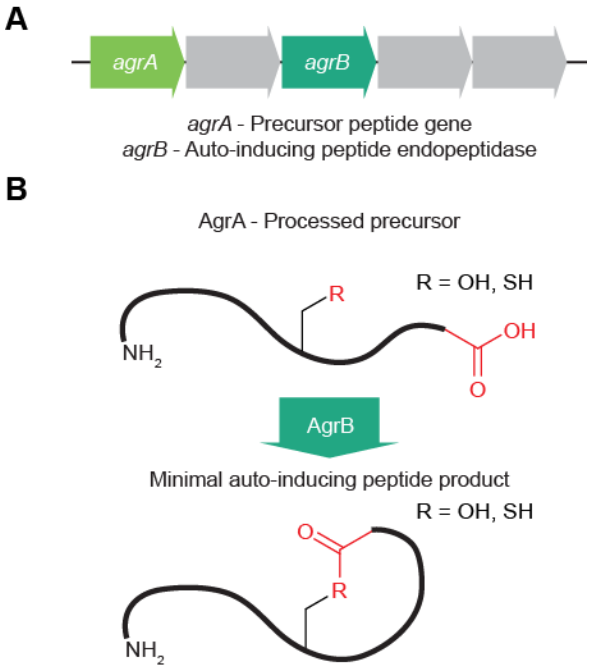
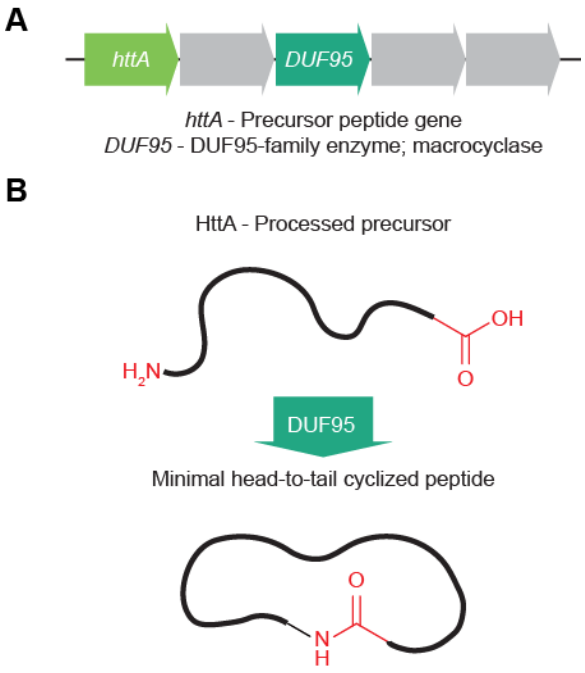
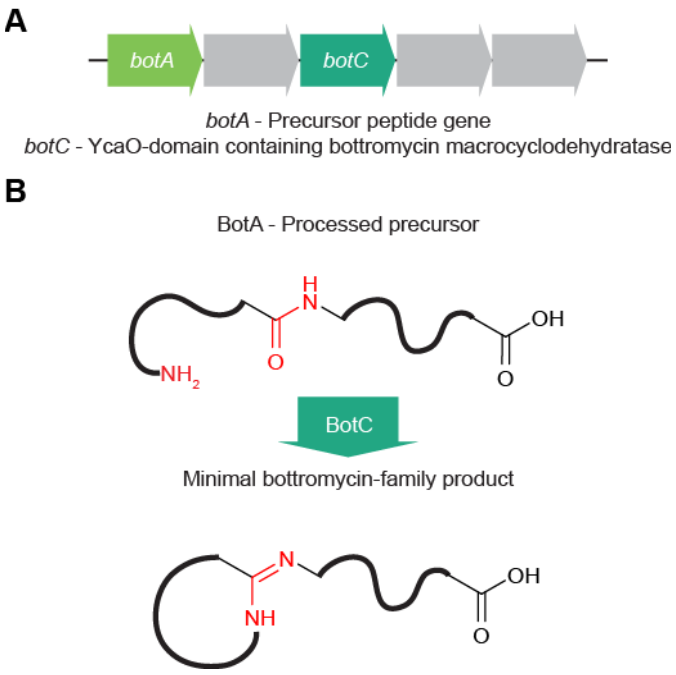
Dataset S7. Related to SI Appendix, Table S1.



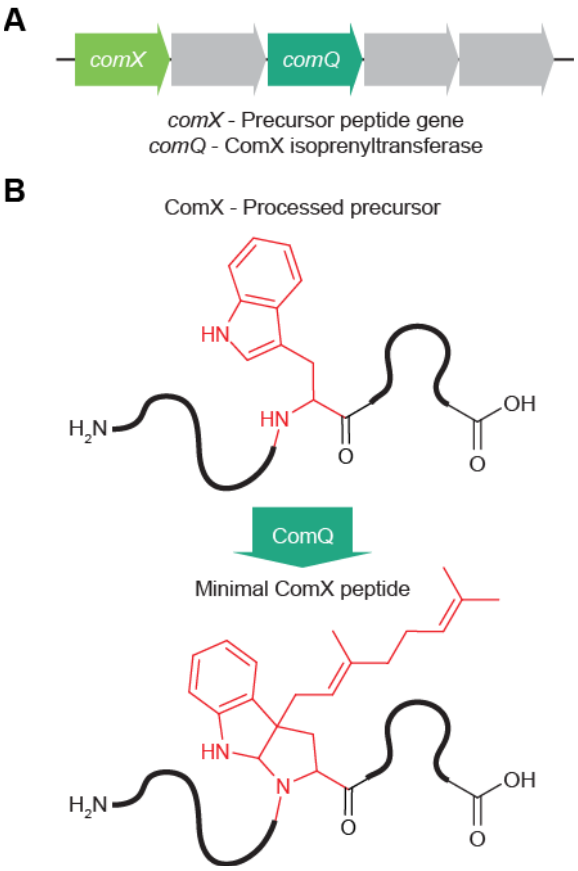
Dataset S7, Fig. 1. Minimal requirements for the detection and prediction of auto-inducing peptide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *agrA* and endopeptidase gene *agrB* for defining auto-inducing peptide gene clusters. (*B*) Detection of the minimal requirements of an auto-inducting peptide gene cluster results in a structure prediction wherein the processed precursor peptide is transformed by the AgrB endopeptidase, forming an intramolecular ester- or thioester-linked macrocycle, based on the presence of serine or cysteine, respectively.



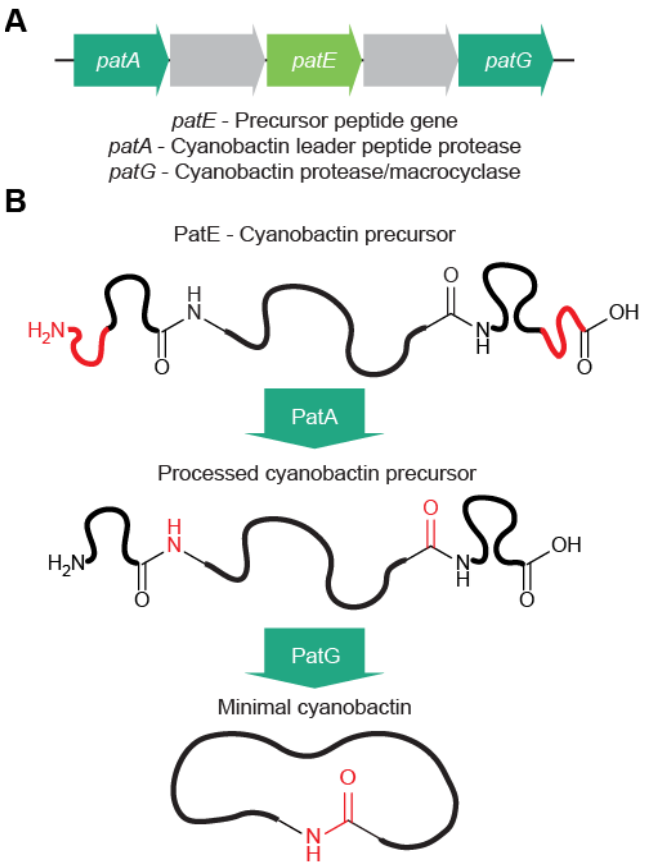
Dataset S7, Fig. 2. Minimal requirements for the detection and prediction of bacterial head-to-tail cyclized peptide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *httA* and macrocyclase gene *DUF95* for defining bacterial head-to-tail cyclized peptide gene clusters. (*B*) Detection of the minimal requirements of a bacterial head-to-tail cyclized peptide gene cluster results in a structure prediction wherein the processed precursor peptide is transformed by the DUF95 macrocyclase, forming an intramolecular amide-linked macrocycle.



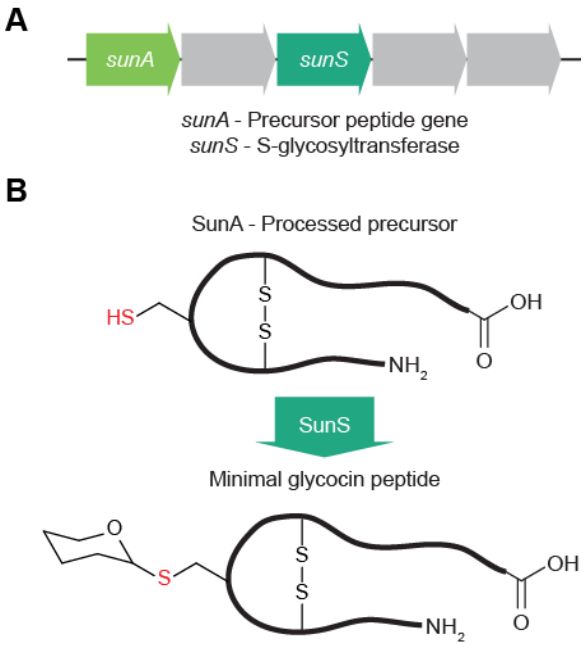
Dataset S7, Fig. 3. Minimal requirements for the detection and prediction of bottromycin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *botA* and macrocyclodehydratase gene *botC* for defining bottromycin gene clusters. (*B*) Detection of the minimal requirements of a bottromycin gene cluster results in a structure prediction wherein the processed precursor peptide is transformed by the BotC macrocyclodehydratase, forming an intramolecular macrolactamidine.



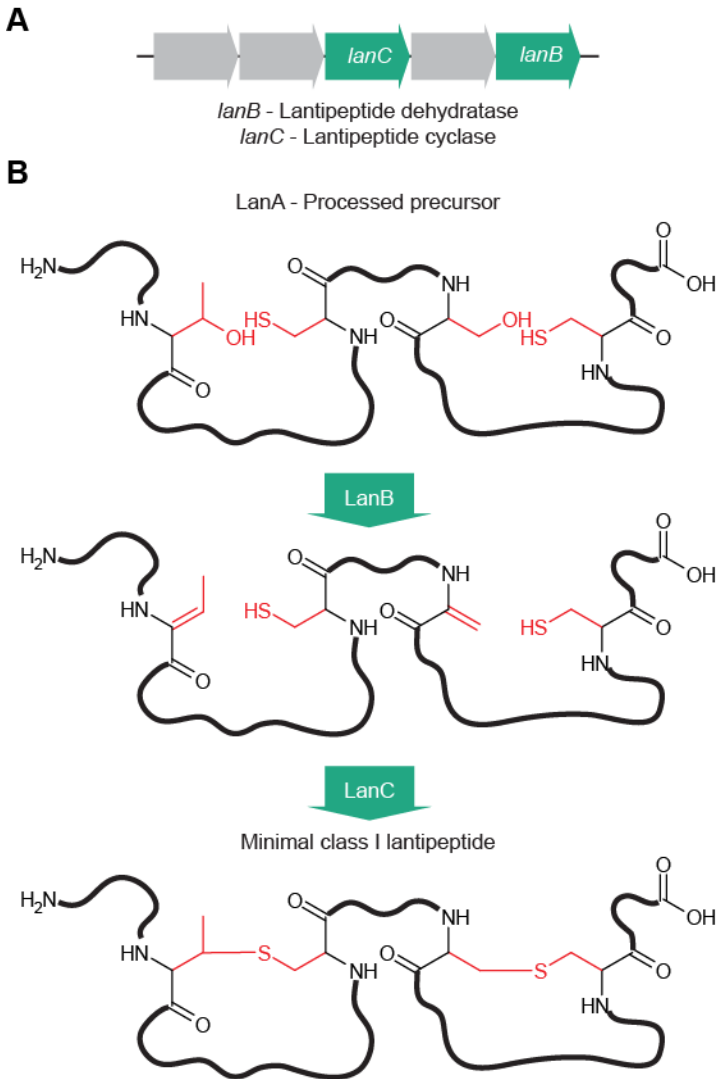
Dataset S7, Fig. 4. Minimal requirements for the detection and prediction of ComX pheromone gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *comX* and isoprenyltransferase gene *comQ* for defining ComX gene clusters. (*B*) Detection of the minimal requirements of a ComX gene cluster results in a structure prediction wherein the processed precursor peptide is transformed by the ComQ isoprenyltransferase, resulting in prenylation and cyclization of a tryptophan residue.



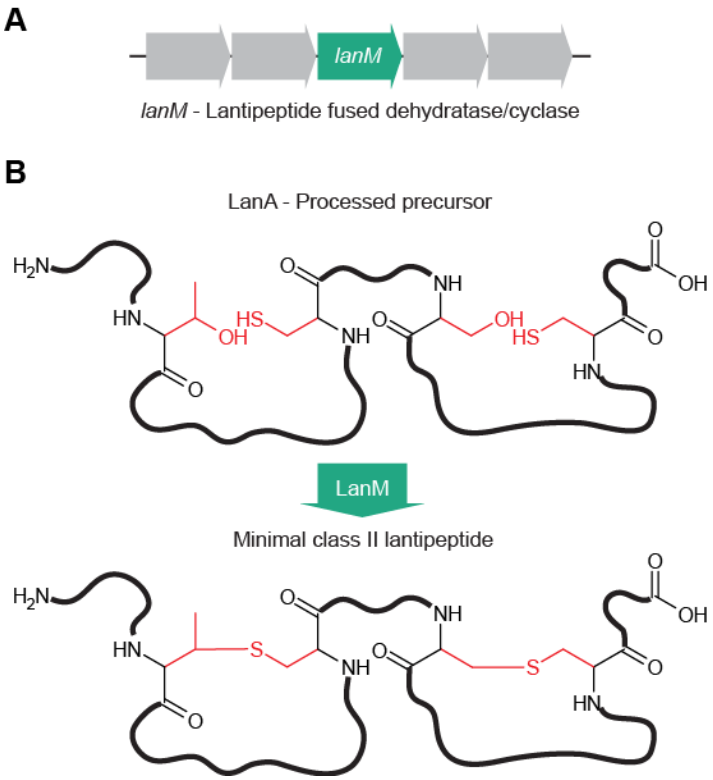
Dataset S7, Fig. 5. Minimal requirements for the detection and prediction of cyanobactin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *patE*, the leader peptide protease gene *patA*, and the protease/macrocyclase gene *patG* for defining cyanobactin gene clusters. (*B*) Detection of the minimal requirements of a cyanobactin gene cluster results in a structure prediction wherein the precursor peptide is processed by the PatA protease, then macrocyclized by the PatG protease/ macrocyclase.



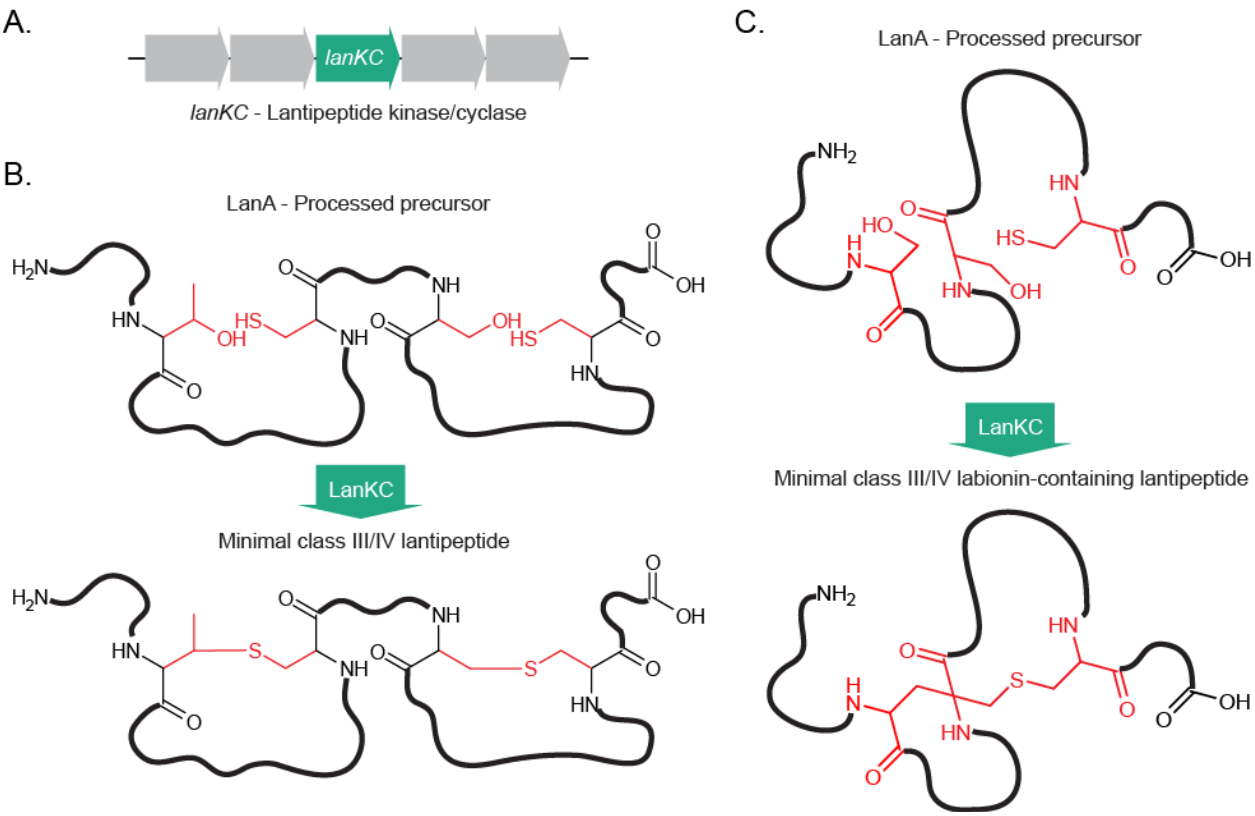
Dataset S7, Fig. 6. Minimal requirements for the detection and prediction of glycocin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *sunA* and S-glycosyltransferase gene *sunS* for defining glycocin gene clusters. (*B*) Detection of the minimal requirements of a glycocin gene cluster results in a structure prediction wherein the processed precursor peptide (containing disulfide bonds) is glycosylated by the SunS S-glycosyltransferase.



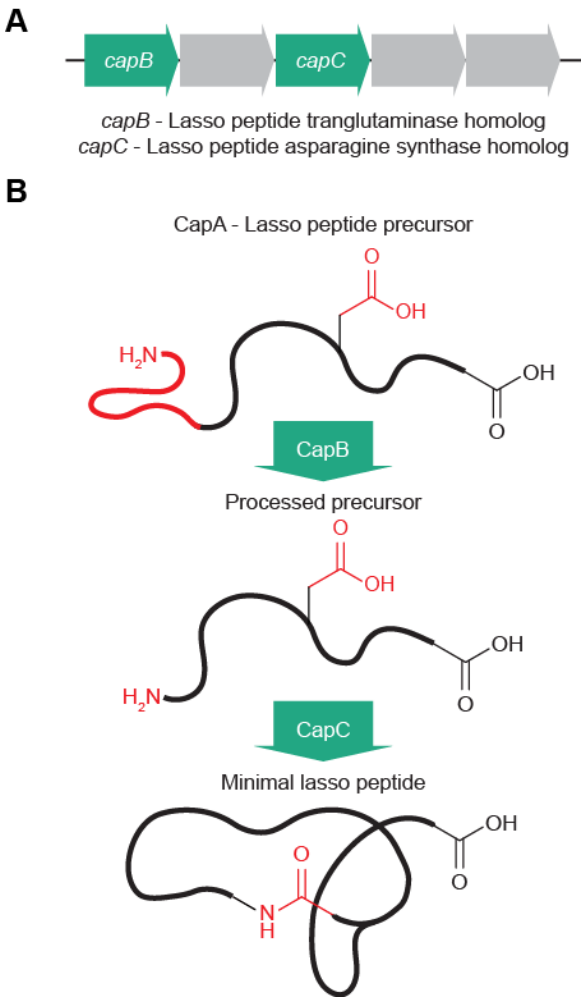
Dataset S7, Fig. 7. Minimal requirements for the detection and prediction of class I lantipeptide gene clusters. (*A*) RiPP-PRISM requires detection of the lantipeptide dehydratase gene *lanB* and lantipeptide cyclase gene *lanC* for defining class I lantipeptide gene clusters. The precursor peptide gene *lanA* can be detected with HMMs or with a defined heuristic. (*B*) Detection of the minimal requirements of a class I lantipeptide gene cluster results in a structure prediction wherein combinations of serines and threonines in the processed precursor peptide are dehydrated by LanB to form Dha and Dhb residues, which can then be left alone or used for lantithione bond formation via the lantipeptide cyclase LanC.



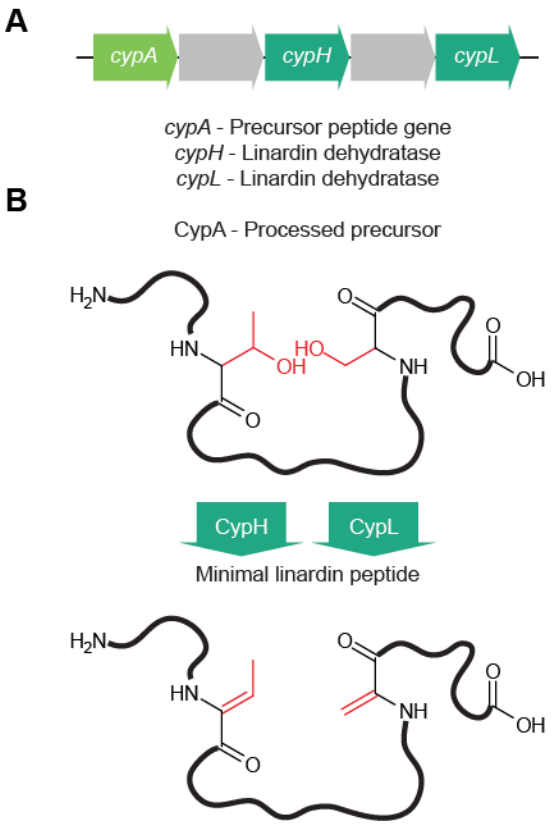
Dataset S7, Fig. 8. Minimal requirements for the detection and prediction of class II lantipeptide gene clusters. (*A*) RiPP-PRISM requires detection of the lantipeptide dehydratase/cyclase gene *lanM* for defining class II lantipeptide gene clusters. The precursor peptide gene *lanA* can be detected with HMMs or with a defined heuristic. (*B*) Detection of the minimal requirements of a class II lantipeptide gene cluster results in a structure prediction wherein combinations of serines and threonines in the processed precursor peptide are dehydrated by LanM to form Dha and Dhb residues, which can then be left alone or used for lantithione bond formation.



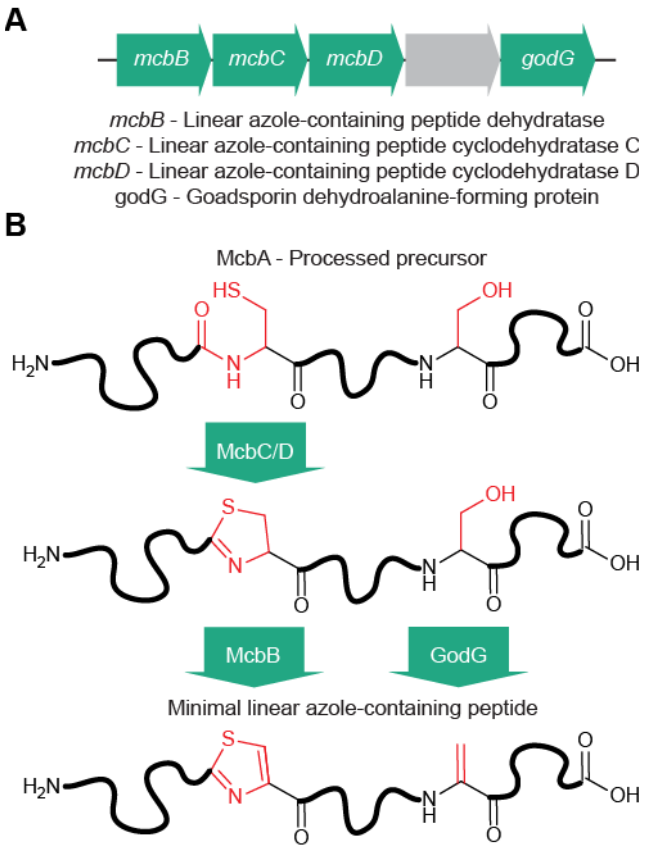
Dataset S7, Fig. 9. Minimal requirements for the detection and prediction of class III/IV lantipeptide gene clusters. (*A*) RiPP-PRISM requires detection of the lantipeptide kinase/cyclase gene *lanKC* for defining class III/IV lantipeptide gene clusters. The precursor peptide gene *lanA* can be detected with HMMs or with a defined heuristic. (*B*) Detection of the minimal requirements of a class III/IV lantipeptide gene cluster results in a structure prediction wherein combinations of serines and threonines in the processed precursor peptide are dehydrated by LanKC to form Dha and Dhb residues, which can then be left alone or used for lantithione bond formation. (C) If spacing of serine, threonine, and cysteine residues is appropriate, LanKC can also form labionins in addition to lantithiones.



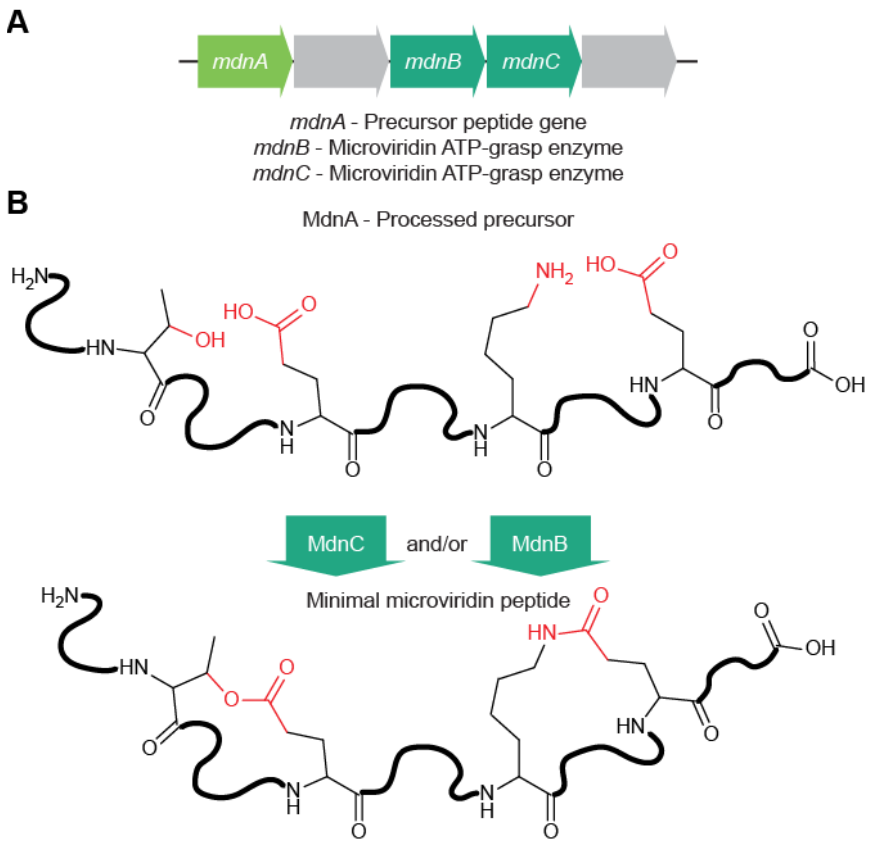
Dataset S7, Fig. 10. Minimal requirements for the detection and prediction of lasso peptide gene clusters. (*A*) RiPP-PRISM requires detection of the lasso peptide transglutaminase homolog gene *capB* and the asparagine synthase homolog gene *capC* for defining lasso peptide gene clusters. The precursor peptide gene *capA* can be detected with HMMs or with a defined heuristic. (*B*) Detection of the minimal requirements of a lasso peptide gene cluster results in a structure prediction wherein the precursor peptide is processed by the CapB transglutaminase homolog, then macrocyclized by the CapC asparagine synthase homolog, condensing the first free amine of the processed precursor with the side chain-associated carboxylic acid of a downstream aspartic or glutamic acid residue to form a macrolactam.



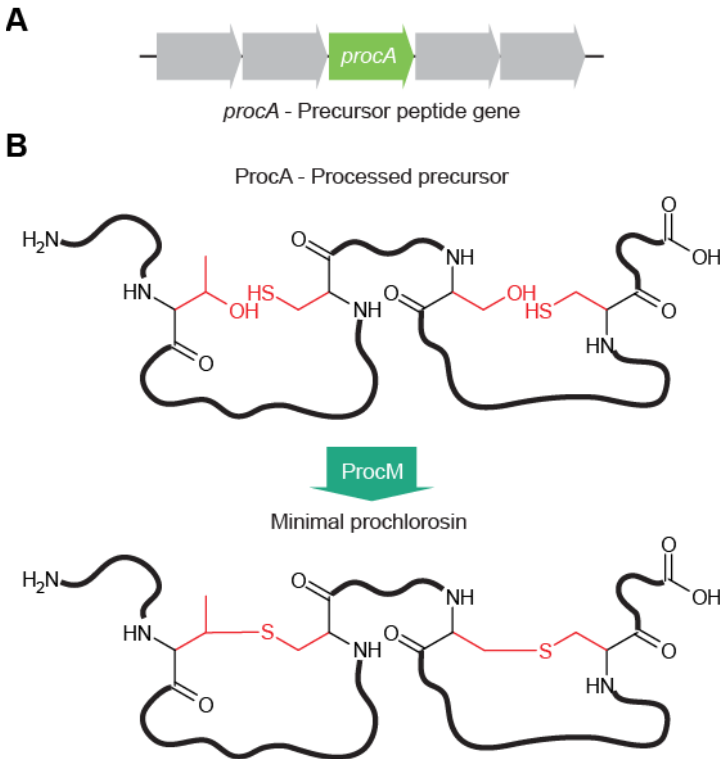
Dataset S7, Fig. 11. Minimal requirements for the detection and prediction of linardin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *cypA* and linardin dehydratase genes *cypH* and *cypL* for defining linardin gene clusters. (*B*) Detection of the minimal requirements of a linardin gene cluster results in a structure prediction wherein threonines and serines in the processed precursor peptide are dehydrated by CypH and CypL to Dhb and Dha residues.



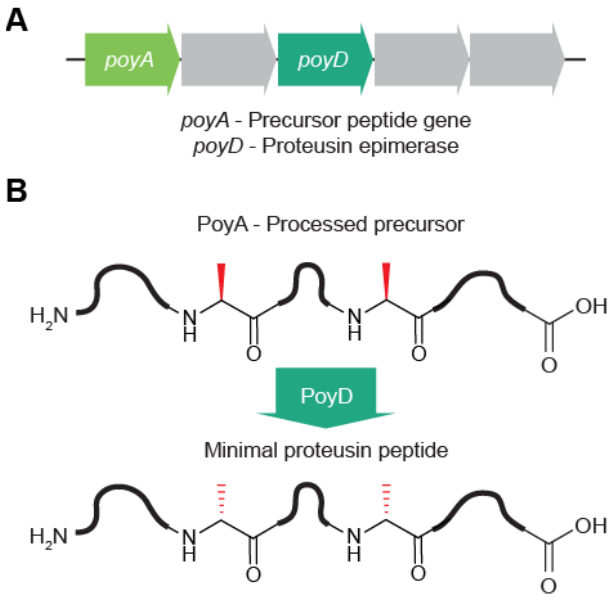
Dataset S7, Fig. 12. Minimal requirements for the detection and prediction of linear azole-containing peptide gene clusters. (*A*) RiPP-PRISM requires detection of the linear azole-containing peptide dehydratase gene *mcbB* and the cyclodehydratase genes *mcbC* and *mcbD* for defining linear azole-containing peptide gene clusters. Alternatively, the goadsporin dehydroalanine-forming protein gene *godG* can be sufficient. The precursor peptide gene *mcbA* can be detected with HMMs or with a defined heuristic. (*B*) Detection of the minimal requirements of a linear azole-containing peptide gene cluster results in a structure prediction wherein serines, threonines, and cysteines in the processed precursor peptide may be heterocyclized by the cyclodehydratases McbC and McbD, potentially followed by oxidation from azolines to azoles via the McbB dehydratase. These transformations of serines and threonines can also be accompanied by dehydration via GodG to form Dha and Dhb residues.



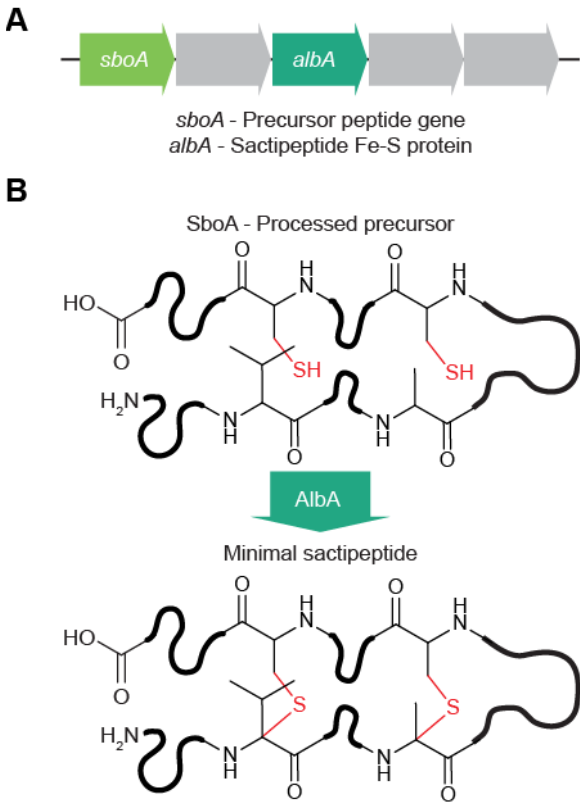
Dataset S7, Fig. 13. Minimal requirements for the detection and prediction of microviridin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *mdnA* and either of the microviridin ATP-grasp enzyme genes *mdnB* and/or *mdnC* for defining microviridin gene clusters. (*B*) Detection of the minimal requirements of a microviridin gene cluster results in a structure prediction wherein side chain-associated carboxylic acids can be used to form esters with threonines and serines via MdnC, or to form lactams with lysine residues via MdnB.



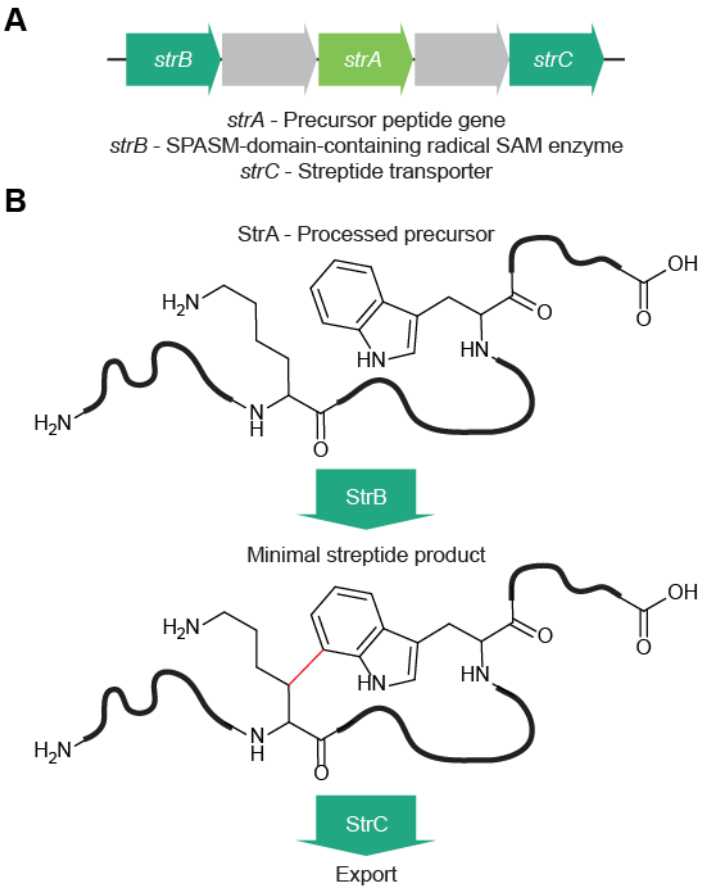
Dataset S7, Fig. 14. Minimal requirements for the detection and prediction of prochlorosin lantipeptide gene clusters. (*A*) RiPP-PRISM only requires detection of the prochlorosin precursor peptide *procA* for defining prochlorosin gene clusters, as the fused dehydratase/cyclase *procM* is often located far from the precursor peptide. (*B*) Detection of the minimal requirements of a prochlorosin gene cluster results in a structure prediction wherein combinations of serines and threonines in the processed precursor peptide are dehydrated by ProcM to form Dha and Dhb residues, which can then be left alone or used for lantithione bond formation.



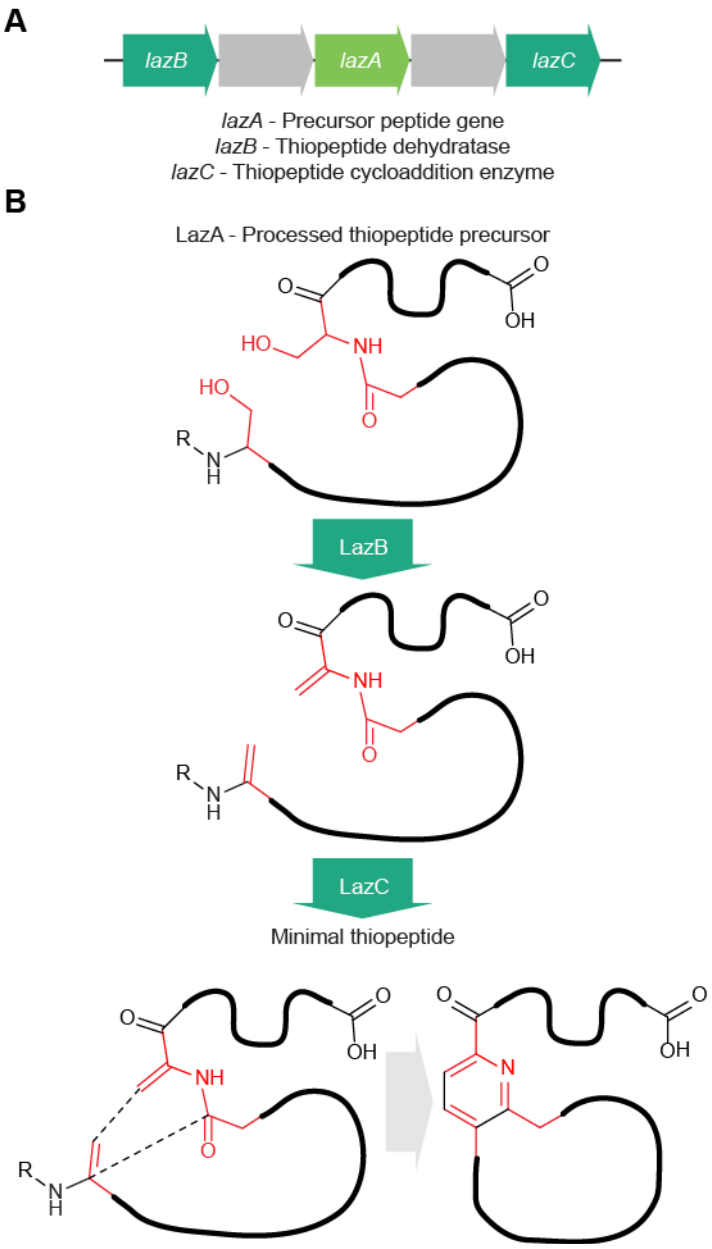
Dataset S7, Fig. 15. Minimal requirements for the detection and prediction of proteusin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *poyA* and the proteusin epimerase *poyD* for defining proteusin gene clusters. (*B*) During proteusin biosynthesis, the proteusin epimerase PoyD epimerizes a series of amino acid residues. As RiPP-PRISM does not display stereochemistry, hypothetical minimal proteusin peptide products are unaffected by the actions of PoyD, although it is still required as a diagnostic for detection of proteusin gene clusters.



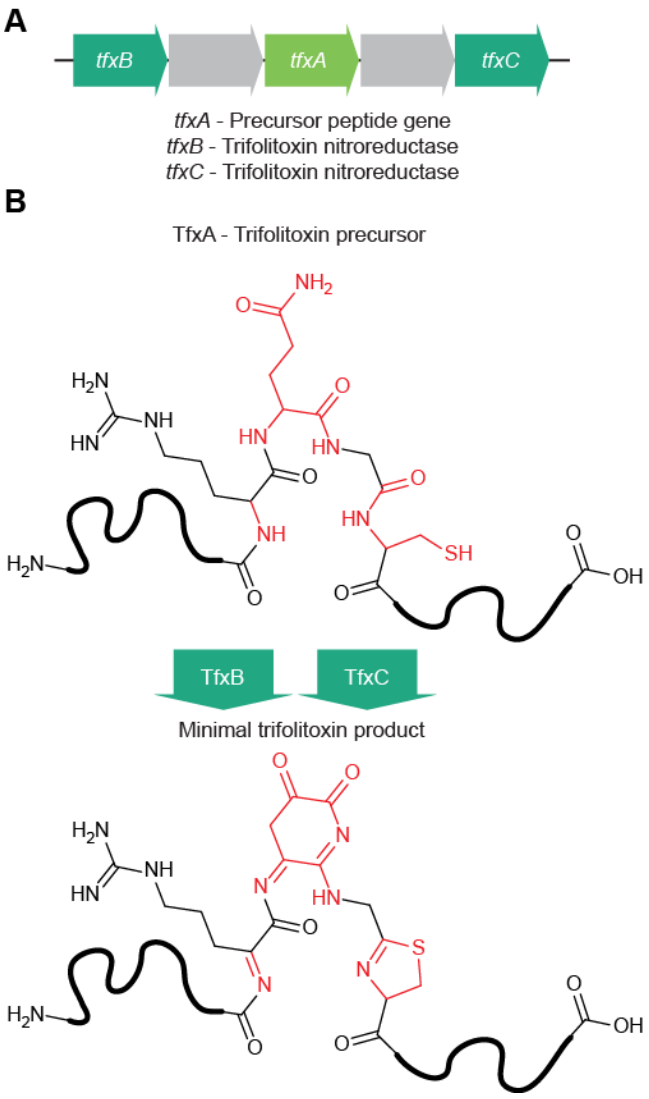
Dataset S7, Fig. 16. Minimal requirements for the detection and prediction of sactipeptide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *sboA* and the sactipeptide Fe-S protein *albA* for defining sactipeptide gene clusters. (*B*) Detection of the minimal requirements of a sactipeptide gene cluster results in a structure prediction wherein the sactipeptide Fe-S protein AlbA facilitates the formation of thioethers between cysteine residues and the α-carbons of downstream amino acids.



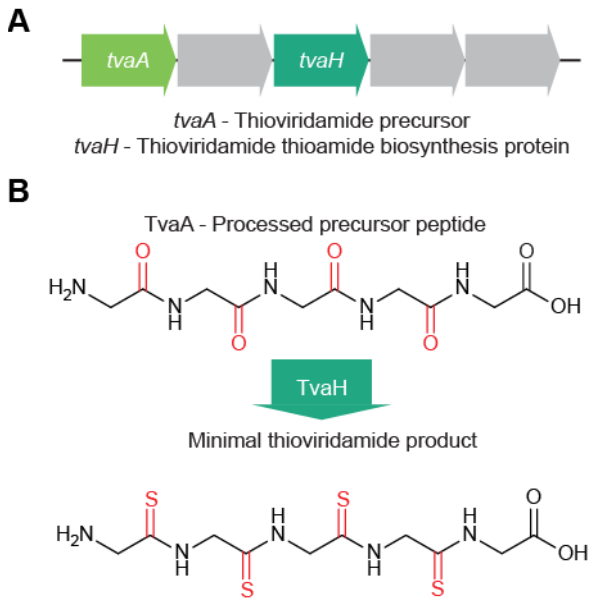
Dataset S7, Fig. 17. Minimal requirements for the detection and prediction of streptide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *strA*, the streptide SPASM-domain containing radical SAM enzyme gene *strB*, and the streptide exporter gene *strC* for defining streptide gene clusters. (*B*) Detection of the minimal requirements of a streptide gene cluster results in a structure prediction wherein the SPASM-domain containing radical SAM enzyme StrB forms a C-C bond between a lysine β-carbon and an aromatic C7 carbon on a tryptophan residue in the processed precursor peptide.



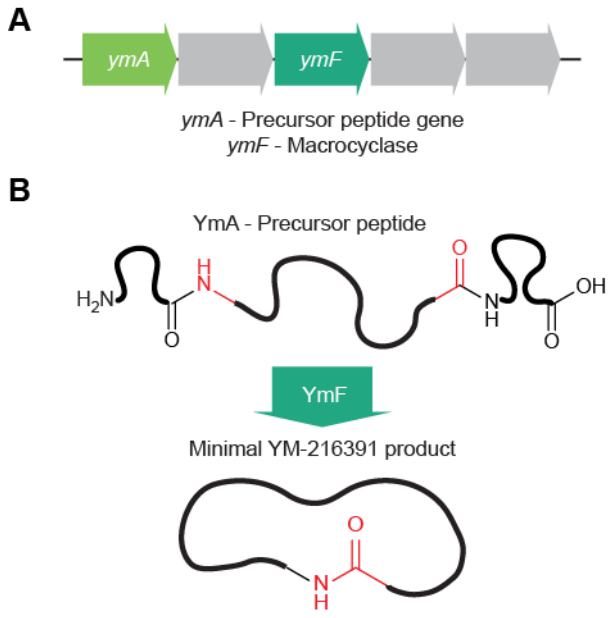
Dataset S7, Fig. 18. Minimal requirements for the detection and prediction of thiopeptide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *lazA*, the thiopeptide dehydratase *lazB*, and the cycloaddition enzyme *lazC* for defining thiopeptide gene clusters. (*B*) Detection of the minimal requirements of a thiopeptide gene cluster results in a structure prediction wherein serines and threonines in the processed precursor peptide can be converted to Dha and Dhb residues by the LazB dehydratase. Two correctly spaced Dha residues can then be acted on by the cycloaddition enzyme LazC, forming a pyridine that results in the elimination of N-terminal residues (in the case of thiocillin-like structures) or forming a dehydropiperidine that retains the N-terminal residues (in the case of thiostrepton-like structures). These two outcomes are determined by detected homology of putative *lazC* genes to either pyridine or dehydropiperidine-forming clades.



Dataset S7, Fig. 19. Minimal requirements for the detection and prediction of trifolitoxin gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *tfxA* and the trifolitoxin nitroreductase genes *tfxB* and *tfxC* for defining trifolitoxin gene clusters. (*B*) Detection of the minimal requirements of a trifolitoxin gene cluster results in a structure prediction that converts a conserved Arg-Gln-Gly-Cys motif into the putative trifolitoxin chromophore, including imine formation and heterocyclizations events. Further characterization of trifolitoxin biosynthesis will hopefully facilitate the assignment of more exact modifications.



Dataset S7, Fig. 20. Minimal requirements for the detection and prediction of thioviridamide gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *tvaA* and the thioviridamide thioamide biosynthesis gene *tvaH* for defining thioviridamide gene clusters. (*B*) Detection of the minimal requirements of a thioviridamide gene cluster results in a structure prediction wherein the TvaH enzyme replaces non-terminal amide ketone oxygen atoms with sulfurs. Ketones must be farther than one residue from the N-terminal beginning of the conserved thioviridamide macrocycle for this reaction to take place.



Dataset S7, Fig. 21. Minimal requirements for the detection and prediction of YM-216391 gene clusters. (*A*) RiPP-PRISM requires detection of the precursor peptide gene *ymA* and the YM-216391 macrocyclase *ymF* for defining YM-216391 gene clusters. (*B*) Detection of the minimal requirements of a YM-216391 gene cluster results in a structure prediction wherein the YmF enzyme performs a proteolytic macrocyclization reaction, removing the N- and C-terminal residues of the YM-216391 precursor while forming a macrolactam.