RSC
Chemical Biology

ROYAL SOCIETY
OF **CHEMISTRY**

PAPER
Max Crüsemann *et al.*
Biosynthesis of the corallorazines, a widespread class of
antibiotic cyclic lipodipeptides

RSC Chemical Biology

Biosynthesis of the corallorazines, a widespread class of antibiotic cyclic lipodipeptides†

Teresa M. Dreckmann,[a] Lisa Fritz,[b] Christian F. Kaiser, [ID][a] Sarah M. Bouhired,[a] Daniel A. Wirtz, [ID][a] Marvin Rausch, [ID][b] Anna Müller, [ID][b] Tanja Schneider, [ID][bc] Gabriele M. König [ID][a] and Max Crüsemann [ID]*[a]

Corallorazines are cyclic lipodipeptide natural products produced by the myxobacterium *Corallococcus coralloides* B035. To decipher the basis of corallorazine biosynthesis, the corallorazine nonribosomal peptide synthetase (NRPS) biosynthetic gene cluster *crz* was identified and analyzed in detail. Here, we present a model of corallorazine biosynthesis, supported by bioinformatic analyses and *in vitro* investigations on the bimodular NRPS synthesizing the corallorazine core. Corallorazine biosynthesis shows several distinct features, such as the presence of a dehydrating condensation domain, and a unique split adenylation domain on two open reading frames. Using an alternative fatty acyl starter unit, the first steps of corallorazine biosynthesis were characterized *in vitro*, supporting our biosynthetic model. The dehydrating condensation domain was bioinformatically analyzed in detail and compared to other modifying C domains, revealing unreported specific sequence motifs for this domain subfamily. Using global bioinformatics analyses, we show that the *crz* gene cluster family is widespread among bacteria and encodes notable chemical diversity. Corallorazine A displays moderate antimicrobial activity against selected Gram-positive and Gram-negative bacteria. Mode of action studies comprising whole cell analysis and *in vitro* test systems revealed that corallorazine A inhibits bacterial transcription by targeting the DNA-dependent RNA polymerase.

Introduction

Soil-dwelling myxobacteria (Myxococcales) have been proven to be prolific producers of bioactive specialized metabolites with promise for pharmaceutical applications.[1,2] Their highly developed secondary metabolism allows them to orchestrate a complex and unusual lifestyle: solitary cells aggregate into a collective preying swarm, with individual cells adopting different morphologies. In nutrient-deficient conditions, these cells form fruiting bodies, where central cells become persistent myxospores.[3] Myxobacteria have large genomes of up to 16 Mbp, that contain large numbers of biosynthetic gene clusters (BGCs), encoding the biosynthesis of various bioactive, specialized metabolites. Many of these are still unexplored, making myxobacteria a popular object of natural product research.[1]

The myxobacterial strain *Corallococcus coralloides* B035, isolated from a Belgian soil sample, is known to synthesize the polyketide/nonribosomal peptide hybrid molecule corallopyronin A,[4] which is a promising antibiotic in the fight against filariasis.[5,6] During investigations on the corallopyronins, a novel group of secondary metabolites named corallorazines was isolated from *C. coralloides* B035 (Fig. 1).[7] The structure of the main metabolite corallorazine A was identified as a dipeptide core, composed of *N*-methylglycine and dehydroalanine (Dha) cyclized *via* a hemiaminal forming a piperazine ring. The Dha residue is additionally acylated with the rare branched fatty acid (2*E*,4*Z*)-iso-octa-2,4-dienoic acid. The less complex derivatives corallorazine B and C were identified as probable intermediates of corallorazine biosynthesis, as both are lacking the glycine residue and the latter compound also the *N*-methyl group. A first investigation into the biosynthetic origin of corallorazine A was carried out by employing feeding experiments with 1-^{13}C labeled glycine, leading to a significant enhancement of the respective glycine carbon signal as well as that of the Dha residue in the ^{13}C NMR spectrum. The authors thus postulated the initial incorporation of serine, which is linked to glycine biosynthesis by the serine-hydroxymethyl transferase, and a subsequent dehydration event to form the final Dha building block. More recently, we fully sequenced

[a] *Institute of Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany. E-mail: cruesemann@uni-bonn.de*

[b] *Institute for Pharmaceutical Microbiology, University Hospital Bonn, University of Bonn, Meckenheimer Allee 168, 53115 Bonn, Germany*

[c] *German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Bonn, Germany*

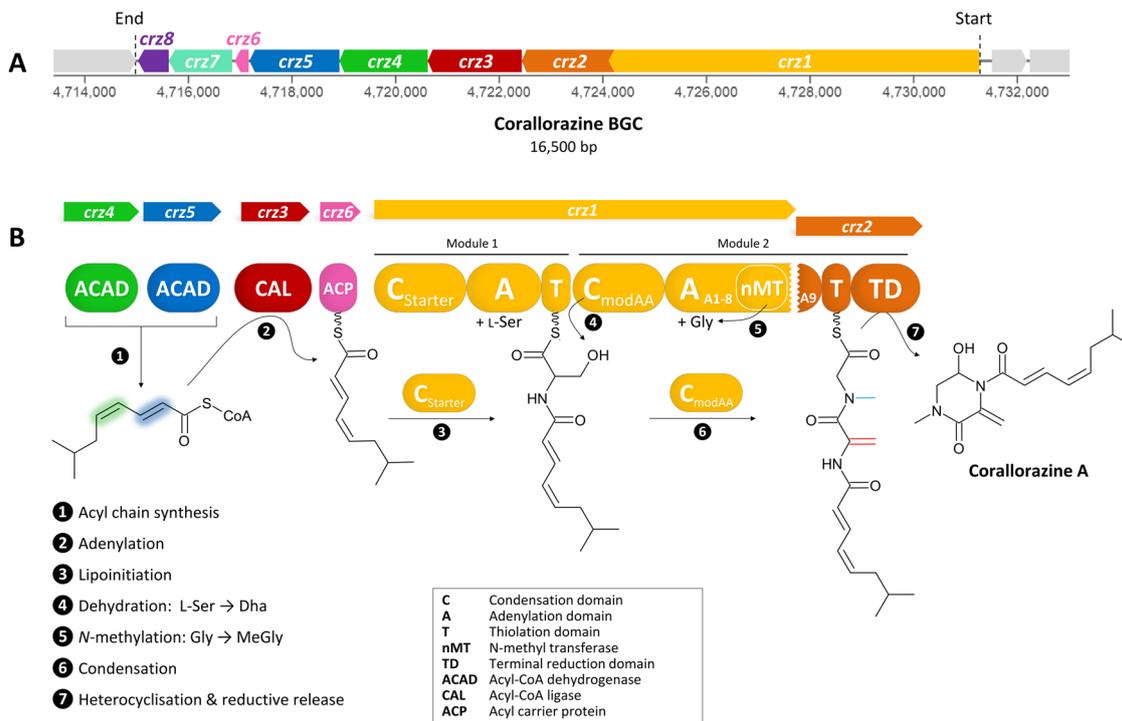† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cb00157e

Fig. 1 (A) The *crz* BGC. (B) Model of corallorazine A–C biosynthesis. The double bond resulting from dehydration of L-serine is marked in red, the *N*-methyl group resulting from SAM-dependent methylation of glycine is marked in blue. Abbreviations are explained in the box. A more detailed model of corallorazine biosynthesis is depicted in Fig. S1 (ESI†).

C. coralloides B035, allowing us to mine its genome for corallorazine biosynthesis genes.[8]

Here, we report the identification and detailed analysis of the corallorazine BGC. We provide a biosynthetic model and investigate corallorazine biosynthesis by *in vitro* experiments on purified non-ribosomal peptide synthetase (NRPS) multidomain proteins. We perform detailed bioinformatic analyses of the dehydrating C domain Crz1C₂ and compare it with other modifying C domains. We analyze the distribution of the corallorazine gene cluster family and its encoded chemical diversity. Furthermore, we re-evaluate the antibacterial activity of corallorazine A, revealing activity towards *Staphylococcus aureus* strains and identify RNA polymerase as antibiotic target.

Results and discussion

Identification of and bioinformatic analysis of the *crz* BGC

To identify corallorazine biosynthesis genes, the genome sequence of *C. coralloides* B035 was analyzed by the bioinformatic pipeline antiSMASH 7.1.0,[9] revealing 37 natural product BGCs. Some of the detected BGCs appear to be superclusters predicted to synthesize more than one set of specialized metabolites. As we anticipated an NRPS being involved in corallorazine A biosynthesis, we focused on three NRPS BGCs and nine NRPS/polyketide synthase hybrid BGCs that were detected. Amongst these, only three contained adenylation (A) domains predicted to activate and incorporate serine and

glycine. Out of these three, only one BGC encoded a bimodular NRPS, predicted to be sufficient for corallorazine A biosynthesis.

The supercluster containing the candidate NRPS genes comprises 38 open reading frames (ORFs), spanning over 87,792 kb. We assigned eight of these, namely *crz1–crz8*, to be putatively involved in the biosynthesis of corallorazine A (Fig. 1(A) and Table S1, ESI†). The core biosynthetic genes *crz1* and *crz2* encode for a bimodular NRPS system (Fig. 1(B)). The encoded megaenzyme Crz1 has the domain sequence C₁-A₁-T₁-C₂-A₂-nMT, suggesting that the second module is split across two ORFs, as its thiolation (T) domain is absent. A preliminary condensation (C) domain analysis with representative sequences retrieved from NaPDoS[10] revealed that C₁ clades with C_Starter domains, catalyzing *N*-acylation of the adjacent T domain-bound amino acids, whereas C₂ is classified as a C_modAA domain. Members of this domain subfamily have been long postulated to catalyze dehydration of carrier protein-bound serine and threonine to yield Dha and dehydrobutyrine, respectively.[10] This hypothesis was recently experimentally confirmed for the biosyntheses of albopeptide[11] and the antibiotic 2-amino-4-methoxy-*trans*-3-butenoic acid,[12] in the latter case followed by structural characterization of the dehydrating domain. The proposed substrate specificities of the Crz A domains A₁ and A₂ were L-Ser and Gly. The A₂ domain additionally harbours a type I SAM-dependent *N*-methyltransferase (nMT) domain.[13] The much smaller ORF *crz2* only encodes the T domain of the second module, as well as a terminating thioreductase (TD) domain,

which is known to catalyse the reductive release and cyclisation of nonribosomal peptides at the expense of NAD(P)H.[14] Taken together, all these features align well with formation of the dipeptidic Dha-N-methyl-Gly core of corallorazine A.

The genes crz3, crz4, crz5, and crz6 encode enzymes that are likely involved in the formation of the unusual acyl residue (2E,4Z)-iso-octa-2,4-dienoic acid. This residue shows significant similarity to branched, dehydrated fatty acid residues found in several lipopeptide antibiotics, such as cadaside, ramoplanin, friulimicin, laspartomycin, and malacidin.[15,16] The substrates for its biosynthesis likely originate from fatty acid metabolism through the condensation of isovaleryl-CoA with two malonyl-CoA units. The genes crz4 and crz5 encode acyl-CoA dehydrogenases, predicted to introduce double bonds in the acyl precursor. Crz4 is predicted to use isovaleryl-CoA as a substrate, while no specific substrate could be predicted for Crz5. In the next step, the acyl-CoA ligase Crz3 is expected to transfer the CoA-bound acyl chain to the acyl carrier protein Crz6, which would ultimately serve as a substrate for Crz1 C_1 (Fig. 1(B) and Fig. S1, ESI†). crz7 and crz8 encode a cytochrome P450 monooxygenase and the corresponding flavine mononucleotide reductase. P450 monooxygenases are known to catalyse different reactions in NRP biosynthesis, such as β-hydroxylation[17] and C–C or C–O–C cyclization.[18] In this case, however, no distinct function could be assigned for Crz7, as the predicted functions for Crz1–6 are sufficient to pose a comprehensive scheme for corallorazine biosynthesis, which is fully depicted in Fig. S1 (ESI†).

In vitro analyses of the corallorazine NRPS

We aimed to verify our biosynthetic hypothesis for corallorazine by *in vitro* reconstitution of single catalytic steps. To this end, first the larger NRPS gene crz1 was cloned and heterologously expressed in *E. coli*. The 7,131 bp gene was separated into two expression constructs, Crz1.1 (containing C_1-A_1-T_1) and Crz1.2 (C_2-A_2-MT) to facilitate the isolated assessment of A and C domain activity, enable efficient expression, and purification. The constructs were cloned into pET28a, expressed in *E. coli* BAP1 with an N-terminal hexahistidinyl tag and purified *via* metal affinity chromatography (Fig. S2, ESI†). In a first step, the substrate specificities of the A domains of Crz1.1 and Crz1.2 were analyzed by the $\gamma^{18}O_4$-ATP exchange assay.[19] We observed specific turnover of L-Ser by Crz1.1, which corroborated the predicted biosynthetic origin of Dha (Fig. 2(A)). However, for Crz1.2, no turnover was observed in presence of Gly. A more detailed examination of the A_2 domain's primary structure revealed that the nMT domain is integrated into the A domain right after core motif A8 and terminates with a stop codon, rendering A_2 on Crz1 inactive. The missing motives A9 and A10 were instead found to be encoded at the beginning of the adjacent ORF crz2 (Fig. S3, ESI†). Thus, we also cloned and heterologously expressed crz2 analogously to the Crz1 constructs. The resulting protein Crz2 (Fig. S2, ESI†) was mixed with Crz1.2 in a 1:1 molar ratio and then assessed for adenylating activity with glycine as substrate. Strikingly, in this setup, we observed an almost quantitative substrate turnover in the $\gamma^{18}O_4$-ATP exchange assay. To our

knowledge, this is the first experimental report of a split, yet functional A domain, encoded on two different genes. This unusual architecture might be the consequence of an integration of the nMT domain or an A subdomain recombination event during evolution, which both were shown to predominantly occur between the motifs A8 and A9.[20]

Next, we conducted a $C_{Starter}$ domain assay, that had previously been established in our laboratory,[21] to test the activity of Crz1-C1. Due to the unavailability of its proposed substrate, (2E,4Z)-iso-octa-2,4-dienoic acid, which is predicted to be bound to acyl carrier protein Crz6, we used octanoyl-CoA as alternative substrate together with L-Ser, and ATP for Crz1.1. Indeed, the activated fatty acid was accepted and we could subsequently detect the formation of octanoyl-serine in the assay (Fig. 2(B)), confirming the proposed function of the Crz-C_1 starter domain and offering the possibility for the targeted generation of corallorazine analogues with altered fatty acid tails.

Bioinformatic analyses of Crz1-C_2 reveal additional features of modifying C domains

We next performed a bioinformatic analysis of Crz1-C_2, which was predicted to be a dehydrating C domain belonging to the C_{modAA} clade.[10] First homologues of this clade were only recently characterized experimentally.[11,12] We first generated a phylogenetic tree including the two C domains of the Crz NRPS and representative C domains of the individual subclasses based on the sequence set of Wheadon and Townsend.[22] To ensure a broad taxonomic coverage and to include additional known subtypes in the analysis, the set was extended with selected C domains from further biosynthetic pathways to yield a total of 275 C domains. The resulting tree is in good agreement with other studies and split into three branches (Fig. S4, ESI†), with an L-clade and a D-clade, representing the C domain family in a narrower sense. In the D-clade, C_{modAA} domains, dual epimerase/condensation (E/C) and ^DC_L domains, condensing D- and L-configured amino acids, each form monophyletic groups.

C_{modAA} domains can further be distinguished by the final modification of their substrate amino acid in addition to simple dehydration. In a previous study, 27 domains from the C_{modAA} subtype were identified and analysed with regard to their type of modification.[12] We aimed to get a more comprehensive picture and expanded this set with 23 additional sequences of potential C_{modAA} domains, including Crz1-C_2, that were identified by a combination of literature and MIBiG[23] analyses (Table S2, ESI†). The resulting phylogenetic tree (Fig. 3) shows that domains catalyzing conjugate addition instead of peptide bond formation and pyrimidine formation each formed their own clades. The pyrrolizidine formation group, which was now extended to six domains, also formed its own clade, while Crz1-C_2 was repeatedly grouped with AmbE C1.

In order to identify conserved amino acid residues and motifs within C_{modAA} domains, potentially offering clues to their mode of action, the alignment (Fig. S5 and S6, ESI†) was
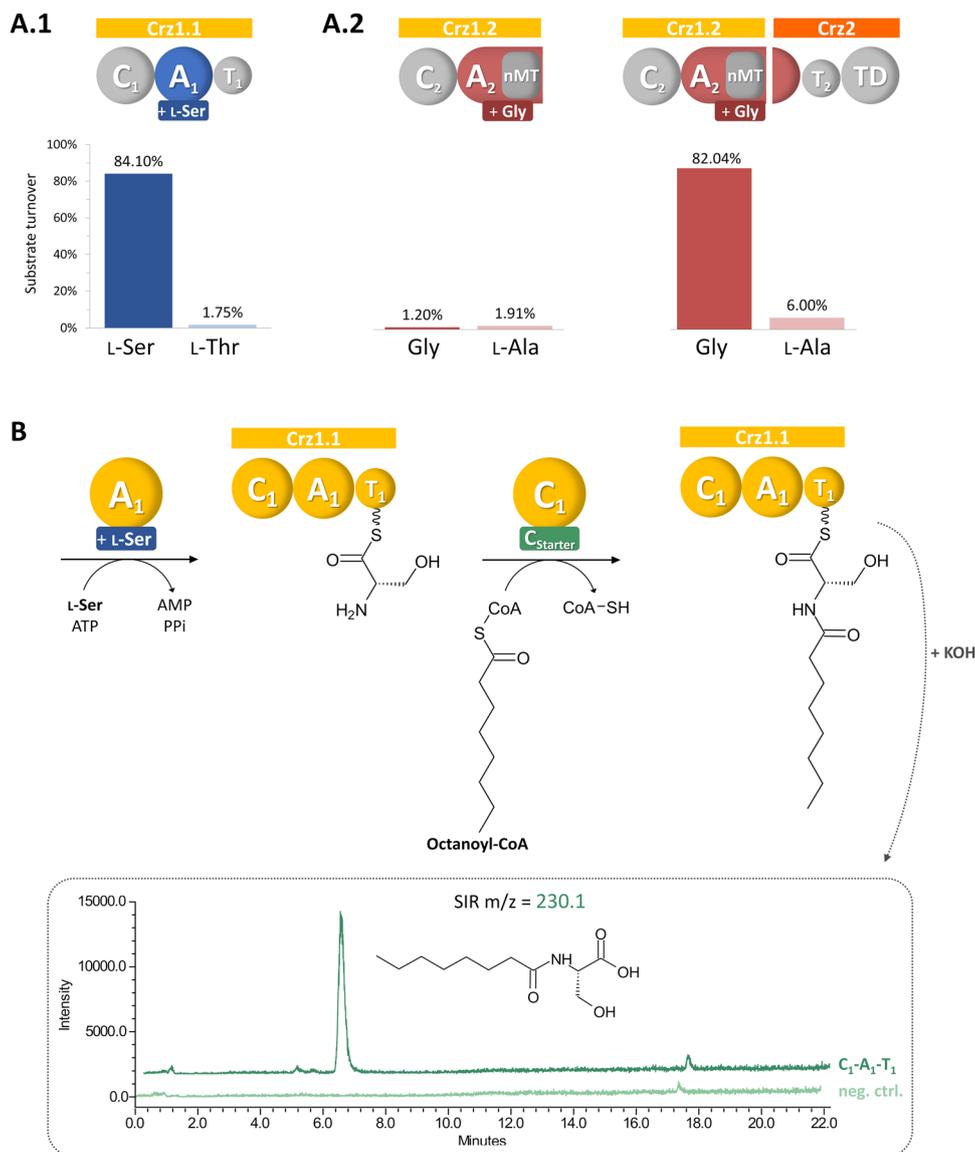
972 | *RSC Chem. Biol.*, 2024, **5**, 970–980

© 2024 The Author(s). Published by the Royal Society of Chemistry

Fig. 2 Biochemical characterization of Crz A domains (A) and Crz1-C_1 (B). Crz A_2 is only active in presence of domain partly encoded on Crz1 and Crz2 (A.2). Crz1-C_1 accepts the alternative substrate octanoyl-CoA, leading to the formation of the biosynthetic intermediate octanoylserine, which is detected in LCMS after hydrolysis in single ion recording (SIR) mode.

further investigated and also compared to $^{D}C_{L}$ and dual E/C domain sequences, related subclasses within the D-clade (Fig. S7, ESI†). Previously, several well conserved residues located near the active centre or the donor side tunnel of AmbE-C_1 were identified.[12] For example, mutagenesis of active site residues R1605 and H1632 significantly reduced the activity of AmbE-C_2, indicating their importance for the function of C_{modAA} domains. In our alignment, R1605 (K1325 in Crz1-C_2), that was predicted to take part in substrate positioning,[12] shows functional conservation in all analysed C_{modAA} domains (78% K, 22% R). Several other residues were found to be highly to moderately conserved among all C_{modAA} domains in the alignment (Fig. 3). Noticeably, some of these were located in inter-motif regions, and partly arranged in motif-like sections that were not present in this form in any of the other

representatives of the D-clade (Fig. S7, ESI†). Particularly striking was a section directly located N-terminal of the C5 motif, which extends over approx. 8 amino acids and, in a structural model generated with AlphaFold, is part of a helix on the donor side of the substrate tunnel (Fig. S8 and S9, ESI†).

This helix contains the highly conserved K/R mentioned above, but also a noteworthy histidine (H1332), which is present in 48 of the 50 analyzed C_{modAA} domains. In our structural model, the side chain of H1332 is oriented in the same direction as that of K1325 (Fig. S9, ESI†). Together, K1325 and H1332 seem to delineate the beginning and end of a motif-like region characterized by the consensus sequence [K/R]SVLLAAH, which is, in the model structure, in close proximity to conserved T (T1347, 82%), E (E1393, 88%), and Y (Y1489, 90%) side chains.
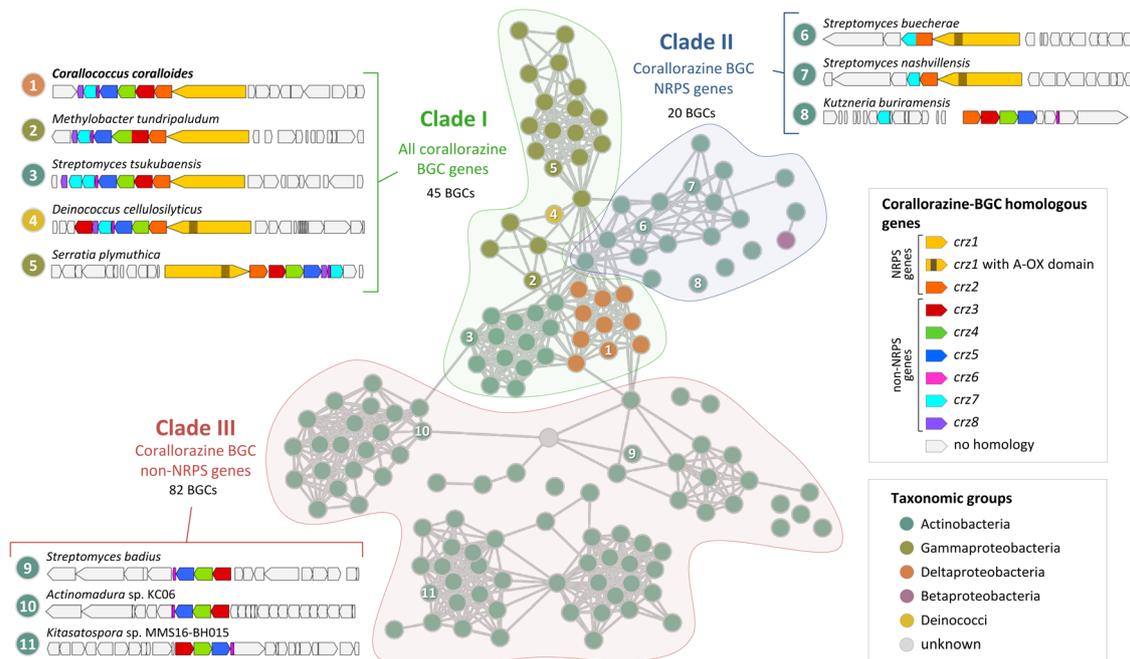
Fig. 3 Phylogenetic analysis of C_{modAA} domains. A maximum likelihood tree of 50 C_{modAA} domains was generated using one randomly selected dual E/C domain as outgroup. A total of 100 bootstrap replicates were performed and bootstrap values above 75 are given on corresponding branches. Domains were assigned colours according to the taxonomic classification of the source organism at phylum level, and by final modification of the amino acid they act on. Conserved residues detected in this study are depicted next to each domain. Visualisation was done with iTol v6.[24]

The HHxxxDG motif in the active site is crucial for the activity of canonical C domains.[25] Our alignment does show strict conservation of the DG motif in C_{modAA} domains, but not of both histidines. In seven cases, the first His, which is thought be crucial for structural integrity, is missing, while in three other cases the second His is mutated. In FclJ-C_2 and Zmn17-C_2, both catalyzing side chain conjugation, the first His is exchanged by a Cys. Given that Cys can also form hydrogen bonds, it might substitute for His in maintaining structural integrity under certain conditions. Interestingly, TnaC-C_3, the third domain in this group, retains both histidines but features a cysteine directly downstream of the second His. Domains involved in conjugate addition (BlmVI-C_2, TlmVI-C_2, and ZbmVI-C_2) lack both histidines, and feature Phe and Gly or Ala instead. Here too, the presence of Phe could induce an active site conformation which promotes addition to the side chain. The absence of the second His, which is attributed a central role in peptide bond formation, aligns with the altered functions of these domains.

Further specific residues for C_{modAA} domains were located in close proximity to the donor substrate tunnel: these include an

R (96%) at the beginning of the C6 motif, followed by a less conserved R (78%), corresponding to R1400/R1401 in Crz1, and R1679/R1680 in AmbE, respectively. The model shows both arginine side chains to be located near the substrate tunnel entrance and are facing the side chain of a conserved E (E1365, 88%) located directly opposite. Mutating R1679 or R1680 to Ala had no significant effect on its activity.[12] However, it is feasible that one Arg might substitute for the other. In fact, each of the C_{modAA} domains contains either both or at least one of the arginines, which supports their predicted importance for C_{modAA} domain functionality.

Distribution of the corallorazine BGC

During our bioinformatic analyses, we identified multiple highly similar homologues of *crz* genes in the NCBI database. To gain insight into the global distribution and evolution of the *crz* BGC, we extracted all corresponding BGCs from NCBI and submitted them to a network analysis with BiG-SCAPE.[26] At a cutoff of 0.7, the *crz* BGC clusters with 146 other BGCs, which originate from diverse bacterial genera, including rare natural product producers such as Deinococcus. Most BGCs of the *crz*

974 | *RSC Chem. Biol.*, 2024, **5**, 970–980

© 2024 The Author(s). Published by the Royal Society of Chemistry

Fig. 4 Cytoscape network visualization of homologous corallorazine BGCs. Nodes correspond to BGCs containing at least one gene that is homologous to a gene in the corallorazine BGC (A1, red circle outline). Information on host strain and presence of *crz1–8* homologues is displayed in Table S3 (ESI†). Classification in clade I, II or III is indicated by colored areas. Examples for characteristic clusters from these clades are shown in this figure and in Fig. S11 (ESI†). Length of the connections between nodes correlates with the degree of relationship according to BiG-SCAPE analysis.

GCF cluster were detected in actinomycete bacteria, which however may be biased by the high abundance of sequenced members of this genus in the genome databases. BGCs in the resulting network were roughly divided into three clades (I–III) (Fig. 4, Fig. S10, S11 and Tables S3, S4, ESI†).

Clade I consists of BGCs harbouring homologues of all eight *crz* genes, with *crz7* duplicated in most *Streptomyces* species. The most intriguing additional feature is an additional mono-oxygenase domain (MOx) within A_2, detected in 38 pathways, notably all from members of the Gammaproteobacteria genera *Serratia* and *Pseudomonas*. This MOx domain was previously reported from myxothiazol and melithiazol biosynthesis, where it is also encoded within glycine-recruiting A domains.[27] In the respective pathways, the MOx domain hydroxylates C_α, leading to the spontaneous dissimilation of an amine due to the unstable $N–C_\alpha$ bond. The remaining T domain-bound glyoxylate is subsequently released by the terminating TE domain. The presence of additional MOx domains in *crz*-like BGCs suggests the formation of derivatives with truncated, linear peptide cores by these pathways, similar to the shunt products corallorazine B/C.

Clade II of the network analysis contains BGCs that encode homologues of *crz1* and *crz2*, but mostly lack genes corresponding to *crz3–6*, suggesting production of molecules with a cyclic dipeptide core, but lacking the acyl side chain. Interestingly, genes encoding for a P450 monooxygenase and the associated FMN reductase, corresponding to *crz7* and *crz8*, respectively, are present in almost all BGCs from clade II. In five cases, *e.g.* cluster 6 from *Streptomyces buecherae* (Fig. 4), the

P450 is even fused to the NRPS. This constellation implies an important function of the P450 for biosynthesis of this compound class. However, our current biosynthetic model does not involve steps catalysed by Crz7 and Crz8. Instead, these enzymes might also play a role in side chain biosynthesis or self-resistance.[28] The BGCs of clade III harbour genes corresponding to *crz3–6* that are located in close proximity to genes encoding various NRPSs- and/or polyketide synthases (PKS). This indicates a conserved interplay of the encoded enzymes and the collective horizontal transfer of the corresponding genes, yielding structurally diverse compounds with branched, unsaturated fatty acid side chains, such as the calcium-dependent antibiotics of cadasides, malacidins, and taromycins.[16,29]

Antibiotic activity and mechanism of action

Finally, we determined the antibacterial activity of corallorazine A against a panel of selected Gram-negative and Gram-positive test strains (Table 1). In our previous study, only the activity against *Bacillus* and *Escherichia coli* was tested,[7] which prompted us to include other pathogenic bacteria in the bioassays. The compound shows moderate activity against Gram-positive pathogens with the highest activity against methicillin-resistant *Staphylococcus aureus* strain COL (MIC 4–8 μg mL^{-1}). Corallorazine A lacks activity against *E. coli*, although the permeable efflux-negative *E. coli* strain MB5746 was susceptible, referring to the outer membrane as a crucial barrier for corallorazine A.[30]

To determine the target pathway, pathway-selective β-galactosidase bioreporter strains were treated with corallorazine A.

© 2024 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2024, **5**, 970–980 | **975**

Table 1 Activity of corallorazine A against selected bacterial strains. MSSA, methicillin-sensitive *S. aureus*; MRSA, methicillin-resistant *S. aureus*, VISA, vancomycin-intermediate *S. aureus*, VRE, vancomycin-resistant enterococci

Strain	MIC [μg mL^{-1}]
Staphylococcus aureus SG511 (MSSA)	16
Staphylococcus aureus RN4220 (MSSA)	32
Staphylococcus aureus HG003 (MSSA)	32
Staphylococcus aureus Mu50 (VISA, rifampin^R)	8
Staphylococcus aureus COL (MRSA)	4–8
Enterococcus faecalis BM4223 (VRE)	16
Bacillus subtilis 168	128
Escherichia coli I-112768	>128
Escherichia coli MB5746	16

Expression of LacZ was specifically induced in *B. subtilis* P$_{yvgS}$-*lacZ* indicating interference with RNA biosynthesis while DNA, protein and cell wall biosyntheses remained unaffected (Fig. 5(A)). In a next step, we used conditional antisense RNA expression to reduce expression of putative target proteins involved in transcription and translation to confirm RNA biosynthesis pathway as a target for corallorazine A (Fig. 5(B)). Reduction of target protein results in hypersensitivity towards the inhibitor. While the reduction of expression levels of RNA polymerase Sigma factor SigA (*sigA*) and translation initiation factor IF-1 (*infA*) had no effect, increased susceptibility towards corallorazine A was observed when the RNA polymerase beta subunit (*rpoC*) and the 30S ribosomal protein S5 (*rpsE*) were downregulated, indicating the latter being molecular targets. The direct effect of corallorazine A on the RNA polymerase from *E. coli* was examined in an *in vitro* transcription assay (Fig. 5(C)). The results show that corallorazine A inhibits transcription in a concentration-dependent manner, suggesting direct interaction with RNA polymerase.

Conclusions

We identified and analyzed the corallorazine BGC from *C. coralloides* B035 and substantiated our hypothesis for corallorazine biosynthesis by assessing the catalytic activity of single NRPS domains *in vitro*. The respective bimodular NRPS is encoded by two genes *crz1* and *crz2*, whereby the gene borders are within the encoded A domain of the second module. Most parts of this A domain, together with a conventionally embedded *N*-methyltransferase are encoded on *crz1*, whereas the smaller A subdomain, however crucial for catalytic activity, is encoded on *crz2*. This unusual constellation raises questions about its evolutionary origin. In case of an erroneous recombination event, this position would indicate a good starting point for artificial reengineering of NRPS assembly lines.

The minor, linear derivatives corallorazine B and C appear to be biosynthetic shunt products of the assembly line, possibly caused by this unusual split, as both are lacking the glycine residue. In addition, our detailed bioinformatic analyses of the dehydrating Crz1-C$_2$ and related C$_{modAA}$ domains revealed previously not recognized conserved sequence features of this group of modifying C domains, including modifications of the highly conserved active site HHxxxDG motif for certain domains.
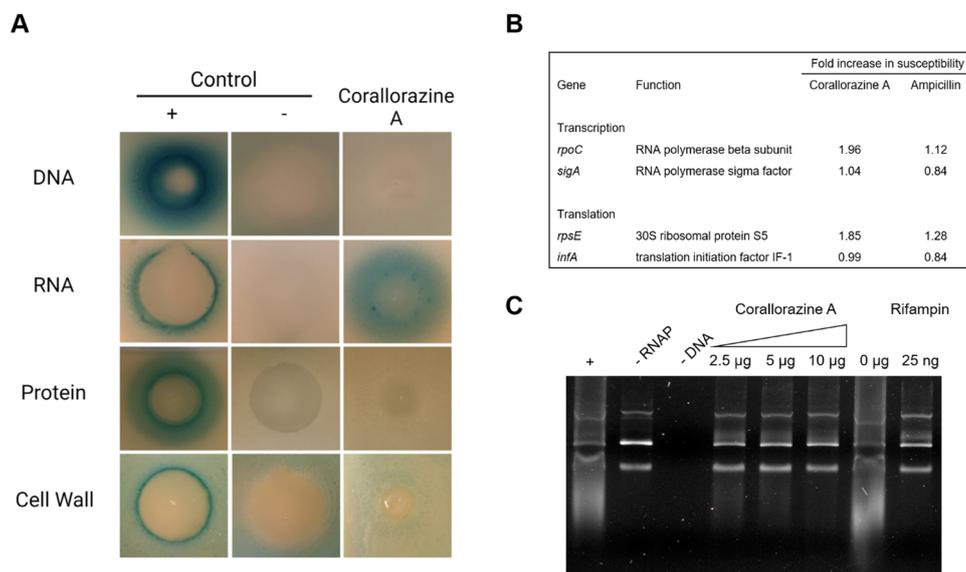

Fig. 5 Corallorazine A targets the DNA-dependent RNA polymerase. (A) The effect of corallorazine A on major macromolecular biosyntheses in *B. subtilis*. *B. subtilis* bioreporter strains contain selected promoter-*lacZ* fusions to identify interference with DNA (P$_{yorB}$), RNA (P$_{yvgS}$), protein (P$_{yheI}$), and cell wall (P$_{ypuA}$) biosynthesis. A blue halo around the inhibition zone indicates induction of a specific stress response and resulting β-galactosidase expression. The antibiotics ciprofloxacin, rifampicin, clindamycin, and vancomycin served as positive controls. Corallorazine A specifically induces P$_{yvgS}$ indicative of interference with RNA biosynthesis. (B) Corallorazine A susceptibility was determined in cells with reduced expression levels of proteins involved in transcription and translation. Expression of *rpoC* and *rpsE* antisense RNA resulted in sensitization towards corallorazine A. The β-lactam antibiotic ampicillin was used as control. (C) Inhibition of RNA polymerase activity by corallorazine A *in vitro*. Agarose gel electrophoresis shows that *E. coli* RNA polymerase *in vitro* transcription activity was reduced in presence of increasing concentrations of corallorazine A.

976 | *RSC Chem. Biol.*, 2024, **5**, 970–980

© 2024 The Author(s). Published by the Royal Society of Chemistry

Our network analysis unveiled the *crz* BGC and closely related homologues thereof to be surprisingly widely distributed among bacteria from different phyla and allowed a comparative global analysis of the corallorazine-like GCF. In several cases, the presence of an additional MOx domain within the glycine-incorporating A domain indicates the production of oxidized or smaller, linear corallorazine derivatives. The widespread occurrence of BGCs for the assembly of corallorazine-like compounds suggests specialized bioactivities and ecological relevance of the encoded molecules, which may also be further investigated in view of their pharmaceutical potential. Finally, unexpected frequent co-occurrence of genes encoding a P450 monooxygenase and its corresponding FMN reductase along with the corallorazine NRPS genes in homologous BGCs give reason for further investigation of their potential impact in corallorazine biosynthesis.

Experimental

Biological material and cultivation

The myxobacterial strain *Corallococcus coralloides* B035 was isolated from a soil sample, collected from Belgium.[31] For bacterial growth, 250 µL of a cryo culture of *C. coralloides* B035 was placed on VY/2 agar plates (5.0 g L^{-1} yeast extract, 1.36 g L^{-1} CaCl$_2$·2H$_2$O, 0.5 mg L^{-1} vitamin B12 and 15.0 g L^{-1} agar). For cultivation in liquid medium, slices of agar, containing fruiting bodies of the myxobacterium, were inoculated in MD-1 + glucose medium (3.0 g L^{-1} casitone, 0.7 g CaCl$_2$·2H$_2$O, 2.0 g L^{-1} MgSO$_4$·7H$_2$O, and 2.2 g L^{-1} glucose·H$_2$O) and grown at 30 °C for 7 days under shaking conditions of 140 rpm. For long term storage at −80 °C, the liquid bacterial cultures were mixed with an equal volume of 1% casitone. Liquid cultures of all *E. coli* were grown in LB medium supplemented with the appropriate antibiotic. The *E. coli* strains were cultivated at 220 rpm and 37 °C. *E. coli* precultures were inoculated from cryo-cultures stored at 180 °C.

Genomic DNA isolation

C. coralloides B035 was grown in liquid MD1 + glucose medium at 30 °C (140 rpm) for seven days. Gel electrophoresis materials were supplied from Neolab, Germany. All other materials and substances were supplied by Roth, Germany. Genomic DNA isolation was performed as described in.[8]

Molecular cloning

Oligonucleotide primers were synthesized by Eurofins Genomics. Primers listed in Table S5 were used for amplification of whole *crz* genes or the truncated NRPS domain constructs of *crz1*. The resulting products of Q5-PCR (NEB) were cloned into the multiple cloning site of pET28a, using the respective endonucleases and T4 ligase (both NEB). The resulting constructs were isolated from *E. coli* alpha-select silver (Bioline), analyzed via restriction digest and verified by Sanger sequencing. The final plasmids were then transformed into *E. coli* BAP1 for overexpression of recombinant hexahistidine-tagged proteins.

Heterologous protein expression and purification

Expression of NRPS constructs was performed in *E. coli* BAP1 to ensure *in vivo* phosphopantetheinylation of the T domains.[32] A glycerol stock of the respective *E. coli* BAP1 culture was used to inoculate 5 mL of LB medium, supplemented with 50 ng mL^{-1} kanamycin for plasmid selection and incubated at 37 °C, 220 rpm overnight. This densely-grown preculture was used to inoculate 250 mL TB in a baffled flask, supplemented with kanamycin. After roughly 2 h incubation (37 °C, 220 rpm), OD$_{600}$ reached 0.7 and the culture was chilled on ice. Protein expression was induced by addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was then incubated for additional 16 h (16 °C, 200 rpm). For extraction of the proteins, cells were harvested via centrifugation (10 000 × *g*, 4 °C, 2 min). The cells were resuspended in 5 mL lysis buffer (50 mM NaH$_2$PO$_4$, 10 mM imidazole, 300 mM NaCl, pH 8) per g pellet. The cells were lysed via sonication and the lysate centrifuged (12 000 × *g*, 4 °C), until lysate was clear. The supernatant, containing the hexahistidine-tagged protein was incubated on ice with 1 mL of Ni-NTA-agarose suspension (Qiagen) for 1 h at 100 rpm. The suspension was finally filtered with a propylene column (Qiagen), washed with 4 mL wash buffer (50 mM NaH$_2$PO$_4$, 20 mM imidazole, 300 mM NaCl, pH 8) and eluted in 2.2 mL elution buffer (50 mM NaH$_2$PO$_4$, 300 mM imidazole, 300 mM NaCl, pH 8). The buffer system of the elution fraction was exchanged with PD-10 columns (Cytiva), following the manufacturers protocol. The desired concentration as set by centrifugation with Amicon membrane filter (Merck KGaA) with a MWCO of 30 kDa or 10 kDa.

γ^{18}O$_4$ATP-exchange assay

2 µL of the A domain containing NRPS (5 µM in 20 mM tris pH 7.5, 5% glycerol) was mixed with 2 µL solution 1 (3 mM amino acid, 15 mM pyrophosphate, 20 mM tris pH 7.5) and 2 µL solution 2 (3 mM γ^{18}O$_4$-ATP, 15 mM MgCl$_2$, 20 mM tris pH 7.5) and incubated at 22 °C for 1 h. The reaction was quenched with 6 µL 9-aminoacridine in acetone (5 mg mL^{-1}). Precipitated proteins were removed via centrifugation. For sample analysis, 1 µL of the sample were spotted on a carrier plate of ground steel and analyzed with a Bruker AutoFlex III MALDI-TOF-MS in negative mode. Absolute substrate conversion in [%] was calculated by dividing the peak area at *m/z* 506 through the combined peak areas at *m/z* 508, 510, 512, and 514, divided by 83.33 for the molar ratio of labelled against unlabeled pyrophosphate in the assay.[19]

C$_{Starter}$ domain assay

Crz1.1, with the domain architecture C$_1$-A$_1$-T$_1$, was isolated from *E. coli* BAP1 and buffered to assay buffer (25 mM tris pH 7.5, 150 mM NaCl, 10 mM MgCl$_2$) using PD-10 columns, following the manufacturers protocol. The assay was performed in a 500 µL one-pot-reaction: 12.5 µM Crz1.1 were incubated with 1 mM ATP, 1 mM L-Ser, and 0.5 mM octanoyl-CoA (CoALA) for 3 h at 22 °C. For the negative control, the enzyme was inactivated beforehand at 80 °C for 10 min. After incubation,

© 2024 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2024, **5**, 970–980 | **977**

proteins were precipitated with 10% trichloroacetic acid (final concentration) and incubated for 30 min on ice. Subsequently, the acid was removed by 3 washing steps with ddH$_2$O. The precipitated proteins were then resuspended in 0.1 M KOH and incubated at 70 °C for 20 min to cleave the thioester-T domain bonds upon alkaline hydrolysis. Samples were dehydrated *via* lyophilization and dissolved in a minimal volume of MeOH for LC–MS analysis using a Waters Alliance e2695 Separation Module, equipped with a RP-18 column (Waters XBridge BEH Shield RP18 Column; 3.5 μm, 2.1 × 100 mm). A binary solvent mixture of A: 90% acetonitrile + 10% ammonium acetate (5 mM) pH 7.4 and B: 10% acetonitrile + 90% ammonium acetate (5 mM) pH 7.4 was used, with a flow rate of 0.3 mL min^{-1}, starting at minute 0–5 with isocratic elution of 10% B, a gradient 10% B → 100% at minute 5–25, and again isocratic conditions of 100% B at minute 25–30. Injection volume of MeOH sample was 5 μL. Signals were detected in negative ion mode for total ion current (TIC) and for selected ion recording (SIR) at m/z = 230.1 for octanoyl-serine.

Bioinformatic analyses of condensation domains

In order to generate a set of representative C domain sequences from bacterial BGCs, mainly a reduced set of sequences from Wheadon and Townsend[22] was used. This set includes E, Cyc, $^L C_L$, $C_{Starter}$, C_{modAA}, dual E/C and $^D C_L$ domains from different biosynthetic pathways deposited on MIBiG.[23] To ensure a broad taxonomic coverage and to include additional known subtypes (hybridC, X and I) in the analysis, this set was extended with C domains from further bacterial biosynthesis pathways, which were also retrieved from clusters deposited on MIBiG or taken from literature sources. Selection of C_{modAA} sequences was based on the set used by Patteson *et al.*[12] and extended by manual searches in the MiBIG repository. Extraction of domain sequences was done *via* InterPro[33] using the general profile entry for C domains (IPR001242) with occasionally extending sequences on the sides, similarly to the procedure in.[22] Multiple sequence alignments of C domain sequences were performed using MUSCLE v3.8.31[34] algorithm implemented in Jalview v2.11.3.2[35] with default settings. Resulting alignments were then manually trimmed at both ends in order to remove overhangs and poorly aligned regions. Alignments were displayed as sequence logos using WebLogo 3.[36] For better visualization and comparability, alignment positions with gaps of more than 90% were removed for generation of logos, with the exception of affected positions in the sequences of Crz1 C$_2$ and AmbE C$_2$ from the C$_{modAA}$ alignment. C domain core motifs were identified manually according to He *et al.*[37] and Rausch *et al.*[38] Phylogenetic reconstructions were then performed using raxmlGUI v2.0.10,[39] an application that provides a graphical interface for the phylogeny tool RAxML v8.0.[40] First, applying ModelTest-NG,[41] which is implemented in raxmlGUI, the best substitution matrix for each alignment was determined. For both the set of different subtypes of the C domain superfamily, as well as the set comprising only C$_{modAA}$ domains, the LG model[42] was the best fit, with the decorations +I +G4 +F being determined for the former set and +G4 +F for the latter. With

these settings, a maximum likelihood tree with 100 bootstrap replicates was created and visualised with ITOL v.6.[24]

Structural model of Crz1-C$_2$

The structure prediction for Crz1-C$_2$ was retrieved with Alpha-Fold v2.3.2 Colab[43] using default settings. For this purpose, Crz1-C$_2$ was specified ranging from position 1076 to 1516 within the protein sequence of Crz1 (QAT85340.1). Visualization was done using UCSF ChimeraX v1.7.1.[43]

Bioinformatic analyses of corallorazine BGC and homologues

The representatively chosen amino acid sequences of the proteins Crz1 (QAT85340.1), Crz2 (QAT85339.1) and Crz4 (QAT85337.1), which were considered characteristic for the corallorazine BGC, were each subjected to a global alignment with sequences from the National Center for Biotechnology Information (NCBI) databank, using protein–protein BLAST[44] at standard parameters with filter setting from 41–100% identity. Nucleotide regions, on which the matched protein's coding regions were localized, were screened for BGCs using anti-SMASH 7.1.0[9] in relaxed mode. Corresponding annotated BGC containing regions where collected and subjected to a BiG-SCAPE analysis[26] using mixed mode at a 0.7 cutoff level. The resulting network was visualized with Cytoscape 3.5.1. Pre-sorted gene clusters were revisited and manually categorized, based on the presence of genes homologous to *crz1–8*. Some examples of gene clusters along neighboring regions were visualized using Gene Graphics[45] and genes were colorized according to their predicted product, if information was available.

Determination of minimum inhibitory concentrations (MICs)

Corallorazine A was isolated from *Corallococcus coralloides* B035 as previously described.[7] MICs were determined by broth micro-dilution according to CLSI guidelines. Test medium was cation-adjusted Mueller–Hinton broth (MHB) and cell concentrations were adjusted to approximately 5 × 10^5 cells per ml. After 20 h of incubation at 37 °C, the MIC was defined as the lowest concentration of antibiotic that inhibited visible growth.

Bacillus subtilis β-galactosidase reporter assays

B. subtilis β-galactosidase reporter assays were performed as previously described.[46] In short, the reporter strains *B. subtilis* 168 pAC6-P$_{ypuA}$ (cell wall), pAC6-P$_{yorB}$ (DNA), pAC6-P$_{yvgS}$ (RNA) and pAC6-P$_{yheI}$ (protein) were grown in MHB supplemented with 5 μg mL^{-1} chloramphenicol at 30 °C to an OD$_{600}$ of 0.5. MHA plates containing 5 μg mL^{-1} chloramphenicol with 150 μg mL^{-1} (cell wall reporter), 250 μg mL^{-1} (DNA reporter), and 500 μg mL^{-1} (RNA and protein reporter) X-gal, respectively, and 1 × 10^7 CFU mL^{-1} of the respective strain were poured at 55 °C. After cooling down, 5 μg corallorazine A as well as positive and negative control antibiotics (6 μg vancomycin and 3 μg clinda-mycin for cell wall, 0.3 μg ciprofloxacin and 3 μg clindamycin for DNA, 6 μg rifampin and 3 μg clindamycin for RNA, 3 μg clindamycin and 6 μg vancomycin for protein reporter) were spotted and allowed to dry before incubating overnight at 30 °C.

978 | *RSC Chem. Biol.*, 2024, **5**, 970–980

© 2024 The Author(s). Published by the Royal Society of Chemistry

Susceptibility profiling of *S. aureus* strains, depleted for proteins involved in transcription and translation mediated by antisense RNA expression

S. aureus RN4220 strains, in which specific antisense RNAs were conditionally expressed to reduce expression levels of proteins involved in transcription and translation, were tested for susceptibility towards corallorazine A as reported previously.[47] First, the xylose concentration which inhibits the visible growth of the respective antisense strains was determined as described above (determination of MICs). Overnight cultures grown in LB containing 34 µg mL^{-1} chloramphenicol were diluted to an OD$_{600}$ of 0.01 in the LB medium supplemented with 0×, 0.25×, 0.5× and 1× MIC of xylose, respectively. Cell suspensions were transferred into a transparent 96 chimney well plate (Greiner) containing serially diluted corallorazine A or ampicillin as control. Growth was monitored for 16 h at 37 °C using a Tecan Spark 10 M microplate reader. The increase in susceptibility was calculated as previously described.[48]

In vitro transcription using *E. coli* RNA polymerase

Unless otherwise stated, transcription was assessed using 0.5 units *E. coli* RNA polymerase holoenzyme (New England Biolabs), 0.5 µg of ΦX174 RF DNA, and corallorazine A in indicated amounts in 1× reaction buffer (40 mM Tris–HCl, 150 mM KCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, 0.01% Triton X-100, 0.5 mM of each NTP) in a total volume of 25 µL. The DNA was transcribed at 37 °C for 2 h. After the addition of 25 µL 2x RNA loading dye (ThermoFisher Scientific), the reaction was incubated at 70 °C for 10 min and immediately cooled down on ice for another 3 min. The samples were loaded onto a 1% agarose gel supplemented with 1× GelRed® (Biotium) and run at 100 V. The gel was imaged using the Molecular Imager® Gel Doc™ XR+ imaging system with the Image Lab Software Version (BioRad Laboratories).

Data availability

The data supporting this article (tables of biosynthetic gene clusters and NRPS domains used for the bioinformatic analyses) have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 C. D. Bader, F. Panter and R. Müller, *Biotechnol. Adv.*, 2020, **39**, 107480.

2 M. A. Bhat, A. K. Mishra, M. A. Bhat, M. I. Banday, O. Bashir, I. A. Rather, S. Rahman, A. A. Shah and A. T. Jan, *Pharmaceutics*, 2021, **13**, 1265.

3 P. Cao, A. Dey, C. N. Vassallo and D. Wal, *J. Mol. Biol.*, 2015, **427**, 3709.

4 H. Irschik, R. Jansen, G. Höfle, K. Gerth and H. Reichenbach, *J. Antibiot.*, 1985, **38**, 145.

5 T. F. Schäberle, A. Schiefer, A. Schmitz, G. M. König, A. Hörauf and K. Pfarr, *Int. J. Med. Microbiol.*, 2014, **304**, 72.

6 A. Schiefer, M. P. Hübner, A. Krome, C. Lämmer, A. Ehrens, T. Aden, M. Koschel, H. Neufeld, L. Chaverra-Muñoz, R. Jansen, S. Kehraus, G. M. König, D. Pogorevc, R. Müller, M. Stadler, S. Hüttel, T. Hesterkamp, K. Wagner, K. Pfarr and A. Hörauf, *PLoS Negl. Trop. Dis.*, 2020, **14**, e0008930.

7 A. Schmitz, S. Kehraus, T. F. Schäberle, E. Neu, C. Almeida, M. Roth and G. M. König, *J. Nat. Prod.*, 2014, **77**, 159.

8 S. Bouhired, O. Rupp, J. Blom, T. F. Schäberle, A. Schiefer, S. Kehraus, K. Pfarr, A. Goesmann, A. Hörauf and G. König, *Microbiol. Resour. Announce.*, 2019, **8**, e00050-19.

9 K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2023, **51**, W46.

10 N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen and P. R. Jensen, *PLoS One*, 2012, 7, e34064.

11 S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law and H. Deng, *Angew. Chem., Int. Ed.*, 2021, **60**, 3229.

12 J. B. Patteson, C. M. Fortinez, A. T. Putz, J. Rodriguez-Rivas, L. H. Bryant, III, K. Adhikari, M. Weigt, T. M. Schmeing and B. Li, *J. Am. Chem. Soc.*, 2022, **144**, 14057.

13 T. A. Lundy, S. Mori and S. Garneau-Tsodikova, *RSC Chem. Biol.*, 2020, **1**, 233.

14 A. Chhabra, A. S. Haque, R. K. Pal, A. Goyal, R. Rai, S. Joshi, S. Panjikar, S. Pasha, R. Sankaranarayanan and R. S. Gokhale, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 5681.

15 M. Strieker and M. A. Marahiel, *ChemBioChem*, 2009, **10**, 607.

16 B. M. Hover, S.-H. Kim, M. Katz, Z. Charlop-Powers, J. G. Owen, M. A. Ternei, J. Maniko, A. B. Estrela, H. Molina and S. Park, *et al.*, *Nat. Microbiol.*, 2018, **3**, 415.

17 S. Uhlmann, R. D. Süssmuth and M. J. Cryle, *ACS Chem. Biol.*, 2013, **8**, 2586.

18 K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch and M. J. Cryle, *Nature*, 2015, **521**, 105; Y. Zhao, E. Marschall, M. Treisman, A. McKay, L. Padva, M. Crüsemann, D. R. Nelson, D. L. Steer, R. B. Schittenhelm, J. Tailhades and M. J. Cryle, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204957.

19 V. V. Phelan, Y. Du, J. A. McLean and B. O. Bachmann, *Chem. Biol.*, 2009, **16**, 473.

20 M. Baunach, S. Chowdhury, P. Stallforth and E. Dittmann, *Mol. Biol. Evol.*, 2021, **38**, 2116; T. A. Lundy, S. Mori,

N. Thamban Chandrika and S. Garneau-Tsodikova, *ACS Chem. Biol.*, 2020, **15**, 282.

21 C. Hermes, R. Richarz, D. A. Wirtz, J. Patt, W. Hanke, S. Kehraus, J. H. Voß, J. Küppers, T. Ohbayashi and V. Namasivayam, *et al.*, *Nat. Commun.*, 2021, **12**, 144.

22 M. J. Wheadon and C. A. Townsend, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2026017118.

23 B. R. Terlouw, K. Blin, J. C. Navarro-Muñoz, N. E. Avalon, M. G. Chevrette and S. Egbert, *et al.*, *Nucleic Acids Res.*, 2023, **51**, D603.

24 I. Letunic and P. Bork, *Nucleic Acids Res.*, 2021, **49**, W293.

25 K. Bloudoff and T. M. Schmeing, *Biochim. Biophys. Acta. Proteins Proteom*, 2017, **1865**, 1587.

26 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. de Los Santos, M. Yeong, P. Cruz-Morales and S. Abubucker, *et al.*, *Nat. Chem. Biol.*, 2020, **16**, 60.

27 B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle and S. Beyer, *et al.*, *J. Biol. Chem.*, 1999, **274**, 37391; S. Weinig, H.-J. Hecht, T. Mahmud and R. Müller, *Chem. Biol.*, 2003, **10**, 939.

28 J. J. Reddick, S. A. Antolak and G. M. Raner, *Biochem. Biophys. Res. Commun.*, 2007, **358**, 363; S. S. Boddupalli, R. W. Estabrook and J. A. Peterson, *J. Biol. Chem.*, 1990, **265**, 4233; N. Liu, E. D. Abramyan, W. Cheng, B. Perlatti, C. J. B. Harvey, G. F. Bills and Y. Tang, *J. Am. Chem. Soc.*, 2021, **143**, 6043.

29 C. Wu, Z. Shang, C. Lemetre, M. A. Ternei and S. F. Brady, *J. Am. Chem. Soc.*, 2019, **141**, 3910; K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein and B. S. Moore, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 1957.

30 S. Kodali, A. Galgoci, K. Young, R. Painter, L. L. Silver, K. B. Herath, S. B. Singh, D. Cully, J. F. Barrett, D. Schmatz and J. Wang, *J. Biol. Chem.*, 2005, **280**, 1669.

31 O. Erol, T. F. Schäberle, A. Schmitz, S. Rachid, C. Gurgui, M. El Omari, F. Lohr, S. Kehraus, J. Piel, R. Müller and G. M. König, *ChemBioChem*, 2010, **11**, 1253.

32 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790.

33 T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B. L. Pinto and G. A. Salazar, *et al.*, *Nucleic Acids Res.*, 2023, **51**, D418.

34 R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792.

35 A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, 1189.

36 G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner, *Genome Res.*, 2004, **14**, 1188.

37 R. He, J. Zhang, Y. Shao, S. Gu, C. Song, L. Qian, W.-B. Yin and Z. Li, *PLoS Comput. Biol.*, 2023, **19**, e1011100.

38 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.

39 D. Edler, J. Klein, A. Antonelli and D. Silvestro, *Methods Ecol. Evol.*, 2021, **12**, 373.

40 A. Stamatakis, *Bioinformatics*, 2014, **30**, 1312.

41 D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel and T. Flouri, *Mol. Biol. Evol.*, 2020, **37**, 291.

42 S. Q. Le and O. Gascuel, *Mol. Biol. Evol.*, 2008, **25**, 1307.

43 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov and O. Ronneberger, *et al.*, *Nature*, 2021, **596**, 583; E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2023, **32**, e4792.

44 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403.

45 K. J. Harrison, V. Crécy-Lagard and R. Zallot, *Bioinformatics*, 2018, **34**, 1406.

46 H. Harms, A. Klöckner, J. Schrör, M. Josten, S. Kehraus, M. Crüsemann, W. Hanke, T. Schneider, T. F. Schäberle and G. M. König, *Planta Med.*, 2018, **84**, 1363.

47 R. A. Forsyth, R. J. Haselbeck and K. L. Ohlsen, *et al.*, *Mol. Microbiol.*, 2002, **43**, 1387.

48 A. I. Chiriac, F. Kloss, J. Krämer, C. Vuong, C. Hertweck and H. G. Sahl., *J. Antimicrob. Chemother.*, 2015, **70**, 2576.