J Ind Microbiol Biotechnol (2014) 41:333–344
DOI 10.1007/s10295-013-1357-4

REVIEW

SIMB
Society for Industrial Microbiology and Biotechnology

Prospecting genomes for lasso peptides

Mikhail O. Maksimov · A. James Link

Received: 4 August 2013 / Accepted: 23 September 2013 / Published online: 19 October 2013

Abstract Genome mining has unlocked a veritable treasure chest of natural compounds. However, each family of natural products requires a genome-mining approach tailored to its unique features to be successful. Lasso peptides are ribosomally synthesized and posttranslationally modified products with a unique three-dimensional structure. Advances in the understanding of these molecules have informed the design of strategies to identify new members of the class in sequenced genomes. This review presents the bioinformatic methods used to discover novel lasso peptides and describes how such analyses have afforded insights into the biosynthesis and evolution of this peptide class.

Keywords Lasso peptides · RiPPs · Genome mining

Introduction

With the advent of abundant, freely available genome sequences, the field of natural products discovery has been transformed. The chemistry that endows natural products with their beautiful and complex structures is increasingly being mapped to specific genes [43]. A related activity, genome mining, involves searching genomes for genes or gene clusters indicative of natural product biosynthesis. A milestone in these genomics-based approaches for the study of natural products is certainly the complete sequencing of the *Streptomyces coelicolor* genome, carried out by a team led by Sir David Hopwood [4]. In the decade-plus since the *S. coelicolor* genome was published, genomic tools have become requisite additions to the natural products researcher's toolbox. In honor of Sir Hopwood's 80th birthday this year and his myriad contributions to natural products research, this article will describe how genome-mining approaches have been used to increase our understanding of lasso peptides, a class of ribosomally derived naturally products.

Lasso peptides belong to a large and ever-growing superfamily of natural products arising from ribosomally synthesized polypeptide precursors. This superfamily has recently been given the moniker RiPPs for ribosomally synthesized and posttranslationally modified peptides [1]. RiPPs lend themselves especially to genome mining approaches [41] because of the gene-encoded nature of their precursors. Many RiPPs include extensive posttranslational modifications [10] including chemical moieties reminiscent of modifications generated by non-ribosomal peptide synthetase (NRPS) tailoring enzymes [29]. In contrast to heavily modified RiPPs such as cyanobactins [8] and bottromycins [15] that look little like their ribosomally synthesized precursors, lasso peptides are relatively bereft of posttranslational modifications. However, lasso peptides do fold into a truly remarkable right-handed, threaded three-dimensional structure reminiscent of a slipknot or lasso (Fig. 1). This structure is held together by a single posttranslationally installed isopeptide bond that joins the N-terminus of the matured peptide to a glutamate or aspartate sidechain eight or nine amino acids down the chain. The C-terminal portion of the peptide, ranging from 7 to 15 amino acids, is threaded through this ring resulting in the lasso structure. Diversity in the lasso peptide family comes from the wide variety of

M. O. Maksimov · A. J. Link (✉)
207 Hoyt Laboratory, Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA
e-mail: ajlink@princeton.edu

A. J. Link
Department of Molecular Biology, Princeton University, Princeton, NJ, USA

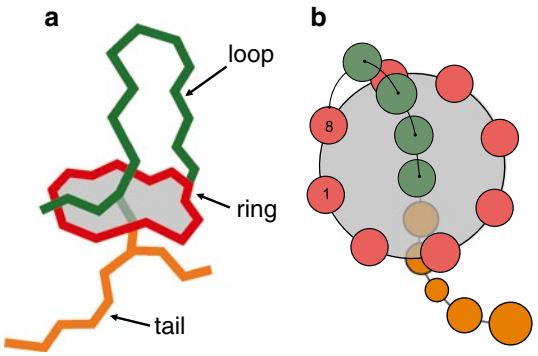

Fig. 1 Lasso peptides have a unique topology resembling a slipknot. **a** An idealized representation of the lasso peptide backbone. Ring, loop and tail portions of the polypeptide are colored *red*, *green* and *orange*, respectively. Residues making up the steric lock are drawn *above* and *below* the plane of the ring (*gray*). Specific sidechain chemical groups on the lock were left off to reflect the diversity of possible steric locks in lasso peptides. **b** *Top* view of an idealized lasso peptide highlights the right-handedness of lasso peptides. Individual residues are represented as *spheres* and coloring is consistent with **a** (color figure online)

sequences (Fig. 2) that exists in family members as well as different lengths and topologies (Fig. 3) that can occur.

Our group has recently written an authoritative review on lasso peptides [27] that describes the broad array of different medicinally relevant activities of known lasso peptides as well as what is known about their structure and biosynthesis. We will touch briefly on these topics here, but most of this article will focus instead on genome-mining approaches to lasso peptide discovery, several of which have appeared quite recently. We will describe how genome mining for new lasso peptides has not only expanded the list of known members of this family, but has also led to new insights into their biological function.

Lasso peptide function

Members of the lasso peptide family have a broad spectrum of natural activities. Several members have antimicrobial [16, 22, 36], anti-viral [7] and anti-metastatic [40] properties, and some function as receptor antagonists [44, 45] and enzyme inhibitors [17, 19]. Antimicrobial activity typically extends to either closely related strains or to organisms competing with the producer strain for resources. Structure–activity relationship studies on MccJ25 have established its mode of action as an RNA polymerase inhibitor. MccJ25 blocks the NTP uptake channel in a "cork-in-bottle" fashion [30], preventing transcription. Recently, RNA polymerase has been shown to also be the target of the lasso peptide capistruin [23]. A more in-depth discussion of the natural activities of lasso peptides can be found in our recent review on the topic [27]. The activities of many of the lasso peptides discovered more recently through genome mining have not been established. However, clues from the genomic context of lasso peptide gene clusters can provide some insights into their function. We will return to this topic later in this review.

Lasso peptide structure

As described in the introduction, lasso peptides get their name from their remarkable threaded structure. Segments of the peptide above and below the ring are referred to as the loop and tail, respectively (Fig. 1a). While all lasso peptides share this general fold, three classes have been defined based on the presence or absence of disulfide bridges (Fig. 2). Class I lasso peptides have two disulfide bonds: one between the N-terminal cysteine and the loop, and the other between the ring and the tail. Class II lasso peptides invariably start with a glycine and lack disulfide bonds. The only class III lasso peptide, BI-32169 [21], contains a single disulfide bond between a cysteine at the C-terminus of the peptide and the ring. All characterized examples of lasso peptides are right-handed, which means that looking into the plane of the ring such that residue number increases in a counter-clockwise direction, the first residue of the loop will rise towards the observer (Fig. 1b). There is as yet no experimental or theoretical evidence to explain the exclusive chirality of lasso peptides [12].

A defining feature of class II lasso peptides is the steric lock, sometimes referred to as the plug, that keeps the tail from escaping the confines of the ring (Fig. 1a). The lock is often a dyad of large amino acids located above and below the ring that define the threading position and consequently the lengths of the loop and the tail. Since class II peptides lack disulfide bonds, they rely entirely on the steric lock to restrain their structures. The fact that various residues have been observed in the steric locks makes it challenging to predict where the lasso peptide is threaded from the sequence (Fig. 2). Whether or not a specific amino acid is able to secure the tail inside the ring can even depend on the threading position itself [13]. Thus, determining the identity of the steric lock is a topic better suited to biochemical, rather than bioinformatic, investigations.

Lasso peptides have variable structural features and amino acid compositions (Fig. 2). Many of the early examples, such as the class I peptides isolated from *Streptomyces* and class II peptides MccJ25, anantin and propeptin, are predominantly hydrophobic. In contrast, many of the recently discovered lasso peptides from proteobacterial species are more polar (bottom half of Fig. 2). While the rings of lasso peptides are either eight or nine residues long, the degree of variability in the relative sizes of loops

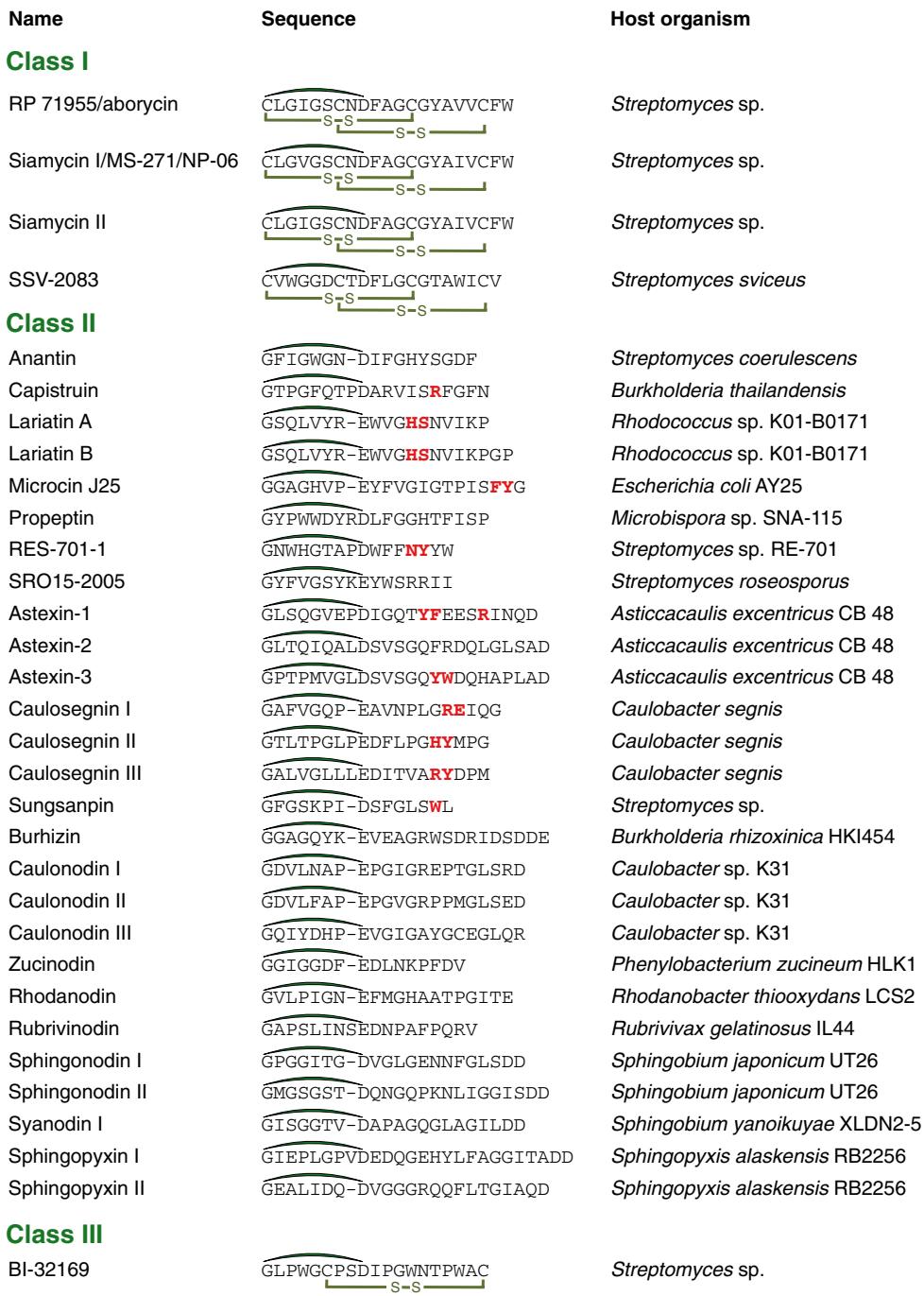

Fig. 2 The lasso peptide family has grown tremendously since the advent of genome mining to total 32 confirmed members. The main feature of the secondary structure of lasso peptides is the macrocycle, represented by a *green arc*, formed between the N-terminal amine and an internal glutamic or aspartic acid residue in position 8 or 9. Disulfide bridges reinforce the structure of class I and class III members, while the tails of class II lasso peptides are held in place by steric locks. Experimentally determined steric locks are *bolded* and shown in *red* (color figure online)

and tails is much greater. Differently sized loops and tails provide most of the topological variability in the lasso peptide class (Fig. 3).

Recalcitrance to thermal, proteolytic and chemical treatment is a hallmark of lasso peptides [35] and a major part of their appeal as a peptide engineering scaffold. These properties are understood to stem from the constrained nature of the lasso fold that prevents the kind of denaturation and aggregation that proteins are so susceptible to. However, recent experiments on the astexins, caulosegnins

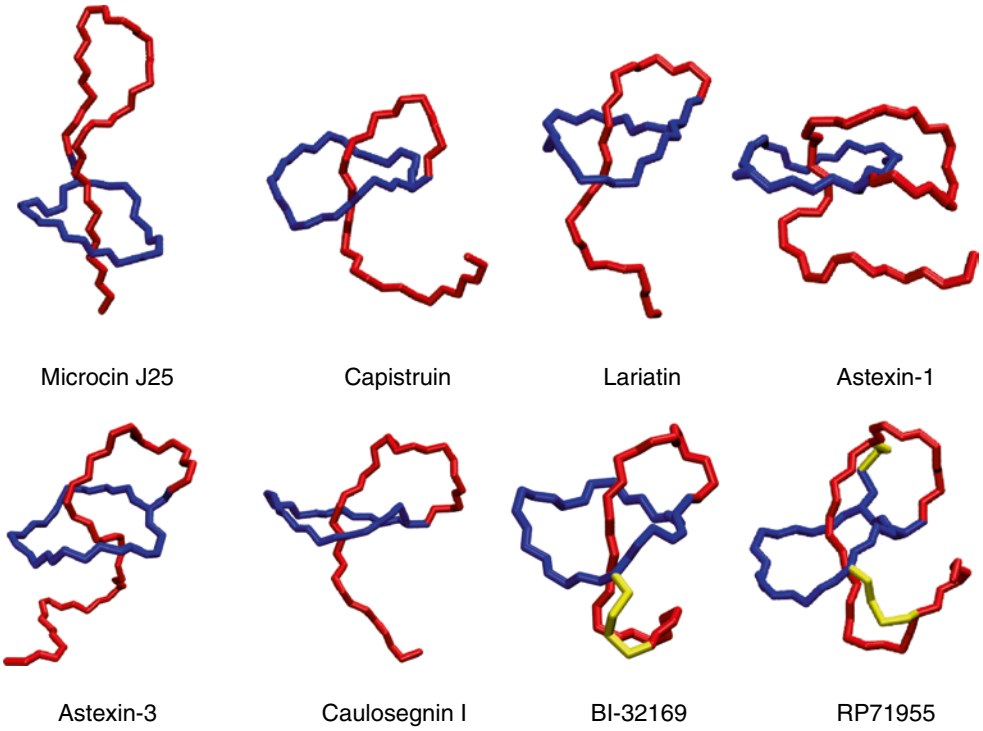

Fig. 3 Molecular models of eight lasso peptides whose solution structure was determined by NMR spectroscopy. Rings are shown in *blue*, disulfide bonds in *yellow* and the rest of the polypeptide chain in *red* (color figure online)

and variants of capistruin have shown that, with regard to thermostability, not all lasso peptides are created equal. Astexins-1 and -2 and caulosegnins I and III can unthread after prolonged heat treatment (50–95 °C). A branched-cyclic (unthreaded lasso) peptide is formed after the tail escapes the ring. The lasso peptide capistruin could be made heat-sensitive through mutagenesis of its steric lock to much smaller amino acids [20]. This highlighted the size of the steric lock relative to the inner diameter of the ring as an important determinant of thermostability. Astexin-3 [26] and caulosegnin II [13], on the other hand, can withstand prolonged heating without unfolding. Five of the 12 newest lasso peptides from proteobacteria [14] were shown to unthread at 95 °C. Interestingly, akin to the producers of the astexins and the caulosegnins, *S. japonicum* and *S. alaskensis* were also found to produce both thermostable and thermolabile lasso peptides.

Lasso peptide biosynthesis and regulation

Three genes are strictly necessary for lasso peptide biosynthesis, and the first gene cluster described for MccJ25 [38, 39] also included a fourth gene encoding an ABC transporter called McjD (Fig. 4). A common "ABCD" nomenclature (e.g., McjABCD for MccJ25 and CapABCD for capistruin) has been recently adopted to consistently describe these clusters [1]. The A gene encodes a precursor peptide that is processed by the B and C maturation enzymes into the lasso fold. Since all lasso peptide gene clusters share these three essential genes, they are commonly referred to simply as the A, B and C homologs. The D gene encodes an ABC transporter that serves as an immunity factor for lasso peptides with antimicrobial activity [32].

As is often observed in the RiPPs superfamily [31], the precursor peptide of lasso peptides contains an N-terminal leader sequence preceding the core peptide (Fig. 5). The B and C homologs act interdependently to process the linear product of the A gene into a lasso. McjB is an ATP-dependent cysteine protease [46] that cleaves off the leader peptide and is distantly related to the transglutaminase family [33]. The McjC enzyme is believed to prime the acid sidechain for cyclization by adenylation and has been shown to be functionally interdependent with McjB [11]. These C enzymes have homology to other enzymes that form amide bonds such as asparagine synthetases and β-lactam synthetases [11, 33]. Due to the difficulties associated with soluble purification of lasso peptide maturation enzymes, the exact sequence of steps in the biosynthesis of lasso peptides and their kinetics remain unknown. However, the ATP dependence of McjB suggests that a step in which the linear precursor peptide is prefolded into the lasso shape may be necessary [46]. No structural information is currently available on any lasso peptide B or C homolog.

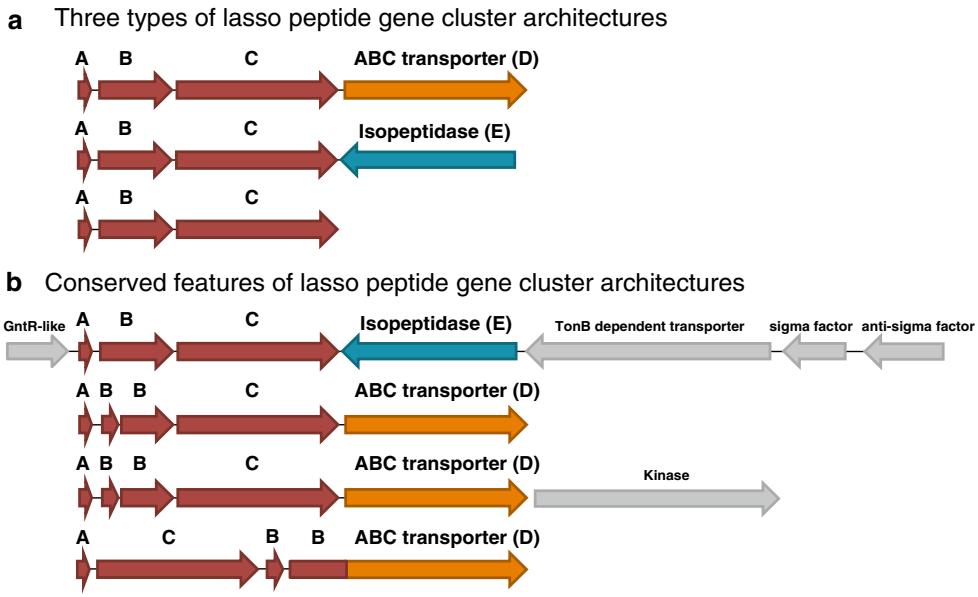

Fig. 4 Lasso peptide gene clusters have been found in three distinct architectures (**a**). Additional conserved features of these architectures have been identified using genome mining techniques (**b**). However, many of the thusly identified conserved genes have not been experimentally linked to lasso peptide biosynthesis and are shown in *gray*. The function of the B and C enzymes is split between two proteins in certain clusters. In other clusters one part of the B protein is fused to the ABC-transporter enzyme

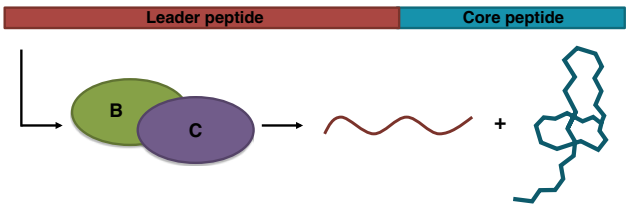

Fig. 5 Schematic of lasso peptide biosynthesis. The B and C maturation enzymes act cooperatively to process the linear precursor into the lasso form. The C enzyme is a cyclase responsible for installing the macrocycle into the core portion of the peptide, liberated from the leader portion by the B protease

Genome mining approaches for lasso peptide discovery

In 2008 with the discovery of capistruin, a lasso peptide made by *Burkholderia thailandensis*, the era of genome mining for lasso peptides was launched [22]. Prior to this, all lasso peptides had been isolated using functional screens and purified from culture broths of the native host. Currently, 21 of the 32 known lasso peptides were identified using genome mining, with many of these examples appearing over the last 2 years. Three distinct approaches have proven successful in attacking this problem, including querying proteins homologous to the maturation enzymes, the precursor-centric approach and mass spectrometry-guided genome mining.

Homology-based genome mining

Initial steps toward finding unknown lasso peptide biosynthesis clusters were taken using McjB and McjC—the only lasso peptide maturation enzymes known at the time—as a template in BLAST searches. Despite the lack of strong homology of McjB and McjC to other enzymes in the global protein database, results of these queries suggested that lasso peptide biosynthesis clusters were present in other bacterial genomes [11, 37]. This analysis informed future lasso peptide genome mining approaches in several ways. First, examination of open reading frames that could encode precursor peptides in the vicinity of the maturation enzymes further reaffirmed that glycine is conserved in the first position of the precursor. Further, the lengths of the precursor and the relative sizes of the leader and core peptides were shown to be consistent despite a low degree of sequence conservation from peptide to peptide.

Prediction of the primary sequence of capistruin and its locus in the genome of *B. thailandensis*, followed by the production of the peptide from both the native host and heterologously in *E. coli*, was the first successful application of lasso peptide genome mining [22]. The precursor peptide for capistruin was not annotated as a putative gene in the global database. The authors identified this locus by manually surveying the upstream flanking region of the McjB homolog gene—CapB. Homologs of McjC and McjD were present downstream of CapB, completing the *capABCD* gene cluster. Due to their short length, precursor peptides often elude automatic gene prediction and annotation algorithms commonly applied to newly sequenced genomes [6]. Capistruin is perhaps the most widely distributed of the identified lasso peptides. Clusters encoding its production can be found in a number of *Burkholderia* strains in addition to *B. thailandensis*, including *B. mallei*, *B. pseudomallei* and *B. oklahomensis* [28].

The strategy of using a McjB-like cysteine protease to nucleate the search for novel lasso peptide biosynthesis

clusters was used again in the discovery of caulosegnins I-III. A homolog of McjB was found in the genome of *Caulobacter segnis*. After confirming the presence of precursor peptides and a McjC-like protein in the putative cluster, production of caulosegnins I and II by *C. segnis* was verified by mass spectrometry, and all three peptides were expressed heterologously in *E. coli*. No ABC transporter has been found with the cluster, and yet production of peptides was detected in both cells and supernatants of *C. segnis*.

Precursor-centric genome mining

Many approaches to genome mining for RiPPs have focused on finding homologs of maturation enzymes arranged in gene cluster architectures associated with their natural product classes. Indeed, this approach successfully identified many new representatives of various RiPP classes [41]. An alternative approach that is both systematic and effective first probes short open reading frames for putative precursor peptides and then surveys the genetic neighborhood for likely maturation enzymes. Application of such a method by our group (Fig. 6) toward lasso peptides identified 79 putative gene clusters out of 3,000+ known genomes at the time of the study. These clusters were distributed across nine bacterial phyla and an archaeal phylum [28]. One of these predicted lasso peptides, astexin-1, was heterologously expressed and its solution structure solved by NMR. This lasso peptide is unusual because of its large size and polar, rather than hydrophobic, composition. It is the first lasso peptide to be isolated from a Gram-negative freshwater organism *Asticcacaulis excentricus*.

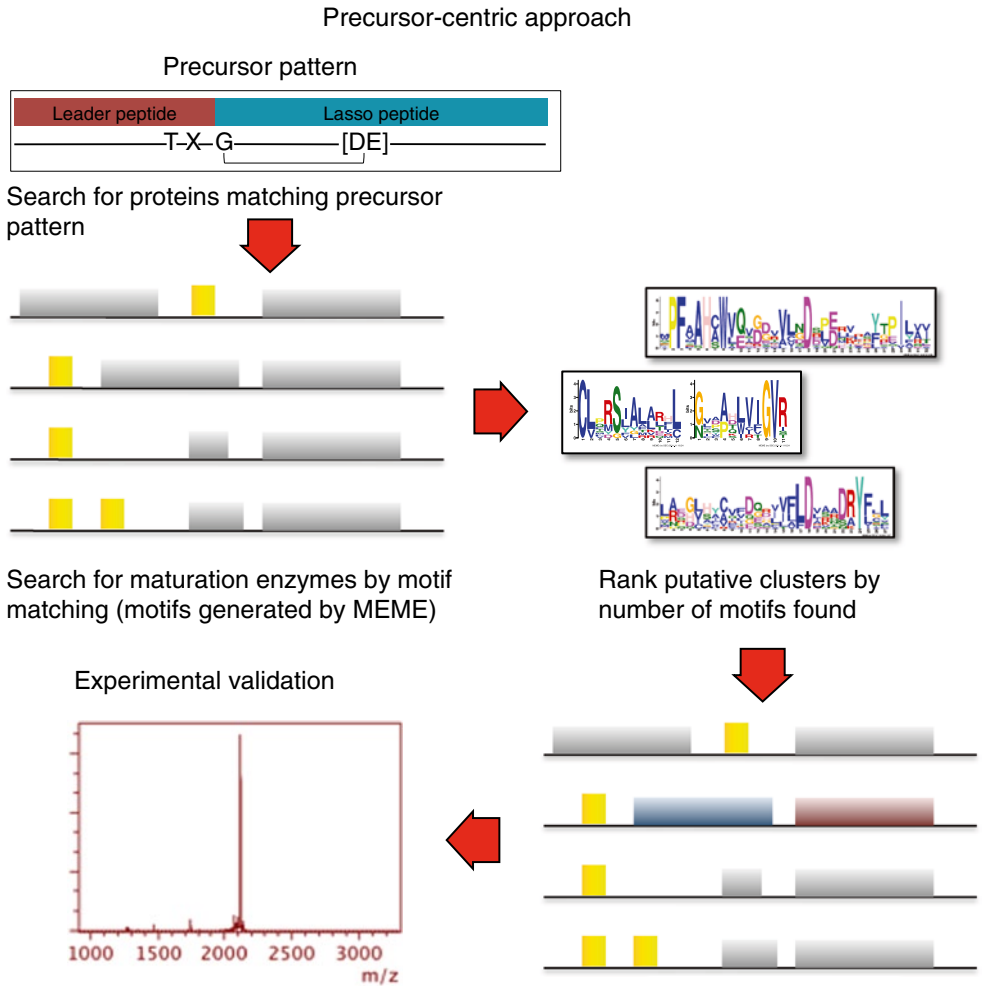

Fig. 6 The precursor-centric approach to lasso peptide genome mining is implemented by first querying all short open reading frames in a bacterial genome for possible lasso peptide precursors using a pattern built from conserved elements in known lasso peptide precursors. The genomic neighborhood of matched precursors is then surveyed for the presence of B and C maturation enzymes using sequence motifs generated using the MEME suite. Best candidate clusters featuring all necessary lasso peptide biosynthesis genes are then selected for experimental validation by expression and mass spectrometric analysis

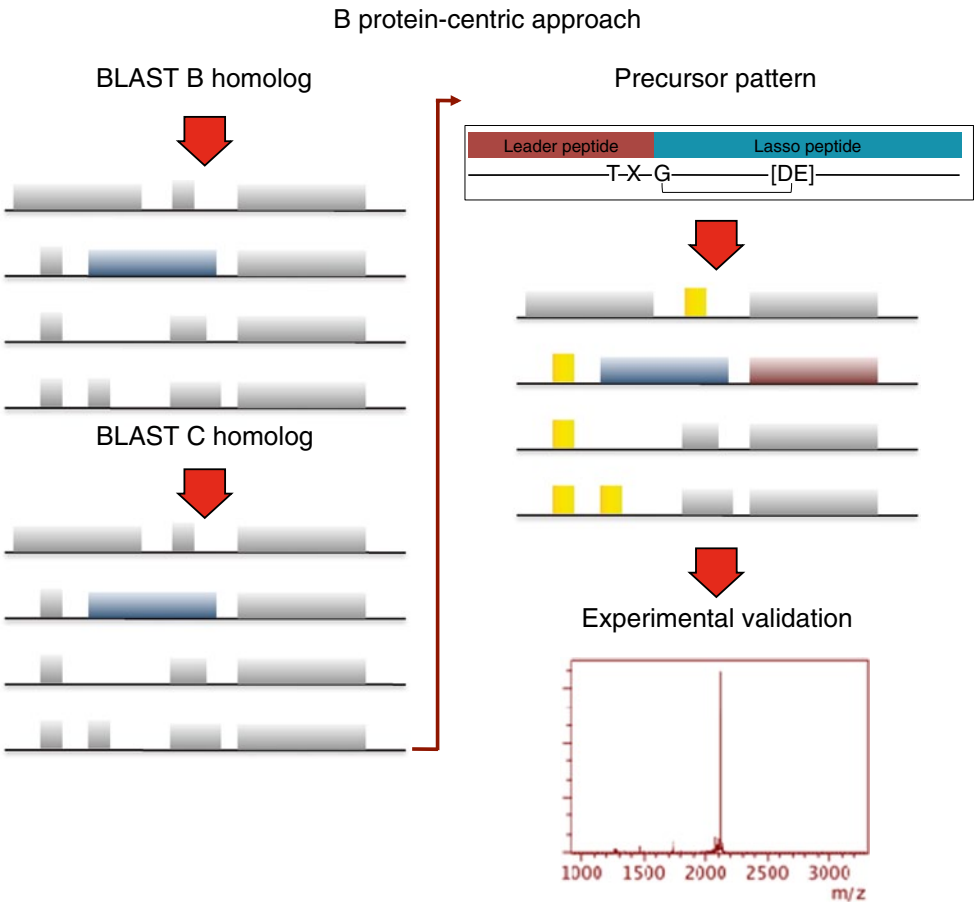

Fig. 7 The B protein-centric method of lasso peptide genome mining uses an iterative BLAST search to find homologs of the B maturation enzyme in bacterial genomes. Precursor peptides and C-homologs are then identified in the vicinity of the putative B-homolog. To validate the likely clusters, putative lasso peptides are then expressed in the native organism and/or heterologously

The crux of the precursor-centric method lies not only in being able to efficiently identify precursor genes, but also in differentiating the false positives from likely precursor hits. To tackle the first task, we constructed a lasso peptide precursor pattern based on their biochemical features. A conserved motif is yet to be identified in lasso peptide leader sequences, and core peptide sequences appear to be hypervariable. Despite this, glycine is highly conserved in the first position of class II lasso peptides and threonine in the penultimate position of their leader peptides. Additionally, there is a requirement for a glutamic or aspartic acid to be present in position 8 or 9 in the ring. These constraints along with sensible restrictions on the lengths of both leader and core peptides were sufficient to significantly narrow the number of putative precursors that needed to be considered. Ranking of putative precursor peptides was based on the presence of short conserved motifs identified in the B and C homologs of lasso peptides using the MEME software suite [2, 3].

An advantage of the precursor-centric approach is that it is highly automated and does not rely on gene annotations. This stands in contrast to integrating results of separate B and C homolog queries to identify putative gene clusters that may or may not have suitable precursor genes. Recent biochemical studies showed that the B homolog of lasso peptides is an ATP-dependent cysteine protease and is thus highly unique to the lasso peptide family [46]. These revelations led Hegemann et al. [14] to implement the reverse of the precursor-centric method and bolster the list of putative lasso peptide gene clusters. This method, titled the B-protein centric approach by the authors (Fig. 7), uses PSI-BLAST to identify lasso B-homologs through an iterative approach and is followed by manual checking for a suitable precursor. In an impressive display of high-throughput validation of genome mining, the authors heterologously expressed 12 novel lasso peptides and characterized them by mass spectrometry. The sequences of the newly validated C gene homologs led them to refine the C-homolog motifs from Maksimov et al. [28].

Mass spectrometry-guided genome mining

The natural product peptidogenomics (NPP) method of genome mining (Fig. 8) uses mass spectrometric data to connect the chemotypes of as yet uncharacterized but expressed natural products to their genotypes [18, 24]. Targets are selected by examining MS spectra of microbial extracts or by MALDI-MS imaging of colonies. The

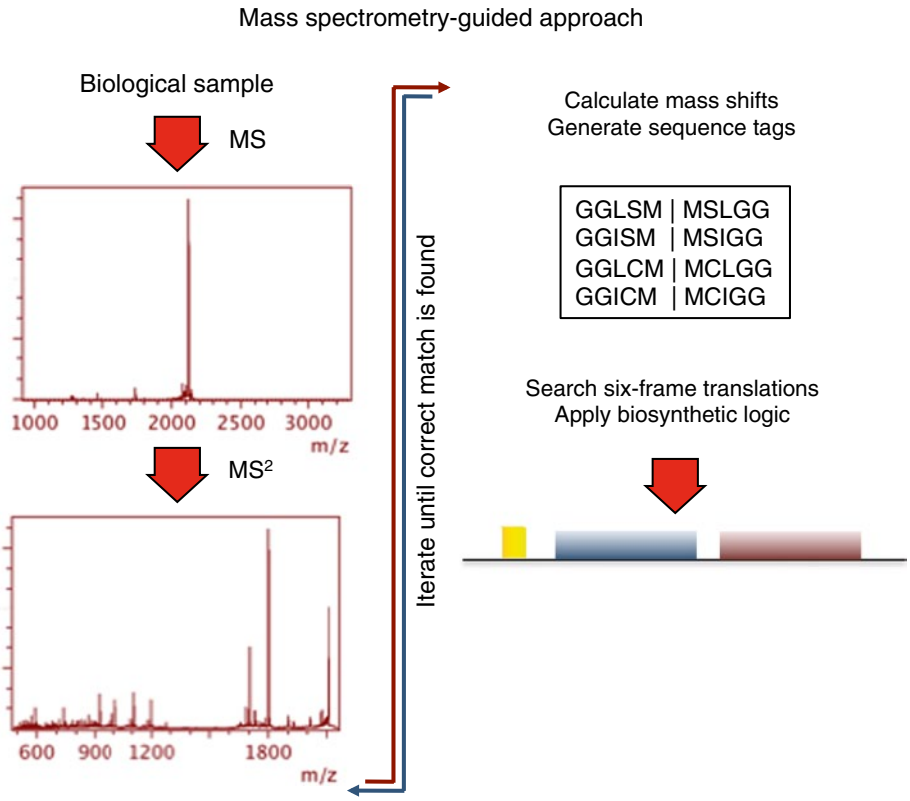

Fig. 8 The mass spectrometry-guided genome mining approach starts with a biological sample such as a fermentation extract, which is subjected to mass spectrometric analysis. Masses in the range characteristic of natural products are then analyzed by tandem mass spectrometry. Short sequence tags are constructed from mass shifts in the MS^2 fragmentation patterns and then iteratively compared to open reading frame translations. Using biosynthetic logic, unknown masses are matched to either ribosomal peptide precursors or non-ribosomal products

products of interest are then analyzed by MS^n during which short peptide sequences present in the product are identified. These are then queried against open reading frame translations or substrates for NRPSs in an iterative way, while applying the biosynthetic logic of known natural products families.

Natural product peptidogenomics has been used to successfully identify both ribosomal and non-ribosomal peptide natural products. Among these are two lasso peptides from *Streptomyces* bacteria: one class I (SSV-2083) and the other class II (SRO15-2005). The sequence of SSV-2083 is significantly different from the highly conserved template of the other three known class I lasso peptides, especially in the ring. At only 15 residues long, SRO15-2005 is a remarkably short lasso peptide. SRO15-2005 was isolated from *n*-butanol extracts of sporulating *Streptomyces roseosporus*, while SSV-2083 was detect by MALDI imaging of sporulating *Streptomyces sviceus*. Their functions remain unknown.

Lessons from genome mining

The first ~17 years of lasso peptide discovery were fueled by functional screens targeting bioactive natural products. This yielded a wealth of information on the structure and function of these molecules and provided some insights into their biosynthesis. Since then, these insights have been reaffirmed or polished by studies performed on new members of the class that were discovered through genome mining. Currently, with 32 confirmed lasso peptides and close to a 100 predicted members of the class having been identified, open questions regarding the distribution, regulation, evolution and in vivo activities of these peptides can begin to be addressed.

Our precursor-centric genome-mining study demonstrated that lasso peptides are widely distributed throughout proteobacterial and actinobacterial phyla [28]. Class I and III lasso peptides have only been identified in actinobacteria, while class II peptides characterized thus far appear primarily in proteobacterial species. In contrast to other RiPPs classes, biosynthesis of lasso peptides is encoded on relatively small gene clusters with several types of architectures.

The existence of multiple gene cluster architectures is one revelation to emerge from genome mining studies (Fig. 4a). Prior to the discovery of astexin-1, all known lasso peptide gene clusters had an associated ABC transporter responsible for exporting its lasso peptide outside of the cell. Minimal "biosynthesis-only" clusters carrying just the precursor gene and the maturation enzymes are observed in genome mining data. However, native or heterologous expression of lasso peptides encoded by them has not been verified experimentally. As an offshoot

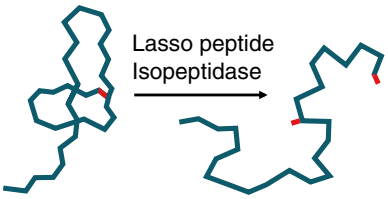

Fig. 9 A lasso peptide isopeptidase linearizes its associated lasso peptide by selectively cleaving the isopeptide bond installed by the B and C maturation enzymes

of our genome mining program, our group has recently reported [26] the discovery of a lasso peptide isopeptidase in the clusters encoding the astexin family of lasso peptides (Fig. 9). Though these clusters lack an ABC transporter, we have labeled this gene "E" to avoid confusion with clusters harboring a D homolog. This enzyme, called AtxE1 in the astexin-1 cluster and AtxE2 in the cluster encoding astexins-2 and -3, specifically hydrolyzes the isopeptide bond installed by the B and C enzymes to unfold the lasso, leaving a linearized core peptide. Both astexin clusters contain an associated peptidase that cannot recognize lasso peptides in the other cluster. The AtxE2 enzyme was purified to homogeneity under native conditions and its kinetic parameters measured. Recognition of the astexins by their respective isopeptidase appears to be based on structure rather than sequence, since AtxE2 cannot hydrolyze thermally unthreaded astexin-2.

Combined, these observations have challenged the premise that the antimicrobial activity of lasso peptides was concomitant with a presence of a dedicated exporter protein for delivering these products into the extracellular medium. Indeed, lasso peptides from isopeptidase containing clusters are either weakly antimicrobial or not at all, which combined with the lack of a self-immunity factor, suggests a bifurcation of activities in the lasso peptide family.

Bioinformatic analysis also revealed several conserved features shared by the three types of lasso peptide gene cluster architectures (Fig. 4b). The isopeptidase clusters share a consistent architecture in which the isopeptidase is followed by a TonB-dependent transporter, and often a sigma/anti-sigma regulatory pair is encoded. A GntR-like transcriptional regulator is often found upstream of the A gene in such clusters. GntR homologs have also been found upstream of lasso peptide clusters without an isopeptidase [14, 26]. The isopeptidase itself is homologous to the prolyl oligopeptidase family of serine proteases, which has been implicated in the processing of other natural products such as amatoxins [25] and the lanthipeptide flavipeptin [42]. Phylogenetic analysis of sequences of B and C homologs of lasso peptide maturation enzymes suggests the existence of two evolutionary clades of lasso peptides (Fig. 10). Clade I is populated by isopeptidase containing lasso peptide clusters, while clade II is made up of the "biosynthesis only" and ABC transporter containing clusters.

Several of the ABC-transporter-containing clusters feature Hpr kinase and sulfotransferase [14], which may carry out further posttranslational modification of their associated lasso peptides (Fig. 4b). Another interesting observation on cluster architectures in the ABC-transporter clade is the rather common occurrence of two separate proteins filling the function performed by a single maturation enzyme in other clusters (Fig. 4b). This phenomenon is hypothesized to have arisen through gene splitting. The biosynthetic cluster of lariatin is an experimentally verified example of this type of architecture. Remarkably, in four putative gene clusters, one of the split maturation enzymes appears to be fused to the ABC transporter (Fig. 4b). Several split C proteins have also been observed. These findings continue to reinforce the idea of cooperativity between lasso peptide maturation enzymes.

The swell in the number of putative and confirmed lasso peptides also allowed for a comparison of their precursors through alignment. Hegemann et al. [14] found that the split between eight- and nine-membered rings is close to even, while a small minority of peptides is predicted to have seven-membered rings. Lasso peptides having such narrow rings have not yet been characterized. Interestingly, all of these peptides are predicted to use the longer glutamate sidechain for cyclization. Among the larger rings, nine-membered rings are preferentially closed by an aspartate, while eight residue rings have no preference. It may be that the nine-member ring cyclized at the glutamate is simply too long to ensure a stable folding of the lasso peptide [9, 13]. In addition to the experimentally verified conserved threonine in the penultimate position of the precursor peptide [5, 34], clade I lasso peptide precursors contain a conspicuous hydrophobic patch in the middle of the leader peptide. While the core peptide sequences are quite variable, most clade I lasso peptides are polar, negatively charged and terminate with one or more aspartic or glutamic acid residue. Conversely, clade II lasso peptides contain moderately conserved prolines in their leader peptides, and their core peptides are mostly hydrophobic with patches of positive charge.

Conclusion: new directions enabled by lasso peptide genome mining

The application of genome mining techniques has clearly opened up the field of lasso peptide discovery. However, there is much more to be learned about this curious class of natural products. For example, recent studies from our group as well as from Marahiel and coworkers have relied on heterologous expression of the lasso peptides predicted

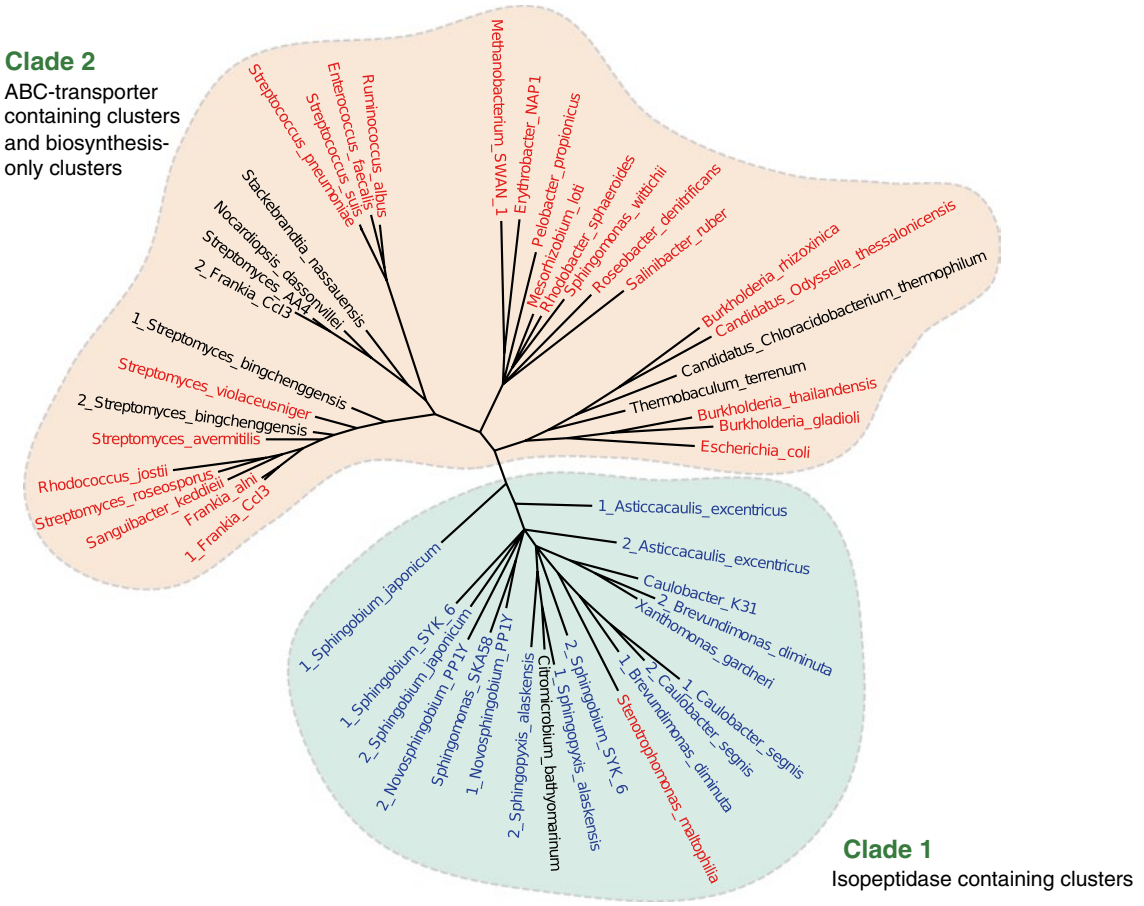

Fig. 10 The phylogenetic tree of known and putative lasso peptide B proteins is split into two distinct clades. Clade I is made up of isopeptidase-containing clusters (*blue*), while clade II consists of clusters with ABC transporters (*red*) and biosynthesis-only clusters (*black*) (color figure online)

from genome mining. It is often the case that the native producers of lasso peptides produce no or very small amounts of the natural product necessitating heterologous production with artificial promoters and ribosome-binding sites [13, 14, 28, 34]. The in vivo regulation of lasso peptides is a topic that has received little consideration so far. In isopeptidase-containing clusters, the GntR homolog and the pair of sigma/anti-sigma regulatory proteins could very well tune the level of biosynthesis of the lasso peptide, as well as the isopeptidase. GntR homologs found in ABCD-type clusters might serve the same purpose. Further, inverted DNA repeats with eight or nine base pair stems have been identified between the A and B genes of a large number of lasso peptide clusters so far. Replacement of these DNA hairpins with optimized ribosome-binding-site sequences can favorably affect the production yields of lasso peptides under heterologous expression in some cases [28], but in others the hairpins have no effect on production [26].

Trading conventional function-based natural product discovery for genome mining is both a blessing, as we have described above, and a curse. The curse side of the equation comes from the fact that genome-mining approaches do not include functional information a priori. Indeed, most of the new lasso peptides discovered over the past 2 years reside squarely in the "function unknown" category. However, careful examination of the genomic context of lasso peptide gene clusters can provide some insights into the native function of natural products. We have hypothesized that the astexins (and by inference potentially all of the clade I lasso peptides) may be functioning as scavenging molecules. Such molecules could be sent out by a producing cell to pick up a cargo, be imported by the associated TonB-dependent transporter and then be hydrolyzed by the isopeptidase to release the cargo.

Another exciting area of future research that is enabled by growing genome-mining data is the possibility of using the conserved motifs identified in lasso peptide maturation enzymes to develop tools for metagenome and microbiome mining. This line of research could lead to the discovery of lasso peptides with new antibiotic activities and

would not rely on having sequenced genomes. Indeed, we view the process of genome mining for lasso peptides as an inherently cyclic process in which large-scale genome searching leads to new testable hypotheses that answer questions about lasso peptide biology. These new morsels of information can of course be fed back into genome mining algorithms to improve them and make them even more fruitful. Today, when it is nearly impossible to think about natural products research without the context of genomics, it is important to remember that these approaches are possible because of the exceptional creativity and hard work by Sir Hopwood and the other researchers that pushed natural products research into the genomic era.

Acknowledgments Research in our laboratory on lasso peptides has been supported by the NSF and Project X. AJL is a DuPont Young Professor and a Sloan Research Fellow.

References

1. 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, Goransson 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, Muller 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, Sussmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30(1):108–160
2. Bailey TL, Elkan C (1995) Unsupervised learning of multiple motifs in biopolymers using expectation maximization. Mach Learn 21(1–2):51–80
3. Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14(1):48–54
4. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147. doi:10.1038/417141a
5. Cheung WL, Pan SJ, Link AJ (2010) Much of the microcin J25 leader peptide is dispensable. J Am Chem Soc 132:2514–2515
6. de Jong A, van Hijum S, Bijlsma JJE, Kok J, Kuipers OP (2006) BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res 34:W273–W279. doi:10.1093/nar/gkl237
7. Detlefsen DJ, Hill SE, Volk KJ, Klohr SE, Tsunakawa M, Furumai T, Lin PF, Nishio M, Kawano K, Oki T, Lee MS (1995) Siamycins I and II, new Anti-HIV-1 peptides. 2. Sequence analysis and structure determination of siamycin I. J Antibiot 48(12):1515–1517
8. Donia MS, Ravel J, Schmidt EW (2008) A global assembly line for cyanobactins. Nat Chem Biol 4(6):341–343. doi:10.1038/nchembio.84
9. Ducasse R, Yan KP, Goulard C, Blond A, Li YY, Lescop E, Guittet E, Rebuffat S, Zirah S (2012) Sequence determinants governing the topology and biological activity of a lasso peptide, microcin J25. ChemBioChem 13(3):371–380. doi:10.1002/cbic.201100702
10. Dunbar KL, Mitchell DA (2013) Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. ACS Chem Biol 8(3):473–487. doi:10.1021/cb3005325
11. Duquesne S, Destoumieux-Garzón D, Zirah S, Goulard C, Peduzzi J, Rebuffat S (2007) Two enzymes catalyze the maturation of a lasso peptide in *Escherichia coli*. Chem Biol 14:793–803
12. Ferguson AL, Zhang SY, Dikiy I, Panagiotopoulos AZ, Debenedetti PG, Link AJ (2010) An experimental and computational investigation of spontaneous lasso formation in microcin J25. Biophys J 99:3056–3065
13. Hegemann JD, Zimmermann M, Xie XL, Marahiel MA (2013) Caulosegnins I–III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. J Am Chem Soc 135(1):210–222. doi:10.1021/ja308173b
14. Hegemann JD, Zimmermann M, Zhu S, Klug D, Marahiel MA (2013) Lasso peptides from proteobacteria: genome mining employing heterologous expression and mass spectrometry. Biopolymers. doi:10.1002/bip.22326
15. Huo LJ, Rachid S, Stadler M, Wenzel SC, Muller R (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. Chem Biol 19(10):1278–1287. doi:10.1016/j.chembiol.2012.08.013
16. Iwatsuki M, Tomoda H, Uchida R, Gouda H, Hirono S, Omura S (2006) Lariatins, antimycobacterial peptides produced by Rhodococcus sp K01-B0171, have a lasso structure. J Am Chem Soc 128(23):7486–7491
17. Katahira R, Yamasaki M, Matsuda Y, Yoshida M (1996) MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from *Streptomyces* sp. 2. Solution structure of MS-271: characteristic features of the 'lasso' structure. Bioorg Med Chem 4(1):121–129. doi:10.1016/0968-0896(95)00176-x
18. Kersten RD, Yang Y-L, Xu Y, Cimermancic P, Nam S-J, Fenical W, Fischbach MA, Moore BS, Dorrestein PC (2011) A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat Chem Biol 7(11):794–802
19. Kimura KI, Kanou F, Takahashi H, Esumi Y, Uramoto M, Yoshihama M (1997) Propeptin, a new inhibitor of prolyl endopeptidase produced by *Microbispora*. 1. Fermentation, isolation and biological properties. J Antibiot 50(5):373–378
20. Knappe TA, Linne U, Robbel L, Marahiel MA (2009) Insights into the biosynthesis and stability of the lasso peptide capistruin. Chem Biol 16(12):1290–1298. doi:10.1016/j.chembiol.2009.11.009
21. Knappe TA, Linne U, Xie XL, Marahiel MA (2010) The glucagon receptor antagonist BI-32169 constitutes a new class of lasso peptides. FEBS Lett 584(4):785–789. doi:10.1016/j.febslet.2009.12.046
22. Knappe TA, Linne U, Zirah S, Rebuffat S, Xie XL, Marahiel MA (2008) Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. J Am Chem Soc 130(34):11446–11454. doi:10.1021/ja802966g
23. Kuznedelov K, Semenova E, Knappe TA, Mukhamedyarov D, Srivastava A, Chatterjee S, Ebright RH, Marahiel MA, Severinov K (2011) The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase. J Mol Biol 412:842–848

24. Liu WT, Kersten RD, Yang YL, Moore BS, Dorrestein PC (2011) Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in *Streptomyces roseosporus*. J Am Chem Soc 133(45):18010–18013. doi:10.1021/ja2040877
25. Luo H, Hallen-Adams HE, Walton JD (2009) Processing of the phalloidin proprotein by prolyl oligopeptidase from the mushroom *Conocybe albipes*. J Biol Chem 284(27):18070–18077. doi:10.1074/jbc.M109.006460
26. Maksimov MO, Link AJ (2013) Discovery and characterization of an isopeptidase that linearizes lasso peptides. J Am Chem Soc. doi:10.1021/ja4054256
27. Maksimov MO, Pan SJ, Link AJ (2012) Lasso peptides: structure, function, biosynthesis, and engineering. Nat Prod Rep 29:996–1006
28. Maksimov MO, Pelczer I, Link AJ (2012) Precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci USA 109(38):15223–15228
29. McIntosh JA, Donia MS, Schmidt EW (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. Nat Prod Rep 26(4):537–559. doi:10.1039/b714132g
30. Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH (2004) Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. Mol Cell 14(6):739–751
31. Oman TJ, van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol 6(1):9–18. doi:10.1038/nchembio.286
32. Pan SJ, Link AJ (2011) Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. J Am Chem Soc 133:5016–5023
33. Pan SJ, Rajniak J, Cheung WL, Link AJ (2012) Construction of a single polypeptide that matures and exports the lasso peptide microcin J25. ChemBioChem 13(3):367–370
34. Pan SJ, Rajniak J, Maksimov MO, Link AJ (2012) The role of a conserved threonine residue in the leader peptide of lasso peptide precursors. Chem Commun 48(13):1880–1882
35. Rosengren KJ, Blond A, Afonso C, Tabet JC, Rebuffat S, Craik DJ (2004) Structure of thermolysin cleaved microcin J25: extreme stability of a two-chain antimicrobial peptide devoid of covalent links. Biochemistry 43(16):4696–4702. doi:10.1021/bi0361261
36. Salomon RA, Farias RN (1992) Microcin-25, a novel antimicrobial peptide produced by *Escherichia coli*. J Bacteriol 174(22):7428–7435
37. Severinov K, Semenova E, Kazakov A, Kazakov T, Gelfand MS (2007) Low-molecular-weight post-translationally modified microcins. Mol Microbiol 65(6):1380–1394. doi:10.1111/j.1365-2958.2007.05874.x
38. Solbiati JO, Ciaccio M, Farias RN, Gonzalez-Pastor JE, Moreno F, Salomon RA (1999) Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. J Bacteriol 181(8):2659–2662
39. Solbiati JO, Ciaccio M, Farias RN, Salomon RA (1996) Genetic analysis of plasmid determinants for microcin J25 production and immunity. J Bacteriol 178(12):3661–3663
40. Um S, Kim YJ, Kwon H, Wen H, Kim SH, Kwon HC, Park S, Shin J, Oh DC (2013) Sungsanpin, a lasso peptide from a deep-sea streptomycete. J Nat Prod 76(5):873–879. doi:10.1021/np300902g
41. Velasquez JE, van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. Curr Opin Chem Biol 15(1):11–21. doi:10.1016/j.cbpa.2010.10.027
42. Voller GH, Krawczyk B, Ensle P, Sussmuth RD (2013) Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. J Am Chem Soc 135(20):7426–7429. doi:10.1021/ja402296m
43. Walsh CT, Fischbach MA (2010) Natural products version 2.0: connecting genes to molecules. J Am Chem Soc 132(8):2469–2493. doi:10.1021/ja909118a
44. Weber W, Fischli W, Hochuli E, Kupfer E, Weibel EK (1991) Anantin—a peptide antagonist of the atrial-natriuretic-factor (ANF) 1. Producing organism, fermentation, isolation and biological activity. J Antibiot 44(2):164–171
45. Wyss DF, Lahm HW, Manneberg M, Labhardt AM (1991) Anantin—a peptide antagonist of the atrial-natriuretic factor (ANF) 2. Determination of the primary sequence by NMR on the basis of proton assignments. J Antibiot 44(2):172–180
46. Yan KP, Li YY, Zirah S, Goulard C, Knappe TA, Marahiel MA, Rebuffat S (2012) Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. ChemBioChem 13(7):1046–1052. doi:10.1002/cbic.201200016