# Accepted Manuscript


Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway

Benjamin A. Schneider, Emily P. Balskus

PII: S0040-4020(18)30308-9

DOI: 10.1016/j.tet.2018.03.043

Reference: TET 29383

To appear in: *Tetrahedron*

Received Date: 18 January 2018

Revised Date: 15 March 2018

Accepted Date: 20 March 2018

Please cite this article as: Schneider BA, Balskus EP, Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway, *Tetrahedron* (2018), doi: 10.1016/j.tet.2018.03.043.

**Title:** Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway

**Authors:** Benjamin A. Schneider and Emily P. Balskus*

**Affiliations**: Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

* Corresponding author. Tel.: +1 617 496 9921; fax +1 617 496 5866; e-mail address: balskus@chemistry.harvard.edu

**Abstract:** Natural products from the human microbiota may mediate host health and disease. However, discovery of the biosynthetic gene clusters that generate these metabolites has far outpaced identification of the molecules themselves. Here, we used an isolation-independent approach to access the probable products of a nonribosomal peptide synthetase-encoding gene cluster from *Ruminococcus bromii*, an abundant gut commensal bacterium. By combining bioinformatics with in vitro biochemical characterization of biosynthetic enzymes, we predicted that this pathway likely generates an *N*-acylated dipeptide aldehyde (ruminopeptin). We then used chemical synthesis to access putative ruminopeptin scaffolds. Several of these compounds inhibited *Staphylococcus aureus* endoproteinase GluC (SspA/V8 protease). Homologs of this protease are found in gut commensals and opportunistic pathogens as well as human gut metagenomes. Overall, this work reveals the utility of isolation-independent approaches

for rapidly accessing bioactive compounds and highlights a potential role for gut microbial natural products in targeting gut microbial proteases.

**Keywords**: gut microbiota; natural product biosynthesis; nonribosomal peptide synthetase; peptide aldehyde; protease inhibitor; protease

## 1. Introduction

Small molecules produced by the human gut microbiota are potential mediators of this microbial community's effects on host health and disease.[1] However, the major inhabitants of the gut have not been extensively investigated as natural product producers. Though genome and metagenome sequencing continues to reveal that human gut microbes have a rich biosynthetic potential, discovering natural products from these organisms has proven challenging, in part because many cannot be cultivated in the laboratory. Moreover, investigations to date have found that gut microbial natural products are often difficult or impossible to isolate or are not produced under standard laboratory conditions.[1–3] Though there are limited examples of isolating small molecules produced by gut microbes in pure culture (e.g., ruminococcin A),[4] alternative strategies such as functional metagenomics[5] and expression of biosynthetic gene clusters in heterologous hosts[6] have also revealed gut microbial natural products (Fig. 1A). Overall, there is clearly a continued need for new approaches that will provide more rapid access to products of gut microbial gene clusters.

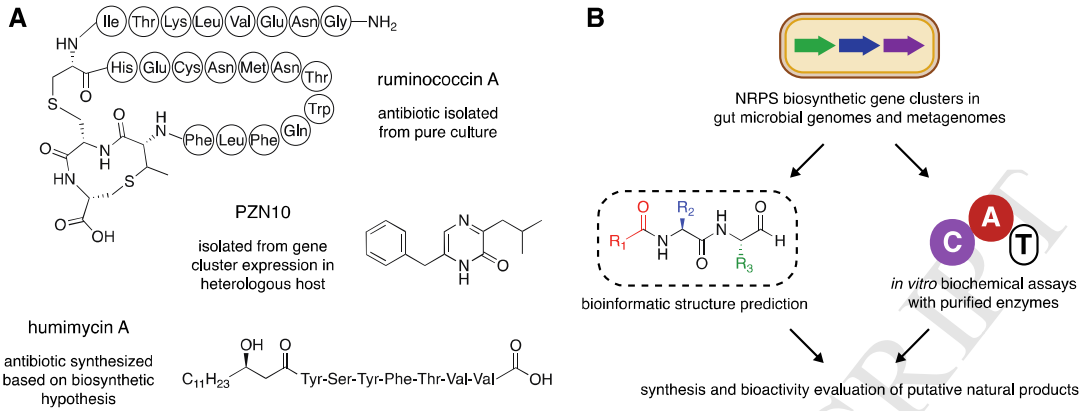



**Fig. 1:** Isolation-independent approaches may accelerate identification of bioactive natural products from the human gut microbiota. (A) Selected natural products from human gut bacteria, including the proposed structure of the lanthionine-containing bacteriocin (lantibiotic) ruminococcin A, PZN10, and the antibiotic humimycin. (B) Our isolation-independent workflow for characterizing small molecules produced by important gut commensals involves first selecting nonribosomal peptide synthetase (NRPS)-encoding biosynthetic gene clusters of interest based on abundance in metagenomic sequencing data and microbial ecology. Bioinformatic predictions and in vitro biochemical assays then provide structural information that informs the chemical synthesis of candidate natural product structures. These focused small molecule libraries can then be evaluated for bioactivity.

We envisioned a strategy for accessing gut microbial secondary metabolites that would combine in vitro characterization of biosynthetic enzymes with chemical synthesis (Fig. 1B). By mining human gut metagenomic sequence data, we could identify small nonribosomal peptide synthetase (NRPS) biosynthetic gene clusters of interest based on metagenomic sequencing data and microbial ecology. These enzymes share a conserved

chemical logic and would therefore be amenable to bioinformatic analyses and prediction of their natural product structures. We could then test our predictions and identify key biosynthetic building blocks using in vitro biochemical assays with purified biosynthetic enzymes. Finally, we would access the candidate natural product structures using chemical synthesis and evaluate these focused small molecule libraries for bioactivity. A key advantage of this approach is that it could provide a more rapid way to access bioactive small molecules compared to traditional isolation- or heterologous expression-based natural product discovery. Indeed, Brady and coworkers recently demonstrated the utility of a related strategy (the "synthetic-bioinformatic natural products", or syn-BNPs, approach) in their discovery of humimycin A (Fig. 1A).[7] By mining sequenced genomes from the human microbiota for NRPS gene clusters, predicting the structures of the likely gene cluster products using bioinformatics, and synthesizing the predicted nonribosomal peptides, they accessed a new antibiotic that is active against methicillin-resistant *Staphylococcus aureus* clinical isolates.

Here, we have used our isolation-independent approach to access the putative products of an NRPS gene cluster from *Ruminococcus bromii*, one of the most abundant commensal microbes in the human gut. We first employed bioinformatic analyses to predict the product of this conserved and widely distributed gene cluster (the *rup* gene cluster) as a reactive, *N*-acylated dipeptide aldehyde (ruminopeptin). We then used in vitro biochemical characterization of the NRPS assembly line enzymes to identify the building blocks of ruminopeptin. Using a short, solution phase synthesis, we accessed a library of ruminopeptin analogues and evaluated their bioactivities. We found these molecules

inhibit *S. aureus* endoproteinase GluC (SspA/V8 protease), which has been implicated in virulence in a mouse abscess model.[8] The human gut microbe and opportunistic pathogen *Enterococcus faecalis* also produces a virulence-related glutamyl endopeptidase,[9] and further bioinformatics analyses revealed additional homologs of this enzyme in gut microbial genomes and metagenomes. We hypothesize that protease inhibitors of this family may be important for mediating microbe-microbe interactions in the human gut.

## 2. Results and discussion

### 2.1 The prominent human gut microbe *Ruminococcus bromii* possesses an abundant and conserved biosynthetic gene cluster

With the goal of discovering bioactive secondary metabolites from the human gut microbiota, we initially focused on the prominent gut commensal *R. bromii*. This organism is one of the most abundant microbes in the human gut across a diversity of environments and diets, [10–13] and it has an important ecological role in the colon as a keystone species in the degradation of resistant starch.[14,15] *R. bromii* is a member of Clostridium cluster IV, which is significantly less abundant in patients with inflammatory bowel disease (IBD) as compared with healthy subjects.[16] This phylogenetic group of Clostridia contains organisms that are generally considered to be beneficial in the gut environment and includes *Faecalibacterium prausnitzii*, which has a well-studied anti-inflammatory role.[17] Though to our knowledge *R. bromii* has not yet been reported to

produce natural products, we hypothesized that this could be a mechanism by which this organism exerts its beneficial effects or maintains its ecological niche in the human gut.

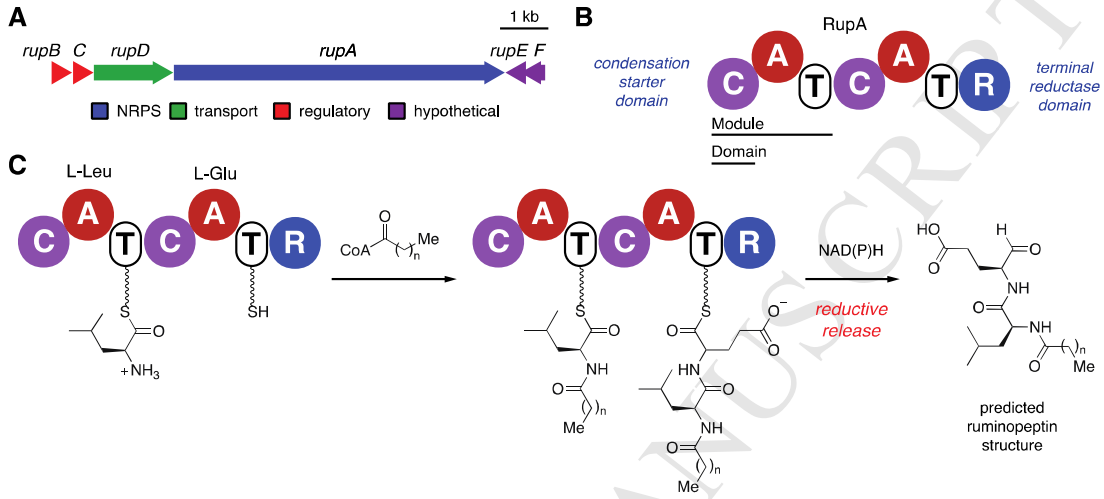



**Fig. 2:** A biosynthetic gene cluster from the abundant gut commensal *Ruminococcus bromii* encodes a putative peptide aldehyde natural product. (A) The *rup* gene cluster from *R. bromii*. The gene cluster encodes a single multi-modular NRPS, a transporter, two regulatory elements, and two hypothetical proteins. (B) The RupA NRPS contains a condensation-starter (C-starter) domain and a terminal reductase (R) domain (A = adenylation domain, T = thiolation domain). (C) Biosynthetic hypothesis for the production of ruminopeptin by the *rup* gene cluster.

*R. bromii* encodes a 10.9 kb biosynthetic gene cluster that encodes a single di-modular NRPS, an efflux pump (ABC transporter), two regulatory elements, and two hypothetical proteins (Fig. 2A, Table S1). The *rup* gene cluster (also known as bgc45) has been identified previously by Fischbach and co-workers in a large survey of biosynthetic gene clusters from the human microbiome and is part of a larger family of NRPS gene clusters

found in gut microbial genomes and metagenomes.[18] This study also revealed the *rup* gene cluster to be one of the most abundant gene clusters found in human microbiome project (HMP) stool metagenomes. Moreover, a highly similar gene cluster (bgc71, 97.2% nucleotide sequence identity) from a closely related, unisolated *Ruminococcus* species was identified in several RNAseq datasets from stool samples of healthy subjects, indicating that this biosynthetic pathway is likely expressed under physiological conditions.[6] Overall, these findings suggest the product of the *rup* gene cluster is likely produced under physiological conditions. Coupled with the established importance of *R. bromii*, this may indicate a particularly important role for this metabolite in the human gut microbiota.

Based on gene content and NRPS biosynthetic logic, we predicted that the *rup* gene cluster would produce a peptide aldehyde natural product. The NRPS (RupA) features a condensation-starter (C-starter) domain, indicating that the N-terminus of the product non-ribosomal peptide is likely *N*-acylated,[19] one complete NRPS module, and a terminal reductase (R) domain (Fig. 2B). This final domain should catalyze release of a nascent thioester intermediate from the NRPS enzyme, generating either an aldehyde or a primary alcohol-containing product.[20] A peptide aldehyde product would likely able to act as an inhibitor of serine, cysteine, or threonine proteases as has been demonstrated for NRPS-derived peptide aldehydes produced by soil microbes (e.g. fellutamide B[21] and the flavopeptins[22]). Notably, *Ruminococceae* are negatively correlated with protease activity in the colon,[23] and production of small molecule protease inhibitors by these organisms is a potential mechanism by which this association could arise.

If the product of the *rup* gene cluster does play a crucial role in *R. bromii*'s ecology and evolutionary history, we might expect it to be highly conserved in this species. To assess the presence of this gene cluster across *R. bromii* strains, we used PCR with specific primers to amplify a fragment of the first NRPS adenylation domain ($RupA_{A1}$) in three available human-derived *R. bromii* isolates (*R. bromii* L2-63, *R. bromii* ATCC 27255, and *R. bromii* 22-5-S 6 FAA NB).[14,24] We observed amplification in each strain (Fig. S1). We then subsequently PCR-amplified and sequenced the full gene clusters to reveal greater than 96% identity on the nucleotide level (Table S2). Conservation of this biosynthetic gene cluster across all known human-derived *R. bromii* isolates provides evidence that this pathway may be important for the organism's native biological role.

### 2.2 Bioinformatic analysis of the *rup* gene cluster predicts production of an *N*-acylated dipeptide aldehyde

In order to gain information about the product of the *rup* gene cluster, we first used bioinformatic analyses to predict the activities and substrate specificities of each of the domains in this two-module NRPS assembly line. RupA lacks an adenylation-thiolation (A-T) didomain loading module and instead contains a predicted C-starter domain. C-starter domains catalyze *N*-acylation of an initially loaded, assembly-line tethered amino acid with a fatty acyl-CoA. Multiple sequence alignments with biochemically and genetically characterized C-starter domains (ClbN, XcnA, and GlbF) revealed the RupA C-starter domain contains key conserved residues indicative of *N*-acylation activity (Fig.

S2).[25–27] We then used the Maryland PKS/NRPS server[28] to predict the substrate specificities of the two A domains of RupA. We found that the first NRPS module likely preferred L-leucine and the second NRPS module likely used either L-aspartate or L-glutamate (Fig. S3). Finally, we generated a structure-based multiple sequence alignment of the final $RupA_R$ domain with other characterized NRPS terminal R domains using PROMALS3d.[29] From this alignment, we could identify all of the key conserved active site residues involved in NAD(P)H binding as well as the Thr/Tyr/Lys catalytic triad required for thioester reduction (Fig. S4).

Together, these analyses allowed us to propose a biosynthetic hypothesis for the *rup* pathway and predict the structure of the final peptide aldehyde product(s), which we named ruminopeptin (Fig. 2C). After post-translational modification of the RupA T domains by a phosphopantetheine (ppant) transferase, initiation of biosynthesis occurs with the activation of L-leucine by the A domain of the first NRPS module and loading onto the ppant arm of the first T domain. The C-starter domain of the first module then acylates the amino group of the tethered L-leucine with a fatty acyl CoA. The resulting *N*-acylated aminoacyl thioester intermediate is then elongated by amide bond formation with the amino acid loaded by the second NRPS module, either L-aspartate or L-glutamate, to generate a nascent *N*-acylated dipeptide thioester intermediate. Finally, reductive offloading of this intermediate by the R domain will give a peptide aldehyde product that we have named ruminopeptin.

## 2.2 The product(s) of the *rup* gene cluster are not readily isolated from *R. bromii* cultures

Having proposed a candidate structure for the *rup* gene cluster product(s), we wondered if it would be possible to isolate these secondary metabolites from *R. bromii*. We began by identifying culture conditions in which the *rup* pathway was expressed. We cultivated two strains of *R. bromii* using a variety of nutrient sources and several unusual culture additives (rumen fluid, chopped meat broth). We extracted RNA from saturated cultures and assessed gene cluster expression using specific primers with single-step RT-PCR. We observed that including fructose as a carbohydrate source in growth media was necessary for *rup* gene cluster expression and that inclusion of additives had no effect (Fig. S5). However, in numerous attempts using cultures grown under conditions where the *rup* genes were expressed (5 mL to 1 L scales) we could not identify candidate masses corresponding to any predicted ruminopeptin peptide aldehyde product by LC-MS. We also attempted comparative metabolite profiling using XCMS,[30] but this analysis did not reveal any candidate masses of interest.

## 2.3 In vitro biochemistry reveals the building blocks of the *rup* gene cluster product ruminopeptin

Since we could not readily isolate the predicted product(s) of the *rup* gene cluster, we sought to reconstitute this pathway in vitro to confirm our biosynthetic hypothesis and identify the preferred amino acid and acyl-CoA building blocks used by the NRPS

assembly line. The individual modules of the RupA NRPS were expressed and purified in *Escherichia coli* as C-$His_6$-tagged constructs ($RupA_{C1\text{-}A1\text{-}T1}$ and $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$) (Fig. S6). We then used a set of standard biochemical assays to verify the activities of the two NRPS modules and determine the substrate specificities of their A domains. The ATP-[$^{32}$P]$PP_i$ exchange assay[31] was used to assess amino acid activation by each individual module of RupA. These experiments revealed that $RupA_{C1\text{-}A1\text{-}T1}$ preferentially activates L-leucine but can also accept L-valine, while $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ preferentially activates L-glutamate over L-aspartate (Figs. 3A, S7 and S8). We then used the promiscuous phosphopantetheinyl transferase Sfp to load BODIPY-CoA onto the T domain of each module, verifying that this enzyme can posttranslationally modify the RupA NRPS[32] (Fig. S9). Finally, T domain loading assays[31] with $^{13}$C-labeled amino acids confirmed that both L-leucine and L-valine were tethered onto $RupA_{C1\text{-}A1\text{-}T1}$ and that both L-glutamate and L-aspartate could be loaded onto $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ (Figs. S10 and S11).

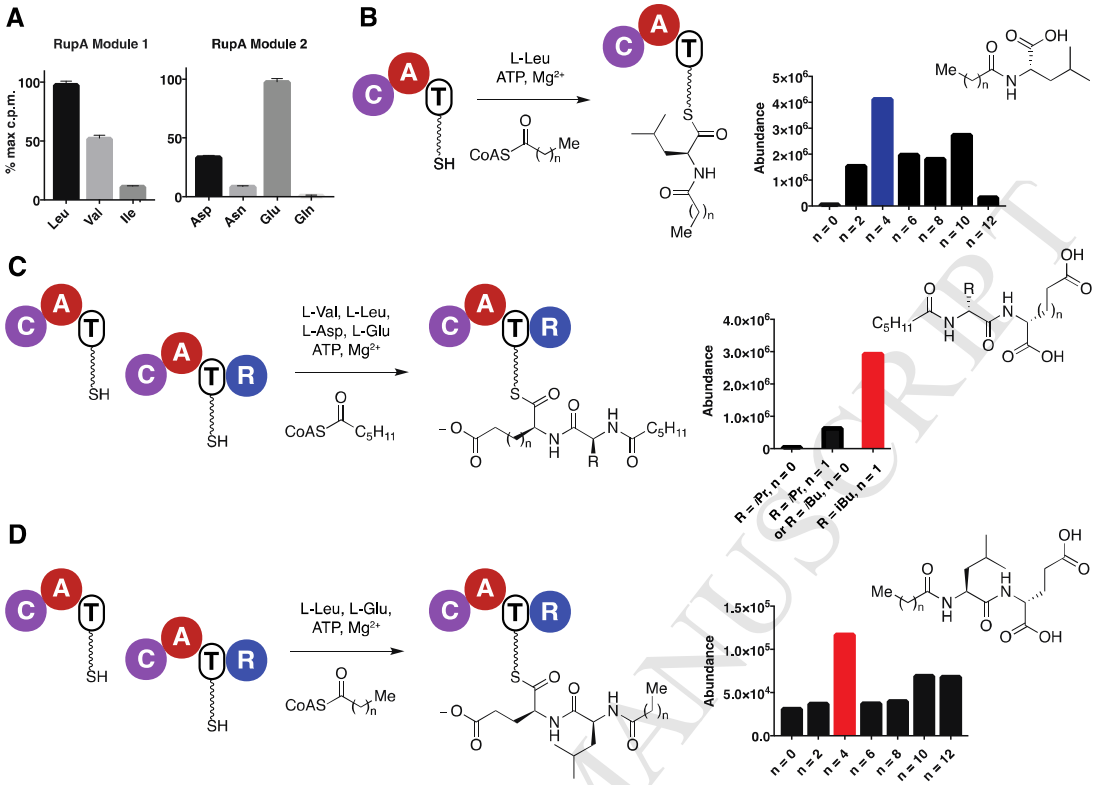



**Fig. 3**: RupA preferentially uses hexanoyl-CoA, L-leucine, and L-glutamate to produce a T-domain tethered *N*-acylated dipeptide thioester intermediate. (A) ATP-[$^{32}$P]PP$_i$ exchange assay for individual RupA modules. (B) LC-MS assay for C-starter domain specificity. Mass abundances (extracted ion chromatogram intensities) are shown for hydrolyzed products obtained from the reaction of RupA$_{C1\text{-}A1\text{-}T1}$, L-leucine, and equimolar amounts of $C_2$−$C_{14}$ fatty acyl-CoA substrates. The mass abundance of *N*-hexanoyl-L-leucine is highlighted in blue. Representative results are shown from at least two independent experiments. (C, D) LC-MS assay for tethered *N*-acylated dipeptide synthesis by RupA. Mass abundances (extracted ion chromatogram intensities) are shown for *N*-acylated dipeptide products. The mass abundance of *N*-hexanoyl-L-leucyl-L-glutamic acid in each experiment is highlighted in red. Representative results are shown from at least two independent experiments. (C) The assay mixture contained RupA$_{C1\text{-}A1\text{-}T1}$, RupA$_{C2\text{-}A2\text{-}T2\text{-}R}$, ATP, hexanoyl-CoA, and equimolar amounts of L-valine, L-leucine, L-aspartate, and L-glutamate (amino acid competition format). (D) The assay mixture contained RupA$_{C1\text{-}A1\text{-}T1}$, RupA$_{C2\text{-}A2\text{-}T2\text{-}R}$,

ATP, L-leucine, L-glutamate, and equimolar amounts of $C_2$−$C_{14}$ fatty acyl-CoA substrates (acyl-CoA competition format).

With the amino acid building blocks established, we next set about identifying the specific acyl-CoA(s) that would be recognized and incorporated on the *N*-terminus of ruminopeptin. Though fatty acids can be incorporated into nascent polyketides and non-ribosomal peptides by several mechanisms,[27,33,34] we predicted that the C-starter domain of the RupA NRPS would *N*-acylate L-leucine using a freely diffusible fatty acyl-CoA co-substrate. In order to determine the preferred fatty acyl-CoAs, we reconstituted the activity of $RupA_{C1\text{-}A1\text{-}T1}$. We incubated $RupA_{C1\text{-}A1\text{-}T1}$ with L-leucine, ATP, and a set of short, medium and long even-chain acyl-CoAs ($C_2$ to $C_{14}$) in a competition assay format. We then hydrolyzed the resulting *N*-acylated aminoacyl thioester intermediates from the NRPS for detection using LC-MS (Figs. 3B and S12, Table S3).[25,27] Using this assay we identified *N*-hexanoyl-L-leucine as the most abundant product, indicating that hexanoyl-CoA and other medium-chain acyl-CoA's are likely preferred substrates of the RupA C-starter domain.

We subsequently modified this assay to probe the activity of both NRPS modules. As we were unable to successfully express full-length RupA, we instead included the individually expressed and purified NRPS modules ($RupA_{C1\text{-}A1\text{-}T1}$ and $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$) in the reaction mixture along with amino acids, acyl-CoA substrates, and ATP. As the NAD(P)H cofactor required for the reductase domain was not provided in these initial attempts, we predicted this assay should generate T-domain-tethered *N*-acylated dipeptide thioesters which could be hydrolyzed from the enzyme and detected by LC-MS. We first

performed a competition experiment to identify the preferred amino acid building blocks, including in the reaction mixture multiple amino acids (L-valine, L-leucine, L-aspartate, and L-glutamate) along with a single fatty acyl-CoA substrate (hexanoyl-CoA) (Figs. 3C and S13, Table S4). The preferred product generated in this experiment incorporated L-leucine and L-glutamate. To verify the identity of the preferred acyl-CoA substrate, we next performed this assay using the preferred amino acid substrates (L-leucine and L-glutamate) and a mixture of even-chain acyl-CoA's (Fig. 3D, Table S5). In this experiment, *N*-hexanoyl-L-leucyl-L-glutamic acid was the most abundant product. From these results, we concluded that RupA can produce a range of nascent T-domain tethered *N*-acylated dipeptide thioester intermediates, and may preferentially use hexanoyl-CoA, L-leucine, and L-glutamate building blocks.

Finally, we sought to completely reconstitute the RupA NRPS in vitro to access putative peptide aldehyde products. To accomplish this, we included NAD(P)H, the cofactor required for R domain activity, in reaction mixtures along with ATP and the preferred substrates hexanoyl-CoA, L-leucine, and L-glutamate. We analyzed the supernatants of reaction mixtures by LC-MS and attempted to detect the expected masses of peptide aldehydes, primary alcohols, truncated products, or molecules that could arise from degradation of the predicted structures. However, after extensive optimization we were unable to detect any putative final products in this experiment. We were also unable to identify final products in the presence of any other combinations of building blocks that we had previously examined. We did observe formation of the hydrolysis products of tethered *N*-acylated dipeptide thioester intermediates, indicating that the NRPS modules

were functional (data not shown). We also confirmed that synthetic standards of the predicted peptide aldehyde products (see section 2.4 for synthesis) could be detected under these assay conditions (data not shown).

Suspecting that our RupA$_{C2\text{-}A2\text{-}T2\text{-}R}$ construct may have purified with an inactive R domain, we individually expressed and purified two additional constructs (RupA$_R$ single domain and RupA$_{T2\text{-}R}$ di-domain) and evaluated their reactivity toward a synthetic *N*-acetylcysteamine substrate **5** that mimics the preferred RupA$_{T2}$-tethered intermediate (Fig. 4A). Monitoring consumption of NAD(P)H by the change in absorbance at 340 nm ($A_{340}$), we could detect activity of neither RupA$_R$, RupA$_{T2\text{-}R}$ nor the full module RupA$_{C2\text{-}A2\text{-}T2\text{-}R}$ toward the synthetic substrate, suggesting that we have not successfully purified an active form of RupA$_R$ (Fig. 4B).

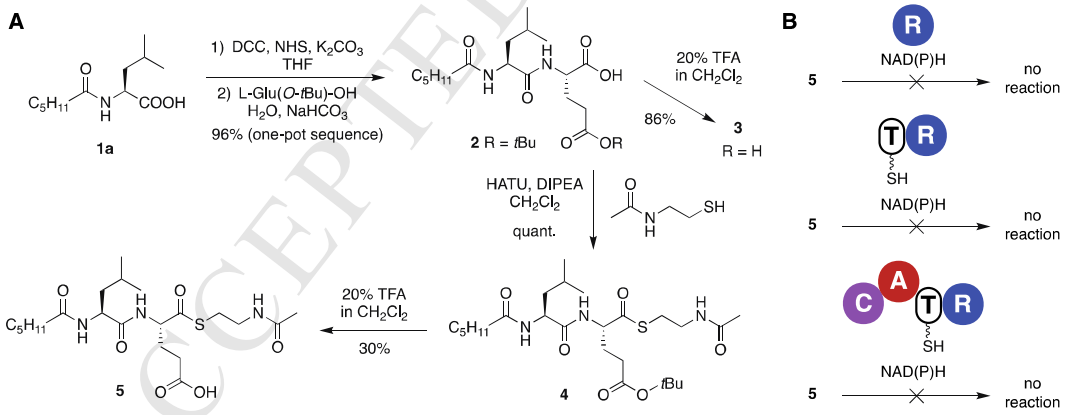



**Fig. 4**: The RupA reductase domain was inactive toward substrate mimic *N*-hexanoyl-L-Leu-L-Glu-SNAc **5** in vitro. (A) Synthesis of an *N*-acetylcysteamine (SNAc) substrate for RupA$_R$ (DCC = *N*,*N*′-dicyclohexylcarbodiimide, NHS = *N*-hydroxysuccinimide). (B) By measuring decrease in

absorbance at 340 nm, no consumption of NAD(P)H was observed when reacting **5** with purified $RupA_R$, $RupA_{T2\text{-}R}$, or $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$.

Though we were unable to reconstitute the activity of the RupA R domain in vitro, bioinformatic analyses suggest this domain should be active in vivo. Though $RupA_R$ shows only low amino acid sequence identity to other biochemically characterized R domains (e.g., 23.5% with $AusA_R$,[35] 18.7% with $MxcG_R$,[36] and 24.6% with the R domain from bgc35, which was previously reconstituted in vitro[6]), we were able to identify the conserved catalytic triad and NAD(P)H binding motifs (Fig. S4). Among the diverse superfamily of short-chain dehydrogenases/reductases, which includes NRPS terminal R domains, proteins with sequences identities as low as 15-30% are reported to share similar three dimensional folds.[37] We generated a homology model of $RupA_R$ with the known two-electron reducing terminal R domain from AusA, using HHPred and MODELLER. This model suggests that the $RupA_R$ motif for NAD(P)H binding and catalytic triad for reduction chemistry are properly oriented (Fig. S14). Therefore, we propose that this R domain is likely active in vivo and involved in producing the final product of the *rup* gene cluster.

In summary, our efforts to reconstitute the activity of the *R. bromii* NRPS RupA strongly suggest that the peptide aldehyde *N*-hexanoyl-L-Leu-L-Glu-H is a likely product of this enzymatic assembly line. This work does not rule out the possibility that RupA may produce additional metabolites in vivo. We observed some promiscuity in the *N*-acylation

activity of the C-starter domain and A domain specificities of RupA. However, this type of promiscuity is often observed for NRPS enzymes reconstituted in vitro, and the preferred products observed in this format typically correspond to the most abundant natural analogues, even in cases where multiple products can also be isolated from cultures.[35,38] It is also possible that there are unusual biosynthetic substrates available to *R. bromii*, particularly acyl-CoA's with unusual acyl-chain modifications, which we did not provide in our in vitro reconstitution experiments. Though RupA may produce additional molecules with different scaffolds, it is reasonable to propose that *N*-hexanoyl-L-Leu-L-Glu-H could be one major biosynthetic product.

Additionally, as we have never directly observed activity of the RupA R domain, we could not determine if this assembly line produces an aldehyde or an alcohol. In previous biosynthetic reconstitutions of NRPS terminal reductase domains, the aldehyde intermediate has not been detected in significant quantities when the final expected product is the peptide alcohol.[39,40] Though the activity of aldehyde-producing NRPS terminal reductase domains has been reconstituted in several cases,[35,41] the natural products generated by these pathways are cyclic imines or pyrazinones, so only trace amounts of free aldehyde intermediates were detected in these experiments. Given the difficulties encountered in resolving this biosynthetic step, we decided to move forward to examine the biological activity of the putative peptide aldehydes we predict could be produced by RupA.

**2.4 Synthesis and biological activity evaluation of ruminopeptin analogues**

We next used the information that we obtained from our bioinformatic and biochemical analyses of the *rup* biosynthetic pathway to inform the chemical synthesis of a focused library of predicted ruminopeptin structures. We designed 12 analogues of the predicted *N*-acyl dipeptide aldehyde scaffold that varied in the *N*-acyl substituent and amino acid components. We then accessed these compounds using a solution-phase synthetic route adapted from previous syntheses of aspartyl and glutamyl peptide aldehydes.[42,43] Peptide synthesis has previously been used as a tool to access natural product peptide aldehydes when isolation efforts yielded insufficient quantities of pure material for activity screening.[22] In our case, we envisioned that accessing a small library could not only provide compounds for assays but also enable structure-activity relationship studies.

From *N*-Cbz- and *O*-*t*Bu-protected L-glutamate and L-aspartate precursors, we accessed key intermediates containing an aldehyde masked as a semicarbazone functional group using the previously reported reaction sequences (**6a**-**b**, Fig. 5).[42,43] The resulting semicarbazone-protected intermediates were then coupled to *N*-acylated L-leucine and L-valine derivatives (**1a**-**i**, Fig. 5) using the peptide coupling reagent HATU to yield **7a**-**l** (Table 1). From these coupling products, deprotection of the *O*-*t*Bu ester proceeded with 20% trifluoroacetic acid in dichloromethane. Finally, transfer of the semicarbazide functional group to formaldehyde under acidic conditions and corresponding regeneration of the aldehyde provided the desired peptide aldehyde products (**8a**-**l**) (Table 2). Using this route, we accessed a small library of ruminopeptin analogues on a multi-milligram scale (6-42% overall yield, 7-36 mg obtained).

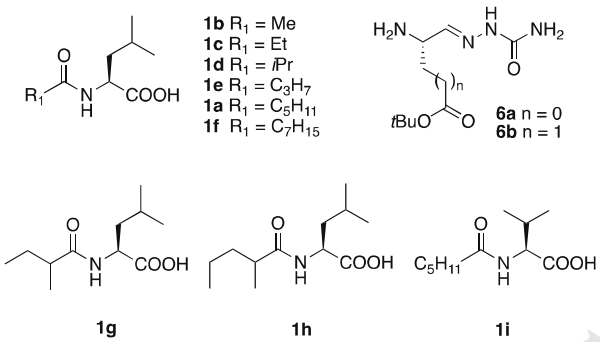
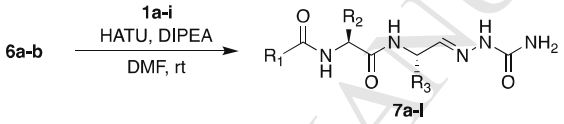
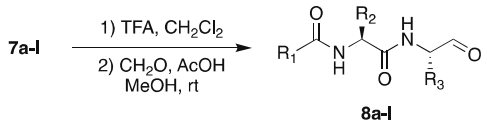



**Fig. 5:** Synthetic precursors used in this study.




| Entry | Product | $R_1$ | $R_2$ | $R_3$ | %Yield |
|---|---|---|---|---|---|
| 1 | **7a** | Me | Leu | Asp (O-*t*Bu) | 31% |
| 2 | **7b** | Me | Leu | Glu (O-*t*Bu) | 28% |
| 3 | **7c** | $C_2H_5$ | Leu | Glu (O-*t*Bu) | 65% |
| 4 | **7d** | $C_3H_7$ | Leu | Glu (O-*t*Bu) | 63% |
| 5 | **7e** | *i*Bu | Leu | Glu (O-*t*Bu) | 48% |
| 6 | **7f** |  | Leu | Glu (O-*t*Bu) | 62% |
| 7 | **7g** |  | Leu | Glu (O-*t*Bu) | 87% |
| 8 | **7h** | $C_5H_{11}$ | Leu | Glu (O-*t*Bu) | 88% |
| 9 | **7i** | $C_5H_{11}$ | Val | Glu (O-*t*Bu) | 67% |
| 10 | **7j** | $C_5H_{11}$ | Leu | Asp (O-*t*Bu) | 39% |
| 11 | **7k** | $C_5H_{11}$ | Val | Asp (O-*t*Bu) | 88% |
| 12 | **7l** | $C_7H_{15}$ | Leu | Glu (O-*t*Bu) | 43% |

**Table 1:** Coupling reaction between *O*-*t*Bu-protected semicarbazones **6a**-**b** and *N*-acyl amino acids **1a**-**i** to give semicarbazone intermediates **7a**-**l**.




| Entry | Product | $R_1$ | $R_2$ | $R_3$ | %Yield |
|---|---|---|---|---|---|
| 1 | **8a** | Me | Leu | Asp | 66% |
| 2 | **8b** | Me | Leu | Glu | 60% |
| 3 | **8c** | $C_2H_5$ | Leu | Glu | 64% |
| 4 | **8d** | $C_3H_7$ | Leu | Glu | 28% |
| 5 | **8e** | *i*Bu | Leu | Glu | 55% |
| 6 | **8f** |  | Leu | Glu | 48% |
| 7 | **8g** |  | Leu | Glu | 16% |
| 8 | **8h** | $C_5H_{11}$ | Leu | Glu | 34% |
| 9 | **8i** | $C_5H_{11}$ | Val | Glu | 23% |
| 10 | **8j** | $C_5H_{11}$ | Leu | Asp | 24% |
| 11 | **8k** | $C_5H_{11}$ | Val | Asp | 17% |
| 12 | **8l** | $C_7H_{15}$ | Leu | Glu | 15% |

**Table 2**: Removal of *O*-*t*Bu protecting groups from **7a-l** and exchange of semicarbazones with formaldehyde to afford the desired ruminopeptin analogues **8a-l**.

With access to sufficient quantities of peptide aldehydes **8a-l**, we could begin to identify potential target(s) of these molecules. Our biosynthetic reconstitution experiments strongly suggest that ruminopeptin contains a glutamate residue in its P1 position. We thus gained insights into potential targets by comparing the predicted structures of the ruminopeptins to the known substrate specificities of secreted microbial serine and cysteine proteases, as well as host proteases. As specific post-glutamyl hydrolyzing activity is rare among microbial proteases and unknown among human proteases, this analysis revealed only one promising candidate: the glutamyl endopeptidases. Glutamyl endopeptidases are a class of secreted serine proteases found in several bacterial species, including the human pathogens *E. faecalis*[44] (SprE) and *S. aureus*[45] (endoproteinase GluC/SspA/V8 protease). These proteases are regularly found in quorum-sensing

regulated operons alongside additional proteases, and glutamyl endopeptidases appear to regulate the action of these enzymes (a metalloprotease in *E. faecalis* and a cysteine protease in *S. aureus*). SspA and SprE are thought to play roles in biofilm formation in these pathogens, though the precise details of their involvement vary substantially among different strains and assay systems.[46,47] Additionally, results from several animal models implicate SspA and SprE in virulence.[8,48–50]

We screened our library of ruminopeptin analogues (**4a-l**) for their ability to inhibit the activity of SspA in vitro. In this assay, protease was pre-incubated with the peptide aldehyde for 10 min. Protease activity was then quantified by measuring the increase in fluorescence corresponding to the release of the AMC fluorophore from the fluorogenic peptide substrate Z-Leu-Leu-Glu-AMC. We found that several of the synthetic compounds inhibit SspA, with approximately 50% inhibition observed for the most potent compounds, medium-chain acyl analogues **8h** and **8l**, at 10 μM (Figs. 6A and S15). Intriguingly, in vitro reconstitution experiments suggested these compounds are also among the mostly likely products generated by the *rup* gene cluster. We observed reduced inhibition with the short-chain acyl analogues **8b-d** and insignificant inhibitory activity with the branched acyl chain analogues **8e-g** or aspartyl analogues **8a**, **8j** and **8k**.

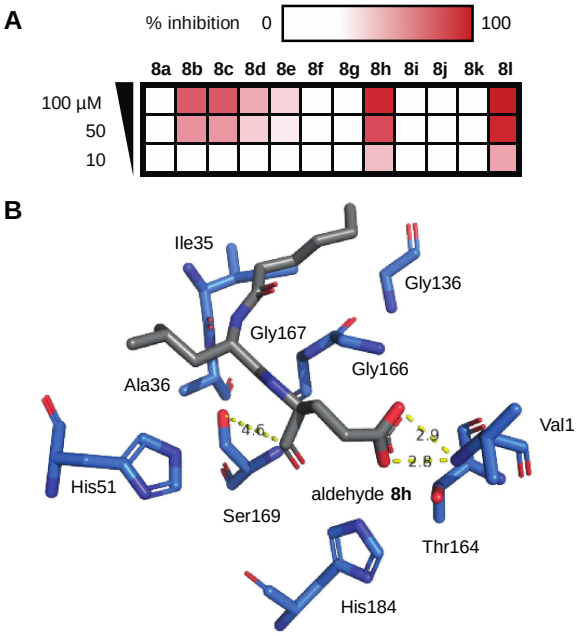



**Fig. 6**: Ruminopeptin-like compounds inhibit SspA from *S. aureus*. A) Inhibition profile of predicted ruminopeptins against SspA. The assays were conducted by pre-incubating 1 ng/µL SspA with inhibitor for 10 min at room temperature followed by addition of fluorogenic peptide substrate Z-Leu-Leu-Glu-AMC to a final concentration of 75 µM. Fluorescence (367 nm excitation/460 nm emission) was then monitored for 20 min at 30 ºC and inhibitor efficiency calculated by comparing the slope of the linear portion of the curve with the negative control (no inhibitor). Reactions were performed in duplicate and inhibitor efficiency is reported as a mean of both trials. B) Potential interaction between peptide aldehyde **8h** (grey) and SspA (blue). The structure was docked into the crystal structure of SspA (PDB: 1qy6) using the induced fit docking algorithm in Glide.

To better understand the interaction between **8h** and SspA, we performed a docking experiment using Glide.[51] Substrate recognition by SspA is reported to rely on an electrostatic interaction between the negatively charged glutamate side chain in position

P1 of the peptide substrate and the positively charged *N*-terminal amine of SspA.[52] We observed a similar interaction between this *N*-terminal amine and the glutamyl side chain of **8h** when we docked this inhibitor into the crystal structure of SspA (PDB: 1qy6). The electrophilic aldehyde warhead of **8h** was also located within reasonable proximity (4.6 Å) to the nucleophilic Ser residue (Fig. 6B), suggesting that this inhibitor binds the protease similarly to a model protein substrate and an orientation that would facilitate formation of a reversible, covalent hemiacetal linkage.

Though synthetic inhibitors of SspA have previously been reported,[53,54] this work provides the first indication of endogenous inhibition of glutamyl endopeptidases by microbial natural products and is also the first report of peptide aldehyde inhibitors of this enzyme class. By surveying the structure-activity relationships (SAR) in the synthetic library of closely related family members, we revealed that the presence of both L-leucine and L-glutamate are important for SspA inhibition.

**2.5 Identification of glutamyl endopeptidase homologs in the human gut microbiota**

The observation that putative peptide aldehydes derived from a gut commensal could inhibit a bacterial glutamyl endopeptidase made us curious about the relevance of these proteases within the human gut. Though *S. aureus* is more commonly associated with the nasal microbiota and can be detected from nasal swabs of approximately 40% of healthy individuals, studies have consistently detected this bacterium in the stool microbiomes of approximately 20% of healthy individuals.[55] Indeed, intestinal carriage of *S. aureus* is

hypothesized to contribute to bacterial dissemination in the environment.[55] In comparison, the opportunistic pathogen *E. faecalis* can be detected in 47% of fecal samples from healthy individuals,[56] and its glutamyl endopeptidase SprE resembles SspA (49% amino acid similarity and 27% identity).[57] Therefore, the homologous *E. faecalis* glutamyl endopeptidase represents an additional possible target that may be more relevant within the habitat of *R. bromii*.

It is also possible that peptide aldehydes produced by *R. bromii* interact with related proteases found in other gut commensal microbes. To our knowledge, the presence and roles of glutamyl endopeptidases in the human gut microbiota has not yet been investigated. Glutamyl endopeptidases have been discovered in *Staphylococcus*, *Bacillus*, and *Streptomyces* species, and many have been biochemically characterized, including SspA, SprE, glutamyl peptidase BL (from *Bacillus licheniformis*),[58] glutamyl peptidase BS (from *Bacillus subtilis*),[59] glutamyl peptidase BI (from *Bacillus intermedius*),[60] and glutamyl endopeptidase II (from *Streptomyces griseus*).[61] These enzymes all belong to the structural chymotrypsin family,[62] and though they exhibit some differences in kinetic parameters and specificity, they all share a preference for cleavage after glutamyl residues and would therefore likely be inhibited by a glutamyl aldehyde. The diversity of these biochemically characterized examples of glutamyl endopeptidases provided a broad starting point for identifying additional potential targets of ruminopeptin in the human gut.

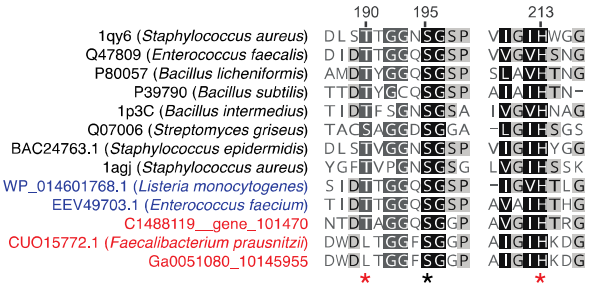



**Fig. 7**: Glutamyl endopeptidase homologs are found in gut microbial genomes and human gut metagenomes. ClustalW2 alignment of biochemically characterized glutamyl endopeptidases (black), homologs from sequenced organisms (blue), and homologs from human gut microbes and gut metagenomes (red). Included are the sequences of characterized glutamyl endopeptidases from *S. aureus* (1qy6),[52] *E. faecalis* (Q47809),[57] *B. licheniformis* (P80057),[59] *B. subtilis* (P39790),[63] *B. intermedius* (1p3c),[64] *S. griseus* (Q07006),[61] *Staphylococcus epidermidis* (BAC24763.1),[65] and epidermolytic toxin A from *S. aureus* (1agj),[66] with additional predicted glutamyl endopeptidases from *L. monocytogenes* (WP_014601768.1), *E. faecium* (EEV49703.1), and *F. prausnitzii* (CUO15772.1). Metagenomic sequences were retrieved using the BLAST tool at JGI Integrated Microbial Genomes & Microbiome Samples.[67] Catalytic Ser195 is indicated with a black asterisk. Positions 190 and 213, which may be involved in conferring substrate specificity, are indicated with red asterisks.

To explore whether additional glutamyl endopeptidases are present in the human gut microbiota, we used BLAST searches to locate members of this family in sequenced gut microbial genomes. Queries of the non-redundant (nr) protein database of NCBI with six representative glutamyl endopeptidase sequences identified putative hits in other prominent residents of the human gut (e-value cutoff 9 $e^{-12}$). Conserved residues Thr190 (or Ser190) and His213 (chymotrypsin numbering) in the S1 binding pocket of crystallized glutamyl endopeptidases have been identified as important for binding

glutamate-containing substrates (Thr164 and His184 in SspA).[62] We were able to identify these residues in the BLAST hits from *Enterococcus faecium* (24.5% ID to SprE) and the pathogen *Listeria monocytogenes* (25.5% ID to glutamyl endopeptidase BL) (Fig. 7). Additionally, we identified a sequence from *F. prausnitzii* that is annotated as a glutamyl endopeptidase precursor. This sequence maintains His213, but not Thr190, in the S1 site. It remains to be determined if these putative glutamyl endopeptidases from prominent gut commensals and human pathogens actually exhibit post-glutamyl hydrolyzing activity. Overall, these proteins may not only represent ecologically relevant targets of the ruminopeptins but also provide a promising starting point for investigating the biological roles of microbial proteases in the human gut.

In order to assess the presence of these proteases in human subjects and determine the distribution of glutamyl endopeptidases among unsequenced members of the gut microbiota, we also performed a BLAST search of representative glutamyl endopeptidase sequences against assembled stool metagenomes available through the JGI (268 metagenomes). After limiting the results based on an e-value cutoff (2e-10), length (188-400 residues to account for the diversity among characterized members of this protease family), and the presence of a candidate histidine residue in the S1 binding pocket, we identified 52 glutamyl endopeptidase homologs in 51 different samples. 47 of these sequences have ≥99% amino acid sequence ID to the putative glutamyl endopeptidase from *F. prausnitzii*. The remaining sequences do not map to sequenced genomes. This analysis suggests that these putative targets of the ruminopeptins may be present in the human gut.

### 2.6 Potential implications of gut microbial peptide aldehyde production

A multitude of human and microbial proteases are present in the human gut environment. While host-derived digestive proteases are not a major contributor to protease activity in the colon, other human proteases are involved in gut barrier maintenance and facilitating migration of immune cells within the mucosal layer.[68] Host proteases are also involved in regulating the immune system, and during inflammation human immune cells secrete proteases which are responsible for degrading extracellular tissues.[69] Gut protease activity is upregulated in ulcerative colitis[70,71] and Crohn's disease,[69] and is also involved in the pathogenesis of colorectal cancer.[72] Host derived proteases are therefore potential therapeutic targets, with the protease inhibitor camostat mesilate previously explored as a treatment for Crohn's disease.[73]

Though much work has focused on the activity of human proteases in gastrointestinal (GI) disease, microbial proteases are also active in this environment. Bacterial proteases in the healthy gut are involved in metabolism and nutrient acquisition,[74,75] but they may also disrupt mucosal barriers,[76] interact with protease-activated receptors,[77] or modulate the host immune response.[75] These enzymes are also hypothesized to have roles in biofilm formation and modification of the microbial or intestinal surface.[75] Additionally, several microbial secreted proteases are implicated as virulence factors in the gut context.[76] Therefore, inhibition of microbial proteases has recently attracted interest as a therapeutic strategy for GI disease.

Our work has uncovered the first evidence that gut microbial natural products may be capable of modulating microbial protease activity. We hypothesize that the peptide aldehyde(s) produced by the *rup* gene cluster likely target a microbial protease found in this environment. Aside from the restriction of glutamyl endopeptidase activity to bacteria, the biogeography of *R. bromii* in the human gut and the potential instability of the ruminopeptins also suggest these natural products likely have a microbial target. In a study of gut microbes associated with insoluble, undigested polysaccharide particles in fecal samples, *R. bromii* was one of the three most enriched species in this phase as opposed to the soluble phase.[78] This observation may indicate that *R. bromii* is located distantly from host cells in comparison to other gut species. Moreover, a potential explanation for our inability to identify putative *rup* gene cluster products in *R. bromii* cultures could the instability of these peptide aldehydes. Indeed, incubation of peptide aldehyde **8h** in an *R. bromii* culture resulted in almost complete degradation in just 15 min when incubated at 37 ºC (data not shown). Overall, given their instability and the localization of *R. bromii*, we hypothesize that peptide aldehydes produced by this organism have evolved to target other microbial species living in close proximity.

As highlighted earlier, the *rup* gene cluster is part of a larger family of NRPS biosynthetic gene clusters found in human gut bacterial genomes and metagenomes. In recent work, Fischbach and coworkers accessed the putative products of several of these gene NRPS gene clusters using a distinct workflow.[6] Relying principally on heterologous expression of these gene clusters in *E. coli* and *B. subtilis*, they were able to identify

primarily cyclic pyrazinones and dihydropyrazinone compounds, along with one *N*-acylated peptide aldehyde (*N*-octanoyl-Met-Phe-H, a product of bgc33 from *Clostridium* sp. CAG:567). This result indicates that the R domains of these NRPS assembly lines can produce aldehyde products. Interestingly, their heterologous expression strategy was not universally effective, as products could be identified for only 7 of the 14 gene clusters investigated.[6] As evidence that these products were not simply artifacts of heterologous expression, they also isolated one cyclic compound from its native producing organism and reconstituted of the biosynthesis of a cyclic pyrazinone *in vitro*.

Fischbach and coworkers hypothesized that the cyclic compounds observed in their study were derived from linear dipeptide aldehyde precursors and that these dipeptide aldehydes may be the relevant, bioactive metabolites in vivo. Therefore, they synthesized several dipeptide aldehydes and evaluated their inhibitory activity against human proteases. They found that free-amino dipeptide aldehydes had potent activity against cysteine protease cathepsins (cathepsins B, L, C and S), which are found in the host lysosome, as well as calpain. Using chemoproteomics to measure the global interactions of a representative dipeptide aldehyde (L-Phe-L-Phe-H) with the human proteome, they concluded that the cathepsins are likely principal targets of this compound. They suggest that inhibition of these lysosomal proteases may disrupt immune recognition of members of the commensal microbiota, as lysosomal cathepsins are involved in antigen processing and presentation.

Though this exciting study illustrates the potential of gut microbial natural products to interact with host targets, some prominent questions remain. It is currently unknown whether the peptide aldehydes investigated by Fischbach and co-workers have alternative targets in the gut microbiota, as their interactions with microbial proteases were not explored. The predominant cyclic dihydropyrazinones and pyrazinones products isolated from heterologous expression and vitro reconstitution experiments were not screened for biological activity, so the potential roles of these molecules are unclear. Finally, the metabolites observed in this work were not completely consistent with the biosynthetic machinery present in the corresponding gene clusters. The cyclic dihydropyrazinones and pyrazinones observed in this study should be derived from dipeptide aldehyde precursors. However, in most cases, bioinformatic analysis of the NRPS gene clusters reveals assembly lines containing either C-starter domains, which should produce *N*-acylated dipeptide aldehydes, or additional upstream loading modules, which should synthesize tripeptide aldehydes. Given that we have observed an active C-starter domain in RupA, this discrepancy suggests that the identities of the metabolites generated by the gut organisms harboring these gene clusters may still be unclear.

It is also important to note that Fischbach and co-workers attempted to heterologously express the *rup* gene cluster (bgc45) in *B. subtilis* but did not observe any product formation. Moreover, heterologous expression was also unsuccessful for the two other NRPS gene clusters most closely related to the *rup* pathway (bgc41 and bgc43). This finding illustrates the continued challenges associated with identifying gut microbial

natural products and highlights how isolation-independent strategies can provide information about natural products that are recalcitrant to other characterization methods.

## 3. Conclusion

In summary, we have used bioinformatics, in vitro biochemistry, and chemical synthesis to access the putative products of the *rup* gene cluster. We demonstrated that the ruminopeptins inhibit a bacterial protease implicated in virulence in several human pathogens and that homologs of this enzyme are also present in commensal gut organisms. The ecological details of the ruminopeptin-glutamyl endopeptidase interaction remain to be determined, as do the broader roles of gut microbial protease inhibitors and gut microbial proteases. Our studies of how ruminopeptin-like protease inhibitors affect the human gut microbiota are currently underway.

## 4. Experimental

### 4.1. General materials and methods

Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA) and Sigma Aldrich (Billerica, MA). Recombinant plasmid DNA was purified with the Qiaprep Kit from Qiagen (Germantown, MD) and the E.Z.N.A. Plasmid Mini Kit from OMEGA kit from Omega Bio-Tek (Norcross, GA). Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and Gel Band Purification Kit from GE Healthcare. DNA sequencing was performed by Beckman Coulter Genomics (Danvers, MA), Genewiz (Cambridge, MA), and Eton Bioscience (Boston, MA). Restriction enzymes were purchased from New

England BioLabs (Ipswich, MA). Nickel-nitrilotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen. SDS-PAGE gels were purchased from BioRad. Protein concentrations were determined by quantifying protein A280 using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific) or by the Bradford assay. Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm.

All chemicals were obtained from Sigma-Aldrich except where noted. Protected amino acids were obtained from Chem-Impex (Dale, IL) and Advanced ChemTech (Louisville, KY). HATU was purchased from Oakwood Chemical (Estill, SC). All NMR solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). NMR spectra were visualized using iNMR version 5.5.7. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as internal standard for $^{1}H$ ($CDCl_3$ = 7.26 ppm, DMSO-$d_6$ = 2.50 ppm, $D_2O$ = 4.79 ppm) and $^{13}C$ ($CDCl_3$ = 77.25 ppm, DMSO-$d_6$ = 39.52 ppm). Data are reported as follows: chemical shift, integration multiplicity (s = singlet, br s = broad singlet, d = doublet, t = triplet, m = multiplet, q = quartet, qt = quintet), coupling constant, and integration.

High-resolution mass spectral data was obtained in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. Enzyme assays were analyzed on a Bruker Impact II qTOF mass spectrometer in negative ion mode coupled to an Agilent 1290 uHPLC. Each LC-MS run was internally calibrated using sodium formate introduced at the end of the run. For liquid chromatography, 5 µL of sample was injected

onto a Phenomenex Kinetex C18 column (100Å pore size, 150 mm x 2.1 mm, 2.6 µm particle size). Mobile phase A was 0.1% formic acid (v/v) in water, and mobile phase B was 0.1% formic acid (v/v) in acetonitrile. The mobile phase composition started at 1% B, which was maintained for 2 min. Samples were then subjected to a linear gradient over 8 min to 100% B. Flow of 100% B was maintained for 4 min, and the column was then re-equilibrated to 1% B over 1.9 min. High-resolution mass spectral (HRMS) data for the synthetic compounds was obtained on an Agilent 6210 TOF. The capillary voltage was set to 4.5 kV and the end plate offset to −500 V, the drying gas temperature was maintained at 190 °C, with a flow rate of 8 l/min and a nebulizer pressure of 21.8 p.s.i. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% $H_2O$ and 50% acetonitrile as solvent.

### 4.2. Cultivation of bacterial strains

*R. bromii* strains were cultivated using several different growth media: M2GSC (which is supplemented with 30% rumen fluid)[79] and RUM media,[80] which were prepared as previously described with the following modifications: supplementary heat-sensitive vitamins were prepared as a 1000x aqueous stock (except for D-pantetheine, which was prepared as a 100x aqueous stock) and separately filtered and sparged with nitrogen to render anaerobic. Carbohydrates were also prepared as 100x aqueous stocks and treated with the same procedure. The media itself was boiled, sparged with nitrogen, dispensed in Hungate tubes under anaerobic conditions, and then autoclaved. Supplementary vitamins and carbohydrates were then added to individual aliquots of the growth media at the time of inoculation.

A lyophilized stock of *R. bromii* ATCC 27255 was purchased from the American Type Culture Collection, Manassas, VA. *R. bromii* L2-63 was provided as a glycerol stock by Harry Flint and coworkers (University of Aberdeen). *R. bromii* 22-5-S 6 FAA NB was provided as a glycerol stock by Emma Allen-Vercoe and coworkers (University of Guelph).

*R. bromii* L2-63, *R. bromii* ATCC 27255, and *R. bromii* 22-5-S 6 FAA NB were inoculated from frozen glycerol stocks as 5 mL cultures in RUM media with fructose and allowed to grow in a 10% hydrogen/10% carbon dioxide/bal. nitrogen atmosphere for approximately 24 h until they reached saturation. These cultures were then passaged as 1:100 dilutions and allowed to reach saturation again before extraction of genomic DNA. Genomic DNA was extracted using the standard protocol of the UltraClean Microbial DNA Isolation Kit form MO BIO (Carlsbad, CA). To confirm strain identities, primers fD1 and rP2[81] were used to amplify and sequence 16S rRNA sequences.

### 4.3. Detection of the *rup* gene cluster by PCR.

PCR reactions for amplification of the *rup* gene cluster from each of the *R. bromii* strains were accomplished using Phusion PCR mix (ThermoFisher). Reactions were performed according to the manufacturer's instructions and contained 0.1 µL template DNA, 10 µM each of forward and reverse primers, half final volume of the 2x master mix, and water to total 25 µL. The reaction mixtures were annealed at 65 ºC. Initially, the gene cluster was

detected by using primers repDetect1 and rupDetect2, designed to amplify the RupA $A_1$ domain from strain L2-63 (Table S6). Subsequently, the remainder of the gene cluster was sequenced by PCR-amplifying overlapping regions of the cluster, using primers designed for strain L2-63, and then assembling the resulting reads using the Geneious 9 assembler. The primers used for sequencing are indicated in Table S6.

### 4.4. RT-PCR for detection of *rup* gene cluster expression.

For detection of *rup* gene cluster expression in *R. bromii* strains, the organism was grown as described above and subjected to various culture conditions (Fig. S5). Two mL of each culture was then mixed with an equal amount of bacterial RNAProtect solution (Qiagen) and processed according to the manufacturer's instructions. Protected cell pellets prepared in this way were stored at –80 ºC for up to 48 h before downstream processing. Total RNA was obtained with the TRIzol reagent using previously published procedures.[82] After extraction, RNA was air dried overnight and then digested with RNAse-free DNAse (Invitrogen) according to the manufacturer's procedure. RT-PCR was conducted using the SuperScript III one-step RT-PCR system with Platinum *Taq* DNA polymerase (ThermoFisher) according to the manufacturer's instructions. Reaction mixtures contained 0.3 µL template RNA, 10 µM each of forward and reverse primers, half final volume of the 2x master mix, 1 µL of SuperScript III RT/Platinum *Taq* enzyme mix, and water to total 25 µL. The reactions were annealed at 63 ºC. Cluster expression was detected using primers rupDetect-1 and rupDetect-2 (Fig. S5). Bands were identified by imaging on a Gel Doc™ EZ Gel Documentation System (BioRad).

### 4.5. Cloning, overexpression and purification of $RupA_{C1\text{-}A1\text{-}T1}$, $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$, $RupA_{T1}$, $RupA_{R}$, and $RupA_{T2\text{-}R}$

Protein expression constructs were PCR amplified from *R. bromii* L2-63 genomic DNA using the primers shown in Table S6. PCR amplification was performed using Phusion PCR mix (ThermoFisher). Reactions were performed according to the manufacturer's instructions and contained 0.1 µL template DNA, 10 µM each of forward and reverse primers, half final volume of the 2x master mix, and water to total 25 µL or 50 µL. Reaction mixtures were divided in 12.5 µL portions and annealed along a gradient from 50 ºC to 70 ºC, with all reaction mixtures showing a band by diagnostic PCR pooled and purified.

Restriction digests were conducted according to the manufacturer's instructions, with the enzymes indicated in Table S6, and were purified directly using agarose gel electrophoresis. Gel fragments were further purified using the Illustra GFX PCR DNA and Gel Band Purification Kit. The digests were ligated into linearized expression vectors using T4 DNA ligase (New England Biolabs). $RupA_{C1\text{-}A1\text{-}T1}$, $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$, and $RupA_{T2\text{-}R}$ were ligated into the pET-29b vector to encode a C-terminal $His_6$-tagged construct. $RupA_{T1}$ and $RupA_{R}$ were ligated into the pET29a vector to encode a N-terminal $His_6$-tagged construct. Ligations were incubated at room temperature for 3 h and contained 3 µL of water, 1 µL of T4 Ligase Buffer (10x), 1 µL of digested vector, 3 µL of digested insert DNA, and 2 µL of T4 DNA Ligase (400 U/ µL). 10 µL of each ligation was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The

identities of the resulting constructs were confirmed by sequencing of purified plasmid DNA.

For protein expression, the vectors containing $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$, $RupA_{T1}$, $RupA_{R}$, $RupA_{T2\text{-}R}$ were transformed into chemically competent *E. coli* BL21 (DE3) cells. The vector containing $RupA_{C1\text{-}A1\text{-}T1}$ was co-transformed into *E. coli* BL21 GOLD (Agilent Technologies) with the addition of chaperone plasmid pGro7 (Takara Bio USA, Mountain View, CA). Cell stocks were stored at –80 °C in LB/glycerol.

General procedure for protein large scale overexpression and purification: A 50 mL starter culture of BL21 or BL21+pGro7 *E. coli* was inoculated from a single colony and grown overnight at 37 ºC in LB medium supplemented with 50 µg/ml kanamycin (and 20 µg/mL chloramphenicol for BL21 + pGro7). Overnight cultures were diluted 1:100 into 2 L of LB medium containing 50 µg/mL kanamycin (and 20 µg/mL chloramphenicol for BL21+pGro7). Cultures were incubated at 37 ºC with shaking at 175 rpm, moved to 15 ºC at OD600 = 0.2-0.3, induced with 500 µM IPTG at OD600 = 0.5-0.6, and incubated at 15 ºC for 19 h. Cells from 2 L of culture were harvested by centrifugation (4,000 rpm x 10 min) and resuspended in 35 mL of lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, pH 7.5, supplemented with 1 mM DTT for purification of $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 10,000 psi, and the lysate was clarified by centrifugation (10,800 rpm x 30 min). The supernatant was supplemented with 1 M imidazole for a final concentration of 5 mM imidazole, treated with 20 µL DNAse I, and passed over 4 mL of Ni-NTA resin (pre-

washed with 3 x 10 mL lysis buffer). The resin-bound protein was washed with 25 mL of 25 mM imidazole elution buffer. Protein was eluted from the column using a stepwise imidazole gradient in elution buffer (50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 200 mM), collecting 2 mL fractions. SDS–PAGE analysis (4–15% Tris- HCl gel) was used to determine which fractions contained the desired protein. Fractions were combined and dialyzed twice against 2 L of storage buffer (20 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 10% (v/v) glycerol, pH 7.5, supplemented with 1 mM DTT for purification of $RupA_{CAT2R}$). Solutions containing protein were frozen in liquid nitrogen and stored at –80 ºC. This procedure afforded yields of 8.6 mg/L for $RupA_{C1-A1-T1}$, 1.7 mg/L for $RupA_{C2-A2-T2-R}$, 3.5 mg/L for $RupA_{T1}$, 2.4 mg/L $RupA_R$, and 7.5 mg/L $RupA_{T2-R}$. The purified proteins are visualized on an SDS-PAGE gel in Fig. S6.

### 4.6. Biochemical characterization of RupA

#### 4.6.1. ATP-[$^{32}$P]PP$_i$ exchange assay

The reaction mixture (100 µL) contained 75 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM amino acid substrate, and 4 mM $Na_4PP_i$/[$^{32}$P]$PP_i$ (stock 1:1500 dilution prepared from Phosporous-32 radionuclide, PerkinElmer, ~6 mCi/mL, in 40 mM $Na_4PP_i$). Reaction mixtures were initiated by the addition of $RupA_{C1-A1-T1}$ or $RupA_{C2-A2-T2-R}$ (1 µM) and incubated at room temperature for 30 min. Reactions were quenched by the addition of 200 µL of charcoal suspension (16 g/L activated charcoal, 100 mM $Na_4PP_i$, 3.5 % (v/v) $HClO_4$). The samples were centrifuged (13,000 rpm x 3 min), and the supernatant was removed. The charcoal pellet was washed two times with 200 µL of wash buffer (100 mM Na4$PP_i$, 3.5 % (v/v) $HClO_4$). The pellet was resuspended in 200

µL of wash buffer and added to 10 mL of scintillation fluid (Ultima Gold, Perkin Elmer). Radioactivity was measured on a Beckman LS 6500 scintillation counter. Full data with negative controls is presented in Figs. S7 and S8.

#### 4.6.2. BODIPY-CoA fluorescent phosphopantetheinylation assay

BODIPY-CoA[32] and Sfp[83] were prepared using previously reported procedures. The reaction mixture (50 µL) contained 5 µM of either $Rup_{C1\text{-}A1\text{-}T1}$ or $Rup_{C2\text{-}A2\text{-}T2\text{-}R}$, 1.0 µM Sfp, 5 µM BODIPY-CoA, 10 mM $MgCl_2$, 25 mM Tris pH 8.5, and 50 mM NaCl. Reaction mixtures were incubated for 1 h in the dark at room temperature and then diluted 1:1 in 2x Laemmli sample buffer (Bio-Rad), boiled for 10 min, and separated by SDS-PAGE (4-15% Tris-HCl gel). The gel was first imaged at λ=365 nm, then stained with Bio-Safe Coomassie Stain (BioRad) and imaged again.

#### 4.6.3. T-domain loading assays with $^{14}$C-labeled amino acids

Reaction mixtures (50 µL) contained 25 mM Tris pH 8.5, 50 mM NaCl, 10 mM $MgCl_2$, 250 µM CoA tri-lithium salt, 500 µM DTT, 30 µM of the indicated amino acid, 3 µM of either $Rup_{C1\text{-}A1\text{-}T1}$ or $Rup_{C2\text{-}A2\text{-}T2\text{-}R}$. For the assay with $Rup_{C1\text{-}A1\text{-}T1}$, the reaction mixture was supplemented with 30 µM $RupA_{T1}$ to amplify signal. Amino acids used were $^{14}$C-L-Leu (0.1 mCi/mL, 328 mCi/mmol), $^{14}$C-L-Val (0.1 mCi/mL, 246 mCi/mmol), $^{14}$C-L-Glu (0.1 mCi/mL, 260 mCi/mmol), and $^{14}$C-L-Asp (0.1 mCi/mL, 201 mCi/mmol). Loading of the phosphopantetheinyl arm onto the T domains of $RupA_{C1\text{-}A1\text{-}T1}$ (and $RupA_{T1}$) or $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ was initiated by the addition of Sfp (1 µM) to the reaction mixture, followed by incubation at room temperature for 1 h. Loading of the T domain with amino

acid was then initiated by the addition of ATP (3 mM). After incubation at room temperature for 1 h, the reaction was quenched by the addition of 100 µL of bovine serum albumin (1 mg/mL) followed by 500 µL of trichloroacetic acid (TCA) (10% (w/v) aqueous solution). The protein was pelleted by centrifugation (10,000 rpm x 8 min). After removal of the supernatant, the protein pellet was washed two times with 250 µL of TCA (10% w/v aqueous solution). The pellet was resuspended in 200 µL of formic acid and added to 10 mL of scintillation fluid (Ultima Gold, Perkin Elmer). Radioactivity was measured on a Beckman LS 6500 scintillation counter.

#### 4.6.4. LC-MS assays for C domain substrate specificity and *N*-acyl dipeptide production

For the assay with the first module $RupA_{C1\text{-}A1\text{-}T1}$ (**Fig. 3B**), the reaction mixture (50 µL) contained 25 mM Tris buffer pH 8.5, 50 mM NaCl, 10 mM $MgCl_2$, 400 µM DTT, 4 mM L-Leu, 250 µM CoA-tri-lithium salt, 6.6 % (v/v) DMSO, and $RupA_{C1\text{-}A1\text{-}T1}$ (10 µM). Loading of the phosphopantetheinyl arm onto the T domain of $RupA_{C1\text{-}A1\text{-}T1}$ was initiated by the addition of Sfp (3 µM) to the reaction mixture, followed by incubation at room temperature for 1 h. ATP (5 mM) was then added to the reaction mixture, and the C domain loading reaction was initiated by the addition of the fatty acyl-CoA substrate (1 mM). For the fatty acyl-CoA competition experiment, a stock solution containing all the fatty acyl-CoA substrates was added, each to a final concentration of 142 µM. This mixture was incubated at room temperature for 2 h and quenched by the addition of methanol (125 µL). After incubation on ice for 10 min, the samples were centrifuged (13,000 rpm x 10 min). The protein pellets were washed two times with 125 µL of

methanol and dried under a stream of nitrogen gas. Products bound to the T domain were hydrolyzed by the addition of 0.1 M KOH (5 µL) followed by heating at 74 °C for 10 min. The samples were cooled on ice, and 0.1 M HCl (25 µL) was added to the solutions. Finally, methanol (60 µL) was added to the samples, which were then incubated at –80 °C for at least 2 h to precipitate protein. The samples were centrifuged (13,000 rpm x 15 min) and the supernatant was analyzed by LC-MS. Masses were not observed in reactions without ATP, without enzyme, or reactions containing boiled enzyme. Full data is presented in Table S3.

For assays including both modules $RupA_{C1\text{-}A1\text{-}T1}$ and $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ (**Figs. 3C** and **3D**), the reaction mixture (50 µL) contained 25 mM Tris buffer pH 8.5, 50 mM NaCl, 10 mM $MgCl_2$, 400 µM DTT, 250 µM CoA-tri-lithium salt, 6.6 % (v/v) DMSO, $RupA_{C1\text{-}A1\text{-}T1}$ (10 µM), and $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ (10 µM). The amino acid competition experiment contained 4 mM each of L-valine, L-leucine, L-aspartate, and L-glutamate, and the fatty acyl-CoA competition experiment contained 4 mM each of L-leucine and L-glutamate. Loading of the phosphopantetheinyl arm onto the T domains of $RupA_{C1\text{-}A1\text{-}T1}$ and $RupA_{C2\text{-}A2\text{-}T2\text{-}R}$ was initiated by the addition of Sfp (3 µM) to the reaction mixture, followed by incubation at room temperature for 1 h. ATP (5 mM) was then added to the reaction mixture, and the C domain loading reaction was initiated by the addition of the fatty acyl-CoA substrate (1 mM). For the fatty acyl-CoA competition experiment, a stock solution containing all the fatty acyl-CoA substrates was added, each to a final concentration of 142 µM. This mixture was incubated at room temperature for 19 h, and an identical workup procedure was followed. Product masses were not observed in reactions without ATP, without

enzyme, or reactions containing boiled enzyme. Full data is presented in Tables S4 and S5.

### 4.7. SspA inhibition assays

For SspA inhibition assays, the reaction mixture (50 µL) contained 50 mM Tris-HCl pH 8, 1 ng/µL endoproteinase GluC (Worthington Biochemical Corporation, Lakewood, NJ), and 75 µM Z-LLE-AMC (Ubiquitin-Proteasome Biotechnologies, Aurora, CO). The assays were conducted in half-area white microplates (Enzo Life Sciences). To set up the reaction mixtures, assay buffer was added to each well, followed by inhibitor compounds and then SspA. The protease was incubated with inhibitor compounds at room temperature for 10 min in order to allow for protease/inhibitor interaction. Substrate was then added and the reaction mixtures were monitored for fluorescence (367 nm excitation/460 nm emission, PMT medium, plate read height 0.81 mm) in a Spectramax i3 Plate Reader once per minute for 20 min at 30 ºC. Reactions were performed in duplicate and inhibitor efficiency is calculated as a mean of both trials. The positive control inhibitor Ac-Glu$^{P}$(OPh)$_2$ was used to validate the assay.[54] Numerical data is presented in Fig. S15.

### 4.8. Chemical synthesis procedures and characterization data

#### 4.8.1. Synthesis of *N*-acyl amino acids:

*N*-acetyl-L-Leucine (**1b**) was purchased from Chem-Impex. Other *N*-acyl amino acids were synthesized using the Schotten-Bauman reaction of acyl chlorides with amino acids in aqueous base. The amino acid (1.0 equiv) was dissolved in 15% aqueous NaOH (0.5

M) and cooled to 0 ºC. The acid chloride was added dropwise and the reaction mixture stirred overnight, allowing to warm to room temperature. 20% aqueous HCl was added to pH = 2 and the resulting solution was extracted with dichloromethane (three portions of 3x reaction volume). The combined organic layers were washed with saturated aqueous sodium chloride (one portion of 1x reaction volume). The solution was then dried over $Na_2SO_4$, filtered, and concentrated in vacuo. The characterization data for compounds **1b**–**1e** and **1g** matched previously reported results.[84,85]

4.8.1.1. Octanoyl-L-leucine (**1f**):

Colorless solid (701 mg, 54.5%) (m.p. 121–123 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 10.63 (br s, 1H), 6.12 (d, *J* = 8.1 Hz, 1H), 4.62 (m, 1H), 2.24 (t, *J* = 7.6 Hz, 2H), 1.71 (m, 2H), 1.63 (m, 2H), 1.60 (m, 1H), 1.28 (m, 8H), 0.94 (m**,** 6H), 0.87 (t, *J* = 7.0 Hz, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 176.8, 174.4, 51.1, 41.5, 36.7, 31.9, 29.4, 29.2, 25.9, 25.1, 23.1, 22.8, 21.1, 14.3. HRMS (ESI): Calc'd for formula $C_{14}H_{26}NO_3^-$ $[M–H]^-$ 256.1918, found 256.1927.

4.8.1.2. (2-Methylpentanoyl)-L-leucine (**1h**):

Colorless oil (710 mg, 62%). $^1$H NMR (500 MHz; $CDCl_3$): δ 10.69 (br s, 1H), 6.10 (t, *J* = 9.1 Hz, 1H), 4.63 (m, 1H), 2.27 (q, *J* = 7.0 Hz, 1H), 1.70 (m, 2H), 1.61 (m, 2H), 1.36–1.29 (m, 4H), 1.16 (m, 1H), 1.13 (m, 3H), 0.94 (t, *J* = 8.2 Hz, 6H), 0.89 (m, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 177.6, 177.3, 50.9, 41.4, 36.5, 35.9, 25.2, 23.0, 22.1, 20.7, 17.9, 14.2. HRMS (ESI): Calc'd for formula $C_{12}H_{22}NO_3^-$ $[M–H]^-$ 228.1605, found 228.1614.

4.8.1.3. Hexanoyl-L-valine (**1i**)

Colorless solid (2.0 g, 93%) (m.p. 121–123 ºC). $^{1}$H NMR (500 MHz; $CDCl_3$): δ 11.37 (s, 1H), 6.31 (d, *J* = 8.7 Hz, 1H), 4.59 (dd, *J* = 8.7, 4.7 Hz, 1H), 2.26 (t, *J* = 7.6 Hz, 2H), 2.22 (m, 1H), 1.63 (qt, *J* = 7.4 Hz, 2H), 1.30 (m, 4H), 0.95 (m, 6H), 0.87 (t, *J* = 6.87, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 175.5, 174.6, 57.2, 36.8, 31.5, 25.7, 22.5, 19.2, 17.9 14.1. HRMS (ESI): Calc'd for formula $C_{11}H_{20}NO_3^-$ $[M–H]^-$ 214.1499, found 214.1453.

#### 4.8.2. Synthesis of enzymatic assay standards

4.8.2.1. (*S*)-5-(*tert*-Butoxy)-2-((*S*)-2-hexanamido-4-methylpentanamido)-5-oxopentanoic acid (**2**):

To an oven-dried flask containing *N*-hexanoyl-L-Leucine (200 mg, 0.872 mmol, 1.00 equiv), was added DCC (1.01 equiv), NHS (1.01 equiv), anhydrous potassium carbonate (1.00 equiv) and anhydrous THF (5 mL). The reaction mixture was stirred at room temperature for 3 h. The mixture was filtered through glass wool into a suspension of L-Glu(*O*-*t*Bu)-OH (177 mg, 0.872 mmol, 1.00 equiv) in 10% aqueous sodium bicarbonate (20 mL). The glass wool was washed with THF (3 x 5 mL). The reaction mixture was stirred for an additional 2 h. The mixture was neutralized with 5% aqueous citric acid to pH = 7 and extracted with ethyl acetate (3 x 25 mL). The combined organic layers were washed with water (25 mL) and brine (25 mL), dried over $Na_2SO_4$, filtered, and concentrated in vacuo. The crude product was purified by flash chromatography on silica gel using $CHCl_3$/MeOH/AcOH (93:5:2) to yield **2** (362 mg, 96%) as a colorless oil. $^{1}$H NMR (500 MHz; $CDCl_3$): δ 7.34 (d, *J* = 5.9 Hz, 1H), 6.63 (m, 1H), 4.60 (t, *J* = 6.7 Hz, 1H), 4.50 (q, *J* = 6.2 Hz, 1H), 2.33 (m, 2H), 2.21 (m, 4H), 2.00 (m, 2H), 1.61 (m, 3H), 1.43 (s, 9H), 1.29 (m, 4H), 0.89 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 175.9, 174.5,

173.4, 172.4, 129.2, 125.4, 81.1, 52.0, 41.4, 36.4, 31.4, 25.5, 24.8, 22.9, 22.3, 21.0, 14.1. HRMS (ESI): Calc'd for formula $C_{21}H_{37}N_2O_6^-$ $[M–H]^-$ 413.2657, found 413.2674.

4.8.2.2. *N*-Hexanoyl-L-leucyl-L-glutamic acid (**3**):

A solution of **2** in 20% trifluoroacetic acid in dichloromethane (4.28 mL) with 5 µL $H_2O$ was stirred for 1 h at room temperature. The reaction mixture was concentrated in vacuo, and residual TFA was removed by forming the azeotrope with anhydrous toluene. The resulting crude product was purified by flash chromatography on silica gel using $CHCl_3$/MeOH/AcOH (88:10:2) to afford **3** (28 mg, 83%) as a colorless oil. $^1$H NMR (500 MHz; 10% $CD_3OD$ in $CDCl_3$, referenced to $CDCl_3$): δ 7.56 (d, *J* = 7.9 Hz, 1H), 6.87 (d, *J* = 8.7 Hz, 1H), 4.54 (m, 2H), 2.39 (m, 2H), 2.20 (m, 2H), 1.59 (m, 4H), 1.51 (m, 1H), 1.24 (m, 6H), 0.88 (m, 9H). $^{13}$C NMR (126 MHz; 10% $CD_3OD$ in $CDCl_3$, referenced to $CDCl_3$): δ 176.1, 174.5, 173.0, 129.2, 128.4, 41.4, 36.5, 31.5, 30.2, 27.0, 25.5, 24.9, 23.0, 22.5, 22.2, 20.8, 14.1. HRMS (ESI): Calc'd for formula $C_{17}H_{29}N_2O_6^-$ $[M–H]^-$ 357.2031, found 357.2053.

4.8.2.3. *tert*-Butyl (*S*)-5-((2-acetamidoethyl)thio)-4-((*R*)-2-hexanamido-4-methylpentanamido)-5-oxopentanoate (**4**):

To a solution of **2** (362 mg, 0.873 mmol) in dry dichloromethane (8.7 mL) were added HATU (1.5 equiv), DIPEA (3.0 equiv), and *N*-(2-mercaptoethyl)acetamide (1.2 equiv) under argon. The reaction mixture was stirred for 18 h at room temperature and then quenched with 10 mL saturated aqueous $NaHCO_3$. The aqueous layer was extracted with dichloromethane (3 x 10 mL). The combined organic layers were washed with 10 mL brine, dried over $MgSO_4$, filtered, and concentrated in vacuo. The crude product was purified by flash chromatography on silica gel using $CH_2Cl_2$/MeOH (9:1) to yield **4** (466

mg, quant.) as a colorless oil. $^{1}$H NMR (500 MHz; $CDCl_3$): δ 4.50 (m, 1H) 3.71 (m, 2H), 3.45 (m, 1H), 3.17 (m, 2H), 3.17 (m, 1H), 3.02 (m, 4H), 2.90 (br s, 3H), 2.33 (m, 2H), 2.12 (m, 2H), 1.64 (m, 2H), 1.49 (m, 6H), 1.44 (s, 9H), 1.31 (m, 2H), 0.92 (m, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 196.7, 173.2, 172.6, 161.9, 81.3, 55.5, 45.1, 43.7, 40.2, 39.0, 36.5, 31.6, 30.9, 29.3, 28.3, 26.8, 25.6, 25.0, 23.2, 22.6, 18.8, 17.3, 14.1, 12.0. HRMS (ESI): Calc'd for formula $C_{25}H_{44}N_3O_6S^-$ $[M–H]^-$ 514.2956, found 514.2974.

4.8.2.4. *N*-hexanoyl-L-Leu-L-Glu-SNAc (**5**):

A solution of **4** in 20% trifluoroacetic acid in dichloromethane (20 mL) with 5 µL $H_2O$ was stirred for 30 min at 0 ºC. The reaction mixture was concentrated in vacuo, and residual TFA was removed by forming the azeotrope with anhydrous toluene. The crude material was purified using a 15.5 g RediSep Rf Gold C18Aq column on a Combiflash Rf Teledyne ISCO Purification System (mobile phase A: 0.1% TFA in water, mobile phase B:0.1% TFA in acetonitrile) to yield **5** (31 mg, 30%) as a colorless oil.

$^{1}$H NMR (500 MHz; DMSO-$d_6$): δ 8.53 (s, 1H), 8.03 (m, 1H), 7.93 (m, 1H), 4.36 (m, 2H), 3.11 (m, 2H), 2.85 (m, 2H), 2.27 (m, 2H), 2.10 (m, 2H), 1.78 (s, 3H), 1.48 (m, 4H), 1.24 (m, 3H), 0.86 (m, 9H). $^{13}$C NMR (126 MHz; DMSO-$d_6$): δ 200.7, 173.7, 172.9, 172.2, 58.3, 50.8, 39.5, 38.1, 35.1, 30.8, 29.6, 27.6, 26.3, 26.2, 25.0, 24.2, 23.1, 22.9, 22.5, 21.9, 21.5, 13.9. HRMS (ESI): Calc'd for formula $C_{21}H_{36}N_3O_6S^-$ $[M–H]^-$ 459.2403, found 459.2414.

**4.8.3. Coupling of *N*-acyl amino acids to semicarbazone-protected aldehydes:**

To a solution of semicarbazone-protected intermediate **6a** or **6b** (1.0 equiv) and acyl L-leucine **1a–i** (1.0 equiv) in DMF (0.6 M) was added HATU (1.1 equiv) and DIPEA (5.1 equiv) with stirring, under argon. After 3 h, the reaction mixture was diluted with ethyl acetate (to 10 x initial volume) and quenched by addition of 1M aqueous NaOH (10 x initial volume). The organic layer was collected and the aqueous layer extracted with three portions of ethyl acetate (each 10 x initial reaction volume). The combined organic layers were washed with water and brine (each 20 x initial reaction volume), dried over $Na_2SO_4$, filtered, and concentrated in vacuo using a Genevac EZ-2 Elite centrifugal evaporator. Products were purified by flash chromatography on silica gel using $CH_2Cl_2$/MeOH (9:1).

4.8.3.1. *tert*-Butyl (*S*,*E*)-3-((*S*)-2-acetamido-4-methylpentanamido)-4-(2-carbamoylhydrazineylidene)butanoate (**7a**):

The product (68 mg, 31%) was isolated as a colorless solid (m.p. 160–162 ºC). $^1$H NMR (500 MHz; DMSO-$d_6$): δ 10.00 (s, 1H), 8.18 (d, *J* = 8.2 Hz, 1H), 7.99 (d, *J* = 8.2 Hz, 1H), 7.08 (d, *J* = 3.2 Hz, 1H), 6.25 (br s, 2H), 4.69 (m, 1H), 4.27 (q, *J* = 7.7 Hz, 1H), 2.65 (m, 1H), 2.49 (m, 1H), 1.82 (s, 3H), 1.56 (m, 2H), 1.40 (m, 2H), 1.37 (m, 9H), 0.86 (m, 6H). $^{13}$C NMR (126 MHz, DMSO-$d_6$): δ 171.9, 169.6, 169.1, 156.6, 140.3, 80.1, 51.0, 47.1, 41.1, 38.0, 27.6, 24.2, 22.9, 22.5, 21.7. HRMS (ESI): Calc'd for formula $C_{17}H_{31}N_5O_5Na^+$ $[M+Na]^+$ 408.2217, found 408.2226.

4.8.3.2. *tert*-Butyl (*S*,*E*)-4-((*S*)-2-acetamido-4-methylpentanamido)-5-(2-carbamoylhydrazineylidene)pentanoate (**7b**):

The product (132 mg, 28%) was isolated as a colorless solid (m.p. 159–162 ºC). $^{1}$H NMR (500 MHz; DMSO-$d_6$): δ 9.90 (s, 1H), 8.04 (d, $J$ = 8.3 Hz, 1H), 7.96 (d, $J$ = 8.1 Hz, 1H), 7.04 (d, $J$ = 4.0 Hz, 1H), 6.25 (br s, 2H), 4.36 (m, 1H), 4.27 (m, 1H), 3.30 (s, 3H), 2.48 (m, 2H), 2.17 (m, 2H), 1.82 (s, 3H), 1.66 (m, 2H), 1.55 (m, 2H), 1.39 (m, 1H), 1.37 (s, 9H), 0.84 (m, 6H). $^{13}$C NMR (126 MHz, DMSO-$d_6$): δ 171.9, 171.8, 169.1, 156.7, 141.3, 79.6, 51.1, 48.9, 41.0, 30.8, 27.7, 27.4, 24.2, 22.9, 22.5, 21.7. HRMS (ESI): Calc'd for formula $C_{18}H_{33}N_5O_5Na^+$ $[M+Na]^+$ 422.2374, found 422.2383.

4.8.3.3. *tert*-Butyl (*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((*S*)-4-methyl-2-propionamidopentanamido)pentanoate (**7c**):

The product (131 mg, 65%) was isolated as a colorless solid (m.p. 124–126 ºC). $^{1}$H NMR (500 MHz; $CDCl_3$): δ 9.69 (m, 1H), 7.78 (m, 1H), 6.57 (m, 1H), 5.45 (br s, 2H), 4.55 (m, 1H), 4.50 (m, 1H), 2.25 (m, 4H), 2.04 (m, 1H), 1.85 (m, 1H), 1.64 (m, 2H), 1.53 (m, 1H), 1.41 (m, 9H), 1.12 (m, 3H), 0.90 (m, 6H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 175.0, 172.7, 158.3, 80.9, 52.3, 50.2, 40.9, 31.3, 29.6, 28.3, 27.7, 25.1, 23.1, 22.9, 22.2, 10.0. HRMS (ESI): Calc'd for formula $C_{19}H_{34}N_5O_5^-$ $[M–H]^-$ 412.2565, found 412.2588.

4.8.3.4. *tert*-Butyl (*S*,*E*)-4-((*S*)-2-butyramido-4-methylpentanamido)-5-(2-carbamoylhydrazineylidene)pentanoate (**7d**):

The product (160 mg, 63%) was isolated as a colorless solid (m.p. 122–124 ºC). $^{1}$H NMR (500 MHz; $CDCl_3$): δ 9.68 (s, 1H), 7.78 (d, $J$ = 7.0 Hz, 1H), 6.57 (br s, 1H), 4.54 (m, 1H), 4.50 (t, $J$ = 7.0 Hz, 1H), 2.24 (m, 4H), 2.04 (m, 2H), 1.84 (m, 2H), 1.64 (m, 2H), 1.54 (m, 1H), 1.41 (m, 9H), 0.95 (m, 6H), 0.87 (m, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 174.3, 172.7, 170.6, 158.3, 80.9, 52.2, 50.2, 40.8, 38.5, 31.3, 28.3, 27.8, 25.0, 23.1, 22.2,

19.3, 13.9. HRMS (ESI): Calc'd for formula $C_{20}H_{37}N_5O_5Na^+$ $[M+Na]^+$ 450.2687, found 450.2675.

4.8.3.5. *tert*-Butyl (*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((*S*)-2-isobutyramido-4-methylpentanamido)pentanoate (**7e**):

The product (124 mg, 48%) was isolated as a colorless oil. $^1$H NMR (500 MHz; $CDCl_3$): δ 9.71 (m, 1H), 7.81 (m, 1H), 7.21 (s, 2H), 6.47 (m, 1H), 4.56 (m, 1H), 4.51 (m, 1H), 2.43 (m, 1H), 2.21 (m, 2H), 2.06 (m, 2H), 1.84 (m, 2H), 1.48 (m, 1H), 1.43 (m, 4H), 1.40 (s, 9H), 1.13 (m, 6H), 0.92 (d, $J$ = 6.4 Hz, 3H), 0.88 (d, $J$ = 6.3 Hz, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 178.2, 172.65, 158.3, 80.8, 54.7, 53.7, 51.9, 51.0, 50.1, 40.6, 35.5, 31.1, 28.3, 25.1, 23.1, 22.2, 19.9, 19.6, 18.8, 17.6. HRMS (ESI): Calc'd for formula $C_{20}H_{37}N_5O_5Na^+$ $[M+Na]^+$ 450.2687, found 450.2712.

4.8.3.6. *tert*-Butyl (4*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((2*S*)-4-methyl-2-(2-methylbutanamido)pentanamido)pentanoate (**7f**):

The product (155 mg, 62%) was isolated as a colorless solid (m.p. 125–126 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.76 (s, 1H), 7.90 (m, 1H), 7.2 (m, 1H), 6.60 (m, 1H), 4.56 (m, 2H), 2.32 (m, 1H), 2.20 (m, 2H), 2.03 (m, 2H), 1.82 (m, 2H), 1.62 (m, 2H), 1.42 (m, 1H), 1.39 (m, 9H), 1.10 (m, 3H), 0.91 (m, 3H), 0.87 (m, 6H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 177.6, 172.6, 158.3, 142.2, 128.7, 80.8, 55.0, 51.93, 51.89, 42.9, 40.5, 31.2, 28.4, 27.5, 25.1, 23.1, 22.1, 18.8, 17.81, 17.66, 12.1. HRMS (ESI): Calc'd for formula $C_{21}H_{39}N_5O_5Na^+$ $[M+Na]^+$ 464.2843, found 464.2845.

4.8.3.7. *tert*-Butyl (4*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((2*R*)-4-methyl-2-(2-methylpentanamido)pentanamido)pentanoate (**7g**):

The product (217 mg, 87%) was isolated as a colorless solid (m.p. 124–126 ºC). $^{1}$H NMR (500 MHz; $CDCl_3$): δ 9.76 (s, 1H), 7.84 (t, *J* = 7.9 Hz, 1H), 7.21 (s, 1H), 6.51 (br s, 1H), 4.55 (m, 2H), 2.28 (m, 1H) 2.22 (m, 2H), 2.05 (m, 2H), 1.83 (m, 2H), 1.47 (m, 1H), 1.40 (m, 9H), 1.28 (m, 4H), 1.09 (m, 3H), 0.93 (m, 3H), 0.88 (m, 6H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 177.9, 172.6, 158.3, 80.8, 77.5, 77.2, 77.0, 54.9, 51.9, 41.2, 40.4, 36.6, 31.19, 31.07, 25.0, 23.1, 22.16, 22.07, 20.8, 18.8, 18.17, 18.01, 17.5, 14.2. HRMS (ESI): Calc'd for formula $C_{22}H_{41}N_5O_5Na^+$ $[M+Na]^+$ 478.3000, found 478.3016.

4.8.3.8. *tert*-Butyl (*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((*S*)-2-hexanamido-4-methylpentanamido)pentanoate (**7h**):

The product (189 mg, 88%) was isolated as a colorless oil. $^{1}$H NMR (399 MHz; $CDCl_3$): δ 9.94 (s, 1H), 7.94 (d, *J* = 6.8 Hz, 1H), 7.23 (s, 1H), 6.84 (d, *J* = 8.1 Hz, 1H), 4.54 (br s, 2H), 2.23–2.14 (m, 4H), 2.00 (m, 2H), 1.80 (m, 2H) 1.58 (m, 5H), 1.39 (s, 9H), 1.27 (m, 4H), 0.89 (m, 3H), 0.85 (s, 6H). $^{13}$C NMR (100 MHz; $CDCl_3$): δ 174.2, 172.78, 172.58, 158.4, 142.2, 80.7, 52.0, 50.1, 46.5, 41.0, 36.5, 31.3, 28.3, 28.0, 25.56, 25.50, 25.0, 23.1, 22.5, 22.2, 14.1, 8.9. HRMS (ESI): Calc'd for formula $C_{22}H_{42}N_5O_5^+$ $[M+H]^+$ 456.318, found 456.3197.

4.8.3.9. *tert*-Butyl (*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((*S*)-2-hexanamido-3-methylbutanamido)pentanoate (**7i**):

The product (204 mg, 67%) was isolated as a yellow glass. $^{1}$H NMR (500 MHz; $CDCl_3$): δ 9.76 (m, 1H), 7.73 (m, 1H), 7.33 (s, 1H), 7.18 (m, 2H), 6.23 (d, *J* = 6.3 Hz, 2H), 4.54 (m, 1H), 3.69 (m, 1H), 3.20–3.13 (m, 1H), 2.23 (m, 2H), 2.06 (m, 2H), 1.86 (m, 2H), 1.63 (m, 4H), 1.41 (m, 9H), 1.30 (m, 4H), 0.94 (m, 6H), 0.87 (m, 3H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 175.2, 174.0, 128.8, 57.35, 57.33, 55.2, 43.3, 36.8, 31.59, 31.54, 28.2, 25.66,

25.62, 22.57, 22.53, 19.3, 18.8, 18.0, 17.4, 14.1, 12.6. HRMS (ESI): Calc'd for formula $C_{21}H_{40}N_5O_5^+$ $[M+H]^+$ 442.3024, found 442.3041.

4.8.3.10. *tert*-Butyl (*S*,*E*)-4-(2-carbamoylhydrazineylidene)-3-((*S*)-2-hexanamido-4-methylpentanamido)butanoate (**7j**):

The product (104 mg, 39%) was isolated as a colorless solid (m.p. 106–108 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.66 (s, 1H), 7.76 (d, *J* = 7.6 Hz, 1H), 6.58 (d, *J* = 8.1 Hz, 1H), 4.85 (m, *J* = 2.8 Hz, 1H), 4.49 (m, 1H), 2.70 (m, 1H), 2.58 (m, 1H), 2.20 (m, 3H), 1.60 (m, 2H), 1.40 (m, 9H), 1.28 (m, 4H), 0.88 (m, 9H). $^{13}$C NMR (125 MHz, $CDCl_3$): δ 174.2, 172.6, 170.3, 158.1, 141.6, 81.6, 55.2, 52.0, 47.8, 43.2, 41.0, 38.1, 36.6, 31.6, 28.2, 25.5, 25.1, 23.2, 22.6, 22.1, 18.8, 17.5, 14.2, 12.6. HRMS (ESI): Calc'd for formula $C_{21}H_{39}N_5O_5Na^+$ $[M+Na]^+$ 464.2843, found 464.2845.

4.8.3.11. *tert*-Butyl (*S*,*E*)-4-(2-carbamoylhydrazineylidene)-3-((*S*)-2-hexanamido-3-methylbutanamido)butanoate (**7k**):

The product (183 mg, 88%) was isolated as a pale yellow solid (m.p. 105–107 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.72 (s, 1H), 7.77 (d, *J* = 7.7 Hz, 1H), 7.28 (s, 1H), 6.76 (m, 1H), 4.86 (m, 1H), 4.30 (m, 1H), 2.70 (m, 1H), 2.60 (m, 1H), 2.22 (m, 2H), 2.04 (m, 1H), 1.61 (m, 2H), 1.41 (m, 9H), 1.29 (m, 4H), 0.91 (m, 9H). $^{13}$C NMR (125 MHz, $CDCl_3$): δ 174.1, 171.8, 170.4, 158.1, 141.6, 81.7, 58.9, 55.0, 47.8, 36.7, 31.6, 31.0, 28.3, 25.6, 22.6, 19.6, 18.7, 17.5, 14.2, 12.5. HRMS (ESI): Calc'd for formula $C_{20}H_{37}N_5O_5Na^+$ $[M+Na]^+$ 450.2687, found 450.2679.

4.8.3.12. *tert*-Butyl (*S*,*E*)-5-(2-carbamoylhydrazineylidene)-4-((*S*)-4-methyl-2-octanamidopentanamido)pentanoate (**7l**):

The product (59 mg, 43%) was isolated as a colorless solid (102.5–104 ºC). $^{1}$H NMR (500 MHz, $CDCl_3$) δ 9.74 (s, 1H), 7.78 (d, *J* = 7.5 Hz, 1H), 7.19 (m, 1H), 6.55 (d, *J* = 8.2 Hz, 1H), 4.53 (m, 2H), 2.21 (m, 4H), 2.03 (m, 1H), 1.84 (m, 1H), 1.58 (m, 6H), 1.40 (s, 9H), 1.26 (m, 8H), 0.86 (m, 9H). $^{13}$C NMR (125 MHz, $CDCl_3$): δ 174.4, 172.7, 158.1, 80.9, 77.5, 77.2, 77.0, 54.9, 52.2, 50.3, 43.0, 40.7, 36.7, 31.9, 31.3, 29.39, 29.19, 28.3, 27.7, 25.9, 25.1, 23.2, 22.8, 22.1, 18.9, 17.6, 14.3, 12.5. HRMS (ESI): Calc'd for formula $C_{24}H_{44}N_5O_5^-$ $[M–H]^-$ 482.3348, found 482.3367.

#### 4.8.4. Removal of *tert*-butyl protecting groups and regeneration of aldehydes:

The coupling product (1.0 equiv) was stirred in 20% trifluoroacetic acid in dichloromethane (0.02 M) under argon, with immediate addition of water (3.0 equiv). The reaction mixtures were stirred for 1 h and concentrated in vacuo using a Genevac EZ-2 Elite centrifugal evaporator, and residual TFA was removed by forming the azeotrope with anhydrous toluene.

The deprotected semicarbazone intermediate (1.0 equiv) was dissolved in MeOH/37% formaldehyde/acetic acid (5:1:1, 16 mM) and stirred for 30 min at room temperature. Water was added to the reaction mixture (to 2 x initial reaction volume), and the reaction mixture was concentrated in vacuo to remove methanol. The reaction mixture was then diluted with water (1 x initial reaction volume) and extracted with three portions of ethyl acetate (each 1 x initial reaction volume). The combined organic layers were washed with

two portions of water and one portion of brine (each 1 x initial reaction volume). The combined organic layers were dried over $Na_2SO_4$, filtered, and concentrated in vacuo using a Genevac EZ-2 Elite centrifugal evaporator. The resulting products were further purified by trituration with two 2 mL volumes of diethyl ether.

For several of the more hydrophilic compounds (**8a**–**e**), the final reaction step diverged from the above procedure. The deprotected semicarbazone intermediate (1.0 equiv) was dissolved in MeOH/37% formaldehyde/acetic acid (5:1:1, 16 mM) and stirred for 30 min at room temperature. Water was added to the reaction mixture (to 2 x initial reaction volume), the reaction mixture was concentrated in vacuo to remove methanol, and water was then removed by lyophilization. The resulting solid was redissolved in water (1 x initial reaction volume) and filtered to remove insoluble particulates. The solution was lyophilized again and the resulting products further purified by trituration with two 2 mL volumes of diethyl ether.

4.8.4.1. (*S*)-3-((*S*)-2-Acetamido-4-methylpentanamido)-4-oxobutanoic acid (**8a**):

The product (24 mg, 66%) was isolated as a colorless solid (m.p 130-132 ºC). $^1$H NMR (500 MHz, 1:1 $CD_3OD/D_2O$, referenced to $D_2O$): δ 4.92 (m, 1H), 4.59 (m, 2H), 4.24 (m, 1H), 3.39 (s, 1H), 3.31 (s, 1H), 2.60 (m, 1H), 2.50 (m, 1H), 2.00 (s, 3H), 1.60 (m, 2H), 1.37 (m, 1H), 0.92 (m, 6H). $^{13}$C NMR (126 MHz; DMSO-$d_6$): δ 174.9, 172.6, 169.1, 102.5, 87.6, 85.3, 81.9, 50.8, 40.9, 24.24, 24.22, 22.9, 22.5, 21.6. HRMS (ESI): Calc'd for $C_{12}H_{19}N_2O_5^-$ $[M–H]^-$ 271.1299, found 271.1298.

4.8.4.2. (*S*)-4-((*S*)-2-Acetamido-4-methylpentanamido)-5-oxopentanoic acid (**8b**):

The product (23 mg, 60%) was isolated as a colorless solid (m.p. 99–101 ºC). $^{1}$H NMR (500 MHz; DMSO-$d_6$): δ 9.33 (s, 1H), 8.06 (m, 1H), 7.36 (m, 1H), 4.32 (m, 1H), 3.97 (m, 1H), 2.09 (m, 2H), 1.69 (m, 2H), 1.59 (m, 1H), 1.46 (m, 2H), 1.23 (m, 3H), 0.84 (m, 6H). $^{13}$C NMR (126 MHz; DMSO-$d_6$): δ 200.7, 178.8, 173.9, 57.4, 50.8, 45.8, 29.6, 24.2, 22.9, 22.5, 21.6, 12.1, 8.6. HRMS (ESI): Calc'd for formula $C_{15}H_{25}N_2O_5^-$ $[M–H]^-$ 285.1456, found 285.1454.

4.8.4.3. (*S*)-4-((*S*)-4-Methyl-2-propionamidopentanamido)-5-oxopentanoic acid (**8c**):

The product (36 mg, 64%) was isolated as an orange solid (m.p. 138–141 ºC). $^{1}$H NMR (500 MHz; DMSO-$d_6$): δ 9.56 (s, 1H), 9.36 (s, 1H), 8.38 (d, $J$ = 7.1 Hz), 7.92 (m, 1H), 4.33 (m, 1H), 4.03 (m, 1H), 2.12 (m, 2H), 1.58 (m, 1H), 1.44 (m, 2H), 1.26 (m, 2H), 0.98 (m, 3H), 0.86 (m, 6H). $^{13}$C NMR (126 MHz; DMSO-$d_6$): δ 200.7, 173.9, 173.1, 125.5, 57.4, 50.7, 39.5, 29.5, 28.3, 24.3, 23.06, 22.95, 21.6, 18.1, 16.7, 9.9. HRMS (ESI): Calc'd for formula $C_{14}H_{23}N_2O_5^-$ $[M–H]^-$ 299.1612, found 299.1611.

4.8.4.4. (*S*)-4-((*S*)-2-Butyramido-4-methylpentanamido)-5-oxopentanoic acid (**8d**):

The product (12 mg, 28%) was isolated as an orange solid (m.p. 124–126 ºC). $^{1}$H NMR (500 MHz; DMSO-$d_6$): δ 9.56 (s, 1H), 9.35 (s, 1H), 8.18 (m, 1H), 7.95 (m, 1H), 4.59 (m, 1H), 4.33 (m, 1H), 4.03 (m, 1H), 2.48 (m, 2H), 2.07 (m, 2H), 1.49 (m, 3H), 1.27 (m, 1H), 0.84 (m, 9H). $^{13}$C NMR (126 MHz; DMSO-$d_6$): δ 200.7, 173.9, 173.1, 172.1, 57.4, 53.5,

50.7, 37.1, 29.5, 24.3, 23.1, 21.6, 18.7, 18.1, 16.7, 13.5. HRMS (ESI): Calc'd for formula $C_{15}H_{25}N_2O_5^-$ $[M–H]^-$ 313.1769, found 313.1775.

4.8.4.5. (*S*)-4-((*S*)-2-Isobutyramido-4-methylpentanamido)-5-oxopentanoic acid (**8e**):

The product (7 mg, 55%) was isolated as a light brown solid (m.p. 117–119 ºC). $^1$H NMR (600 MHz; $CDCl_3$): δ 9.54 (s, 1H), 7.48 (s, 1H), 7.35 (s, 1H), 6.49 (d, *J* = 8.1 Hz, 1H), 4.60 (d, *J* = 6.1 Hz, 1H), 4.49 (m, 1H), 2.42 (m, 2H), 1.90 (m, 1H), 1.65 (m, 2H), 1.57 (m, 1H), 1.25 (m, 2H), 1.14 (m, 6H), 0.92 (m, 6H). $^{13}$C NMR (126 MHz; $CDCl_3$): 198.7, 178.3, 176.2, 173.5, 110.2, 77.2, 58.2, 56.1, 55.5, 51.7, 51.0, 41.2, 35.7, 29.9, 25.1, 23.0, 22.4, 19.81, 19.76, 19.4. HRMS (ESI): Calc'd for formula $C_{15}H_{25}N_2O_5^-$ $[M–H]^-$ 313.1769, found 313.1768.

4.8.4.6. (4*S*)-4-((2*S*)-4-Methyl-2-(2-methylbutanamido)pentanamido)-5-oxopentanoic acid (**8f**):

The product (9 mg, 48%) was isolated as a yellow solid (m.p. 140–142 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.55 (s, 1H), 6.39 (m, 1H), 4.60 (m, 1H), 4.50 (m, 1H), 2.41 (m, 2H), 2.17 (m, 2H), 1.63 (m, 3H), 1.43 (m, 1H), 1.25 (m, 2H), 1.12 (m, 3H), 0.92 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 177.8, 173.5, 125.7, 53.7, 51.7, 30.6, 29.9, 27.5, 25.1, 23.0, 22.4, 22.3, 17.4, 12.0. HRMS (ESI): Calc'd for formula $C_{16}H_{27}N_2O_5^-$ $[M–H]^-$ 327.1925, found 327.1924.

4.8.4.7. (4*S*)-4-((2*S*)-4-Methyl-2-(2-methylpentanamido)pentanamido)-5-oxopentanoic acid (**8g**):

The product (7 mg, 16%) was isolated as a light brown solid (m.p. 127–129 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.54 (s, 1H), 7.57 (m, 1H), 6.81 (m, 1H), 4.63 (m, 1H), 4.47 (m,

1H), 2.40 (m, 2H), 2.26 (m, 2H), 1.89 (m, 1H), 1.57 (m, 3H), 1.25 (m, 4H), 1.08 (m, 3H), 0.89 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 198.6, 178.2, 176.4, 175.8, 128.8, 110.2, 60.7, 58.2, 56.1, 51.7, 41.0, 36.5, 25.0, 22.3, 21.8, 20.7, 17.91, 17.72, 14.2. HRMS (ESI): Calc'd for formula $C_{17}H_{29}N_2O_5^-$ $[M–H]^-$ 341.2082, found 341.2082.

4.8.4.8. (*S*)-4-((*S*)-2-Hexanamido-4-methylpentanamido)-5-oxopentanoic acid (**8h**):

The product (11 mg, 34%) was isolated as a colorless solid (m.p. 127–129 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.53 (s, 1H), 7.67 (d, *J* = 7.0 Hz, 1H), 6.74 (d, *J* = 8.2 Hz, 1H), 4.68 (br s, 1H), 4.44 (br s, 1H), 2.39 (br s, 2H), 2.20 (m, 4H), 1.90 (m, 1H), 1.59 (m, 2H), 1.28 (m, 4H), 0.91 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 198.5, 176.0, 174.9, 171.5, 60.7, 58.3, 51.8, 41.2, 36.5, 31.5, 29.8, 25.5, 22.9, 22.5, 21.0, 14.4, 14.1. HRMS (ESI): Calc'd for formula $C_{17}H_{29}N_2O_5^-$ $[M–H]^-$, 327.1925; Found, 327.1924 .

4.8.4.9. (*S*)-4-((*S*)-2-Hexanamido-3-methylbutanamido)-5-oxopentanoic acid (**8i**):

The product (8 mg, 23%) was isolated as an orange solid (m.p. 111–114 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.55 (s, 1H), 6.60 (m, 1H), 6.11 (m, 1H), 4.35 (m, 1H), 4.13 (m, 1H), 2.42 (m, 2H), 2.24 (m, 4H), 1.61 (m, 3H), 1.28 (m, 4H), 0.89 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 176.6, 175.2, 77.2, 56.1, 36.5, 31.5, 25.6, 22.5, 21.0, 19.3, 18.7, 14.1. HRMS (ESI): Calc'd for formula $C_{16}H_{27}N_2O_5^-$ $[M–H]^-$ 327.1925, found 327.1924.

4.8.4.10. (*S*)-3-((*S*)-2-Hexanamido-4-methylpentanamido)-4-oxobutanoic acid (**8j**):

The product (10 mg, 24%) was isolated as a colorless glass. $^1$H NMR (500 MHz; $CDCl_3$): δ 7.34 (m, 1H), 6.21 (m, 1H), 4.59 (m, 2H), 2.26 (m, 3H), 1.63 (m, 4H), 1.30 (m, 4H),

0.96 (m, 2H), 0.88 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 176.8, 176.1, 174.3, 57.2, 36.8, 31.56, 31.51, 31.2, 25.6, 22.56, 22.51, 21.0, 19.2, 17.9, 14.1. HRMS (ESI): Calc'd for formula $C_{16}H_{27}N_2O_5^-$ $[M–H]^-$ 327.1925, found 327.1926.

4.8.4.11. (*S*)-3-((*S*)-2-Hexanamido-3-methylbutanamido)-4-oxobutanoic acid (**8k**):

The product (7 mg, 17%) was isolated as a colorless solid (m.p. 126–128 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.72 (s, 1H), 8.70 (br s, 1H), 8.18 (m, 1H), 7.14 (m, 1H), 4.59 (m, 1H), 4.50 (m, 1H), 2.20 (m, 2H), 1.57 (m, 5H), 1.26 (m, 6H), 0.88 (m, 9H). $^{13}$C NMR (151 MHz; $CDCl_3$): δ 176.2, 175.1, 174.0, 77.2, 56.0, 53.6, 51.9, 41.1, 36.4, 31.5, 25.5, 25.0, 22.9, 22.5, 22.0, 21.0, 14.1. HRMS (ESI): Calc'd for formula $C_{15}H_{25}N_2O_5^-$ $[M–H]^-$ 313.1769, found 313.1769.

4.8.4.12. (*S*)-4-((*S*)-4-Methyl-2-octanamidopentanamido)-5-oxopentanoic acid (**8l**):

The product (7 mg, 15%) was isolated as a colorless solid (m.p. 106–107 ºC). $^1$H NMR (500 MHz; $CDCl_3$): δ 9.54 (s, 1H), 7.45 (s, 1H), 6.29 (br s, 1H), 6.17 (m, 1H), 4.49 (m, 1H), 4.35 (m, 1H), 2.42 (m, 2H), 2.23 (m, 4H), 1.95 (m, 1H), 1.62 (m, 4H), 1.28 (m, 8H), 0.92 (m, 9H). $^{13}$C NMR (126 MHz; $CDCl_3$): δ 175.0, 174.6, 170.5, 128.7, 125.7, 66.0, 56.0, 53.6, 31.9, 29.32, 29.15, 25.9, 25.0, 23.5, 22.8, 15.4, 14.2. HRMS (ESI): Calc'd for formula $C_{19}H_{33}N_2O_5^-$ $[M–H]^-$ 369.2395, found 369.2395.

**Acknowledgements**

We gratefully acknowledge Drs. Harry J. Flint and Sylvia H. Duncan (University of Aberdeen) for proving *R. bromii* L2-63 and gDNA from this organism and for helpful

discussions, Dr. Emma Allen-Vercoe (University of Guelph, Ontario) for providing *R. bromii* 22-5-S 6 FAA NB, and Sunia Trauger for mass spectrometry analyses. This work was supported by DARPA (HR0011-16-2-0013) and a George W. Merck Fellowship.

Conflicts of interest: none

**References**


1. Garg N, Luzzatto-Knaan T, Melnik AV, Caraballo-Rodríguez AM, Floros DJ, Petras D, Gregor R, Dorrestein PC, Phelan V V. *Nat Prod Rep*. 2017.
2. Crost EH, Ajandouz EH, Villard C, Geraert PA, Puigserver A, Fons M. *Biochimie*. 2011; 93: 1487–1494.
3. Wilson MR, Zha L, Balskus EP. *J Biol Chem*. 2017; 292: 8546–8552.
4. Dabard J, Bridonneau C, Phillipe C, Anglade P, Molle D, Nardi M, Ladiré M, Girardin H, Marcille F, Gomez A, Fons M. *Appl Environ Microbiol*. 2001; 67: 4111–4118.
5. Cohen LJ, Kang H-S, Chu J, Huang Y-H, Gordon EA, Reddy BVB, Ternei MA, Craig JW, Brady SF. *Proc Natl Acad Sci*. 2015; 112: E4825–E4834.
6. Guo C, Chang F, Wyche TP, Backus KM, Acker TM, Funabashi M, Taketani M, Donia MS, Nayfach S, Pollard KS, Craik CS, Cravatt BF, Clardy J, Voigt CA, Fischbach MA. *Cell*. 2017; 168: 517–526.e18.
7. Chu J, Vila-Farres X, Inoyama D, Ternei M, Cohen LJ, Gordon EA, Reddy BVB, Charlop-Powers Z, Zebroski HA, Gallardo-Macias R, Jaskowski M, Satish S, Park S, Perlin DS, Freundlich JS, Brady SF. *Nat Chem Biol*. 2016: 1–5.
8. Coulter SN, Schwan WR, Ng EYW, Langhorne MH, Ritchie HD, Westbrock-Wadman S, Hufnagle WO, Folger KR, Bayer AS, Stover CK. *Mol Microbiol*. 1998; 30: 393–404.
9. Qin X, Singh KV, Weinstock GM, Murray BE. *Infect Immun*. 2000; 68: 2579–2586.
10. Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, Ugarte E, Muñoz-

Tamayo R, Paslier DLE, Nalin R, Dore J, Leclerc M. *Environ Microbiol*. 2009; 11: 2574–2584.

11. Moore WEC, Moore LH. *Appl Environ Microbiol*. 1995; 61: 3202–3207.
12. Flint HJ, Scott KP, Louis P, Duncan SH. *Nat Rev Gastroenterol Hepatol*. 2012; 9: 577–589.
13. Lay C, Sutren M, Rochet V, Saunier K, Doré J, Rigottier-Gois L. *Environ Microbiol*. 2005; 7: 933–946.
14. Ze X, Duncan SH, Louis P, Flint HJ. *ISME J*. 2012; 6: 1535–1543.
15. Englyst HN, Macfarlane GT. *J Sci Food Agric*. 1986; 37: 699–706.
16. Kabeerdoss J, Sankaran V, Pugazhendhi S, Ramakrishna BS. *BMC Gastroenterol*. 2013; 13: 20.
17. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, Blugeon S, Bridonneau C, Furet J-P, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottière HM, Doré J, Marteau P, Seksik P, Langella P. *Proc Natl Acad Sci U S A*. 2008; 105: 16731–16736.
18. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA. *Cell*. 2014; 158: 1402–1414.
19. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. *BMC Evol Biol*. 2007; 7: 78.
20. Fischbach MA, Walsh CT. *Chem Rev*. 2006; 106: 3468–3496.
21. Yeh H-H, Ahuja M, Chiang Y-M, Oakley CE, Moore S, Yoon O, Hajovsky H, Bok J-W, Keller NP, Wang CCC, Oakley BR. *ACS Chem Biol*. 2016; 11: 2275–2284.

22. Chen Y, McClure RA, Zheng Y, Thomson RJ, Kelleher NL. *J Am Chem Soc*. 2013; 135: 10449–56.

23. Carroll IM, Ringel-Kulka T, Ferrier L, Wu MC, Siddle JP, Bueno L, Ringel Y. *PLoS One*. 2013; 8: e78017.

24. Moore WEC, Cato EP, Holdeman LV. *Int J Syst Bacteriol*. 1972; 22: 78–80.

25. Brotherton CA, Balskus EP. *J Am Chem Soc*. 2013; 135: 3359–3362.

26. Reimer D, Pos KM, Thines M, Grün P, Bode HB. *Nat Chem Biol*. 2011; 7: 888–890.

27. Imker HJ, Krahn D, Clerc J, Kaiser M, Walsh CT. *Chem Biol*. 2010; 17: 1077–1083.

28. Bachmann BO, Ravel J. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data., Elsevier Inc., 1st edn., 2009, vol. 458.

29. Pei J, Kim BH, Grishin N V. *Nucleic Acids Res*. 2008; 36: 2295–2300.

30. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. *Anal Chem*. 2012; 84: 5035–5039.

31. Linne U, Marahiel MA. *Methods Enzymol*. 2004; 388: 293–315.

32. La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD. *Chem Biol*. 2004; 11: 195–201.

33. Nakamura H, Hamer HA, Sirasani G, Balskus EP. *J Am Chem Soc*. 2012; 134: 18518–18521.

34. Watanabe CMH, Townsend CA. *Chem Biol*. 2002; 9: 981–988.

35. Wilson DJ, Shi C, Teitelbaum AM, Gulick AM, Aldrich CC. *Biochemistry*. 2013; 52: 926–937.

36. Gaitatzis N, Kunze B, Müller R. *Proc Natl Acad Sci U S A*. 2001; 98: 11136–41.

37. Kallberg Y, Oppermann U, Jörnvall H, Persson B. *Eur J Biochem*. 2002; 269: 4409–4417.

38. Wyatt MA, Wang W, Roux CM, Beasley FC, Heinrichs DE, Dunman PM, Magarvey NA. *Science*. 2010; 329: 294–296.

39. Read JA, Walsh CT. *J Am Chem Soc*. 2007; 129: 15762–15763.

40. Li Y, Weissman KJ, Müller R. *J Am Chem Soc*. 2008; 130: 7554–5.

41. Kopp F, Mahlert C, Grünewald J, Marahiel MA. *J Am Chem Soc*. 2006; 128: 16478–9.

42. Cardozo C, Chen WE, Wilk S. *Arch Biochem Biophys*. 1996; 334: 113–120.

43. Graybill TL, Dolle RE, Helaszek CT, Miller RE, Ator MA. *Int J Pept Protein Res*. 1994; 44: 173–182.

44. Kawalec M, Potempa J, Moon JL, Travis J, Murray BE. *J Bacteriol*. 2005; 187: 266–275.

45. Nemoto TK, Ohara-Nemoto Y, Ono T, Kobayakawa T, Shimoyama Y, Kimura S, Takagi T. *FEBS J*. 2008; 275: 573–587.

46. Thomas VC, Thurlow LR, Boyle D, Hancock LE. *J Bacteriol*. 2008; 190: 5690–5698.

47. Martí M, Trotonda MP, Tormo-Más MÁ, Vergara-Irigaray M, Cheung AL, Lasa I, Penadés JR. *Microbes Infect*. 2010; 12: 55–64.

48. Singh KV, Nallapareddy SR, Nannini EC, Murray BE. *Infect Immun*. 2005; 73: 4888–4894.

49. Sifri CD, Mylonakis E, Singh K V, Qin X, Garsin DA, Murray BE, Ausubel FM,

Calderwood SB. *Infect Immun*. 2002; 70: 5647–5650.

50. Qin X, Singh KV, Weinstock GM, Murray BE. *Infect Immun*. 2000; 68: 2579–2586.

51. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. *J Med Chem*. 2004; 47: 1739–1749.

52. Prasad L, Leduc Y, Hayakawa K, Delbaere LTJ. *Acta Crystallogr Sect D Biol Crystallogr*. 2004; 60: 256–259.

53. Hamilton R, Walker B, Walker BJ. *Bioorganic Med Chem Lett*. 1998; 8: 1655–1660.

54. Burchacka E, Skoreński M, Sieńczyk M, Oleksyszyn J. *Bioorg Med Chem Lett*. 2013; 23: 1412–1415.

55. Acton DS, Tempelmans Plat-Sinnige MJ, Van Wamel W, De Groot N, Van Belkum A. *Eur J Clin Microbiol Infect Dis*. 2009; 28: 115–127.

56. Blaimont B, Charlier J, Wauters G. *Microb Ecol Health Dis*. 1995; 8: 87–92.

57. Kawalec M, Potempa J, Moon JL, Travis J, Murray BE. *J Bacteriol*. 2005; 187: 266–275.

58. Svendsen I, Breddam K. *Eur J Biochem*. 1992; 204: 165–71.

59. Kakudo S, Kikuchi N, Kitadokoro K, Fujiwara T, Nakamura E, Okamoto H, Shin M, Tamaki M, Teraoka H, Tsuzuki H. *J Biol Chem*. 1992; 267: 23782–23788.

60. Leshchinskaya IB, Shakirov EV, Itskovitch EL, Balaban NP, Mardanova AM, Sharipova MR, Viryasov MB, Rudenskaya GN, Stepanov VM. *FEBS Lett*. 1997; 404: 241–244.

61. Yoshida N, Tsuruyama S, Nagata K, Hirayama K, Noda K, Makisumi S. *J Biochem*. 1988; 104: 451–456.

62. Demidyuk IV, Chukhontseva KN, Kostrov SV. *Acta Naturae*. 2017; 9: 214–216.

63. Niidome T, Yoshida N, Ogata F, Ito A, Noda K. *J Biochem*. 1990; 108: 965–970.

64. Meijers R, Blagova EV, Levdikov VM, Rudenskaya GN, Chestukhina GG, Akimkina TV, Kostrov SV, Lamzin VS, Kuranova IP. *Biochemistry*. 2004; 43: 2784–2791.

65. Moon JL, Banbula A, Oleksy A, Mayo JA, Travis J. *Biol Chem*. 2001; 382: 1095–1099.

66. Cavarelli J, Prévost G, Bourguet W, Moulinier L, Chevrier B, Delagoutte B, Bilwes A, Mourey L, Rifai S, Piémont Y, Moras D. *Structure*. 1997; 5: 813–824.

67. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. *Nucleic Acids Res*. 2012; 40: 115–122.

68. Biancheri P, Di Sabatino A, Corazza GR, MacDonald TT. *Cell Tissue Res*. 2013; 351: 269–280.

69. Vergnolle N. *Gut*. 2016; 65: 1215–1224.

70. Bustos D, Negri G, de Paula JA, Di M, Yapur V, Facente A, de Paula A. *Medicina*. 1998; 58: 262–264.

71. Róka R, Rosztóczy A, Leveque M, Izbéki F, Nagy F, Molnár T, Lonovics J, Garcia–Villar R, Fioramonti J, Wittmann T, Bueno L. *Clin Gastroenterol Hepatol*. 2007; 5: 550–555.

72. Herszényi L, Barabás L, Hritz I, István G, Tulassay Z. 2014; 20: 13246–13257.

73. Senda S, Fujiyama Y, Bamba T, Hosoda S. *Intern Med*. 1993; 32: 350–4.

74. Macfarlane GT, Allison C. *FEMS Microbiol Lett*. 1986; 38: 19–24.

75. Carroll IM. *World J Gastroenterol*. 2013; 19: 7531.

76. Steck N, Mueller K, Schemann M, Haller D. *Gut*. 2012; 61: 1610–1618.

77. Steck N, Hoffmann M, Sava IG, Kim SC, Hahne H, Tonkonogy SL, Mair K, Krueger D, Pruteanu M, Shanahan F, Vogelmann R, Schemann M, Kuster B, Sartor RB, Haller D. *Gastroenterology*. 2011; 141: 959–971.

78. Walker AW, Duncan SH, Harmsen HJM, Holtrop G, Welling GW, Flint HJ. *Environ Microbiol*. 2008; 10: 3275–3283.

79. Miyazaki K, Martin J, Marinsek-Logar R, Flint H. *Anaerobe*. 1997; 3: 373–381.

80. Ze X, Ben David Y, Laverde-Gomez JA, Dassa B, Sheridan PO, Duncan SH, Louis P, Henrissat B, Juge N, Koropatkin NM, Bayer EA, Flint HJ. *MBio*. 2015; 6: e01058-15.

81. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. *J Bacteriol*. 1991; 173: 697–703.

82. Rio DC, Ares M, Hannon GJ, Nilsen TW. *Cold Spring Harb Protoc*. 2010; 5: 1–4.

83. Mofid MR, Marahiel MA, Ficner R, Reuter K. *Acta Crystallogr Sect D Biol Crystallogr*. 1999; 55: 1098–1100.

84. Grahl-Nielsen O, Solheim E. *J Chromatogr A*. 1975; 105: 89–94.

85. Yang, X., Lubian, E., Renes, H., Tondeur, A. P., Haiber, S., Liu, X., Fu, X.US2014/0127144 2014.

**Graphical Abstract:**

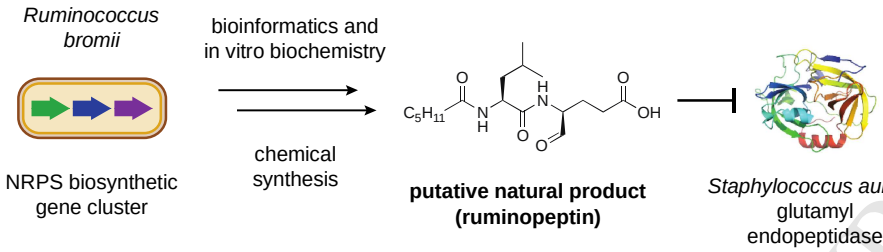
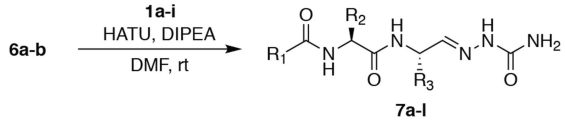
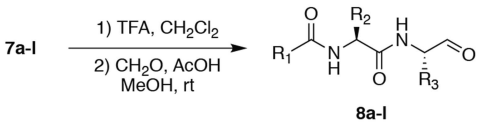


Table 1:




| Entry | Product | $R_1$ | $R_2$ | $R_3$ | %Yield |
|---|---|---|---|---|---|
| 1 | **7a** | Me | Leu | Asp (*O*-*t*Bu) | 31% |
| 2 | **7b** | Me | Leu | Glu (*O*-*t*Bu) | 28% |
| 3 | **7c** | $C_2H_5$ | Leu | Glu (*O*-*t*Bu) | 65% |
| 4 | **7d** | $C_3H_7$ | Leu | Glu (*O*-*t*Bu) | 63% |
| 5 | **7e** | *i*Bu | Leu | Glu (*O*-*t*Bu) | 48% |
| 6 | **7f** |  | Leu | Glu (*O*-*t*Bu) | 62% |
| 7 | **7g** |  | Leu | Glu (*O*-*t*Bu) | 87% |
| 8 | **7h** | $C_5H_{11}$ | Leu | Glu (*O*-*t*Bu) | 88% |
| 9 | **7i** | $C_5H_{11}$ | Val | Glu (*O*-*t*Bu) | 67% |
| 10 | **7j** | $C_5H_{11}$ | Leu | Asp (*O*-*t*Bu) | 39% |
| 11 | **7k** | $C_5H_{11}$ | Val | Asp (*O*-*t*Bu) | 88% |
| 12 | **7l** | $C_7H_{15}$ | Leu | Glu (*O*-*t*Bu) | 43% |

Table 2:




| Entry | Product | $R_1$ | $R_2$ | $R_3$ | %Yield |
|---|---|---|---|---|---|
| 1 | **8a** | Me | Leu | Asp | 66% |
| 2 | **8b** | Me | Leu | Glu | 60% |
| 3 | **8c** | $C_2H_5$ | Leu | Glu | 64% |
| 4 | **8d** | $C_3H_7$ | Leu | Glu | 28% |
| 5 | **8e** | *i*Bu | Leu | Glu | 55% |
| 6 | **8f** |  | Leu | Glu | 48% |
| 7 | **8g** |  | Leu | Glu | 16% |
| 8 | **8h** | $C_5H_{11}$ | Leu | Glu | 34% |
| 9 | **8i** | $C_5H_{11}$ | Val | Glu | 23% |
| 10 | **8j** | $C_5H_{11}$ | Leu | Asp | 24% |
| 11 | **8k** | $C_5H_{11}$ | Val | Asp | 17% |
| 12 | **8l** | $C_7H_{15}$ | Leu | Glu | 15% |