Enzyme-Primed Native Chemical Ligation Produces Autoinducing Cyclopeptides in Clostridia

Evelyn M. Molloy, Maria Dell, Veit G. Hänsch, Kyle L. Dunbar, Romy Feldmann, Ansgar Oberheide, Lydia Seyfarth, Jana Kumpfmüller, Therese Horch, Hans-Dieter Arndt, and Christian Hertweck*

Abstract: Clostridia coordinate many important processes such as toxin production, infection, and survival by density-dependent communication (quorum sensing) using autoinducing peptides (AIPs). Although clostridial AIPs have been proposed to be (thio)lactone-containing peptides, their true structures remain elusive. Here, we report the genome-guided discovery of an AIP that controls endospore formation in Ruminiclostridium cellulolyticum. Through a combination of chemical synthesis and chemical complementation assays with a mutant strain, we reveal that the genuine chemical mediator is a homodetic cyclopeptide (cAIP). Kinetic analyses indicate that the mature cAIP is produced via a cryptic thiolactone intermediate that undergoes a rapid S→N acyl shift, in a manner similar to intramolecular native chemical ligation (NCL). Finally, by implementing a chemical probe in a targeted screen, we show that this novel enzyme-primed, intramolecular NCL is a widespread feature of clostridial AIP biosynthesis.

Introduction

The Clostridia are a heterogeneous class of anaerobic bacteria that play major roles in human and animal health, industrial biotechnology, and the environment.[1] There is growing evidence that Clostridia coordinate a broad range of crucial group behaviors that affect their physiology and pathogenicity by sensing the accumulation of secreted peptidic signals that regulate gene expression.[2] Since the concentration of these chemical mediators correlates with cell density (quorum sensing, QS), gene expression can be synchronized with specific environmental conditions and growth phases.[3] An impressive number of genetic studies have revealed that the accessory gene regulator (Agr) system is central to QS in diverse Clostridia.[2a–m] For example, Agr-based QS is linked to the virulence potential of the infamous human pathogens *Clostridium difficile*[2a,c] and *Clostridium perfringens*,[2m] and regulates toxin production in *C. difficile*,[2a,b] *C. perfringens*[2d–j] and *Clostridium botulinum*.[2k] Furthermore, Agr-mediated QS is implicated in the regulation of clostridial sporulation[2f,k,l] as well as granulose production in the industrially important solvent producer *Clostridium acetobutylicum*.[2l]

The QS signals derived from the Agr system are called autoinducing peptides (AIPs) and belong to the ribosomally synthesized and post-translationally modified peptide (RiPP) class of natural products.[4] The mechanistic details of AIP biosynthesis and signaling were deciphered in pathogenic *Staphylococcus* spp.,[5] with the prototypical Agr system consisting of four components: a gene-encoded precursor peptide (AgrD), an integral membrane macrocyclase (AgrB), and a sensor kinase (AgrC) and response regulator (AgrA) pair that transduce the signal (Figure 1 A).[6] AgrD is comprised of three regions: an *N*-terminal leader peptide that aids in membrane localization,[7] a core peptide that encodes the QS signal,[8] and a *C*-terminal follower peptide[9] that is recognized by AgrB. The biosynthesis of the mature QS signal from this unmodified precursor peptide occurs over two steps (Figure 1 B). First, the AgrB transmembrane cysteine protease[8–10] removes the follower peptide and catalyzes intramolecular thiolactone bond formation between a conserved cysteine in the core peptide and the liberated *C*-terminus.[11] The leader peptide is then removed by an extracellular peptidase,[12] leaving the thiolactone macrocycle with an *N*-terminal extension (NTE) of varying length (Figure 1 A and Figure 1 B).[5a,13] In rare cases, cyclization involves a serine residue, resulting in the formation of a lactone macrocycle.[5c,14] Akin to other QS systems, AIP production and sensing are mutually enhancing, leading to a positive-feedback autoinduction circuit.[5a]

By analogy to their staphylococcal counterparts, it has been assumed that clostridial AIPs generally possess thiolactone moieties.[2a,b,h,j,l] Considering the prevalence[2l,15] and essential roles of Agr signaling circuits in Clostridia,[2a–m] this assumption should be substantiated. Until now, the chemical

[*] Dr. E. M. Molloy, M. Dell, V. G. Hänsch, Dr. K. L. Dunbar,
R. Feldmann, Dr. J. Kumpfmüller, T. Horch, Prof. Dr. C. Hertweck
Dept. of Biomolecular Chemistry, Leibniz Institute for Natural,
Product Research and Infection Biology, HKI
Beutenbergstr. 11a, 07745 Jena (Germany)
E-mail: christian.hertweck@leibniz-hki.de

A. Oberheide, L. Seyfarth, Prof. Dr. H.-D. Arndt
Institute for Organic Chemistry and Macromolecular Chemistry,
Friedrich Schiller University Jena
Humboldtstr. 10, 07743 Jena (Germany)

Prof. Dr. C. Hertweck
Faculty of Biological Sciences, Friedrich Schiller University Jena
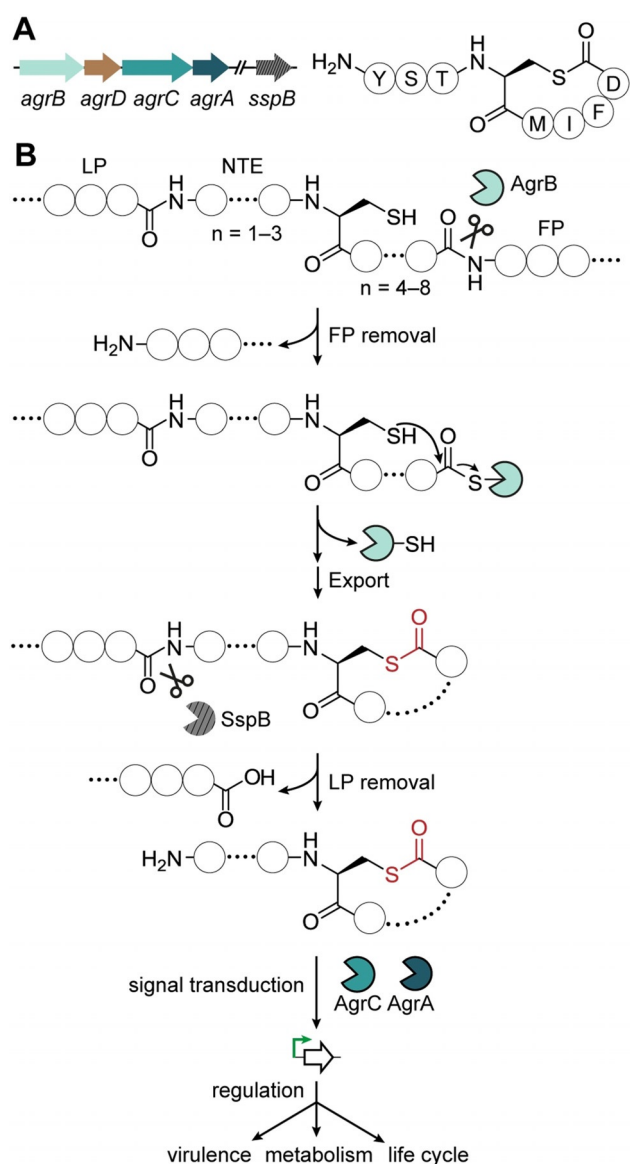07743 Jena (Germany)

Figure 1. The prototypical Agr system of *Staphylococcus aureus*: A) Representation of an *S. aureus* AIP biosynthetic gene cluster showing genes encoding the integral membrane macrocyclase (AgrB), the precursor peptide (AgrD), the cognate two-component transduction system (AgrC-AgrA), as well as the distal gene encoding the membrane protease (SspB). Also shown is an example of a mature staphylococcal thiolactone-containing AIP. B) Overview of the biosynthetic pathway to staphylococcal thiolactone-containing AIPs. The AIP is derived from an internal fragment of the ribosomally synthesized precursor peptide AgrD. First, AgrB catalyzes *C*-terminal proteolysis of the follower peptide (FP) and subsequent macrocyclization, followed by secretion. Then, the *N*-terminal leader peptide (LP) is removed by SspB, leaving an *N*-terminal extension (NTE) of varying length. The mature AIP, which varies among strains, consists of a five-membered thiolactone ring with a 2–4 residue NTE. In response to binding of the AIP, the sensor kinase AgrC is autophosphorylated and the signal is transduced by the response regulator AgrA leading to global alterations in gene expression.

structures of clostridial AIPs have been inferred from indirect experimental evidence such as signal degradation,[2b] and by chemical complementation of mutant phenotypes with syn-

thetic AIPs designed from bioinformatics-based structure predictions.[2j,l] Here, we report the first complete structural assignment of a clostridial AIP and provide compelling evidence that the previously suggested structures are transient precursors of the true signaling molecules. Our findings not only prompt a re-examination of some previously reported AIP structures, but also reveal a novel pathway to cyclic RiPPs by natural native chemical ligation (NCL).

Results and Discussion

Genomics-guided discovery of an AIP from a cellulolytic, antibiotic-producing anaerobe

Ruminiclostridium cellulolyticum DSM 5812 has been extensively studied for its ability to degrade cellulose[16] and to produce the DNA-gyrase inhibiting antibiotic closthioamide,[17] but there has been no report on AIP-mediated QS in this genetically tractable strain.[18] Indeed, AIP-dependent signaling pathways have not been studied in any *Ruminiclostridium* species to date. We therefore chose *R. cellulolyticum* as a starting point for the characterization of clostridial AIPs. To gain insight into the potential of *R. cellulolyticum* to produce AIPs, we mined its genome for genes encoding members of the AgrB protein family (PF04647). We identified four homologs of the AgrB peptidase gene (Table S1). Next, we scanned the genome region surrounding each candidate *agrB* for genes tentatively encoding a AIP precursor peptide (AgrD) and detected a candidate *agrD* in all four loci (Table S1 and Figure S1). In theory, the four *agrBD* gene pairs could enable *R. cellulolyticum* to produce four different AIPs (*Rc*-AIP1-4) (Figure S1). This finding is unusual since previous bioinformatic analyses detected just one or two AIP gene clusters in members of the Clostridia.[2b,c,l,15]

The core peptide sequences of each possible *R. cellulolyticum* AIP were predicted by aligning *Rc*-AgrD1-4 to AgrD homologues for which the corresponding products have been structurally characterized (Figure S2). Based on previous synthetic studies involving complementation of mutant phenotypes in various Clostridia[2j,l] and a preliminary MS-based detection of an AIP from *Hungateiclostridium thermocellum*,[19] we predicted that the AIPs from *R. cellulolyticum* would not have NTEs, *i.e.* they should present a free *N*-terminus at the ultimate cysteine residue.

In order to assess the production of AIPs by *R. cellulolyticum*, the strain was cultivated, the culture was extracted with ethyl acetate, and the extract was analyzed by HPLC-HRMS. The metabolite profile was searched for the molecular ion masses calculated for each macrocyclic AIP (*Rc*-AIP1-4) without an NTE. No signals were detected that could account for the *Rc*-AIP2-4 predictions, regardless of NTE length (0 to 4 residues). However, we detected a metabolite with an exact mass consistent with the prediction for mature *Rc*-AIP1 devoid of an NTE (calc. m/z 873.3389 $[M+H]^+$; found m/z 873.3398) $[M+H]^+$) (Figure 2A and Figure 2B). We also checked the metabolite profile for the exact masses of *Rc*-AIP1 bearing NTEs of up to four residues; however, no
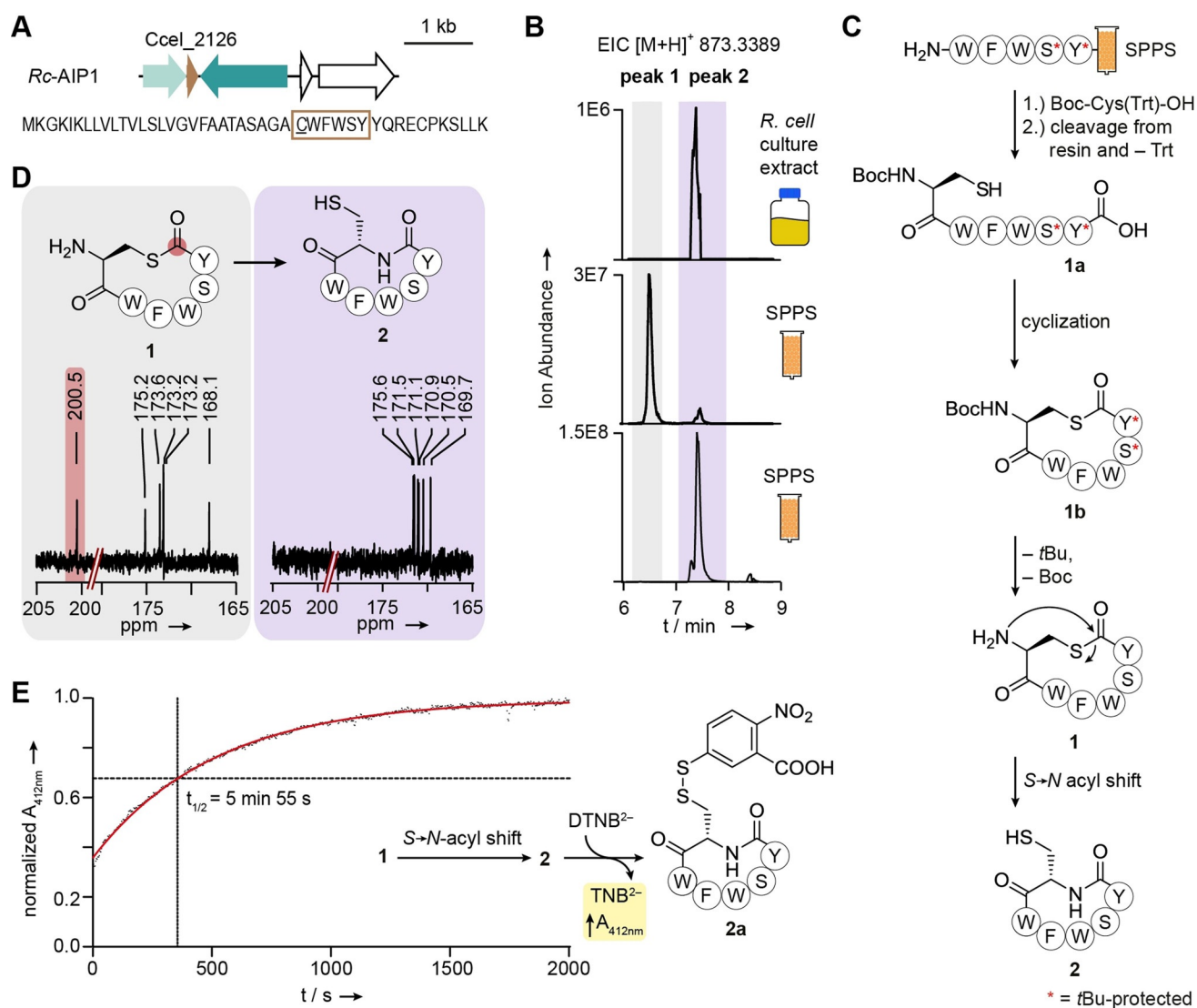
Figure 2. *Ruminiclostridium cellulolyticum* produces a homodetic cyclopeptide (cAIP) derived from an instable thiolactone: A) Representation of the *Rc*-AIP1 biosynthetic gene cluster. The predicted core peptide sequence is surrounded by a brown box, with the cysteine residue at the putative cyclization position underlined. B) Extracted ion chromatogram (EIC) showing the detection of the mass corresponding to *Rc*-AIP1 in the metabolite profile of *R. cellulolyticum* and the synthetic references of **1** and **2** (see Figure S7 for the mass spectra). SSPS, solid-phase peptide synthesis. C) Synthetic route to synthetic references **1** and **2**. The asterisk indicates the *t*Bu-protection of a hydroxyl group. D) ^{13}C-NMR spectra recorded for synthetic references **1** and **2**, with a signal (highlighted in red) in the characteristic region for a thiolactone in the case of **1** (see Figure S6 and Figure S10 for full spectra). E) Determination of the stability of **1** by reaction of the spontaneously formed free thiol in **2** with Ellman's reagent. The increase in absorbance at 412 nm, resulting from the free anion (highlighted in yellow), was measured over time and the half-life ($t_{1/2}$) of **1** was calculated based on the reaction rate. The graph shows a representative example (see Figure S11 for complete data set).

corresponding signals were found (Figure S3). HRMS^2-based fragmentation of the ion corresponding to the *Rc*-AIP1 candidate resulted in a pattern in which all observed fragment masses were in agreement with a macrocyclic peptide of the predicted amino acid sequence (CWFWSY) (Figure S4). Notably, *Rc*-AIP1 is only the second example of a native clostridial AIP detected to date and, in both cases, the signaling molecules lack an NTE.[19]

Rc-AIP1 is a homodetic cyclopeptide derived from a thiolactone-containing intermediate

The possibility of spontaneous $S \rightarrow N$ acyl migrations occurring in AIPs with thiolactones formed between the *C*- and *N*-termini was only recently posited.[20] The identification of *Rc*-AIP1, in which an NTE is absent, provided us with an opportunity to investigate this prospect. We attempted to isolate *Rc*-AIP1 from *R. cellulolyticum*, but this proved challenging due to low production titers. As an alternative approach, we synthesized a reference compound for comparison to *Rc*-AIP1. Specifically, we employed Fmoc-based solid-

phase peptide synthesis (SPPS, for more experimental details see SI) to prepare the linear *t*Bu-protected peptide, WFWSY. The *N*-terminal amino acid (Cys) was coupled to the linear protected peptide as a *S*-Trt- and *N*-Boc-protected version by using ethyl cyanohydroxyiminoacetate (Oxyma), (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)-dimethylamino-morpholino-carbenium hexafluorophosphate (COMU), and diisopropylethylamine (DIPEA) as coupling reagents. The Trt group was removed simultaneously as the Boc- and *t*Bu-protected peptide was liberated from the resin by treatment with hexafluoroisopropanol (HFIP) and purified by preparative HPLC. The obtained thiol peptide acid **1a** was cyclized with 1-hydroxybenzotriazole hydrate (HOBt) and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDCI) as coupling reagents, and DIPEA and 4-dimethylaminopyridine (DMAP) as bases in DMF,[2j] to yield the corresponding thiolactone **1b** (Figure 2C and Figure S5). After removal of the *t*Bu-based protecting groups with trifluoroacetic acid (TFA), the labile thiolactone **1**, which could be handled under strictly acidic conditions only, was purified by preparative HPLC (3.4% overall yield starting from SPPS). The characteristic ^{13}C-NMR signal at ≈ 200 ppm confirmed the successful formation of the thiolactone-containing peptide **1** (Figure 2D and Figure S6).

Immediately after deprotection, we subjected the synthetic thiolactone **1** to HPLC-HRMS analysis in order to compare its retention time, accurate mass, and fragmentation patterns with those of *Rc*-AIP1 from *R. cellulolyticum* culture extracts. We observed a divergence in the retention times of **1** (peak 1) and *Rc*-AIP1 (Figure 2B and Figure S7). However, we noted that a minor congener (peak 2) in the reference chromatogram had the same retention time as *Rc*-AIP1. We also found that both compounds show the same *m/z* values and very similar HRMS^2 fragmentation patterns, yet differ in their stabilities to the collision energies used (Figure S8). These observations led us to conclude that **1** is unstable and undergoes an $S \rightarrow N$ acyl shift to form the corresponding homodetic cyclopeptide **2** with a more stable amide bond (Figure 2C).

To scrutinize this conjecture, we used HPLC-HRMS to monitor the stability of thiolactone **1** in MeOH. We noted that **1** was indeed unstable, since the corresponding peak decreased upon formation of a new peak that came to predominate in the HPLC-HRMS profile (Figure S9). In order to obtain a synthetic reference, we incubated **1** in 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS) at pH 7 until peak 2 became the major signal in the chromatogram. The corresponding compound **2** was purified by preparative HPLC and its lactam structure was confirmed by NMR analyses (Figure 2D and Figure S10). Notably, the retention time of **2** proved to be identical to that of the natural product (*Rc*-AIP1) (Figure 2B and Figure S7). Consequently, *Rc*-AIP1 must exclusively possess amide bonds and hence deviates from the hitherto assumed (thio)lactone-containing architecture of AIPs. In light of this difference, we designate *Rc*-AIP1 as a "cAIP" to emphasize that it is a homodetic cyclopeptide.

To exclude the possibility that lactam **2** arose in the *R. cellulolyticum* metabolite profile as an artifact of the extraction procedure, we freshly prepared thiolactone **1** and, immediately after deprotection, determined its rearrangement kinetics under conditions mimicking those of the cultivation (at 37 °C in 50 mM MOPS, pH 7.1). As the conversion of **1** to **2** results in the formation of a thiol, the transformation was monitored by using a spectrophotometer and a thiol detection assay based on Ellman's reagent (Figure 2E).[21] We determined an average half-life of 5.4 ± 0.6 min for **1** (Figure 2E and Figure S11), indicating that the NTE-free thiolactone-containing AIP would not persist over the estimated generation time (6 h) of *R. cellulolyticum* determined under similar cultivation conditions.[22] We propose that by the time the threshold concentration needed to trigger a QS-dependent response would be reached, the spontaneously formed **2** would dominate over **1**. We therefore propose that **2** is the native structure of *Rc*-AIP1.

Rc-AIP1 controls endospore formation in *Ruminiclostridium cellulolyticum*

Our elucidation of the structure of *Rc*-AIP1 revealed an unprecedented structural deviation for an AIP family member, but the question remained as to whether the homodetic cyclopeptide **2** acts as a competent signal. If so, it should be sensed by *R. cellulolyticum* and elicit a response. However, no physiological role has been reported for an *R. cellulolyticum* AIP. To this end, we created a null mutant by using CRISPR-nCas9 genome editing[18] to incorporate a nonsense mutation in *Rc-agrD1* (Ccel_2126), the gene encoding *Rc*-AIP1 (Figure 3A and Figure S12). As expected, no signal for *Rc*-AIP1 could be detected in the metabolite profile of *R. cellulolyticum* Δ*agrD1* (Figure 3B). This biosynthetic deficiency could be complemented by re-introducing an intact copy of *Rc-agrD1* on a plasmid (Figure 3B).

Since Agr-based QS has been implicated in the initiation of sporulation in Clostridia,[2f,k,l] we assessed the sporulation efficiency of *R. cellulolyticum* Δ*agrD1* (Figure 3C). We found that the number of heat-resistant endospores formed by *R. cellulolyticum* Δ*agrD1* was indeed reduced by greater than two orders of magnitude compared to the wild-type strain (Figure 3D). Exogenous supplementation of *R. cellulolyticum* Δ*agrD1* with **2** could rescue the spore formation deficiency in a concentration-dependent manner (Figure 3D). Importantly, **2** restored wild-type levels of sporulation at 200 nM, a concentration considered physiologically relevant[23] and at least an order of magnitude lower than previously reported for clostridial Agr-dependent phenotypes.[2j,l] Combined with the inherent instability of **1** over a biologically relevant timescale, these chemical complementation experiments corroborate that **2** is an authentic signaling molecule involved in the orchestration of endospore formation in *R. cellulolyticum*.
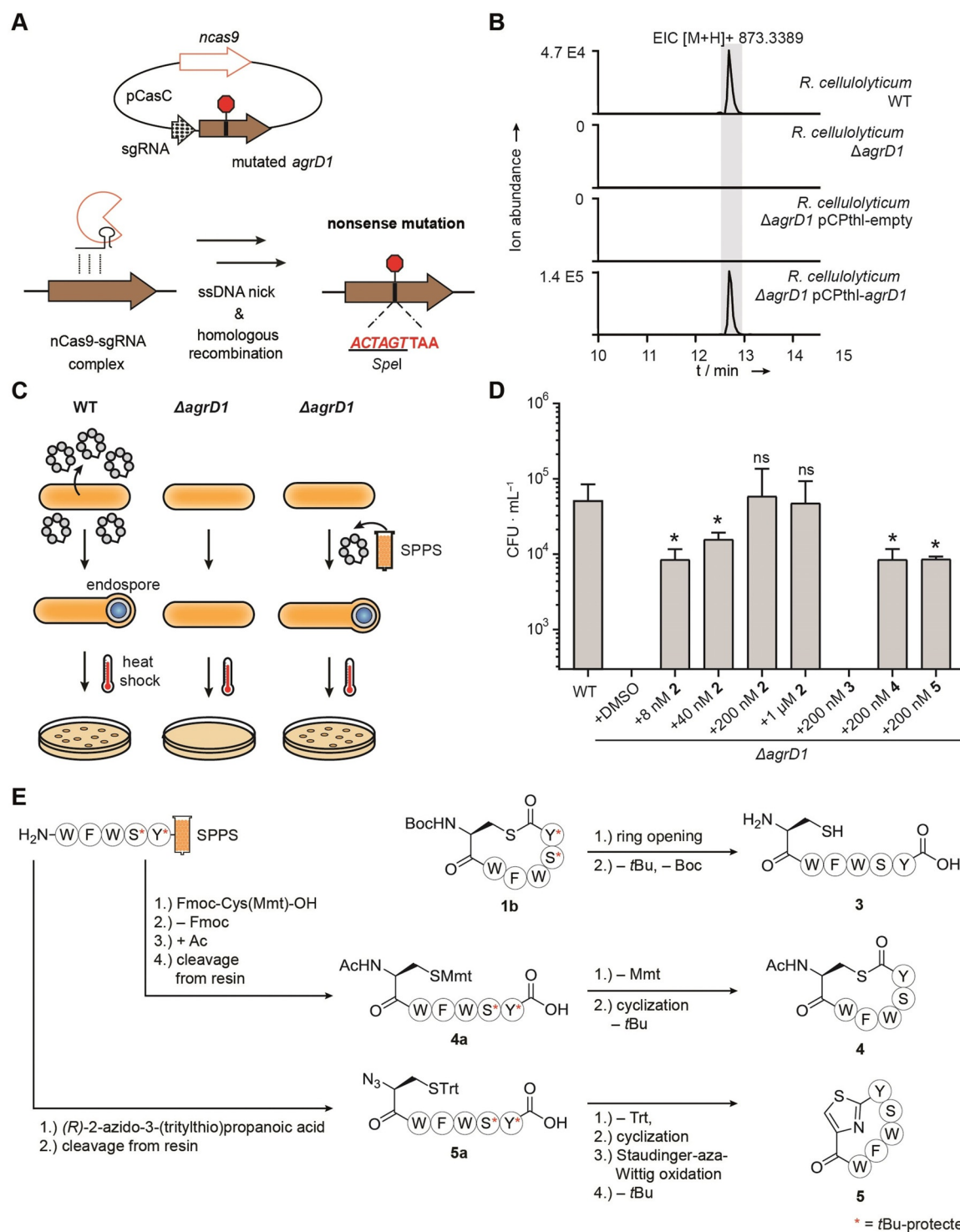
Figure 3. Control of endospore formation in *Ruminiclostridium cellulolyticum* by the cAIP *Rc*-AIP1 and evaluation of synthetic surrogates: A) Scheme depicting the incorporation of a nonsense mutation into *Rc-agrD1* (Ccel_2126) of *R. cellulolyticum* to generate the null mutant *R. cellulolyticum* Δ*agrD1*. The simultaneous introduction of a *Spe*I recognition site enabled the differentiation of successfully edited *R. cellulolyticum* from the wild type by restriction digest of colony PCR products. B) Extracted ion chromatogram (EIC) showing the presence or absence of *Rc*-AIP1 in the metabolite profiles of *R. cellulolyticum* wild type (WT), *R. cellulolyticum* Δ*agrD1*, *R. cellulolyticum* Δ*agrD1* pCPthI-empty (vector control) and *R. cellulolyticum* Δ*agrD1* pCPthI-*agrD1* (complemented mutant). C) Overview of the endospore formation assay. SSPS, solid-phase peptide synthesis. D) Quantification of endospore formation (colony forming units per milliliter, CFU · mL^{-1}) by heat-treated cultures of *R. cellulolyticum* wild type, *R. cellulolyticum* Δ*agrD1*, and *R. cellulolyticum* Δ*agrD1* supplemented with the listed concentrations of **2**, **3**, **4** and **5**. The results depicted are derived from at least three independent experiments. The colony counts for *R. cellulolyticum* Δ*agrD1* + DMSO (vehicle control) and *R. cellulolyticum* Δ*agrD1* + 200 nM **3** were below the quantification limit (300 CFU · mL^{-1}) of the assay. For the other assay conditions, the displayed values are the mean colony counts (CFU · mL^{-1}) with error bars showing the standard deviation. All values were compared to that of the wild type (ns, not significant; *, $p \leq 0.05$). DMSO, dimethyl sulfoxide. E) Synthetic route to compounds **3**–**5**. The linear sequences for **4** and **5** were produced in the same manner as for **1** and **2** (see Figure 2 C). The asterisk indicates the *t*Bu-protection of a hydroxyl group.

Synthetic surrogates reveal the importance of a macrocycle to Rc-AIP1 recognition

The rapid transformation of the thiolactone **1** into lactam **2** under culturing conditions precludes a direct comparison of their potencies in the endospore formation assay. In order to gain preliminary insights into the structure–activity relationship of *Rc*-AIP1, we prepared three analogues: a linear derivative **3**, an *N*-terminally acetylated derivative **4**, and a derivative **5** containing a peptidomimetic thiazole moiety at the macrocyclization site (Figure 3 E and Figure S5). *N*-acetylated analogue **4** features an increased ring size compared to the naturally occurring lactam **2**, but preserves the thiolactone moiety since it cannot undergo an $S \rightarrow N$ acyl shift. Thiazole-containing analogue **5** has the same ring size as **2**, but the risk of disulfide formation is eliminated. Moreover, it represents a stabilized analogue of the cAIP intermediate thiolactone **1** and is not prone to hydrolysis. Heterocyclic motifs are present in many natural products and their incorporation in cyclic peptides has led to molecules with improved properties.[24]

In brief, the linear peptide **3** was obtained by hydrolysis of the protected, SPPS-derived thiolactone **1b** with aqueous NaOH, followed by TFA treatment to remove the protecting groups (0.8 % overall yield starting from SPPS) We prepared the *N*-acetylated analogue **4** from the corresponding *N*-acetylated, fully-protected linear peptide **4a**, which was assembled on solid support. Treatment of **4a** with TFA and Et₃SiH yielded the ω-mercapto carboxylic acid, which was transformed into the thiolactone by using benzotriazole-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (Py-BOP) and DIPEA.[25] Global cleavage of the protecting groups was performed under acidic conditions to give thiolactone **4** (8.4 % overall yield starting from SPPS). The thiazole analogue **5** was synthesized from the fully protected linear peptide **5a**, which was assembled by SPPS in analogy to **1a** and **4a**, but with (*R*)-2-azido-3-(tritylthio)propanoic acid replacing the protected cysteine.[25] Upon deprotection, the ω-mercapto carboxylic acid was cyclized to **5** by using PyBOP as a coupling reagent and DIPEA as a base. The heterocyclization was promoted by means of a Staudinger-*aza*-Wittig sequence[25c] by using PPh₃ in 2,6-lutidine. The resulting thiazoline was oxidized by using 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and bromotrichloromethane. Subsequently the protecting groups were cleaved by TFA/Et₃SiH/H₂O (16 % overall yield starting from SPPS).[26]

With these structural analogues at hand, we tested their potencies in the endospore formation bioassay with *R. cellulolyticum* Δ*agrD1*. In stark contrast to the response to **2**, supplementation with 200 nM of the linear peptide **3** was indistinguishable from the phenotype of the untreated mutant (Figure 3 D). It is likely that the conformation needed for effective binding of *Rc*-AIP1 to its cognate receptor is favored by its cyclic nature compared to the conformationally more flexible linear peptide. Both macrocyclic analogues **4** and **5** were active in the endospore formation assay, albeit less so than **2**. Specifically, both **4** and **5** were only capable of partially complementing the sporulation deficient phenotype of *R. cellulolyticum* Δ*agrD1* at 200 nM, a level at which **2** completely

restored sporulation. As well as further supporting that a macrocyclic nature is important for recognition, the partial functionality of analogues **4** and **5** indicate that there is some variation tolerated in terms of linkage moiety and ring size of synthetic surrogates. However, the relative impact of these structural variations on biological function is difficult to distinguish. More elaborate structure-activity studies will be necessary to precisely evaluate the structural features that are crucial for *Rc*-AIP1 to adopt the conformation required for optimal interaction with its receptor.

Clostridial AIPs commonly undergo intramolecular $S \rightarrow N$ rearrangements

Prompted by the discovery that *Rc*-AIP1 is a potent QS signal, we wondered if the products of other clostridial Agr clusters might be homodetic cyclopeptides. We therefore mined the genomes of diverse Clostridia spanning four genera for genes encoding homologous Agr systems (Figure S13A and Table S1). From this bioinformatic analysis, we selected four strains from two genera for cultivation and extraction. After performing HPLC-HRMS analyses of the culture extracts, we scrutinized the metabolite profiles for signals consistent with the relevant mature AIPs having NTEs of up to four residues. We identified a number of candidates, all of which lacked an NTE. The sequences of the candidate AIPs were verified by HRMS²-based fragmentation (Figure S4 and Figure S14–18). Our detection of the AIP produced by *C. acetobutylicum* confirms the prediction that it is a 6-membered macrocyclic peptide without an NTE.[21]

Since the absence of an NTE is a prerequisite for the $S \rightarrow N$ acyl shift to occur, we presumed that the newly detected AIPs should be subject to NCL-like rearrangement to form cAIPs. However, such a connectivity difference would again be indistinguishable by HRMS². Although it has been proposed that the AIPs of *C. acetobutylicum*[2l] and *C. perfringens*[2j] would not have NTEs based on chemical complementation experiments, the resulting possibility of rearrangement was not explored. Our detection of numerous NTE-free AIPs allowed us to investigate whether the $S \rightarrow N$ acyl shift observed for *Rc*-AIP1 is a more general phenomenon guided by the innate chemical reactivity of the α-amino thioester.

In order to circumvent the need for synthetic references to finalize the structures for all candidate AIPs, we developed a method that could distinguish between a thiolactone or lactam moiety (Figure S19). We employed *N*-benzylmaleimide to chemoselectively label the free thiols that would only be present in the lactam forms and exploited the fact that cAIPs would be resistant to basic hydrolysis (Figure 4 A). The method was validated by way of synthetic standards (Figure S20–23) and by showing that *Rc*-AIP1 in a crude extract behaved as expected for a cAIP (Figure S24). When applied in a screen of the crude extracts containing candidate AIPs, a thiol adduct was detected in all (Figure 4 B–E), indicating that the cysteine side chain of each AIP had indeed been liberated by an $S \rightarrow N$ acyl shift. In accordance with the results obtained from *N*-benzylmaleimide labeling, no signals corresponding to the linearized forms were detected upon NaOH
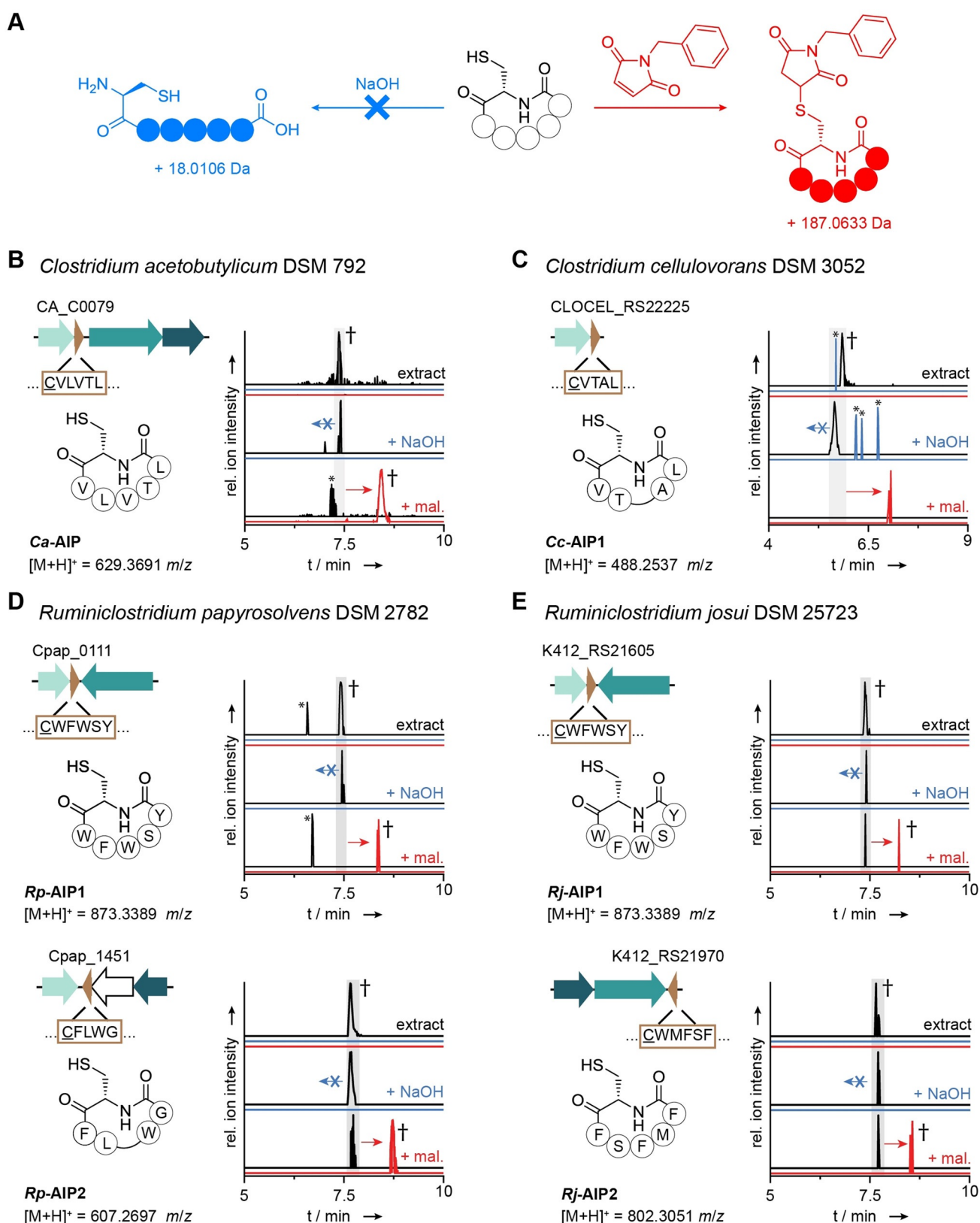
Figure 4. Identification of cAIPs in extracts of diverse Clostridia: A) General scheme depicting the reaction of a cAIP with *N*-benzylmaleimide resulting in the formation of a thiol adduct (red) and the absence of hydrolysis by NaOH (blue). B–E) After treatment with *N*-benzylmaleimide (+ mal) and NaOH, the AIPs detected in extracts of the named Clostridia behaved in a manner characteristic of homodetic peptides (cAIPs) (Figure 4A). Specifically, new species corresponding to the mal-adducts were detected (red), whereas no hydrolytic ring-openings were observed (blue). EICs are shown at either a 1.0 ppm window (panel B) or a 1.5 ppm window (panels C–E). † Further analysis by HRMS^2 experiments (untreated: Figure S4 and Figure S14–18, + mal: Figure S25–28). * Unrelated peaks detected in the crude extract.

treatment of the extracts (Figure 4B–E). Hence, we conclude that, like *Rc*-AIP1, these NTE-free AIPs from Clostridia are homodetic cyclopeptides and should be classified as cAIPs.

A biosynthetic route to cyclopeptides involving enzyme-primed native peptide ligation

Two in vivo routes for the formation of the cAIPs may be conceived (Figure 5A). By analogy to the staphylococcal Agr system, AgrB should cleave the follower peptide of AgrD and catalyze the formation of a thiolactone-containing macrocycle, which is then exported. Concomitant proteolytic cleavage of the leader peptide would liberate a free *N*-terminus that can undergo a spontaneous *S*→*N* acyl shift (route (i)). Alternatively, the leader peptide may be removed before the core peptide is processed by AgrB. The resulting free amine would then liberate the AgrB-bound intermediate by directly forming the amide bond (route (ii)), mirroring other RiPP maturation routes (e.g. cyanobactins).[27] Whereas there is no direct proof of a cryptic thiolactone intermediate, multiple lines of evidence support route (i) as the more probable pathway. Firstly, a cysteine residue at the cyclization position is indispensable for the formation of a cryptic thiolactone in route (i), whereas any amino acid would be suitable for route (ii). We and others noted that the cysteine residue is a conserved feature of clostridial AgrD peptides (Figure S13B).[2b,g,k,l,15] Secondly, in characterized Agr systems the leader peptide is removed after the thiolactone-containing

peptide is secreted from the cell.[5a] Route (ii) would be inconsistent with these data.

Taken together, these findings reveal a specialized biosynthetic pathway to cAIPs in Clostridia. By cleaving the follower peptide, AgrB generates a thioester-bound peptide and promotes an entropy-driven cyclization reaction to release the corresponding thiolactone. Liberation of the amine by cleavage of the NTE region sets the stage for a consecutive, maturing *S*→*N* acyl shift. This mechanism is strikingly similar to NCL strategies employed for the synthesis of (cyclo)peptides (Figure 5B).[28] In particular, the AgrB-primed pathway mirrors intramolecular variants of NCL that permit the on-resin synthesis of cyclopeptides.[29] The reactivities and mechanisms are identical in both the synthetic and biosynthetic routes, yet the synthetic peptide is linked to a bead, whereas the cAIP precursor is tethered to an enzyme. Spontaneous *S*→*N* acyl migrations have been reported in protein splicing pathways,[30] as well as in a recently discovered compensatory pathway for glutathione biosynthesis in bacteria lacking the GshA glutathione ligase.[31] By drawing on chemical logic, it had been suspected that the thiolactone formed from an unprotected *N*-terminal cysteine in a subset of AIPs could rearrange to yield an amide bond.[20,32] Our work describing the intramolecular, enzyme-primed NCL route to clostridial cAIPs now provides support for this proposal.

Notably, the propensity of thioesters to undergo exchange reactions and *S*→*N* acyl shifts was recently exploited in a discovery strategy for canonical thiolactone-possessing
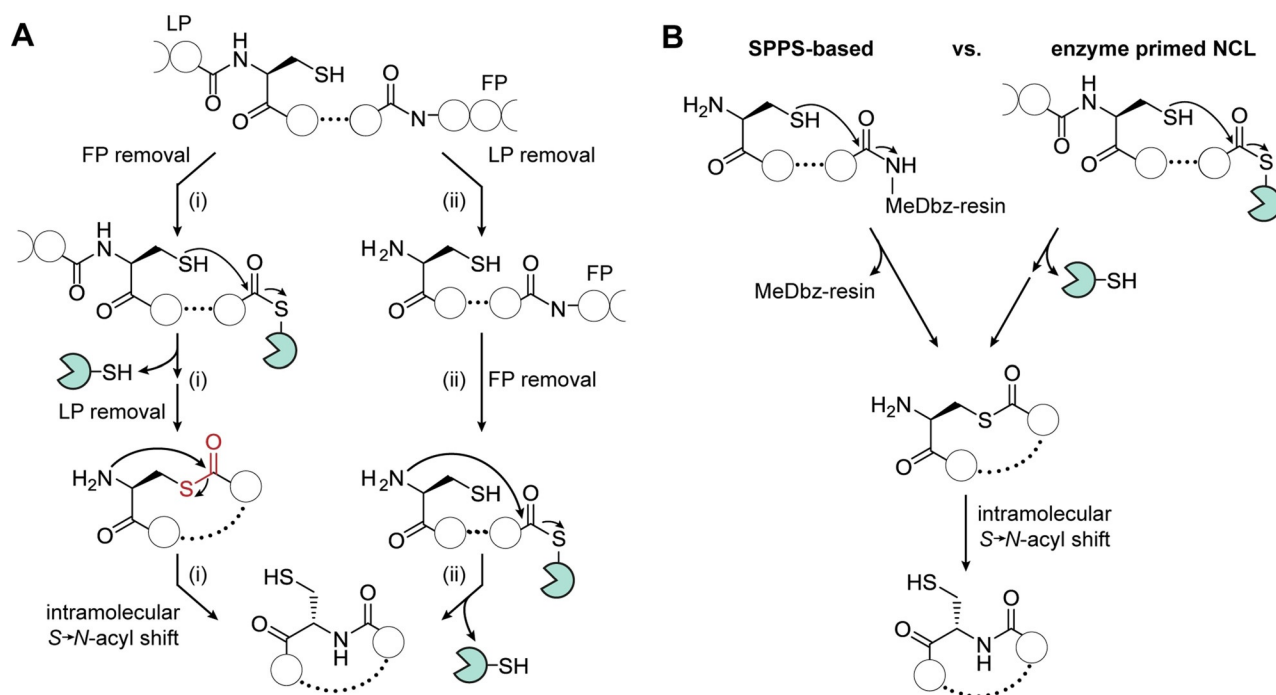

Figure 5. Model for cAIP biosynthesis: A) Two possible routes for in vivo cAIP formation involving either (i) a native chemical ligation (NCL)-like *S*→*N* shift or (ii) the direct formation of the lactam. FP, follower peptide; LP leader peptide. B) Comparison of of solid-phase peptide synthesis (SPPS)-based and enzyme-primed NCL strategies for cyclopeptide production. In chemical synthesis, direct attack of a free thiol leads to self-cleaving from the 3-amino-4-(methylamino)benzoic acid (MeDbz)-linked resin yielding the thiolactone. In cAIP biosynthesis, the same product is produced by the attack of a free thiol to the thioester. The intermediary thiolactones undergo intramolecular *S*→*N* acyl shifts to give the corresponding cyclopeptides.

AIPs produced in culture.[20,32] Efforts to apply the NCL-based technique to identify the NTE-free AIP of the pathogen *Listeria monocytogenes*[33] were unsuccessful.[20] On the basis of our findings, we propose, as did the authors, that the mature AIP presents as the lactam form, meaning it would be incompatible with chemoselective trapping by NCL. We speculate that the NTE-free AIP of the human-associated *Lactobacillus plantarum*,[34] an important anaerobe used for food fermentation, could also be a cAIP. The fate of the lactone in NTE-free AIPs cyclized through a serine residue, such as that of *H. thermocellum*, remains to be clarified.[19]

Finally, it should be highlighted that this previously overlooked natural NCL-like cyclopeptide formation adds to the known pathways for RiPP maturation. Previously, ATP-dependent and protease-dependent avenues have been reported for cyclopeptide biosynthesis in diverse RiPP families,[27,35] with a maturation enzyme installing the lactam bond in each case. The mechanism for cAIP macrolactamization described herein markedly differs from these routes and represents a novelty in the RiPP biosynthetic toolbox.

Conclusion

Many genetic studies have indicated that Agr-derived peptides mediate important processes in Clostridia. However, the true nature of the implicated signaling molecules has remained uncertain. In this study, we provide the first rigorously determined structure of a clostridial AIP. On the basis of chemical synthesis and meticulous analysis, we ascertained that the signals detected in clostridial cultures do not correspond to thiolactones; they are homodetic cyclopeptides or cAIPs that result from enzyme-primed $S \rightarrow N$ acyl shifts. In the case of a model clostridial cAIP, *Rc*-AIP1 produced by *R. cellulolyticum*, kinetic analyses exclude the possibililty that the thiolactone precursor could play a significant biological role, which is also supported by chemical complementation experiments.

Until now, it was believed that cyclization through a thioester or ester was a conserved feature of AIPs.[4,5c] In expanding the chemical diversity of the AIP family to include cAIPs, this discovery highlights the importance of rigorous chemical analyses and synthesis, and serves as a cautionary tale regarding cursory bioinformatics-based assumptions about the structures of natural products. Moreover, the structures and pathways described herein are an unexpected divergence from the canonical AIP biosynthetic routes and prompt a revision of the clostridial AIPs and their formation. This novel route to cyclic RiPPs via cryptic thiolactone intermediates may be considered as the natural equivalent of an intramolecular NCL. From a translational point of view, identification of the cAIP structures opens new possibilities to control anaerobic QS-regulated processes using synthetic surrogates in biotechnology and remediation. Providing the homodetic cyclopeptide architecture is conserved in the products of the Agr systems implicated in the virulence of pathogenic Clostridia,[2a–k,m] this knowledge could also assist current research on QS antagonists as therapeutics.[36]

Conflict of interest

The authors declare no conflict of interest.

[1] P. Cruz-Morales, C. A. Orellana, G. Moutafis, G. Moonen, G. Rincon, L. K. Nielsen, E. Marcellin, *Genome Biol. Evol.* **2019**, *11*, 2035–2044.

[2] a) C. Darkoh, C. Odo, H. L. DuPont, *mBio* **2016**, *7*, e01237-16; b) C. Darkoh, H. L. DuPont, S. J. Norris, H. B. Kaplan, *mBio* **2015**, *6*, e02569-14; c) M. J. Martin, S. Clare, D. Goulding, A. Faulds-Pain, L. Barquist, H. P. Browne, L. Pettit, G. Dougan, T. D. Lawley, B. W. Wren, *J. Bacteriol.* **2013**, *195*, 3672–3681; d) J. Chen, B. A. McClane, *Infect. Immun.* **2012**, *80*, 3008–3017; e) J. Chen, J. I. Rood, B. A. McClane, *mBio* **2011**, *2*, e00275-11; f) J. Li, J. Chen, J. E. Vidal, B. A. McClane, *Infect. Immun.* **2011**, *79*, 2451–2459; g) K. Ohtani, Y. Yuan, S. Hassan, R. Wang, Y. Wang, T. Shimizu, *J. Bacteriol.* **2009**, *191*, 3919–3927; h) J. E. Vidal, M. Ma, J. Saputo, J. Garcia, F. A. Uzal, B. A. McClane, *Mol. Microbiol.* **2012**, *83*, 179–194; i) J. E. Vidal, J. Chen, J. Li, B. A. McClane, *PLoS One* **2009**, *4*, e6232; j) M. Ma, J. Li, B. A. McClane, *J. Bacteriol.* **2015**, *197*, 1807–1818; k) C. M. Cooksley, I. J. Davis, K. Winzer, W. C. Chan, M. W. Peck, N. P. Minton, *Appl. Environ. Microbiol.* **2010**, *76*, 4448–4460; l) E. Steiner, J. Scott, N. P. Minton, K. Winzer, *Appl. Environ. Microbiol.* **2012**, *78*, 1113–1122; m) M. A. Navarro, J. Li, J. Beingesser, B. A. McClane, F. A. Uzal, *mSphere* **2020**, *5*, e00500-20; n) T. Kosaka, S. Nakayama, K. Nakaya, S. Yoshino, K. Furukawa, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 58–68; o) J. Feng, W. Zong, P. Wang, Z. T. Zhang, Y. Gu, M. Dougherty, I. Borovok, Y. Wang, *Biotechnol. Biofuels* **2020**, *13*, 84; p) A. K. Kotte, O. Severn, Z. Bean, K. Schwarz, N. P. Minton, K. Winzer, *Microbiology* **2020**, *166*, 579–592.

[3] a) C. M. Waters, B. L. Bassler, *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 319–346; b) P. Williams, K. Winzer, W. C. Chan, M. Cámara, *Philos. Trans. R. Soc. London Ser. B* **2007**, *362*, 1119–1134.

[4] P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Göransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Müller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P.

Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Süssmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey, W. A. van der Donk, *Nat. Prod. Rep.* **2013**, *30*, 108–160.

[5] a) R. P. Novick, E. Geisinger, *Annu. Rev. Genet.* **2008**, *42*, 541–564; b) V. Monnet, V. Juillard, R. Gardan, *Crit. Rev. Microbiol.* **2016**, *42*, 339–351; c) M. Thoendel, J. S. Kavanaugh, C. E. Flack, A. R. Horswill, *Chem. Rev.* **2011**, *111*, 117–151; d) J. K. Vasquez, K. H. J. West, T. Yang, T. J. Polaske, G. Cornilescu, M. Tonelli, H. E. Blackwell, *J. Am. Chem. Soc.* **2020**, *142*, 750–761; e) T. Yang, Y. Tal-Gan, A. E. Paharik, A. R. Horswill, H. E. Black-well, *ACS Chem. Biol.* **2016**, *11*, 1982–1991; f) Y. Tal-Gan, M. Ivancic, G. Cornilescu, H. E. Blackwell, *Org. Biomol. Chem.* **2016**, *14*, 113–121.

[6] S. Y. Queck, M. Jameson-Lee, A. E. Villaruz, T. H. Bach, B. A. Khan, D. E. Sturdevant, S. M. Ricklefs, M. Li, M. Otto, *Mol. Cell* **2008**, *32*, 150–158.

[7] L. Zhang, J. Lin, G. Ji, *J. Biol. Chem.* **2004**, *279*, 19448–19456.

[8] G. Ji, R. C. Beavis, R. P. Novick, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12055–12059.

[9] M. Thoendel, A. R. Horswill, *J. Biol. Chem.* **2009**, *284*, 21828–21838.

[10] a) L. Zhang, L. Gray, R. P. Novick, G. Ji, *J. Biol. Chem.* **2002**, *277*, 34736–34742; b) R. Qiu, W. Pei, L. Zhang, J. Lin, G. Ji, *J. Biol. Chem.* **2005**, *280*, 16695–16704.

[11] B. Wang, A. Zhao, R. P. Novick, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10679–10684.

[12] J. S. Kavanaugh, M. Thoendel, A. R. Horswill, *Mol. Microbiol.* **2007**, *65*, 780–798.

[13] M. Otto, R. Süssmuth, G. Jung, F. Götz, *FEBS Lett.* **1998**, *424*, 89–94.

[14] a) J. Nakayama, Y. Cao, T. Horii, S. Sakuda, A. D. Akkermans, W. M. de Vos, H. Nagasawa, *Mol. Microbiol.* **2001**, *41*, 145–154; b) J. Nakayama, S. Chen, N. Oyama, K. Nishiguchi, E. A. Azab, E. Tanaka, R. Kariyama, K. Sonomoto, *J. Bacteriol.* **2006**, *188*, 8321–8326.

[15] A. Wuster, M. M. Babu, *J. Bacteriol.* **2008**, *190*, 743–746.

[16] M. Desvaux, *FEMS Microbiol. Rev.* **2005**, *29*, 741–764.

[17] a) T. Lincke, S. Behnken, K. Ishida, M. Roth, C. Hertweck, *Angew. Chem. Int. Ed.* **2010**, *49*, 2011–2013; *Angew. Chem.* **2010**, *122*, 2055–2057; b) A. I. Chiriac, F. Kloss, J. Krämer, C. Vuong, C. Hertweck, H. G. Sahl, *J. Antimicrob. Chemother.* **2015**, *70*, 2576–2588; c) K. L. Dunbar, M. Dell, F. Gude, C. Hertweck, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 8850–8858.

[18] a) I. Fedorova, A. Arseniev, P. Selkova, G. Pobegalov, I. Goryanin, A. Vasileva, O. Musharova, M. Abramova, M. Kazalov, T. Zyubko, T. Artamonova, D. Artamonova, S. Shmakov, M. Khodorkovskii, K. Severinov, *Nucleic Acids Res.* **2020**, *48*, 2026–2034; b) T. Xu, Y. Li, Z. Shi, C. L. Hemme, Y. Li, Y. Zhu, J. D. Van Nostrand, Z. He, J. Zhou, *Appl. Environ. Microbiol.* **2015**, *81*, 4423–4431.

[19] T. J. Verbeke, R. J. Giannone, D. M. Klingeman, N. L. Engle, T. Rydzak, A. M. Guss, T. J. Tschaplinski, S. D. Brown, R. L. Hettich, J. G. Elkins, *Sci. Rep.* **2017**, *7*, 43355.

[20] B. H. Gless, M. S. Bojer, P. Peng, M. Baldry, H. Ingmer, C. A. Olsen, *Nat. Chem.* **2019**, *11*, 463–469.

[21] P. W. Riddles, R. L. Blakeley, B. Zerner, *Anal. Biochem.* **1979**, *94*, 75–81.

[22] K. C. B. Jennert, C. Tardif, D. I. Young, M. Young, *Microbiology* **2000**, *146 Pt 12*, 3071–3080.

[23] P. MDowell, Z. Affas, C. Reynolds, M. T. Holden, S. J. Wood, S. Saint, A. Cockayne, P. J. Hill, C. E. Dodd, B. W. Bycroft, W. C. Chan, P. Williams, *Mol. Microbiol.* **2001**, *41*, 503–512.

[24] I. V. Smolyar, A. K. Yudin, V. G. Nenajdenko, *Chem. Rev.* **2019**, *119*, 10032–10240.

[25] a) A. Oberheide, S. Pflanze, P. Stallforth, H.-D. Arndt, *Org. Lett.* **2019**, *21*, 729–732; b) S. Schwenk, C. Ronco, A. Oberheide, H.-D. Arndt, *Eur. J. Org. Chem.* **2016**, 4795–4799; c) M. Riedrich, S. Harkal, H.-D. Arndt, *Angew. Chem. Int. Ed.* **2007**, *46*, 2701–2703; *Angew. Chem.* **2007**, *119*, 2755–2758.

[26] A. J. Phillips, Y. Uto, P. Wipf, M. J. Reno, D. R. Williams, *Org. Lett.* **2000**, *2*, 1165–1168.

[27] C. Ongpipattanakul, S. K. Nair, *Biochemistry* **2018**, *57*, 3201–3209.

[28] a) H. M. Burke, L. McSweeney, E. M. Scanlan, *Nat. Commun.* **2017**, *8*, 15655; b) C. J. White, A. K. Yudin, *Nat. Chem.* **2011**, *3*, 509–524; c) J. P. Tam, C. T. Wong, *J. Biol. Chem.* **2012**, *287*, 27020–27025; d) V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, *Chem. Rev.* **2019**, *119*, 7328–7443; e) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776–779.

[29] a) J. Tulla-Puche, G. Barany, *J. Org. Chem.* **2004**, *69*, 4101–4107; b) G. Serra, L. Posada, H. Hojo, *Chem. Commun.* **2020**, *56*, 956–959; c) C.-F. Liu, J. P. Tam, *J. Am. Chem. Soc.* **1994**, *116*, 4149–4153.

[30] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447–496.

[31] K. Veeravalli, D. Boyd, B. L. Iverson, J. Beckwith, G. Georgiou, *Nat. Chem. Biol.* **2011**, *7*, 101–105.

[32] D. N. McBrayer, Y. Tal-Gan, *Nat. Chem.* **2019**, *11*, 398–399.

[33] M. Zetzmann, A. Sanchez-Kopper, M. S. Waidmann, B. Blombach, C. U. Riedel, *Front. Microbiol.* **2016**, *7*, 989.

[34] M. H. Sturme, J. Nakayama, D. Molenaar, Y. Murakami, R. Kunugi, T. Fujii, E. E. Vaughan, M. Kleerebezem, W. M. de Vos, *J. Bacteriol.* **2005**, *187*, 5224–5235.

[35] M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, M. J. Grande Burgos, C. Hill, S. Kim, J. Koehnke, J. A. Latham, A. J. Link, B. Martínez, S. K. Nair, Y. Nicolet, S. Rebuffat, H. G. Sahl, D. Sareen, E. W. Schmidt, L. Schmitt, K. Severinov, R. D. Süssmuth, A. W. Truman, H. Wang, J. K. Weng, G. P. van Wezel, Q. Zhang, J. Zhong, J. Piel, D. A. Mitchell, O. P. Kuipers, W. A. van der Donk, *Nat. Prod. Rep.* **2021**, *38*, 130–239.

[36] a) B. LaSarre, M. J. Federle, *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 73–111; b) A. M. Salam, C. L. Quave, *mSphere* **2018**, *3*, e00500-17; c) Y. Tal-Gan, D. M. Stacy, M. K. Foegen, D. W. Koenig, H. E. Blackwell, *J. Am. Chem. Soc.* **2013**, *135*, 7869–7882; d) J. K. Vasquez, H. E. Blackwell, *ACS Infect. Dis.* **2019**, *5*, 484–492; e) Y. Tal-Gan, M. Ivancic, G. Cornilescu, T. Yang, H. E. Blackwell, *Angew. Chem. Int. Ed.* **2016**, *55*, 8913–8917; *Angew. Chem.* **2016**, *128*, 9059–9063; f) J. K. Vasquez, Y. Tal-Gan, G. Cornilescu, K. A. Tyler, H. E. Blackwell, *ChemBioChem* **2017**, *18*, 413–423.