

Biosynthesis-guided discovery reveals enteropeptins as alternative sactipeptides containing *N*-methylornithine

Received: 25 November 2021

Accepted: 14 September 2022

Published online: 31 October 2022

 Check for updatesKenzie A. Clark¹, Brett C. Covington¹ and Mohammad R. Seyedsayamdost^{1,2}✉

The combination of next-generation DNA sequencing technologies and bioinformatics has revitalized natural product discovery. Using a bioinformatic search strategy, we recently identified ~600 gene clusters in otherwise overlooked streptococci that code for ribosomal peptide natural products synthesized by radical *S*-adenosylmethionine enzymes. These grouped into 16 subfamilies and pointed to an unexplored microbiome biosynthetic landscape. Here we report the structure, biosynthesis and function of one of these natural product groups, which we term enteropeptins, from the gut microbe *Enterococcus cecorum*. We show three reactions in the biosynthesis of enteropeptins that are each catalysed by a different family of metalloenzymes. Among these, we characterize the founding member of a widespread superfamily of Fe–S-containing methyltransferases, which, together with an Mn²⁺-dependent arginase, installs *N*-methylornithine in the peptide sequence. Biological assays with the mature product revealed bacteriostatic activity only against the producing strain, extending an emerging theme of fratricidal or self-inhibitory metabolites in microbiome firmicutes.

Natural products have provided a rich source of complex enzymatic chemistries, therapeutic lead compounds and paradigms in biology. Some of the most well-known natural products, such as vancomycin and penicillin, are synthesized by large assembly-line non-ribosomal peptide synthetase complexes¹. An alternative strategy is provided by ribosomally synthesized and post-translationally modified peptides (RiPPs), in which the ribosome synthesizes a short, genetically encoded peptide that is then transformed into the final mature product by, usually, a small number of tailoring enzymes². This strategy is particularly widespread in organisms with small, host-adapted genomes such as the firmicutes, as RiPP biosynthetic gene clusters (BGCs) have small genomic footprints. Aside from hosting numerous unexplored RiPP BGCs, the firmicutes are also important due to their abundance in diverse animal microbiomes where the molecules they secrete have the potential to influence host health^{3–5}.

Because of their small size and lack of unifying genetic features, RiPP BGCs initially proved difficult to identify with standard approaches, and methods have recently emerged for this purpose^{2,6–9}. We have contributed an approach that utilizes a biosynthesis–regulation co-occurrence strategy⁵. Using the streptide biosynthetic operon as a reference^{10,11}, we searched for RiPP gene clusters that are built by one or more radical *S*-adenosylmethionine (RaS) enzymes and regulated by quorum sensing (QS), reasoning that the resulting natural products would contain unprecedented chemical motifs, due to the remarkable versatility of the RaS enzyme superfamily^{12,13}, and be physiologically relevant, as QS has been shown to regulate important microbial behaviours, such as virulence and biofilm formation^{14,15}. This search revealed ~600 RiPP BGCs, which grouped into 16 subfamilies based on precursor peptide sequence similarity, in mammalian microbiome streptococci alone (Fig. 1a)⁵. Our subsequent mining of this network has revealed new RaS enzyme chemistry and

¹Department of Chemistry, Princeton University, Princeton, NJ, USA. ²Department of Molecular Biology, Princeton University, Princeton, NJ, USA.

✉ e-mail: mrseyed@princeton.edu

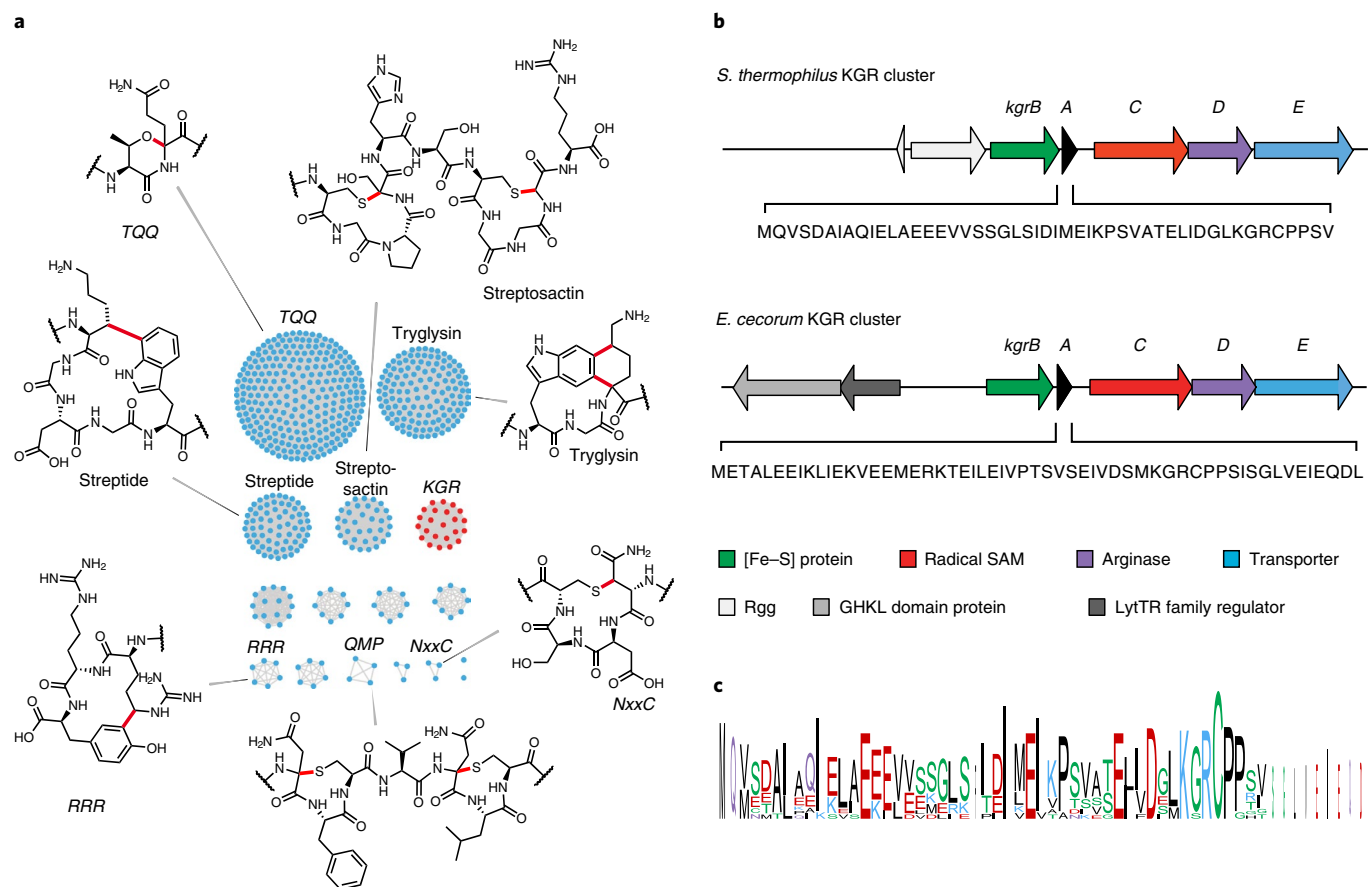


Fig. 1 | Streptococcal RaS-RiPPs and the *kgr* gene cluster. a, Sequence similarity network of RaS-RiPP BGCs regulated by QS. Each node represents a unique BGC, and the edges indicate sequence similarity of the precursor peptide. Subfamilies are named based on conserved motifs in the precursor. Known products of RaS enzymes in each subfamily are shown with the newly installed

bond in red. For the streptide, streptosactin and tryglysin subfamilies, the mature product is known. **b**, The *kgr* BGC from *S. thermophilus* and *E. cecorum*. Sequences of precursor peptides are shown. **c**, Sequence logo plot for all known KgrA peptides. The KGRCPP motif is conserved.

natural products, including tetrahydro[5,6]benzindole formation, in the tryglysin^{5,16}, atypical sactonine topology in streptosactin and the *QMP* operon^{17,18}, α -ether and β -thioether linkages by the *TQQ* and *NxxC* operons, respectively^{19,20}, and an arginine–tyrosine crosslink by a RaS enzyme in the *RRR* cluster²¹. Interestingly, mature products from these clusters have revealed compelling bioactivities, with tryglysin proving to be nanomolar inhibitors of the pathogen *Streptococcus pneumoniae*¹⁶, and streptosactin exhibiting fratricidal activity with nanomolar potency¹⁷.

The gene clusters examined thus far contained simple architectures, consisting of a single RaS enzyme, with or without a discrete RiPP recognition element²², and a precursor peptide. In this article we turn our attention to the most complex subfamily in our network, which contains an RaS enzyme, a hypothetical Fe–S cluster protein, and a predicted manganese-dependent arginase homologue. Using in vitro biochemistry, we elucidate the reaction carried out by each of these metalloenzymes and find that the RaS enzyme installs an α -thioether bond joining neighbouring Cys and Arg residues to form an unusual six-membered thiomorpholine heterocycle, an alternative to the macrocyclic topology of previously identified sactipeptides. The modified arginine residue is then deguanidinated to ornithine by the arginase and *N*-methylated by the Fe–S protein with the use of *S*-adenosylmethionine (SAM), resulting in *N*-methylornithine in the RiPP. The Fe–S-dependent methyltransferase is the first characterized member of this prevalent superfamily of enzymes. Knowledge regarding the biosynthetic reactions allowed us to identify the

mature natural products, which we name enteropeptin A–C, from the gut bacterium *Enterococcus cecorum*. Biological activity assays with enteropeptin A show species-specific bacteriostatic activity only against the producing strain, extending an emerging theme of self-inhibitory RiPPs produced by firmicutes, notably those in animal microbiomes.

Results and discussion

The *KGR* subfamily

The *KGR* subfamily from our original network, named for the conserved C-terminal KGR motif, contains 23 unique clusters from *Streptococcus thermophilus*. It encodes a 49mer precursor peptide (KgrA), a putative Fe–S cluster protein (KgrB), which does not belong to the RaS superfamily and exhibits little discernible homology to characterized enzyme families, an RaS enzyme (KgrC), a protein with sequence similarity to arginase (KgrD), which uses an Mn^{2+} cofactor to convert free arginine to ornithine in the urea cycle^{23,24}, and finally a combination protease/exporter (KgrE) (Fig. 1b). We sought to elucidate the reactions of these enzymes by an in vitro approach in which the enzymes and precursor peptide would be generated recombinantly and reaction products elucidated with high-resolution mass spectrometry (HR-MS), tandem HR-MS (HR-MS/MS) and multidimensional NMR spectroscopy.

After repeated attempts to obtain soluble RaS enzyme (KgrC) from *S. thermophilus* failed, we performed additional bioinformatic searches and identified a highly homologous *kgr* cluster from several

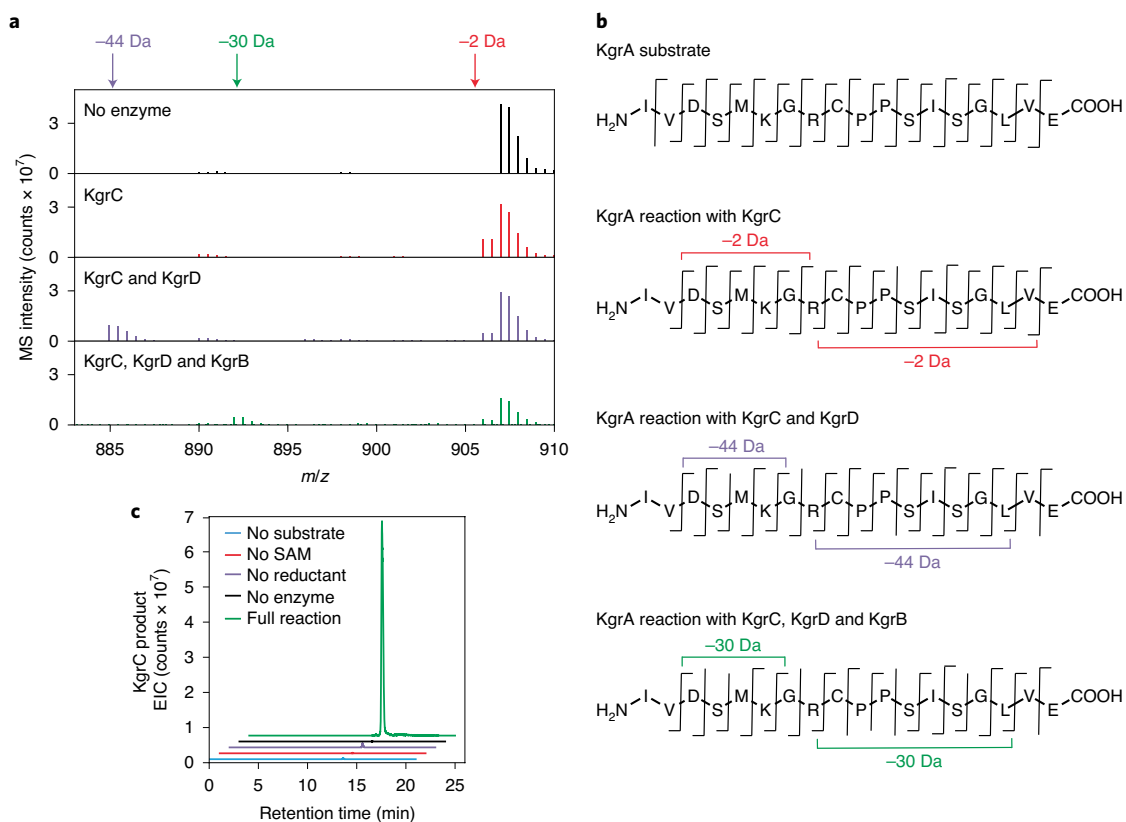


Fig. 2 | Reactions of KgrBCD characterized by HR-MS and HR-MS/MS. a, MS profiles of the reactions of KgrA with KgrC, KgrD and KgrBCD. Products for each reaction are marked and colour-coded. **b**, HR-MS/MS profiles of the GluC-cleaved KgrA (amino acids 34–56) following reaction with KgrC, KgrD and

KgrBCD. Observed fragments and the mass difference relative to the unreacted KgrA peptide are marked. **c**, HPLC–MS analysis of the reaction of KgrC with KgrA. The extracted ion chromatogram (EIC) of the KgrC product is shown. Product formation is only observed in the full reaction.

E. cecorum strains (Supplementary Table 1 and Supplementary Fig. 1). *E. cecorum* is a commensal strain isolated from chicken digestive tracts and has increasingly acted as an opportunistic human pathogen, causing outbreaks of enterococcal spondylitis in commercial poultry production²⁵. We focused on the *kgr* cluster from *E. cecorum* ATCC 43198, which encodes the same set of modification enzymes as the orthologous cluster in *S. thermophilus* with a 56mer precursor peptide (Fig. 1b). The cluster is regulated by an upstream divergently transcribed LytTR-type transcriptional regulator²⁶. A logo plot of the precursor peptide, generated from both the *Enterococcus* and *Streptococcus* clusters, shows that the KGRCPP motif is conserved (Fig. 1c).

Characterization of the KGR biosynthetic enzymes

To determine the nature of the modifications installed by KgrBCD onto KgrA, we produced each protein individually as a hexaHis-tagged construct via recombinant expression in *Escherichia coli* (Supplementary Tables 2 and 3). As KgrB and KgrC were predicted to contain Fe–S clusters, they were purified anaerobically. The ultraviolet–visible absorption spectra of these proteins were similar and contained a broad peak at 410 nm consistent with the presence of at least one $[4\text{Fe}-4\text{S}]^{2+}$ rather than a $[2\text{Fe}-2\text{S}]^{2+}$ (Supplementary Fig. 2)²⁷. Quantification of iron and labile sulfide revealed 2.9 ± 0.1 Fe and 3.5 ± 0.1 S^{2-} per KgrB protomer, while KgrC contained 3.9 ± 0.5 Fe and 4.8 ± 0.3 S^{2-} (Supplementary Table 4). A sequence alignment of KgrC and SuiB, a RaS enzyme for which the crystal structure has been solved^{28,29}, revealed a SPASM domain in KgrC that may bind two $[4\text{Fe}-4\text{S}]$ clusters in addition to the active site cluster (Supplementary Fig. 3). Our preparation of KgrC was likely not fully reconstituted, although chemical reconstitution was not necessary to obtain active protein. The number and nature of Fe–S clusters

in KgrB remain to be determined. The third biosynthetic enzyme, KgrD, a predicted arginase homologue, was purified aerobically with supplemental Mn^{2+} .

With all components available, *in vitro* reactions were carried out by incubating KgrA with each enzyme alone, with each combination of two enzymes or with all three enzymes. All reactions contained SAM, dithiothreitol (DTT) and reductant (titanium citrate). Following overnight incubation under anaerobic conditions, reactions were digested with trypsin or GluC to give peptide fragments small enough to allow for analysis by high performance liquid chromatography–quadrupole time-of-flight mass spectrometry (HPLC–QTOF-MS) (Supplementary Tables 5 and 6). No new products were observed following incubation of unmodified KgrA with KgrB, KgrD, or KgrC and KgrD (Supplementary Fig. 4). However, reaction with KgrC alone gave a product 2 Da lighter than substrate (Fig. 2a); notably, no proteolysis was observed at the Arg residue in the KGR motif after the modification. While KgrC and KgrB only revealed the -2 Da product (Supplementary Fig. 4), the reaction with KgrC and KgrD yielded product that was 44 Da lighter than the substrate in addition to a small amount of remaining -2 Da product (Fig. 2a). Finally, the reaction with all three biosynthetic enzymes resulted in formation of a product peak that was 30 Da lighter than the original KgrA substrate, or 14 Da heavier than the product of the KgrCD reaction (Fig. 2a). This species was only observed in the presence of all three biosynthetic enzymes and was assigned as the final reaction product. While we were unable to obtain soluble KgrC from the *S. thermophilus* system, the *S. thermophilus* KgrD and KgrB were found to carry out the same reactions as their homologues when incubated with KgrA and KgrC from *E. cecorum* (Supplementary Fig. 4).

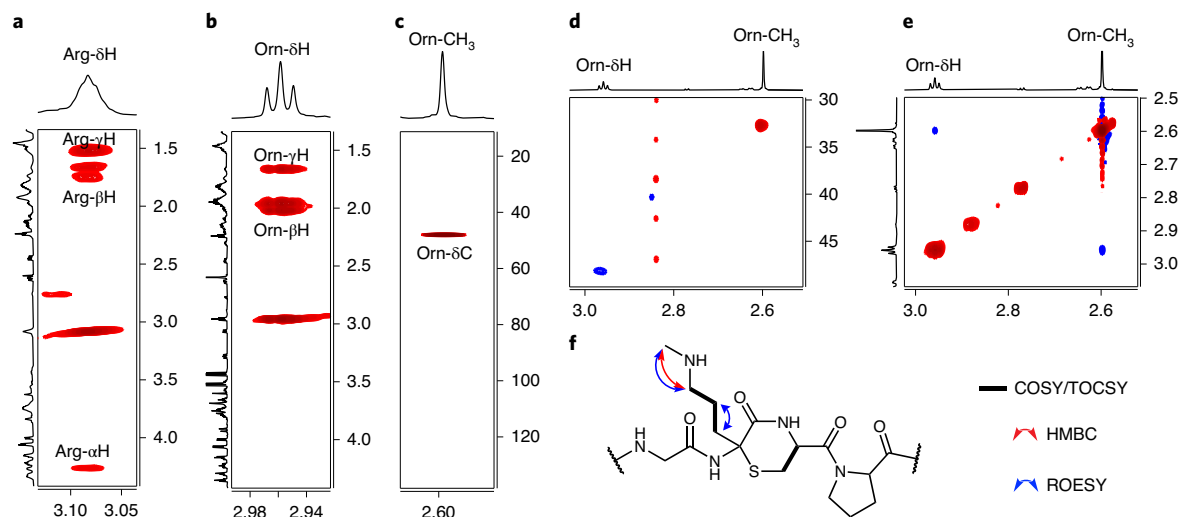


Fig. 3 | NMR spectral analysis of the product of the reaction of KgrA with KgrBCD. **a, b**, TOCSY slice of the Arg41-δH (precursor peptide numbering) in unreacted KgrA (**a**) compared to the corresponding Orn-δH in product (**b**). An α-H is not observed in the product. **c**, HMBC slice of the *N*-methyl group of Orn shows a cross-peak to the Orn-δC. **d**, Heteronuclear single-quantum coherence

spectrum highlighting the Orn *N*-methyl-H/C and Orn-δH/δC correlations.

e, ROESY showing cross-peaks between the Orn-δH and *N*-methyl group.

f, Structure of the product of the KgrABCD reaction focusing on the modified region. Relevant NMR correlations are shown. The absolute configuration at the Orn-αC remains to be determined.

KgrC, a thiomorpholine-forming RaS enzyme

We first used HR-MS/MS to analyse the GluC-treated products after reaction of KgrA with KgrC, KgrCD or KgrBCD (Supplementary Tables 7–10). With KgrA alone, nearly all expected *b* and *y* ions of the unmodified peptide were observed. Upon reaction with KgrC, all *b* ions upstream of Arg41 were unaltered, whereas those downstream were –2 Da relative to untreated KgrA. Likewise, *y* ions downstream of Arg41 were unaltered, whereas those upstream were –2 Da (Fig. 2b). An α-thioether bond with the adjacent Cys residue would be consistent with the data observed as these linkages undergo fragmentation under collision-induced dissociation^{30,31}. In the absence of KgrA, KgrC produced 5'-deoxyadenosine (5'-dA), a futile cycling reaction of cofactor SAM that is diagnostic for RaS enzymes. Moreover, in the presence of KgrA, the release of 5'-dA and product occurred with similar kinetics (Supplementary Fig. 5); the reaction was strictly dependent on SAM and reductant (Fig. 2c). We conclude that KgrC is a RaS enzyme that installs the smallest sactonine linkage to date, leading to a thiomorpholine modification in the KgrA backbone. The absolute configuration of the newly formed quaternary centre at the Arg α-carbon remains to be determined.

KgrD, an RiPP arginase

Next, we analysed the product resulting from the reaction of KgrA with KgrCD. All *b* ions upstream of Arg41 were unaltered, whereas those downstream were now 44 Da lighter than unreacted KgrA (Fig. 2b). The *y* fragments mirrored this pattern and isolated residue 41 as the target of the KgrD-catalysed modification. The HR-MS data were entirely consistent with loss of the guanidinium group, giving rise to an ornithine side chain at this position. We examined the modification further by purifying KgrD in the absence or presence of Mn²⁺ and incubating it with the product of the KgrAC reaction; the Mn²⁺-reconstituted enzyme showed the highest activity (Supplementary Fig. 6). Together, these results suggest that KgrD is an Mn²⁺-dependent enzyme that hydrolyses the Arg41 guanidinium group, only after thiomorpholine formation, giving rise to an ornithine side chain.

KgrB, an Fe–S-dependent methyltransferase

Finally, we analysed the product of the KgrBCD reaction by HR-MS/MS and noted that the *b* and *y* ion fragments point to a 14 Da increase at

the newly generated Orn residue, suggesting KgrB may install a methyl group (Fig. 2b). To determine the exact nature of the modification and provide additional support for the products of KgrC and KgrD, we carried out the KgrABCD reaction on a large scale and conducted detailed one-dimensional/two-dimensional (1D/2D) NMR analysis after trypsinolysis of the product to a 17mer peptide. An unmodified 17mer was synthesized via solid-phase peptide synthesis as a reference and analysed by 1D/2D NMR in a similar manner (Supplementary Table 11 and Supplementary Figs. 7 and 8). ¹H and total correlation spectroscopy (TOCSY) NMR analysis revealed correlations from the Arg-δH to the α-, β- and γ-Hs in the substrate (Fig. 3a). In the product, however, the correlation from the Orn-δH to an α-H was missing, and the β-, γ- and δ-carbons were clearly methylenes, as indicated by heteronuclear single-quantum coherence, all consistent with a modification at the Orn-αC (Fig. 3b). In addition, the peaks assigned to the adjacent Cys-αH and βHs shifted from 4.88 and 2.75/3.10 ppm in the substrate to 4.76 and 3.07/3.28 ppm, respectively, in the product. These shifts are expected for an α-thioether, thereby establishing the proposed thiomorpholine installation. The ¹H NMR spectrum also revealed a new methyl singlet peak at 2.59 ppm with an associated 32.7 ppm ¹³C shift (Supplementary Fig. 8). Heteronuclear multiple-bond correlations (HMBC) from the methyl protons to the Orn-δC (Fig. 3c,d) and a rotating-frame nuclear Overhauser effect spectroscopy (ROESY) correlation (Fig. 3e) to the Orn-δH established this feature as the *N*-methyl group of Orn. These data, together with the HR-MS and HR-MS/MS results above, point to three unusual transformations, consisting of thiomorpholine formation by the RaS enzyme, followed by Mn²⁺-dependent deguanidination by KgrD, and finally Orn *N*-methylation by the Fe–S-containing KgrB (Fig. 3f).

A superfamily of Fe–S methyltransferases

We were intrigued by the Fe–S-dependent *N*-methylation and investigated this reaction further in vitro and bioinformatically. The latter showed that KgrB does not significantly match any Pfam in the available protein database. A PSI-BLAST search with the KgrB sequence allowed us to generate a sequence similarity network with the 5,000 top hits (Supplementary Fig. 9). Nearly all identified proteins were annotated as a domain of unknown function, uncharacterized protein or hypothetical protein. Interestingly, 40 of the 5,000 proteins were

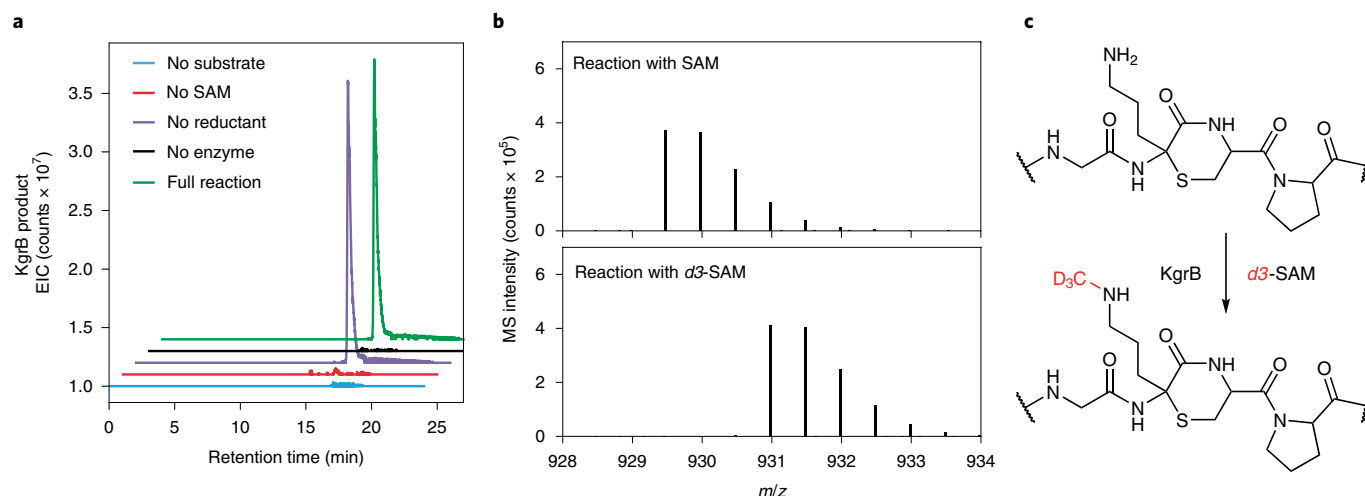


Fig. 4 | Orn *N*-methylation catalysed by KgrB. **a**, HPLC–MS analysis of the reaction of KgrB with the product of KgrACD. Reductant is not required but SAM is. **b**, Reaction with d3-SAM gives a corresponding +3 Da increase relative to the product with protonated SAM. **c**, Reaction carried out by KgrB with d3-SAM.

annotated as Flagellin *N*-methylase^{32,33}, which would indicate a similar activity to KgrB. The Flagellin *N*-methylase protein, FliB, has been studied through genetic approaches, although not yet biochemically, and was proposed to methylate flagellar Lys residues in *Salmonella typhimurium*^{32,33}. While FliB and KgrB share only 11% sequence similarity, they appear to catalyse similar reactions; a sequence alignment of the two proteins shows a conserved CxxxCC motif along with several other conserved Cys residues (Supplementary Fig. 10), while a Log-oPlot of the top 5,000 PSI-BLAST hits shows that this CxxxCC motif is conserved among the entire family (Supplementary Fig. 11). Furthermore, an AlphaFold model of the structure of KgrB predicts that these conserved Cys residues are positioned in an orientation that can bind an Fe–S cluster (Supplementary Fig. 11)³⁴. Unfortunately, attempts to assess these conserved Cys residues in a quadruple site-directed mutant (C22A/C25A/C29A/C30A-KgrB) and a double mutant (C25A/C30A-KgrB) resulted in insoluble protein. The Fe–S cluster appears to be important for solubility. Together, these analyses suggest that enzymes in the Fe–S methyltransferase family are widespread without a single biochemically characterized member.

To characterize KgrB further, we first prepared sufficient amounts of the KgrB substrate using a heterologous coexpression system in which a maltose-binding protein-tagged KgrA was coexpressed with KgrC and KgrD on a pRSFDuet vector in *E. coli*. Upon purification, we confirmed the presence of the Orn-thiomorpholine modification (Supplementary Table 12). After reaction with KgrB, we observed time- and enzyme-dependent formation of the *N*-methyl product (Supplementary Fig. 12), which required SAM, substrate and KgrB, but not reductant (Fig. 4a), suggesting that SAM served as the methyl donor. To verify, we repeated the reaction with S-(methyl-d3)-S-adenosylmethionine (d3-SAM) and observed a product that was now +3 Da relative to the control reaction with protonated SAM (Fig. 4b,c). HR-MS/MS analysis confirmed that the deuterated methyl group was located on the Orn side chain (Supplementary Table 13). The other product, S-adenosylhomocysteine (SAH), was difficult to identify directly because of significant levels of SAH in the commercial SAM preparations and continuous degradation of SAM to SAH, a previously reported side reaction^{35,36}, even after purification of SAM. We also monitored 5'-dA formation and found that KgrB could generate small amounts in the presence of reductant. However, 5'-dA formation was minimal in the absence of reductant and did not correlate with product formation (Supplementary Fig. 13). Together, our results provide the first

biochemical characterization of this superfamily of Fe–S-dependent methyltransferases.

Discovery of enteropeptins from *E. cecorum*

We next used insights into the reactions of the biosynthetic enzymes to search for the mature product of the *kgr* cluster in *E. cecorum* ATCC 43198. Based on general RiPP biosynthetic logic, we compiled a list of m/z values of potential mature peptides containing the modifications elucidated above and trimmed either *N*-terminal to Arg41 and/or *C*-terminal to Pro43. *E. cecorum* was then cultured to stationary phase, and the supernatant cleared by solid-phase extraction and subsequently analysed by HPLC–Qtof-MS. Any observed ion that matched an accurate m/z in the compiled list was further characterized by HR-MS/MS. Three candidates emerged, the most abundant comprising an 8mer peptide containing the *N*-methylOrn-thiomorpholine modification. MS/MS analysis showed b and y ions consistent with the sequence of the peptide and the fragmentation pattern observed in the enzymatic reactions (Supplementary Table 14). In addition, two minor 7mer and 9mer variants, ~10-fold less abundant than the 8mer product, were also observed and confirmed by HR-MS/MS to carry the *N*-methylOrn-thiomorpholine alteration (Supplementary Tables 15 and 16 and Supplementary Fig. 14). These products could result from the predicted transporter, KgrE, cleaving *C*-terminally to both Ser37 and Ser45 to give the 8mer product but occasionally cleaving *N*-terminally, yielding the 7mer or 9mer products. Alternatively, other proteases may be involved.

To provide further support that these peptides are the mature *kgr* products, we prepared a synthetic standard via an *in vitro* reaction with the purified biosynthetic enzymes and a modified KgrA substrate. Site-directed mutagenesis was used to insert a GluC-site and an AspN-site upstream of Met38 and downstream of Ser45, respectively. The disubstituted KgrA was purified and used as a substrate for an *in vitro* reaction with KgrC, KgrD and KgrB. The resulting product was treated with GluC/AspN to deliver the modified 8mer peptide. This standard gave the expected HR-MS and HR-MS/MS fragments (Supplementary Table 17), indicating it contained the modifications outlined above. It was then compared to the authentic 8mer isolated from *E. cecorum*. The two peptides showed identical chromatographic properties and coeluted when injected onto the HPLC–MS in a 1:1 ratio (Fig. 5a). We conclude that the mature product of the *kgr* gene cluster is this 8mer RiPP which we have named enteropeptin A (1) (Fig. 5b). It is synthesized by a sequence of four modifications, thiomorpholine

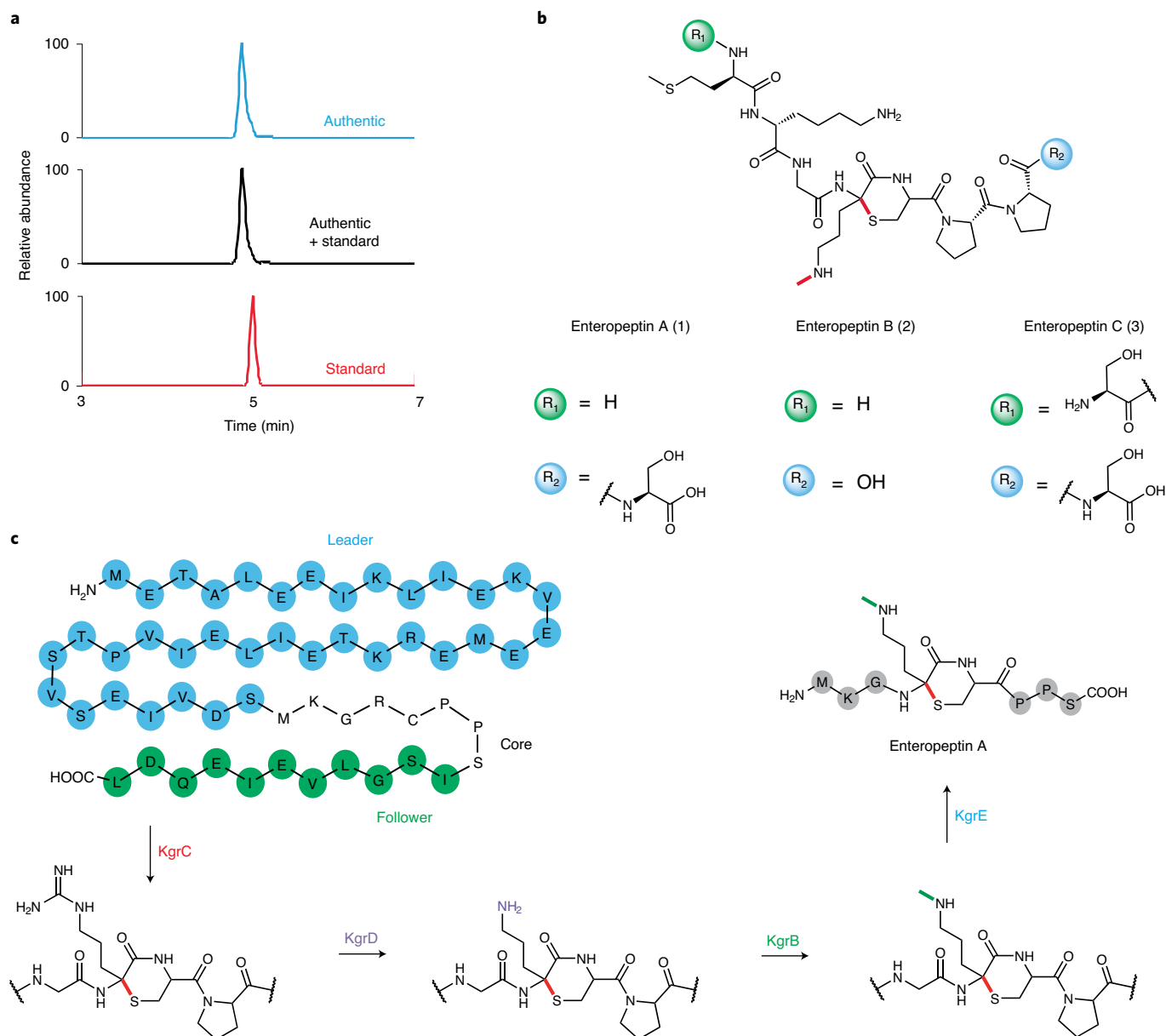


Fig. 5 | Discovery of enteropeptins, the mature products of the *kgr* cluster, from *E. cecorum*. **a**, HPLC–Qtof–MS analysis of authentic and synthetic enteropeptin A. Shown are extracted ion chromatograms for authentic enteropeptin from *E. cecorum* (blue, top), heterologously produced enteropeptin (red, bottom), and a 1:1 coinjection of these two samples (black, middle), which

coelute. **b**, Structures of enteropeptin A, the major product of the *kgr* cluster, and of minor products enteropeptin B and C. **c**, Biosynthetic pathway for enteropeptins. Leader and follower sequences are indicated with blue and green circles, respectively. Unmodified amino acids in the final product are shown in grey circles.

formation by KgrC, deguanidination by KgrD, methylation of the resulting Orn group by KgrB and finally removal of the leader and follower sequences by KgrE and perhaps another protease (Fig. 5c). The 7mer and 9mer products, termed enteropeptin B (2) and C (3), form two less-abundant products of the *kgr* operon.

Narrow-spectrum bacteriostatic activity

As a starting point to examine the activity of enteropeptin A, we first determined its timing of production. The RaS–RiPPs in our network are controlled by an *shp–rgg* QS operon and their products accumulate starting in mid-exponential phase^{37,38}. Similarly, the *kgr* cluster in *E. cecorum* is controlled by a LytTR-type regulator²⁶, and enteropeptin A synthesis as a function of growth, as determined by optical density at 600 nm (OD₆₀₀), revealed production in mid-exponential phase that

persisted into stationary phase, reaching a maximal concentration of 0.7 μ M (Fig. 6a).

With pure material at hand and clues regarding production titres, we tested the effect of 0.1–6 μ M enteropeptin A on a number of human commensal and opportunistic pathogens, including *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *E. coli*, *Enterococcus faecalis*, *Streptococcus thermophilus* and *E. cecorum* (Supplementary Table 18). Enteropeptin A only inhibited the growth of *E. cecorum* but did not affect the other strains. A growth curve of *E. cecorum* in the presence of various concentrations of enteropeptin A, or the linear 8mer as a control, showed a concentration-dependent growth delay or cessation only with mature product; the linear 8mer had no effect (Fig. 6b). When used at the detected production titres (~1 μ M), we observed recovery of growth over time. Significant recovery did

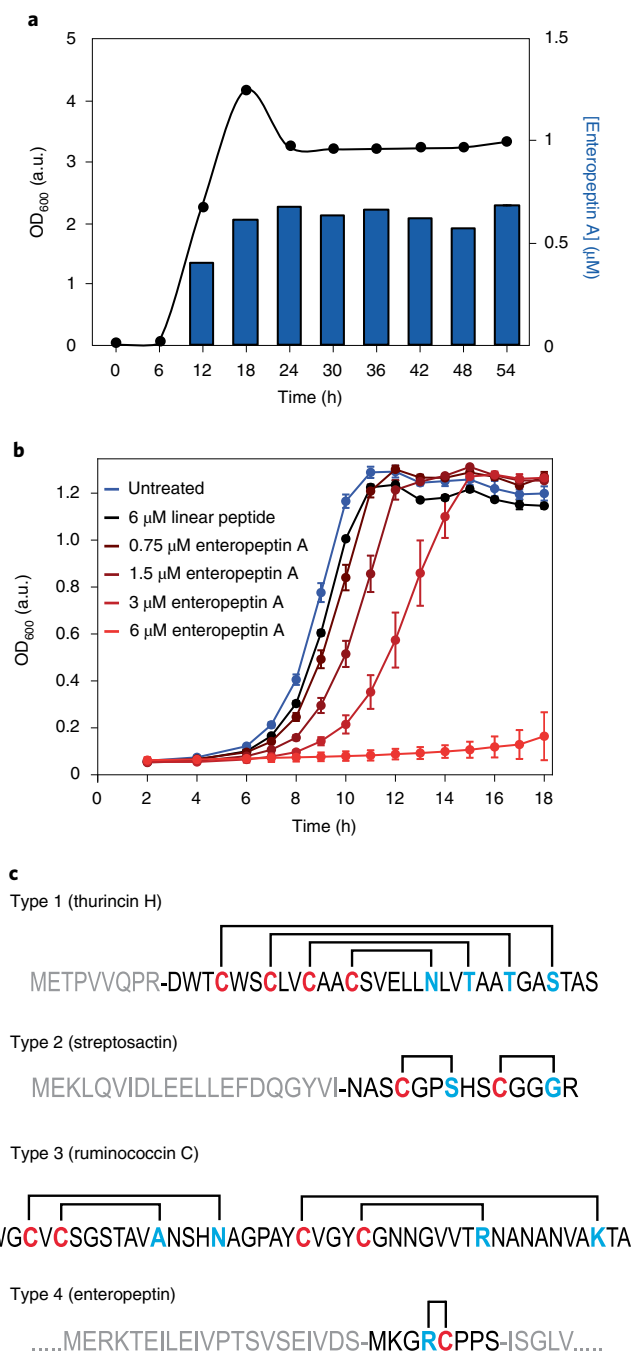


Fig. 6 | Enteropeptin activity and compound class. **a**, Enteropeptin A is synthesized starting in the exponential phase and reaches a maximal concentration of ~0.7 μM in the stationary phase. Data are presented as mean values of two independent measurements. **b**, Enteropeptin bacteriostatically inhibits *E. cecorum* growth. Shown are growth curves of *E. cecorum* in the presence of 0.75–6 μM of enteropeptin A. No effect is seen with 6 μM of unmodified, linear core peptide. Data are presented as mean values of three independent measurements; error bars represent standard deviation. **c**, Comparison of the topology of known Type 1, Type 2 and Type 3 sactipeptides, represented by thurincin H, streptosactin and ruminococcin C, respectively, with enteropeptin as a Type 4 sactipeptide. Leader/follower sequences are shown in grey; for brevity, only partial leader/follower sequences are shown for ruminococcin C and enteropeptin.

not occur at 6 μM after 18 h. Whether enteropeptin A is fratricidal or plays a role in regulating growth by different means remains to be determined.

Conclusion

Enteropeptins add to the growing list of sactipeptides but are distinct from those discovered thus far by containing a small sactonine linkage in the form of a thiomorpholine heterocycle. The majority of mature sactipeptides belong to Type 1 with multiple nested thioether macrocycles and at least two residues between the acceptor and donor residues (Fig. 6c)³⁹. Type 2 sactipeptides, streptosactin and the product from the QMP cluster, carry an unnested bicycle topology¹⁷. The Type 3 sactipeptide, ruminococcin C, contains a bicyclic nested topology^{40,41}. Enteropeptins now add another distinct chemotype, a thiomorpholine ring, for which we propose the Type 4 sactipeptide designation (Fig. 6c).

A further distinguishing feature from previously discovered sactipeptides are the additional modifications in enteropeptins. Thuricin α and β (ref. ⁴²), thurincin H (ref. ⁴³), huazacin/thuricin Z (refs. ^{31,44}) and ruminococcin C (refs. ^{40,41}) contain only thioether linkages while SKF and subtilisin carry head-to-tail amide cyclizations^{45–49}, with the former additionally harbouring a disulfide linkage. The presence of ornithine is well documented in non-ribosomal peptides; *N*-methylornithine, however, has not been reported in RiPP natural products. The incorporation of ornithine was very recently reported in the heterologous landornamide product^{50,51}. The use of an arginase homologue to generate ornithine in RiPPs and to modify it further represents a backdoor route for the incorporation and diversification of unnatural amino acids into ribosomally generated products⁵².

Perhaps the most surprising discovery from this study is KgrB, a member of an enzyme superfamily of Fe–S-dependent methyltransferases that was not previously known. Only one other example of an Fe–S methyltransferase has previously been described; RumA catalyses the methylation of U1939 of 23S ribosomal RNA to yield 5-methyluridine or ribothymidine⁵³. A crystal structure of the enzyme shows that the Fe–S cluster is located in the RNA-binding domain of the protein and is ligated by four cysteines with a CX₂CX₂CX₂C motif⁵⁴. The Fe–S cluster was proposed to facilitate protein folding or RNA binding rather than a catalytic role. Although KgrB and RumA contain Fe–S clusters and catalyse methylation reactions, they are unrelated and belong to distinct structural families and enzyme phylogenies. The two proteins share only 8% sequence identity and RumA does not contain the characteristic CxxxCC motif found in KgrB and other superfamily members (Supplementary Fig. 15). Although it is tempting to speculate catalytic roles for the KgrB Fe–S cluster, such as Lewis acid catalysis to activate the Orn amino group for nucleophilic attack, additional studies are required to elucidate the mechanism of KgrB and provide insights into the chemistry underlying this unusual methylation strategy.

Enteropeptins contribute to an emerging theme of RaS-RiPPs in firmicutes that show species-specific growth-inhibitory activity against the host. Aside from the streptococcal RiPPs streptosactin and tryglysin, the *Bacillus subtilis* products SKF, subtilisin and YdF also show varying levels of inhibitory activity against the host^{55,56}. Whether these molecules act as true fratricidal agents that cause cell lysis or merely inhibit growth through other mechanisms remains to be delineated^{57,58}. As the first natural product identified from *E. cecorum*, enteropeptins point the way towards additional microbiome natural products. Host-associated firmicutes have not been thought of as prolific producers of secondary metabolites but the discovery of these RiPPs show that these bacteria, and others associated with mammalian microbiomes, harbour a vast and poorly explored biosynthetic capacity that is now ripe to be mined.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41557-022-01063-3>.

References

1. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
2. Montalbán-López, M. et al. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **38**, 130–239 (2021).
3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
4. Donia, M. S. et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).
5. Bushin, L. B., Clark, K. A., Pelczar, I. & Seyedsayamdost, M. R. Charting an unexplored streptococcal biosynthetic landscape reveals a unique peptide cyclization motif. *J. Am. Chem. Soc.* **140**, 17674–17684 (2018).
6. Russell, A. H. & Truman, A. W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* **18**, 1838–1851 (2020).
7. Hetrick, K. J. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr. Opin. Chem. Biol.* **38**, 36–44 (2017).
8. Kloosterman, A. M., Medema, M. H. & van Wezel, G. P. Omics-based strategies to discover novel classes of RiPP natural products. *Curr. Opin. Biotechnol.* **69**, 60–67 (2021).
9. Tietz, J. I. & Mitchell, D. A. Using genomics for natural product structure elucidation. *Curr. Top. Med. Chem.* **16**, 1645–1694 (2016).
10. Schramma, K. R., Bushin, L. B. & Seyedsayamdost, M. R. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nature Chem.* **7**, 431–437 (2015).
11. Ibrahim, M. et al. Control of the transcription of a short gene encoding a cyclic peptide in *Streptococcus thermophilus*: a new quorum-sensing system? *J. Bacteriol.* **189**, 8844–8854 (2007).
12. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
13. Landgraf, B. J., McCarthy, E. L. & Booker, S. J. Radical S-adenosylmethionine enzymes in human health and disease. *Annu. Rev. Biochem.* **85**, 485–514 (2016).
14. Bassler, B. L. & Losick, R. Bacterially speaking. *Cell* **125**, 237–246 (2006).
15. Whiteley, M., Diggle, P. S. & Greenberg, E. P. Progress in and promise of bacterial quorum sensing research. *Nature* **551**, 313–320 (2017).
16. Rued, B. E. et al. Quorum sensing in *Streptococcus mutans* regulates production of tryglysin, a novel RaS-RiPP antimicrobial compound. *mBio* **12**, e02688–20 (2021).
17. Bushin, L. B., Covington, B. C., Rued, B. E., Federle, M. J. & Seyedsayamdost, M. R. Discovery and biosynthesis of streptosactin, a sactipeptide with an alternative topology encoded by commensal bacteria in the human microbiome. *J. Am. Chem. Soc.* **142**, 16265–16275 (2020).
18. Caruso, A. & Seyedsayamdost, M. R. Radical SAM enzyme QmpB installs two 9-membered ring sactonine macrocycles during biogenesis of a ribosomal peptide natural product. *J. Org. Chem.* **86**, 11284–11289 (2021).
19. Clark, K. A., Bushin, L. B. & Seyedsayamdost, M. R. Aliphatic ether bond formation expands the scope of radical SAM enzymes in natural product biosynthesis. *J. Am. Chem. Soc.* **141**, 10610–10615 (2019).
20. Caruso, A., Bushin, L. B., Clark, K. A., Martinie, R. J. & Seyedsayamdost, M. R. Radical approach to enzymatic β -thioether bond formation. *J. Am. Chem. Soc.* **141**, 990–997 (2019).
21. Caruso, A., Martinie, R. J., Bushin, L. B. & Seyedsayamdost, M. R. Macrocyclization via an arginine-tyrosine crosslink broadens the reaction scope of radical S-adenosylmethionine enzymes. *J. Am. Chem. Soc.* **141**, 16610–16614 (2019).
22. Burkhardt, B. J., Hudson, G. A., Dunbar, K. L. & Mitchell, D. A. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* **11**, 564–570 (2015).
23. Kanyo, Z. F., Scolnick, L. R., Ash, D. E. & Christianson, D. W. Structure of a unique binuclear manganese cluster in arginase. *Nature* **383**, 554–557 (1996).
24. Ash, D. E., Cox, J. D. & Christianson, D. W. Arginase: a binuclear manganese metalloenzyme. *Met. Ions Biol. Syst.* **37**, 407–428 (2000).
25. Jung, A., Chen, L. R., Suyemoto, M. M., Barnes, H. J. & Borst, L. B. A review of *Enterococcus cecorum* infection in poultry. *Avian Dis.* **62**, 261–271 (2018).
26. Do, T., Page, J. E. & Walker, S. Uncovering the activities, biological roles, and regulation of bacterial cell wall hydrolases and tailoring enzymes. *J. Biol. Chem.* **295**, 3347–3361 (2020).
27. Agar, J. N. et al. IscU as a scaffold for iron–sulfur cluster biosynthesis: sequential assembly of [2Fe–2S] and [4Fe–4S] clusters in IscU. *Biochemistry* **39**, 7856–7862 (2000).
28. Davis, K. M. et al. Structures of the peptide-modifying radical SAM enzyme SuiB elucidate the basis of substrate recognition. *Proc. Natl Acad. Sci. USA* **114**, 10420–10425 (2017).
29. Schramma, K. R. & Seyedsayamdost, M. R. Lysine–tryptophan-crosslinked peptides produced by radical SAM enzymes in pathogenic streptococci. *ACS Chem. Biol.* **12**, 922–927 (2017).
30. Himes, P. M., Allen, S. E., Hwang, S. & Bowers, A. A. Production of sactipeptides in *Escherichia coli*: probing the substrate promiscuity of subtilisin A biosynthesis. *ACS Chem. Biol.* **11**, 1737–1744 (2016).
31. Hudson, G. A. et al. Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new C α , C β , and C γ -linked thioether-containing peptides. *J. Am. Chem. Soc.* **141**, 8228–8238 (2019).
32. Tronick, S. R. & Martinez, R. J. Methylation of the flagellin of *Salmonella typhimurium*. *J. Bacteriol.* **105**, 211–219 (1971).
33. Stocker, B. A. D., McDonough, M. W. & Ambler, R. P. A gene determining presence or absence of e-N-methyllysine in *Salmonella* flagellar protein. *Nature* **189**, 556–558 (1961).
34. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
35. Lankau, T., Kuo, T. N. & Yu, C. H. Computational study of the degradation of S-adenosyl methionine in water. *J. Phys. Chem. A* **121**, 505–514 (2017).
36. Salyan, M. E. K. et al. A general liquid chromatography/mass spectroscopy-based assay for detection and quantitation of methyltransferase activity. *Anal. Biochem.* **349**, 112–117 (2006).
37. Fleuchot, B. et al. Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol. Microbiol.* **80**, 1102–1119 (2011).
38. Chang, J. C., LaSarre, B., Jimenez, J. C., Aggarwal, C. & Federle, M. J. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. *PLoS Pathog.* **7**, e1002190 (2011).
39. Flühe, L. & Marahiel, M. A. Radical S-adenosylmethionine enzyme catalyzed thioether bond formation in sactipeptide biosynthesis. *Curr. Opin. Chem. Biol.* **17**, 605–612 (2013).

40. Chiumento, S. et al. Ruminococcin C, a promising antibiotic produced by a human gut symbiont. *Sci. Adv.* **5**, eaaw9969 (2019).
41. Balty, C. et al. Ruminococcin C, an anti-clostridial sactipeptide produced by a prominent member of the human microbiota *Ruminococcus gnavus*. *J. Biol. Chem.* **294**, 14512–14525 (2019).
42. Rea, M. C. et al. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc. Natl Acad. Sci. USA* **107**, 9352–9357 (2010).
43. Lee, H., Churey, J. J. & Worobo, R. W. Biosynthesis and transcriptional analysis of thurincin H, a tandem repeated bacteriocin genetic locus, produced by *Bacillus thuringiensis* SF361. *FEMS Microbiol. Lett.* **299**, 205–213 (2009).
44. Mo, T. et al. Thuricin Z: a narrow-spectrum sactibiotic that targets the cell membrane. *Angew. Chem. Int. Ed.* **58**, 18793–18797 (2019).
45. Kawulka, K. E. et al. Structure of subtilisin A, an antimicrobial peptide from *Bacillus subtilis* with unusual posttranslational modifications linking cysteine sulfurs to alpha-carbons of phenylalanine and threonine. *J. Am. Chem. Soc.* **125**, 4726–4727 (2003).
46. Kawulka, K. E. et al. Structure of subtilisin A, a cyclic antimicrobial peptide from *Bacillus subtilis* with unusual sulfur to α -carbon cross-links: formation and reduction of α -thio- α -amino acid derivatives. *Biochemistry* **43**, 3385–3395 (2004).
47. Flöhe, L. et al. Two [4Fe–4S] clusters containing radical SAM Enzyme SkfB catalyze thioether bond formation during the maturation of the sporulation killing factor. *J. Am. Chem. Soc.* **135**, 959–962 (2013).
48. Lui, W.-T. et al. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **107**, 16286–16290 (2010).
49. Flöhe, L. et al. The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilisin A. *Nat. Chem. Biol.* **8**, 350–357 (2012).
50. Bösch, N. M. et al. Landornamides: antiviral ornithine-containing ribosomal peptides discovered through genome mining. *Angew. Chem. Int. Ed.* **59**, 11763–11768 (2020).
51. Mordhorst, S., Morinaka, B. I., Vagstad, A. L. & Piel, J. Posttranslationally acting arginases provide a ribosomal route to non-proteinogenic ornithine residues in diverse peptide sequences. *Angew. Chem. Int. Ed.* **59**, 21442–21447 (2020).
52. Walsh, C. T. Blurring the lines between ribosomal and nonribosomal peptide scaffolds. *ACS Chem. Biol.* **9**, 1653–1661 (2014).
53. Agarwalla, S., Kealey, J. T., Santi, D. V. & Stroud, R. M. Characterization of the 23S ribosomal RNA m⁵U1939 methyltransferase from *Escherichia coli*. *J. Biol. Chem.* **277**, 8835–8840 (2002).
54. Lee, T. T., Agarwalla, S. & Stroud, R. M. Crystal structure of RumA, an iron–sulfur cluster containing *E. coli* ribosomal RNA 5-methyluridine methyltransferase. *Structure* **12**, 397–407 (2004).
55. Benjdia, A. et al. Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in *Bacillus subtilis*. *Nature Chem* **9**, 698–707 (2017).
56. Chen, Y., Wang, J., Li, G., Yang, Y. & Ding, W. Current advancements in sactipeptide natural products. *Front. Chem.* **9**, 595991 (2021).
57. Claverys, J.-P., Martin, B. & Håvarstein, L. S. Competence-induced fratricide in streptococci. *Mol. Microbiol.* **64**, 1423–1433 (2007).
58. Claverys, J.-P. & Håvarstein, L. S. Cannibalism and fratricide: mechanisms and raisons d'être. *Nat. Rev. Microbiol.* **5**, 219–229 (2007).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Materials and strains

All materials were obtained from MilliporeSigma or Fisher Scientific unless otherwise specified. Restriction enzymes, Gibson Assembly Master Mix, T4 DNA ligase, Q5 high-fidelity DNA polymerase, Q5 Site-Directed Mutagenesis Kit, Trypsin-ultra mass spectrometry grade, endoproteinase GluC, shrimp alkaline phosphatase (rSAP), and the corresponding buffers were purchased from New England BioLabs (NEB). DNA purification kits were purchased from Qiagen. The *kgrA*, *kgrB*, *kgrC* and *kgrD* genes, codon-optimized for expression in *E. coli*, were purchased from GENEWIZ as linear gene fragments. *S*-(Methyl-*d*3)-*S*-adenosylmethionine was purchased from CDN Isotopes. *E. cecorum* ATCC 43198 was purchased from ATCC.

General procedures

Ultraviolet-visible spectra were acquired on a Cary 60 ultraviolet-visible spectrophotometer (Agilent). Low-resolution HPLC-MS analysis was performed on an Agilent instrument consisting of a liquid autosampler, a 1260 Infinity Series HPLC system coupled to a photodiode array detector and a 6120 Series ESI mass spectrometer. HR-HPLC-MS for the *in vitro* reactions was carried out on an Agilent UHD Accurate Mass Qtof liquid chromatography-mass spectrometry (LC-MS) system, equipped with a 1260 Infinity Series HPLC, an automated liquid sampler, a photodiode array detector, a JetStream ESI source and a 6540 Series Qtof. HR-HPLC-MS for the identification and analysis of enteropeptin was carried out on an Agilent UHD Accurate Mass Qtof LC-MS system, equipped with a 1290 Infinity II Series HPLC, an automated liquid sampler, a photodiode array detector, a JetStream ESI source and a 6546 Series Qtof. HPLC purifications were carried out on an Agilent 1260 Infinity Series analytical or preparative HPLC system equipped with a temperature-controlled column compartment, a diode array detector, an automated fraction collector and an automated liquid sampler. NMR spectra were acquired at the Princeton University Department of Chemistry Facilities. $1\text{D }^1\text{H}$ and 2D NMR spectra were collected in the triple-resonance cryoprobe of a Bruker Avance III 800 MHz NMR spectrometer or Bruker Avance III 500 MHz NMR spectrometer. KgrA substrate and product peptide samples were prepared in D_2O in a 2.5-mm-diameter tube. All NMR data were analysed with MestReNova software.

Construction of pET28b(+)*_kgrA*, pET28b(+)*_kgrB*, pET28b(+)*_kgrC* and pET28b(+)*_kgrD*

Protein sequences for each gene in the *kgr* gene cluster were identified from *E. cecorum* ATCC 43198 (Uniprot ID: [S1RLF6](#); NCBI reference sequence: WP_016250708.1). The codon-optimized gene fragments for each gene were purchased from GENEWIZ with 5'- and 3'-terminal extensions containing the necessary overlap sequence to be used in DNA assembly reactions for insertion into pET28b(+) to allow for expression with an N-terminal His tag. The *kgrD* gene began with a GTG start codon within the end of the *kgrC* gene so an ATG start codon was added to the beginning of the sequence to allow for expression in *E. coli*. The pET28b(+) vector was digested with NdeI and XhoI and was treated with rSAP prior to DNA assembly. DNA assembly reactions were performed with HiFi DNA Assembly Master Mix (NEB). Reactions were incubated for 1 h at 50 °C and then transformed into chemically competent *E. coli* DH5 α cells. The transformants were selected on an LB agar plate containing kanamycin (50 $\mu\text{g ml}^{-1}$). The target vectors were isolated from a liquid culture of a single colony using the QIAprep Spin Miniprep Kit (Qiagen). Insertion of the gene fragment was confirmed by Sanger sequencing with the T7 promoter and terminator primers. Vectors pET28b(+)*_kgrA* and pET28b(+)*_kgrD* were then transformed into chemically competent *E. coli* BL21(DE3) cells for expression. Vectors pET28b(+)*_kgrB* and pET28b(+)*_kgrC* were each co-transformed with pDB1282, which carries the *isc* operon required for proper assembly

of the Fe-S clusters, into chemically competent *E. coli* BL21(DE3) cells for expression.

Expression of KgrA

Eight 4 l flasks each containing 1.5 l of LB supplemented with kanamycin (50 $\mu\text{g ml}^{-1}$) were inoculated with 1% of a culture grown overnight from a single colony of *E. coli* BL21(DE3) cells carrying the pET28b(+)*_kgrA* plasmid. Cells were grown at 37 °C, 180 r.p.m. to an OD_{600} of 0.7 and before induction with isopropylthiogalactoside (IPTG) to a final concentration of 0.1 mM. Cells were then grown for 8 h at 37 °C, 180 r.p.m. before being harvested by centrifugation. A typical yield was 2.5 g of pellet per litre of culture.

Purification of KgrA

The purification was carried out at 4 °C in a cold room. Lysis buffer consisted of 50 mM HEPES, 300 mM KCl, 10% glycerol, 0.05% Triton X-100, 2 mM imidazole, pH 7.5, and 2 mM tris(2-carboxyethyl)phosphine (TCEP). Wet cell paste was resuspended in lysis buffer (5 ml per g cell paste). Phenylmethanesulfonyl fluoride was added to a final concentration of 0.1 mM and Protease Inhibitor Cocktail was added to a concentration of 10 $\mu\text{l ml}^{-1}$. Lysozyme and DNaseI were added to final concentrations of 1 mg ml^{-1} and 9.5 U ml^{-1} of lysis buffer, respectively, and the suspension stirred for 30 min at room temperature. The cells were then incubated for 30 min stirring on ice and subsequently sonicated for a total of 4 min in 15 s on/15 s off cycles at a 30% power rating inside the glovebox. The cells were then allowed to rest for 4 min, and the cycle was repeated. The crude cell lysate was then transferred to centrifuge tubes and cell debris pelleted by centrifugation (4 °C, 35,000g, 65 min). The supernatant was applied to a nickel metal affinity resin (5 ml) which had been equilibrated with lysis buffer. The column was washed with 30 ml of lysis buffer followed by 30 ml of wash buffer (50 mM HEPES, 300 mM KCl, 20 mM imidazole, 10% glycerol, 2 mM TCEP, pH 7.5). His-KgrA eluted within 1.5–2 column volumes of elution buffer (50 mM HEPES, 300 mM KCl, 300 mM imidazole, 15% glycerol, 2 mM TCEP, pH 7.5). His-KgrA was subsequently desalted on a Sephadex G-25 column (~50 ml, diameter = 1.25 cm, length = 25 cm) which had been equilibrated in G-25 buffer (100 mM HEPES, 300 mM KCl, 10% glycerol, 5 mM DTT, pH 7.5). Fractions containing His-KgrA, as judged by the Bradford assay, were pooled and concentrated. To remove glycerol and buffer components to allow for lyophilization and quantification by weight, KgrA was further purified on a semipreparative Phenomenex Jupiter C18 column (300 Å, 5 μm , 250 $\times$ 10 mm) with gradient from 20% to 100% MeCN + formic acid over 20 min. A typical yield for KgrA was 2–3 mg of pure peptide per litre of culture grown.

Expression of KgrB

Expression of KgrB was carried out using previously published procedures for the expression of radical SAM enzymes without modifications⁵⁹. Four 4 l flasks each containing 2.5 l of M9 media (supplemented with 2 mM $\text{Mg}(\text{SO}_4)_2$, 0.1 mM CaCl_2 , 20 mM glucose, 25 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 50 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ ampicillin) were inoculated with 1% of a culture grown overnight from a single colony of *E. coli* BL21(DE3) cells carrying the pET28b(+)*_kgrC* and pDB1282 vectors. Cells were grown at 37 °C, 180 r.p.m. to an OD_{600} of ~0.3 and were supplemented with 25 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 0.3 mM L-Cys and 0.2% (w/v) L-(+)-arabinose. They were then grown to an OD_{600} of ~0.7 and were cooled to 18 °C before induction with 100 μM IPTG. Cells were grown overnight at 18 °C with no shaking and were harvested by centrifugation after 18 h. A typical yield was 2 g of pellet per litre of culture.

Expression of KgrC

Expression of soluble KgrC was found to be best using modified conditions from our standard, previously published radical SAM enzyme expression conditions. Four 4 l flasks each containing 2.5 l of M9 media (supplemented with 2 mM $\text{Mg}(\text{SO}_4)_2$, 0.1 mM CaCl_2 , 20 mM glucose,

50 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ ampicillin) were inoculated with 1% of a culture grown overnight from a single colony of *E. coli* BL21(DE3) cells carrying the pET28b(+), *kgrC* and pDB1282 vectors. Cells were grown at 37 °C, 180 r.p.m. to an OD₆₀₀ of ~0.3 and were supplemented with 0.2% (w/v) L-(+)-arabinose. They were then grown to an OD₆₀₀ of ~0.7 and were cooled to 18 °C and supplemented with 5 μM FeCl₃ before induction with 25 μM IPTG. Cells were grown overnight at 18 °C, 110 r.p.m. and were harvested by centrifugation after 18 h. A typical yield was 2 g of pellet per litre of culture.

Purification of KgrB and KgrC

Purification of KgrB and KgrC were carried out using previously published procedures without modifications⁵⁹. Purification was carried out as described for the purification of KgrA through the G-25 column except all steps were carried out inside an MBraun glovebox and cobalt metal affinity resin was used rather than nickel. KgrB and KgrC were both found to be active without chemical reconstitution using Fe²⁺ and S²⁻.

Expression of KgrD

Two 4 l flasks each containing 2.5 l of M9 supplemented with 50 $\mu\text{g ml}^{-1}$ kanamycin, 2 mM Mg(SO₄)₂, 0.1 mM CaCl₂, 20 mM glucose, 0.1 mM MnCl₂ and 25 μM Fe(SO₄)₂, were inoculated with 1% of a culture grown overnight from a single colony of *E. coli* BL21(DE3) cells carrying the pET28b(+)-*kgrD* plasmid. Cells were grown at 37 °C, 200 r.p.m. to an OD of 0.7 and were cooled to 18 °C before induction with 0.1 mM IPTG and supplementation with 0.1 mM MnCl₂. Cells were grown overnight at 18 °C, 200 r.p.m. and were harvested by centrifugation after 18 h. A typical yield was 1.4 g of pellet per litre of culture.

Purification of KgrD

The purification of KgrD was carried out exactly as described for the purification of KgrA with two modifications: (1) DNA precipitation was carried out rather than use of DNase I; (2) all buffers were supplemented with 0.1 mM MnCl₂.

Enzymatic activity assays

Enzymatic assays were performed in an inert atmosphere in a glovebox (MBraun). The KgrA peptide substrate was brought into the glovebox as aliquots of lyophilized powder for the appropriate reaction volume. Reactions were typically carried out in 1.5 ml Eppendorf tubes on a 50 μl scale and contained assay buffer (100 mM HEPES, 300 mM KCl, 10% glycerol, pH 7.5), 14 μM KgrB, 6 μM KgrC, 13 μM KgrD, 200 μM titanium citrate, 5 mM DTT, 1 mM SAM and 250 μM KgrA. After incubation overnight, reactions were removed from the glovebox and digested with trypsin. For digestion by trypsin, CaCl₂ was added to the sample from a 2 M stock to achieve a final concentration of 20 mM. Trypsin-ultra mass spectrometry grade or endoproteinase GluC (NEB) was added to a final concentration of 100:1 (Trypsin-ultra) or 50:1 (GluC) substrate to protease. Digestion reactions were incubated for 5 h at 37 °C. Precipitated protein was removed by centrifugation. Samples were then analysed by HR-HPLC-MS as described above.

Large-scale enzymatic assay

A large-scale reaction to generate enough product for NMR analysis was carried out on a 5 ml scale with the same assay components and concentrations as described above. Reactions were allowed to incubate overnight and were then brought out of the glovebox and digested with trypsin as described above.

Purification of reaction product

Following digestion with trypsin, the 17mer KgrA product fragment was isolated by HPLC. Repeated injections were performed on a semipreparative Phenomenex Synergi Fusion-RP column (80 Å, 4 μm , 250 × 10 mm), which was equilibrated in 20% MeCN in H₂O (+ 0.1% formic acid). The peptide was eluted with a gradient of 20–65% MeCN

over 20 min at a flow rate of 2.0 ml min⁻¹. The product and remaining substrate eluted in a single peak around 18 min. Fractions were analysed by HR-HPLC-MS and those containing the expected product fragment were pooled and dried by lyophilization. The lyophilized material was resuspended in 30% MeCN in H₂O (+ 0.1% formic acid) and repeated injections were performed on an analytical Phenomenex Kinetex C18 column (100 Å, 2.6 μm , 100 × 4.6 mm), which was equilibrated with 20% MeCN in H₂O (+ 0.1% formic acid). The peptide was eluted with a gradient of 20–35% MeCN over 22 min at a flow rate of 0.5 ml min⁻¹. Fractions containing the pure product were pooled and lyophilized.

Synthesis of KgrA 17mer substrate standard

The 17mer KgrA substrate standard was synthesized via Fmoc-based solid-phase chemistry using an automated microwave peptide synthesizer (CEM). Preloaded Leu-Wang resin (0.05 mmol) and standard Fmoc and side-chain-protected amino acids were used. The resin was swelled in dimethylformamide (DMF) (10 ml) for 3 min. Fmoc deprotection reactions were performed using 10% piperazine (w/v) in a 10:90 solution of ethanol:*N*-methylpyrrolidinone (NMP) with 0.1 M 1-hydroxybenzotriazole (HOBt). The activator solution consisted of 0.5 M *N,N'*-diisopropylcarbodiimide (DIC) in DMF and activator base solution of 1.0 M Oxyma with 0.1 M *N,N'*-diisopropylethylamine (DIPEA) in DMF. Amino acids were coupled with a single coupling reaction for 4 min at 90 °C, with 5 equiv. of each amino acid. Leu at position 10 and the final three amino acids were double-coupled. Arg was coupled at 75 °C. Upon completion of the synthesis, the resin was removed from the reaction vessel of the peptide synthesizer and transferred to an Econo-Pac column (BioRad). The resin was washed several times with DMF, followed by DCM, then dried thoroughly under vacuum. The peptide was cleaved from the resin by incubation with freshly prepared cleavage cocktail (5 ml) consisting of 92.5% trifluoroacetic acid (TFA), 2.5% H₂O, 2.5% triisopropylsilane (TIS) and 2.5% 2,2'-(ethylenedioxy)diethanethiol (DODT). The reaction was stirred for 3 h at room temperature. The mixture was drained from the reaction tube and the resin was rinsed several times with TFA. The filtrate and rinses were combined and subsequently concentrated by evaporation of TFA under a stream of nitrogen. The peptide was precipitated by addition of 10 volumes of ice-cold diethyl ether. The ether/peptide suspension was incubated at –20 °C for 20 min before isolating the precipitated peptide by centrifugation (4 °C, 3,500g, 15 min). The ether was poured off and the peptide was dried overnight.

Purification of KgrA 17mer substrate standard

Crude KgrA was dissolved in 40% MeCN and 50 mM ammonium bicarbonate. Diluted ammonium hydroxide was added to bring the solution to pH 7. The peptide was purified by preparative HPLC on a preparative Phenomenex Luna C18 column (100 Å, 5 μm , 250 × 21.20 mm) which was equilibrated in 8% MeCN in H₂O (+ 0.1% formic acid). The peptide was eluted with a gradient of 8–58% MeCN over 18 min. KgrA was further purified on a semipreparative Phenomenex Synergi Fusion-RP column (80 Å, 4 μm , 250 × 10 mm) which was equilibrated in 20% MeCN in H₂O (+ 0.1% formic acid). The peptide was eluted with a gradient of 20–35% MeCN over 25 min.

Time-course assays

Time-course assays with KgrC and KgrB were performed with assay conditions as described above. Time-course reactions were carried out in duplicate. Two 200 μl reactions were set up and at each time point, 20 μl was removed from each reaction and immediately quenched with an equal volume of MeCN + 1% formic acid. Samples were removed from the glovebox and 200 μl of water was added to each. Reactions were frozen and lyophilized. Once dry, samples were resuspended in trypsin buffer and digested with trypsin as described above. Samples were analysed by HR-HPLC-MS using a Phenomenex Jupiter C18 column (300 Å, 5 μm , 100 × 4.6 mm) with a flow rate of 0.5 ml min⁻¹ and an isocratic step at 5% for 5 min, followed by a gradient of 5–70% MeCN over 13 min.

Metal-dependence of KgrD

KgrD was expressed on a small scale (50 ml M9) in the presence or absence of Mn^{2+} under conditions equivalent to the large-scale expression. For the condition with Mn^{2+} , 100 μM MnCl_2 was added before inoculation and again at the time of IPTG induction. Small-scale purifications were carried out using HisPur Ni-NTA spin columns. The lysis, wash and elution buffers used were identical to those used for the large-scale purification of His-KgrD. Cell pastes were resuspended in lysis buffer (5 ml per 200 mg cell paste), which consisted of 50 mM HEPES, 300 mM KCl, 10% glycerol, 0.05% Triton X-100, 2 mM imidazole, pH 7.5, 2 mM TCEP and 0.1 mM MnCl_2 for the condition with MnCl_2 . They were supplemented with 10% (v/v) BugBuster 10 $\times$ Protein Extraction Reagent. Benzonase-nuclease (25 U μl^{-1}) and protease inhibitor cocktail were added at 1 μl ml^{-1} and 10 μl ml^{-1} , respectively. The suspensions were incubated on ice on a rocking platform for 30 min. Crude cell lysate was centrifuged to remove cell debris. Clarified lysate was loaded onto a HisPur Ni-NTA spin column (0.2 ml resin bed volume, Thermo-Fisher), which had been equilibrated with lysis buffer. Once loaded, the resin was washed with five resin bed volumes of lysis buffer followed by 5 volumes of wash buffer. His-KgrD was eluted from the resin using 2–4 volumes of elution buffer. Protein elution was monitored by Bradford reagent. Isolated His-KgrD was exchanged into G-25 buffer (100 mM HEPES, 300 mM KCl, 10% glycerol, 5 mM DTT, pH 7.5) using a PD-10 desalting column (GE Healthcare) following the manufacturer's instructions. Following purification, His-KgrD was brought into the glovebox for assays. Assays were set up containing assay buffer (100 mM HEPES, 300 mM KCl, 10% glycerol, pH 7.5), 4 μM KgrC, KgrD from the small-scale expression with or without Mn^{2+} (concentration could not be quantified accurately), 200 μM titanium citrate, 5 mM DTT, 1 mM SAM, 250 μM KgrA, and were supplemented with 50 mM MnCl_2 or no additional metal. Samples were then analysed by HR-HPLC–MS as described above.

Construction of pRSFDuet-1_6HMBP*kgrA_kgrCD*

The *kgrA* gene was PCR-amplified from its synthetic DNA fragment using primers Duet_ *kgrA*_F (BamHI) and Duet_ *kgrA*_R (PstI) (Supplementary Table 3). The PCR product was then used in a HiFi assembly reaction to insert it into the first multiple cloning site (MCS1) of pRSFDuet-1, which had been linearized using the same restriction enzymes. Reactions were incubated for 1 h at 50 °C and then transformed into chemically competent *E. coli* DH5 α cells. The transformants were selected on an LB agar plate containing kanamycin (50 μg ml^{-1}). The target vector was isolated from a liquid culture of a single colony using the QIAprep Spin Miniprep Kit (Qiagen). Insertion of the gene fragment was confirmed by Sanger sequencing with the DuetDOWN1 primer. The pRSFDuet-1_6HMBP*kgrA* vector was then digested with NdeI and XhoI, opening up MCS2. The *kgrC* gene was PCR-amplified from its synthetic DNA fragment using primers Duet_ *kgrC*_F and Duet_ *kgrC*_RBS_R to add the proper overlap regions for HiFi assembly, and an artificial ribosome binding site between the *kgrC* and *kgrD* genes. The *kgrD* gene was also PCR-amplified from its synthetic DNA fragment using primers Duet_RBS_ *kgrD*_F and Duet_ *kgrD*_R to add the proper overlap regions for HiFi assembly. The resulting PCR products were then used in a HiFi DNA assembly reaction with vector pRSFDuet-1_6HMBP*kgrA* to give pRSFDuet-1_6HMBP*kgrA_kgrCD*. The target vector was isolated from a liquid culture of a single colony using the QIAprep Spin Miniprep Kit (Qiagen). Insertion of the gene fragment was confirmed by Sanger sequencing with the DuetUP1 and T7 terminator primers.

Analytical-scale coexpression of KgrC, KgrD and 6HMBP-KgrA proteins

The pRSFDuet-1_6HMBP*kgrA_kgrCD* vector, along with pDB1282 (*isc* operon), was transformed into *E. coli* BL21(DE3) by heat-shock and plated onto LB agar supplemented with 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} kanamycin. A 125 ml flask containing 50 ml LB (+ 50 μg ml^{-1}

kanamycin and 100 μg ml^{-1} ampicillin) was inoculated with a single colony and grown overnight at 37 °C, 200 r.p.m. The following morning, 7.5 ml of overnight culture was used to inoculate each of six 2 l flasks containing 0.75 l Terrific Broth (kanamycin/ampicillin). This 50 ml culture was grown at 37 °C, 200 r.p.m. to an OD_{600} of 0.4, at which point it was supplemented with 0.05 mM FeCl_3 and 0.05% arabinose to induce expression of the *isc* operon. The culture was then grown to an OD_{600} of $\sim$ 0.8 at which point it was supplemented with 0.1 mM IPTG to induce coexpression of 6HMBP-KgrA, KgrC and KgrD. Growth was continued at 37 °C, 200 r.p.m. for 18 h. Cells were harvested by centrifugation (30,000g, 30 min, 4 °C).

Analytical-scale purification and protease digest of 6HMBP-KgrA

The lysis, wash and elution buffers used were identical to those used for the purification of His-KgrA. Cell paste was resuspended in lysis buffer (5 ml per 200 mg cell paste), which consisted of 50 mM HEPES, 300 mM KCl, 10% glycerol, 0.05% Triton X-100, 2 mM imidazole, pH 7.5, and 2 mM TCEP. It was supplemented with 10% (v/v) BugBuster 10 $\times$ Protein Extraction Reagent. Benzonase-nuclease (25 U μl^{-1}) and protease inhibitor cocktail were added at 1 μl ml^{-1} and 10 μl ml^{-1} , respectively. The suspension was incubated on ice on a rocking platform for 30 min. Crude cell lysate was centrifuged to remove cell debris. Clarified lysate was loaded onto a HisPur Ni-NTA spin column (0.2 ml resin bed volume, Thermo-Fisher), which had been equilibrated with lysis buffer. Spin purification was carried out as per instructions. Briefly, once loaded, the resin was washed with 5 resin bed volumes of lysis buffer followed by 5 volumes of wash buffer. 6HMBP-KgrA was eluted from the resin using 2–4 volumes of elution buffer. Protein elution was monitored by Bradford reagent. Isolated 6HMBP-KgrA was exchanged into cleavage buffer (50 mM Tris–HCl, 150 mM NaCl, pH 8) or water using a PD-10 desalting column (GE Healthcare) following the manufacturer's instructions. Once isolated, the 6HMBP-KgrA fusion proteins were proteolysed to facilitate analysis by HPLC–Qtof-MS. First, the fusion protein was treated with HRV3C protease to remove the 6HMBP tag. Next, the HRV3C-cleaved KgrA peptide was further digested using trypsin. All protease reactions were performed in cleavage buffer composed of 50 mM Tris–HCl, 150 mM NaCl, pH 8. For trypsin cleavage, the buffer was supplemented with CaCl_2 at a final concentration of 20 mM. Proteolysed samples were then analysed by HPLC–Qtof-MS using a Phenomenex Jupiter C18 column (300 Å, 5 μm , 100 $\times$ 4.6 mm) at a flow rate of 0.5 ml min^{-1} with an isocratic step at 5% for 5 min, followed by a gradient of 5–70% MeCN over 13 min.

Large-scale coexpression and purification 6HMBP-KgrA

E. coli containing the pRSFDuet-1_6HMBP*kgrA_kgrCD* vector, along with pDB1282 (*isc* operon) was grown exactly as described above but in six 2 l flasks, each containing 0.75 l of TB. The yield was 30 g of cell pellet per 4.5 l of culture. The purification was carried out at 4 °C in a cold room exactly as described for purification of His-KgrA. The yield was 101 mg of protein.

Enzymatic activity assays with KgrB

Enzymatic assays were performed in an inert atmosphere in a glovebox (MBraun). The 6HMBP-KgrA purified after coexpression with KgrC and KgrD was brought into the glovebox as aliquots of lyophilized powder for the appropriate reaction volume. Reactions were typically carried out in 1.5 ml Eppendorf tubes on a 50 μl scale and contained assay buffer (100 mM HEPES, 300 mM KCl, 10% glycerol, pH 7.5), 12 μM KgrB, 200 μM titanium citrate, 5 mM DTT, 1 mM SAM and 250 μM substrate. After incubation overnight, reactions were removed from the glovebox and digested with trypsin. For digestion by trypsin, CaCl_2 was added to the sample to achieve a final concentration of 20 mM. Trypsin-ultra mass spectrometry grade was added to a final concentration of 100:1 substrate to trypsin. Digestion reactions were incubated for 5 h at 37 °C.

Precipitated protein was removed by centrifugation. Samples were then analysed by HPLC–Qtof–MS using a Phenomenex Jupiter C18 column (300 Å, 5 µm, 100 × 4.6 mm) with a flow rate of 0.5 ml min^{−1} and with an isocratic step at 5% for 5 min, followed by a gradient of 5–70% MeCN over 13 min. Assays with isotopically labelled SAM were carried out exactly as described with 1 mM S-(methyl-*d*3)-S-adenosylmethionine. Assays to look for production of 5'-dA were carried out as described above but were heat quenched rather than digested with trypsin. The assays were centrifuged, and the supernatant was analysed by HPLC–MS on an Analytical Phenomenex Synergi Fusion-RP column (80 Å, 4 µm, 100 × 4.6 mm).

Bioinformatic analysis of KgrB

A PSI-BLAST (Position-Specific Iterated BLAST) search was carried out on the KgrB protein sequence. Three iterations were carried out, and after each first round, sequences with an *E* value better than the default threshold were selected for the next round. After the third round, the top 5,000 hits were used to generate a sequence similarity network using the EFI-EST web tool with the default settings⁶⁰. An alignment score of 75 was used to generate the final network, which was visualized in Cytoscape⁶¹. The 5,000 sequences were aligned using the Clustal Omega online software and this alignment was utilized to generate a LogoPlot using the WebLogo3 website^{62,63}.

Mutagenesis of KgrA

Mutagenesis of the KgrA substrate was carried out on vector pET28b(+)_kgrA using the Q5 Site-Directed Mutagenesis Kit (NEB) to generate an S57E/I66D mutant. Primers (Supplementary Table 3) were designed using NEBaseChanger software. The standard protocol was followed with two modifications: a gel-extraction step was included to isolate the correct plasmid band after the PCR step, and the KLD reaction was incubated at room temperature for 3 h rather than 15 min before transformation. Single colonies were selected for sequencing and plasmids with the correct double mutations were subsequently transformed into chemically competent *E. coli* BL21(DE3) cells for expression. Substrate variants were expressed, purified and isolated exactly as described for the wild-type substrate.

Identification of enteropeptin from *E. cecorum*

Identification and isolation of enteropeptin from *E. cecorum* cultures, and bioactivity assessments, were performed in chemically defined media (CDM)⁶⁴. *E. cecorum* was inoculated into Todd-Hewitt yeast extract (THY) broth from glycerol stocks and cultured overnight at 37 °C and 5% CO₂. Overnight cultures were centrifuged at 4,000g for 5 min, and bacteria were resuspended in prewarmed CDM. A 2 ml portion of this was then used to inoculate 200 ml (1% inoculum) of CDM in a 250 ml Erlenmeyer flask which was then incubated at 37 °C and 5% CO₂. After 18 h, the culture was centrifuged and filtered, and the supernatant was passed through a washed and equilibrated 0.5 g Oasis HLB solid-phase extraction cartridge. The HLB cartridge was then washed with 2 column volumes of water + 0.1% formic acid and eluted with 4 column volumes of 50% MeCN + 0.1% formic acid. The elution was dried in vacuo, resuspended in 200 µl water + 0.1% formic acid (1,000× concentrated) and characterized by HPLC–Qtof–MS. HPLC–Qtof–MS utilized a Phenomenex Omega Polar C18 column (5 µm, 2.1 × 150 mm) with a gradient from 100% water + 0.1% formic acid to 100% MeCN over 16 min with a flow rate of 0.75 ml min^{−1}. The HPLC–Qtof–MS data were analysed as described previously¹⁷. Briefly, molecular formulae were first generated for all possible products of the *kgr* locus from *E. cecorum* based on the available information from in vitro studies, and the Agilent MassHunter software “Find Compounds by Formula” feature was used to identify possible matches within the LC–MS data. These matches were then selected for MS/MS fragmentation, and resultant spectra were inspected for fragmentation spectra consistent with the candidate KGR peptide sequence. Candidate ions for three products

(SMGKOCPPS, MGKOCPPS and MGKOCPP) were observed with MS/MS fragmentation spectra consistent with their respective peptide sequences. KGR A was the most abundant product observed with the sequence MGKOCPPS. A synthetic KGR A was prepared as described below and the authentic and synthetic KGR A were found to coelute when mixed and analysed by HPLC–Qtof–MS as described above.

Enteropeptin standard preparation

KgrA(S57E/I66D) was expressed and purified exactly as described for the wild-type KgrA. It was used as a substrate in an in vitro reaction with KgrB, KgrC and KgrD as described for the wild-type reactions. Following the reaction, the peptide was co-digested with AspN and GluC in a 50:50 mixture of the AspN and GluC cleavage buffers to give the final product with an *m/z* of 845.39 as analysed by HPLC–Qtof–MS.

Enteropeptin time-course of production

E. cecorum was inoculated into THY media from glycerol stocks and cultured overnight at 37 °C and 5% CO₂. Overnight cultures were centrifuged at 4,000g for 5 min, and bacteria were resuspended in prewarmed CDM. This was then used to inoculate two 125 ml Erlenmeyer flasks, each with 100 ml of CDM at an OD₆₀₀ of 0.01. Aliquots of 1 ml were collected from each flask every 6 h. For each aliquot, the absorbance at 600 nm was measured. The concentration of enteropeptin was measured at each time point by HPLC–Qtof–MS through direct injection of filtered culture supernatants. Sample supernatants were passed through a Phenomenex Omega Polar C18 column (5 µm, 2.1 × 150 mm) with a gradient from 100% water + 0.1% formic acid to 100% MeCN over 8 min with a flow rate of 0.75 ml min^{−1}. The concentration of enteropeptin in the supernatant was estimated by MS through comparison to a standard curve made with an unmodified linear 8mer peptide standard using the same method.

Large-scale isolation of enteropeptin from *E. cecorum*

Overnight cultures containing 10 ml of brain heart infusion (BHI) medium in a 14 ml culture tube were inoculated from a glycerol stock of *E. cecorum* ATCC 43198. The culture was grown statically overnight at 37 °C in an incubator containing 5% CO₂. Two 2 l bottles containing 2 l of chemically defined medium were each inoculated with 20 ml of overnight culture and grown statically at 37 °C for 14–18 h. Cultures were centrifuged for 30 min, and the supernatant was filtered using 0.2 µm Corning filter units. The filtered supernatant was passed through a porous graphitic carbon column (30 ml bed volume), which had been washed with NaOH, rinsed with water, washed with MeCN and equilibrated with H₂O supplemented with 0.1% formic acid. Following application of the filtered supernatant, the column was washed with H₂O supplemented with 0.1% formic acid and bound material eluted with 50% MeCN in H₂O supplemented with 0.1% formic acid. The elution was dried by rotary evaporation, resuspended in H₂O and purified on a semipreparative Phenomenex Luna Omega Polar column (100 Å, 5 µm, 250 × 10 mm) running at 2 ml min^{−1}. Elution was carried out with a gradient of 2.5–6% MeCN over 20 min. Both MeCN and H₂O contained 0.1% formic acid. Enteropeptin eluted in a broad peak between 16.5 and 19 min. Fractions containing enteropeptin, as determined by HPLC–MS, were pooled and dried, giving approximately 0.5 mg of material. Material was resuspended in H₂O and was further purified on an analytical Phenomenex Kinetex Polar C18 column (100 Å, 2.6 µm, 150 × 4.6 mm) running at 0.75 ml min^{−1}. Elution was carried out with a gradient of 0.5–6% MeCN over 20 min. Both MeCN and H₂O contained 0.1% formic acid. Enteropeptin eluted in a broad peak between 12 and 14 min. Fractions containing enteropeptin were pooled and dried. This did not yield enough material to accurately quantify by weight so the concentration of enteropeptin was again estimated by comparison to a standard curve generated with an unmodified 8mer peptide standard. From this, we measured the final mass to be 0.013 mg from 4 l of culture.

Bioactivity of enteropeptin

To monitor the effect of enteropeptin, glycerol stocks of bacterial strains were inoculated into 10 ml of either LB (*E. coli*, *S. aureus* and *P. aeruginosa*), BHI (*E. faecalis* and *S. thermophilus*) or THY (*E. cecorum*) broth. *E. coli*, *S. aureus*, *E. faecalis* and *P. aeruginosa* were incubated overnight at 37 °C, 200 r.p.m., while *S. thermophilus* and *E. cecorum* were incubated in an atmosphere of 5% CO₂ without shaking. The next morning, cultures were centrifuged at room temperature at 4,000g for 5 min. Strains were resuspended in prewarmed CDM, and then for each strain 100 µl was added to 10 ml of CDM (1% inoculum). Ninety-six-well plates were prepared using the 1% inoculum cultures with 100 µl of culture per well. Enteropeptin was serially diluted down the plates with a highest concentration of 3 µM. The 96-well plates were then cultured at either 37 °C, 200 r.p.m. (*E. coli*, *S. aureus*, *E. faecalis* and *P. aeruginosa*) or 37 °C, 5% CO₂ without shaking (*S. thermophilus* and *E. cecorum*) with a Breathe-Easy membrane. Growth was measured by OD₆₀₀ with a SynergyH1 microplate reader (BioTek) after 8 h.

Growth curve of *E. cecorum* with enteropeptin

Glycerol stocks of *E. cecorum* were streaked onto a plate containing BHI medium. A single colony was inoculated into 10 ml THY broth and were incubated overnight at 37 °C in an atmosphere of 5% CO₂ without shaking. The next morning, cultures were centrifuged at room temperature at 4,000g for 5 min. Strains were resuspended in prewarmed CDM. Resuspended strains were then inoculated into 5 ml CDM at an OD₆₀₀ of 0.01 and 100 µl of this culture was added to each well of a 96-well plate that had been sterilized by ultraviolet light. An additional 100 µl of culture and a 6 µM final concentration of enteropeptin were added to the required wells and this culture was serially diluted down the column to give a gradient of enteropeptin concentrations. All conditions were carried out in triplicate. Sterile water or a 6 µM final concentration of the linear 8mer peptide were added to the control wells. The plate was grown at 37 °C in an atmosphere of 5% CO₂ with a Breathe-Easy membrane. Growth was monitored approximately every 1 h, and at each time point strains were monitored by measuring OD₆₀₀ with a SynergyH1 microplate reader (BioTek).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Uniprot ID for KgrC is [S1RLF6](#). Plasmids and strains used in this study are described in Supplementary Table 2. All oligonucleotides are shown in Supplementary Table 3. The sequences of codon-optimized gene fragments are provided in Supplementary Note 1. Other relevant data supporting the findings of this study are available within the paper and the supplementary material. NMR spectra are included in Supplementary Information. Raw NMR data used to elucidate natural product structures are available from the corresponding author upon reasonable request.

References

59. Bushin, L. B. & Seyedsayamdost, M. R. Guidelines for determining the structures of radical SAM enzyme-catalyzed modifications in the biosynthesis of RiPP natural products. *Methods Enzymol.* **606**, 439–460 (2018).
60. Gerlt, J. A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochimica et Biophysica* **1854**, 1019–1037 (2015).
61. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
62. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
63. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
64. van de Rijn, I. & Kessler, R. E. Growth characteristics of group A streptococci in a new chemically defined medium. *Infect Immun.* **27**, 444–448 (1980).

Acknowledgements

We thank the Edward C. Taylor-Eli Lilly Fellowship in Chemistry (to K.A.C.) and the National Science Foundation (NSF CAREER Award 1847932 to M.R.S.) for support of this work.

Author contributions

K.A.C. and M.R.S. conceived of the study. K.A.C. carried out in vitro characterization and structural elucidation of the reactions of all enzymes. B.C.C. identified the mature natural products from the native organism. K.A.C. and B.C.C. carried out bioactivity assays. K.A.C. and M.R.S. wrote the manuscript, with contributions from B.C.C.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41557-022-01063-3>.

Correspondence and requests for materials should be addressed to Mohammad R. Seyedsayamdost.

Peer review information *Nature Chemistry* thanks Qi Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Agilent MassHunter Workstation Data Acquisition
Agilent OpenLab
BioTek Synergy/H1 microplate reader Gen5 2.08
Bruker Biospin Topspin

Data analysis

Agilent MassHunter Workstation Qualitative Analysis 10.0
MestReNova x64 14.1
Microsoft Office 16.60
PerkinElmer ChemOffice Suite 2019
Adobe Illustrator 2021
JGI IMG (<https://img.jgi.doe.gov>)
Benchling (<https://www.benchling.com/>)
Cytoscape 3.8.2
EFI-EST (<https://efi.igb.illinois.edu/>)
PSI-BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
WebLogo 3
EMBL-EBI Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data supporting the findings of this study are available within the paper and the supplementary material. NMR data used to characterize structures are available from the corresponding author upon reasonable request.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Bioactivity assays was performed in triplicates as is routine for these measurements and as justified by the low error and high degree of reproducibility observed.

Data exclusions

No data were excluded from any analyses.

Replication

All experimental results could be reliably replicated. All experiments were conducted independently at least twice.

Randomization

No experimental groups were used in this study, and randomization was therefore not relevant.

Blinding

Group allocation was not relevant to the experiments in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging