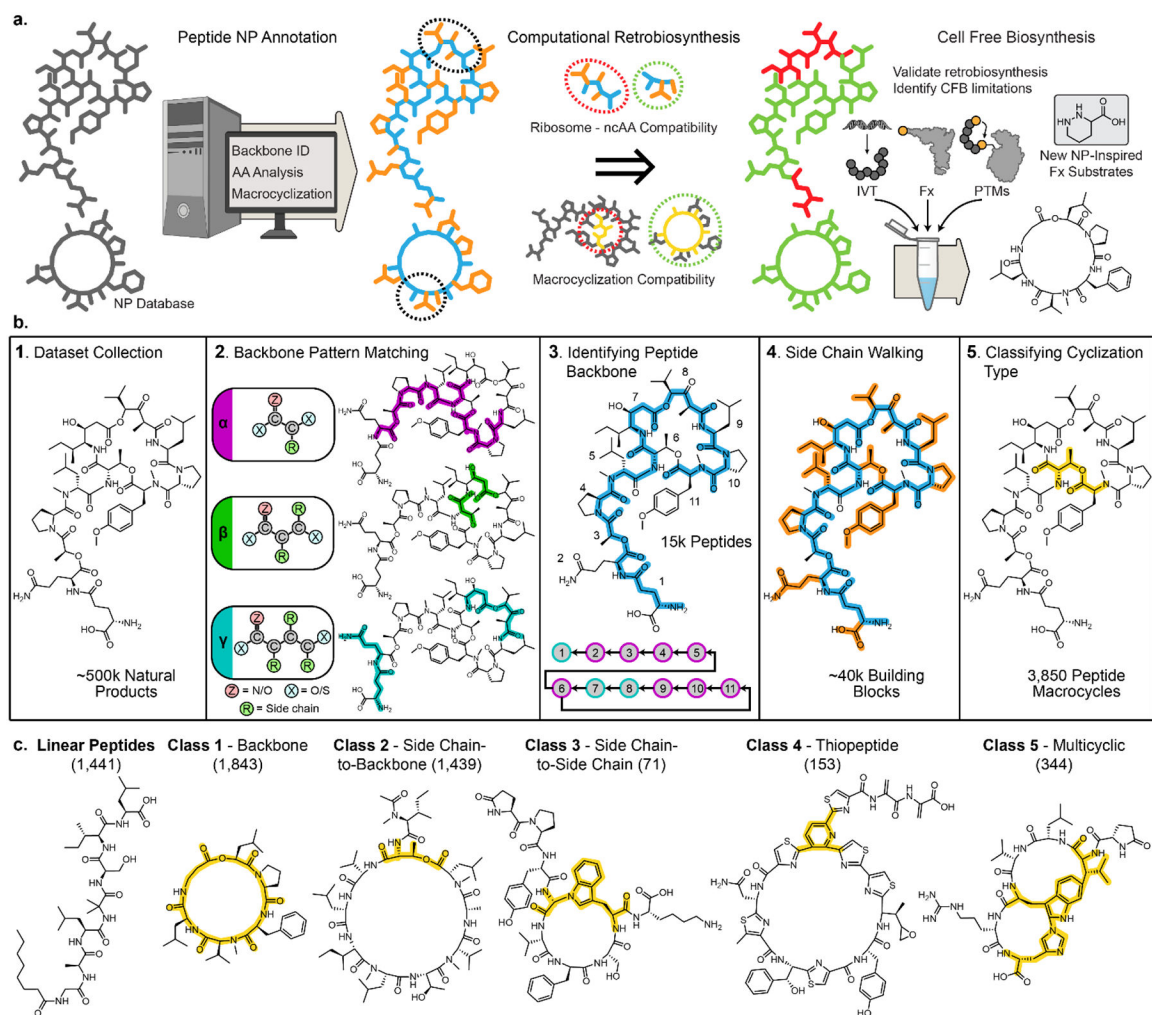
Figure 1.

(a) Overview of workflow for the cheminformatics-guided cell free biosynthesis of macrocyclic peptide natural products. (b) Development of macrocyclic peptide natural product (NP) parsing algorithm. The algorithm proceeds through five key steps beginning with (1) the collection of NP structures from the Lotus, Supernatural II, and MiBiG online datasets. (2) These structures are searched to identify component amino acid building blocks matching the patterns of α-, β- and γ-amino acids. (3) These building blocks are combined using an optimized scoring function described in the SI to identify the longest and simplest peptide backbone. (4) The side chains of each individual building block are parsed to capture information on macrocycle composition. (5) Sites of macrocyclization are identified by connections between the peptide backbone. (c) The NPs are classified based on the type of macrocyclization.
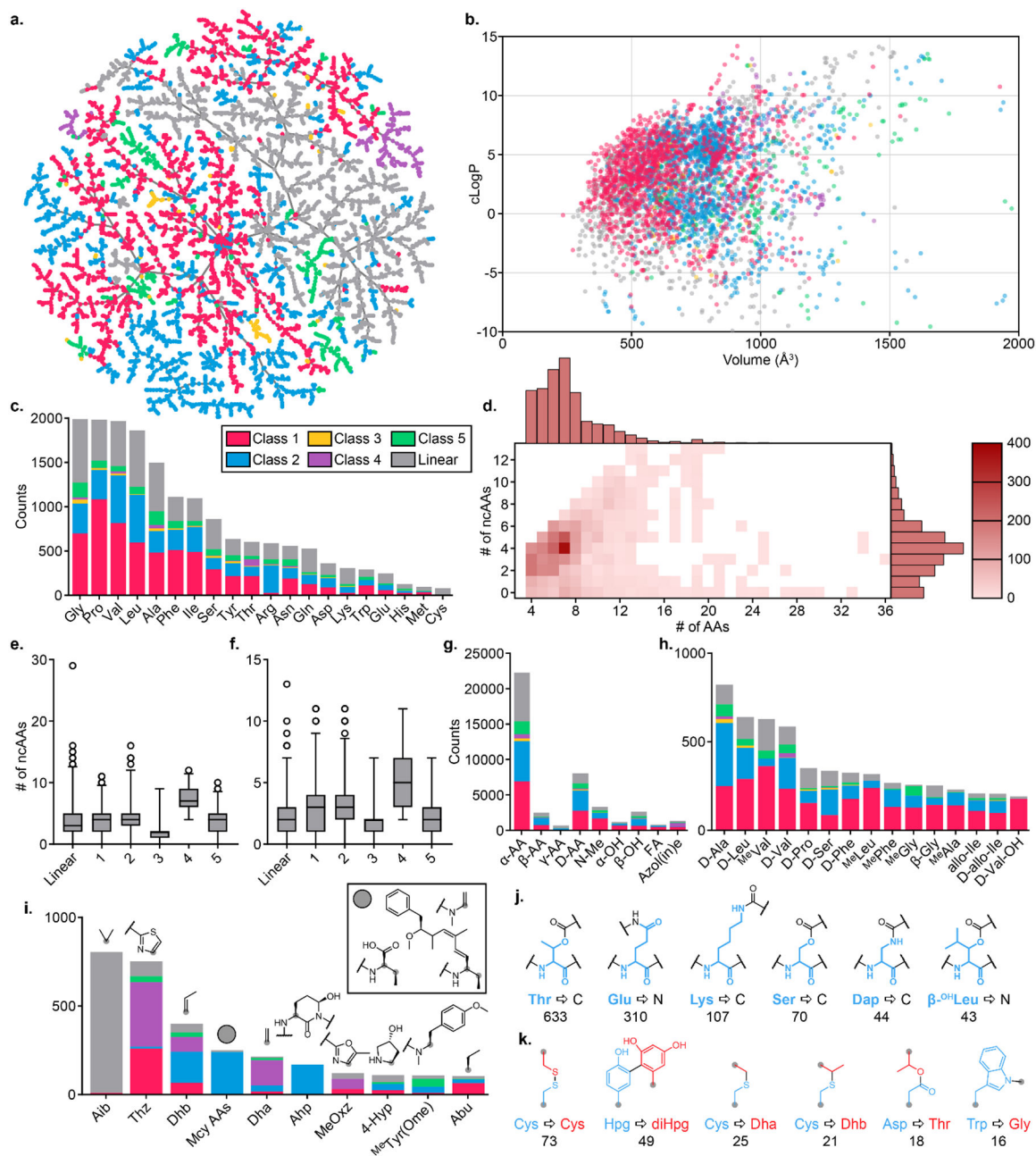
Figure 2.

Computational analysis of peptide natural products. (a) TreeMap visualization of peptide natural product chemical space colored by macrocycle classification. (b) Comparison of the size of macrocycles (volume) to the calculated partition coefficient. (c) Occurrence frequency of the canonical amino acids in peptide natural products. (d) Heatmap of the abundance of non-canonical amino acids (ncAAs) compared to the total number of amino acids in a peptide natural product. (e-f) Box-and-whisker plots of the (e) number of ncAAs and (f) longest consecutive ncAA run in a peptide natural product split up per class. (g-k) Occurrence frequency of (e) ncAA categories, (f) modified canonical amino acids, (g) common non-canonical amino acids, (h) side chain-to-backbone amine (N) or carboxyl (C)

macrocyclization linkages, and (i) side chain-to-side chain macrocyclization linkages in peptide natural products.
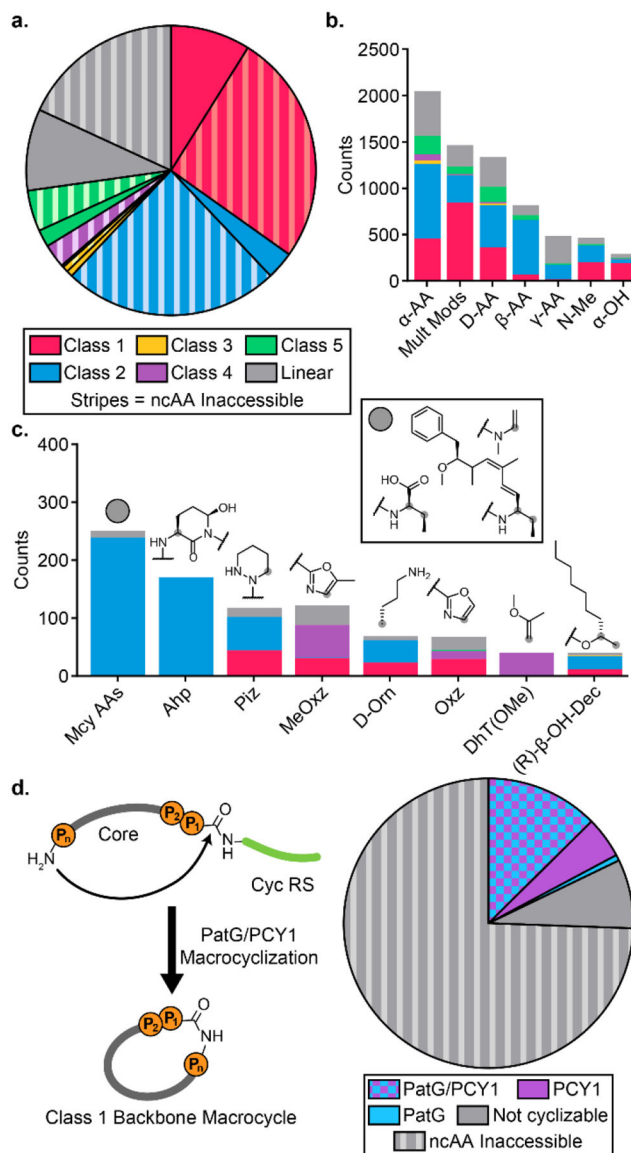
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.
Retrobiosynthetic analysis of macrocycle accessibility for cell free biosynthesis. (a) Pie chart showing the predicted building block accessibility of each peptide natural product class. Predictions are based on the collection of building blocks previously demonstrated to be incorporated using CFB. (b-c) Occurrence frequency of (b) inaccessible ncAA categories and (c) individual inaccessible building blocks. (d) PatG and PCY1 macrocyclization scheme and pie chart showing the predicted macrocyclization accessibility for the Class 1 backbone macrocycles using these enzymes.
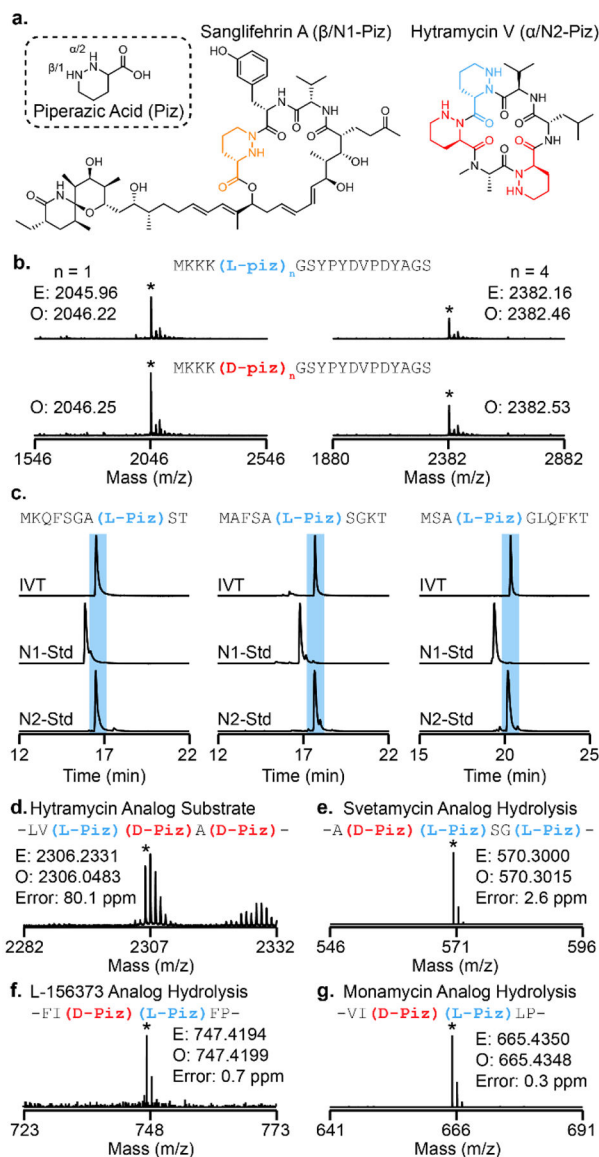
Figure 4.
Expanding cell free biosynthesis capabilities into natural product chemical space by the ribosomal translation of piperazic acid. (a) Structure of piperazic acid (piz) and examples of natural products containing the different regioisomers of piz. (b) MALDI mass spectra showing the optimized incorporation of a single and four consecutive L- and D-piz residues. The optimized translation reaction utilized altered concentrations of translation factors (EF-P, EF-G, EF-Tu, EF-Ts, IF-2). (c) nano-LCMS characterization of the regiochemistry of ribosomal piz translation comparing synthetic standards of both regioisomers to a translation sample (IVT) for three standard peptides. (d) MALDI mass spectrum for the CFB of a Hytramycin piz-natural product analog. This substrate was unable to be modified at all by a backbone macrocyclase enzyme when trying to access the natural backbone macrocyclic structure. (e-g) nano-LCMS mass spectra for the CFB of piz-containing natural product

analogs (e) Svetamycin, (f) L-156373, and (g) Monamycin. Macrocyclization with PCY1 led strictly to hydrolysis in all three cases.

Table 1.

Results of the cell free biosynthesis of a library of predicted accessible peptide natural products.

#	ID	MSA**YPYDVPDY**AS**GIEGR**–XXXXXXXXXX Core	IVT	Acyl	Tier
1	LTS0004249	–GYDPETGTWG	✓		
2	SN00383716	–GF(Hyp)GP	✓		
3	LTS0246876	–YaFEVVG	✓		
4	SN00114511	–YAGF1R	✓		
5	LTS0065646	–HPK(Hyp)NSfW	✓		
6	LTS0039580	–(Ac)TvPYᴹᵉFA	✓	✓	
7	LTS0102086	–(Hex)ELLvDLL	✓	✓	
8	LTS0000231	–(Hex)GALZSTL	✓	✓	
9	SN00327076	–(Hex)TaAAyV	✓	✓	
10	LTS0226082	–(Hex)yS1WR	✓	✓	
11	LTS0224244	–(Hex)fTyRyT	✓	✗	
12	LTS0040024	–(Hex)ZPZZPZZPAZ(βG)AZZ(βG)ZZZAG	✗	✗	

#	ID	MSA**YPYDVPDY**AS**GIEGR**–XXXXXXX–**AYDGE** Core Cyc RS	IVT	Cyc	Tier
13	LTS0001598	–SPFLGTLP–	✓	✓	
14	SN00379332	–GLVMFGVP–	✓	✓	
15	LTS0154672	–AILGGWLA–	✓	✓	
16	SN00345807	–GIVYFP–	✓	✓	
17	LTS0136602	–LFFP**L**P–	✓	✓	
18	LTS0171612	–SSGV**t**WYP–	✓	✓	
19	SN00376531	–MVa**V**y**G**TP–	✓	✓	
20	LTS0105107	–SDFSP**y**SNKP–	✓	–	
21	SN00397245	–G**v**FSIP–	✓	–	
22	SN00235782	–M**t**ATVGYP–	✓	–	
23	LTS0113968	–YDFWK**v**YP–	✓	–	
24	LTS0048526	–WL**v**ING–	✓	–	
25	SN00236278	–QᴹᵉFGFᴹᵉFG–	✓	–	
26	LTS0247550	–(Abu)S(βF)SP–	✓	–	
27	Isaridin 1	–IVᴹᵉA(βG)**L**P–	✓	✓	
28	Isaridin 2	–FVᴹᵉA(βG)**L**P–	✓	✓	
29	Isaridin 3	–FVᴹᵉF(βG)**L**P–	✓	✓	
30	Isaridin 4	–YVᴹᵉF(βG)**L**P–	✓	✓	
31	Isaridin 5	–FᴹᵉVL(βG)**L**P–	✓	✓	
32	Isaridin 6	–FᴹᵉAL(βG)**L**P–	✓	–	
33	LTS0167534	–Lᴹᵉ**L**ALᴹᵉ**L**Aᴹᵉ**I**ᴹᵉ**L**A–	✗	✗	
34	LTS0168356	–Sk(βG)sS(cpG)(Hyp)–	✗	✗	
35	Isaridin 7	–Fᴹᵉ**F**ᴹᵉ**A**(βG)**L**P–	✓	–	
36	Isaridin 8	–Fᴹᵉ**A**ᴹᵉ**F**(βG)**L**P–	✓	–	
37	LTS0040243	–**V**L(Yᴼᴹᵉ)ᴹᵉ**L**TS–	✗	✗	
38	LTS0157285	–(βG)l(Pip)1ᴹᵉ**V**ᴹᵉ**A**–	✗	✗	
39	LTS0027135	–(ᴹᵉYᴼᴹᵉ)Aᴹᵉ**V**ᴹᵉ**L**ᴹᵉ**I**ᴹᵉ**A**ᴹᵉ**G**ᴼᴹᵉ**L**P–	✗	✗	

#	ID	ML**AELSEEALGDA**S**GIEGR**–XXXXXX–**AYDGE** Core Cyc RS			
40	SN00402980	–ATVTI(Thz)–	✓	✓	
41	LTS0246454	–V(Thz)MP(Thz)YP–	✓	✓	
42	SN00365841	–LQ(Thz)G(Thz)IP–	✓	–	
43	LTS0004025	–WWGa(Dha)ᴹᵉGG(Thz)–	✗	✗	

*Green checks indicate the product was detected by EIC, a red "X" indicates the product was not detected, an orange dash indicates hydrolysis was the major species resulting from the macrocyclization reaction. Cyc = cyclization.

**ncAAs are highlighted red; D-AA = lower case; α-

IVT Confidence: High / Medium / Low