Natural Trojan horse inhibitors of aminoacyl-tRNA synthetases†

Dmitrii Y. Travin,Konstantin Severinov and Svetlana Dubiley*

For most antimicrobial compounds with intracellular targets, getting inside the cell is the major obstacle limiting their activity. To pass this barrier some antibiotics mimic the compounds of specific interest for the microbe (siderophores, peptides, carbohydrates, etc.) and hijack the transport systems involved in their active uptake followed by the release of a toxic warhead inside the cell. In this review, we summarize the information about the structures, biosynthesis, and transport of natural inhibitors of aminoacyl-tRNA synthetases (albomycin, microcin C-related compounds, and agrocin 84) that rely on such “Trojan horse” strategy to enter the cell. In addition, we provide new data on the composition and distribution of biosynthetic gene clusters reminiscent of those coding for known Trojan horse aminoacyl-tRNA synthetases inhibitors. The products of these clusters are likely new antimicrobials that warrant further investigation.

Introduction

Aminoacyl-tRNA synthetases (aaRSs) are essential cellular enzymes that catalyze the attachment of tRNAs to cognate amino acids and thus ensure correct translation of genetic information into proteins. The tRNA charging proceeds in two steps: first, an amino acid is activated by the formation of an aminoacyl-adenylate; second, the aminoacyl group is transferred to the 3'-terminal adenosine of cognate tRNA (Fig. 1A). Both steps are performed in a single catalytic center, and some aaRSs bind all three substrates – amino acid, ATP, and tRNA – simultaneously. While the mechanism of tRNA charging is conserved, two distinct aaRS classes with multiple subclasses are distinguished (Fig. 1B). In class I aaRSs, the catalytic domain adopts the Rossmann fold; most enzymes from this class function as monomers. The catalytic domains of class II aaRSs are organized as seven β-strands flanked by α-helices and most of these enzymes form dimers or oligomers.

As aaRSs perform an essential function, inactivation of any one of them leads to growth inhibition and, eventually, cell death. A number of natural antibiotics targeting tRNAs are

Received 16th November 2020, Accepted 10th February 2021
DOI: 10.1039/d0cb00208a
rsc.li/rsc-chembio

Dmitrii Y. Travin received his MSc degree in bioengineering and bioinformatics from Lomonosov Moscow State University in 2018. He started his PhD studies at Skolkovo Institute of Science and Technology under the supervision of Prof. Konstantin Severinov. Dmitrii performs genome mining and characterization of novelazole-containing peptides produced by bacteria, investigates their biosynthesis, mode of action, and ecological role in microbial communities.

Konstantin Severinov is a Professor of Molecular Biology and Biochemistry at Rutgers, the State University of New Jersey and at Skolkovo Institute of Science and Technology, and a Head of the Laboratory at the Waksman Institute of Microbiology. His research interests include peptide antibiotics, toxin-antitoxin systems, CRISPR-Cas and other bacterial defence systems, and transcription regulation in bacteriophage-infected cells.

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cb00208a

* Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia. E-mail: dmitrii.travin@skoltech.ru, severik@waksman.rutgers.edu, svetlana.dubiley@gmail.com

† Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

‡ Waksman Institute for Microbiology, Rutgers, Piscataway, New Jersey, USA
known, two of which are approved for medical\(^2\) and veterinary\(^4\) use. With the exception of purpuromycin,\(^5\) which acts on multiple aARSs, all other natural inhibitors are specific to certain aARSs.\(^6\)–\(^18\) (Fig. 1B). The majority of aARS-targeting antibiotics bind in the catalytic center and mimic either the cognate amino acid (e.g., indolomycin),\(^19\) aminoacyl-adenylate (e.g., mupirocin),\(^20\) or charged tRNA (e.g., halofuginone).\(^21\) For reviews covering the diversity of natural aARS inhibitors and their mechanisms of action see Cochrane \textit{et al.},\(^22\) Chopra \& Reader\(^23\) and Ho \textit{et al.}\(^24\)

![Fig. 1](image)

**Svetlana Dubiley**

Svetlana Dubiley is a Head of the Laboratory at the Institute of Gene Biology, Moscow. Svetlana received her PhD under supervision of Professor Andrey Mirzabekov at the Enghelgardt Institute of Molecular Biology, Moscow. Her research interests encompass molecular microbiology, biosynthesis of post-translationally modified peptides, and the social behavior of bacteria.
has a remarkably low (4 to 62 nM) minimal inhibitory concentration. The mouse model is similar to that of amoxicillin. Extensive S. pneumoniae elaboration of total chemical synthesis allows large-scale production of albomycin through a peptide linker (Fig. 2A). The three forms of albomycin initially detected by Waksman and later designated as albomycin, micromycin C-related compounds, and agrocin 84. In addition, we reveal several previously overlooked gene clusters that likely encode new aaRS inhibitors.

1. Albomycins

History of discovery and biological activity. In 1947, during the screening of a Streptomyces griseus strain collection for the production of antibacterial substances, a new antibiotic, which showed no cross-resistance with previously identified streptomycin, was discovered in the Selman Waksman laboratory. The compound was designated grisein. In parallel, a compound with potent antibacterial activity against a number of clinically relevant microorganisms was purified from Streptomyces subtropicus (former Actinomyces subtropicus) by the Gause group in the USSR. The compound was named albomycin and found to be effective against a wide range of Gram-positive and Gram-negative bacteria including clinically relevant staphylococci, streptococci and various enterobacteria. Albomycin has a remarkably low (4 to 62 nM) minimal inhibitory concentrations (MIC) for Streptococcus pneumoniae strains. The efficacy of albomycin treatment of S. pneumoniae infection in a mouse model is similar to that of amoxicillin. Extensive preclinical and clinical studies carried out in the early 1950s showed exceptional safety of albomycin and its efficacy in treatment of meningitis and pulmonary infections, including those caused by penicillin-resistant pneumococci. Recent elaboration of total chemical synthesis allows large-scale production of albomycin and its analogs, opening way for their development into effective antibacterial drugs.

Structure and mechanism of action. The albomycin molecule consists of 4'-thioxyloluranosyl pyrimidine covalently bound to an iron-chelating tri-N5-hydroxy-N5-acetyl-L-ornithine siderophore through a peptide linker (Fig. 2A). The three forms of albomycin initially detected by Waksman and later designated as δ₁, δ₂, and ε differ in substituents at the nucleobase. The major and most potent δ₂ congener contains a carbamoyl group at N4 and a methyl group at the N3 position of cytosine. Albomycin ε is a biosynthesis intermediate lacking the carbamoyl modification. The δ₁ congener is a 3-methyluridin analog of albomycin ε and is believed to be a synthesis byproduct.

The iron-chelator portion serves as a vehicle for active delivery of albomycin warhead inside both Gram-positive and Gram-negative bacterial cells through the ferrichrome-specific transporter system. In E. coli, albomycin first binds to the outer membrane receptor FhuA. FhuA is a β-strand barrel buried in the outer membrane. An N-terminal domain located inside the barrel serves as a cork. Albomycin competes for the binding site with ferrichrome, the natural FhuA substrate (Fig. 2C). Binding of a ligand promotes the FhuA interaction with TonB, which in turn stimulates a rearrangement of the cork domain and the release of the ligand into the periplasmic space. The energy required for translocation is provided by the proton motive force transduced by the TonB-ExbB-ExbD complex. Once in the periplasm, albomycin binds FhuD, a periplasmic subunit of the inner membrane ABC transporter FhuB/FhuC (Fig. 2D). Inside the cell, Fe³⁺ is reduced to Fe²⁺, which dissociates from the chelator moiety, and the siderophore part is cleaved off by the PepN peptidase. As a result, the toxic nucleoside part is accumulated in the cytoplasm of E. coli in ~500-fold excess over the concentration of antibiotic in the medium. When added directly to bacterial culture, the nucleoside portion of albomycin does not inhibit cell growth. Thus, albomycin is a Trojan horse agent. The initial rate of Fe³⁺ transport by albomycin inside the cell is only two times lower than that by ferrichrome. Thus, the antibiotic not only kills competing bacteria but is also an efficient iron-scavenger for the producer (Fig. 2B).

The target of albomycin was discovered serendipitously during a screening for inhibitors of bacterial aaRSs. SB-217452, a metabolite of Streptomyces sp. ATCC 700974, was shown to inhibit Staphylococcus aureus seryl-tRNA synthetase (SerRS) at low nanomolar concentration. The compound was identified to be the seryl-linked nucleoside moiety released upon the cleavage of the iron carrier part of albomycin δ₂. Molecular modeling of SB-217452 in complex with SerRS shows that the N4 carbamoyl group modified cytosine structurally mimics adenine of the natural substrate seryl-AMP. A more extensive network of hydrogen bonding formed by SB-217452 compared to seryl-AMP within the active center of the enzyme ensures its higher affinity.

Biochemistry and immunity. The biosynthetic gene cluster (BGC) of albomycin in the genome of Streptomyces sp. ATCC 700974 comprises 18 genes (Fig. 3B). Biosynthesis of albomycin begins with the assembly of ferrichrome, a siderophore based on a nonribosomally synthesized tripeptide. Ferrichrome is produced on nonribosomal peptide synthetase (NRPS) AbmQ, which contains one adenylylating (A) and two condensation (C) domains linked via thiolation domains (T) (Fig. 3A). The presence of a single adenylylating domain is a unique feature of AbmQ and implies that this domain supplies substrates for both condensation domains. The substrate for AbmQ, N5-acetyl-N5-hydroxy-L-ornithine, is generated from l-ornithine by AbmB, a flavin-dependent ornithine monooxygenase, and AbmA, an acetyltransferase. In contrast to most NRPSs, AbmQ does not have a thioesterase domain and the newly synthesized tripeptide is directly transferred from the C-terminal peptidyl-carrier domain to seryl-nucleoside SB-217452. The biosynthetic pathway for SB-217452 is not fully established. It is thought to include several condensation and epimerization steps resulting in 6'-amino-4'-thioheptosyl cytosine linked to l-serine via an amide bond.
Neither precursors for the synthesis of 5′-oxo-4′-thioribonucleoside nor enzymes involved are known.

The most remarkable part of the albomycin synthesis pathway is the formation of an amide bond between the thionucleoside and the L-serine, which results in SB-217452, a highly toxic compound. In vitro reconstruction of this step shows that AbmF uses seryl-adenylate produced by seryl-tRNA synthetases as a seryl donor for the amide bond formation, although a possibility that seryl-tRNA can also serve as an aminoacyl donor was not ruled out. Interestingly, the albomycin BGC encodes AbmK, which is a functional SerRS. Unlike the housekeeping SerRS, AbmK is insensitive to SB-217452 in vitro and confers resistance to albomycin in *E. coli* when overexpressed. In vitro *Streptomyces* sp. ATCC 700974 housekeeping SerRS is substantially less efficient in the coupled reaction of thionucleoside biosynthesis since it is inhibited by the end product, SB-217452. The $K_M$ for serine of AbmK is 20 times higher than that of housekeeping SerRS. Thus, AbmK may serve a triple function in the producing strain – providing the substrate for albomycin biosynthesis, conferring immunity to the toxic compound produced, and,
regulating albomycin biosynthesis by restricting it to conditions of high concentrations of intracellular serine.

**BGCs of putative seryl-nucleosides.** To date, albomycin is the only known representative of its class. We performed a

**Fig. 3** (A) Known steps in the biosynthesis of albomycin and the enzymes involved. Domains of NRPS AbmQ are indicated: A – adenylating domain, C – condensation domain, T – thiolation domain. Fragments of the molecule attached at different steps of the biosynthetic pathway are shown in different colors. (B) Biosynthetic gene cluster of albomycin (abm, red background) and found BGCs containing subsets of abm-like genes. Nomenclature of genes in the ctj gene cluster is from Zeng et al. The color scheme of enzymes in panel A matches that of genes in B. Predicted functions of the gene products are listed below. Protein IDs are shown for AbmF homologs. ABC – ATP binding cassette, CoA – coenzyme A, HAD – haloacid dehalogenase-like, MFS – major facilitator superfamily, PLP – pyridoxal phosphate, SAM – S-adenosyl methionine.
BLASTP search of the NCBI protein RefSeq database using the sequence of AbmF as a bait. Top 12 hits with E-values less than $1 \times 10^{-35}$ and query coverage above 85% were hypothetical proteins whose genes were located close to SerRS encoding genes. Within these 12 hits, in addition to trivial cases of BGCs encoding a complete pathway for albomycin biosynthesis from various Streptomyces, two additional groups of BGCs were identified (Fig. 3B). The first group contains just two very similar clusters, one from S. vitomphilius ATCC 31673, and another, ctj from Streptomyces sp. C, that was earlier discovered by Zeng et al.38 Though almost a full set of genes required for seryl-thiofuranoside cytosine biosynthesis is present in both clusters, the abmC and abmL orthologs are absent (Fig. 3B). While the function of AbmL in albomycin biosynthesis is not established, AbmC transfers the ferrichrome moiety to seryl-nucleoside SB-217452 and is thus essential for the production of the Trojan horse compound.31 The NRPS enzymes encoded in both clusters are highly similar and contain an adenylating domain, a thiolation domain but no condensation domains. 38

In contrast to albomycin BGCs, clusters from Enterobacteriaceae such as Escherichia, Salmonella, and Shigella, and some strains of Klebsiella, are susceptible to Mc at low micromolar concentrations; Pseudomonas and Gram-positive bacteria are resistant.58 McC may provide competitive advantage to producing strains in natural environments. Moreover, the probiotic activity of E. coli strain H22 in a mouse model was attributed to Mc production.59

**Structure and mechanism of action.** The E. coli McC is the best-studied representative of a family of McC-like compounds that were either experimentally identified or bioinformatically predicted in both Gram-positive and Gram-negative bacteria.60,61 The E. coli McC consists of a formylated heptapeptide MRTGNAsOSN coupled with an adenyl moiety through a phosphoramidate linkage (Fig. 4A).62 The phosphate group is decorated with an aminopropyl. Many McC-like compounds are peptidyl-adenylates with no additional decorations.61 Though peptide parts of some McC-like compounds can be as long as 56 amino acids (Table S1, ESI†) most McC-like compounds are based on heptapeptides.61 The nucleotide part of McC-like compounds can be either an adenyl or cytidyl (Fig. 4A) with various additional modifications.63,64

All studied McC-like compounds share the Trojan horse mechanism of action. The peptide part serves as a vehicle to deliver the isosparaginyl-nucleotide warhead inside the cell.11 The route of E. coli McC entrance into the cell is not fully established. Despite multiple mentions in the literature,65 no direct evidence that OmpF is the primary porin responsible for McC passage through the outer membrane of Gram-negative bacteria exists. Most probably, several porins are involved. Once E. coli McC reaches the periplasmic space it interacts with YejA, the substrate-binding subunit of the YejABEF ABC transporter (Fig. 4B).66 Though exact cellular function of this transporter in E. coli is not established, homologs from Salmonella, Brucella and Pseudomonas may be involved in resistance to antimicrobial peptides.67–70

Presumably, the fact that peptide parts of most McC-like compounds are seven amino acids long is explained, at least in part, by the binding preferences of YejA. Using chemically synthesized E. coli McC analogs it was shown that shortening the peptide length to six amino acids dramatically reduces bioactivity due to inefficient uptake through YejABEF.71 Increase of the peptide part length up to 10–12 amino acids stimulated the uptake.61 Peptidyl-nucleotides with the peptide part lengths around 20–25 amino acids can enter E. coli not only through YejABEF but also through SbmA,72 a transporter of peptide antibiotics such as microcin B17 (43 amino acids) or microcin J25 (21 amino acids).73,74 McC-like compounds with even longer peptide parts, i.e., the 42 amino acid long McC from Versinia pseudotuberculosis (Fig. 4A), undergo endoproteolytic processing prior to export, which leads to bioactive compounds with shorter (11 amino acids in the case of Y. pseudotuberculosis McC) peptide parts.75

Once E. coli McC gets inside the bacterial cytoplasm, its peptide part is deformylated and six N-terminal residues are decoded by a specific peptidyl-tRNA synthetase with the concomitant release of the isosparaginyl-nucleotide from the seryl-tRNA peptidyl-tRNA synthetase, which carries the isosparaginyl warhead. Upon delivery of the isosparaginyl-nucleotide inside the cell, it interacts with the substrate-binding subunit of the transport system and transferred to the sensitive proteins, leading to the cells’ death. The isosparaginyl residue acts as a prodrug, releasing the active component of the compound, i.e., the isosparaginyl nucleotide.76

**Microcin C and related compounds**

**History of discovery and biological activity.** Microcin C (McC, initially named “microcin C7”) was purified from Escherichia coli BM7006,52 a strain first isolated by Chabbert in 1950.53 The same compound was independently discovered and described under the name “microcin C51”.54 Eventually, it was shown that 1–2% of E. coli strains isolated from humans are McC producers.55–57

The product of NRPS may be transferred to the seryl-nucleoside part either by carboxylate-amine ligase family protein (CtjK in Streptomyces sp. C), or by CtjH, a protein of unknown function. Presumably, Streptomyces sp. C and S. vitomphilius ATCC 31673 produce different SB-217452-like nucleosides, since the ctj cluster lacks one of the rSAM enzymes, an ortholog of AbmJ, which is responsible for D-ribo to D-xylo epimerization of the furanose ring in albomycin. It is worth mentioning that the D-ribo form of albomycin has significantly reduced antimicrobial activity.48 Both clusters are apparently functional, as the specific enzymatic activity of carbamoyl transferase CtjE and methyltransferase CtjF toward cytidine was confirmed.38 Moreover, ctjE complemented the deletion of the abmE gene from the albomycin cluster. Interestingly, both clusters encode an HAD-family hydrolase (nucleosidase), which is lacking from other clusters and may serve as an additional immunity gene.31

The second group of putative seryl-thiofuranoside cytosine synthesizing BGCs was not previously described (Fig. 3B). Its representatives are found in the genomes of Proteobacteria from different classes. Despite major rearrangements in their architecture, these BGCs contain roughly the same set of genes. These include genes encoding cytosine-modifying carbamoyl-transferase (AbmE-like), cytosine methyltransferase (AbmI-like), and thiofuranosyl-assembly and epimerization enzymes homologous to AbmD, AbmH, and AbmF. The BGC from Trinickia caryophylli DSM 50341 also contains an abmJ ortholog. In contrast to albomycin BGCs, clusters from Proteobacteria entirely lack genes required for assembly of the siderophore part. This implies either a different mechanism of end product uptake by the sensitive cells or the coupling to a siderophore molecule whose biosynthesis is encoded elsewhere in the genome of the producer.

### 2. Microcin C and related compounds

**Microcin C and related compounds**

**History of discovery and biological activity.** Microcin C (McC, initially named “microcin C7”) was purified from Escherichia coli BM7006,52 a strain first isolated by Chabbert in 1950.53 The same
removed by aminopeptidases. In *E. coli*, any one of the three aminopeptidases, PepA, PepB, or PepN, is capable of intracellular McC processing (Fig. 4B). The final product of processing is an aminopropylated isoasparaginyl-adenylate. Processed *E. coli* McC mimics aspartyl-adenylate, the intermediate of the tRNA<sup>Asp</sup> aminoacylation reaction (Fig. 4A), and inhibits aspartyl-tRNA synthetase (AspRS). Inhibition of AspRS results in the accumulation of uncharged tRNA<sup>Asp</sup>, translation inhibition, stringent response and cessation of cell growth.

Although the structures of AspRS in complex with several chemically synthesized analogs of processed McC have been determined, the exact binding mode of processed McC to AspRS is yet to be revealed. Given the variety of known and proposed auxiliary modifications in the nucleoside part of McC-like compounds, the catalytic center of AspRS should be flexible enough to accommodate structurally diverse aminoacyl-nucleotides.

**Biosynthesis and regulation.** McC-like compounds belong to the ribosomally synthesized post-translationally modified peptides (RiPPs) family since they are produced by modification of a gene-encoded peptide precursor by dedicated enzymatic machinery. Peptide precursors of most RiPPs comprise the leader part, which is required for the specific recognition by the enzymes introducing the post-translational modification, and the core part where modifications are introduced to. Uniquely, the core peptide of McC-like compounds is represented by a single C-terminal amino acid residue. In contrast to other RiPPs, the fully modified McC-like compounds are exported from the producer cell with complete or partial leader peptide attached to the miniaturized core.

The minimal microcin C BGC (*mcc*) consists of just three genes: *mccA*, encoding the precursor peptide; *mccB*, coding for a Thf-like nucleoside transferase; and *mccC*, encoding an export pump (Fig. 6 (1)). Auxiliary genes present in many *mcc* BGCs encode enzymes responsible for secondary post-translational modifications or self-immunity, however, the function of most of them is not identified yet (Fig. 6).

When the sequence of the *E. coli mcc* gene cluster was determined (Fig. 6 (2)), it became apparent that the *mcca* gene codes for a MRTGNA<sub>7</sub>C0<sub>7</sub>N<sub>7</sub> heptapeptide, while mature McC has the...
fMRTGNAβsoN peptide part.80 This discrepancy was resolved by work from the Walsh lab, when they showed that nucleoside coupling to the precursor peptide by adenyllytansf-ferase MccB proceeds in two steps through a succinimide intermediate.81,82 The opening of the succinimide ring results in the isomerization of the asparagine residue, converting it to isoasparagine. The reaction consumes two molecules of ATP per molecule of peptide-nucleotide synthesized (Fig. 5B). While one would expect that a similar mechanism could be employed with peptide precursors containing C-terminal Gln, however, at least the E. coli MccB is unable to modify such a peptide.72 In fact, the presence of a C-terminal asparagine is the only strict requirement shared by known McC-like precursor peptides.

The crystal structure of MccB reveals a homodimer with two active centers (Fig. 5A). Both steps of the adenylation reaction are catalyzed within the same active center. The N-terminal part of one subunit (so-called RII Recognition Element, “RRE”) and the central “crossover loop” of the opposite protomer form the peptide clamp domain responsible for MccA binding.82,83 The formyl group on the N-terminal methionine of the precursor heptapeptide strongly improves the efficiency of adenylation by promoting productive binding of the precursor peptide to MccB.84

Fig. 5  (A) The structure of the dimeric nucleoside transferase MccB from E. coli in complex with N-formylated microcin C (PDB ID: 6OM4).84 Adenylation domains and N-terminal RREs are shown in different colors, black arrows indicate “crossover loops”. Close-up views: (left) the active site of the adenylation domain with key residues required for efficient catalysis (D214 and Y239) is shown in stick representation, black arrow indicates the P-N bond; (right) the binding site of MccA peptide N-terminus formed by RRE and “crossover loop” (violet), residues 313–347 (catalytic domain) are removed for clarity. (B) The mechanism of P–N bond formation by nucleoside transferase MccB. R stands for the fMRTGNA peptide. Two molecules of ATP consumed per molecule of McC produced are shown in violet, released AMP is blue. (C) Installation of aminopropyl decoration by the MccD and MccE1 enzymes. SAM – S-adenosyl methionine, MTA – 5′-methylthioadenosine.

© 2021 The Author(s). Published by the Royal Society of Chemistry
Strikingly, MccB is rather promiscuous in its choice of NTP substrates. In vitro, the *E. coli* precursor peptide can be modified with either AMP, CMP, GMP or UMP moieties. In vivo, the reaction is highly specific and, depending on the enzyme, exclusively yields either peptidyl-adenylates (e.g., when catalyzed by *E. coli* MccB) or peptidyl-cytidylates (e.g., when catalyzed by *Bacillus amyoliquefaciens* MccB). All known MccB enzymes catalyzing in vivo cytidylation contain a C-terminal methyltransferase domain responsible for the additional decoration of the nucleobase with carboxymethyl. This reaction requires carboxy-S-adenosyl methionine (cxSAM) as a substrate. The cxSAM is synthesized from S-adenosyl methionine (SAM) and prephosphorylated by MccS, and enzyme encoded exclusively in BGCs producing cytidylated McC-like compounds. The exact position of the carboxymethyl group on the cytidine base remains unknown.

Auxiliary post-translational modifications of RIPPs increase their bioactivity and/or are required to overcome the immunity of target bacteria. The aforementioned aminopropyl decoration of the phosphate found in *E. coli* MCC and in some other MCC-like compounds serves both purposes: it increases the toxicity of peptidyl-nucleotides and allows to escape the action of some immunity proteins. Noteworthy, most *Salmonella* species are resistant to MCC-like peptidyl-adenylates but sensitive to aminopropylated compounds. The aminopropyl is introduced via a two-step reaction. During the first step, a class I methyltransferase MccD transfers the 3-carboxy-3-aminopropyl moiety from SAM to the peptidyl-nucleotide substrate, releasing methylthioadenosine (MTA) as a byproduct (Fig. 5C). Since MTA is a potent inhibitor of SAM-binding enzymes, the reaction efficiently proceeds only in the presence of the MtM nucleosidase, which is encoded in the core genome of many bacteria. The second step is decarboxylation of the 3-carboxy-3-aminopropylated peptidyl-nucleotide intermediate by MccE1 (in the case of *E. coli* – the N-terminal domain of a fused protein MccE) (Fig. 5C).

*E. coli* MCC is produced in the early stationary phase of growth, and its synthesis is tightly regulated. Transcription of the *mccABCDE* operon is initiated by the σ7 factor-containing RNA polymerase in a CRP-dependent manner. A ribosome-dependent transcription terminator is commonly found in the *mccA-mccB* intergenic regions of *mcc* operons (Fig. 4C). In *E. coli*, the function of this terminator leads to the production of two types of mRNA molecules: a short one, which contains the *mccA* ORF only, and a long one, comprising the entire operon. Co-transcriptional recognition of the Shine–Dalgarno sequence of *mccA* by the ribosome stimulates transcription termination. The ribosome binding to short mRNA is virtually irreversible, which allows multiple rounds of translation without ribosome dissociation and ensures that enough peptide precursor substrate is produced for the action of the maturation enzymes.

**Immunity.** The MCC-producing bacteria are vulnerable to the action of isoasparaginyl-nucleotide that accumulates inside their cytoplasm. To prevent the damage to the producer, diverse genes responsible for MCC immunity are often incorporated in *mcc*-like BGCs. The Gcn5-related N-acetyltransferase (GNAT) MccE2 provides immunity by acetylating processed MCC and many other aminocyl nucleosides at the alpha-aminogroup, which likely prevents the binding to cognate aminocyl-tRNA synthetases (Fig. 7B). MccE2 enzymes appear to be specific to the nucleoside part of their substrates. For instance, *E. coli* MccE2 protects cells from peptidyl-adenylate compounds, while the *Y. pseudotuberculosis* homolog provides resistance to MCCs containing carboxymethylated cytosine.

The MccF protein encoded in the *E. coli mcc* cluster provides self-immunity by cleaving off the aspartate residue or the oligopeptide from, respectively, processed or mature MCC (Fig. 7A). MccF is the S66 family serine peptidase and is similar to LdcA, a carboxypeptidase, which cleaves amide bonds between L- and D- amino acids during murein recycling. However, the latter enzyme is unable to protect *E. coli* from MCC-like compounds. MccF homologs are encoded in several *mce*-like BGCs (Fig. 6) and by numerous stand-alone bacterial genes. The stand-alone MccF ortholog from *B. anthracis* can hydrolyze peptidyl-adenylates and isoasparaginyl-adenylates with or without aminopropyl decoration and provides MCC resistance to *E. coli* when overexpressed.

Another immunity protein, MccH targets the phosphoramid linkage of isoasparaginyl-adenylates. It is a member of a vast family of Histidine-Triad Hydrolases (HITs). MccH is similar to highly conserved HinT phosphoraminidases present in almost all organisms. The function of HinT is poorly understood and its natural substrates are unknown. HinT from *E. coli* or *Hyalangium minutum* are unable to hydrolyze isoasparaginyl-adenylate. In contrast, MccH, which was first identified in the *mcc* BGC from *H. minutum*, evolved to specifically cut the N-P bond in isoasparaginyl-adenylate, releasing AMP and isoasparagine (Fig. 7C). A number of bacterial MccH orthologs not linked to *mcc*-like BGCs were identified. When overexpressed, these enzymes provided resistance to simple MCC-like peptidyl-adenylates but not to aminopropylated forms.

**BGCs of *mcc*-like compounds.** Recently, a number of *mcc*-like BGCs were predicted in bacterial plasmids and chromosomes (Table S1, ESI†). Some of these were validated earlier; validation of several others is shown in Fig. S1 (ESI†). The majority of the BGCs consist of just three core genes, *mccABC*, and presumably encode the pathway of biosynthesis of simplest peptidyl-nucleotides (Fig. 6 [1]). However, many *mce*-like BGCs display more complex architectures (Fig. 6).

As already mentioned, clusters encoding MccB proteins with C-terminal carboxymethyltransferase domains also encode cxSAM synthases MccS, which are absent from other *mcc* clusters. Experimentally validated products of such clusters are cytidine-containing antibiotics with carboxymethyl decoration installed by C-terminal carboxymethyltransferase domains of their MccBs (Fig. 6 [14–20], see also above). Two genes *mccX* (Fig. 6 [15, 18]) and *mccY* (Fig. 6 [16, 17]) are also found exclusively in this group of clusters, suggesting that the proteins they encode may specifically target the carboxymethylated cytosine nucleobase. While MccX has no sequence similarity with any proteins with established function, MccY is homologous to AbmE from the albomyacin.
biosynthesis pathway, MccY may add a carbamoyl decoration to the amino group of the carboxymethylated cytosine nucleobase.

SAM-dependent enzymes comprise the largest and most diverse group of tailoring enzymes encoded by mcc-like operons. Besides the already described 3-carboxy-3-aminopropyl transferase (MccD) and carboxymethyltransferase forming the C-terminal domain of some MccB proteins, genes encoding putative SAM-dependent transferases are identified in many mcc-like BGCs (Fig. 6 (4–6)). However, as there are no close homologs with known substrate specificity for these enzymes, secondary post-translational modifications they perform are impossible to predict.

Analysis of the mcc-like BGCs composition suggests that the self-immunity strategies employed by McC producers extend beyond those already discussed. Some BGCs encode AspN-like
peptidases (Fig. 6 (18)) or ClpA homologs (Fig. 6 (8)), which may hydrolyze isoasparaginyl-nucleotides accumulating inside the producing cell. Several BGCs from Actinobacteria include asnS (aspartate tRNA synthetase) genes (Fig. 6 (19)), which, similarly to abmk from albomycin biosynthetic pathway (see above) or agnB2 from agrocin 84 BGC (see below) may provide resistance to processed McC-like compounds.

In summary, the abundance and diversity of known or putative strategies of avoidance of self-intoxication by intracellularly accumulating processed McC-like compounds parallels the diversity of secondary post-translational modifications of the basic peptide-nucleotide backbone and suggest that no single strategy is universal: additional decorations attached to the isoasparaginyl-nucleotide core allow some non-cognate
McC-like compounds to overcome the immunity of specific producers.

3. Agrocin 84

**History of discovery, biological activity, structure, and mode of action.** Crown gall disease caused by soil-borne tumorigenic strains of *Rhizobium radiobacter* (formerly classified as *Agrobacterium tumefaciens*) affects a wide spectrum of dicotyledonous plants and leads to significant losses in nursery stock of stone fruits and nut trees. An nonpathogenic strain *Rhizobium rhizogenes* K84 (formerly *Agrobacterium radiobacter* K84) isolated from a soil sample derived from an Australian plant nursery in 1970, was shown to prevent the development of agrobiological infections when added to planting material. For almost half a century, this strain has been used commercially as a biocontrol agent against crown gall disease.

The activity of *R. rhizogenes* K84 against tumor-inducing strains is due to the production of an antimicrobial compound agrocin 84, a disubstituted adenine nucleotide containing 3'-deoxarabinose instead of the ribose. The two moieties attached to the nucleotide core via phosphoramide bonds are D-glucopyranose-2-phosphate at the N6 position of adenine and 2,3-dihydroxy-4-methylpentanamide at the phosphate group (Fig. 8A, left panel). The D-glucose-2-phosphate (Fig. 8A, red dashed frame) is required for the uptake by susceptible strains. Once the uptake moiety is removed, a toxic moiety (TM84, Fig. 8A, right panel) is released, making agrocin 84 a Trojan horse agent.

Plant-pathogenic agrobacteria insert a specific region (T-DNA) derived from their Ti (Tumor inducing) plasmids into the nuclear genome of the plant cell. Some T-DNA genes encode enzymes of the biosynthesis of opines – metabolites produced by the transformed plant cells and subsequently imported into the bacterial cells. These compounds serve as a source of carbon, phosphorus and nitrogen for the agrobacteria. The import of the precursor of pentanamide is attached to the adenosine monophosphate via the phosphoramide bond. The import of opines agrocinopines A and B is determined by the AccABCDE transporter encoded on the Ti plasmid (Fig. 8B). Agrocin 84 hijacks this transport system to get inside the cell of tumor inducing agrobacteria. Agrocinopine A is made of sucrose bound to L-arabinose-2-phosphate (Fig. 8A), while agrocinopine B has fructose instead of sucrose. Crystal structures (Fig. 8C) show that the periplasmic binding protein AccA recognizes both agrocin and agrocinopine A through their pyranose-2-phosphate groups (D-glucose-2-phosphate in the case of agrocin and L-arabinose-2-phosphate in the case of agrocinopine, Fig. 8A, red dashed frames). The subsequent processing step required for the release of agrocin 84 toxic moiety is catalyzed by phosphodiesterase AccF (Fig. 8B), which also performs the first step of agrocinopines catabolism. Since the accABCDEFG (agrocinopine catabolism) locus is part of the Ti-plasmid, agrocin-producing *R. rhizogenes* strains specifically inhibit the growth of tumor inducing strains but have no effect on strains that do not induce tumors. In other words, by mimicking the tumor-derived substance, agrocin 84 is able to find its way specifically into phytopathogenic bacteria harboring the Ti-plasmid.

TM84, the toxic moiety of agrocin 84, inhibits tRNA<sub>Leu</sub> aminoacylation by leucyl-tRNA synthetase (LeuRS, Fig. 8B). TM84 is a structural mimic of leucyl-adenylate (Fig. 8A, grey background) with a nonhydrolyzable phosphoramidate bond (Fig. 8A, black arrow) connecting the nucleotide and the modified pentanamide. Unlike other stable leucyl-adenylate analogs, TM84 binds weakly to the LeuRS active site in the absence of tRNA<sub>Leu</sub>. The 3’-terminal adenosine of tRNA<sub>Leu</sub> forms a hydrogen bond with TM84, stabilizing the ternary complex (Fig. 8D).

**Biosynthesis and immunity.** The production of agrocin 84 by *R. rhizogenes* K84 is associated with the presence of a ~45 kbp plasmid pAgK84, which can be transferred by conjugation to other rhizobia, turning them into agrocin producers and making them resistant to the action of the compound. As the transfer of pAgK84 to phytopathogenic strains can break down biocontrol, a strain lacking a 5.9 kbp fragment of pAgK84 required for the transfer was constructed. This strain, named *R. rhizogenes* K1026, had similar biocontrol properties as the parent strain K84 but was unable to transfer the resistance determinants. K1026 serves as an active ingredient in commercial biological pesticide NOGALL. The exact pathway of agrocin 84 biosynthesis is unknown. The sequence of the pAgK84 plasmid reported by Kim et al. allowed the identification of a 17-gene agn biosynthetic cluster required for agrocin 84 production and immunity (Fig. 9, pink background). The function of several Agn proteins was proposed based on sequence similarity with known enzymes. AgnA harbors a conserved adenylation domain characteristic of aaRSs and was thus proposed to catalyze the key step in which the precursor of pentanamide moiety is attached to the adenosine monophosphate via the phosphoramide bond. Our analysis suggests that AgnC1 has remote structural similarity to prolyl-tRNA synthetases (according to Phyre2 and HHpred) and may thus be another candidate for the enzyme attaching the modified pentanamide to the nucleotide moiety. AgnB1, AgnC3 and AgnC7 are SAM-dependent enzymes proposed to be involved in modified pentanamide biosynthesis. The involvement of SAM in agrocin 84 biosynthesis is supported by an observation that *R. rhizogenes* K84 with mutations in the chromosomal achY gene encoding S-adenosyl-L-homocysteine (SAH) hydrolase (an enzyme involved in SAH recycling back to SAM) do not produce agrocin.

The agn BGC (Fig. 9, pink background) includes two self-immunity genes. *agnG* encodes a transmembrane pump involved in the export of the antibiotic; *agnB2* encodes a leucyl-tRNA-synthetase. Compared to its housekeeping counterpart, the AgnB2 enzyme is ~1000-fold less sensitive to TM84, providing the producer with alternative machinery for tRNA<sub>Leu</sub> charging.

**BGCs of putative agrocin 84-like compounds.** Agrocin 434 is a narrow-spectrum antimicrobials produced by the strains of the former genus *Agrobacterium* and active against other agrobacteria (Fig. S2, ES1†). However, these compounds are structurally distinct from agrocin 84 and no information about their modes of action is available, making
agrocin 84 the only compound of its class. We performed BLASTP searches for orthologs of proteins encoded in the agn biosynthetic gene cluster in the bacterial genomes from the RefSeq database. A search using AgnA, AgnB1-B2 and AgnC1-C7 sequences as baits found homologs encoded by a BGC in the genome of Ensifer sp. ENS09, a representative of Rhizobiales. This BGC is analogous to...
Fig. 9  (A) Comparison of agrocin 84 (agn) biosynthetic gene cluster (pink background) with five clusters found in the genomes of Actinobacteria and a cluster from the genome of Ensifer sp. ENS09. The proposed functions of encoded proteins are listed below. Levels of sequence identity (in %) between Agn proteins and their homologs encoded by the cluster from Ensifer sp. ENS09 are indicated. Protein IDs are provided for AgnC1 homologs. ABC – ATP binding cassette, GlcNAc – N-acetylglucosamine, PLP – pyridoxal phosphate, SAM – S-adenosyl methionine, SDR – short-chain dehydrogenase/reductase. (B) Maximum likelihood phylogenetic tree of AgnC1 homologs, built using PhyML. Numbers at branching points indicate bootstrap support values from 100 replicates (scores above 60 are shown). The presence of agn genes homologs in corresponding BGCs is indicated by painted circles on the right. The number of distinct genes encoding putative MetRSs in the genome of the given strain is shown in the ## column. As the whole-genome sequence of the strain Streptomyces bottropensis DO-45 is missing, the number of MetRS genes is unknown (question mark in the ## column).
the agn cluster but lacks homologs of agnE1-agnF (Fig. 9A). The cluster encodes an export pump that differs from AgnG in the number and composition of predicted transmembrane regions. Strikingly, despite the huge number of various Rhizobiales isolated and sequenced to date, the Ensifer sp. ENS09 BGC is the first cluster coding for an agrocin 84-like compound identified in the genomes of Rhizobiales outside R. rhizogenes.

Another interesting finding came from the results of a BLASTp search, which used the sequence of AgnC1 protein as a bait. In addition to trivial hits from R. rhizogenes K84 and Ensifer sp. ENS09, the top hits with E-values lower than 1e-8 contained sequences of 19 hypothetical proteins encoded in actinobacterial genomes. Manual analysis of the genomic context of agnC1 orthologs retrieved by the search revealed genes encoding homologs of other Agn proteins (Fig. 9B). Interestingly a 420-bp ORF preceding agnC1, which previously was not annotated as a gene,123 is present in BGCs from Ensifer and Actinobacteria, suggesting that it may encode a functional partner of AgnC1 (we annotated it as agnC0). We propose that identified actinobacterial BGCs encode the biosynthesis of nucleotide-containing natural products structurally close to agricin 84. The newly identified BGCs share an additional set of genes not found in the agn BGC (Fig. 9A, genes with black numbers). The products of these genes may be responsible for additional novel modifications of the nucleotide scaffold. Strikingly, a gene encoding a methionyl-tRNA-synthetase (MetRS) is present in most actinobacterial BGCs (Fig. 9A, red dashed frames). In most of the studied cases, the gene of MetRS associated with the BGC is not the only MetRS encoding gene present in the genome (Fig. 9B). The product of this gene may play the same role as AgnB2, providing immunity to the produced compound. If true, this would imply that unlike agricin 84, the products of some of the identified BGCs target the housekeeping MetRS.

Conclusions

Albomycin, microcin C-like compounds, and agricin 84 are aaRSs inhibitors exploiting the “Trojan horse” strategy to enter cells. The transport moieties of these antibiotics mimic three distinct types of compounds actively imported into bacterial cells: a siderophore, peptidines, and an opine. The spontaneous resistance to the Trojan horse inhibitors is almost exclusively associated with the loss of function of a dedicated transporter.12,66,131 Although the frequency of such resistance mutations is relatively high (within 10^{-4}–10^{-5} range), it may not undermine the potential of Trojan horse inhibitors to be developed into drugs. For example, the loss of YejABEF or ferrichrome transporters reduces the virulence of bacteria and makes them prone to clearance by the immune system of the host.12,68,69 Likewise, the loss of the Ti-plasmid encoding AccABCDE transporter required for the internalization of agricin 84, leads to a non-phytopathogenic phenotype.131

Trojan horse inhibitors containing covalently bound siderophore moiety – sideromycins – have inspired researchers to develop numerous conjugates, each being a combination of a siderophore, a linker, and a known synthetic or natural antibiotic (see Negash et al.132 or Wenczewicz & Miller133 for recent reviews of synthetic sideromycins). In McC-like compounds, the C-terminal residue (isoAsn) enzymatically attached to the nucleoside core through the N–P bond limits the specificity of processed compounds towards AspRS. However, McC analogs containing various terminal amino acids attached to adenosine-5’-sulfamoyl with nonhydrolyzable sulfonamide bond were obtained via total chemical synthesis. These compounds retain antibacterial activity and are potent inhibitors of corresponding aaRSs, thus expanding the targeting potential of natural McC variants and validating peptide-nucleotide scaffolds as a platform for the creation of new synthetic antibacterials.134,135

The results of our limited bioinformatics analyses show that the diversity of the three classes of molecules reviewed here is not yet completely tapped. When validated experimentally, these compounds may become viable antibiotics. Most interestingly, the products of putative agn-like BGCs revealed by our searches in the genomes of Actinobacteria may have different intracellular targets that the known compounds and thus could provide alternative means of biocontrol in agriculture.

Conflicts of interest

Authors declare no conflict of interests.

Acknowledgements

The work was supported by RFBR grant No 20-34-90098 to DYT and NIH AI1172190 to Satish K. Nair and KS. We would like to thank Dr Daria Tsibulskaya and Dr Marina Serebryakova for invaluable help and Dr Peter Mergaert, Dr Solange Morera, Konstantin Gilep and Dr Alexey Kulikovsky for critical reading of the manuscript and discussions.

References


