Natural Product Discovery from Solventogenic Clostridia

By

Jeffrey S Li

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wenjun Zhang, Chair
Professor Sanjay Kumar
Professor Matt F. Traxler

Spring 2020

Abstract

Natural Product Discovery from Solventogenic Clostridia

by

Jeffrey S Li

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Wenjun Zhang, Chair

Secondary metabolites, or natural products, are small molecules which frequently possess biological activities and are a critically important resource for the development of new pharmaceuticals, pigments, and agrochemicals. The need for new natural product sources has led to the genomics-driven identification of the anaerobes as a promising and relatively untapped reservoir of natural product diversity. Soil isolates such as the industrially significant solventogenic clostridia are obligate anaerobes which can ferment carbohydrates to produce acetone and butanol. These bacteria are an especially promising subset of the anaerobes for discovery of natural products. This work begins with a description of the limited repertoire of known natural products isolated from anaerobes, showcasing their chemical diversity and broad range of bioactivity. Then, we describe genome mining efforts in different solventogenic clostridia. In *Clostridium saccharoperbutylacetonicum*, we describe genome and transcriptome profiling of secondary metabolite biosynthetic gene clusters (BGCs) to enable in-depth characterization of a nonribosomal peptide synthetase conserved in other solventogenic clostridia. This led to the discovery of a novel compound, an *N*-acyl dipeptidyl alcohol. We further discuss the association of the biosynthetic gene with butanol tolerance, a phenotype with industrial significance. In *Clostridium roseum*, we describe methods for gene delivery to enable targeting of BGCs for knockout and characterization. The products of one BGC are reported, a novel family of unusual solvatochromic compounds named the clostyrylpyrones. The biosynthesis of these compounds is proposed. In a broader survey of secondary metabolism in clostridia, we describe efforts in a panel of ten species of clostridia to identify antimicrobial potential using a traditional disc-diffusion approach. We also describe, for two of the strains, implementation of a high-throughput media screening study to identify BGC-specific chemical inducers of secondary metabolism. Overall, this work expands the known chemical diversity of natural products derived from anaerobes, supports the value of targeting these bacteria for genome mining, and extends our understanding of secondary metabolism in solventogenic clostridia.

Table of Contents

List of Figures ..... **iii**
List of Tables ..... **v**
Acknowledgements ..... **vi**
Chapter 1. Introduction **1**
1.1. Anaerobic bacteria as producers of natural products 1
1.2. Chemical diversity of the anaerobes 4
1.2.1. Compounds discovered by genomics-independent methods 4
1.2.2. Compounds discovered by genomics-dependent methods 8
1.3. Solventogenic clostridia as producers of secondary metabolites 12
Chapter 2. Investigation of secondary metabolism in *Clostridium saccharoperbutylacetonicum* N1-4 ..... **14**
2.1. Introduction 14
2.2. Results 15
2.2.1. Bioinformatic analysis of PK/NRP BGCs in *Csa* 15
2.2.2. Transcriptional profiling of PK/NRP BGCs during ABE fermentation 17
2.2.3. Characterization of *nrps3*-associated secondary metabolite 17
2.3. Discussion 24
2.4. Materials and Methods 26
Chapter 3. Genome mining in *Clostridium roseum* **30**
3.1. Introduction 30
3.2. Results 31
3.2.1. Bioinformatics of *Cro* 31
3.2.2. Transcriptional analyses of *Cro* 37
3.2.3. Establishing a genetic toolkit for *Cro* BGC knockouts 39
3.2.4. Identification of metabolites associated with BGC 1 46
3.2.5. Discovery and structural elucidation of metabolites associated with BGC 14a 50
3.2.6. Biosynthesis of the clostyrylpyrone scaffold 54
3.2.7. Preliminary antioxidant activity of clostyrylpyrones 58
3.2.8. Preliminary growth rate inhibition assay of clostyrylpyrones 59
3.3. Discussion 60

3.4. Materials and Methods 61

Chapter 4. Culture variation methods for cryptic biosynthetic gene cluster activation in clostridia for novel antibiotic discovery 68

4.1. Introduction 68

4.2. Results 70

4.2.1. Selection of *Clostridium* strains 70

4.2.2. Antibiotic activity screening 72

4.2.3. Design of a high-throughput media additive screening strategy for BGC activation in *Csa* and *Cbei* 75

4.3. Discussion 87

4.4. Materials and Methods 88

Chapter 5. Conclusions 93

References 96

Appendix A 112

Appendix B 130

List of Figures

Figure 1-1. Representation of bacterial anaerobic taxa by percentage of overall genomic sequence and of overall PKS/NRPS-containing BGC sequence. 3

Figure 1-2. Secondary metabolites discovered from anaerobes by traditional methods 5

Figure 1-3. Secondary metabolites discovered from anaerobes by genome mining. 9

Figure 2-1. BGCs of *Csa* which contain PK or NRP biosynthetic genes. 15

Figure 2-2. Homologs of *nrps3*. 16

Figure 2-3. Expression level of the BGCs of *Csa*. 17

Figure 2-4. Mutant generation in *Csa*. 18

Figure 2-5. HRMS characterization of compound 1. 19

Figure 2-6. Characterization of *nrps3*-associated secondary metabolite, 1. 20

Figure 2-7. SDS-PAGE gel of the purified NRPS. 21

Figure 2-8. Residual glucose in ABE fermentation cultures of *Csa* wild-type and Δ*nrps3*. 22

Figure 2-9. Comparison of wild-type *Csa* and Δ*nrps3* in batch cultures under ABE fermentation conditions. 23

Figure 2-10. Butanol challenge assays with feeding of compound 1 to Δ*nrps3* groups 24

Figure 2-11. Expression level of BGCs of *Csa*. 25

Figure 3-1. BGCs of *Cro* which contain PKS or NRPS genes. 31

Figure 3-2. Genes of BGC 1. 32

Figure 3-3. Genes of BGC 2. 33

Figure 3-4. Genes of BGC 7. 34

Figure 3-5. Genes of BGC 9. 34

Figure 3-6. Genes of BGC 14a. 35

Figure 3-7. Genes of BGC 14b. 36

Figure 3-8. Genes of BGC 24. 36

Figure 3-9. Genes of BGC 84 37

Figure 3-10. Transcriptional analysis of *Cro* BGCs in colony. 38

Figure 3-11. Generation of *Cro pyrF* by ClosTron system 41

Figure 3-12. Tween 80 surfactant reduces cell aggregation of *Cro*. 42

Figure 3-13. Isolation of ClosTron targeted BGC knockouts. 44

Figure 3-14. Generation of *Cro* ΔBGC14a. 45

Figure 3-15. Metabolites associated with BGC 1.. 46
Figure 3-16. HRMS characterization of BGC 1 metabolites after precursor feeding..... 47
Figure 3-17. Tandem HRMS characterization of BGC 1 associated metabolites.. 48
Figure 3-18. Identification of secondary metabolites associated with BGC 14a..... 50
Figure 3-19. Chromism of BGC 14a associated Peak I metabolites..... 51
Figure 3-20. Proposed biosynthesis of clostyrylpyrones. 55
Figure 3-21. Aryl acid biosynthesis pathway from *Cro*..... 57
Figure 3-22. Heterologous expression of *Cro* BGC 14a in *Csa*. 58
Figure 3-23. Hydrogen peroxide disc diffusion assays..... 58
Figure 3-24. Growth rate inhibition assay of compound 1 and 2. 59
Figure 4-1. Antibiotics isolated from clostridia..... 69
Figure 4-2. Workflow for genetic reporter mediated high-throughput media screening..... 75
Figure 4-3. Relative performance of fluorescent reporters in *Cbei* and *Csa*. 81
Figure 4-4. Fluorescence of tLOV over time in cultures of *Cbei* or *Csa*..... 82
Figure 4-5. Benchmarking of fluorescence against known *nrps3* metabolite..... 83
Figure 4-6. Elicitor screening results in *Cbei*.. 85
Figure 4-7. Elicitor screening results in *Csa*..... 86

List of Tables

Table 1-1. Examples of large PKS- and/or NRPS-encoding BGCs detected in the genomes of anaerobes 2

Table 1-2. Examples of secondary metabolite biosynthetic potential in anaerobic bacteria 4

Table 2-1. Strains and plasmids used in Chapter 2 26

Table 2-2. Synthetic oligonucleotides used in Chapter 2 27

Table 3-1. Combinations of plasmid parts tested in *Cro* 39

Table 3-2. *Cro* BGC knockout targets. 43

Table 3-3. NMR spectroscopic data for compounds 1, 2 in DMSO-d_6 and compounds 3, 4 in CD_3OD (900 MHz) 53

Table 3-4. Strains used in Chapter 3 61

Table 3-5. Synthetic oligonucleotides used in Chapter 3 62

Table 3-6. PCR reactions used in Chapter 3 64

Table 3-7. Plasmid constructs used in Chapter 3. 65

Table 4-1. Predicted BGCs in the *Clostridium* strains utilized in this study. 71

Table 4-2. Growth characteristics of selected *Clostridium* strains 72

Table 4-3. Disc diffusion assay results for extracts from *Clostridium* cultures. 73

Table 4-4. Selected BGC promoter sequences for screening. 76

Table 4-5. Selected elicitors for screening. 78

Table 4-6. Primers used in Chapter 4 90

Table 4-7. PCR reactions used in Chapter 4 91

Acknowledgements

Foremost, I would like to acknowledge my advisor, Professor Wenjun Zhang, for her years of support and guidance in my research and scientific training. I have learned so much from her firm, understanding, and effective leadership, and I aspire to attain her scientific acumen in my professional career.

I would like to acknowledge my colleagues and friends from the Zhang lab for good company, sound advice, various scientific contributions, and persistent invitations to coffee hour: Frederick Twigg, Will Skyrud, Colin Barber, Nick Zill, Zhongrui Li, Antonio Del Rio Flores, and Di Gu. I would like to thank Drs. Nico Herman, Joyce Liu, Nick Harris, and Xuejun Zhu for their early mentorship and supportive presence in the lab. I would like to thank Drs. Ella Zafrir, Seong Jong Kim, Wenlong Cai, Will Porterfield, Zhijuan Hu, and Yongle Du for their chemistry expertise and supportive interactions. I would like to acknowledge my undergraduate students who demonstrated great enthusiasm, curiosity, and diligence: Katherine Zhou, Ted Gao, Alex Leung, Sam Ng, and Alli Green. Additional thanks to my SYIP students for their contagious energy and outrageous memes: Zach Danner, Varsha Subramanyam, Andrew Chan, and Stephen Jiang.

I would like to thank George McClellan and the late Dr. John Utting for instilling my early interest in biology. I would like to acknowledge Drs. Jonathan Gastel and Andy Ng and Professors Shengping Zheng (Hunter College), Anthony Hay (Cornell University) and Virginia Cornish (Columbia University) for providing me with formative experiences in scientific research.

I would like to thank my parents, my sister, and my grandma for giving me the motivation and encouragement to complete my graduate journey.

For many fond memories of the Bay Area, I would like to thank my Berkeley family: Steven DelaCruz and Ivón Padilla-Rodriguez, Sam Leung, Frederick Twigg and Qinlei Liu (and Watson Twigg), Philomena Weng, and Marianne Sleiman. Special thanks to Steven and Ivón for their company during the pandemic lockdown. I would like to thank Zhi Jian Lin and Oliver Lu for witty and refreshing life commentary.

I would like to thank my partner Jessie and dedicate this work to her.

Chapter 1. Introduction

Parts of this chapter have been adapted from the following with permission: Li, J. S., Barber, C. C., Zhang, W. "Natural products from anaerobes." *J. Ind. Microbiol. Biotechnol.* **46**, 375-383 (2019).

1.1. Anaerobic bacteria as producers of natural products

Natural products, often known as secondary metabolites, are structurally diverse small molecules possessing diverse and potent biological activities. Due to the metabolic burden inherent to natural products biosynthesis, these compounds are believed to be shaped by evolutionary pressure to fulfill specific biological functions.[1] For example, natural products produced by microbes have been demonstrated to influence processes as diverse as virulence, motility, stress response, biofilm formation, morphological differentiation, nutrient acquisition, and defense.[2] Elucidating the identity and function of these microbial natural products may yield valuable biological insights or tools for modulating critical biological processes and promoting biotechnological applications of producing microbes. Historically, natural products are also valuable as medicinal compounds, and have been used to treat various human health conditions such as cancer, infectious disease, autoimmune disorder, cardiovascular disease, and neurological disease.[3]

Modern natural product discovery can be guided by genomics, which enables detection of physically colocalized biosynthetic gene clusters (BGCs) which inform a microbe's ability to produce natural products. The major classes of BGCs include the thiotemplated assembly-line natural products (polyketides/nonribosomal peptides), terpenes, and ribosomally produced and post-translationally modified products (RiPPs). From a discovery standpoint, these classes of natural products are appealing because their BGCs possess conserved and detectable genomic signatures.[4] Despite this shared biosynthetic origin, natural products of these classes retain high diversity of both chemical structure and biological function. As genomic data continues to become widely available, new classes of microbes have been discovered to be promising reservoirs of natural products, including anaerobes, pathogens, and symbionts of humans or insects or nematodes.[5] These new reservoirs are of interest for their potential to address the high incidence of compound rediscovery, a problem in natural product discovery which may be partially attributed to screening bias toward well-known natural product producers such as filamentous actinomycetes and fungi. Anaerobes especially have been largely neglected in natural product discovery efforts, with few compounds identified, despite the first report of anaerobic production of antibiotics by bacteria over half a century ago.[6]

The idea of anaerobes as promising natural product producers is supported by genomic analysis,[7,8] although their total genetic capacity for secondary metabolite biosynthesis lags that of the most "gifted" aerobes.[9] A 2013 survey of 211 complete anaerobic bacterial genomes revealed that natural product BGCs, in particular those involved in polyketide and non-ribosomal peptide biosynthesis, could be found in 33% of the analyzed genomes.[7] As exemplified in **Table 1-1**, some BGCs contain over 60 kb of genes encoding polyketide synthases (PKSs) and/or non-ribosomal peptide synthetases (NRPSs), making them comparable in size to well-known antibiotic BGCs characterized in aerobes.[7] It is notable that these BGCs are not equally distributed in various phyla, with Firmicutes of the genus *Clostridium* and Deltaproteobacteria possessing a relatively greater

potential in anaerobes for natural product biosynthesis (**Figure 1-1**). Additional examples of bacteria from these taxa which contain many secondary metabolite BGCs are presented in (**Table 1-2**). Interestingly, genomic potential for secondary metabolism is also correlated with the isolation site of the anaerobic organism; strains with the most BGCs predominantly originated from soil.[7] Despite the established genomic potential, chemical data on natural products from the anaerobic world are very limited and most BGCs identified from genomic analysis do not have an associated product. To date, only a handful of natural products have been isolated from anaerobes and structurally characterized.

Table 1-1. Examples of large PKS- and/or NRPS-encoding BGCs detected in the genomes of anaerobes. PKS, polyketide synthase. NRPS, non-ribosomal peptide synthetase. BGC, biosynthetic gene cluster.

Organism	BGC Type	# of PKS Modules	# of NRPS Modules	Total PKS/NRPS Sequence (kb)
Ruminiclostridium cellulolyticum H10	Hybrid	14	1	63.75
	Hybrid	3	8	48.1
	Hybrid	6	9	60.98
Clostridium cellulovorans 743B	NRPS	-	18	65.71
	Hybrid	1	9	33.72
Clostridium botulinum A2 BoNT/Kyoto-F	NRPS	-	9	36.83
Clostridium botulinum H04402 065	NRPS	-	7	29.53
Clostridium kluyveri DSM 555	Hybrid	2	6	35.99
Ruminococcus albus 7 ATCC 27210	Hybrid	3	6	35.16
Geobacter uraniireducens $Rf4$	PKS	9	-	35.12
Opitutus terrae PB90-1	Hybrid	6	11	68.88

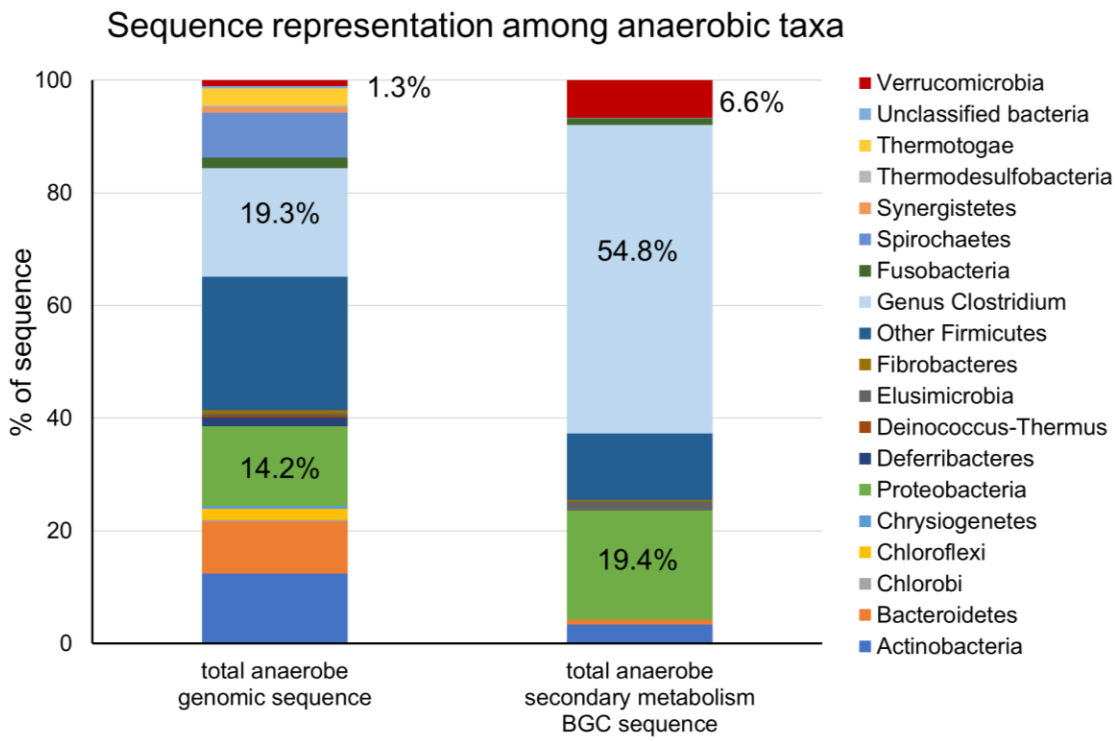


Figure 1-1. Representation of bacterial anaerobic taxa by percentage of overall genomic sequence and of overall PKS/NRPS-containing BGC sequence. PKS/NRPS biosynthetic potential is enriched in the genus *Clostridium*, in Deltaproteobacteria, and in Verrucomicrobia (relative). Calculated from results of a survey of 211 complete anaerobe genomes[7]

Table 1-2. Examples of secondary metabolite biosynthetic potential in anaerobic bacteria. BGCs from published genomes were detected using AntiSMASH 4.0 with ClusterFinder "off".[10] The selected species are from two of the most promising phyla, the Firmicutes and Deltaproteobacteria. PK, polyketide. NRP, non-ribosomal peptide. RiPP, ribosomally produced and post-translationally modified peptide. BGC, biosynthetic gene cluster. SM, secondary metabolite. [a]Denotes BGCs for PK/NRPs identified from draft genomes, which are often fragmented or misassembled in draft genomes so the number of BGCs can be overrepresented.[9,11] [b]As *Clostridium* is highly polyphyletic, these species have been reassigned to the suggested genus *Ruminiclostridium* according to the NCBI. [c]Putative ladderane biosynthesis clusters. [d]Putative AMP-binding domain containing clusters. [e]Putative arylpolyene biosynthesis clusters. [f]Putative acyl amino acid biosynthesis cluster. [g]Putative homoserine lactone biosynthesis cluster

Organism	Genome Size (Mb)	Predicted PK/NRP	Predicted RiPP	Other	Predicted Total SM	Total BGC Sequence (Mb)	% of Genome
Clostridium papyrosolvens DSM 2782[a,b]	4.92	13	3	2[d]	18	0.77	15.7
Clostridium roseum DSM 6424[a]	4.94	8	3	2[c]	13	0.40	8
Clostridium sp. BNL1100	4.61	9	4	1[c]	14	0.73	15.9
Clostridium aurantibutyricum DSM 793[a]	4.92	9	3	1[c]	13	0.43	8.7
Clostridium termitidis CT1112, DSM 5398[a,b]	6.42	14	3	1[d]	18	0.77	12
Clostridium cellulovorans 743B, ATCC 35296	5.26	7	10	0	17	0.67	12.7
Clostridium cellobioparum ATCC 15832[a,b]	6.13	11	2	1[c]	14	0.47	7.7
Clostridium saccharoperbutylacetonicum N1-4	6.67	6	2	1[d]	9	0.41	6.1
Clostridium beijerinckii NRRL B-598	6.19	3	3	0	6	0.26	4.1
Desulfofaba hansenii DSM 12642[a]	6.71	15	4	4[e]	23	0.52	7.8
Desulfospira joergensenii DSM 10085[a]	6.12	3	1	2[f,g]	6	0.22	3.6

1.2. Chemical diversity of the anaerobes

The following presents examples of several known natural products isolated from anaerobic bacteria, highlighting compounds discovered by traditional discovery techniques as well as recent genome mining efforts to discover unique natural products, particularly polyketides and non-ribosomal peptides with diverse activities.

1.2.1. Compounds discovered by genomics-independent methods

Genomic analysis suggests that nearly all major families of natural products, including polyketides, non-ribosomal peptides, RiPPs, and terpenes, can be produced by anaerobes.[7] However, very few metabolites have been isolated, perhaps because the lower efficiency of fermentative metabolism in anaerobes has led to stronger evolutionary pressures to strictly regulate the biosynthesis of secondary metabolites, precluding expression in typical laboratory conditions.[12] This section summarizes the known natural products discovered from anaerobes through traditional methods (**Figure 1-2**), with a particular focus on their structures, biological activities, and biosynthesis. Examples of RiPPs are excluded, although quite a few have been isolated and characterized from anaerobes.[13–19]

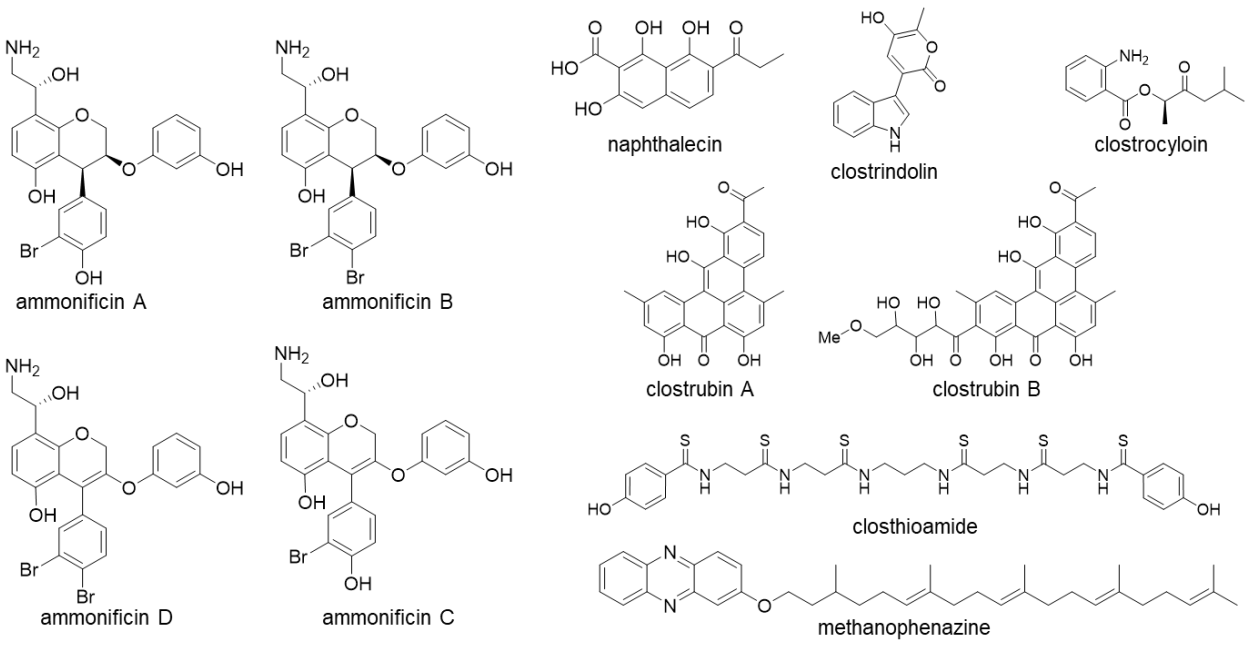


Figure 1-2. Secondary metabolites discovered from anaerobes by traditional methods

Methanophenazine

The phenazine family of natural products is known to be synthesized by diverse bacterial genera, including many pseudomonads and actinomycetes.[20] Additionally, several anaerobic archaeal *Methanosarcina* species are also known phenazine producers. The first such compound, methanophenazine, was isolated from lyophilized membranes of *Methanosarcina mazei* Gö1 by extraction with isooctane.[21] Methanophenazine is a 2-hydroxyphenazine derivative that is connected to a polyisoprenoid tail via an ether bridge (**Figure 1-2**). Proposed to play an important role in membrane-bound electron transport, methanophenazine represents the first example of such bioactivity by a phenazine compound.[21–23] Although phenazine biosynthesis in bacteria has been linked to the *phz* operon, *M. mazei* Gö1 does not have any identifiable *phz* homologs, suggesting that methanophenazine biosynthesis might proceed via a mechanism different to that of other bacteria.[20,24,25] The biosynthetic pathway of methanophenazine currently remains unclear despite the availability of genome sequences of the producers.[20]

Naphthalecin

One of the early studies on an obligate anaerobic bacterium isolated from a symbiotic growth with an aerobic bacterium in soil led to discovery of a new small molecule antibiotic.[26] In particular, the producing anaerobe was identified as a new species of the genus *Sporotalea,* and activity-guided isolation of a cell-associated compound led to the discovery of naphthalecin, a naphthalene derivative substituted with hydroxyl, propionyl, and acetyl groups (**Figure 1-2**). This compound demonstrated broad antibacterial activity against Gram-positive bacteria, including the symbiotic aerobic bacterium, but not against tested Gram-negative bacteria or fungi. The biosynthetic pathway for naphthalecin was not reported.

Ammonificins

Deep-sea hydrothermal vent systems are a promising source of new natural products because these systems contain a broad diversity of microbial organisms that have adapted their biochemical machinery to cope with extremely harsh environment conditions, likely resulting in altered metabolic pathways to produce structurally unusual metabolites. Aerobic isolates from these systems have already yielded interesting new natural products with antibiotic activities.[27,28] Additionally, one family of novel secondary metabolites, named ammonificins, have been purified from *Thermovibrio ammonificans*, an anaerobic and chemolithoautotrophic bacterium isolated from a hydrothermal vent system on the East Pacific Rise.[29] A total of four ammonificins, A-D, were purified from apoptosis-inducing cell extracts.[30,31] Ammonificins A-D were structurally determined to be novel chroman (A and B) or chromene (C and D) derivatives substituted with hydroxyethylamine and phenol or brominated phenol (**Figure 1-2**). Interestingly, ammonificins C and D induced apoptosis at micromolar concentrations while ammonificins A and B were inactive in an apoptosis induction assay. The biosynthetic pathway for ammonificin biosynthesis has not been reported.

Closthioamides

The genus *Clostridium* occurs widely in soil and in the gastrointestinal tract of higher organisms and includes several notorious human pathogens as well as non-pathogenic species useful for industrial biotechnology.[32] Although recent genomic analysis has revealed that natural product biosynthetic genes are widespread among clostridia,[7] the first secondary metabolite from clostridia was isolated in 2010. This compound, the antibiotic closthioamide, was purified from *Clostridium cellulolyticum* (recently renamed to *Ruminiclostridium cellulolyticum*), a cellulose-degrading organism isolated from decayed grass compost.[33] The initial discovery of closthioamide involved extensive culture condition screening; it was only successful upon addition of soil extracts to the culture medium to mimic its natural habitat. The production of closthioamide was further improved by overexpression of an anti-terminator gene (*nusG*) in *C. cellulolyticum* that induced the biosynthesis of closthioamide and related thioamides without the need for soil extracts,[34] and further improved again by the addition of inorganic sulfide to the culture medium.[35] Closthioamide possesses an unprecedented structure with multiple thioamide groups (**Figure 1-2**) that are critical for its potent antibiotic activity toward methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant enterococci (VRE), and drug-resistant *Neisseria* in a variety of disease models.[33,36] The mode of action of closthioamide was shown to inhibit DNA gyrase activity, with a different molecular mechanism from that of the quinolones and aminocoumarins.[37]

The biosynthesis of closthioamide was proposed to go through stepwise assembly using a hydroxybenzoate starting unit followed by elongation with β-alanine units and thionation of the intermediates, finished through the fusion of two intermediates via a diaminopropane linker.[34] The recently identified BGC for closthioamide suggested that instead of a typical modular NRPS, free-standing enzymes such as an ATP-grasp enzyme, an AMP-dependent ligase, and an amide synthase may be involved in activation and ligation of building monomers in a thiotemplated assembly line.[35] Dunbar *et al*. reported extensive *in vitro* work and proposed a mechanism of biosynthesis for the monomers of closthioamide: β-alanine biosynthesis is derived from CtaF-

dependent decarboxylation of Asp; para-hydroxybenzoic acid (PHBA) is derived from the the activity of CtaA chorismate lyase; diaminopropane may be derived from the activities of CtaKBF (aminotransferase, reductase, decarboxylase, respectively) acting upon an Asp substrate.[38] The mechanism of closthioamide biosynthesis then proceeds by two parallel peptidyl carrier protein (PCP) loading pathways: one pathway utilizes PHBA and ATP with CtaA, CtaH, and CtaI to form PHBA-PCP; the other functions iteratively, utilizing Asp and ATP with CtaD, CtaE, and CtaF to load a trimer of β-alanine onto a PCP. The pathways converge with CtaG transacylation of PHBA from its PCP (CtaH) onto the β-alanine trimer on its respective PCP (CtaE).[38] Then, iterative thioamidation by CtaC, an alpha-adenine nucleotide hydrolase homolog, incorporates sulfur into the amide bonds.[35] The biosynthesis of diaminopropane and dimerization remains speculative. This recently discovered example of a noncanonical thiotemplated assembly line biosynthesis of a nonribosomal peptide attests to further potential of anaerobes to produce secondary metabolites that may be beyond the current detection capabilities of bioinformatics.

Clostrubins

Clostrubin A was initially isolated as a deep red to purple pigment from cultures of *Clostridium beijerinckii* HKI0724.[39] Its production was also detected in cultures of the potato pathogen *Clostridium puniceum*, along with a related compound clostrubin B.[40] Both compounds feature the same unusual pentacyclic polyphenol scaffold, and clostrubin B contains an extra sugar-like linear side chain (**Figure 1-2**). The biosynthesis of clostrubins has been linked to a type II PKS gene cluster that is rare in anaerobes, and the polyphenol scaffold seems to emerge from a non-canonical polyketide folding, distinct from the conserved folding patterns of aerobic bacteria. Activity assays demonstrated potent antibiotic activity of clostrubin A against human pathogens, with minimum inhibitory concentrations (MIC) of 0.12 µM against MRSA, 0.97 µM against VRE, and 0.12-0.48 µM against various mycobacteria.[39] In addition, clostrubins A and B demonstrated antibacterial activity against a few common microbial potato pathogens, with MIC values in the range of 14-95 nM, suggesting that clostrubins may be used as chemical weapons to fight against competitors in a resource-limited environment. Intriguingly, these aromatic polyketides were also shown to enable the plant pathogen *C. puniceum* to survive in an oxygen-rich environment, adding to the growing number of examples of antibiotics with dual functions.[41]

Clostrindolin

Clostrindolin, a novel pyrone-containing alkaloid, was isolated from *Clostridium beijerinckii* HKI805.[42] This compound was purified from 50 L of culture and NMR data led to the proposal of an unusual 5-hydroxy-6-methyl-2*H*-pyran-2-one substructure conjugated to an indole to form 5-hydroxy-3-(1*H*-indol-3-yl)-6-methyl-2*H*-pyran-2-one. The closest related compound is the volatile compound 5-hydroxy-6-methyl-2*H*-pyran-2-one, which is isolated from tropical plants and plant pathogenic fungi. Schieferdecker *et al*. propose a biosynthesis through vinylogous addition of indole to a tautomer of the preformed 5-hydroxy-6-methyl-2*H*-pyran-2-one ring, followed by oxidation.[42] No BGC has been reported for clostrindolin. Activity assays with clostrindolin demonstrated strong and highly selective inhibition of *Mycobacterium vaccae* IMET 10670 and low cytotoxicity in a variety of human cell lines, suggesting a promising starting point for a antimycobacterial therapy agent. The MIC was determined to be 15.8 µM against *M. vaccae*. Agar disc diffusion assays also demonstrated moderate activity against *E. coli*. Structure-activity

relationship studies determined the unusual pyrone partial structure was significant for antimycobacterial activity.

Acyloins

A family of antimicrobial and antiproliferative acyloins was isolated from three strains of *Clostridium beijerinckii*: NCIMB 8052, HKI805, and HKI806.[43] These compounds include sattazolin A and B (known antiviral metabolites discovered from *Bacillus subtilis*) and three new congeners, as well as the novel compound clostrocyloin. The novel congeners were determined to be dehydrosattazolin, sattazolin B anthranilic acid ester, and hydroxysattazolin. Clostrocyloin, a white to yellowish solid, was unique to extracts of *C. beijerinckii* HKI805. Its structure was determined to contain the same α-hydroxy ketone characteristic of the acyloins, with the same isobutyl moiety on one end of the molecule, but a methyl rather than indole derivative on the other end of the molecule. In addition, clostrocyloin features an anthranilic acid ester at the α-hydroxy moiety. Extensive agar disc-diffusion assays demonstrated the antibiotic activities of these natural products. Clostrocyloin was found to be an antifungal agent capable of inhibiting the growth of *Sporobolomyces salmonicolor* and *Penicillium notatum*. Dehydrosattazolin, hydroxysattazolin, and clostrocyloin all possessed inhibitory activity against the mycobacterium *M. vaccae*. Dehydrosattazolin and clostrocyloin also demonstrated growth inhibition against *Bacillus subtilis*. Dehydrosattazolin, hydroxysattazolin, and clostrocyloin all demonstrated low cytotoxicity, suggesting their promise as scaffolds for antibiotic development. Meanwhile, the sattazolin B anthranilic acid ester was discovered to possess cytotoxic properties in HUVEC, K-562, and HeLa cells, with concentration of half-maximal growth inhibition (GI_{50}) of 13-22 μM.

The biosynthesis of these compounds was attributed to thiamine diphosphate dependent sattazolin synthase homologs, which were identified *in silico* and characterized *in vivo* and *in vitro* in enzymatic assays.[43] A related acyloin synthase was identified in association with clostrocyloin biosynthesis. The biosynthesis of clostrocyloin was proposed to initiate by attachment of pyruvate to a thiamine pyrophosphate to form a tertiary alcohol, followed by decarboxylation and nucleophilic attack by the resulting enol to the keto group of 4-methyl-2-oxopentanoic acid (which is derived from transaminated Leu). This intermediate is decarboxylated, possibly by the carboxymuconolactone decarboxylase found in proximity to the sattazolin synthase, and the resulting ene-diol intermediate is tautomerized to the acyloin, which is further esterified to anthranilic acid. The synthase demonstrates broad substrate specificity which may have value in producing enantiopure acyloin compounds.

1.2.2. Compounds discovered by genomics-dependent methods

Considering the occurrence of many putative secondary metabolite BGCs in anaerobes, genome mining to identify these BGC-associated metabolites can be a powerful approach to rapidly access new compounds.[4] In the following examples (**Figure 1-3**), we describe recent applications of genome mining to discover new anaerobic natural products that are associated with PKS and/or NRPS gene clusters. Depending on strain availability, strain cultivability, and available genetic tools, these natural products were identified and characterized by different methods including bioinformatics, gene expression analyses, biosynthetic gene disruption, heterologous expression, comparative metabolomics, biochemical characterization of biosynthetic enzymes, and chemical synthesis.

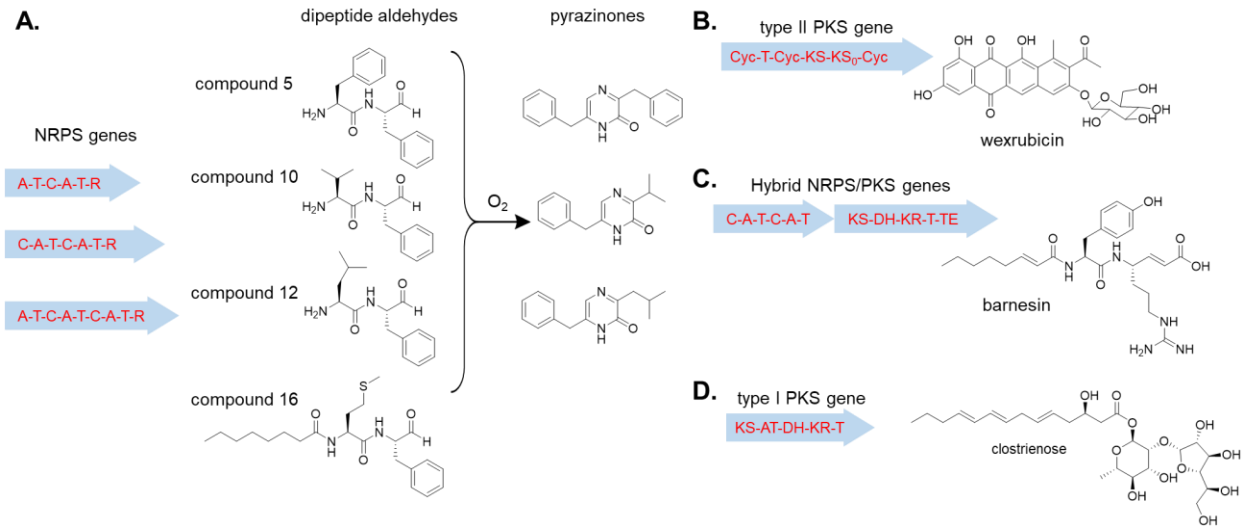


Figure 1-3. Secondary metabolites discovered from anaerobes by genome mining. (A) Dipeptide aldehydes from gut *Clostridium* spp. (B) Wexrubicin from *Blautia wexlerae*. (C) Barnesin from *Sulfurospirillum barnesii*. (D) Clostrienose from *Clostridium acetobutylicum*. Domain abbreviations: A, adenylation; T, thiolation; C, condensation; R, reduction; Cyc, aromatase/cyclase; KS, ketosynthase; KS_0, chain-length factor; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; TE, thioesterase.

Dipeptide aldehydes

The human gut harbors a broad diversity of anaerobes, but they are often difficult to culture, hindering metabolic and functional study of gut residents in isolation. Nevertheless, a metagenomic analysis by Zhu *et al.*[44] supports the idea that the gut microbiome collectively possesses significant biosynthetic coding potential with thousands of identified BGCs, although it is challenging to ascertain whether these genes truly correspond to secondary metabolism due to the use of ClusterFinder in the analysis.[45] One family of NRPS gene clusters was studied because it is widely distributed among healthy humans based on stool analysis and resides nearly exclusively in gut bacterial genome sequences.[46] The core NRPS enzymes typically have domains organized into A-T-C-A-T-R (A, adenylation; T, thiolation; C, condensation; R, reduction), although an extra module or domain (in particular an N-terminal C domain) can be present (**Figure 1-3a**). These gene clusters are prevalent in clostridia and a few other organisms from the gut. A total of 14 of these clusters were selected for metabolite analysis through heterologous expression in *Escherichia coli* and *Bacillus subtilis*, seven of which yielded detectable compounds belonging to the family of pyrazinones and dihydropyrazinones (**Figure 1-3a**). At least one native strain harnessing this gene cluster was able to produce the same metabolites as the engineered heterologous host, confirming that heterologous expression was a valid means to probe the metabolites produced by these gene clusters. Intriguingly, as the dipeptide aldehydes are plausible biosynthetic precursors to the isolated pyrazinones and dihydropyrazinones, and *N*-acylated metabolites (promoted by the first C domain of NRPS) retain the predicted C-terminal aldehyde moiety generated by the R domain, the dipeptide aldehydes immediately released from the NRPS assembly were proposed to be the active metabolites found in the gut under physiological

conditions (**Figure 1-3a**). Since peptide aldehydes are well known to inhibit proteases,[47–49] a few dipeptide aldehydes were tested and confirmed to be potent and selective protease inhibitors. A further unbiased target identification of the metabolite Phe-Phe-H using the isotopic tandem orthogonal proteolysis-activity-based protein profiling (isoTOP-ABPP) approach identified the cathepsins (specifically cathepsin L) as the principal targets.[46]

The list of dipeptide aldehydes produced by gut anaerobes was further expanded by recent work to identify the product of a conserved NRPS gene cluster from the abundant gut commensal *Ruminococcus bromii*.[50] It is notable that this cluster failed to yield any product during heterologous expression in *B. subtilis*.[50] The core NRPS enzyme has domains organized into C-A-T-C-A-T-R, and bioinformatics coupled with biochemical analysis of the NRPS suggested that an *N*-acylated dipeptide aldehyde (named ruminopeptin) could be the active metabolite, although the function of the R domain could not be reconstituted in vitro; nor could the proposed product be isolated from *R. bromii* cultures. Nonetheless, putative ruminopeptin scaffolds were chemically synthesized and several of them inhibited *Staphylococcus aureus* endoproteinase GluC (SspA/V8 protease), homologs of which were found in gut commensals, opportunistic pathogens, and the human gut metagenome.[50]

Wexrubicin

The human gut also harbors obligate anaerobes capable of producing polyketide secondary metabolites. One type II PKS containing BGC was discovered to be highly prevalent in human fecal metagenomes and was represented in samples from 7% of subjects from Fiji, 17% from Denmark or Spain, 23% from China, and 28% from the United States.[51] This BGC is also found in *Blautia wexlerae* DSM 19850, from the taxonomic class Clostridia. The BGC sequence was amplified from the genome of *B. wexlerae* and integrated into the genome of the heterologous expression strain *Bacillus subtilis* 168-sfp for characterization of its associated metabolite. This led to discovery of wexrubicin, a tetracyclic, anthracycline C21 ring system (**Figure 1-3b**). The polyketide is linked to a β-glucose moiety at C4, consistent with the presence of an encoded glycosyl transferase in the BGC. Wexrubicin is structurally related to clinically established anticancer drugs like doxorubicin and daunorubicin, as well as antitumor antibiotics tetracenomycin and elloramycin. Despite the similarity to these drug-like compounds, cytotoxicity assays against HeLa cells and MIC assays against pathogenic and commensal bacteria revealed no detectable activity.[51] The function of wexrubicin remains under investigation given the prevalence of this BGC in the human gut.

Barnesin

Anaerobic proteobacteria may also be promising sources for future natural product discovery based on genomic analysis.[7] A recent genome mining effort in *Sulfurospirillum barnesii*, an arsenate-reducing Epsilonproteobacterium isolated from a freshwater marsh, demonstrated for the first time that bioactive natural products could be produced from this genus.[52] In this study, an NRPS/PKS hybrid biosynthetic gene cluster found in *S. barnesii* was targeted for metabolite discovery (**Figure 1-3c**). A homologous unexplored NRPS/PKS cluster was also found on the genome of *Geovibrio* sp. L21-Ace-BES, but not in any other sequenced *Sulfurospirillum* spp. After confirming gene expression under laboratory growth conditions, metabolite comparison between

S. barnesii and two closely related *Sulfurospirillum* spp. that lack the NRPS/PKS cluster led to the identification of a new metabolite, named barnesin A, which is unique to *S. barnesii*. Barnesin A was revealed to be an *N*-acylated dipeptide carboxylate containing a vinylogous side chain (**Figure 1-3c**). In addition to antibiotic activity against a few human pathogens, barnesin A showed selective and nanomolar inhibitory activity against cysteine proteases including cathepsin B. Considering that small peptides and vinylogous systems are well known to inhibit proteases, it was reasonable to propose that barnesin A has a mode of action via a 1,4-Michael-type addition mechanism, which was supported by structure-activity-relationship studies.[52]

Clostrienose

In contrast to clostridia from the gastrointestinal tract of higher organisms, clostridia isolated from soil environments have greater natural product biosynthetic potential according to genomic analysis (**Table 1-2**).[7] Many of these *Clostridium* spp. are useful for industrial biotechnology, such as biomass degradation and industrial-scale production of organic acids or solvents. However, despite extensive research and industrial application of these organisms, knowledge of the molecular identity and function of their secondary metabolites remains limited. Modern genome mining of secondary metabolites could be a promising approach to accelerate the discovery of new natural products from these anaerobes, provide insights in biological functions of secondary metabolites, and possibly improve these strains for industrial applications by manipulating their secondary metabolism. A recent genome mining effort in *Clostridium acetobutylicum* exemplified such opportunities.[53]

C. acetobutylicum is an organism well-known for its use in historical production of organic solvents such as acetone, butanol, and ethanol.[54] One PKS gene was identified in all sequenced *C. acetobutylicum* strains and the encoding enzyme has domains organized into KS-AT-DH-KR-T (KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; T, thiolation) (**Figure 1-3d**). Expression of this PKS gene was significantly upregulated during early stationary phase, suggesting that the corresponding polyketide product could be associated with morphological development and/or solventogenesis.[32] Metabolomic comparison between the wild-type and mutant with the disrupted pks gene led to the identification of the polyketide metabolite, named clostrienose. The structure of clostrienose was revealed to be a 2-hydroxy-5,8,10-tetradecenoic acid linked to a disaccharide, α-D-galactofuranosyl(1→2)-α-L-rhamnopyranoside, via an ester linkage (**Figure 1-3d**). This molecular scaffold could never be predicted by bioinformatics due to the iterative activity of this single module PKS as well as the apparent non-clustered nature of the biosynthetic genes on the genome. This also raises concerns about using heterologous expression to probe BGC-associated metabolites since all required biosynthetic genes may not be clustered together. Clostrienose was shown to be important in stimulating sporulation and granulose accumulation in *C. acetobutylicum*, and the PKS deletion strain exhibited improved traits for industrial solvent production, such as reduced sporulation, reduced granulose accumulation, and increased butanol titer and productivity.[53]

1.3. Solventogenic clostridia as producers of secondary metabolites

The anaerobes most enriched in PKS and NRPS containing BGCs are soil isolates from the genus *Clostridium* (**Figure 1-1**). Another group of soil isolates from the genus *Clostridium* are the acetone-butanol-ethanol (ABE) producers, which are capable of fermentative conversion of carbohydrates into ABE. Indeed, many of the known examples of anaerobe-derived natural products were isolated from these solventogenic clostridia, including the clostrubins, clostrienose, clostrocyloin, and clostrindolin. These observations point to the ABE fermenting clostridia as a promising source of anaerobes for natural products discovery efforts.

The historical prominence of the ABE fermentation process is closely tied to technological and geopolitical developments.[55] ABE fermenting microbes were first discovered by Pasteur in 1862. In 1915, the first patent was filed describing a process for conversion of starch to acetone and butanol by *C. acetobutylicum.* Thereafter, a slew of research developments enabled cheaper feedstocks, more complete utilization of feedstock, lower processing temperature, improved butanol selectivity, decreased contamination incidence. ABE fermentation became a major commercial process and supplied ~66% and 10% of world demand for butanol and acetone, respectively.[55] The process supplied acetone for cordite (smokeless gunpowder) production through both World Wars, and butanol for production of quick-drying lacquers for automobile manufacturing. The process was improved by strains that could accommodate new feedstocks like molasses. After World War II, ABE fermentation products were largely replaced in the market by cheaper petrochemical-derived solvents in the United States and Britain, although operations continued in South Africa and the Soviet Union into the 1980s.[56] Research interest in ABE fermenting clostridia has experienced a resurgence due to the global need for sustainable solvents and drop-in[57] biofuels which can be produced from renewable resources.

Recent developments in ABE fermentation research have advanced both technical processing and, to a larger extent, understanding of clostridial physiology and genetics.[54] Studies have leveraged modern advancements in molecular biology to make improvements at the strain level,[58] and untargeted multi-omics approaches to understand the biology from a systems level.[59] Strain engineering has enhanced solvent production by manipulating metabolic flux, addressed product toxicity, and changed outputs to be more compatible with combustion engines.[58] Other efforts have focused on expanding the input range for fermentation to utilize non-food substrates by combining solventogenic clostridia with the metabolic capabilities of either cellulose- or syngas-utilizing microbes.[58] Systems biology approaches have provided a holistic perspective of ABE biology, using untargeted characterizations such as transcriptomics, proteomics, and metabolomics, especially in *C. acetobutylicum*.[59] These gene expression profiling methods have captured the processes of acidogenesis and solventogenesis and uncovered new biology, such as regulatory functions of small non-coding RNAs.[60] Metabolomics using isotopically labeled tracer feeding experiments also contributes to studies of metabolic flux for *in silico* modeling.[61,62] Much of what is known about ABE biology stems from what is characterized in the model organism *C. acetobutylicum*, but it is important to also study other ABE fermenters which differ in patterns of product formation and substrate tolerance.[63] Indeed, many new ABE fermenters have been recently sequenced and provide opportunities to identify new targets for genetic engineering.[64] Other developments have characterized population heterogeneity of fermentation cultures[63] or identified unique chemistry such as the role of cyclopropyl fatty acids which modulate cell membrane fluidity,[65] demonstrating the continued need to research the fundamental biology of ABE producers.

The fact that *C. acetobutylicum* and many other ABE fermenters contain BGCs encoding natural products suggests an important role of natural products in the evolutionary history of these organisms. It follows that the presumed biological processes mediated by these natural products can impact their performance in an ABE fermentation context. This is clearly demonstrated by the recent discovery of clostrienose, a signaling molecule from *C. acetobutylicum* which was associated with physiological processes competing with solventogenesis.[53] In contrast, the clostrubins of *C. puniceum* and *C. beijerinckii* are an example of natural products which confer desirable fermentation traits as they mediated aerotolerance,[39,40] although the direct impact of clostrubin on ABE performance has yet to be reported. Thus, it is important to study the ABE fermenting clostridia from the lens of natural products biosynthesis not only to discover novel chemical diversity but also to understand the fundamental biology of these industrially important organisms.

Chapter 2. Investigation of secondary metabolism in *Clostridium saccharoperbutylacetonicum* N1-4

Parts of this chapter have been adapted from the following with permission: Li, J. S., Barber, C. C., Herman, N. A., Cai, W., Zafrir, E., Du, Y., Zhu, X., Skyrud, W., Zhang, W. "Investigation of secondary metabolism in the industrial butanol hyper-producer *Clostridium saccharoperbutylacetonicum* N1-4." *J. Ind. Microbiol. Biotechnol.* **47**, 319-328 (2020).

2.1. Introduction

Clostridium saccharoperbutylacetonicum N1-4 (*Csa)* is a Gram-positive, spore-forming obligate anaerobe. After its isolation from soil in 1959, it was patented by the Sanraku Distillers Company in 1960[66] for use in saccharolytic fermentations to produce organic solvents, including acetone, butanol, and ethanol (ABE). In subsequent decades, ABE fermentation largely fell out of favor due to competition from the petrochemical industry.[56] Today, interest in renewably produced butanol as a drop-in biofuel[57] has revived investigations into ABE fermentation as a sustainable source of chemical energy.[55,67,68] Despite the advantages of historical precedent,[69] substrate flexibility,[8-22] and butanol hyper-production for *Csa*, challenges remain to improve ABE productivity, titer, and yield of this organism. A deeper understanding of the biology of *Csa* could provide insights necessary to further improve its industrial fermentation traits.[63]

Microbial secondary metabolites are known to possess diverse functions relating to the metabolism, physiology, differentiation, interspecies competition, etc.[2] For example, the plant pathogen, *C. puniceum*, produced an aromatic polyketide, clostrubin, which enabled *C. puniceum* to both survive in an oxygen-rich environment and inhibit other plant pathogenic bacteria.[39,40] An ABE model organism, *C. acetobutylicum*, produced a glycosylated polyketide, clostrienose, which promoted sporulation and granulose accumulation.[53] A mutant of *C. acetobutylicum* lacking the capacity to produce clostrienose downregulated these differentiation processes, resulting in improved butanol titer and productivity. The secondary metabolism of clostridia could thus be manipulated to improve traits relevant for industrial applications. Among all secondary metabolites, polyketides (PKs) and non-ribosomal peptides (NRPs) are two major families noted for their chemical diversity, range of potent bioactivities, and well-studied mechanism of biosynthesis.[85,86] PKs and NRPs are biosynthesized by polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), the thio-templated assembly-line enzymes, together with diverse tailoring enzymes. The biosynthetic genes for a particular metabolite often co-localize on the genome in a biosynthetic gene cluster (BGC), facilitating discovery in silico.[87] Notably, while *Csa* has a relatively large genome (6.6 Mb)[88,89] among clostridia and putative PKS and NRPS encoding genes are widespread in its genome,[90] little is known regarding the regulation and function of these genes, and no PK and NRP metabolites have been reported.

We here study the PK/NRP-related secondary metabolism in *Csa* for the first time. First, an in-depth bioinformatic analysis is performed to profile BGCs encoding putative PKSs and NRPSs. Then, an untargeted transcriptomics approach is utilized to probe BGC expression in ABE fermentation culture conditions. Finally, one of the identified highly expressed BGCs is investigated to reveal its associated metabolite and possible role in ABE fermentation.

2.2. Results

2.2.1. Bioinformatic analysis of PK/NRP BGCs in *Csa*

As an initial step in exploring the secondary metabolism of *Csa*, we performed *in silico* analysis of the published genome of the type strain (NCBI accessions NC_020291.1 and NC_020292.1). The *Csa* genome consists of a 6.53 Mb circular chromosome and a 136 kb circular megaplasmid, giving this organism the largest reported genome of sequenced clostridia to date.[88,89] The bioinformatic pipelines AntiSMASH[91] and PRISM[92] were used to identify putative BGC loci, and MultiGeneBLAST[93] and BiG-SCAPE[94] were used to query for homologous BGCs in other organisms. We identified seven BGCs which contain genes characteristic of PK or NRP biosynthesis. These can be categorized into four predicted NRPS gene clusters and three hybrid PKS-NRPS gene clusters (**Figure 2-1**). We trimmed the BGCs conservatively to define putative boundaries by removing genes that were not in operon with a biosynthetic gene or that had predicted functions other than biosynthetic/transporter/regulatory/hypothetical. In the case of *hyb2* and *hyb3*, one locus was defined as two adjacent BGCs based on the fractured nature of biosynthetic genes in *hyb2*, the *cis*- vs *trans*-acyltransferase PKS modules, and the presence of distinct termination domains. More in-depth cluster-specific analyses are presented in **Table A-1** (**Appendix A**).

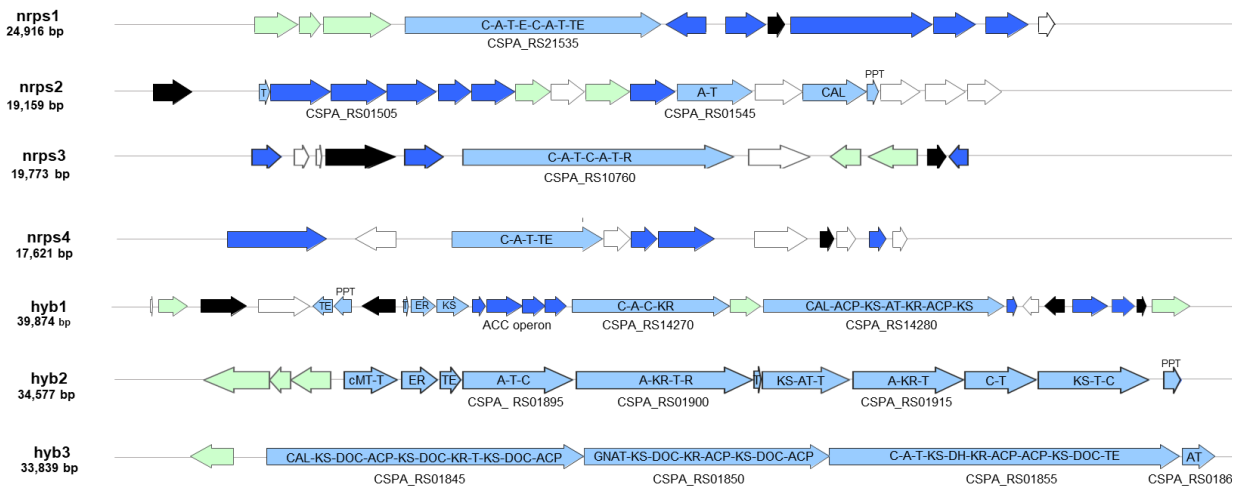


Figure 2-1. BGCs of *Csa* which contain PK or NRP biosynthetic genes. Gene color indicates putative function: light blue, core biosynthetic genes; dark blue, additional biosynthetic gene; black, transcriptional regulator; light green, transporter; white, other. Domain organization is indicated for PKS/NRPS genes. Abbreviations: C, condensation; A, adenylation; T, thiolation/peptide carrier; E, epimerization; TE, thioesterase; R, reductive release; KR, ketoreductase; CAL, Acyl-CoA ligase; KS, ketosynthase; cMT, C-methyltransferase; ER, enoyl-reductase; DOC, *trans*-AT docking; ACP, acyl-carrier protein; GNAT, Gcn5-related *N*-acetyltransferase; DH, dehydratase; PPT, 4'-phosphopantetheinyl transferase domain

The relative abundance of these secondary metabolism genes in *Csa* compared to other *Clostridium* species may suggest an important role of secondary metabolites in the evolutionary biology of *Csa*. Several of the BGCs appear to be conserved in other Firmicutes, particularly *hyb1* (**Figure A-1, Appendix A**) which is found in a wide diversity of Firmicutes and extensively among *Clostridium spp.* The *hyb2* BGC has one identifiable homolog in a *Paenibacillus* species (**Figure A-2, Appendix A**) with almost identical pfam domain architecture. Notably, the megasynthetase gene of *nrps3* has homologs in two other ABE-producing Clostridia (**Figure 2-2**), although their respective neighborhoods are not conserved. This suggests the true boundaries of the BGC encompass the singular gene.

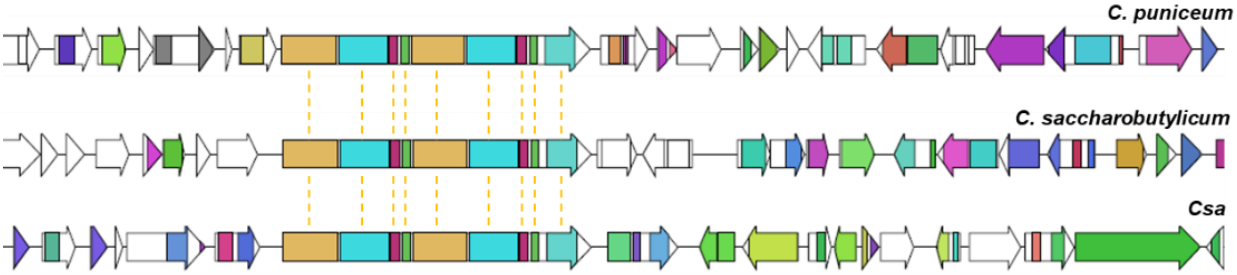


Figure 2-2. Homologs of *nrps3*. BiG-SCAPE analysis of genomes of *Clostridia* found on NCBI revealed examples of the characteristic NRPS gene (yellow dotted lines denoting homologous pfam domains) which are found in other ABE fermenters

2.2.2. Transcriptional profiling of PK/NRP BGCs during ABE fermentation

A transcriptomic approach was used to study gene expression levels of the BGCs in an ABE fermentation context. High quality RNA (RIN ~8) was prepared from *Csa* batch cultures grown to log phase with four biological replicates. The resulting RNA-seq dataset (SRA: PRJNA551507) was quality controlled using FastQC. An average of 91.4% or 86.3 million reads per biological replicate were used as input for downstream analysis, well above the 5-10 million reads suggested to be sufficient for bacterial transcriptome analysis.[95] Next, input reads were mapped to either the *Csa* chromosome or the megaplasmid. Reads mapping to multiple loci were counted once for each locus. FPKM values were calculated for each locus and compiled to assess relative gene expression.

Most of the genome is expressed, with 99.8% of genes having at least one mapped read and 97.8% of genes having at least 10 mapped reads. For each BGC, the core PKS/NRPS gene with the minimum average expression level was used to represent overall BGC expression. The *gyrB* (DNA gyrase subunit B) housekeeping gene was used as an expression benchmark.[96] The resulting profile of expression is shown in **Figure 2-3**. Five of seven BGCs demonstrated some baseline expression (FPKM > 1), with *nrps2* and *hyb2* falling below the expression cutoff. Two BGCs, *nrps3* and *nrps4*, were expressed at a level comparable to that of *gyrB*. A detailed compilation of BGC gene expression levels is presented in **Table A-2** (**Appendix A**).

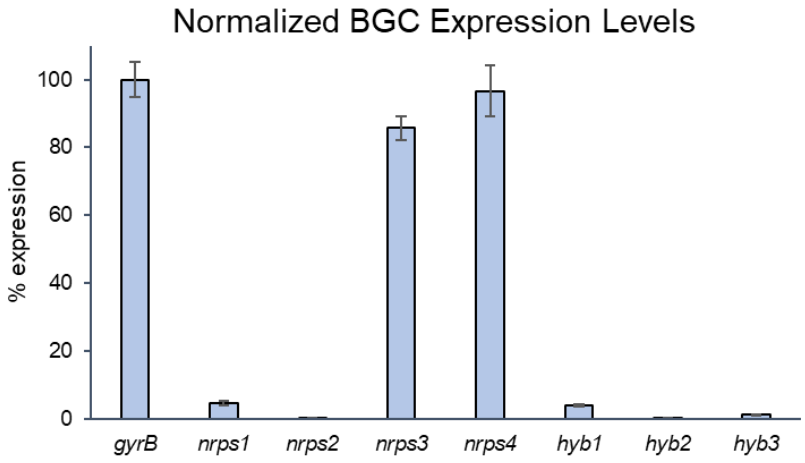


Figure 2-3. Expression level of the BGCs of *Csa*. Floor expression values for PKS/NRPS are normalized to that of *gyrB*, a housekeeping benchmark

2.2.3. Characterization of *nrps3*-associated secondary metabolite

As *nrps3* is both highly expressed and conserved in other solvent-producing *Clostridium* species, we selected this BGC for metabolite interrogation. In order to discover the secondary metabolite associated with *nrps3*, CSPA_RS10760 was disrupted using a CRISPR/Cas9-nickase targeted homologous recombination strategy.[97,98] The resulting strain, Δ*nrps3*, was genotyped by PCR to confirm successful editing at the target locus (**Figure 2-4**).

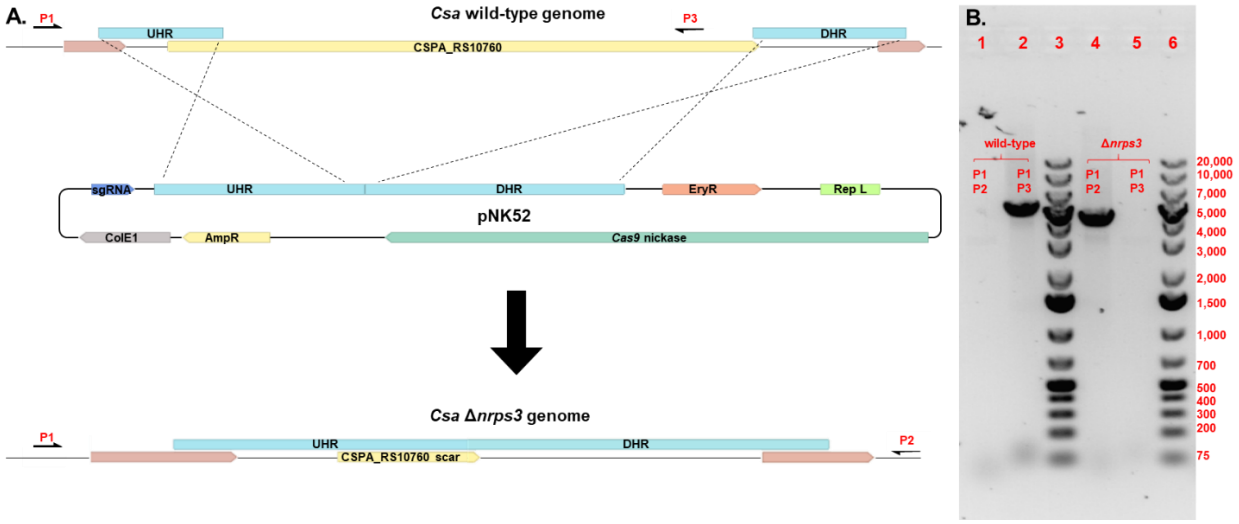


Figure 2-4. Mutant generation in *Csa*. (A) Schematic of pNK52 and the *nrps3* locus in both *Csa* wild-type and *Csa* Δ*nrps3*. After serial passaging, *Csa* pNK52 was streaked to identify clones with the Δ*nrps3* genotype. Dotted lines represent homologous recombination events between the wild-type chromosome and pNK52. Single-barbed arrows P1-P3 mark the primer annealing sites for PCR verification. Key: UHR, upstream homology region; DHR, downstream homology region; sgRNA, single-guide RNA; EryR, erythromycin resistance gene; Rep L, Gram-positive origin of replication; ColE1, Gram-negative origin of replication; AmpR, ampicillin/carbenicillin resistance gene. (B) DNA electrophoresis gel showing colony PCR products produced by either P1-P2 (lanes 1 and 4; expected sizes: wild-type, no band; mutant, 5.6 kb) or P1-P3 (lanes 2 and 5; expected sizes: wild-type, 5 kb; mutant, no band). Lanes 3 and 6 contain O'GeneRuler 1 kb Plus DNA Ladder (Invitrogen, Thermo Fisher Scientific)

Next, chemical extracts of the wild-type and Δ*nrps3* cultures were analyzed using LC-HRMS. Untargeted metabolomic comparison of the strains was performed using XCMS, enabling identification of a new compound, **1**, with molecular formula $C_{15}H_{30}N_2O_3$ (calculated for $C_{15}H_{31}N_2O_3^+$: 287.2329; found: 287.2327) (**Figure 2-5**). A majority of **1** was found in spent culture medium rather than pelleted cell mass, suggesting that it is secreted upon biosynthesis.

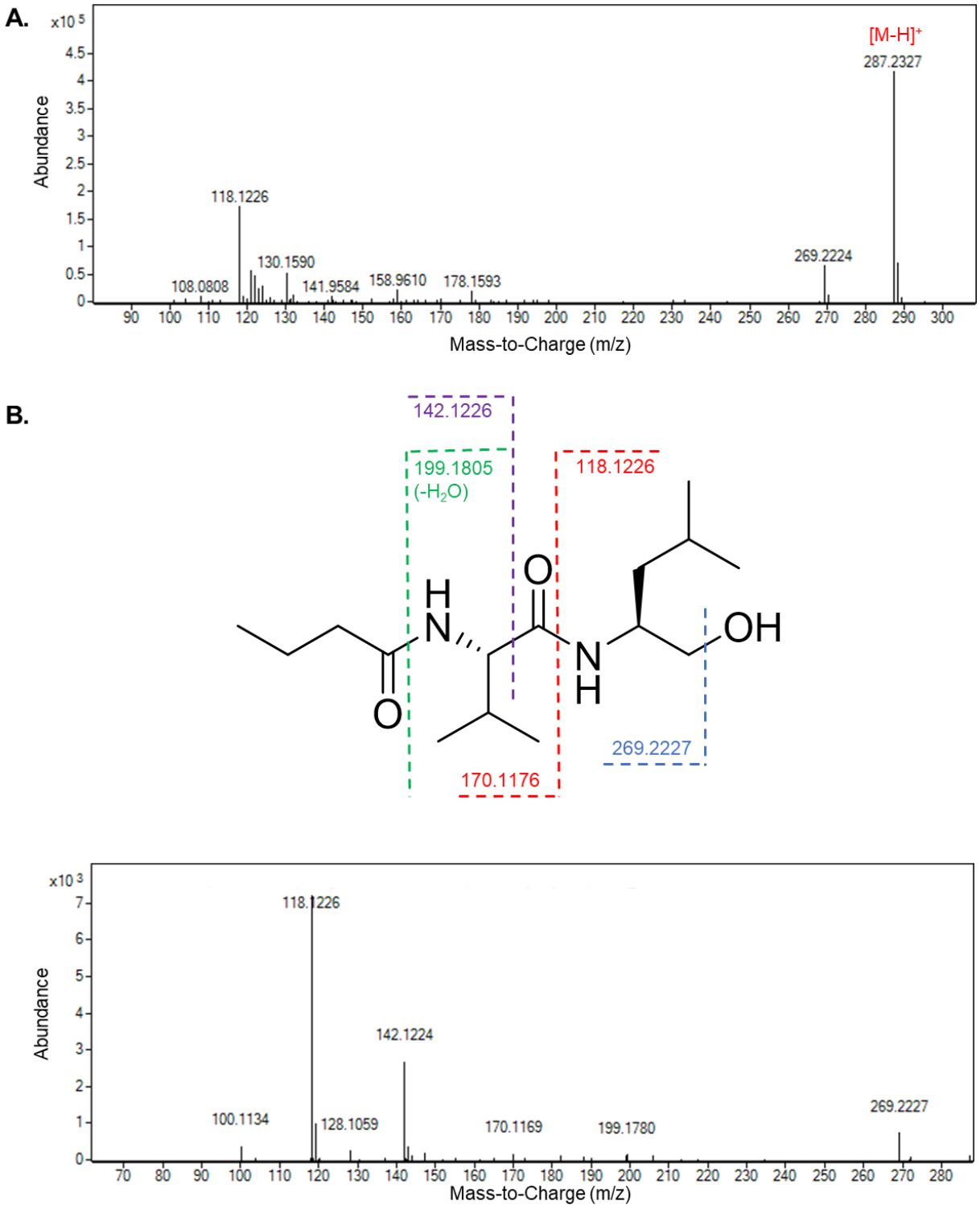


Figure 2-5. HRMS characterization of compound 1. (A) HRMS (positive mode) of 1 (calc., 287.2329 m/z). (B) MS/MS analysis (positive mode) of 1.

To obtain enough material for structural characterization, three liters of *Csa* batch culture was extracted with ethyl acetate and purified using a size-exclusion column packed with Sephadex LH-20, followed by several rounds of high-performance liquid chromatography (HPLC) purification. A detailed method is available under Supplementary Methods. This yielded 2 mg of pure compound that was subjected to 1D and 2D NMR analysis for structural elucidation, including ^{1}H, ^{13}C, HSQC, COSY, and HMBC (**Figure A-3, Appendix A**). Compound **1** was determined to be an *N*-acylated dipeptidyl alcohol derived from butyric acid, valine and leucine monomers. A chemical standard was synthesized (**Figure A-4, Appendix A**) and the proposed structure was verified by HPLC retention time and tandem MS spectrum.

The domain organization of the di-modular NRPS offers insights into the biosynthesis of **1**. The proposed mechanism (**Figure 2-6a**) channels substrates through successive modules to form two peptide bonds between the butyryl starting unit and subsequent L-valine and L-leucine monomers. Then, the reductase domain releases the alcohol product from the assembly line through a four-electron reduction.

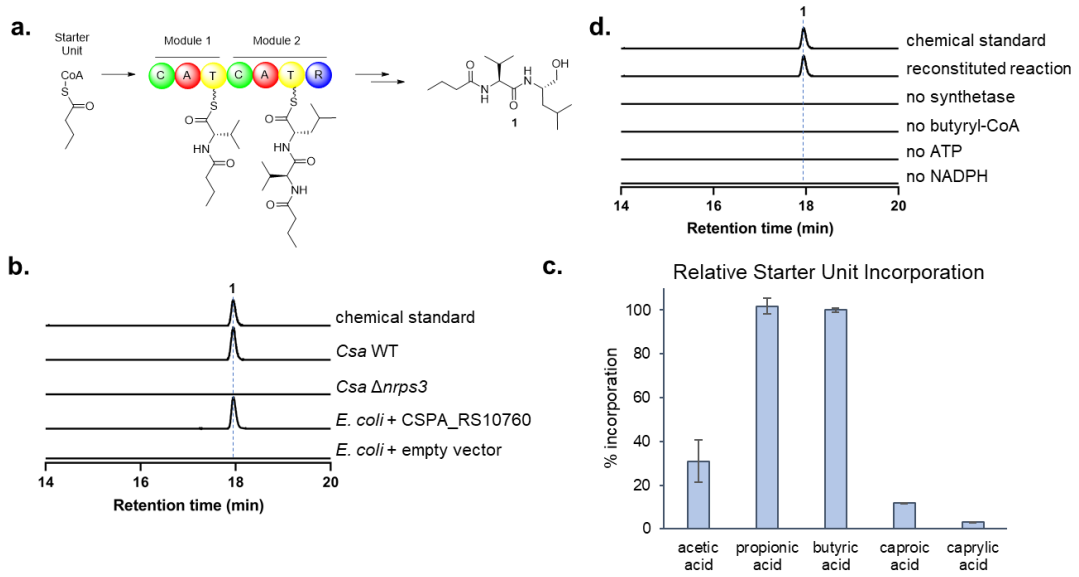


Figure 2-6. Characterization of *nrps3*-associated secondary metabolite, 1. (A) Proposed biosynthesis of the secondary metabolite 1, an *N*-acylated dipeptidyl alcohol. (B) HRMS extracted ion chromatograms demonstrating requirement of CSPA_RS10760 in *Csa* for compound biosynthesis, and heterologous production of 1 in a heterologous E. coli host expressing CSPA_RS10760. (C) Relative starter unit promiscuity of the NRPS demonstrated by substrate feeding of varying-length short-chain fatty acids in *E. coli* + CSPA_RS10760. (D) HRMS extracted ion chromatograms demonstrating biosynthesis of 1 in vitro. The calculated mass for 1 (287.2329) with 10 ppm mass error tolerance was used for each trace in b and d

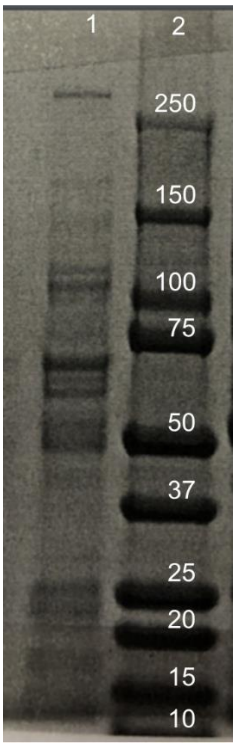


Figure 2-7. SDS-PAGE gel of the purified NRPS. Lane 1: the intact his-tagged gene product of CSPA_RS10760 (290 kDa); Lane 2: Precision Plus Protein Dual Color Standard (Bio-Rad)

To confirm that the NRPS encoded by CSPA_RS10760 is sufficient to produce the identified compound, we overexpressed the gene in *E. coli* to probe for possible heterologous metabolite production (**Figure 2-7**). The recombinant *E. coli* strain produced the expected product **1**, consistent with the predicted function of the di-modular NRPS (**Figure 2-6b**). We further probed the starter unit substrate promiscuity of the assembly line by feeding fatty acids of varying chain length (**Figure 2-6c**). New products corresponding to the two-, three-, six-, and eight-carbon head groups were identified upon feeding of the corresponding fatty acids, demonstrating that the C domain of the NRPS has a relaxed substrate specificity toward short- to medium-chain fatty acyl groups, albeit with C3-4 substrates preferred. Finally, to unequivocally confirm the function of the NRPS, the intact megasynthetase was purified from *E. coli* for in vitro activity reconstitution (**Figure 2-6d**). The enzymatic reaction mixture containing butyryl-CoA, L-valine, L-leucine, ATP and NADPH successfully produced compound **1** as a product, confirming that the NRPS is sufficient to make this product.

The product of CSPA_RS10760 resembles a class of reported dipeptidyl aldehyde compounds associated with various human gut commensal microorganisms.[46,50] This class of compounds represents potent inhibitors of a variety of human and bacterial proteases. Both these human-gut derived metabolites and **1** are biosynthesized by small NRPS genes terminating in R domains. However, several key differences can be observed between these secondary metabolites. The majority of NRPSs associated with protease inhibition lack the *N*-terminal C domain, with the resulting lack of *N*-acylation in the product leading to the formation of pyrazinone shunt metabolites. Moreover, two NRPSs with an identical domain organization to the NRPS encoded

by CSPA_RS10760 have been expressed in heterologous *E. coli* hosts, and *N*-acylated dipeptidyl aldehydes, instead of alcohols, have been reported as products. In contrast, the major products of CSPA_RS10760 seemed to be alcohols from both anaerobic culture of *Csa* and aerobic heterologous expression in *E. coli*, and compound **1** did not have obvious protease inhibition activity toward cathepsin B. Thus, the R domain encoded by CSPA_RS10760 is a more efficient reductase which catalyzes the iterative reduction of the peptidyl carboxyl to a terminal alcohol, distinct from the homologous NRPSs from human gut microbes.

We then examined the possible role of *nrps3* in solvent production. Initial batch fermentations of the wild-type *Csa* and Δ*nrps3* showed that *nrps3* does not have a direct impact on either glucose consumption (**Figure 2-8**) or batch ABE titers (**Figure 2-9a-b**) after a 48 hour fermentation. In addition, the wild-type *Csa* and Δ*nrps3* strains exhibited similar colony morphologies and were indistinguishable in assays assessing for swimming-motility and granulose accumulation (method described in **Appendix A**). We next turned to an untargeted method to identify a possible impact of *nrps3* disruption; differential gene expression analysis was carried out using RNA-seq data representing cultures of *Csa* wild-type and Δ*nrps3*. Filtering the data for two-fold differential gene expression and p-value < 0.001 yielded CSPA_RS10760 (the NRPS) and six protein-coding genes which were revealed by STRING[99] analysis to comprise a putative glycerol metabolism operon (**Figure 2-9c**). While *Csa* cannot use glycerol as a sole carbon source,[100] glycerol metabolism genes have been reported as part of the solvent tolerance response in a wide variety of fermentation hosts, including *E. coli* (*glpBCFQ* in response to hexanes stress,[101] *glpC* in response to xylene and cyclohexane,[102] and *C. acetobutylicum* (up-regulation of *glpA* and *glpF* in response to 3.7 g/liter butanol).[103] To probe whether *nrps3* mediates a solvent stress response in *Csa*, we collected growth data comparing wild-type and Δ*nrps3* after butanol challenge.[103] We found that Δ*nrps3* exhibited a growth defect during the exponential phase relative to the wild-type (Fig. 4d), although the optical densities of Δ*nrps3* cultures converged with *Csa* wild-type cultures during the stationary phase. These results support an association between *nrps3* and *glp*-mediated solvent tolerance revealed through transcriptomic analysis, although further work will be required to determine the underlying molecular mechanism.

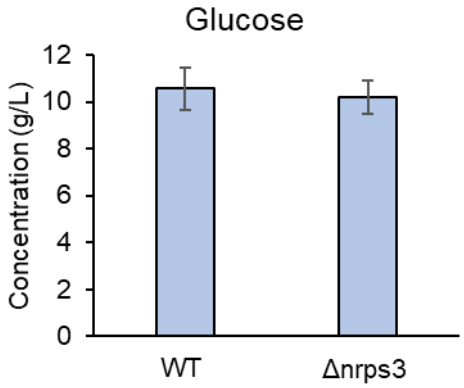
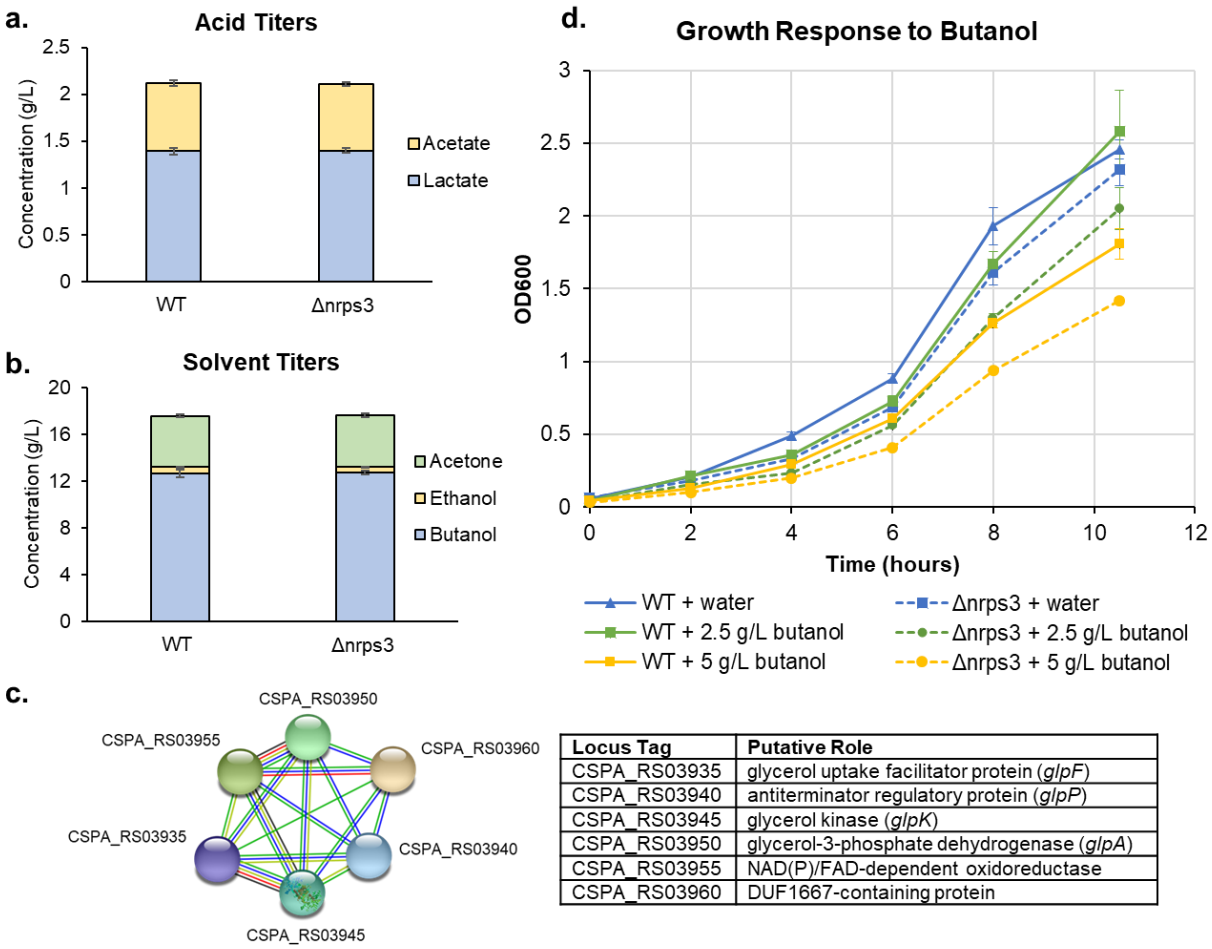


Figure 2-8. Residual glucose in ABE fermentation cultures of *Csa* wild-type and Δ*nrps3*. Samples were collected at the end of batch fermentation (48 h), resolved using HPLC and quantified by refractive index detector



Locus Tag	Putative Role
CSPA_RS03935	glycerol uptake facilitator protein (*glpF*)
CSPA_RS03940	antiterminator regulatory protein (*glpP*)
CSPA_RS03945	glycerol kinase (*glpK*)
CSPA_RS03950	glycerol-3-phosphate dehydrogenase (*glpA*)
CSPA_RS03955	NAD(P)/FAD-dependent oxidoreductase
CSPA_RS03960	DUF1667-containing protein

Figure 2-9. Comparison of wild-type *Csa* and Δ*nrps3* in batch cultures under ABE fermentation conditions. (A) Production of organic acids. (B) Production of ABE. (C) STRING network analysis of RNA-seq data comparing wild-type and Δ*nrps3* fermentations identified a differentially regulated genetic pathway in Δ*nrps3*. The nodes represent the six down-regulated genes, CSPA_RS03935- CSPA_RS03960, summarized in the accompanying table. This *glp* operon has been associated with solvent stress response in another ABE producer, *C. acetobutylicum*. The edge colors reflect different lines of evidence for gene functional relationships: light green, gene neighborhood; red, gene fusions; blue, gene co-occurrence; olive green, textmining; black, co-expression. (D) Butanol challenge assay demonstrating a growth defect in Δ*nrps3* relative to wild-type *Csa*. Each assay was repeated at least three times independently.

One possible hypothesis for the mechanism of action of this small molecule is that it is a signaling molecule such as those associated with quorum sensing systems. In a preliminary study, we tested the ability of purified **1** to chemically complement the growth defect observed in Δ*nrps3* during the butanol challenge assay (**Figure 2-10**). This experiment was inconclusive due to the use of dimethylsulfoxide solvent, which likely altered the growth profiles collected.

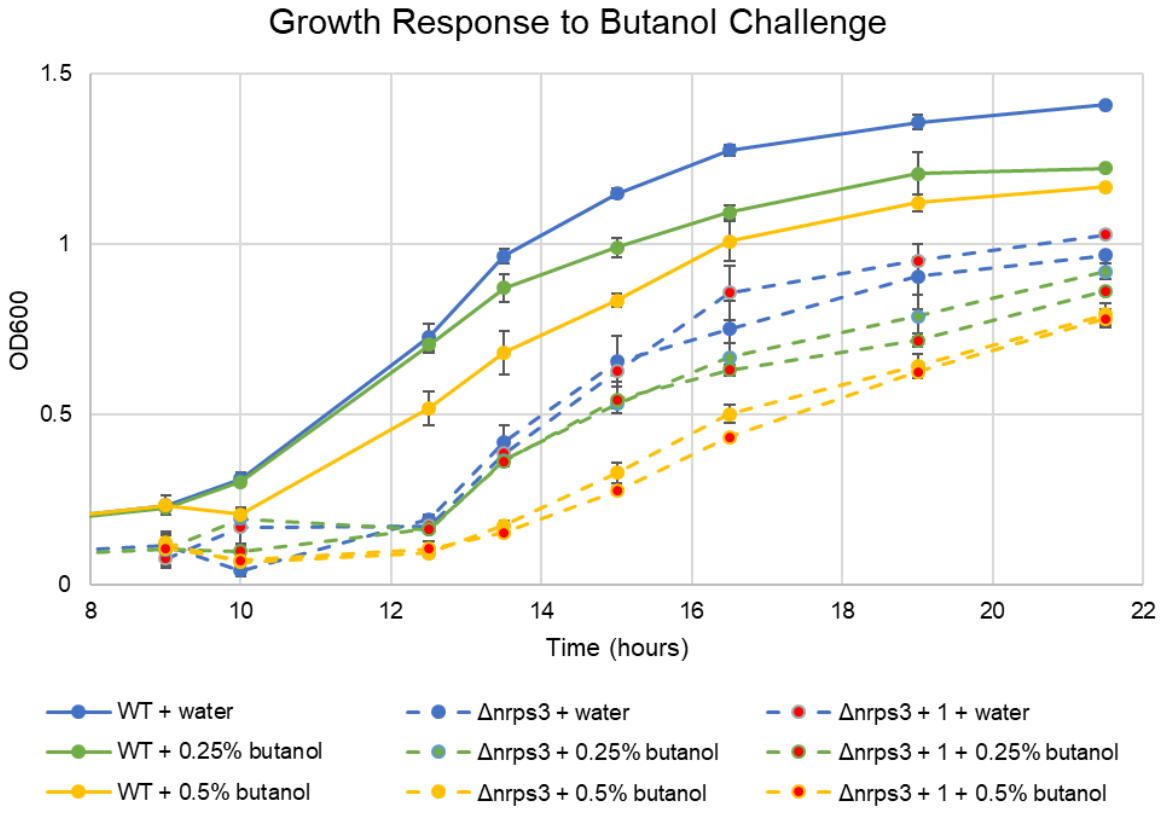


Figure 2-10. Butanol challenge assays with feeding of compound 1 to Δ*nrps3* groups

2.3. Discussion

This work provides first insights into the secondary metabolism of *Csa*, an anaerobic bacterium historically valued for industrial solvent production. An in-depth bioinformatic analysis of *Csa* revealed seven uncharacterized PK/NRP BGCs in its genome, of which five were expressed under solvent-production conditions, as shown by transcriptomic analysis. Of these, two were very highly expressed, namely *nrps3* and *nrps4*. The highest-expressed BGC, *nrps4*, has no homologs among known ABE producers. In addition, we observed that by expanding our BGC expression analysis to consider floor expression across putative PKS/NRPS-containing operons rather than just PKS/NRPS genes, we obtained a different profile of expression with *nrps3* as the highest-expressed BGC (**Figure 2-11**). Thus, we selected *nrps3* for further investigation due to its high level of expression as well as its observed occurrence in other known ABE fermenters Further investigation of one of the highly expressed BGCs, *nrps3*, led to the identification of its associated

metabolite, an *N*-acylated dipeptidyl alcohol, and its biosynthetic mechanism. While phenotypic comparisons between the *Csa* wild-type and Δ*nrps3* showed no difference in the batch culture solvent production titers, a comparative transcriptomic analysis followed by butanol challenge assays suggested a possible role of *nrps3* in *glp*-mediated butanol tolerance. This work thus demonstrated another example of a small-molecule secondary metabolite affecting traits relevant for microbial industrial applications, as well as the value of transcriptomics as a powerful untargeted tool for associating secondary metabolites with ABE biology.

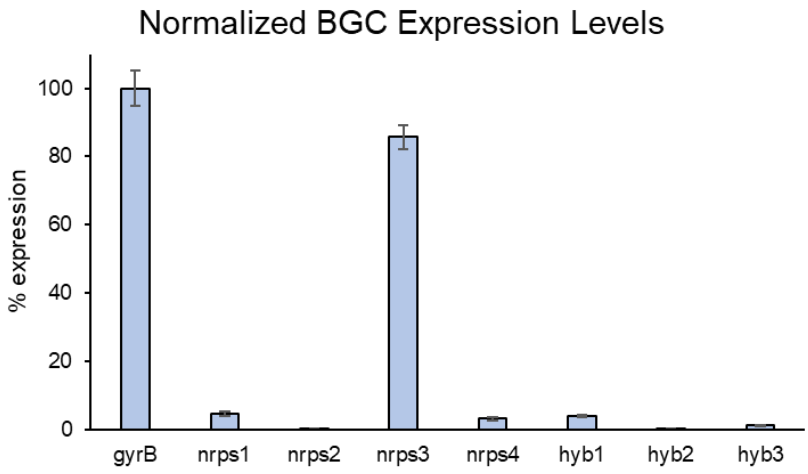


Figure 2-11. Expression level of BGCs of *Csa.* Floor gene expression values across PKS/NRPS-containing operons are normalized against a housekeeping benchmark, *gyrB*

Further studies are necessary to determine the function of compound **1** in solvent stress response. One hypothesis for the function of **1** is that it is a signaling molecule. As a relatively hydrophobic small molecule it could feasibly diffuse through the cell membrane or utilize promiscuous dipeptide transporters found in bacteria. Discovery of a dedicated receptor for **1** would corroborate this hypothesis. During feeding, however, we observed that the wild-type growth (**Figure 2-10**) was less affected than that of Δ*nrps3* regardless of exogenously supplied compound. This is likely due to the high additional solvent stress presented by the DMSO necessary to apply **1** to the cultures. An alternative interpretation of this result is that compound **1** has an indirect benefit and the biosynthesis itself confers the solvent tolerance benefit. For example, high expression of *nrps3* early in fermentation could result in biosynthesis of **1** during the subsequent acidogenic stage of fermentation. This could provide a means of balancing redox cofactors or depleting butyrate, which has been proposed to mediate toxic effects of solvent stress.[104] Understanding the mechanism of *nrps3*-mediated solvent tolerance could lead to insights for engineering more robust ABE-fermenting clostridia.

2.4. Materials and Methods

Bacterial strains and growth conditions. All strains used in this study are listed in **Table 2-1**. *E. coli* strains were cultured in lysogeny broth (LB) containing the appropriate antibiotics at 37°C. Cloning was performed in an *E. coli* XL1-blue background. Protein work and *E. coli in vivo* assays were performed in *E. coli* BAP1.[105] *Csa* was cultured at 37°C in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) containing an atmosphere of 97% nitrogen and 3% hydrogen. For routine culture and genetics work, *Csa* was incubated in 2x YTG (16 g/liter tryptone, 10 g/liter yeast extract, 5 g/liter NaCl, 10 g/liter glucose, 15 g/liter agar for solid media) supplemented with 40 µg/ml erythromycin when necessary and adjusted to pH 6.5 with 1 N HCl.

Table 2-1. Strains and plasmids used in this chapter

Bacterial strain or plasmid	Relevant characteristics	Source or reference(s)
E. coli strains		
XL1-blue	Cloning strain	Agilent
BAP1	Heterologous expression strain	Pfeifer *et al*.[105]
Csa strains		
N1-4 (HMT)	Wild-type strain	ATCC 27021
N1-4 Δ*nrps3*	CSPA_10760 deletion strain	this study
Plasmids		
pXZ247	pETDuet-1 vector expressing CSPA_RS10760	Novagen/EMD Millipore
pNK52	*E. coli-Clostridium* shuttle vector; Ery^R; *repL*; Amp^R; colE1; P_{bdh}:spCas9; P_{sCbei_5830}:sgRNA; homologous flanking sequences	this study

Transcriptomic analysis. Fermentation samples were obtained 12 h after inoculation. 1 ml triplicate samples were pelleted and suspended in 1 ml TRIzol reagent (Ambion). These were stored at -80°C before processing based on a procedure modified from a previous method.[106] Samples were thawed and processed according to the manufacturer's instructions; samples were spun down and supernatants were extracted in chloroform. The aqueous phase of the three-phase mixture was pipetted into a new RNase-free centrifuge tube and mixed with 1 volume 70% ethanol, then applied to a RNeasy Mini Kit (Qiagen North America, Germantown, MD) for cleanup. The mixture was filtered by spinning through another Qiagen mini kit column, followed by the standard RW1 and RPE washes, and RNA was obtained in 60 µl DEPC-treated water. Residual genomic DNA was depleted using RQ1 DNase (Promega, Madison, WI). The samples were adjusted to 100 µl with DEPC-treated water, mixed with 350 µl RLT buffer and 250 µl ethanol, and returned to Qiagen spin columns for desalting according to the manufacturer's cleanup procedure. This yielded 60 µl samples containing 0.3-2.4 µg total RNA. RNA quality control, library construction, and library sequencing were performed by the University of California-Berkeley QB3 Functional Genomics Laboratory and Vincent J. Coates Genomic Sequencing Laboratory. RNA quality and concentration were assessed using a nanochip on an Agilent 2100 Bioanalyzer. Bacterial 16S and 23S rRNA was depleted using a RiboZero Kit (Illumina, San Diego, CA). The remaining RNA was converted to an RNA-seq library using an Illumina mRNA-Seq library construction kit. RNA library sequencing was performed on an Illumina HiSeq4000 instrument using 100 bp paired-end

reads. After adapter trimming, reads were quality controlled using FastQC.[107] Base calls with a Phred score > 30 were kept for downstream analysis. Quality-controlled reads were mapped to the *Csa* chromosome (Refseq: NC_020291.1) and megaplasmid (Refseq: NC_020292.1) using Bowtie2.[108] Read counts were extracted using HTSeq[109] and normalized to gene length. Differential expression analysis between wild-type and *Δnrps3* strains was carried out using DESeq2.[110]

Plasmid construction. Oligonucleotides were provided by IDT (Coralville, IA). Relevant synthetic oligonucleotides are presented in **Table 2-2**. Phusion polymerase (NEB, Ipswich, MA) was used for all PCR reactions. Detailed cloning procedures are provided under Supplementary Methods. All synthetic oligonucleotide sequences are described in Supplementary Table S2. After Gibson assembly,[111] the reaction mix was transformed into chemically competent *E. coli* XL1-blue (Agilent Technologies, Santa Clara, CA). Clones were isolated and DNA was extracted using a plasmid Qiagen Miniprep kit. Constructs were validated by restriction digest patterning and Sanger sequencing. Detailed construction notes for pNK52 and pXZ247 are presented in **Appendix A**.

Table 2-2. Synthetic oligonucleotides used in this chapter

Oligonucleotide name	Sequence
Pcbei_F	ATGCTCTGACGCTTAAATGC
Pcbei_R_nrps3	AAACTCTGAAACTATGTGATGTATCTCGAGATGGTGGAATGATA
gRNA_F_nrps3	CGAGATACATCACATAGTTTCAGAGTTTTAGAGCTATGCTGTTTTGG
gRNA_R	TCTTCAAGATCTTTTAGGTGGTC
UHR_nrps3_Fo	TTTTTGCGGCCGCGACCACCTAAAAGATCTCAGGAGCTCAAGTTGGTTTAGA
UHR_nrps3_Ro	ACTTTCATCTAGTTTACCCCATTCCCCGGGCTCGCAGCTCACTATTTAATACG
DHR_nrps3_Fo	GAATGGGGTAAACTAGATGAAAGT
DHR_nrps3_Ro	GGAGTTACCTTAAATGGTAACTCAACTAGTTCTGCATCTTCCTTATACCTTTG
US_nrps3_F	CTTAGATAGTGACAATACCCATGAA
wt_nrps3_R	ACTTCTATTAAAGCAGGCAGCA
DS_nrps3_R	GATGCTGATTGCTCACTAGTTG
p45_cas9n_Fo	AAATGTACAATGGATAAGAAATACTCAATAGGCTTAGCTATCGGCACAAATAGC
p45_cas9n_Ro	AAAGTTTAAACTCAGTCACCTCCTAGCTGACTCAAATCAAT
p43_bk_F	TAAACCCGCCTAAACTGC
p43_bk_Ro	TTAATGTTTTTTTAAGGCATTAGTACTAGTTCTAGAGCATTTAAGCGTCAGAGCAT
p43_bk_F2	ACTAATGCCTTAAAAAAACATTAAAG
p43_bk_R2	ACATTACCGTACTGGCGCC
p43_Pbdh_Fo	CTGCAGGCGCCAGTACGGTAATGTGGTAAGACGAACAGCAGAACT
p43_Pbdh_R	TTCTTTTCCTCCTCTTACACAC
p43_cas9_Fo	GCGTGTGTAAGAGGAGGAAAAGAATGTACAATGGATAAGAAATACTCAATAGGCT
p43_cas9_Ro	CGGGCAGTTTAGGCGGGTTTAAACTCAGTCACCTCCTAGCTGACT
p44_UHR_pyrF_Fo	AGTCGGTGCTTTTTTTGCGGCCGCGACCACCTAAAAGATCTTGAAGA
p44_UHR_pyrF_R	GCTTCACTTGGAGTTCTTCCT
p44_DHR_pyrF_Fo	AGAAGGAAGAACTCCAAGTGAAGCCCCGGGGGAAATGGTGGAGTCGTTAA
p44_DHR_pyrF_Ro	TTAATGTTTTTTTAAGGCATTAGTACTAGTGTCCTTTAGATGATATGCCTCC
synthetic gRNA	GTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
sCbei_5830 promoter sequence	ATAATCTTTAATTTGAAAAGATTTAAGGCTTATTTAAATAAAAAATATGAGGGAAGAAT TGATATAAATTTAATTTTGTTATTGTATTATGGTATGTATGGAATAAATTTAACATAAAG ACAGTAATAATGTTCTTGAATTTAGACTTTTTATGTGTTATCATTAACAAGTATCAAAA ATGACATTTAATAAATTAATAATAATTTTAAAAATATATTTTTGATAAAAGCAATGATTA ACATGGTTTGACGTCTGAGAAGAGACGATTTTCTCAATAGGAGAAATTAAGGTGCAA ACCCTTATCATTCCACCAT
pXZ247_F_BamHI	gcagccatcaccatcatcaccacagccaggatccgatgaacagtgttaatgaaaaatttg
pXZ247_R1	catatttaatttaggaatttctcc
pXZ247_F1	tattggcttaatatgtttaaagg
pXZ247_R_PstI	aagcattatgcggccgcaagcttgtcgacctgcagttataagaaattaacttttcacaa
P1	gctataatggagtgtaatggaaa
P2	gatgctgattgctcactagttg
P3	cctgaccatctaataacccatt

Generation of *Csa nrps3* knockout mutant. Wild-type *Csa* was transformed with pNK52 using a previously reported electroporation method,[112] with some adjustments. Briefly, competent cell stocks of *Csa* were prepared from overnight cultures (37°C, PL7 media) from glycerol stocks stored at -80°C. After reaching an OD_{600} of 0.6, overnight cultures were subcultured in 60 ml liquid 2x YTG (10% inoculum) and incubated for 3 to 5 h until reaching an OD_{600} of 0.6. The subcultures were centrifuged (room temperature, 3,500×g, 15 min), decanted, and the pellet was resuspended in 6 ml room-temperature EPB (270 mM sucrose, 5 mM NaH_2PO_4 pH 7.4). We found that competent cell stock could be stored at -80°C in 20% glycerol, albeit with some loss of transformation efficiency. Next, 500 µl aliquots of electrocompetent cells were transferred into 4 mm Bio-Rad (Hercules, CA) cuvettes pre-chilled to 4°C, and 2 µg plasmid DNA was added. Electric pulses were delivered by a Bio-Rad Gene Pulser Xcell with parameters as follows: mode, exponential pulse; voltage, 2.0 kV; resistance, 200 Ω; capacitance, 25 µF. Following electroporation (yielding time constants of ~4 ms), cells were immediately resuspended in 10 ml 2x YTG and allowed to recover for 16 h at 37°C. Recovery cultures were centrifuged again and concentrated in 1 ml fresh media. Next, 100 µl cells were plated on 2x YTG plates containing 40 µg/ml erythromycin. Colonies were picked and transferred into 10 ml liquid 2x YTG with antibiotic for 24-48 hours to allow Cas9-nickase mediated homologous recombination to delete 90% of the gene, using the recombination template present on pNK52 (Supplementary Fig. S1a). Dilutions were then plated on nonselective 2x YTG plates to allow curing of pNK52. After two days' incubation, colonies were replica plated on both nonselective and selective 2x YTG to detect successful loss of the plasmid. Colony PCR was used to screen for the genotype of the *Csa* Δ*nrps3* CRISPR mutants (Supplementary Fig. S1b).

Flask fermentations of *Csa*. Overnight cultures of *Csa* were prepared in a derivative of clostridial growth medium (CGM),[113] PL7 (30 g/liter glucose, 5 g/liter yeast extract, 2.67 g/liter ammonium sulfate, 1 g/liter NaCl, 0.75 g/liter monobasic sodium phosphate, 0.75 g/liter dibasic sodium phosphate, 0.5 g/liter cysteine-HCl monohydrate, 0.7 g/liter magnesium sulfate heptahydrate, 20 mg/liter manganese sulfate monohydrate, and 20 mg/liter iron sulfate heptahydrate, with the initial pH adjusted to 6.3 using 1 N HCl) from a single colony from a 2x YTG plate. At exponential phase, with OD_{600} of 0.4 to 0.6, a 4% inoculum was added to 25 ml of PL7G (PL7 with glucose increased to 35 g/liter, and 6 g/liter $CaCO_3$ for pH control) in loosely capped 50 ml centrifuge tubes to avoid pressurization. All fermentations were performed as biological triplicates in static batch culture. 1 ml samples were drawn at intervals of 12, 24, 48, and 72 h after inoculation. An additional 1 ml sample was drawn for metabolome analysis at the same time intervals.

Fermentation analytical procedures. Concentrations of acetone, butanol, ethanol, acetate, butyrate, lactate, and residual glucose were quantified using calibration curves generated on a Shimadzu Prominence UFLC system with refractive index and diode array detection (Shimadzu America, Inc., Columbia, MD). Prior to analysis, samples of culture supernatant were filtered using 0.22 µm polyvinylidene difluoride syringe (PVDF) filters (Restek, Bellefonte, PA). The resulting filtrate was resolved using a Bio-Rad Aminex HPX-87H column (300 mm by 7.8 mm) and detected by refractive index (for glucose, butanol, ethanol, acetate, and lactate), or by UV absorbance (for acetone, 265 nm; for butyrate, 208 nm). The method used a column temperature of 35°C, a 35 min run duration, and the manufacturer-recommended mobile phase (0.01 N H_2SO_4) at a flow rate of 0.7 ml/min.

Metabolomic analysis. For untargeted metabolomic analyses, samples were analyzed, in biological quadruplicate, using liquid chromatography-high resolution mass spectrometry (LC-HRMS). Samples containing 1 ml culture were extracted in 1 volume ethyl acetate. Mixtures were

vortexed and spun down (6000×g, 1 min). The upper phase solvent layer was pipetted into a new centrifuge tube and dried under N_2, then resuspended in 100 µl methanol. 10 µl injections were analyzed on an Agilent Technologies 6545 Accurate-Mass QTOF LC-MS instrument fitted with an Agilent Eclipse Plus C18 column (4.6x100 mm). The run method used a linear gradient of 2-98% CH_3CN (v/v) over 57 min in H_2O with 0.1% formic acid (v/v) at a flow rate of 0.5 ml/min. Metabolomics data were analyzed using XCMS[114] to identify metabolites unique to either strain. The following parameters were used: p-value < 0.0005, fold change > 10, peak intensity > 10,000.

***In vitro* reconstitution of the CSPA_RS10760 NRPS.** For *in vitro* assays, N-terminally his-tagged NRPS was overexpressed and purified from *E. coli* BAP1 pXZ247. Protein was purified as follows: A single colony was inoculated into 10 ml LB + 100 µg/ml carbenicillin for overnight growth at 37°C. About 5 ml was used to inoculate 700 ml LB + 100 µg/ml carbenicillin, and the culture was shaken at 240 rpm and 37°C until the OD_{600} reached 0.4. The culture was iced for ten minutes and isopropyl thio-β-D-1-galactopyranoside (IPTG) was added to a final concentration of 120 µM to induce protein expression. The culture was incubated at 16 °C for 16 h. Cells were harvested by centrifugation (5500×g, 4°C, 20 min), resuspended in 25 ml lysis buffer (50 mM HEPES, pH 8.0, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization over ice. Cell debris was removed by centrifugation (17,700×g, 4 °C, 60 min), and Ni-NTA agarose resin (Thermo Fisher Scientific, Waltham, MA) was added to the supernatant (2 ml/liter culture). The mixture was nutated at 4 °C for 1 h, loaded onto a gravity flow column, and the NRPS protein was eluted with increasing concentrations of imidazole in Buffer A (50 mM HEPES, pH 8.0, 1 mM DTT). Purified NRPS protein was concentrated and buffer exchanged into Buffer A + 10% glycerol using an Amicon Ultra-15 spin filter (MilliporeSigma, Burlington, MA) with nominal molecular weight cutoff of 100 kDa. Aliquots of purified NRPS protein were aliquoted and flash frozen in liquid nitrogen. The full *in vitro* reaction mixture contained 50 mM HEPES pH 8, 2 mM $MgCl_2$, 1 mM TCEP, 5 mM ATP, 4 mM NADPH, 5 mM L-valine, 5 mM L-leucine, 10 mM butyric acid, and 5 mM CoA, 0.01 mM Orf35 (a promiscuous CoA ligase)[115] and 0.01 mM NRPS in a 50 µl reaction volume. After 30 min, reactions were quenched with 2 volumes methanol and spun down (14000×g, 1 min) and the supernatant was collected for LC-MS analysis.

Substrate feeding in a heterologous host . *E. coli* BAP1 pXZ247 was cultured in 30 ml volumes of LB supplemented with 100 µg/ml carbenicillin at 37°C until the OD_{600} reached between 0.4 and 0.6. Cultures were cooled to 16°C and supplemented with 120 µM IPTG, 1 mM L-valine, 1 mM L-leucine, and 1 mM of various short-chain fatty acids (acetic acid, propionic acid, butyric acid, hexanoic acid, or octanoic acid). After 48 h incubation, cultures were spun down (4000×g, 5 min) and 10 ml of supernatant was extracted twice using 5 ml ethyl acetate. The resulting products were detected by LC-HRMS under the conditions described above.

Butanol challenge assays. Butanol challenge assays were modified from a method used in *C. acetobutylicum*.[116] *Csa* wild-type and Δ*nrps3* cultures (in biological quadruplicate) were incubated overnight at 37°C in 50 ml centrifuge tubes containing 30 ml PL7G. At the onset of exponential growth at 8 h, cultures were challenged with the addition of water or butanol to a final concentration of 0, 2.5, or 5 g/liter butanol. Sampling was performed at 2-3 h time intervals by taking 200 µl volumes. OD_{600} was monitored in 96-well plates (Corning, NY) using a SpectraMax M2 instrument (Molecular Devices, San Jose, CA).

Chemical complementation assays. Butanol challenge assays were performed as described above with two modification: cultures were scaled down to 10 ml and experimental cultures were supplemented with either 50 µl dimethylsulfoxide (as a control) or an equivalent volume of compound **1** solution, to a final concentration of 6 mg/liter.

Chapter 3. Genome mining in *Clostridium roseum*

3.1. Introduction

Clostridium roseum (*Cro*) is a Gram-positive, spore-forming obligate anaerobe. First discovered in 1935,[117] it was noted for its pink-orange pigmentation and studied as a model endospore-forming bacterium.[118–120] Isolates of this species have demonstrated broad industrial significance. In one study, *Cro* was detected and isolated from a Chinese strong-flavor baijiu fermentation operation.[121] Other reported strains can produce biofuels such as H_2 from diverse agricultural wastes including beer lees,[122] beet molasses,[77] and cattle waste sludge.[123] Still others have been detected in mixed microbial consortia utilizing wheat straw hydrolysate,[124] brewery yeast waste or rice straw compost.[125] *Cro* is also closely related to the Acetone-Butanol-Ethanol (ABE) fermenting type organism *Clostridium acetobutylicum* and has garnered interest for next-generation ABE bioprocessing,[64] with one strain reaching titers of 13.87 g/L butanol and 20.2 g/L total solvents when fermenting date fruits.[126]

Anaerobes, especially soil-associated clostridia such as *Cro*, have been highlighted as an untapped reservoir of genomic potential to biosynthesize natural products.[7,8,127] These remarkable small molecules are structurally and functionally diverse, making them an attractive source of bioactive scaffolds for development into drugs, agrochemicals, nutraceuticals, etc. Since the characterization of the first natural product from a *Clostridium* species in 2010,[33] the pace of discovery has burgeoned. More recent discoveries include the clostrubin antibiotics from *Clostridium beijerinckii* HKI0724 and *Clostridium puniceum*,[39,40] the clostrienose signaling molecules from *Clostridium acetobutylicum*,[53] and clostrindolin and clostrocyloin antibiotics from *Clostridium beijerinckii* HKI805.[42,43] The discovery of the clostrubins and clostrienoses support earlier suggestions that the clostridia are a promising and novel source for the discovery for polyketides.[127]

Clostridia are remarkable among the anaerobes for their enriched biosynthetic potential to produce thiotemplated "assembly-line" natural products, comprised of polyketides and nonribosomal peptides.[7] These secondary metabolites are produced by polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) genes, frequently in concert with additional tailoring enzymes. PKS/NRPS and associated biosynthetic genes often co-localize in biosynthetic gene clusters (BGCs) on the genome, facilitating *in silico* genome mining approaches for discovery of novel chemical.[128] One strain of *Cro* (DSM 6424) was highlighted for possessing several large PKS/NRPS BGCs in a recent review of secondary metabolism in anaerobes,[90] but little is known regarding the regulation or function of these BGCs, and no chemical structures have been reported for their associated metabolites.

Herein, we describe natural product discovery efforts by genome mining in *Cro* DSM 6424. We present bioinformatic and transcriptional analyses of the BGCs found in this anaerobe. We discuss the development of genetic manipulations to facilitate strain construction for *Cro* metabolomics studies, which leads to the discovery of a novel family of compounds. We describe purification of the major congeners associated with one BGC for structural elucidation. We then propose a biosynthetic pathway for the novel compounds which is supported by a heterologous expression study.

3.2. Results

3.2.1. Bioinformatics of *Cro*

Bioinformatic analyses were performed in AntiSMASH[91] and PRISM[92] to detect PKS/NRPS-containing BGCs in the 4.94 Mb genome of *Cro* (NCBI accession: LZYU00000000). A summary of the eight detected BGCs is presented in **Figure 3-1**. BGCs were trimmed conservatively to define predicted boundary limits by removing genes from the ends if they had non-biosynthetic functions or were not in a putative operon with a biosynthetic gene. MultiGeneBLAST[93] and BiG-SCAPE[94] were used to query for homologous BGCs from other organisms. Several of these BGCs are conserved in the genomes of other ABE fermenters species such as *Cro* DSM 7320, *Clostridium felsineum*, and *Clostridium aurantibutyricum*. The predictions for the number of substrates and their identities are described below for each BGC. Many BGCs have homologs in other ABE-fermenting *Clostridium* species or other bacteria, suggesting evolutionarily conserved functions within their shared clades or environmental niches.

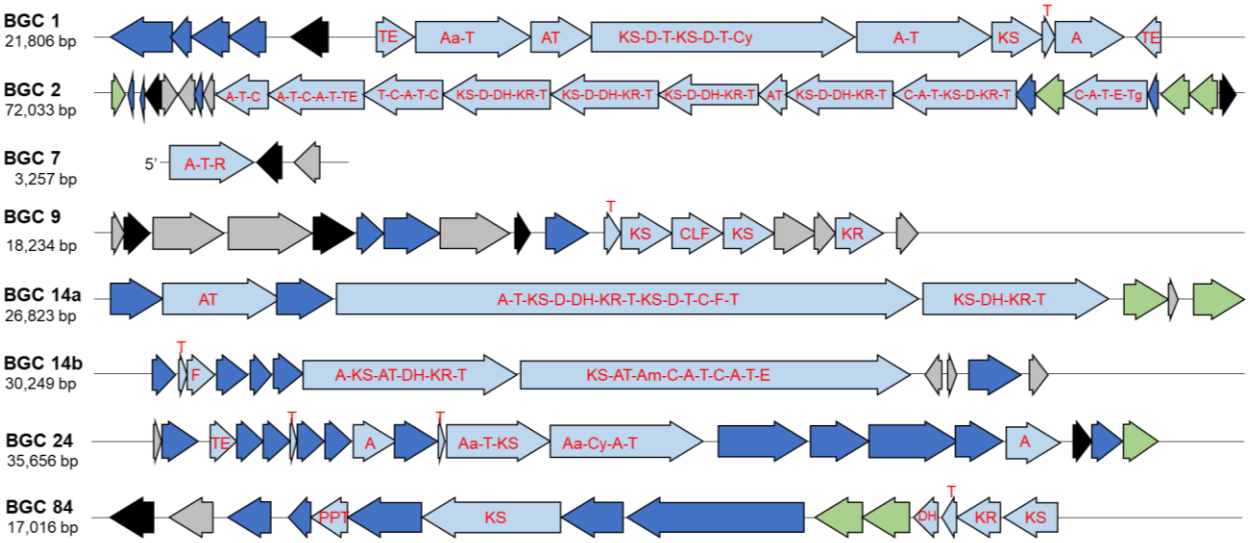


Figure 3-1. BGCs of *Cro* which contain PKS or NRPS genes. BGC numbering refers to the associated contig. Gene color represents putative function: light blue, PKS/NRPS gene; dark blue, other biosynthetic gene; black, regulatory; green, transporter; grey, hypothetical or other. BGC 7 is found at a contig edge. PKS/NRPS genes have the indicated domain architecture. Abbreviations: TE, thioesterase domain; Aa, acyl-activating domain; T, thiolation domain; E, epimerization domain; AT, acyltransferase domain; KS, ketosynthase domain; D, *trans*-AT docking domain; Cy, heterocyclization domain; A, adenylation domain; C, condensation domain; DH, dehydratase domain; KR, ketoreductase domain; Tg, TIGR01720 (a domain of unknown function); R, reductive release domain; CLF, chain-length factor; F, FkbH-like domain; Amino, aminotransferase domain; PPT, 4'-phosphopantetheinyltransferase domain

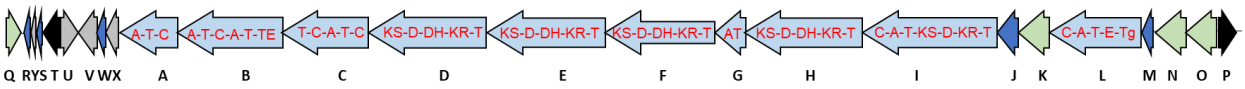
BGC 1 (**Figure 3-2**) features a hybrid PKS/NRPS system spread over nine open reading frames (ORFs). The thiotemplate assembly line is predicted to incorporate five substrates; the two *trans*-AT docking domains for malonyl-CoA incorporation by the free AT (ORF G), and ORF E is predicted to activate a Cys residue which may be cyclized by the Cy domain of ORF F. The other two substrate specificities are not clear from bioinformatic analyses. The other biosynthetic genes may be involved in biosynthesis of malonyl-CoA extender units.

ORF	Locus Tag	Putative Function
A	CLROS_00170	TE
B	CLROS_00160	A
C	CLROS_00150	T
D	CLROS_00140	KS
E	CLROS_00130	A-T
F	CLROS_00120	KS-D-T-KS-D-T-Cy
G	CLROS_00110	AT

ORF	Locus Tag	Putative Function
H	CLROS_00100	Aa-T
I	CLROS_00090	TE
J	CLROS_00080	LytR-like sensory transduction protein
K	CLROS_00070	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
L	CLROS_00060	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha
M	CLROS_00050	biotin carboxyl carrier protein of acetyl-CoA carboxylase
N	CLROS_00040	2-oxoglutarate carboxylase small subunit

Figure 3-2. Genes of BGC 1. Genes are colored by function: light blue, PKS/NRPS genes; dark blue, other biosynthetic; black, regulator. PKS/NRPS genes are labeled by domain organization: TE, thioesterase; Aa, acyl-activating; T, thiolation; A, adenylation; AT, acyltransferase; KS, ketosynthase; D, *trans*-AT docking; Cy, heterocyclization

BGC 2 (**Figure 3-3**) features a hybrid PKS/NRPS system spanning ten ORFs. The thiotemplate assembly line architecture suggests the product is composed of 12 substrates. ORF G is predicted to utilize malonyl-CoA at the five *trans*-AT docking domains (ORFs D, E, F, H, I). The predicted A-domain specificities are as follows: ORF B (C-terminal A-domain), Val; ORF C, ornithine; ORF H, Tyr; ORF I, His; ORF L, Asn. The remaining A-domains in ORFs A and B are for hydrophilic substrates. ORF J encodes a putative acetyltransferase which may have a biosynthetic or even self-resistance function if the product is a secreted antibiotic. ORF M encodes a putative reductase which may have a biosynthetic function.

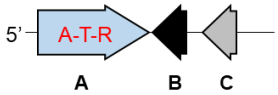


ORF	Locus Tag	Putative Function
Q	CLROS_01190	D-serine/D-alanine/glycine transporter
R	CLROS_01200	glucose 1-dehydrogenase
Y	CLROS_01210	hypothetical protein
S	CLROS_01220	transporter
T	CLROS_01230	putative diguanylate cyclase
U	CLROS_01240	endoglucanase E precursor
V	CLROS_01250	endoglucanase
W	CLROS_01260	succinyl-diaminopimelate desuccinylase
X	CLROS_01270	CapP-like capsule biosynthesis protein
A	CLROS_01280	A-T-C
B	CLROS_01290	A-T-C-A-T-TE
C	CLROS_01300	T-C-A-T-C
D	CLROS_01310	KS-D-DH-KR-T

ORF	Locus Tag	Putative Function
E	CLROS_01320	KS-D-DH-KR-T
F	CLROS_01330	KS-D-DH-KR-T
G	CLROS_01340	AT
H	CLROS_01350	KS-D-DH-KR-T
I	CLROS_01360	C-A-T-KS-D-KR-T
J	CLROS_01370	chloramphenicol acetyltransferase
K	CLROS_01380	ABC transporter
L	CLROS_01390	C-A-T-E-Tg
M	CLROS_01400	L-glyceraldehyde 3-phosphate reductase
N	CLROS_01410	ABC transporter
O	CLROS_01420	ABC transporter
P	CLROS_01430	AraC family transcriptional regulator

Figure 3-3. Genes of BGC 2. Genes are colored by function: green, transporter; light blue, PKS/NRPS; dark blue, other biosynthetic genes; black, regulator; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: TE, thioesterase; A, adenylation; T, thiolation; AT, acyltransferase; KS, ketosynthase; D, *trans*-AT docking; Cy, heterocyclization; C, condensation; E, epimerization; DH, dehydratase; KR, ketoreductase; Tg, TIGR01720 (a domain of unknown function, associated with condensation domains)

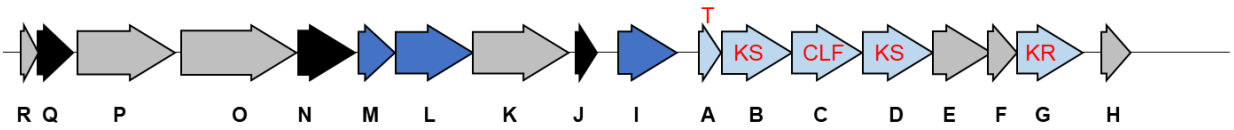
BGC 7 (**Figure 3-4**) comprises a one-module NRPS gene encoding a terminal reductive release mechanism. These may be commonly found in clostridia.[46] The A-domain of ORF A has uncertain substrate specificity. Due to its localization at the edge of a contig, this sequence may be a fragment of a larger BGC.



ORF	Locus Tag	Putative Function
A	CLROS_05460	A-T-R
B	CLROS_05470	regulator
C	CLROS_05480	hypothetical

Figure 3-4. Genes of BGC 7. This BGC may be incomplete due to its proximity to a contig edge. Genes are colored by function: light blue, NRPS; black, regulator; grey, hypothetical/other. Abbreviations: A, adenylation domain; T, thiolation; R, reductive release

BGC 9 (**Figure 3-5**) features a putative type II PKS spanning five ORFs. These have been suggested to be rare in bacteria outside of the Actinomycetes.[40] ORF C encodes a KS-like CLF domain characteristic of an iterative type II PKS. ORF I may be responsible for product release.



ORF	Locus Tag	Putative Function
R	CLROS_07990	hypothetical
Q	CLROS_08000	ferric uptake regulation protein
P	CLROS_08010	ribonuclease J
O	CLROS_08020	GTP-binding protein
N	CLROS_08030	YneA-like protein
M	CLROS_08040	O-methyltransferase
L	CLROS_08050	YhbU-like protease
K	CLROS_08060	penicillin-binding protein
J	CLROS_08070	RNA polymerase sigma-28 factor precursor

ORF	Locus Tag	Putative Function
I	CLROS_08080	hydrolase
A	CLROS_08090	T
B	CLROS_08100	KS
C	CLROS_08110	CLF
D	CLROS_08120	KS
E	CLROS_08130	hypothetical
F	CLROS_08140	hypothetical
G	CLROS_08150	KR
H	CLROS_08160	hypothetical

Figure 3-5. Genes of BGC 9. Genes are colored by function: light blue, PKS/NRPS; dark blue, other biosynthetic genes; black, regulator; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: T, thiolation; KS, ketosynthase; CLF, chain-length factor; KR, ketoreductase

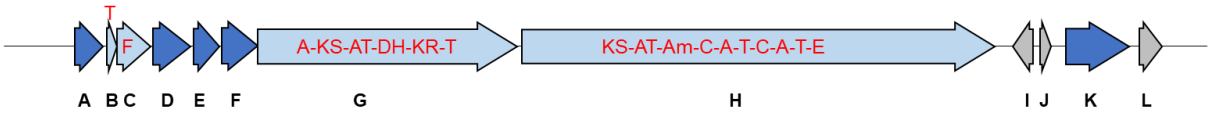
BGC 14a (**Figure 3-6**) encodes a five-module hybrid thiotemplate assembly line spanning three ORFs. The AT of ORF B is predicted to utilize malonyl-CoA in the two *trans*-AT docking domains of ORF D. The fifth module, found in ORF E, lacks a detectable AT or a *trans*-AT docking domain. ORF A encodes a putative aldolase from shikimic acid/aromatic acid metabolism. ORF C encodes a putative alcohol dehydrogenase. This BGC lacks a detectable terminal domain.

ORF	Locus Tag	Putative Function
A	CLROS_12250	phospho-2-dehydro-3-deoxyheptonate aldolase, Tyr-sensitive
B	CLROS_12260	AT
C	CLROS_12270	YjmD-like Zn-dependent alcohol dehydrogenase
D	CLROS_12280	A-T-KS-D-DH-KR-T-KS-D-T-C-F-T

ORF	Locus Tag	Putative Function
E	CLROS_12290	KS-DH-KR-T
F	CLROS_12300	transporter
G	CLROS_12310	hypothetical
H	CLROS_12320	ABC transporter

Figure 3-6. Genes of BGC 14a. Genes are colored by function: light blue, PKS/NRPS; dark blue, other biosynthetic genes; green, transporter; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: AT, acyltransferase; A, adenylation; T, thiolation; KS, ketosynthase; D, *trans*-AT docking; DH, dehydratase; KR, ketoreductase; F, FkbH-like

BGC 14b (**Figure 3-7**) encodes a five-module hybrid thiotemplate assembly line spanning four ORFs. The AT domains (ORFs G and H) have predicted substrate specificity for malonyl-CoA. The upstream ORFs A, D, E, and F may be related to precursor biosynthesis. None of the A domains have clear substrate specificities. This BGC lacks a recognizable terminal domain. A homologous BGC is found in *Lysinibacillus sphaericus* (Accession: JOTQ01000011), with identical domain predictions for ORFs A-G. The ORF H homolog appears to be fractured into three genes in the homologous BGC. Additionally, the homologous BGC possesses an extra NRPS module with predicted A domain specificity for Gly, and a downstream TE and PPT. This suggests the *Cro* BGC 14b may be incomplete or nonfunctional.

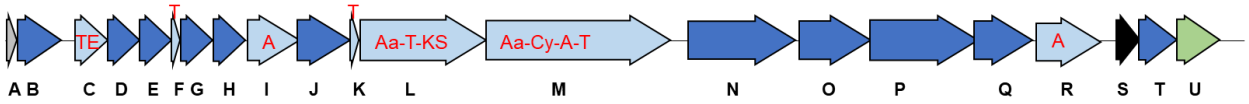


ORF	Locus Tag	Putative Function
A	CLROS_12330	3-hydroxybutyryl-CoA dehydrogenase
B	CLROS_12340	D-alanine-poly(phosphoribitol) ligase subunit
C	CLROS_12350	F
D	CLROS_12360	acyl-CoA dehydrogenase
E	CLROS_12370	acryloyl-CoA reductase electron transfer subunit gamma
F	CLROS_12380	acryloyl-CoA reductase electron transfer subunit beta

ORF	Locus Tag	Putative Function
G	CLROS_12390	A-KS-AT-DH-KR-T
H	CLROS_12400	KS-AT-Am-C-A-T-C-A-T-E
I	CLROS_12410	hypothetical
J	CLROS_12420	thioredoxin
K	CLROS_12430	SAM-dependent methyltransferase
L	CLROS_12440	hypothetical

Figure 3-7. Genes of BGC 14b. Genes are colored by function: light blue, PKS/NRPS; dark blue, other biosynthetic genes; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: T, thiolation; F, FkbH-like; AT, acyltransferase; A, adenylation; KS, ketosynthase; D, *trans*-AT docking; DH, dehydratase; KR, ketoreductase; Am, aminotransferase; C, condensation; E, epimerization

BGC 24 (**Figure 3-8**) encodes a hybrid assembly line spanning 13 ORFs. The substrate of ORF M is predicted to be a Cys which is likely cyclized by the Cy domain. The remaining substrates are unclear. ORFs NOPQ constitute a PKS-like polyunsaturated fatty acid biosynthesis cassette.[129]

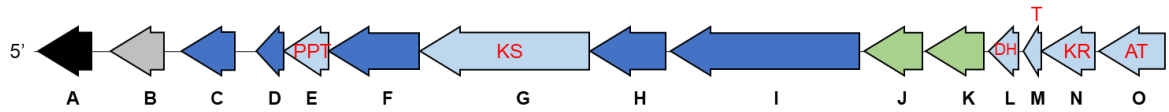


ORF	Locus Tag	Putative Function
A	CLROS_18090	hypothetical
B	CLROS_18100	2,6-dihydropseudooxynicotine hydrolase
C	CLROS_18110	TE
D	CLROS_18120	acryloyl-CoA reductase electron transfer subunit gamma
E	CLROS_18130	acryloyl-CoA reductase electron transfer subunit beta
F	CLROS_18140	T
G	CLROS_18150	FabH-like KS
H	CLROS_18160	acyl-CoA dehydrogenase
I	CLROS_18170	A
J	CLROS_18180	oxygen-independent coproporphyrinogen-III oxidase
K	CLROS_18190	T

ORF	Locus Tag	Putative Function
L	CLROS_18200	Aa-T-KS
M	CLROS_18210	Aa-Cy-A-T
N	CLROS_18220	PfaA-like protein
O	CLROS_18230	PfaB-like protein
P	CLROS_18240	PfaC-like protein
Q	CLROS_18250	PfaD-like protein
R	CLROS_18260	A
S	CLROS_18270	MarR family regulator
T	CLROS_18280	4Fe-4S dicluster domain
U	CLROS_18290	transporter

Figure 3-8. Genes of BGC 24. Genes are colored by function: light blue, PKS/NRPS; dark blue, other biosynthetic genes; green, transporter; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: TE, thioesterase; T, thiolation; A, adenylation; Cy, Aa, acyl-activating; KS, ketosynthase; Cy, heterocyclization

BGC 84 (**Figure 3-9**) contains genes which constitute a PKS assembly line. ORF G is a didomain of an intact and a non-functional (C576G point mutation in the active site) KS which may function analogously to a CLF.



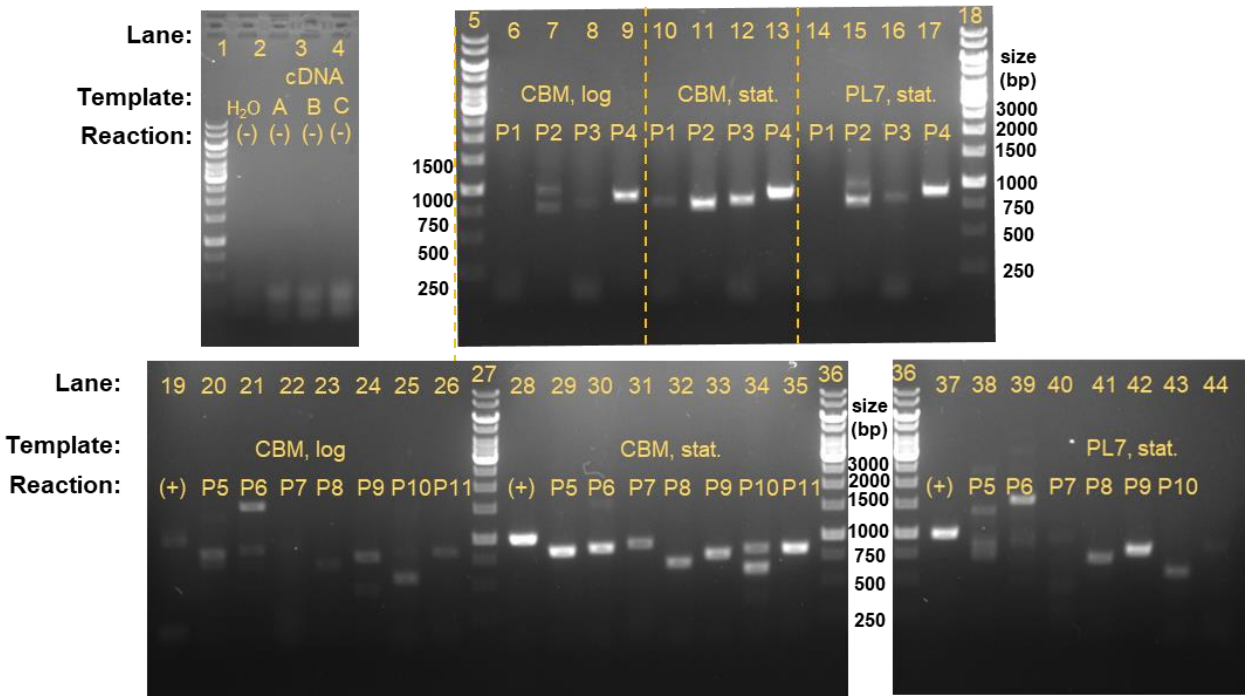
ORF	Locus Tag	Putative Function
A	CLROS_37260	HTH-type transcriptional regulator
B	CLROS_37270	hypothetical
C	CLROS_37280	4Fe-4S domain
D	CLROS_37290	esterase
E	CLROS_37300	PPT
F	CLROS_37310	glycosyltransferase
G	CLROS_37320	FabF-like KS
H	CLROS_37330	aminomethyltransferase

ORF	Locus Tag	Putative Function
I	CLROS_37340	N-acetyltransferase
J	CLROS_37350	ABC transporter
K	CLROS_37360	ABC transporter
L	CLROS_37370	FabZ-like DH
M	CLROS_37380	T
N	CLROS_37390	FabG-like KR
O	CLROS_37400	FabD-like AT

Figure 3-9. Genes of BGC 84. This BGC lies on a contig edge. Genes are colored by function: light blue, PKS/NRPS; dark blue, other biosynthetic genes; green, transporter; grey, hypothetical/other. PKS/NRPS genes are labeled by domain organization: PPT, 4'-phosphopantetheine transferase; KS, ketosynthase; DH, dehydratase; T, thiolation; KR, ketoreductase; KS, ketosynthase

3.2.2. Transcriptional analyses of *Cro*

Targeted transcriptional analyses were undertaken to determine whether the BGCs were expressed during culture.[130] *Cro* was cultured in PL7 media to late exponential phase and in CBM to mid-exponential and late stationary phase (to characterize transcription before and after biofilm formation). RNA was extracted for reverse-transcriptase PCR analysis. As a negative control, the purified RNA was targeted for PCR amplification of *hsdS* (locus tag CLROS_01700) prior to reverse transcription to check for contaminating genomic DNA. Next, the gene expression at different loci in the BGCs was assayed. The results (**Figure 3-10**) demonstrate that the majority of the BGCs are expressed during culture, with band intensities of BGCs 1, 9, 14a, 14b, 24, and 84 comparable to that of housekeeping gene *gyrB* (locus tag CLROS_30180). BGC 2 demonstrates conditional expression in liquid culture, with transcripts from both targeted loci seen only during stationary phase growth in CBM.



PCR Reaction	Target	ORF	Exp. Size (kb)	Primers		Relative Expression: CBM log	CBM stationary	PL7 stationary
(-)	*hsdS*	-	1.25	SN17	SN18	0	0	0
(+)	*gyrB*	-	0.99	JL276	JL277	1	4	4
P5	BGC 1	F	0.83	JL252	JL253	2	4	2
P6	BGC 1	B	0.87	JL254	JL255	1	4	1
P7	BGC 2	L	0.92	JL256	JL257	0	3	1
P1	BGC 2	B	0.82	JL258	JL259	0	2	0
P2	BGC 9	B	0.78	JL260	JL261	2	4	3
P8	BGC 9	G	0.69	JL262	JL263	1	3	2
P9	BGC 14a	D	0.78	JL264	JL265	2	4	4
P3	BGC 14b	H	0.82	JL266	JL267	1	4	2
P10	BGC 24	L	0.84	JL268	JL269	2	3	2
P11	BGC 24	R	0.83	JL270	JL271	1	4	1
P4	BGC 84	G	0.89	JL274	JL275	4	4	4

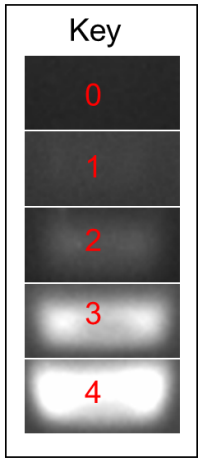


Figure 3-10. Transcriptional analysis of *Cro* BGCs in colony. Reverse-transcriptase PCR targeting different loci give a semi-quantitative assessment of BGC expression relative to that of a housekeeping gene, *gyrB*.

3.2.3. Establishing a genetic toolkit for *Cro* BGC knockouts

Initial efforts to introduce DNA into *Cro* tested electroporation-based methods. The protocols described for *Clostridium saccharoperbutylacetonicum* (**Chapter 2**) and *Clostridium beijerinckii* B-598 (**Chapter 4**) yielded no transformants despite strong growth of *Cro* in the media tested. Electroporation efficiency can often be boosted in Gram-positive organisms by digesting the cell wall or disrupting its assembly[131] to reduce the barrier to entry for DNA. Well known clostridial cell wall disruption agents such as lysostaphin,[132] ampicillin,[133] and lysozyme[133] were tested independently and in combination before transformation in PEG-containing medium. Glycine supplementation, known to interfere with D-Ala incorporation into peptidoglycan during culture,[134] was also tested. Heat shock treatment, demonstrated to inactivate native DNase activity in *Clostridium acetobutylicum* SA-1,[134] was also tested. Protoplast stabilization agents such as choline,[135] sodium polyanethol sulfonate,[133] and spermidine[136] were tested during recovery of cell cultures. Although cultures grew during recovery in nonselective media, no antibiotic-resistant transformants were recovered. One possible explanation for the results is presence of a putative type I restriction-modification enzyme complex (locus tags CLROS_1680 to CLROS_1720) in the *Cro* genome.

Bacterial conjugation methods can bypass native restriction-modification barriers to DNA delivery,[137] possibly due to the single-stranded nature of the DNA payload.[138] DNA constructs featuring a variety of modular plasmid parts[139] were successfully conjugated into *Cro* using *E. coli* WM6026, a donor strain with *meso*-2,6-diaminopimelic acid (DAP) auxotrophy to facilitate counter-selection.[140] The optimized procedure uses a *Cro* recipient culture prepared in PL7 liquid media to facilitate even spreading; this is then plated onto CBM agar during the mating step. The nutritionally sparser CBM was crucial for successful conjugation. Mating on agar rather than liquid media was required for detectable conjugation, consistent with observations that RK2-based transfer efficiency increases 3-4 orders of magnitude on agar.[141] DNA from *Cro* conjugants was extracted and transformed back into *E. coli* XL1-blue to confirm intact plasmid transfer. Conjugation efficiency ranged from 1×10^{-6} to 1.8×10^{-7} (**Table 3-1**). The chloramphenicol/thiamphenicol resistance marker *catP* was selected for subsequent work due to lower background growth after conjugation.

Table 3-1. Combinations of plasmid parts tested in *Cro*. Average conjugation efficiencies are presented

Plasmid	Gram-Positive Ori	Marker	Gram-Negative Ori	Efficiency
pMTL83353	**pCB102**	***aad9***	**ColE1**	**6.1×10^{-7}**
pMTL82151	**pBP1**	***catP***	**ColE1**	**1.3×10^{-6}**
pMTL84151	**pCD6**	***catP***	**ColE1**	**1.8×10^{-7}**
pSN1	**pCB102**	***catP***	**ColE1**	**1.4×10^{-6}**
pSN4	**pCB102**	***mls***	**ColE1**	**1.0×10^{-6}**

Next, the ClosTron[142] system was evaluated for generation of targeted knockout mutations in *Cro* due to the scalability afforded by facile retargeting. This targeted retro-transposon system requires two genes: *ltrB*, encoding a group II self-splicing intron; and *ltrA*, encoding a protein with DNA nicking and reverse transcription activity. Together, the *ltrAB* gene products produce a ribonuclear protein complex, targeted by a short sequence on *ltrB*, to undergo genomic integration of the intron at a specific locus. Recent iterations of this concept utilize a retrotransposition activated selectable marker to facilitate identification and isolation of successful gene knockout events.[143] As a proof-of-concept, the pyrimidine biosynthesis gene *pyrF* (CLROS_12550) was targeted for knockout because it confers a well-known phenotype, resistance to 5-fluorouracil.[144] The two genes *ltrA* and *ltrB* (bearing a retrotransposition-activated kanamycin marker) were introduced to *Cro* on a plasmid, pJL77. *Cro* is sensitive to kanamycin, but either the promoter, marker, or activation mechanism (a nested type I intron) proved non-functional in this strain and therefore no mutants could be isolated by kanamycin selection. Thus, while the plasmid and the intron insertion events were detected at the target locus by PCR screening, it proved challenging to resolve insertion mutants from unedited plasmid-bearing cells by restreaking. We isolated *Cro pyrF::ltrB* mutants by 5-fluorouracil counter-selection of the wild-type. After curing of pJL77, *pyrF::ltrB* was rigorously genotyped relative to an unedited *Cro* strain bearing an empty vector, pMTL82151 (**Figure 3-11**).

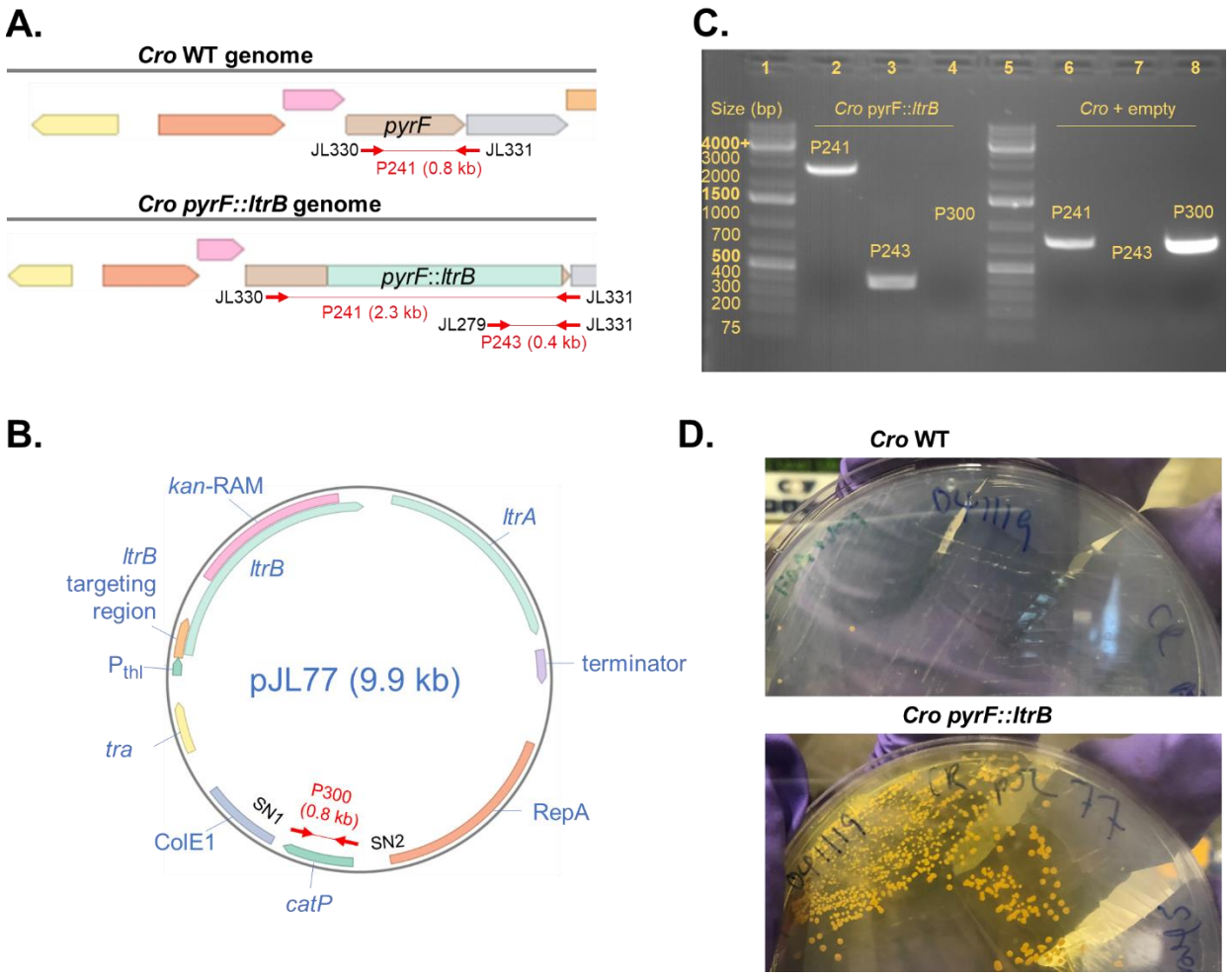


Figure 3-11. Generation of *Cro pyrF* by ClosTron system. (A) Genotype of wild-type *Cro* and *Cro pyrF::ltrB* mutant. Red arrows denote primer binding sites. (B) Map of knockout vector pJL77. Abbreviations: *kan*-RAM, retrotransposition activated kanamycin resistance marker; P_{thl}, constitutive promoter; *tra*, conjugation-associated sequences; ColE1, Gram-negative origin; *catP*, chloramphenicol/thiamphenicol resistance marker; RepA, Gram-positive origin. (C) PCR confirmation of mutant chromosomal genotype and plasmid curing. (D) Robust growth of *Cro pyrF::ltrB* on CBM + 5-fluorouracil

Due to the lack of a selectable marker for knockout mutations, we explored strategies to improve genetic screening by reducing aggregation of *Cro* in culture. Colonies of *Cro* form a cohesive, elastic mass as they mature, presenting a major barrier to the isolation of isogenic colonies after mutagenesis. Inoculation of liquid cultures from single colonies yielded mucus-like biofilms. At early growth (OD_{600} < 1), the film could be vortexed into a macroscopically homogenous culture capable of being restreaked to yield single colonies. However, colonies invariably had mixed populations containing both the wild-type and knockout genotypes detectable by PCR. These observations suggested that tight cell-cell associations resulted in deposition of multicellular colony-forming units during restreaking. Microscopic images

confirmed this hypothesis, showing end-to-end joining of cells into filamentous forms (**Figure 3-12**). Inspired by mycobacterial culture methods, we tested the effect of surfactants on dispersion of *Cro* liquid culture.[145] Among tested surfactants, tween 80 was able to improve the relative dispersion of *Cro* cultures (**Figure 3-12**) and enable recovery of *Cro* mutants by screening rather than selection. The benefit of tween 80 was demonstrable in liquid cultures sub-cultured from vortexed liquid culture, but not in liquid culture inoculated from colony, suggesting that the surfactant interferes with biofilm formation rather than disrupting biofilm structure.

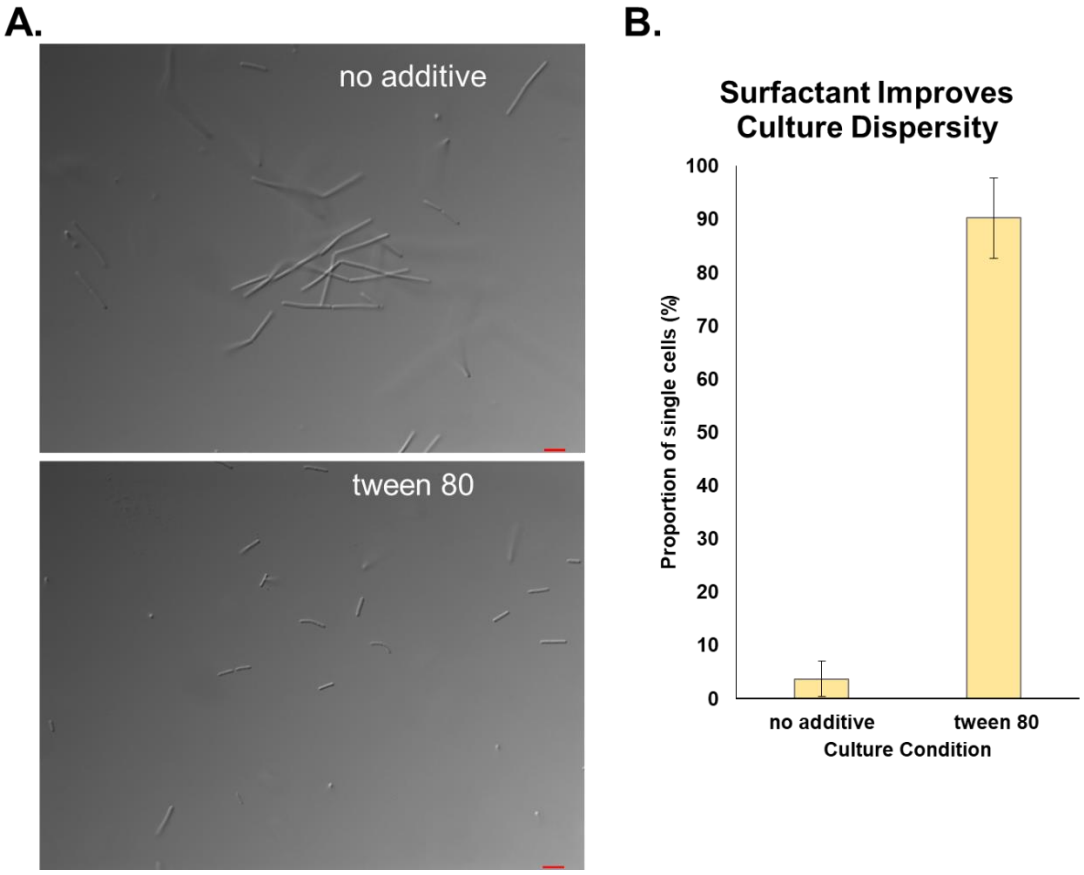
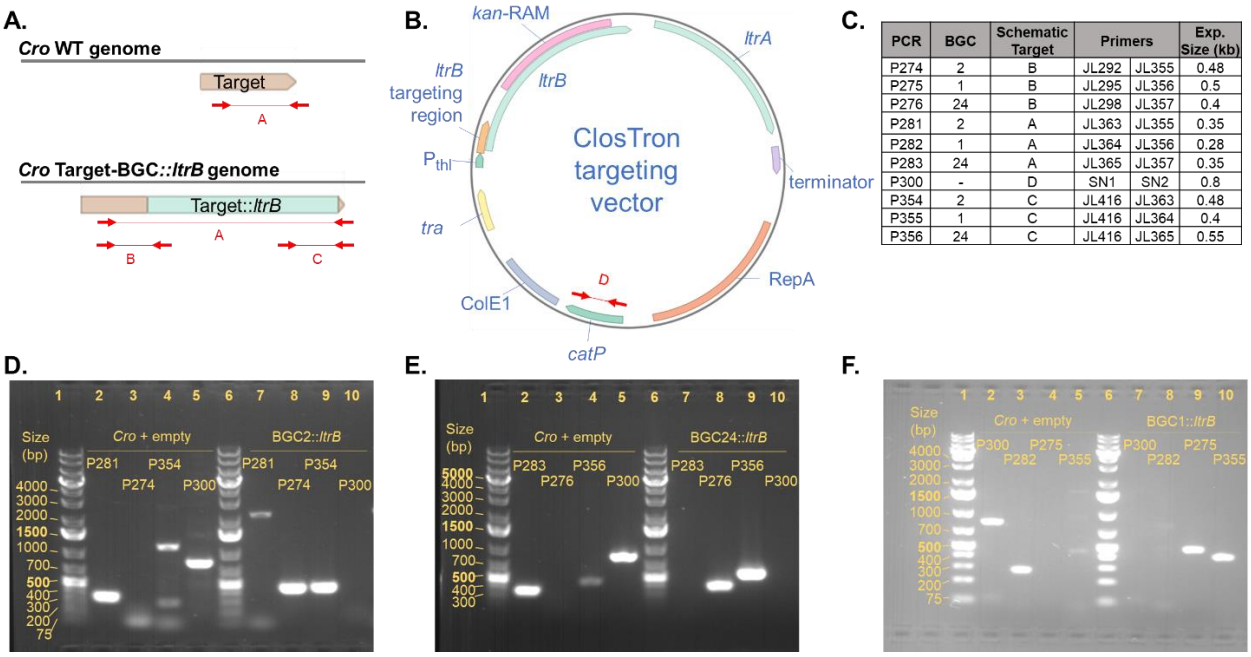


Figure 3-12. Tween 80 surfactant reduces cell aggregation of *Cro.* (A) Representative phase-contrast micrographs of cultures with and without surfactant. Red scale bars represent 5 μm distances. (B) Fraction of cells not found in aggregates

Next, we targeted seven *Cro* BGCs for intron-mediated knockout (**Table 3-2**). Each knockout vector was conjugated into wild-type *Cro*. Then, conjugants were passaged in liquid media (selective for the plasmid) supplemented with tween 80 and plated once per day. Colonies from these plates were screened by PCR. After two to three rounds of subculture, mutants for BGC 1, 2, and 24 were successfully isolated, cured of plasmid, and genotyped by PCR (**Figure 3-13**). The remaining mutants could not be isolated from unedited wild-type background despite many rounds of surfactant-supplemented subculture, suggesting another bottleneck in either intron targeting efficiency or mutant culture properties such as growth or aggregation. We reasoned that the latter is more likely because higher Perutka score[146] for the designed intron targeting sequences did not correlate with successful mutant isolation.

Table 3-2. *Cro* BGC knockout targets. Insertion site describes locus and sense/antisense orientation of the transposon.

Construct	BGC	Locus Tag	Domain Architecture	Insertion Site
pJL89	1	CLROS_00120	KS-D-T-KS-D-T-Cy	4663s
pJL88	2	CLROS_01290	A-T-C-A-T-TE	3934s
pJL85	9	CLROS_08090	T	244s
pJL106	14a	CLROS_12290	KS-DH-KR-T	1534s
pJL86	14b	CLROS_12400	KS-AT-Am-C-A-T-C-A-T-E	618a
pJL90	24	CLROS_18200	Aa-T-KS	3088s
pJL87	84	CLROS_37320	FabF-like KS	2176s



PCR	BGC	Schematic Target	Primers		Exp. Size (kb)
P274	2	B	JL292	JL355	0.48
P275	1	B	JL295	JL356	0.5
P276	24	B	JL298	JL357	0.4
P281	2	A	JL363	JL355	0.35
P282	1	A	JL364	JL356	0.28
P283	24	A	JL365	JL357	0.35
P300	-	D	SN1	SN2	0.8
P354	2	C	JL416	JL363	0.48
P355	1	C	JL416	JL364	0.4
P356	24	C	JL416	JL365	0.55

Figure 3-13. Isolation of ClosTron targeted BGC knockouts. (A) Schematic of the genotypes of edited and unedited *Cro* at a given target locus. Red markers represent PCR amplicons used for genotyping. (B) Schematic of vectors for ClosTron-mediated gene knockout. Abbreviations: *catP*, antibiotic resistance marker; RepA, Gram-positive origin; ColE1, Gram-negative origin; P_{thl}, constitutive promoter (C) Summary of BGC validation PCRs. (D) Genotyping of BGC2::*ltrB.* (E) Genotyping of BGC24::*ltrB*. (F) Genotyping of BGC1::*ltrB*

One of the unsuccessful BGC knockout targets (14a) was selected to test a suicide vector knockout strategy. The non-replicating plasmid can undergo targeted double-crossover allelic exchange to generate a selectable chromosomal genotype (**Figure 3-14a**).[147] After conjugation, the resulting variance in *Cro* colony size and morphology (**Figure 3-14b**) suggested the presence of distinct populations of cells and supported the hypothesis that disrupting this BGC resulted in a growth defect. We observed larger colonies with attenuated pigmentation and smaller colonies with abolished pigmentation; these were hypothesized to be single- and double-crossover mutants, respectively. Both colony types were picked for screening, and double-crossover mutants (ΔBGC14a) were isolated from the small colonies after one round of surfactant-supplemented subculture (**Figure 3-14c**).

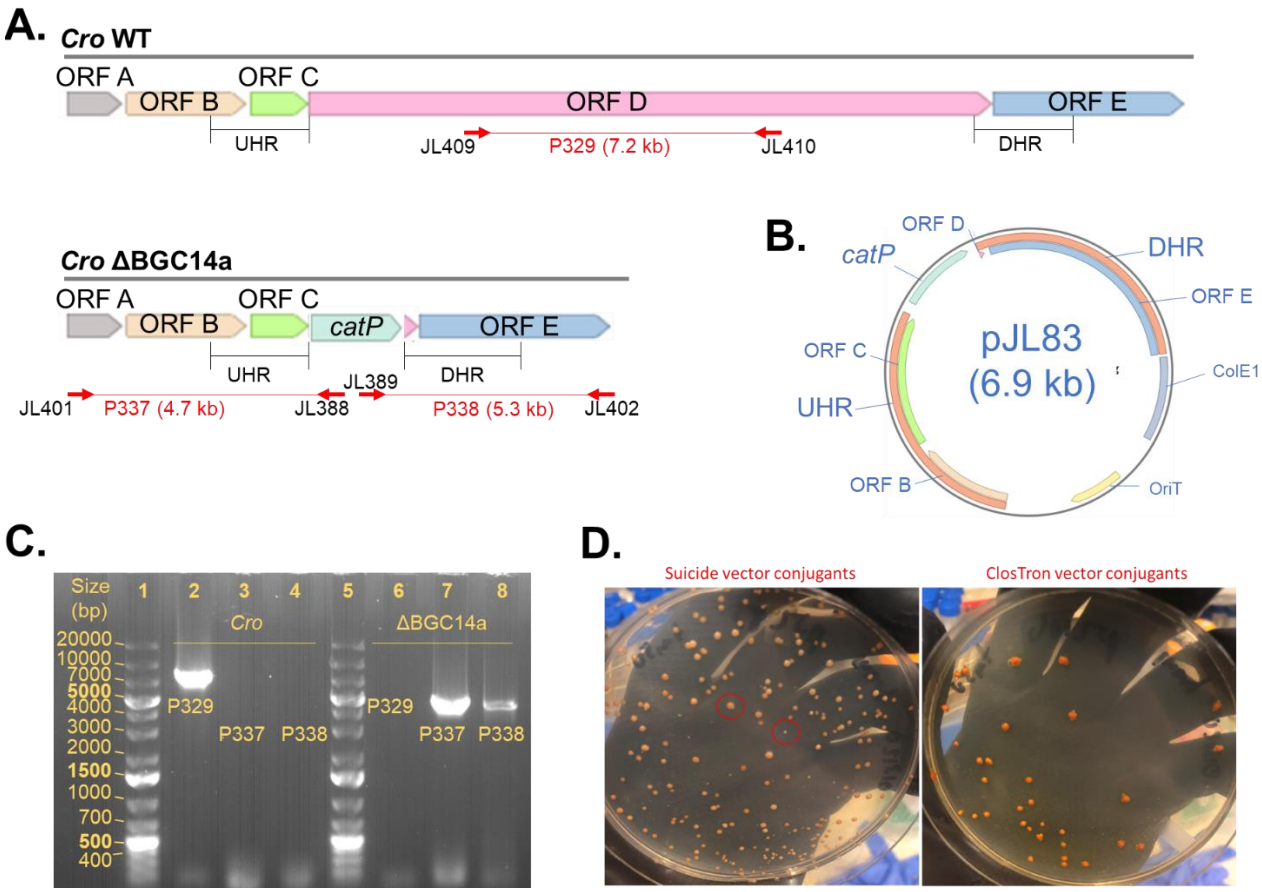


Figure 3-14. Generation of *Cro* ΔBGC14a. (A) Genotype of *Cro* wild-type and ΔBGC14a. Red arrows denote PCR primer binding sites. UHR, upstream homology region; DHR, downstream homology region. (B) Map of suicide vector pJL83. (C) PCR genotyping of *Cro* mutant. (D) Morphology of conjugant colonies differ by knockout method. Representative large and small colonies associated with single- and double crossover mutantions, respectively, are indicated by red circles

3.2.4. Identification of metabolites associated with BGC 1

The metabolites associated with BGC 1 knockout mutant, *Cro* BGC1::*ltrB*, were identified by comparative metabolomics. Chemical extracts from *Cro* wild-type and mutant were analyzed by liquid chromatography high resolution mass spectrometry (LC-HRMS). In positive ion mode, three peaks were identified with the corresponding predicted chemical formulas: $C_{22}H_{25}NO_6S$ (calculated for $C_{22}H_{26}NO_6S^+$: m/z 432.1481; found: 432.1467), $C_{29}H_{37}NO_{11}S$ (calculated for $C_{29}H_{38}NO_{11}S^+$: m/z 608.2166; found: 608.2135), and $C_{35}H_{47}NO_{16}S$ (calculated for $C_{35}H_{48}NO_{16}S^+$: m/z 770.2694; found: 770.2707) (**Figure 3-15**). These compounds are provisionally named BGC1-A, BGC1-B, and BGC1-C respectively. These formula predictions are consistent with earlier bioinformatic predictions that the products contain a Cys substrate, likely cyclized into a thiazoline moiety.

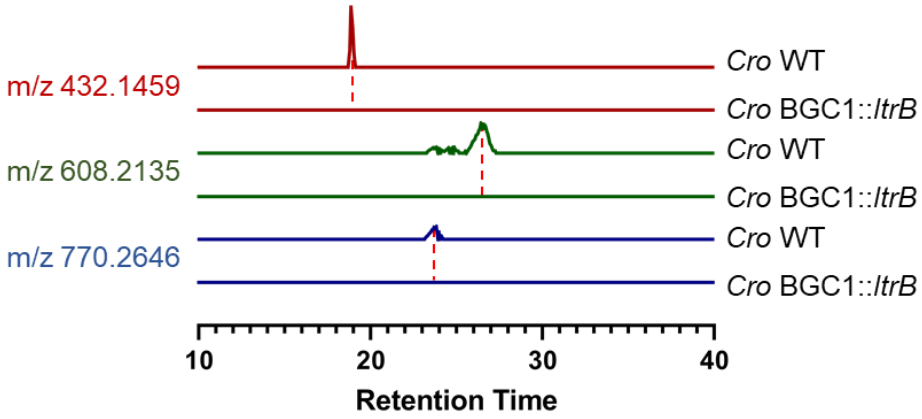


Figure 3-15. Metabolites associated with BGC 1. Shown are extracted HRMS chromatograms in positive ion mode.

Stable isotope labeled precursor feeding confirmed that Cys is a building monomer in the products. LC-HRMS analysis of *Cro* wild-type extracts from cultures fed with [1-^{13}C]-Cys demonstrated enrichment consistent with single incorporation in the mass spectra of all three BGC 1 products (**Figure 3-16**).

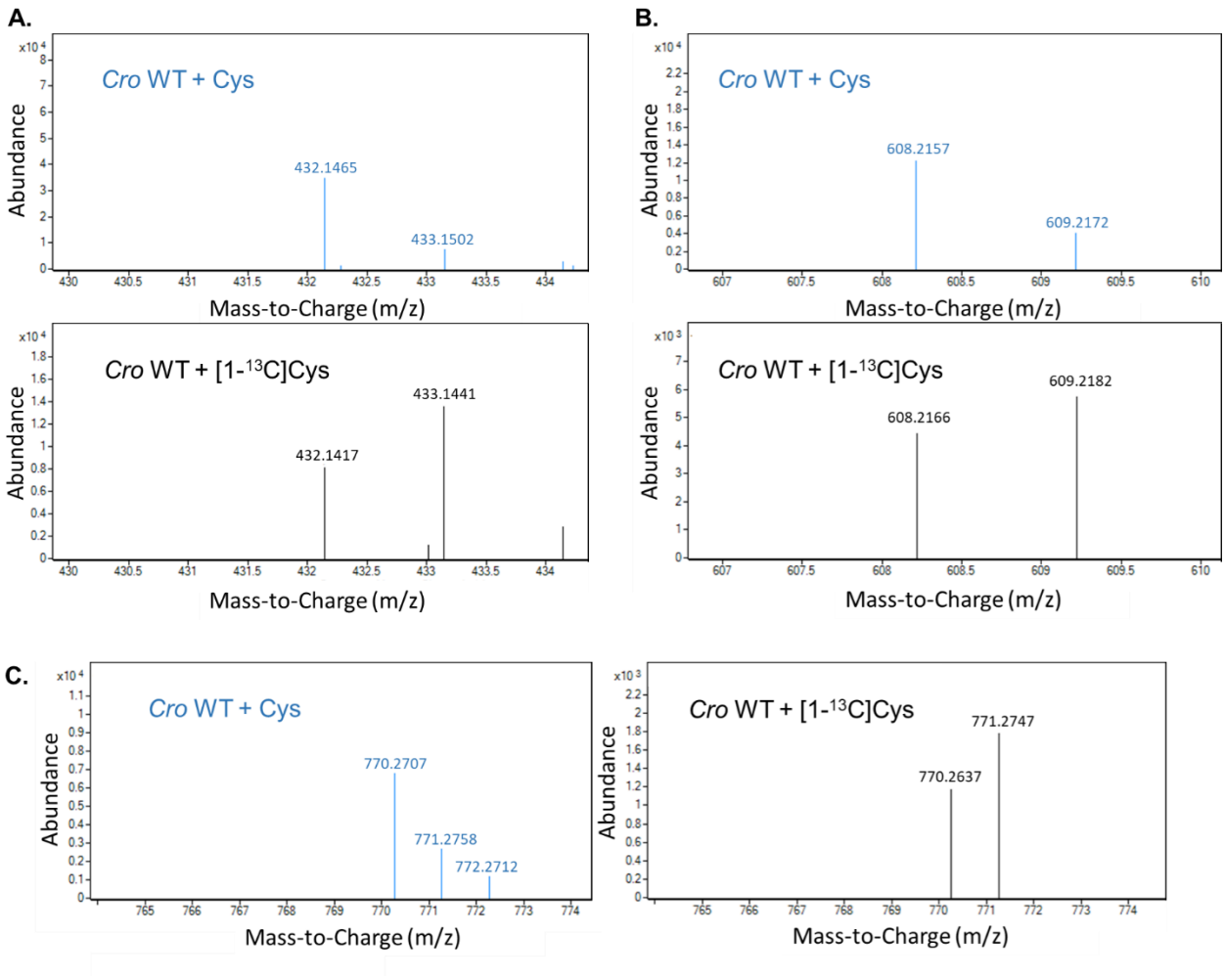


Figure 3-16. HRMS characterization of BGC 1 metabolites after precursor feeding. (A) Isotope labeling of BGC1-A. (B) Isotope labeling of BGC1-B. (C) Isotope labeling of BGC1-C

Tandem HRMS characterization of the labeled and unlabeled compounds (**Figure 3-17**) led to additional observations concerning the structural relationship between these compounds. In positive ion mode, a fragment with m/z identical to that of the parent ion of BGC1-B is found in the BGC1-C spectrum. The 162 Da mass difference between the two could be explained by glycosylation of BGC1-B. The two compounds also share two characteristic fragments with predicted chemical formulas of $C_{15}H_{21}NO_3S$ (calculated for $C_{15}H_{22}NO_3S^+$: 296.13204; found: 296.1332) and $C_{16}H_{21}NO_4S$ (calculated for $C_{16}H_{22}NO_4S^+$: 324.12695; found: 324.1277). In both spectra, these ions are labeled by [1-^{13}C]-Cys feeding. Analogously, the major fragment ions of BGC1-A corresponding to predicted chemical formulas of $C_{15}H_{19}NO_3S$ (calculated for $C_{15}H_{20}NO_3S^+$: 294.1163; found: 294.114) and $C_{15}H_{21}NO_4S$ (calculated for $C_{15}H_{22}NO_4S^+$: 312.12695; found: 312.1245) are also labeled by [1-^{13}C]-Cys feeding.

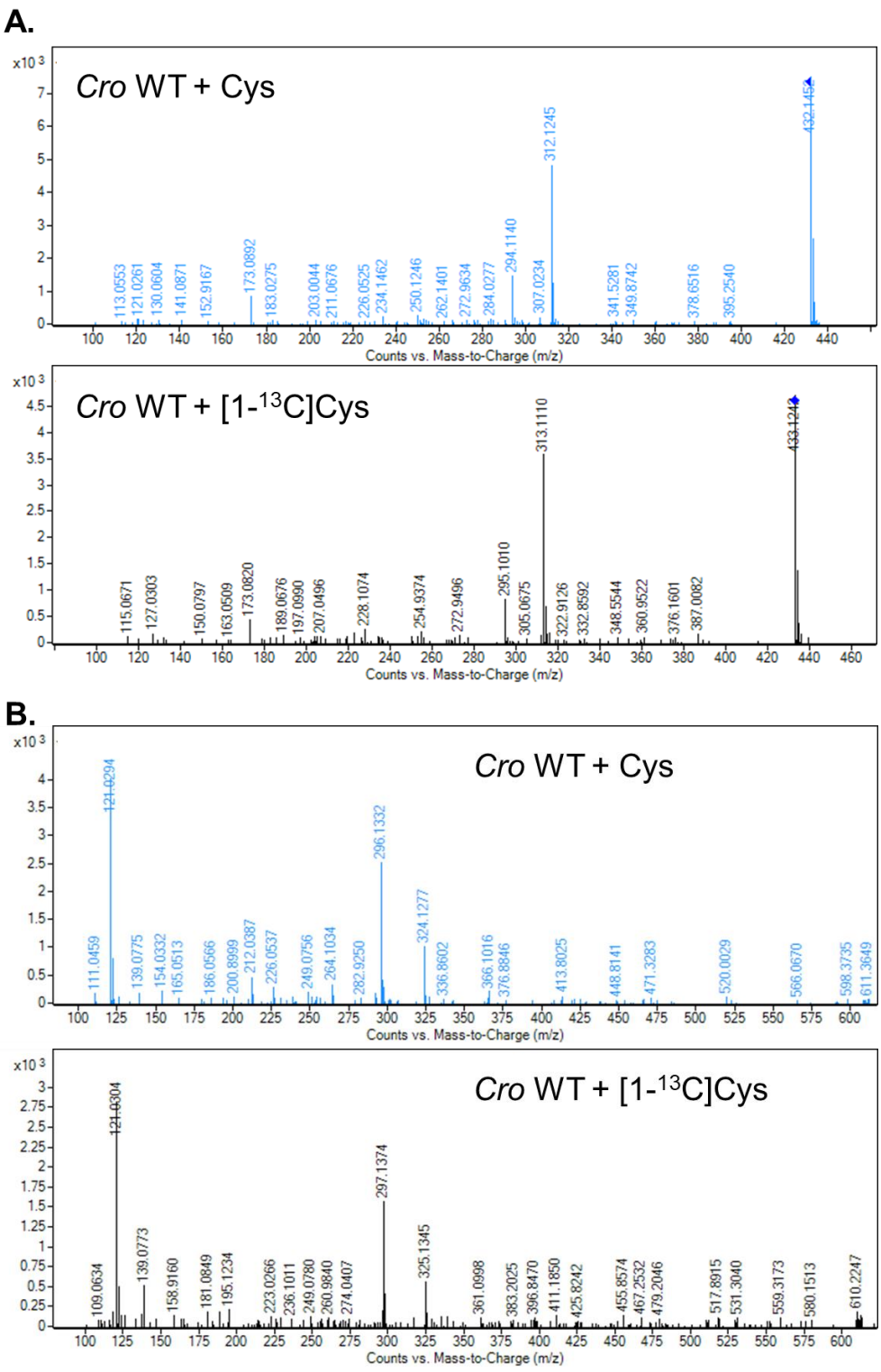
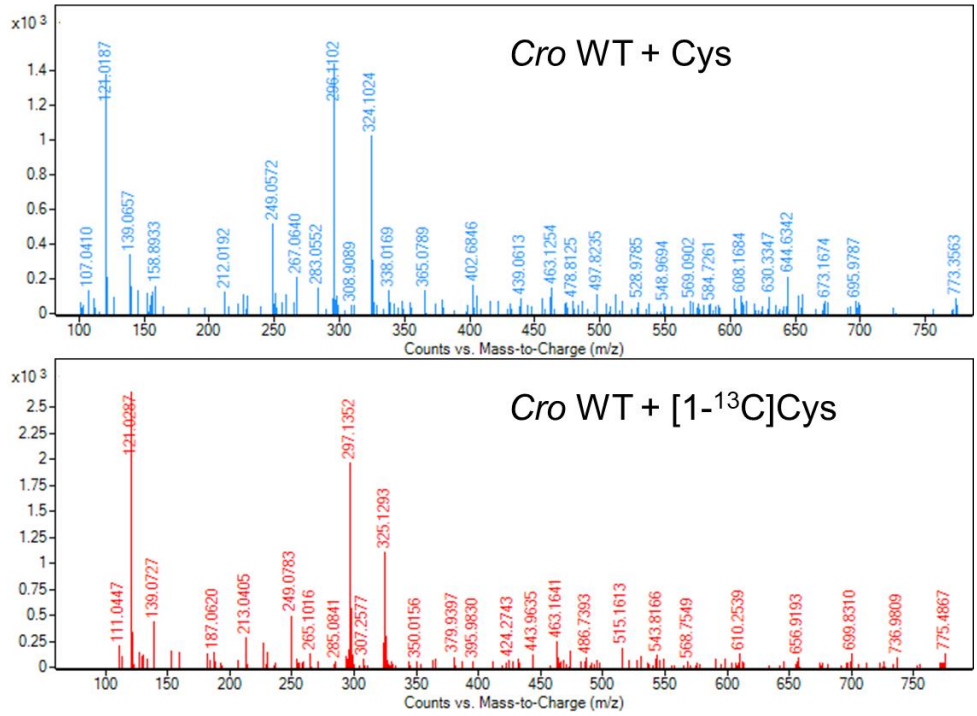


Figure 3-17. Tandem HRMS characterization of BGC 1 associated metabolites. (A) [1-^{13}C]Cys enrichment of BGC1-A fragment ions. (B) [1-^{13}C]Cys enrichment of BGC1-B fragment ions.

Figure 3-17 (continued). (C) [1-^{13}C]Cys enrichment of BGC1-C fragment ions

C.



3.2.5. Discovery and structural elucidation of metabolites associated with BGC 14a

The metabolites associated with the *Cro* BGC 14a knockout strain (*Cro* KO 14a) were identified by a comparative metabolomics approach (**Figure 3-18**). Chemical extracts from *Cro* wild-type and *Cro* ΔBGC14a were analyzed by HPLC-UV/vis to identify peaks unique to the *Cro* wild-type extracts. Several candidate peaks all had related UV/vis spectra with characteristic absorbance maximum in the 400-450 nm range, suggesting they were related congeners. HRMS analysis of the three major peaks (I-III) in positive ion mode enabled prediction of their associated chemical formulas: I, $C_{16}H_{10}O_7$ (calculated for $C_{16}H_{11}O_7^+$: m/z 315.0499; found: m/z 315.0496); II, $C_{16}H_{12}O_7$ (calculated for $C_{16}H_{13}O_7^+$: m/z 317.0656; found: m/z 317.0669); III, $C_{16}H_{12}O_8$ (calculated for $C_{16}H_{13}O_8^+$: m/z 333.0605; found: m/z 333.0612).

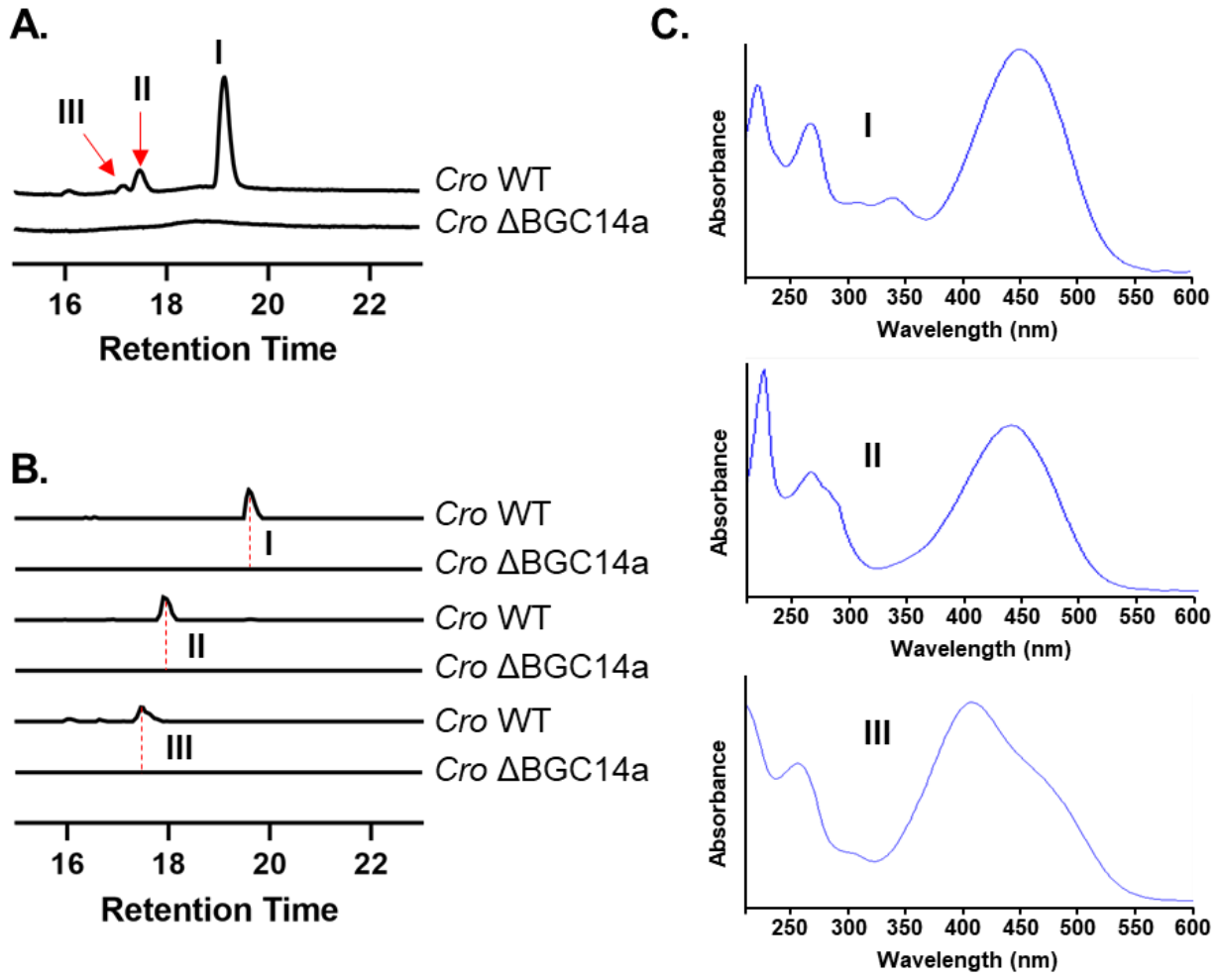


Figure 3-18. Identification of secondary metabolites associated with BGC 14a. (A) UV/vis chromatograms, at 450 nm, comparing *Cro* wild-type and ΔBGC14a. (B) Positive-mode HRMS extracted ion chromatograms. I: 315.0499 m/z; II: 317.0656 m/z; III: 333.0605 m/z. (C) UV/vis spectra of I-III in 30% acetonitrile in water

The major congeners were purified for structural elucidation. A four-liter batch culture of *Cro* was extracted with ethyl acetate and purified by size-exclusion chromatography with a Sephadex LH-20 column packing, followed by one round of reverse-phase HPLC purification. Peak I was isolated as a brownish amorphous solid; Peak II as an orange amorphous solid. Peak I samples demonstrated remarkable thermochromic and solvatochromic properties after isolation (**Figure 3-19**).[148,149] Purification yielded ~1 mg of material corresponding to I and II for NMR characterization.

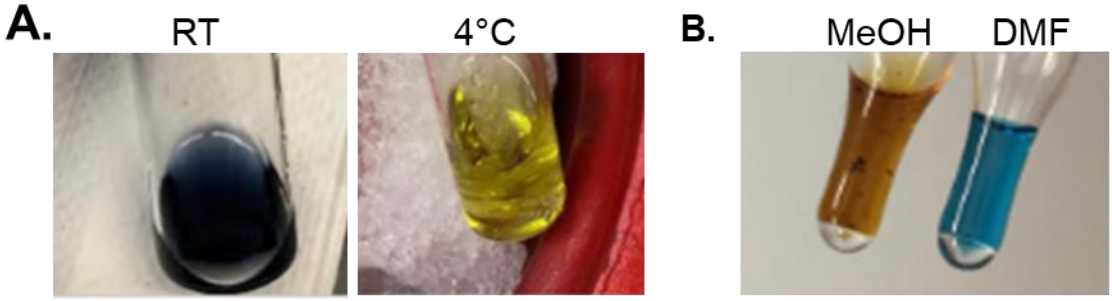


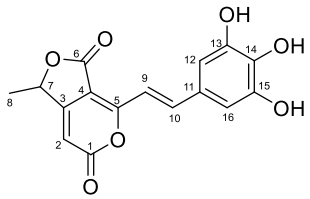
Figure 3-19. Chromism of BGC 14a associated Peak I metabolites. (A) Thermochromism in 30% acetonitrile in water. (B) Solvatochromism in methanol and dimethylformamide

A detailed analysis of the 1D and 2D NMR spectra (^{1}H, ^{13}C, gHSQC, and gHMBC) led to the discovery of four chemical structures (**Table 3-3**). Peak II was discovered to be a mixture of compounds **1** and **2**. Peak I was discovered to be a mixture of two structural isomers, compounds **3** and **4**. These deductions are corroborated by the following evidence, summarized in (**Table 3-3**). The purified Peak II material was confirmed by HRMS to have m/z of 317.0665, consistent with initially predicted chemical formula of $C_{16}H_{12}O_7$. The obtained ^{1}H NMR and HSQC spectra displayed signals for four hydroxyl groups (δ_H 9.34, 13/15-OH of compound **1**; δ_H 9.11, 13-OH of compound **2**; δ_H 9.08, 14-OH of compound **2**), two trans-disubstituted olefins (δ_H 7.62, H-10 of compound **1**; δ_H 7.55, H-10 of compound **1**; δ_H 7.30, H-9 of compound **1**; δ_H 7.27, H-9 of compound **2**), six aromatic or olefinic protons (δ_H 6.73, H-12 and H-16 of compound **1**; δ_H 6.70, H-12 of compound **2**; δ_H 6.58, H-15 of compound **2**; δ_H 6.30, H-2 of compound **1**; δ_H 6.23, H-2 of compound **2**), two oxygen bearing methines (δ_H 5.53, H-7 of compound **1** and **2**), and two methyl groups (δ_H 1.53, H_3-8 of compound **1** and **2**). Furthermore, the ^{13}C NMR and HSQC spectra showed signals for eighteen quaternary sp^2 carbons, ten aromatic or olefinic methines, two oxygen bearing alkyl methines, and two methyl groups. The presence of each dilactone fragment was proposed through their proton and carbon chemical shifts and supported by their ^{3}J-HMBC cross-peaks from H-2 to C-4, H-7 to C-2 and H-8 to C-3, as well as ^{2}J-HMBC from H-2 to C-1 and H-7 to C-3. They were independently connected to a double bond based on ^{3}J-HMBC correlations from H-9 to C-4 and H-10 to C-5. In addition, the other side of each double bond was linked to pyrogallol or hydroxyquinol by the observation of each HMBC correlation. The purified Peak I compounds were confirmed by HRMS to have m/z of 315.0501, consistent with initially predicted chemical formula of $C_{16}H_{10}O_7$. The difference of 2 Da between the molecular weights of **1**/**2** and **3**/**4** suggested that **3**/**4** might be dehydrogenated products of **1**/**2**. Dehydrogenation of the C-7 C-8 double bond in **1**/**2** was confirmed by the deshielding of the chemical shifts of C-7 and C-8 from

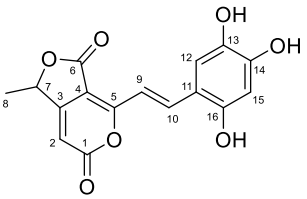
δ_C 76.1 (CH) and 19.3 (CH_3) in **1** and **2** to δ_C 150.7 (C) and 94.8 (CH_2), respectively, in **3** and **4**. The NMR spectroscopic data of **3** and **4** were similar to those of **1** and **2** except as noted above and were fully consistent with its structural assignment as a dehydrogenation of **1** and **2** at the C-7 and C-8 protons, as confirmed by the HMBC cross-peaks. All NMR spectra are available in **Figure B-1** (**Appendix B**).

The structures, named the clostyrylpyrones A-D (compounds **1**-**4** respectively), share a planar scaffold featuring a dilactone with a trihydroxystyrene substituent. There are modest structural similarities to psychoactive plant phenylpropanoid natural products such as the kavalactones,[150] or the *Clostridium*-derived indole alkaloid clostrindolin,[42] although they are biosynthetically unrelated. The solvatochromic properties of **3** and **4** are unusual, as other compounds reported to have this property contain either inorganic components or N atoms in their structures.[148,149]

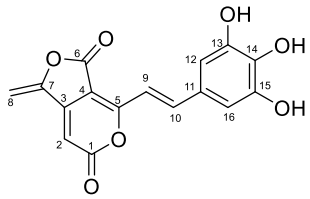
Table 3-3. NMR spectroscopic data for compounds 1, 2 in DMSO-d_6 and compounds 3, 4 in CD_3OD (900 MHz)



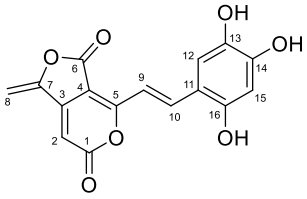
Compound **1**



Compound **2**



Compound **3**



Compound **4**

	1		**2**		**3**		**4**	
position	δ_H (*J* in Hz)	δ_C, type	δ_H (*J* in Hz)	δ_C, type	δ_H (*J* in Hz)	δ_C, type	δ_H (*J* in Hz)	δ_C, type
1		159.7, C		159.9, C		NT		160.8, C
2	6.30 s	101.8, CH	6.23 s	100.7, CH	6.39 s	97.9, CH	6.50 s	99.6, CH
3		163.6, C		163.9, C		NT		NT
4		102.9, C		102.0, C		99.7, C		101.4, C
5		162.7, C		163.2, C		164.8, C		164.0, C
6		166.4, C		166.5, C		NT		NT
7	5.53 q (6.8)	76.1, CH	5.53 q (6.8)	75.9, CH		150.7, C		NT
8	1.53 d (6.8)	19.3, CH_3	1.53 d (6.8)	19.3, CH_3	5.52 d (3.5) 5.36 d (3.5)	94.8, CH_2	5.56 d (3.5) 5.39 d (3.5)	95.2, CH_2
9	7.30 d (15.6)	110.3, CH	7.27 d (15.6)	108.7, CH	7.32 d (15.5)	108.2, CH	7.41 d (15.8)	110.6, CH
10	7.55 d (15.6)	142.3, CH	7.62 d (15.6)	143.9, CH	7.80 d (15.5)	146.4, CH	7.74 d (15.8)	144.2, CH
11		124.8, C		124.4, C		NT		125.3, C
12	6.73 s	108.1, CH	6.70 s	110.7, CH	6.79 s	108.0, CH	6.70 s	111.6, CH
13		146.4, C		141.0, C		143.8, C		NT
14		138.0, C		143.5, C		138.2, C		NT
15		146.4, C	6.58 s	101.4, CH		143.8, C	6.70 s	101.2, CH
16	6.73 s	108.1, CH		152.8, C	6.79 s	108.0, CH		146.2, C
13-OH	9.34 br s		9.11 s					
14-OH	9.08 s							
15-OH	9.34 br s							

3.2.6. Biosynthesis of the clostyrylpyrone scaffold

The domain organization of the BGC 14a thiotemplate assembly line suggests a biosynthetic route for the compound scaffold through the analogous hydroxylated intermediates, **5** and **6** (**Figure 3-20**). These products would have a m/z of 333.0605, consistent with the that of the Peak III minor product identified in association with this BGC (**Figure 3-18b**). The predicted biosynthesis is initiated in module 1 by ATP-dependent activation of 3,4,5-trihydroxybenzoic acid (gallic acid) or 2,4,5-trihydroxybenzoic acid. Modules 2 and 3 make the predicted diketide extensions. Module 4 incorporates an unusual three-carbon extender unit, the first example of such biosynthetic logic from an anaerobic bacterium. The FkbH-like domain[151,152] loads an activated substrate from primary metabolism, D-1,3-bisphosphoglycerate, onto the acyl-carrier protein domain and releases both phosphates, as demonstrated by deuterium-labeled substrate feeding[153] and *in vitro* reconstitution.[154,155] The C domain generates an ester linkage at the 2-hydroxyl, which enables inference of (S) configurations for compounds **1** and **2**. Module 5, found in ORF E, catalyzes a final ketide extension without a detectable *trans*-AT docking domain. Based on the product structures, this module incorporates an unreduced monomer despite the presence of KR a domain. Bioinformatic analysis detected an inactivating point mutation in the KR active site, which yields YxxxG rather than the conserved YxxxN motif. Additional BLAST analyses identified a gene (*pyxF*, NCBI Accession: ASA76633) encoding a PKS module with identical domain architecture, although the KR domain is not inactivated. The *pyxF* module, found in the pyxipyrrolone BGC of the myxobacterium *Pyxidicoccus* sp. MCy9557,[156] has similarities to ORF E. This module is similarly found in a *trans*-AT assembly line, acts downstream of a module containing a condensation domain, and is reportedly functional without a detectable *trans*-AT docking domain (suggesting a limitation of *in silico* prediction).[156] This observation suggests a contextual association between this atypical PKS module and an upstream condensation module. The next module, module 5, lacks an obvious terminal domain as well as the α-pyrone functionality common to all the clostyrylpyrones, which suggests a spontaneous release mechanism.[157] The five-membered lactone is predicted to form by a Michael addition. This forms an alcohol that is eliminated to form part of the pyrone, which is formed during product release.

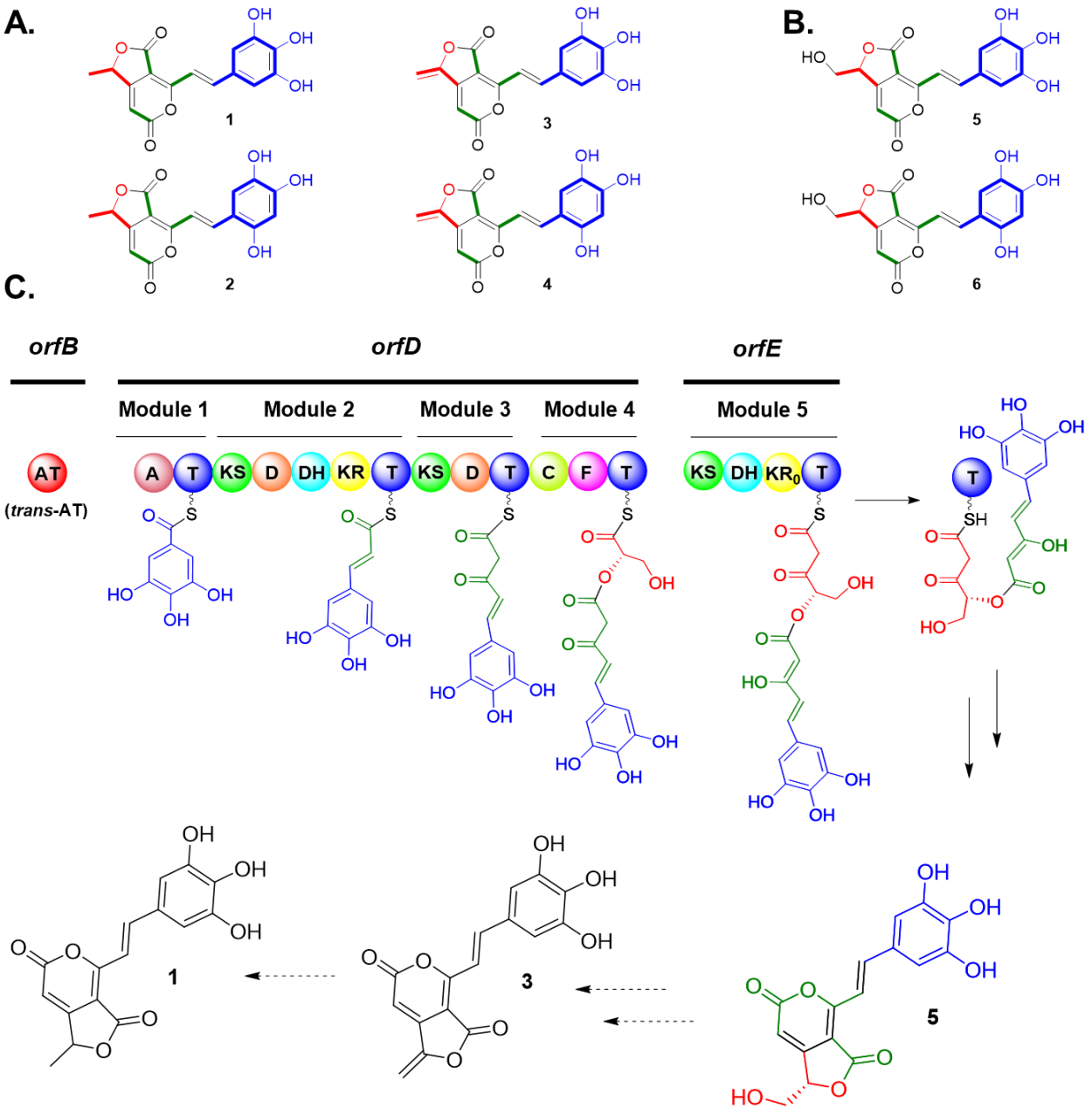
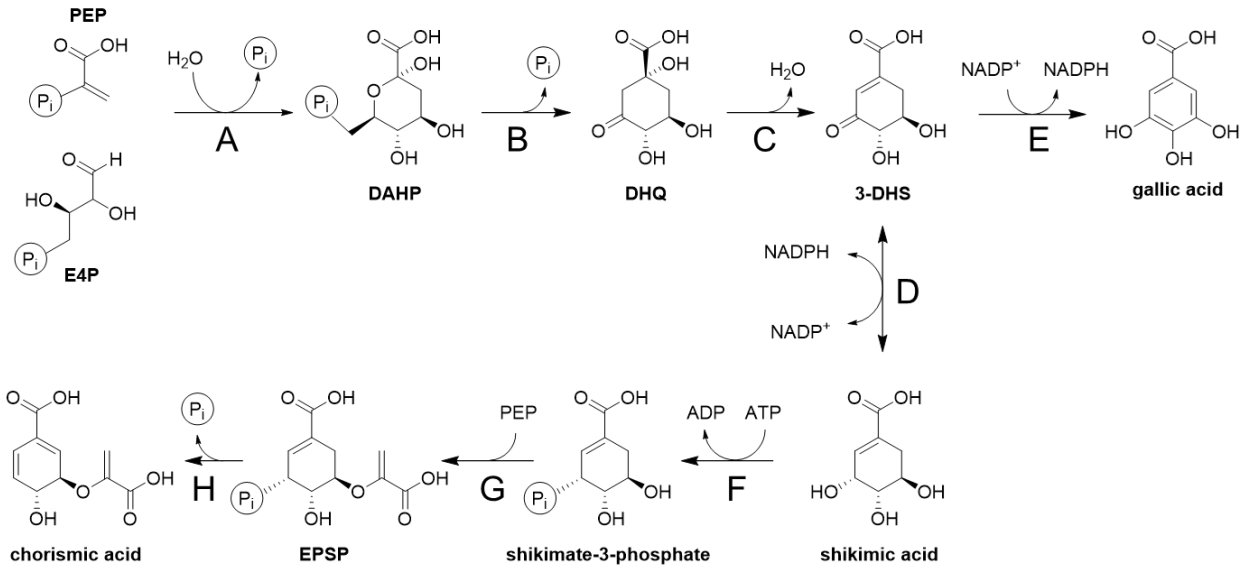


Figure 3-20. Proposed biosynthesis of clostyrylpyrones. (A) Retro-biosynthetic analysis of the clostyrylpyrone scaffold. Monomer predictions: malonyl-CoA in green; D-1,3 bisphosphoglycerate in red; trihydroxybenzoic acids in blue. (B) Predicted hydroxylated intemediates. (C) Thiotemplated assembly line biosynthesis of BGC 14a, using gallic acid as a representative starting monomer. Abbreviations: AT, acyltransferase; A, adenylation; T, thiolation; KS, ketosynthase; D, *trans*-AT docking; DH, dehydratase; KR, ketoreductase; F, FkbH-like; KR_0, inactivated KR

The trihydroxybenzoic acid starting units of this biosynthetic pathway are, to our knowledge, unprecedented in bacterial secondary metabolism. Trihydroxybenzoic acids are usually characteristic of plant tannins and lichen depsides/depsidones.[158] Other aryl acid substrates have been reported in bacterial thiotemplate assembly-line biosynthesis, with well-known examples including 2,3-dihydroxybenzoic acid, 3,4-dihydroxybenzoic acid, salicylic acid, benzoic acid, 3-amino-5-hydroxybenzoic acid, and 4-aminobenzoic acid.[159–162] In this context, both gallic acid and 2,4,5-trihydroxybenzoic acid are unprecedented aryl starter units. The biosynthesis of most of these aromatic substrates stem from the shikimic acid or chorismic acid pathways. We identified genes involved in aryl acid biosynthesis in *Cro* (**Figure 3-21**). Most of the chorismic acid primary metabolism pathway is situated at one locus on contig 6 in *Cro*, as well as in the closely related *Clostridium acetobutylicum* (*Cac*). *Cro* lacks the type I 3-dehydroquinate (DHQ) dehydratase of *Cac* but possesses a type II DHQ dehydratase; thus, it has an intact pathway for aryl compound biogenesis. Additionally, the *Cro* genome contains two apparent duplicate genes which may relate to trihydroxybenzoic acid biosynthesis. One duplicate, CLROS_12250, is BGC 14a ORF A and encodes a putative AroF-like enzyme, the committed step for aromatic compound biosynthesis. The closest known homolog (Accession: RNC29451.1, 63% identity) is found in an uncultured anaerobe, *Ca. Dichloromethanomonas elyunquensis*, suggesting a recent horizontal gene transfer event. The other duplicate, CLROS_3410, encodes an AroE-like dehydrogenase which has been shown to convert 3-dehydroshikimate to gallic acid in *E. coli* and plants.[163] This gene can be distinguished from its primary metabolism homolog (CLROS_4640) by the lack of the N-terminally fused type II chorismate mutase sequence. The biosynthetic origin of 2,4,5-trihydroxybenzoic acid is unclear.

The five-membered lactone of **3** and **4** with the reactive exocyclic olefin has been reported in the biosynthesis of the thiotemplated tetronomycin-like antibiotics.[164,165] In the tetronomycins, the functional group is introduced via D-1,3 bisphosphoglycerate by a module containing the C-F-T domain architecture. The hydroxyl group is acetylated and then eliminated by dedicated enzymes to generate the alkene.[166] However, no homologs of the known Agg4 and Agg5 acetylation and elimination enzymes are detectable in *Cro*, suggesting a different mechanism would responsible for generating **3** and **4** from **5** and **6**. The biosynthesis of these exocyclic olefins is of interest in synthetic biology for programmed thiotemplate assembly of dienophiles capable of *in vivo* Diels-Alder reactions.[165]



Step	Putative Function	*Cac*		*Cro*	
		Homolog	Locus Tag	Homolog	Locus Tag
A	DAHP synthase	*aroF*	CA_RS04760	*aroF*	CLROS_04690
				aroF	CLROS_12250
B	DHQ synthase	*aroB*	CA_RS04770	*aroB*	CLROS_04670
C	type I DHQ dehydratase			*aroD*	CLROS_03400
	type II DHQ dehydratase	*aroQ*	CA_RS04795		
D	shikimate dehydrogenase	*aroE*	CA_RS04785	*aroE*	CLROS_04640
E	3-DHS dehydrogenase			*aroE*	CLROS_03410
F	shikimate kinase	*aroK*	CA_RS04790	*aroK*	CLROS_04630
G	EPSP synthase	*aroA*	CA_RS04775	*aroA*	CLROS_04660
H	chorismate synthase	*aroC*	CA_RS04780	*aroC*	CLROS_04650

Figure 3-21. Aryl acid biosynthesis pathway from *Cro*. Both *Cro* and *Cac* have genes corresponding to a functional chorismic acid metabolic pathway. *Cro* has apparent duplicates of *aroF* and *aroE* which could be involved in biosynthesis of the unusual trihydroxybenzoate starter units of BGC 14a. Abbreviations: DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; DHQ, 3-dehydroquinate; 3-DHS, 3-dehydroshikimate; EPSP, 5-enolpyruvylshikimate-3-phosphate

In order to confirm that BGC 14a is sufficient to produce the predicted hydroxylated clostyrylpyrones, the 22 kb BGC was cloned into a plasmid for heterologous expression in *Clostridium saccharoperbutylacetonicum* N1-4 (*Csa*). Comparative metabolomics experiments led to discovery of a product detected by HRMS in extracts of the heterologous expression strain (**Figure 3-22**). The peak demonstrated identical retention time and mass-to-charge ratio to Peak III of the wild-type *Cro*. Successful production required exogenous supplementation of gallic acid, confirming that the free acid is utilized as the starting unit in thiotemplated biosynthesis. Its production was increased 100-fold in *Csa* cultured at 30°C rather than 37°C (data not shown),

suggesting an improvement in protein solubility at the culture temperature typically used for *Cro*. Notably, the heterologous expression strain did not produce detectable quantities of compounds associated with Peaks I or II, which corroborates our hypothesis that another biosynthetic mechanism is needed for subsequent derivatization of the products **5** and **6**.

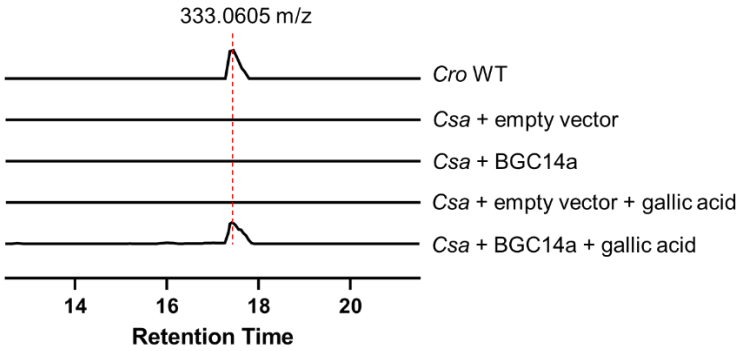


Figure 3-22. Heterologous expression of *Cro* BGC 14a in *Csa*. Positive-mode extracted ion chromatograms demonstrate the gallic acid dependent appearance of the predicted product in *Csa*

3.2.7. Preliminary antioxidant activity of clostyrylpyrones

Highly conjugated phenolic compounds are well known to be powerful antioxidants that can stabilize free-radical chain reactions in foods and other biochemical systems.[167] Thus, we were interested in assessing the function of clostyrylpyrones as antioxidants in the context of *Cro*. We incubated established colonies of *Cro* and *Cro* ΔBGC14a on CBM agar under aerobic conditions to assay for aerotolerance. No growth was observed under these conditions. Next, we tested the ability of pigmented cultures to tolerate hydrogen peroxide stress using a disc diffusion assay. Hydrogen peroxide created an inhibition zone effect in the mutant culture (**Figure 3-23**), although the overall morphology of the lawns was different, suggesting that other phenotypic differences could explain these observations as well, such as higher sporulation of wild-type *Cro* relative to mutant.

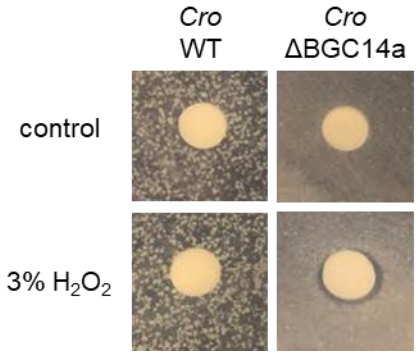


Figure 3-23. Hydrogen peroxide disc diffusion assays. A slight inhibition zone is visible in the *Cro* mutant deficient in clostyrylpyrone production

3.2.8. Preliminary growth rate inhibition assay of clostyrylpyrones

We tested the mixture of clostyrylpyrones A and B (**1** and **2**) for cytotoxicity against MCF-7, a human breast cancer cell line. The MTT-based tetrazolium staining assay[168] was used to quantify cell viability and activity, and the growth rate inhibition (GR)[169] was calculated and plotted (**Figure 3-24**). The GR value ranges from -1 to 0 for cell death, and 0 to 1 for growth rate inhibition. However, all tested concentrations of **1** and **2** yielded values > 1. This can be interpreted as either stimulation of growth rate or a chemical interaction with the tetrazolium dye. Tetrazolium has been used to quantify the reductive potential of antioxidant natural products,[170] but further testing will be required to attribute this observation to either phenomenon.

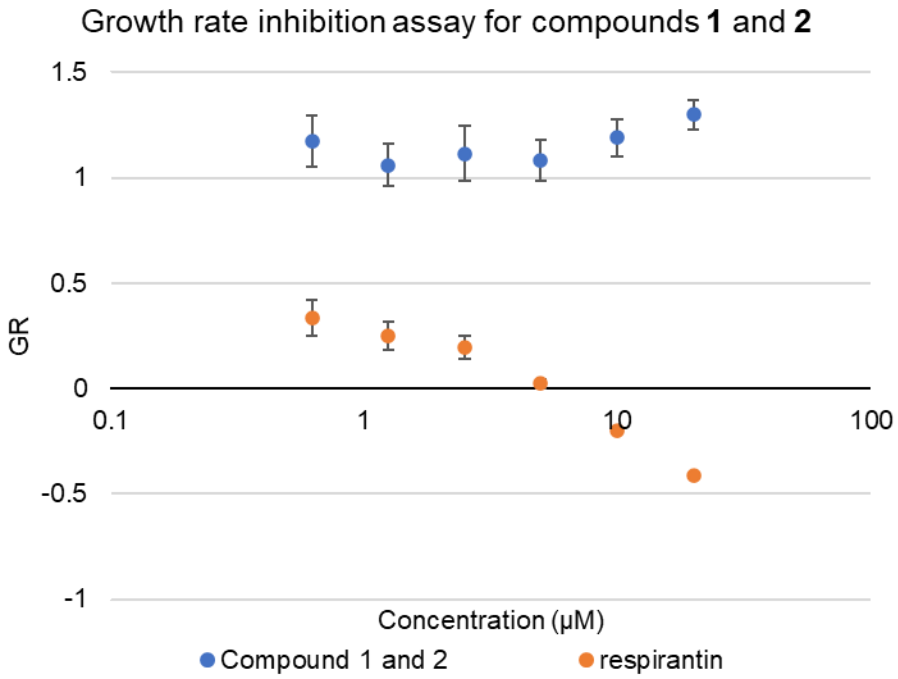


Figure 3-24. Growth rate inhibition assay of compound **1** and **2**. Inhibition is detected in the positive control, performed with respirantin, but not in compounds **1** and **2**.

3.3. Discussion

Anaerobes of the genus *Clostridium* are a promising new source of natural product diversity, crucial for development of bioactive small molecule products. *Cro,* an organism with relevance in industrial biotechnology, has previously been highlighted for its potential to produce thiotemplated secondary metabolites. We identified eight BGCs in *Cro* and profiled transcriptional expression under different growth conditions, showing that seven BGCs are expressed during culture. Further efforts to genetically domesticate *Cro* led to successful gene delivery, clonal selection, and construction of genomic knockout mutants. Comparative metabolomic analyses of wild-type and knockout mutants led to identification of products associated with BGCs 1 and 14a. The BGC 14a metabolites were selected for compound purification, which enabled structural elucidation of a novel family of natural products, the clostyrylpyrones **1**-**4**. Two of the chemical species, **3** and **4**, possess reactive dienophilic olefins characteristic of natural products capable of undergoing Diels-Alder maturation, but their biosynthesis in *Cro* is likely to be through a noncanonical pathway. The biosynthetic pathway of the hydroxylated clostyrylpyrone scaffold **5** and **6** was proposed. The existence of product **5** was supported by heterologous expression of BGC 14a in another *Clostridium* species. Production in the heterologous expression host required exogenous addition of gallic acid, supporting the hypothesis that free trihydroxybenzoic acids, an unprecedented building monomer in thiotemplated natural products, are the starting point of biosynthesis. This study expands the limited repertoire of known anaerobe-derived natural products and describes the first characterized *trans*-AT PKS product from an anaerobic bacterium, demonstrating the promise of this non-traditional source for discovery of new natural products.

3.4. Materials and Methods

Bacterial strains and growth conditions. All strains used in this study are listed in **Table 3-4**. *E. coli* strains were cultured in lysogeny broth (LB) at 37°C and supplemented with antibiotics when appropriate. Cloning was performed in an *E. coli* XL1-blue. *Csa* was maintained in PL7 media (30 g/liter glucose, 5 g/liter yeast extract, 2.67 g/liter ammonium sulfate, 1 g/liter NaCl, 0.75 g/liter monobasic sodium phosphate, 0.75 g/liter dibasic sodium phosphate, 0.5 g/liter cysteine-HCl monohydrate, 0.7 g/liter magnesium sulfate heptahydrate, 20 mg/liter manganese sulfate monohydrate, and 20 mg/liter iron sulfate heptahydrate, with the initial pH adjusted to 6.5 using 1 N HCl). *Cro* was maintained at 30°C in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) containing an atmosphere of 97% nitrogen and 3% hydrogen. For routine culture it was incubated in CBM[171] containing 30 g/liter glucose, 0.5 g/liter monobasic potassium phosphate, 0.5 g/liter dibasic potassium phosphate, 4 g/liter tryptone, 0.2 g/liter magnesium sulfate heptahydrate, 10 mg/liter manganese sulfate heptahydrate, 10 mg/liter ferrous sulfate heptahydrate, 1 mg/liter para-aminobenzoic acid, 1 mg/L thiamine hydrochloride, and 2 µg/liter biotin with the pH adjusted to 6.5 with 1 N HCl. The media was prepared as a filter-sterilized 2x concentrate and mixed with either water or 3% agar for liquid or solid media. For long-term storage, cultures were kept at -80°C in 20% glycerol.

Table 3-4. Strains used in this chapter

Bacterial strains	Relevant characteristics	Source or reference
Escherichia coli		
XL1-blue	Cloning strain	Agilent
WM6026	Conjugation strain, DAP auxotroph	Blodgett *et al*.[140]
Cro		
wild-type	DSM 6424 = NRRL B-575	NRRL
pyrF::ltrB	5-fluorouracil resistance	this study
BGC1*::ltrB*	insertion in ORF F	this study
BGC2*::ltrB*	insertion in ORF B	this study
BGC24*::ltrB*	insertion in ORF L	this study
ΔBGC14a	*catP* insertion in ORF D	this study
Csa ATCC 27021		
pWIS_empty	*eryR*	this study
pJL109	*eryR*, BGC 14a	this study

Transcriptional analyses. *Cro* cultures were inoculated from overnight cultures of either CBM or PL7. After 40 h, cells were subcultured into a 20 ml volume and harvested the next day at OD_{600} ~1 for the PL7 and exponential-phase CBM cultures. The CBM stationary phase cultures harvested after the appearance of thick biofilm and yellow pigments. Cell mass was harvested by centrifugation (4°C, 3,500×g, 15 min) and resuspended in 4.5 ml TRIzol (Ambion, Austin, TX). Samples were stored at -80°C, then thawed and processed according to the manufacturer's instructions. Aqueous phase extracts were precipitated with 1.5 ml isopropanol and centrifuged (4°C, 3,500×g, 15 min). The pellet was washed in 1 ml 70% ethanol and dried under nitrogen. Samples were re-dissolved in 50 µl DEPC-treated water at 60°C for 15 min before treatment with RNase-free DNase I (NEB, Ipswich, MA). Samples were mixed to 800 µl 70% ethanol and transferred to a RNeasy Mini Kit column (Qiagen North America, Germantown, MD) for further purification following the manufacturer's instructions, including the optional on-column DNase I

step. The First Strand cDNA Synthesis kit (NEB) was used to generate reverse transcripts for PCR. Phusion DNA polymerase (NEB, Ipswich, MA) was used for all PCR reactions.

Plasmid construction. All synthetic oligonucleotides (**Table 3-5**) were provided by IDT (Coralville, IA). Phusion polymerase was used for all PCR reactions (**Table 3-6**). FastDigest restriction enzymes were obtained from ThermoFisher. FseI and AscI were obtained from NEB. Plasmid constructs (**Table 3-7**) were assembled by the method of Gibson[111] and reaction mixtures were transformed into chemically competent *E. coli* XL1-blue. Clones were isolated and DNA was extracted using a Zyppy Plasmid Miniprep Kit (Zymo Research, Irvine, CA). Constructs were validated by restriction digest patterning and Sanger sequencing.

Table 3-5. Synthetic oligonucleotides used in this chapter

Primer	Sequence
JL252	AATGGATCCTCACCAGAGAC
JL253	TTTATTACAGCCGCAATTCC
JL254	ATTATATGCGAGAACTTGGC
JL255	AGCCTTATATCCAGGCAC
JL256	ATTATTCGCCATGATATTCC
JL257	TGAAGCTGTTAAGGAAGCAG
JL258	CTGTTATCTCTCCACTTCCG
JL259	TTATGTGTTGGTGGAGAAGG
JL260	TATGGAATTCACGTTGGAG
JL261	TGTAGACGTGCCATGAGC
JL262	TAACAGGTGGGAGTGGAG
JL263	CCCCATCTAATATTATGTTGC
JL264	ATGAATGAATGTCTAACTGTGC
JL265	GCATTACTCTTCAATTTGCC
JL266	TATCCAAGTAGGATGGAAGC
JL267	TCTTCAAATGGATACTCTTGG
JL268	CAGTGCGAAGTGACACC
JL269	GAAGCTTCTGTAGCTCCTCC
JL270	CTTAGCATGAAAGGGATTG
JL271	CACTCCGTCATACCATATCC
JL274	CTCCAAGAAGACCAAATCC
JL275	CAAAATATATGGGGAAATGC
JL276	CACAATTGAAGGTGGAACAC
JL277	CTGCCATAGCATCCTCTATG
JL279	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTCCATTTTAACTTACCTTTCTTTGT
JL281	CGAAATTAGAAACTTGCGTTCAGTAAAC
JL282	AAAAAAGCTTATAATTATCCTTAGTATTCGATGAAGTGCGCCCAGATAGGGTG
JL283	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATGAAGTTAACTTACCTTTCTTTGT
JL284	TGAACGCAAGTTTCTAATTTCGATTAATACTCGATAGAGGAAAGTGTCT
JL285	AAAAAAGCTTATAATTATCCTTACATGCCGTATCAGTGCGCCCAGATAGGGTG
JL286	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTATCACATAACTTACCTTTCTTTGT
JL287	TGAACGCAAGTTTCTAATTTCGGTTGCATGTCGATAGAGGAAAGTGTCT
JL288	AAAAAAGCTTATAATTATCCTTAAAAGCCATTGAAGTGCGCCCAGATAGGGTG
JL289	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATTGAAAGTAACTTACCTTTCTTTGT
JL290	TGAACGCAAGTTTCTAATTTCGGTTGCTTTCCGATAGAGGAAAGTGTCT

Table 3-5 (continued).

Primer	Sequence
JL291	AAAAAAGCTTATAATTATCCTTACCTAACAATGTAGTGCGCCCAGATAGGGTG
JL292	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAATGTAGCTAACTTACCTTTCTTTGT
JL293	TGAACGCAAGTTTCTAATTTCGGTTTTAGGTCGATAGAGGAAAGTGTCT
JL294	AAAAAAGCTTATAATTATCCTTAGGTACCCAGCATGTGCGCCCAGATAGGGTG
JL295	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCAGCATATTAACTTACCTTTCTTTGT
JL296	TGAACGCAAGTTTCTAATTTCGATTGTACCTCGATAGAGGAAAGTGTCT
JL297	AAAAAAGCTTATAATTATCCTTAGCTGGCATATTTGTGCGCCCAGATAGGGTG
JL298	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATATTTGGTAACTTACCTTTCTTTGT
JL299	TGAACGCAAGTTTCTAATTTCGATTCCAGCTCGATAGAGGAAAGTGTCT
JL300	aaacagctatgaccgcggccgctgtatccatatgCATAAGTTTAATTTTTTTGTTAAAAA
JL301	tctcccctaatttttagacttaagggcgggcgcgccTAGATATGACGACAGGAAGAG
JL304	TAAGGATAATTATAAGCTTTTTTAAACAATCTATTTCATAAGTTCC
JL305	GACTTATCTGTTATCACCACATTTGTACAATCTGTAGGAGAACC
JL330	GGTCATGTTTGTCTTGGACTAG
JL331	TTCCCCCCTTAAGCTTCC
JL332	TAAGGATAATTATAAGCTTTTTT AAACAATC
JL355	GAAATAGAAATGTTAAGTGAAAGTGAACG
JL356	TTTGAAATGGGAGGTGTAGC
JL357	GCAATTTTGTGCAATAACGTCTG
JL363	CTCTTCTGGATATTCTGAATCAATAGG
JL364	TTAACGGCCATTCATCTGCC
JL365	GAGTGCTTCAATGTTATTTGCTG
JL387	GAATTCGAGCTCGGTACAGTGGGCAAGTTGAAAAAT
JL388	TTACACTATCAAATAATCTATCTATAATcatCTAAGTTCCCTCTCAAATT
JL393	AAAAAAGCTTATAATTATCCTTACGTTCCGTTTCTGTGCGCCCAGATAGGGTG
JL394	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTTTCTGATAACTTACCTTTCTTTGT
JL395	TGAACGCAAGTTTCTAATTTCGGTTGAACGTCGATAGAGGAAAGTGTCT
JL399	CAGGAAACAGCTATGACCGCGGTAAGACGAACAGCAGAAC
JL400	CTTATTTTTCAATTCCAATTCCATGTACAATTCTTTTCCTCCTCTTACACAC
JL401	GTAAGAGGAGGAAAAGAATTGTACATGGAATTGGAATTGAAAAATAAGTTAC
JL402	TAAAACGACGGCCAGTGCCACTAATTATTCTTTCCATTGTTAACAAC
JL406	GCCAAAATTCGGTCAAATTTTTACATC
JL407	GATAATTCAGCTATGCTTAACATTCC
JL409	ATGTAAAAATTTGACCGAATTTTGG
JL410	GGAATGTTAAGCATAGCTGAATTATC
JL416	TATCGACGGAGCCGATTTTG
N112	GGACGACTTCATTATCTCTG
N113	GCTTCCTTTAACAGAGATAATG
SN1	tatggattataagcggccggccagtgggcaagttgaaaaattc
SN17	actgaaggaggATTAcatatggacaaaaataataataaccgaa
SN18	ctcatatgTAATcctcctttatacaaacatttgttgtagtagtg
SN2	tatcaaaaaggagtttaaacttagggtaacaaaaaacaccg
SN7	gattgttatggattataagcggccggttcatatttat
SN8	ggtcatgagattatcaaaaaggagttttaacttacttattaaat

Table 3-6. PCR reactions used in this work

PCR Reaction	Template	Function	Primer 1	Primer 2
P224	*Clostridium acetobutylicum* genomic DNA	Cloning	JL300	JL304
P226	pACD4 (Sigma TargeTron kit)	Cloning	N112	JL301
P227	pACD4 (Sigma TargeTron kit)	Cloning	JL305	N113
P241	*pyrF*	Screening	JL330	JL331
P243	*pyrF::ltrB*	Screening	JL279	JL331
P246	pJL77	Cloning	JL282	JL281
P247	pJL77	Cloning	JL283	JL284
P248	pJL77	Cloning	JL285	JL281
P249	pJL77	Cloning	JL286	JL287
P250	pJL77	Cloning	JL288	JL281
P251	pJL77	Cloning	JL289	JL290
P252	pJL77	Cloning	JL291	JL281
P253	pJL77	Cloning	JL292	JL293
P254	pJL77	Cloning	JL294	JL281
P255	pJL77	Cloning	JL295	JL296
P256	pJL77	Cloning	JL297	JL281
P257	pJL77	Cloning	JL298	JL299
P259	pJL77	Cloning	N112	JL332
P274	BGC2::*ltrB*	Screening	JL292	JL355
P275	BGC1::*ltrB*	Screening	JL295	JL356
P276	BGC24::*ltrB*	Screening	JL298	JL357
P281	BGC 2	Screening	JL363	JL355
P282	BGC 1	Screening	JL364	JL356
P283	BGC 24	Screening	JL365	JL357
P300	*CatP*	Screening	SN1	SN2
P317	pJL77	Cloning	JL393	JL281
P318	pJL77	Cloning	JL394	JL395
P321	pNK52 for P_{bdh} promoter (**Chapter 2.4**)	Cloning	JL399	JL400
P323	*Cro* genomic DNA	Cloning	JL401	JL406
P325	*Cro* genomic DNA	Cloning	JL407	JL402
P329	*Cro* genomic DNA	Cloning	JL409	JL410
P329	*Cro* BGC 14a	Screening	JL409	JL410
P337	BGC14a*::catP*	Screening	JL401	JL388
P338	BGC14a*::catP*	Screening	JL387	JL402
P354	BGC2::*ltrB*	Screening	JL416	JL363
P355	BGC1::*ltrB*	Screening	JL416	JL364
P356	BGC24::*ltrB*	Screening	JL416	JL365
P372	pMTL82151	Cloning	SN1	SN2
P373	pWIS_empty	Cloning	SN7	SN8

Table 3-7. Plasmid constructs used in this chapter. Listed assembly fragments refer to restriction digests or PCR reactions

Plasmid	Relevant characteristics	Source or assembly fragments
pMTL83353	*E. coli-Clostridium* shuttle vector; pCB102, *specR*, ColE1, *tra*	CHAIN Biotech[139]
pMTL82151	*E. coli-Clostridium* shuttle vector; pBP1, *tmR*, ColE1, *tra*	CHAIN Biotech[139]
pMTL84151	*E. coli-Clostridium* shuttle vector; pCD6, *tmR*, ColE1, *tra*	CHAIN Biotech[139]
pSN1	*E. coli-Clostridium* shuttle vector; pCB102, tmR, ColE1	pMTL83353 digested by FseI and PmeI to yield the 3.8 kb plasmid backbone; P372
pSN4	*E. coli-Clostridium* shuttle vector; pCB102, *eryR*, ColE1	pMTL83353 digested by FseI and PmeI to yield the 3.8 kb plasmid backbone; P373
pJL109	P_{bdh} BGC 14a (heterologous expression construct)	pSN4 digested by NotI and HindIII to yield the 4.4 kb plasmid backbone; P321, P323, P325, P329
pJL77	for ClosTron mutagenesis targeting *pyrF*	pMTL84151 digested by AscI and NdeI to yield the 6 kb plasmid backbone; P224, P226, and P227
pJL85	for ClosTron mutagenesis targeting BGC 9	P227, P259, P246, P247
pJL86	for ClosTron mutagenesis targeting BGC 14b	P227, P259, P248, P249
pJL87	for ClosTron mutagenesis targeting BGC 84	P227, P259, P250, P251
pJL88	for ClosTron mutagenesis targeting BGC 2	P227, P259, P252, P253
pJL89	for ClosTron mutagenesis targeting BGC 1	P227, P259, P254, P255
pJL90	for ClosTron mutagenesis targeting BGC 24	P227, P259, P256, P257
pJL106	for ClosTron mutagenesis targeting BGC 14a	P227, P259, P317, P318

Tested electroporation/transformation methods. Two electroporation protocols previously reported in *Clostridium* species were tested.[112,172] Several strategies were tested for protoplast transformation. Media additives were obtained from Fisher unless otherwise stated. Lysostaphin was obtained from VWR International. *Cro* culture was initiated in T69 and CBM-gly media. T69 media (pH 6.5) contained in g/L: glucose, 10; KH_2PO_4, 0.5; ammonium acetate, 2; $MgSO_4 \cdot 7H_2O$, 0.3; $FeSO_4 \cdot H_2O$, 0.01; Cys•HCl, 0.5; yeast extract, 1; casamino acids, 0.5; tryptone, 0.5. CBM-gly contained 4 g/L glycine. After 1 day (in T69) and 2 days (in CBM-gly) the cultures were and at 2 ml into 90 ml. After 20 h, the following cell wall disruption agents were added for for 1 h incubation at 30°C: ampicillin (20 µg/ml), lysostaphin (10 µg/ml), lysozyme (1 mg/ml), or a mixture of all three.[132,133] The cells were then centrifuged (room temperature, 3,000×g, 1 min), and gently washed twice in 2 ml CBM + 10% PEG 4000 or T69 + 10% PEG 4000 to a final volume of 1 ml in two aliquots each. DNA (1 µg of pWIS_empty) was pipetted into each aliquot. Half of the samples were subjected to heat shock at 55°C in a water bath for 15 min. For recovery, 150 µl aliquots of T69 sample were transferred into 5 ml T69 containing either 4 g/L choline, 1 mg/ml sodium polyanethol sulfonate, 1 mM spermidine. CBM samples were recovered in CBM

supplemented with 300 mM sucrose, 25 mM $MgCl_2$, and 25 mM $CaCl_2$. After overnight recovery,100 µl of recovery cultures were plated on selective and nonselective media.

Conjugation procedure. Chemically competent *E. coli* WM6026[140] was freshly transformed with plasmid to serve as the conjugation donor. From this, overnight cultures of donor were incubated in LB with DAP and appropriate antibiotics at 37°C. *Cro* wild-type overnight culture was prepared in PL7 at 30°C. At ~20 h, *Cro* cultures reached OD_{600} 0.6-1. *E. coli* donors were washed twice in LB and transferred into the anaerobic chamber as a cell pellet. Pellets were resuspended in the PL7 conjugation recipient cultures, to concentrate donor to OD_{600} of 6. Donor/recipient mixtures were plated (100 µl) onto CBM agar and incubated at 30°C for 3-8 h (optimally 4 h) before overlaying with 2.5 mg thiamphenicol. Conjugant colonies appeared in 1-3 days. *Cro* conjugation efficiency was calculated as the ratio of conjugants to recipient colony forming units, tested in at least biological duplicates.

Identification of culture dispersion additives. Liquid cultures of *Cro* were prepared from either a colony or vortex-dispersed liquid culture in CBM, and from liquid culture in PL7. Cultures were supplemented with either a water control, 0.1% v/v tween 80, or 0.001% v/v Triton X-100 and incubated for 2 days at 30°C. Detergent stock solutions prepared at 100× concentration and filter sterilized. To test detergent interaction with established biofilm, a colony-inoculated culture was incubated 5 days in CBM before detergent treatment for 1 h. Cultures were vortexed to suspend them before 10 µl samples were transferred to glass slides (VWR). Phase contrast microscopy images were taken using a Zeiss AxioImager M1 microscope fitted with a Hamamatsu C8484 & Qimaging Micropublisher camera and a Sutter Instruments Lambda LS light source. Relative dispersion was calculated by dividing the number countable single (unattached) cells by the number of total cells. Biological duplicates were used and at least 27 cells were counted per group.

Isolation of knockout mutants. Knockout vectors were introduced into *Cro* by conjugation from *E. coli* WM6026. After 2-3 days, colonies were picked and cultured in 10 ml CBM and 100 µg/ml thiamphenicol. Liquid media was supplemented with 30 µg/ml uracil (for *pyrF::ltrB*) and 0.1% tween 80 (for BGC knockouts). After 24 h, the culture was vortexed to homogenize, passaged at 0.1% into 10 ml subculture, and spread onto CBM agar + thiamphenicol. Agar plates were overlaid with uracil and 500 µg/ml 5-fluorouracil (for *pyrF::ltrB*). After 2-3 days, colonies were screened by touchdown PCR[173] using Phusion polymerase. Cultures were passaged and restreaked until mutants were isolated. After isolation, mutants were cured of plasmid by 1 day culture in non-selective CBM (uracil was supplemented for *pyrF::ltrB*) followed by plating on CBM agar. Colonies were again screened by PCR to confirm plasmid loss.

Metabolomic analysis. Metabolomics samples were analyzed in biological quadruplicate using LC-UV-HRMS. Chemical extracts were prepared by 1:1 extraction of 1 ml culture with 1 ml ethyl acetate. Mixtures were vortexed and spun down (6000×g, 1 min). The upper phase solvent layer was pipetted into a centrifuge tube and dried under N_2, then resuspended in 100 µl methanol. Injections (5 µl) were analyzed on an Agilent Technologies 6545 Accurate-Mass QTOF LC-MS instrument fitted with a 1290 Infinity II DAD for UV/vis and an Agilent Eclipse Plus C18 column (4.6×100 mm). The run method used a linear gradient of 2-98% CH_3CN (v/v) over 40 min in H_2O with 0.1% formic acid (v/v) at a flow rate of 0.5 ml/min. Data analysis was performed in MS-DIAL.[174]

Precursor feeding assay. Cys (Spectrum Chemical, New Brunswick, NJ) and [1-^{13}C]Cys (Cambridge Isotope Laboratories, Tewksbury, MA) were prepared as aqueous 100 mM stock solutions and filter sterilized. *Cro* cultures were prepared in 3 ml duplicate cultures of CBM

supplemented with Cys or [1-^{13}C]Cys to final concentration of 1 mM. After 2 days' incubation at 30°C, 1 ml samples were extracted for metabolomic analyses.

Purification of clostyrylpyrones. *Cro* inoculum was prepared in CBM (20 ml). After 2 days at 30°C, the densely grown, lightly pigmented cultures were aliquoted into two 2 liter bottles of freshly prepared CBM which had been allowed to deoxygenate in the glovebox. Cultures were incubated without agitation for 5 days before harvesting. Culture was extracted twice with 1 volume ethyl acetate and gently centrifuged (2000×g, 10 min) to facilitate phase separation. The upper solvent layer was collected and removed under reduced pressure. The brown oily residue was suspended in 15 ml methanol and partially re-dissolved, leaving behind a red precipitate. The extract was clarified by centrifugation and loaded onto a size-exclusion column packed with Sephadex LH-20 (Sigma-Aldrich) and manually fractionated using MeOH mobile phase. Fractions were screened by LC-UV-MS using an Agilent Technologies 6120 Quadrupole instrument with a 1260 series DAD and Agilent Eclipse Plus C18 column (4.6×100 mm). The run method used a linear gradient of 2-98% CH_3CN (v/v) over 15 min in H_2O with 0.1% formic acid (v/v) at a flow rate of 0.5 ml/min. Fractions containing compounds **1** and **2** or **3** and **4** were consolidated, concentrated under vacuum, and purified using reverse-phase HPLC (using an Agilent 1260 HPLC with DAD) fitted with a semi-preparative Phenomenex Luna C18 column (5 µm, 10×250 mm, 100Å). Compounds **1** and **2** products were resolved with an isocratic elution in 28% CH_3CN (v/v). Compounds **3** and **4** products were resolved with an isocratic elution in 30% CH_3CN (v/v).

Heterologous expression of BGC 14a. The BGC 14a heterologous expression construct (pJL109) and an empty vector pWIS_empty were introduced into *Csa* by electroporation (**Chapter 2.4**). Transformants of the two strains were cultured in 10 ml PL7 + 40 µg/ml erythromycin. Cultures were incubated at 30°C or 37°C, with or without the addition of 1 mM gallic acid (Fisher), added from a 100× stock solution in ethanol. After 3 days, 1 ml samples were collected for metabolomic analysis.

Aerobic growth assay. *Cro* wild-type and ΔBGC14a were spread onto CBM plates to obtain single colonies and cultured at 30°C for 7 days. Plates were removed from the glovebox and further incubated aerobically for 30°C to measure growth activity or other aerobic stress response.[175]

Peroxide disc diffusion assay. Assays were based on a reported method.[176] *Cro* wild-type and ΔBGC14a CBM liquid cultures were inoculated from plates. At 18 h the cultures were turbid and OD_{600} was diluted to 0.2 and 100 µl was plated on CBM to form a lawn. Filter discs of 6 mm diameter (GE Healthcare, Chicago, IL) were placed on top of the agar and dilutions of hydrogen peroxide solution were spotted at 10 µl. After 3 days, plates were examined for inhibition zones.

Growth rate inhibitory assay. Human MCF-7 cell cultures were obtained from the Berkeley Cell Culture Facility. Cell culture and quantification was performed using the MTT Cell Growth Assay Kit (Sigma Aldrich) using the recommended procedure.

Chapter 4. Culture variation methods for cryptic biosynthetic gene cluster activation in clostridia for novel antibiotic discovery

4.1. Introduction

The rising incidence of antibiotic-resistant pathogens is an impending and global problem which affects the treatment options and curability of diseases.[177] One projection estimates that 700,000 annual deaths are attributable per year to antibiotic resistance, and this could grow to 10 million by 2050 if left unchecked.[178,179] New drug discovery efforts are important for maintaining an arsenal of antibiotics for medical practitioners, yet new antibiotic discovery has fallen dramatically. Two important reasons for this decline include the recent industry de-emphasis on natural products in favor of high-throughput synthetic chemistry approaches,[180] as well as compound rediscovery, especially when screening traditional sources such as *Streptomyces* spp.[181] The study of anaerobic bacteria as a new source of natural product scaffolds for antibiotic discovery is an attractive strategy to overcome both of these challenges.[5]

The clostridia represent a relatively untapped reservoir of genomic potential to biosynthesize structurally and functionally diverse natural products.[7] The clostridia with the best potential are non-pathogenic soil isolates,[7] which comprises many of the industrially significant strains such as cellulose degraders and acetone-butanol-ethanol fermenters.[69] Several secondary metabolites recently isolated from Clostridia have demonstrated potent antibiotic activities against a variety of microbial targets (**Figure 4-1**). Prominent examples include compounds discovered from *Clostridium beijerinckii* HKI805 such as clostrindolin, which inhibits growth of *Mycobacterium vaccae*,[42] and clostrocyloin which demonstrated inhibition of *M. vaccae*, *Bacillus subtilis*, and the fungal isolate *Sporobolomyces salmonicolor*.[43] One thioamide-containing natural product, closthioamide, was isolated from *Clostridium cellulolyticum* ATCC 35319 and shown to be a broad-spectrum antibiotic with potent activity against vancomycin-resistant enterococci (VRE) and methicillin-resistant *Staphylococcus aureus* (MRSA),[33] and *Neisseria gonnorheae*.[36] A family of polyketide secondary metabolites, the clostrubins, was found in extracts of both *Clostridium beijerinckii* HKI0724 and *Clostridium puniceum*.[39] These compounds demonstrated activity against a variety of presumed ecological competitors such as *Bacillus pumilus*, *Clavibacter michiganensis* subsp. *sepedonicus*, and *Streptomyces scabies*, as well as against pathogens such as VRE and MRSA.[40] Despite the relative paucity of known anaerobe-derived secondary metabolites,[90] the structural diversity, potency, and spectrum of activity of these recently described compounds supports the idea that clostridial natural products have potential medicinal value.

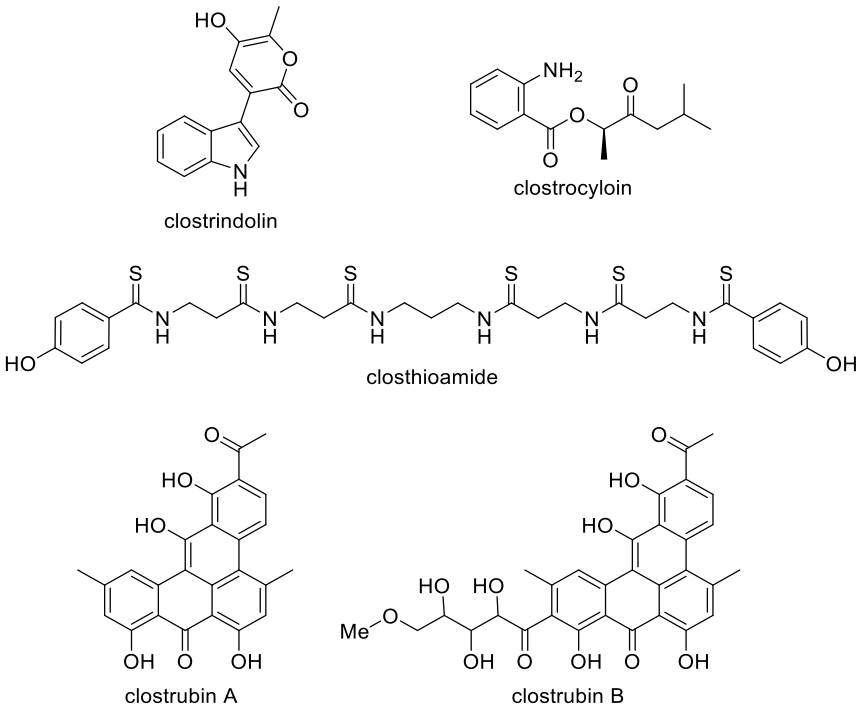


Figure 4-1. Antibiotics isolated from clostridia

Genome mining leverages genomic information to guide compound discovery efforts.[128] Many common tools[91,92] are available for parsing genomes and identifying physically co-localized sets of biosynthetic genes, termed biosynthetic gene clusters (BGCs), which encode production of families of related natural products. Combined with the ever-expanding genomic information available, they reveal that the chemical space that organisms can access is not fully represented in standard laboratory conditions due to the phenomenon of gene cluster silence,[130] even in well-studied model organisms such as *Streptomyces coelicolor* A3(2). Current methods for genome mining can be categorized as pleiotropic or pathway-specific.[130] Pleiotropic approaches induce organism-wide changes to trigger BGC expression and can involve variation of growth conditions, co-culture with environmental competitors, upregulation of global transcriptional regulators, and epigenetic perturbation; they are generally higher throughput but offer lower specificity or predictability. Pathway-specific approaches, on the other hand, are generally lower throughput and involve perturbation of pathway-specific regulators, BGC refactoring, or heterologous expression. Of the pathway-specific strategies, refactoring and heterologous expression have the advantage of being generalizable to all BGCs. *In situ* promoter exchange methods activate BGCs by replacing native promoters with constitutive or inducible ones. CRISPR-Cas9 based platforms can enable robust genetic manipulations and has been applied for this purpose in *Streptomyces*.[182–185] Limitations for these methods include the need for advanced genetic manipulations and the laborious cloning of large DNA constructs, respectively. Methods also exist that combine pleiotropic and genetics-based BGC-targeting to enable powerful and specific BGC activation, namely the reporter-guided mutant selection (RGMS) method[186,187] and the high-throughput elicitor screening (HiTES) method.[188] Both methods involve integration of a reporter gene into an organism under the expression of a native promoter associated with a BGC. In RGMS, the reporter strain is exposed to UV or chemical mutagens to generate genome-scale perturbations, followed

by screening for target-activated mutants. Two reporters are used to reduce background antibiotic resistance resulting from the mutagenesis. In HiTES, the reporter is integrated into the BGC of interest, and then the media is varied by the addition of small molecule inducers in a high-throughput format. The latter method was selected for this work due to its generalizability, pleiotropic activation mechanism yet BGC-specific readout, and simpler requirement for genetic manipulations.

Herein, we describe efforts to survey natural products discovery potential from clostridia. First, we assess clostridial extracts for bioactive components using an antibiotic activity assay. We select promising strains from different sources, culture them in a variety of conditions using the one-strain many compounds (OSMAC) approach,[189] and assay extracts for antibiotic activity. For two genetically tractable isolates, *Clostridium beijerinckii* B-598 (*Cbei*) and *Clostridium saccharoperbutylacetonicum* N1-4 (*Csa*), we design and implement a fluorescent reporter-dependent approach to facilitate high-throughput media elicitor screening to activate and explore secondary metabolism. These studies represent systematic approaches for characterizing secondary metabolism in clostridia at the level of crude cell extract as well as gene transcription.

4.2. Results

4.2.1. Selection of *Clostridium* strains

We initially surveyed available clostridial genomes using AntiSMASH[91] to identify promising isolates (**Table 4-1**) based on their abundance of detectable BGCs, genome availability, genetic tractability, non-pathogenicity, and availability for study. BGC content was tallied based on detectable polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), hybrid PKS-NRPS, and ribosomal peptides and associated post-translational modification (RiPP) biosynthetic elements. Eight sequenced organisms were selected, with an average genome size of 5.3 Mb. This is significantly higher than the 3.3 Mb average for all 83 sequenced genomes from the phylum Firmicutes in a 2013 survey of anaerobic secondary metabolism potential.[7] It has been postulated that larger genomes correlate with more frequent occurrence of large biosynthetic gene clusters.[190] Two additional organisms, *Clostridium felsineum* B-41126 and *Clostridium beijerinckii* B-592, were selected due to hypothesized phylogenetic similarity to their respective type strains, which are sequenced and have large genomes containing numerous BGCs. Two strains, *Clostridium termitidis* CT1112 and *Clostridium cellulovorans* 743B, are valued for their cellulolytic metabolic capabilities. It may be noted that *Clostridium termitidis* CT1112 was recently re-classified as *Ruminiclostridium cellobioparum* subsp. *termitidis* CT1112 on NCBI. *Clostridium kluyveri* is under study for its ability to assimilate byproducts of fermentation such as ethanol and acetate and upgrade to longer-chain products such as caproate.[191] The other species are all capable of fermenting sugars into acetone, butanol, and ethanol.[64]

Table 4-1. Predicted BGCs in the *Clostridium* strains utilized in this study. The number of PKS, NRPS, hybrid PKS/NRPS, and RiPP gene clusters identifiable in each strain is indicated. N/A, not available

Strain	Genome Accession	Genome Size (Mb)	Genetics Reported	PKS	NRPS	Hybrid	RiPP	Origin
C. roseum B-575	LZYU00000000	4.94	No	0	4	4	3	soil
C. felsineum B-41126	N/A	N/A	No	N/A	N/A	N/A	N/A	lake mud
C. kluyveri B-23501	CP000673	3.96	No	0	0	3	1	canal mud
C. beijerinckii B-598	CP011966	6.19	Yes	0	0	3	3	soil
C. saccharobutylicum B-643	CP006721	5.11	Yes	0	4	0	7	soybean
C. acetobutylicum ATCC 824	NC_003030	3.94	Yes	1	0	0	1	soil
C. saccharoperbutyl-acetonicum N1-4(HMT)	NC_020291	6.53	Yes	0	4	3	2	soil
C. beijerinckii B-592	N/A	N/A	Yes	N/A	N/A	N/A	N/A	soil
C. termitidis CT1112	NZ_AORV00000000	6.42	No	0	3	4	2	termite gut
C. cellulovorans 743B	NC_014393	5.26	Yes	0	5	2	11	poplar

4.2.2. Antibiotic activity screening

Different media combinations were tested to generate a variety of cell culture extracts for antibiotic activity assays (**Table 4-2**). Cultures were harvested after 3-4 days. Some groups demonstrated extended lag phases after inoculation and were extracted at 9 days instead. Extracts were tested against a variety of representative target organisms, including Gram-negative bacteria (*Escherichia coli* XL1-blue, *Pseudomonas syringae* pv tomato DC3000), a Gram-positive bacterium (*Amycolatopsis* sp. AA4), a fungus (*Candida albicans* ATCC 96901), and a mycobacterium (*Mycobacterium smegmatis*).

Table 4-2. Growth characteristics of selected Clostridium strains. Abbreviations: Csa, Clostridium saccharoperbutylacetonicum N1-4; Cac, Clostridium acetobutylicum ATCC 824; Cro, Clostridium roseum DSM 6424; Cfe, Clostridium felsineum B-41126; Ckl, Clostridium kluyveri B-23501; Csc, Clostridium saccharobutylicum B-643; Cbe, Clostridium beijerinckii B-592; Cbei, Clostridium beijerinckii B-598; Cce, Clostridium cellulovorans 743B; Cte, Clostridium termitidis; N, no growth; G, growth; nt, not tested; T, turbid culture; S, sedimented culture; B, bubbly culture; W, very slow growth (> 6 days)

Strain	Corn Meal Agar	2xYTG Agar	Potato Dextrose Agar	PLD Agar	Medium C Agar	CBM Agar	Reinforced Clostridial Medium	2xYTX	2xYTG	PL7	DSM 52	ATCC 1191	PYG	CBM-70S	DSMZ 520
Csa	N	G	G	G	G	G	nt	T,B	T,B	T,B	N	T,B	T,B	nt	G
Cac	N	G	N	G	G	G	nt	T,B	T,B	T,B	N	G	G	nt	nt
Cro	N	G	nt	N	N	G	nt	T,B	T,B	T,B	N	W	N	nt	G
Cfe	N	N	N	N	N	N	nt	S	S	N	N	N	T	nt	nt
Ckl	N	G	N	N	G	N	nt	T	T	T	N	W	T	nt	nt
Csc	N	N	N	N	N	S	T,B	N	N	N	N	N	N	N	nt
Cbe	N	G	G	G	G	G	nt	T,B	T,B	S,B	N	T,B	N	N	G
Cbei	N	G	G	G	G	G	nt	T,B	T,B	T,B	N	T,B	T,B	N	G
Cce	N	N	N	N	N	G	nt	W	N	N	N	W	N	T	G
Cte	N	N	N	N	N	G	nt	G	G	T	N	W	N	N	nt

The results of preliminary disc-diffusion assays are presented in **Table 4-3**. None of the extracts inhibited growth of *C. albicans*. Clear associations can be made between certain *Clostridium* strains and target organisms. For example, several *Csa* extracts from different media all inhibited *M. smegmatis*. *Cac* extracts prepared from CBM media, which is known to activate production of its polyketide natural product clostrienose,[53] produce apparent inhibition zones against both Gram-negative and Gram-positive bacteria. Similar inhibition zones were not detected when testing extract from a PKS deletion mutant (Δ*pks*). *Cro* extracts appear to be active against *M. smegmatis* when prepared from CBM and PLD, both media lacking in yeast extract. On the other hand, *Cro* extracts from PL7 culture are associated with inhibition in *P. syringae* and does not affect *M. smegmatis*. This observation could represent differentially expressed metabolism in response to nutritional stimuli. All tested *Cfe* extracts inhibited *M. smegmatis* to varying degrees. *Ckl* extracts showed no evidence of inhibitory activity. *Cbe* extracts from DSMZ 520 media consistently inhibited *P. syringae* in two separate experiments. This effect was lost when the carbon source was switched to cellobiose or xylose. No strong inhibition zones were detected from *Cbei* extracts. *Cce* extracts from CBM inhibited both *P. syringae* and *M. smegmatis*. *Cte* extracts were observed to inhibit *P. syringae*. Notably, all inhibition zones were relatively small (up to 9 mm) relative to the 6 mm diameter of the filter disc itself, which could explain poor reproducibility for some of the replicates shown (for example for *Csa* in 2xYTG). Nonetheless, the detectable inhibition supported the choice of the selected strains as a starting point for antibiotic discovery.

Table 4-3. Disc diffusion assay results for extracts from *Clostridium* cultures. Colors represent assay results: white, not tested; grey, no inhibition (6 mm diameter); yellow, inconclusive; light green, moderate inhibition (7-8 mm diameter); green, strong inhibition (9+ mm diameter) Abbreviations: *Csa*, *C. saccharoperbutylacetonicum* N1-4; *Cac*, *C. acetobutylicum* ATCC 824; *Cac*Δ, *C. acetobutylicum* Δ*pks*; *Cro*, *C. roseum* DSM 6424; *Cfe*, *C. felsineum* B-41126; *Ckl*, *C. kluyveri* B-23501; *Csc*, *C. saccharobutylicum* B-643; *Cbe*, *C. beijerinckii* B-592; *Cbei*, *C. beijerinckii* B-598; *Cce*, *C. cellulovorans* 743B; *Cte*, *C. termitidis*; A, ATCC 1191 medium; AC, ATCC 1191 with cellobiose; AX, ATCC 1191 with xylose; CV, DSMZ 520 derivative; CVC, DSMZ 520 with cellobiose; CVX, DSMZ 520 with xylose; Y, 2xYTG; C, CBM; -Ca, added $CaCO_3$ buffer; PD, potato dextrose; X, 2xYTX; RCM, Reinforced Clostridial Medium

Culture duration	Strain	Medium	*Escherichia coli*	*Pseudomonas syringae*	*Candida albicans*	*Amycolatopsis* sp. AA4	*Mycobacterium smegmatis*
3-4 days	*Csa*	A					
3-4 days	*Csa*	AC					
3-4 days	*Csa*	AX					
3-4 days	*Csa*	CV					
3-4 days	*Csa*	CVC					
3-4 days	*Csa*	CVX					
3-4 days	*Csa*	PLD					
3-4 days	*Csa*	Y					
3-4 days	*Csa*	Y					
3-4 days	*Cac* Δ	C					
3-4 days	*Cac*	**C**					
3-4 days	*Cac*	C					
3-4 days	*Cac*	C-Ca					
3-4 days	*Cac*	P					
3-4 days	*Cac*	PD					
9 days	*Cro*	AC					
9 days	*Cro*	AX					
3-4 days	*Cro*	**C**					
3-4 days	*Cro*	**C**					
3-4 days	*Cro*	CV					
3-4 days	*Cro*	CVC					
3-4 days	*Cro*	CVX					
3-4 days	*Cro*	P					
9 days	*Cro*	PD					
3-4 days	*Cro*	PLD					
3-4 days	*Cfe*	CV					
3-4 days	*Cfe*	**PYG**					
9 days	*Cfe*	**PYG**					
3-4 days	*Cfe*	PYG-Ca					
3-4 days	*Ckl*	P					
9 days	*Ckl*	PD					
3-4 days	*Ckl*	X					
3-4 days	*Cbe*	A					
3-4 days	*Cbe*	A					
3-4 days	*Cbe*	AC					

Table 4-3. (continued).

Culture duration	Strain	Medium	*Escherichia coli*	*Pseudomonas syringae*	*Candida albicans*	*Amycolatopsis* sp. AA4	*Mycobacterium smegmatis*
3-4 days	*Cbe*	AX					
3-4 days	*Cbe*	**CV**					
3-4 days	*Cbe*	**CV**					
3-4 days	*Cbe*	CVC					
3-4 days	*Cbe*	CVX					
3-4 days	*Cbe*	PD					
3-4 days	*Cbe*	PLD					
3-4 days	*Cbe*	X					
3-4 days	*Cbe*	X					
3-4 days	*Cbei*	AC					
3-4 days	*Cbei*	AX					
3-4 days	*Cbei*	C					
3-4 days	*Cbei*	CV					
3-4 days	*Cbei*	CVC					
3-4 days	*Cbei*	CVX					
3-4 days	*Cbei*	PD					
3-4 days	*Cbei*	PLD					
3-4 days	*Cbei*	PLD					
3-4 days	*Cbei*	PLD-Ca					
3-4 days	*Cbei*	RCM					
3-4 days	*Cbei*	X					
3-4 days	*Cce*	AC					
3-4 days	*Cce*	AX					
3-4 days	*Cce*	**C**					
3-4 days	*Cce*	CVC					
3-4 days	*Cce*	CVX					
9 days	*Cce*	PD					
3-4 days	*Cte*	A					
9 days	*Cte*	AC					
9 days	*Cte*	AX					
3-4 days	*Cte*	CV					
3-4 days	*Cte*	CVC					
3-4 days	*Cte*	CVX					
3-4 days	*Cte*	X					

4.2.3. Design of a high-throughput media additive screening strategy for BGC activation in *Csa* and *Cbei*

Several reported genome mining strategies leverage the genetic tractability of the organisms under study.[192] In particular, the HiTES approach has been applied to activate secondary metabolism and discover natural products in aerobic genera such as *Burkholderia* and *Streptomyces*.[188,193] This method utilizes chromosomal integration of a fluorescent reporter gene to generate a facile readout for successful BGC activation in the presence of variable media additives. In order to speed up strain construction, reduce the number of necessary genetic parts, and boost signal by increasing reporter gene copy number, we modified this approach to utilize a plasmid-based reporter for anaerobic *Clostridium* isolates (**Figure 4-2**).

This study utilizes a plasmid based HiTES approach. The workflow begins with identification of a genetically tractable strain and identification of key BGC regulatory sequences, ideally a promoter associated with a well-organized operon of biosynthetic genes. This regulatory sequence can be PCR-amplified and cloned into a modular reporter shuttle plasmid. After introducing the plasmid into the target strain, the promoter of the plasmid will activate the reporter gene in conditions of expression for the target BGC. Each strain can be cultured in 96-well plates with varying chemical elicitors in the wells to induce different metabolic pathways. Then, a fluorescent readout can be used to visualize the gene expression level under different conditions without the need for RNA extraction. Screening hits can be singled out for validation in individual culture and subsequent metabolomic analyses.

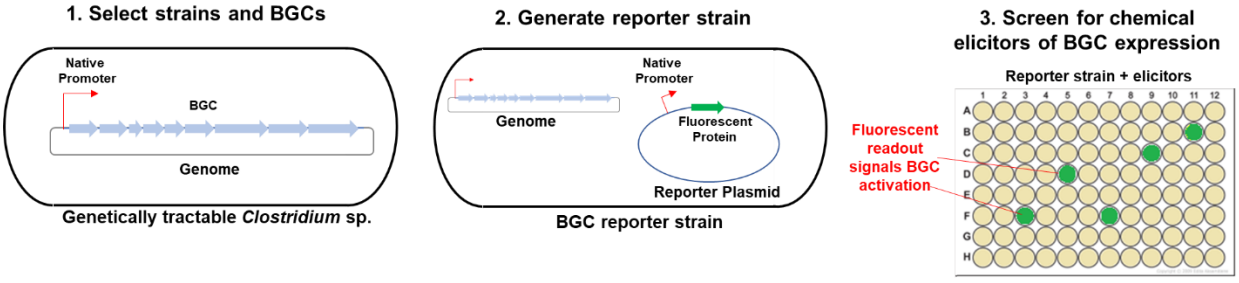


Figure 4-2. Workflow for genetic reporter mediated high-throughput media screening

4.2.3.1. Selection of BGC targets in *Csa* and *Cbei*

Bioinformatic analyses for both *Csa* and *Cbei*, led to selection of fifteen BGC targets. A variety of BGCs were selected, including PKS/NRPS, bacteriocin RiPPs, and one fatty acid BGC in *Cbei*. The upstream sequences (200-300 bp) preceding biosynthetic genes were scanned for promoter sequences. The iPro54-PseKNC webtool[194] was used to identify conserved -12 bp and -24 bp sequences associated with σ^{54} type promoters. BPROM[195] was used to identify conserved -10 bp and -35 bp sequences associated with σ^{70} type promoters. Sequences were selected upstream of obvious gene operons when possible (**Table 4-4**). Several BGCs appeared to possess both σ^{70} and σ^{54} promoters suggesting different patterns of gene expression may be relevant during different cell contexts. Fifteen plasmid constructs were designed and constructed for subsequent elicitor screening.

Table 4-4. Selected BGC promoter sequences for screening. The AntiSMASH index refers to the order of detection and annotation under normal detection strictness. CF denotes BGCs identifiable under loose detection strictness (which uses the ClusterFinder algorithm[196]). Designations of 2a and 2b refer to distinct upstream and downstream BGCs of the same AntiSMASH hit (see Chapter 2 for description of *hyb2* and *hyb3*). Bolded rows correspond to constructs which were later tested during chemical elicitor screening. Promoter predictions with < 95% confidence are indicated. Domain architecture of PKS/NRPS genes is indicated with the following abbreviations: AT, acyltransferase; KS, ketosynthase; DH, dehydratase; T, thiolation; TE, thioesterase; C, condensation; A, adenylation; D, *trans*-AT docking; KR, ketoreductase; CAL, CoA ligase; R, reductive release domain; E, epimerization; PPT, 4-phosphopantetheinyltransferase

Strain	BGC name	AntiSMASH index	Promoter name	Promoter Type	Plasmid Construct	Locus Tag	Putative function
Cbei	**fa1**	CF4	**P70**	**σ^{70}, σ^{54}**	**pJL21**	**X276_21485**	**fabK-like enoyl-ACP reductase**
Cbei	hyb1	2	P71	σ^{54}	pJL22	X276_10360	AT
Cbei	**hyb1**	**2**	**P72**	**σ^{70}**	**pJL23**	**X276_10535**	**KS-DH-T-TE**
Cbei	**hyb2**	**3**	**P73**	**σ^{54}**	**pJL24**	**X276_07695**	**TE**
Cbei	**bac1**	**4**	**P74**	**σ^{70}**	**pJL25**	**X276_05065**	**sactipeptide**
Cbei	hyb3	5	P75	σ^{70}	pJL26	X276_04690	C-A-T-C-A-T-C-A-T
Cbei	hyb3	5	P76	σ^{54}	pJL27	X276_04690	C-A-T-C-A-T-C-A-T
Cbei	**bac2**	**6**	**P77**	**σ^{70}**	**pJL28**	**X276_01900**	**peptidase-containing ABC transporter**
Csa	nrps2	1	P78	σ^{70}	pJL29	CSPA_RS01545	misc nrps mod.
Csa	**hyb2**	**2a**	**P79**	**σ^{70}**	**pJL30**	**CSPA_RS01845**	**CAL-KS-D-T-KS-D-KR-T-KS-D-T**
Csa	**hyb3**	**2b**	**P80**	**σ^{70}**	**pJL31**	**CSPA_RS01885**	**crotonyl-CoA carboxylase/ reductase**
Csa	nrps3	4	P81	σ^{70}	pJL32	CSPA_RS10760	C-A-T-C-A-T-R
Csa	hyb1	5	P82	σ^{54}-like (88%)	pJL33	CSPA_RS14270	hyb1
Csa	nrps4	6	P83	σ^{54}-like (91%)	pJL34	CSPA_RS18035	C-A-T
Csa	nrps1	7	P84	σ^{70}	pJL35	CSPA_RS21535	CAT-E-CAT-TE
Csa	hyb2	2b	P86	σ^{70}, σ^{54}	pJL37	CSPA_RS01930	PPT

4.2.3.2. Selection of media additives for elicitor screening

The selected elicitors, presented in **Table 4-5**, were loosely categorized in terms of their functions representing nutritional stimuli (such as carbon sources, nitrogen sources, vitamins), environmental competition/contamination (such as antibiotics and signaling molecules), variability of mineral content (from salts and chelators), other chemical stressors (such as solvents, detergents, varying pH), and plant-biomass hydrolysis byproducts (various oxidized sugars and aromatic byproducts of lignin or phenylpropanoid degradation)[197] which may be of interest in next-generation bioprocessing feedstocks and are known to impact microbial growth. Additional context is provided to describe either the structural features of the elicitor or predicted interactions with the target bacteria (such as mode of action for antibiotic compounds).

Table 4-5. Selected elicitors for screening. Row and column coordinates denote the arrangement of samples in the 96-well library plate

Row	Col.	Solvent	Reagent	Conc. (mg/ml)	Category	Context
1	A	water	cellulose, micro-crystalline colloidal	120	C source	polymer
2	A	water	lignin	120	C source	polymer
3	A	water	amylopectin, corn	120	C source	polymer
4	A	water	amylose, corn	120	C source	polymer
5	A	water	chitosan, low mw	600	C source	polymer
6	A	water	xylan	60	C source	polymer
7	A	water	D-mannitol	120	C source	sugar derivative
8	A	water	water	N/A	Control	control
1	B	0.3 N NaOH	nalidixic acid	0.001	Antibiotic	DNA replication
2	B	DMSO	trimethoprim	8.7	Antibiotic	folate metabolism
3	B	water	ampicillin	0.005	Antibiotic	cell wall biogenesis
4	B	EtOH	chloramphenicol	0.005	Antibiotic	ribosome
5	B	water	kanamycin	0.01	Antibiotic	ribosome
6	B	0.01 N NaOH	xanthine	0.01	N source	nucleobase
7	B	MeOH	rifampicin	0.001	Antibiotic	RNA polymerase
8	B	water	egg white lysozyme	0.5	Antibiotic	cell wall degradation
1	C	DMSO	amphotericin B	2.5	Antibiotic	membrane disruption
2	C	water	soil extract	N/A	Environmental	mixture
3	C	water	soil extract	N/A	Environmental	mixture
4	C	water	soil extract	N/A	Environmental	mixture
5	C	EtOH	tetracycline	0.005	Antibiotic	ribosome
6	C	water	zeocin	25	Antibiotic	DNA damage
7	C	pure	caproic acid	N/A	C source	fatty acid
8	C	5N NaOH	saccharic acid•K/glucaric acid•K	210.14	C source	sugar derivative
1	D	EtOH	syringaldehyde	36.4	Plant biomass derived	toxic lignin-derived byproducts
2	D	EtOH	vanillic acid	16.8	Plant biomass derived	toxic lignin-derived byproducts
3	D	EtOH	p-coumaric acid	16.4	Plant biomass derived	toxic lignin-derived byproducts
4	D	EtOH	trans-cinnamic acid	14.81	Plant biomass derived	toxic lignin-derived byproducts
5	D	water	phenazine methosulfate	150	Antibiotic	oxidizing agent
6	D	EtOH	benzoic acid	12.2	Plant biomass derived	toxic lignin-derived byproducts
7	D	EtOH	decanoic acid	1	C source	fatty acid
8	D	pure	octanoic acid	N/A	C source	fatty acid
1	E	water	trans-aconitic acid	5	Plant biomass derived	toxic lignin-derived byproducts
2	E	EtOH	2,5 dihydroxybenzoic acid	15.4	Plant biomass derived	toxic lignin-derived byproducts
3	E	EtOH	3,4-dihydroxybenzoic acid/protocatechuate	15.4	Plant biomass derived	toxic lignin-derived byproducts
4	E	water	disodium cytidine monophosphate	91.5	N source	nucleobase
5	E	EtOH	gallic acid	17	Plant biomass derived	aromatic
6	E	EtOH	trans-ferulic acid	19.4	Plant biomass derived	toxic lignin-derived byproducts
7	E	DMSO	2-heptyl-4-quinolone	4.86	Environmental	signaling molecule
8	E	0.01N H_2SO_4 10% EtOH	S-adenosyl methionine (SAM)	19.92	N source	amino acid derivative

Table 4-5. (continued).

Row	Col.	Solvent	Reagent	Conc. (mg/ml)	Category	Context
1	F	EtOH	oleic acid	0.001	C source	fatty acid
2	F	water	pyridoxine HCl	100	Vitamin	aromatic
3	F	water	$K_2TeO_3•H_2O$	1	Inorganic	trace element
4	F	1 N HCl	thymidine	30	N source	nucleoside
5	F	water	D-galacturonic acid	5	C source	sugar derivative
6	F	water	formic acid	1 µL/mL	Plant biomass derived	toxic lignin-derived byproducts
7	F	DMSO	quinazoline	0.01	Antibiotic	structural feature of several drugs
8	F	water	quinic acid	19.2	Plant biomass derived	toxic lignin-derived byproducts
1	G	water	$CuCl_2•2H_2O$	600	Inorganic	heavy metal
2	G	water	$FeCl_3•6H_2O$	135.2	Inorganic	heavy metal
3	G	water	$NiCl_2•6H_2O$	0.01	Inorganic	heavy metal
4	G	water	EDTA	29.2	Chelator	mineral stress
5	G	water	D-gluconic acid•Na,	160	Chelator	mineral stress, sugar derivative
6	G	water	D-gluconic acid•Na,	480	Chelator	mineral stress, sugar derivative
7	G	water	$PbCl_2$	10	Inorganic	heavy metal
8	G	water	citric acid	19.2	Chelator	C source, mineral stress
1	H	water	$CoCl_2•6H_2O$	71.4	Inorganic	heavy metal
2	H	water	betaine•HCl	117	N source	amino acid derivative
3	H	water	VSO_4	14.7	Inorganic	trace mineral
4	H	water	zinc(II) chloride	38	Inorganic	heavy metal
5	H	water	LiCl	0.001	Inorganic	trace mineral
6	H	water	silver sulfate	0.001	Inorganic	heavy metal
7	H	water	$AlK(SO_4)_2•12H_2O$	50	Inorganic	trace mineral
8	H	water	$Na_2WO_4•2H_2O$	4	Inorganic	trace mineral
1	I	water	$Na_2MoO_4•2H_2O$	36	Inorganic	trace mineral
2	I	water	Na_2SeO_3	3	Inorganic	trace mineral
3	I	water	boric acid	6	Inorganic	trace mineral
4	I	water	KI	300	Chemical stress	halide
5	I	water	$CdSO_4•H_2O$	226	Inorganic	heavy metal
6	I	water	$LaCl_3•7H_2O$	245	Inorganic	rare-earth metals
7	I	water	$PrCl_3•H_2O$	0.025	Inorganic	rare-earth metals
8	I	water	$ScCl_3•6H_2O$	30.26	Inorganic	rare-earth metals
1	J	water	NaCl	300	Chemical stress	osmolarity
2	J	pure	ethanolamine	1 µL/mL	N source	amino acid derivative
3	J	water	guanidine HCl	23.9	N source	nucleobases
4	J	water	levulinic acid	101.85 µL/mL	Plant biomass derived	toxic lignin-derived byproducts
5	J	1 N HCl	adenine	33.7825	N source	nucleobases
6	J	water	5 N HCl	N/A	Chemical Stress	acid
7	J	water	hydrogen peroxide	2	Chemical Stress	oxidative shock
8	J	water	5 N NaOH	N/A	Chemical Stress	base
1	K	EtOH	methyl levulinate	1 µL/mL	Plant biomass derived	toxic lignin-derived byproducts
2	K	pure	mineral oil	1 µL/mL	Chemical stress	oil
3	K	EtOH	palmitic acid	200	C source	fatty acid
4	K	water	Triton X-100	1 µL/mL	Chemical Stress	detergent

Table 4-5. (continued).

Row	Col.	Solvent	Reagent	Conc. (mg/ml)	Category	Context
5	K	water	L-Ser	1050	N source	amino acid derivative
6	K	water	glycolic acid	760.5	Plant biomass derived	toxic lignin-derived byproducts
7	K	25% EtOH	R-(-)-leucinol	0.75 µL/mL	N source	amino acid derivative
8	K	water	Tween 20	1 µL/mL	Chemical Stress	detergent
1	L	water	L-(+)-arabinose	4	C source	sugar
2	L	water	beta-lactose	1	C source	sugar
3	L	water	D-galactose	3	C source	sugar
4	L	water	D-(+)-cellobiose	0.25	C source	sugar
5	L	water	D-(+)-xylose	5	C source	sugar
6	L	water	D-(+)-mannose	6	C source	sugar
7	L	water	DMSO	10 µL/mL	Inorganic	membrane stress
8	L	water	N-acetyl-D-glucosamine	111	C source	sugar derivative

4.2.3.3. Selection of a robust gene expression reporter and expression conditions

Several potential fluorescent reporter systems were considered for *Cbei* and *Csa*. The commonly used green fluorescent protein (GFP) and derivatives[198] can be expressed in anaerobic isolates but requires molecular oxygen for spontaneous generation of the imidazolinone-containing[199] fluorophore as a post-translational modification to the protein. This oxygen-based maturation precludes use of GFP in experiments continuously monitoring live cultures of obligate anaerobes. A recently developed alternative to GFP are the cofactor-dependent fluorescent proteins.[200] The best-known example of this anaerobically compatible class of fluorescent proteins is based on engineered LOV (Light Oxygen Voltage) blue light photoreceptors. These fluorescent proteins function by binding to an FMN cofactor, which stabilizes its properties as a cyan-green fluorophore. Proteins with this function are thus described as flavin-binding fluorescent proteins (FbFPs).[200] Other oxygen-agnostic reporters include engineered RNA aptamers that can activate exogenously added chemical fluorophores.[201–203] The disadvantage for these systems is the cost and the dependence on fluorophore feeding as well as the transcription-level rather than the translation-level reporting of gene expression.

For preliminary reporter benchmarking studies, a super-folding green fluorescent protein (sfGFP)[204] and several FbFPs were selected for expression in *Cbei* and *Csa*. The FbFPs included PpFbFPm (the PpFbFP protein derived from the YtvA homolog of *Pseudomonas putida*[205]), CreiLOV (derived from the LOV homolog of *Chlamydomonas reinhardtii*[206]), and tLOV[207] (derived from engineering the iLOV-based reporter from *Arabidopsis thaliana*[208]). Each reporter was cloned onto the pWIS-empty shuttle vector under the expression of a constitutive P_{bdh} promoter from *Csa*. Fluorescence of strains bearing each reporter construct was compared to that of a negative control strain bearing the empty vector. The relative reporter quality of the tested fluorescence genes is shown in **Figure 4-3** for both *Cbei* and *Csa*. The tLOV reporter was the most robust in terms of dynamic range (defined as the difference in fluorescence values between the constitutively expressed reporter and the negative controls). In *Csa*, PpFbFPm was the second-best reporter. Culturing the *Csa*+PpFbFPm-RibM strain in the presence of exogenously fed riboflavin did not improve fluorescence signal, suggesting that either flavin is not limiting in fluorescence yield or RibM is not functional in *Csa*. Thus, we selected tLOV for subsequent screening validation studies.

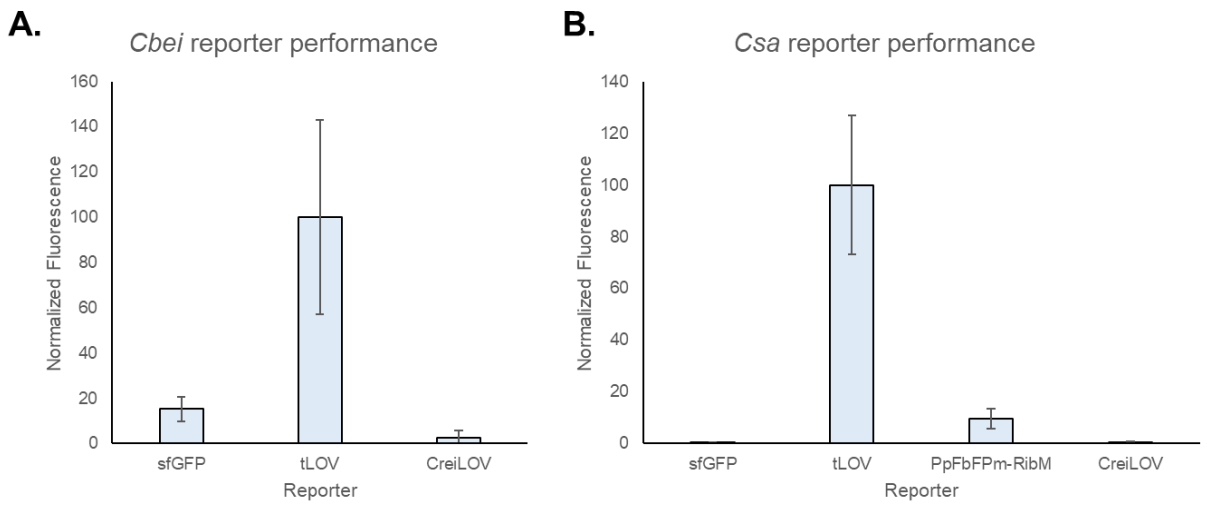


Figure 4-3. Relative performance of fluorescent reporters in *Cbei* and *Csa*. (A) Comparison of sfGFP, tLOV, and CreiLOV in *Cbei*. (B) Comparison of sfGFP, tLOV, PpFbFPm-RibM, and CreiLOV in *Csa*. Dynamic range was calculated as the difference in fluorescence measurement of cultures bearing the indicated reporter gene relative to cultures bearing a negative control. The results are normalized to those of the best reporter, tLOV. Fluorescence for sfGFP was measured at 485/535 nm excitation/emission. Fluorescence for FbFP type reporters was measured at 450/480 nm in *Cbei* and 450/490 nm in *Csa*.

Next, both *Cbei* and *Csa* were assessed for reliability of tLOV reporter performance during screening conditions (in 96-well plates). TYA liquid cultures of *Csa* containing either a negative control or tLOV were harvested at different time points and quantified for fluorescence at 450/490 nm excitation/emission (**Figure 4-4**). Fluorescence was detectable at all tested time points and was much higher in *Csa*, likely due to the use of the P_{bdh} promoter which originates from *Csa*, as well as the relatively higher growth densities of *Csa* cultures. The dynamic range was found to peak between 30 and 46 h in *Csa*, when the cultures had reached stationary phase. The results from *Cbei* cultures showed a decrease between 41 and 46 h, suggesting a peak before 46 h despite the lower growth density at this time. Thus, we selected an incubation time of 38-40 h for subsequent screening experiments.

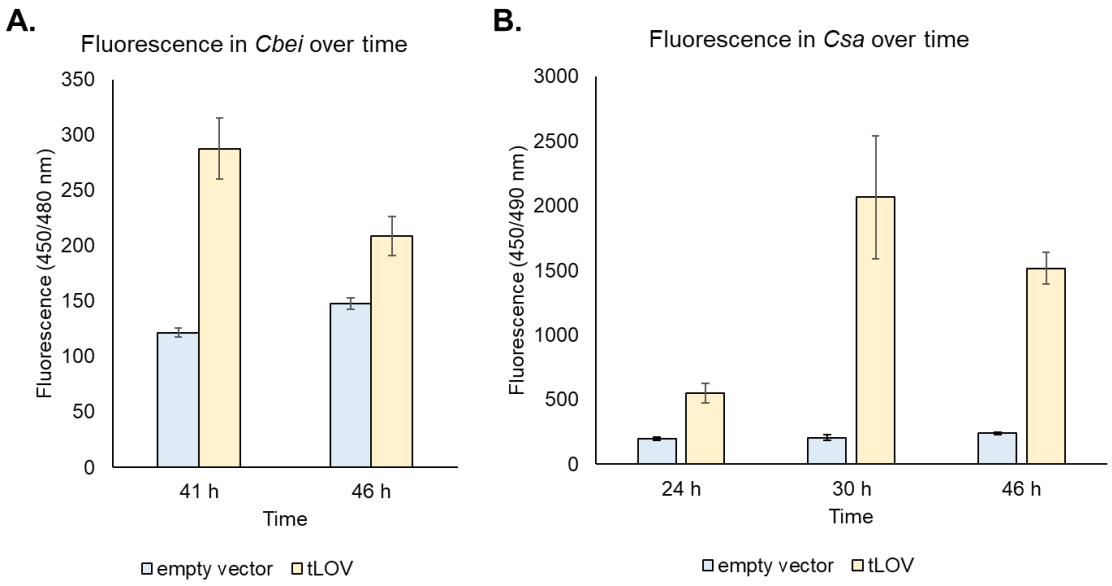


Figure 4-4. Fluorescence of tLOV over time in cultures of *Cbei* or *Csa*. Fluorescence values were obtained from liquid cultures of *Cbei* or *Csa*, containing either tLOV or a negative control, after incubation in 96-well plates.

4.2.3.4. Validating sensitivity for BGC activation

Because *nrps3* is a characterized BGC (**Chapter 2**) with a known product, it was selected for benchmarking to determine if gene expression reporter signal could predict detectable compound production from LC-MS. Liquid cultures of *Csa* with negative control (empty vector), positive control (pJL14, which bears P_{bdh}-tLOV), or pJL32 (P81-tLOV) were prepared in CBM media and quantified in terms of fluorescence at 450/490 nm, OD_{600}, and detectable mass abundance of the 287 m/z ion associated with the *nrps3* BGC. The relationship between fluorescence and growth is plotted in **Figure 4-5a**. Both the positive control and *Csa*+pJL32 demonstrate some monotonic relationship with optical density due to the cell-associated nature of the fluorescent protein. The intrinsic fluorescence of the negative control was fit to a linear function ($R^2 = 0.972$), and this relationship was used to correct for the background fluorescence from all measured fluorescence values. Next, the fluorescence and mass abundance of different samples was plotted (**Figure 4-5b**). Because both the *nrps3* promoter and the P_{bdh} promoter expression are correlated with cell growth, a correlation can also be observed between fluorescence and mass counts as well in both groups. The fluorescence of the P_{bdh} driven tLOV expression can be visualized at much earlier mass count (and cell density) due to the strength of the constitutive promoter, whereas the P81-driven tLOV expression deviates from the baseline only at very high mass signal, suggesting a higher limit of detection due to lower promoter strength. Iterative improvements (**Figure 4-5c**) to pJL32 suggested that elimination of the restriction site (between promoter and reporter) used in modular plasmid construction to create pJL58 could increase relative fluorescence signal, in exchange for a higher cost and a non-modular construction strategy. An insertion of a synthetic ribosomal binding site between promoter and reporter could further boost the signal, at the cost of masking the true efficiency of translation *in vivo*.

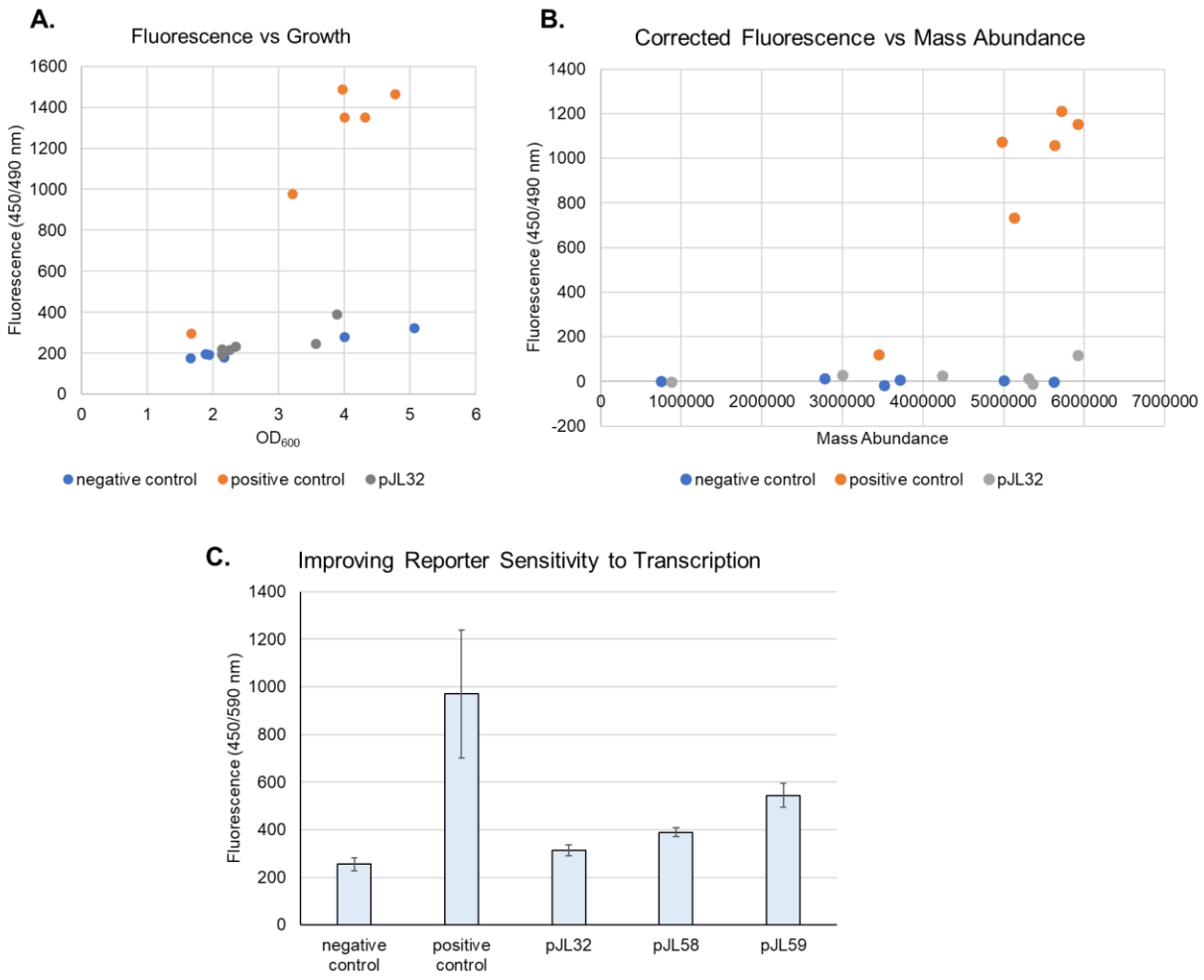


Figure 4-5. Benchmarking of fluorescence against known *nrps3* metabolite. Liquid cultures at different stages of growth were quantified in terms of OD_{600}, fluorescence at 450/490 nm, and mass signal detectable from extracts. Plots represent negative control (*Csa*+empty vector), positive control (*Csa*+pJL14), or *nrps3* screening strain (Csa+pJL32). (A) Fluorescence vs growth density in liquid culture. (B) Corrected fluorescence vs mass abundance demonstrating the relatively high detection limit from pJL32. (C) Fluorescence signal is improved in variants of pJL32

4.2.3.5. Preliminary chemical elicitor screening for BGC activation

Eleven strains harboring the screening constructs indicated in **Table 4-4** were tested for successful chemical elicitor activation in a 96-well plate format. The strains spanned both *Cbei* and *Csa*; five targeted *Cbei* BGCs (pJL21, pJL23, pJL24, pJL25, pJL28), two targeted *Csa* BGCs (pJL30 and pJL31), and a negative and positive control (pWIS-empty, pJL14 respectively) was included for each species. Cultures prepared in liquid CBM were transferred into 96-well plates and treated with the selected chemical elicitors (**Table 4-5**). The growth and fluorescence were measured after 40 h incubation.

Several elicitors resulted in significant growth inhibition in the negative control, including amylopectin (corn), chitosan, most of the antimicrobials (nalidixic acid, ampicillin, chloramphenicol, kanamycin, rifampicin, lysozyme), caproic acid, phenazine methosulfate, formic acid, $CuCl_2$, sodium molybdate, cadmium sulfate, NaCl, ethanolamine, levulinic acid, leucinol, palmitic acid, Triton X-100, and tween 20. These were omitted from subsequent data analysis. The remaining 75 fluorescence reads were blanked against the fluorescence value observed in the appropriate negative control (containing empty vector) and then normalized against the value for the positive control strain (containing pJL14) as an expression benchmark (**Figure 4-6** and **Figure 4-7**). The expression level of the no-elicitor (water) control was also included to represent the baseline expression level of each BGC target.

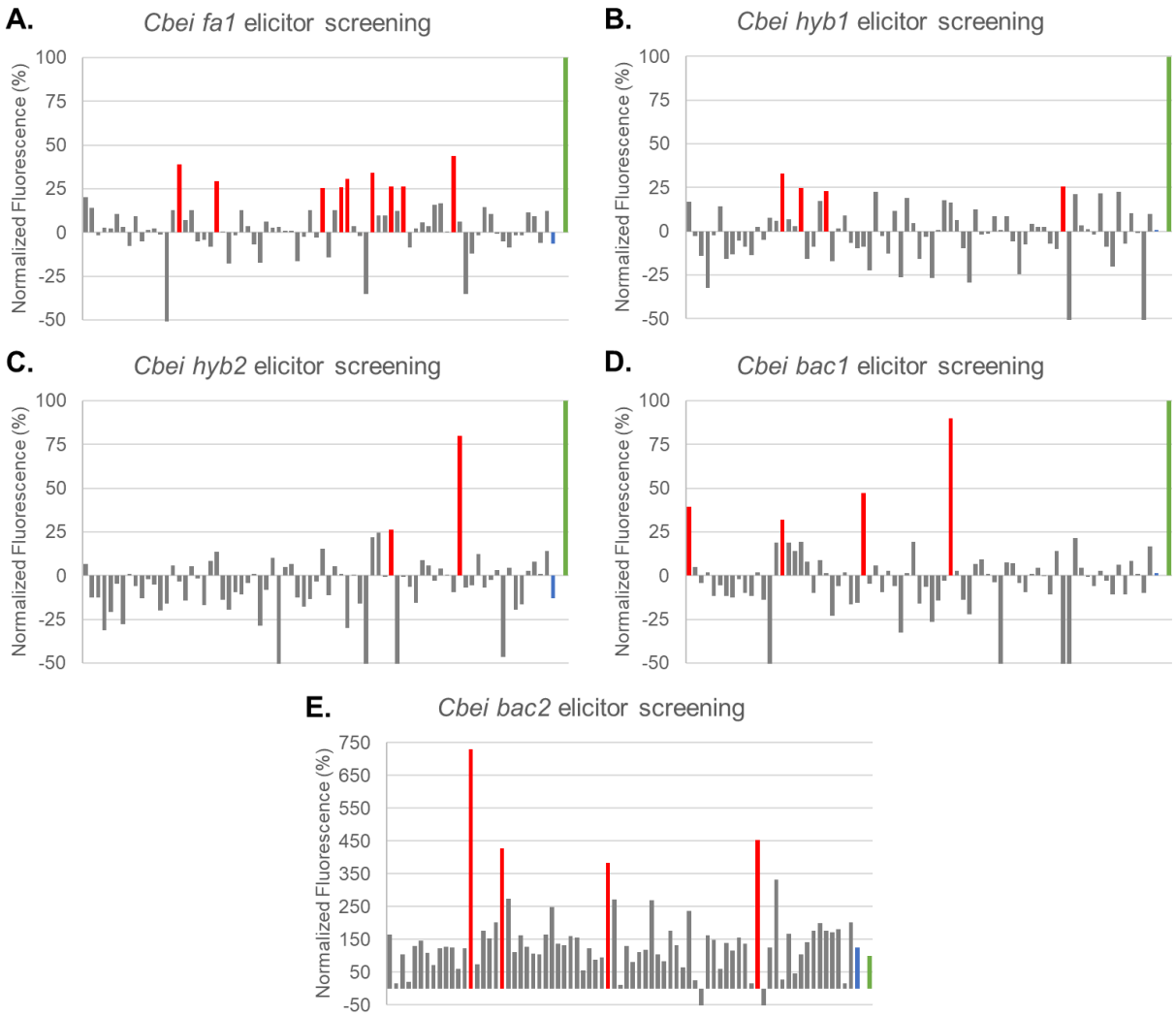


Figure 4-6. Elicitor screening results in *Cbei*. Fluorescence values are blanked against that of a negative control strain (*Cbei*+empty vector) cultured in the same elicitor conditions, and then normalized to that of *Cbei*+pJL14 positive control (green). Expression level of test strain without any elicitor is represented at the right of each graph (blue). Strong elicitors identified from each experiment are highlighted (red). (A) Strong *Cbei fa1* elicitors included aromatics (vanillic acid and 2,5 dihydroxybenzoic acid), D-gluconic acid, $CoCl_2$, LiCl, Na_2WO_4, boric acid, and NaOH. (B) Strong *Cbei hyb1* elicitors included aromatic acids such as vanillic acid, benzoic acid, and 3,4-dihydroxybenzoic acid as well as methyl levulinate. (C) Strong *Cbei hyb2* elicitors included Na_2WO_4 and methyl levulinate. (D) Strong *Cbei bac1* elicitors included cellulose, vanillic acid, oleic acid, and $CoCl_2$. (E) Strong *Cbei bac2* elicitors included saccharic acid, benzoic acid, $FeCl_3$, and NaOH.

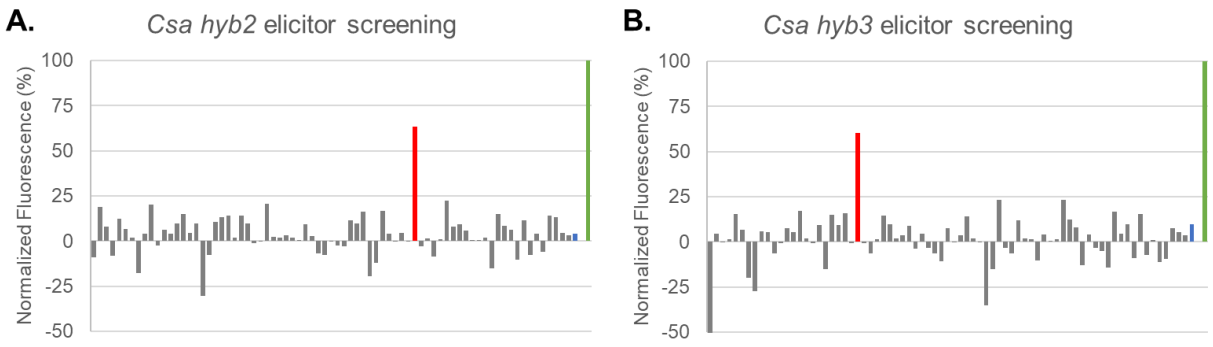


Figure 4-7. Elicitor screening results in *Csa*. Fluorescence values are blanked against that of a negative control strain (*Csa*+empty vector) cultured in the same elicitor conditions, and then normalized to that of *Csa*+pJL14 positive control (green). Expression level of test strain without any elicitor is represented at the right of each graph (blue). The strongest elicitors identified from each experiment are highlighted (red). (A) Na_2SeO_3 was identified as an elicitor of *Csa hyb2*. (B) Cytidine monophosphate was identified as an elicitor of *Csa hyb3*

Many of the conditions yielded negative expression values. These can arise from a variety of factors which result in lower fluorescence from the experimental group than the negative control. One interpretation is that the fluorescent protein tLOV has a fitness effect, which results in synthetic lethal interaction with certain conditions. This is best seen in the very negative (> 50%) values observed in all five *Cbei* strains and the *Csa hyb3* strain, which are all attributable to lack of growth in the experimental strain.

Successfully elicited reporter gene expression was observed in all seven tested strains representing a variety of BGC varieties. In *Cbei*, all five strains had multiple hits with > 25% upregulation. Notably, three had hits with > 50% upregulation relative to the pJL14 positive control. The expression profile of *Cbei bac2* differs from the others because the selected promoter appears to be a constitutively expressed strong promoter, with most tested conditions resulting in > 25% of the pJL14 threshold. Indeed, 56 conditions (including the no-elicitor control) resulted in expression with greater signal than the positive control. Four hits were notably up-regulated over three-fold relative to the no-elicitor control. In *Csa*, meanwhile, each of *hyb2* and *hyb3* had one hit with relatively high (> 50%) expression.

Some overlap was observed in the identity of elicitors which yielded screening hits. For example, vanillic acid was a shared hit in *Cbei* for *fa1*, *hyb1*, and *bac1*; benzoic acid for *hyb1* and *bac2*; $CoCl_2$ for *fa1* and *bac1*; methyl levulinate for *hyb1* and *hyb2*; NaOH for *fa1* and *bac2*; Na_2WO_4 for *fa1* and *hyb2*. Several elicitors were hits unique to one BGC, such as D-gluconic acid (low concentration), LiCl, and boric acid for *Cbei fa1*. At higher concentration of D-gluconic acid, growth was inhibited, resulting in negative expression value; this demonstrates the importance of testing at multiple concentrations. Other unique hits included cellulose and oleic acid for *Cbei bac1*, Na_2SeO_3 for *Csa hyb2*, and cytidine monophosphate for *Csa hyb3*. Notably, several of the hits for *Cbei hyb1* (including 3,4-dihydroxybenzoic acid, benzoic acid, and vanillic acid) have related structures as aromatic acids, suggesting a mechanism of regulation of expression of *hyb1*.

4.3. Discussion

Natural products are crucial for drug discovery efforts—from the 1930s to 2000s, 73% of FDA-approved antimicrobials were from natural products, and 97% of those from microbes.[209] The historical success natural products in antibiotic discovery and development programs, combined with recent trends in natural products research towards non-traditional microbial sources for discovery, has led to the selection of anaerobic bacteria for next-generation natural products discovery. Non-pathogenic clostridia represent one of the most promising anaerobes for this purpose.[33,39]

Ten non-pathogenic clostridia were selected and obtained to assess their potential to produce bioactive natural products. After media compatibilities were established, antibiotic activity assays yielded many potential hits suggesting inhibitory activity against other Gram-negative bacteria, Gram-positive bacteria, and mycobacteria. As predicted, screening hits varied by source strain, medium, and target strain, suggesting that among varying source organisms, varying culture conditions can lead to variable metabolomic states and bioactivities. However, the relatively small sizes of the observed inhibition zones suggest low concentration, low potency, or even low aerobic stability of the hypothesized antibiotics, yielding high noise among the results. Further characterization using traditional fractionation strategies could establish the chemical origin of the different inhibition results observed.

Two promising and genetically tractable strains of clostridia, *Cbei* and *Csa*, were selected for implementation of a plasmid based HiTES method in order to introduce high-throughput screening into this genus for the purpose of activating silent BGCs. Such methods are powerful for accessing a greater portion of the secondary metabolism potential of an organism for new compound discovery.[188] From *Cbei* and Csa, promoters were selected representing a variety of BGC types to target for activation. Elicitors were carefully selected to reflect a variety of potential nutritional, environmental stresses, or industrial bioprocessing contexts. Next, a panel of fluorescent reporters was tested in *Cbei* and *Csa*, and tLOV[207] demonstrated consistently higher performance in both strains. Initial benchmarking studies of tLOV in the regulatory context of the *Csa nrps3* BGC (**Chapter 2**) demonstrated that fluorescence associated with the *nrps3* screening construct could indeed predict compound production, albeit with a high limit of detection. The signal from the reporter construct could be boosted by adjustments to the plasmid design, at the cost of construction scalability and expense.

The results of a preliminary round of chemical elicitor screening was able to identify potential chemical elicitors for all seven of the tested *Cbei* and *Csa* reporter constructs. The nature of the elicitors uncovered will provide a unique opportunity to learn additional biological context of their associated BGCs. In several examples, the identification of the same elicitor for multiple BGCs of *Cbei* suggests those BGCs could be activated through common pleiotropic regulation pathways. For the *Cbei hyb1* BGC, the structural similarity of several identified elicitors suggests the mechanism of regulation is sensitive to aromatic acids.

In summary, this work provides evidence that clostridia can produce antibiotic secondary metabolites, and that a genetic reporter assisted high-throughput elicitor screening approach can facilitate discovery of BGC activation conditions for a more targeted genome mining approach. Further studies will aim to validate the gene activation results and to identify differential metabolism in association with elicitors. If the associated products of the BGCs have demonstrable medicinal value, they could themselves be value-added products of clostridial industrial fermentations. Another direction for the work is to test the other BGCs found in *Cbei* and *Csa* for

a more complete picture of secondary metabolism in these organisms. Finally, the apparent activation of some of the *Cbei* BGCs in response to elicitors such as the benzoic acid derivatives, levulinic acid derivatives, cellulose, and heavy metals may suggest a role of these BGCs in *Cbei* in tolerating or valorizing bioprocessing feedstock contaminants.

4.4. Materials and Methods

Bacterial strains and growth conditions. All strains were obtained commercially from either NRRL or ATCC. All culture work was performed at 37°C in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) containing an atmosphere of 97% nitrogen and 3% hydrogen. Media was equilibrated in the chamber for several days before use to eliminate residual oxygen. DSMZ 52 medium was prepared with the resazurin omitted. A derivative[210] of DSMZ 520 medium was prepared with the following adjustments: $FeSO_4 \cdot 7H_2O$ and $MnSO_4 \cdot H_2O$ was supplemented (20 mg/L each from 1000x stock solution in 0.1 N HCl); glucose was increased to 15 g/liter; where specified, glucose was replaced with cellobiose or xylose; pH was adjusted to 6 before filter sterilization. A derivative[211] of ATCC 1191 was prepared with the following adjustments: resazurin, reducing solution, and $AlK(SO_4)_2$ were omitted; where specified, glucose was replaced with cellobiose or xylose as the carbon source; vitamin content was adjusted to include, per liter, 300 ng biotin, 10 mg PABA, 300 ng folic acid, 10 mg pantothenic acid, 10 mg nicotinic acid, 300 ng vitamin B12, 10 mg thiamine-HCl, 3 mg pyridoxal-HCl. Medium C was prepared as described by Tamburini *et al*.[212]. Reinforced Clostridial Medium was purchased as a dry mix (Oxoid, UK) and prepared as recommended. Potato Dextrose Agar[213] was prepared from a premix (BD, Franklin Lakes, NJ). 2xYTG was prepared with 16 g/liter tryptone, 10 g/liter yeast extract, 5 g/liter NaCl, 10 g/liter glucose and 15 g/liter agar for solid media. 2xYTX is 2xYTG with xylose substituted for the glucose as the carbon source. PL7 contains 30 g/liter glucose, 5 g/liter yeast extract, 2.67 g/liter ammonium sulfate, 1 g/liter NaCl, 0.75 g/liter monobasic sodium phosphate, 0.75 g/liter dibasic sodium phosphate, 0.5 g/liter cysteine-HCl monohydrate, 0.7 g/liter magnesium sulfate heptahydrate, 20 mg/liter manganese sulfate monohydrate, and 20 mg/liter iron sulfate heptahydrate, with the initial pH adjusted to 6.3 using 1 N HCl. PLD medium consists of the following components: 20 g/liter glucose, 15 g/liter agar, 2.0 g/ liter ammonium sulfate, 2.2 g/liter ammonium acetate, 0.5 g/liter monobasic potassium phosphate, 0.5 g/liter dibasic potassium phosphate, 0.2 g/liter magnesium sulfate heptahydrate, 27 mg/liter alanine, 16 mg/liter isoleucine, 23 mg/liter leucine, 13 mg/liter, proline, 19 mg/liter valine, 10 mg/liter manganese sulfate heptahydrate, 10 mg/liter iron sulfate heptahydrate, 10 mg/liter NaCl, 10 mg/liter PABA, 10 mg/liter thiamine, 10 mg/liter pantothenic acid, 10 mg/liter nicotinic acid, 3 mg/liter pyridoxamine-HCl, 0.3 mg/liter biotin, and 0.3 mg/liter folic acid. PYG was prepared with 20 g/liter peptone, 10 g/liter glucose, 10 g/liter yeast extract, 0.4 g/liter sodium bicarbonate, 40 mg/liter dipotassium phosphate, 40 mg/liter monopotassium phosphate, 10 mg/liter hemin, 10 mg/liter vitamin K, 8 mg/liter L-Cys HCl, and 8 mg/liter $MgSO_4$. Clostridial basal medium (CBM)[171] contained 10 g/liter glucose, 0.5 g/liter monobasic potassium phosphate, 0.5 g/liter dibasic potassium phosphate, 4 g/liter tryptone, 0.2 g/liter magnesium sulfate heptahydrate, 10 mg/liter manganese sulfate heptahydrate, 10 mg/liter ferrous sulfate heptahydrate, 1 mg/liter para-aminobenzoic acid, 1 mg/L thiamine hydrochloride, and 2 µg/liter biotin with the pH adjusted to 6.5. For CBM-70S, the glucose was replaced by 70 g/liter sucrose as a carbon source. Where indicated, solid calcium carbonate was added at 6 g/liter for pH control during long incubations.

Preparation of extracts. Cultures were prepared as described above, in 25 ml volumes in 50 ml conical tubes. Each tube was filled to 50 ml with ethyl acetate, vortexed, and spun down. The upper solvent phase was collected into a new tube, dried under nitrogen, and resuspended in 250 µl dimethyl sulfoxide for disc-diffusion assays.

Disc-diffusion assays. Test organisms were prepared in liquid overnight cultures as follows: *E. coli*, *P. syringae*, and *C. albicans*, 37°C in lysogeny broth; *M. smegmatis*, 37°C in Middlebrook 7H9 broth (BD); *Amycolatopsis* sp. AA4, 30°C in ISP2 media (BD). Turbid overnight cultures were diluted 50x in fresh medium, and then 400 µl was plated onto the equivalent agar medium and allowed to dry. Plates were divided into quadrants and sterile 6 mm Whatman filter discs were set with forceps. 20 µl extract was spotted onto each disc and allowed to dry. Plates were incubated 1-2 days to allow lawn formation.

Molecular cloning. All promoter sequences were amplified directly from bacterial colonies suspended in water as PCR template. All restriction digests were performed with FastDigest restriction enzymes (ThermoFisher Scientific, Waltham, MA). PCR templates and restriction digestion products were purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA). Templates for CreiLOV,[206] tLOV,[207] sfGFP,[204] and RibM[214] were obtained as a kind gift from John Dueber (Bioengineering, University of California, Berkeley). Template for the codon-optimized PpFbFP, PpFbFPm, was chemically synthesized (IDT, Newark, NJ). Primer sequences are presented in **Table 4-6**. All reactions (**Table 4-7**) were performed using Phusion polymerase (NEB) under recommended conditions. Negative controls were based on pWIS empty shuttle vectors (**Appendix A**). The pJL14 *E. coli-Clostridium* shuttle vector construct was derived from pNK58 which is based on the pWIS construct and contains a P_{bdh} promoter; it was assembled by the method of Gibson[111] from the tLOV PCR fragment and from the 5.1 kb plasmid backbone of pNK58 obtained by digestion using PmeI and Bsp1407I. Analogous constructs for sfGFP (pJL12) and CreiLOV (pJL15) were assembled using the same strategy. The tested PpFbFPm construct (pJL19) was assembled in operon with an additional fragment, a downstream gene encoding a RibM or PnuC type riboflavin importer (codon-optimized for *E. coli* expression) from *Corynebacterium glutamicum*. Modular screening plasmids (pJL21 to pJL37) were assembled in two pieces from the appropriate promoter and linearized pJL14 featuring sticky ends generated by BcuI and Bsp1407I restriction enzymes, using a T4-ligase based Rapid DNA Ligation Kit (ThermoFisher Scientific). Ligation mixtures were transformed into *E. coli* XL1-blue and colonies were screened by restriction digestion patterning and sequencing the promoter region to verify successful assembly. A derivative of pJL32, pJL58, was constructed by Gibson Assembly using linearized pJL14 and the PCR product P123 (amplified from pJL32). The pJL59 construct was prepared from the 120 bp P124 (amplified from pJL14) and ligated into pJL58. The restriction enzymes used were Kpn2I and XhoI.

Table 4-6. Primers used in this study

Primer	Sequence
JL41	gtaagaggaggaaaagaatgtacaATGCGTAAAGGCGAAGAG
JL42	gcagtttaggcgggtttaaacttaTTTGTACAGTTCATCCATACCATG
JL43	aagaggaggaaaagaatgtacaatGTCCATCGAAAAAAACTTCG
JL44	cgggcagtttaggcgggtttaaactTAATGGTGATGGTGATGGTG
JL45	taagaggaggaaaagaatgtacaATGGCGGGGCTTCG
JL48	cgggcagtttaggcgggtttaaacTCAGACGGTAACGCTTTCTTG
JL49	taagaggaggaaaagaatgtacaatgataaatgcaaaacttcttcag
JL50	GTTCATttcttttcctcctttagttagttaatgttttgcctgaccc
JL52	attaactaactaaaggaggaaaagaaATGAACCCTATCACAGAATTACTG
JL57	taatgtACTAGTTTGTTTATAGAATGCGCAATTC
JL58	ACTACTtgtacaaAAAATCCTCCTTTTTAATTCAC
JL59	taatgtACTAGTAGGAGAACTGAAGAGGTTGTC
JL60	ACTACTtgtacaaAAAAACTATTGGCTTTCTCAT
JL61	taatgtACTAGTATGTGTTTTTTTAGTTAAAATTGC
JL62	ACTACTtgtacaaTAATTTAAGTCACTTTGTTTACCC
JL63	taatgtACTAGTGAAAACAATCAATATTAAGCATG
JL64	ACTACTtgtacaaTTTTTATCGTTACTGAAAATTTA
JL65	taatgtACTAGTGAAATATAAATAAAAAGAGGGCC
JL66	ACTACTtgtacaaTACAAAATCCTCCATTTTAAATATATC
JL67	taatgtACTAGTGACTATTATCATCTGGTTATATTTG
JL68	ACTACTtgtacaaAATATAATCTCCTTATTATTATTTTATTG
JL69	taatgtACTAGTTAAAAATCACCTCTGTATTTATTAA
JL70	taatgtACTAGTCCAGCTTTTTCTAATATAATTATATGG
JL71	ACTACTtgtacaaACTTTCTCCTTCTTCTTTAATAAATATTA
JL72	taatgtACTAGTggaatatggttatatatagg
JL73	ACTACTtgtacaaTTTTATTTCTCCTATTCATTAT
JL74	taatgtACTAGTTTCAAAGAATTATAAAGTGTTTTTG
JL75	ACTACTtgtacaaattctttttatgatactaaatcttatatatc
JL76	taatgtACTAGTaagctgtaataattacagag
JL77	ACTACTtgtacaatcttaaccatactccttaca
JL78	taatgtACTAGTCTTCACAAGAAAATAGTAATTTTAATG
JL79	ACTACTtgtacaaAAAAAAAACACCTCCCAATATAT
JL80	taatgtACTAGTGAATAAATAATAAAGGGCTGAGC
JL81	ACTACTtgtacaaTTCCATGTCCCCCTTTTA
JL82	taatgtACTAGTAAAAGTAAGTTTACATCAGTTAAGGC
JL83	ACTACTtgtacaaattctacatctccttaatcaactcg
JL84	taatgtACTAGTGGTAAGGTTGTAATCACTTTTTAC
JL85	ACTACTtgtacaaATTGTAAGCCCCTTTCATTTAC
JL86	taatgtACTAGTATTTTATGTTTAATTAAAAAATGG
JL87	ACTACTtgtacaaATATACAATTAGTGACTAATTAAACC
JL163	CagtacggtaatgtACTAGTCTTCACAAGAAAATAGTAATTTTAATG
JL164	CGAAGTTTTTTTCGATGGACATtccggaAAAAAAAACACCTCCCAATATAT
JL165	tttccggaaggaggATTAcatatGTCCATCGAAAAAAACTTC
JL166	TAGCTCGAGAAAACCATCTG

Table 4-7. PCR reactions used in this study

Product	Primer 1	Primer 2
sfGFP	JL41	JL42
tLOV	JL43	JL44
CreiLOV	JL45	JL44
PpFbFPm	JL49	JL50
RibM	JL52	JL48
P70	JL57	JL58
P71	JL59	JL60
P72	JL61	JL62
P73	JL63	JL64
P74	JL65	JL66
P75	JL67	JL68
P76	JL69	JL68

Product	Primer 1	Primer 2
P77	JL70	JL71
P78	JL72	JL73
P79	JL74	JL75
P80	JL76	JL77
P81	JL78	JL79
P82	JL80	JL81
P83	JL82	JL83
P84	JL84	JL85
P86	JL86	JL87
P123	JL163	JL164
P124	JL165	JL166

Construction of *Csa* reporter strains. Wild-type *Csa* electrocompetent glycerol stocks prepared as described in **Chapter 2.3**. Cell stocks were thawed on ice and 500 µl aliquots of electrocompetent cells were transferred into 4 mm Bio-Rad (Hercules, CA) cuvettes pre-chilled to 4°C. Each cuvette was supplemented with 2 µg plasmid DNA. Electric pulses were delivered by a Bio-Rad Gene Pulser Xcell with the following parameters: mode, exponential pulse; voltage, 2.0 kV; resistance, 200 Ω; capacitance, 25 µF. Following electroporation (yielding time constants of ~4 ms), cells were immediately resuspended in 10 ml 2x YTG and allowed to recover for 5 h at 37°C. Recovery cultures were plated (at 100 µl volumes) onto 2x YTG plates containing 40 µg/ml erythromycin and incubated at 37°C for 2-3 days.

Construction of *Cbei* reporter strains. Electrocompetent cells were prepared from fresh from liquid overnight cultures in 2x YTG. At OD_{600} of 1.55, they were subcultured 1:10 into 40 ml 2x YTG and further incubated for ~3 h to OD_{600} of 0.6. Next, cells were pelleted by centrifugation (room temperature, 3,500×g, 15 min), washed twice in 10% PEG 8000, pelleted again, and resuspended in 2 ml 10% PEG 8000. Aliquots of 500 µl were transferred into 4 mm cuvettes pre-chilled to 4°C. Each cuvette was supplemented with 2 µg plasmid DNA (this was prepared in Dam/Dcm background although Kolek *et al*.[172] noted that this greatly decreases transformation efficiency) and pulsed in a Bio-Rad Gene Pulser Xcell with the following parameters: mode, exponential pulse; voltage, 1 kV; resistance, 100 Ω; capacitance, 50 µF. Each cuvette was recovered in 10 ml 2x YTG at 37°C for 6 h. Recovery cultures were then spun down, resuspended in 1 ml 2x YTG, and plated at 100 µl volume on 2x YTG plates containing 40 µg/ml erythromycin.

Preparation of elicitor library. All chemicals were obtained from Sigma Aldrich (St. Louis, MO) unless otherwise specified. Ampicillin, D-cycloserine, formic acid, D-mannitol, $AlK(SO_4)_2 \cdot 12H_2O$, $Na_2WO_4 \cdot 2H_2O$, $PrCl_3 \cdot H_2O$, $LaCl_3 \cdot 7H_2O$, $ScCl_3 \cdot 6H_2O$, ethanolamine, oleic acid, mineral oil, palmitic acid, glycolic acid, hydrogen peroxide, DMSO, D-(+)-xylose, and L-(+)-arabinose were obtained from ThermoFisher. Kanamycin was obtained from VWR International (Radnor, PA). Zeocin and amphotericin B were obtained from Neta Scientific (Hainesport, NJ). S-adenosylmethionine was obtained from NEB (Ipswich, MA). Tween 20 was obtained from Bio-Rad Laboratories. Quinazoline was obtained from Grainger (San Leandro, CA). Soil extracts C2-C4 were prepared using local soil from Berkeley, CA by steeping in water overnight and filter-sterilizing using a PES bottle-top filter (Corning, Corning, NY). In order to prepare the library, Row A chemicals were aliquoted into a 2 ml square well 96-well plates (E&K Scientific, Santa Clara, CA) and sterilized by autoclaving. Row A chemicals were then suspended in sterile water as colloids. Stock solutions were prepared using the solvents of **Table 4-5** and

sterilized using 0.2 µm PVDF filters (Restek, Bellefonte, PA). Each well constituted a 1000x concentrate. Rows L and K were 100x concentrates.

Assaying reporter fluorescence in liquid cultures. *Cbei* and *Csa* strains containing empty vectors or fluorescent reporter genes were cultured in TYA medium[215] containing the following components, in g/L: tryptone, 6; yeast extract, 2; glucose, 20; ammonium acetate, 3; magnesium sulfate heptahydrate, 0.3; potassium phosphate monobasic, 0.5; ferrous sulfate heptahydrate, 0.01. The pH was adjusted to 6.5 using HCl before filter-sterilization. Each stationary phase overnight culture was diluted at 100 µl into 40 ml and then transferred in 2 ml aliquots into a row of a 96-well plate. After 30 h incubation at 37°C, the plate was centrifuged (room temperature, 3,500×g, 15 min), decanted and tapped dry on a paper towel. The pellets were resuspended in 250 µl TYA and 200 µl was transferred to a clear-bottomed black 96-well plate for quantification in a SpectraMax M2 instrument (Molecular Devices, San Jose, CA). During measurement, the OD_{600} of *Csa* was ~1.8±0.1 and the OD_{600} of *Cbei* was ~0.7±0.1.

Reporter validation studies using the nrps3 BGC. Cultures of Csa were prepared in 10 ml CBM supplemented with erythromycin (40 µg/ml). After two days, the OD_{600} was measured from 200 µl samples after ten-fold dilution. In addition, 5 ml samples were centrifuged and concentrated 20-fold in fresh medium. Fluorescence was read from 200 µl samples at 450/490 nm. To quantify mass signal of nrps3 metabolite, 3 ml samples were extracted 1:1 in ethyl acetate, dried under nitrogen, resuspended in 150 µl methanol. Twenty µl extract was analyzed using an Agilent 6120 single quadrupole LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm) and a linear gradient of 2-98% acetonitrile in water over 40 minutes and a flow rate of 0.5 ml/min. The mobile phase was supplemented with 0.1% formic acid. For pJL58 and pJL59 study, fluorescence measurements were made using a similar method, but all cultures were adjusted to OD_{600} of ~2.4±0.1 before concentration.

Elicitor screening. Screening strains were all plated onto TYA + erythromycin (40 µg/ml) from glycerol stocks. After 5 days, whole colonies were used to start overnight cultures in CBM + erythromycin. After 18 h, cultures had reached OD_{600} of ~0.9 (*Cbei*) and ~3 (*Csa*). Each seed culture was diluted into 100 ml of the same medium to an OD_{600} of 0.01. and aliquoted in 1 ml volumes into the 96 well plate. Elicitor was applied aseptically from a prearranged plate of elicitors in either 2 µl (rows A-J) or 20 µl volumes (rows K-L) and mixed in. After 40 h incubation, the plates were removed from the anaerobic chamber, centrifuged (room temperature, 3,500×g, 15 min), resuspended to 250 µl (four-fold concentration), and 200 µl sample was used for quantification as described above.

Chapter 5. Conclusions

Natural products are a critically important resource for the development of pharmaceuticals and other bioactive small molecules. The search for new natural product sources has led to the identification of the anaerobes, especially clostridia, as a rich and starkly underutilized reservoir of natural product diversity. The ABE fermenting soil isolates of clostridia appear to be particularly promising sources of polyketide and nonribosomal peptide natural products. This work expands the known chemical space of anaerobe derived natural products, with the discovery of five novel chemical structures spanning two novel natural product families (**Chapters 2** and **3**). It also describes the challenges of implementing advanced genome mining strategies to enable targeted and high-throughput screening for BGC activation to elicit natural products from *Clostridium* spp. (**Chapter 4**). As genomic data become increasingly available, and more advanced genetic tools are developed for these organisms, these methods will enable a powerful and generalizable method for gene-cluster targeted natural product discovery from clostridia. Natural products also mediate biological processes in an inter- and intraspecies fashion. In the context of ABE fermenting clostridia, these compounds may effect biological functions that can shape metabolic engineering strategies or potentially be value-added products during industrial fermentation.

The novel compounds reported in this work demonstrated diverse new structures and functions. **Chapter 2** describes an *N*-acyl dipeptidyl alcohol whose BGC has homologs in other ABE fermenting clostridia, prompting an investigation of its role in fermentation related processes. In *C. saccharoperbutylacetonicum*, *e*xpression of the NRPS gene was associated with butanol tolerance, providing the second reported example of secondary metabolism mediated impact on ABE biology. **Chapter 3** describes characterization of secondary metabolism in *C. roseum*, which led to discovery of the novel clostyrylpyrone family of natural products. These unique compounds, the first reported *trans*-AT PKS derived natural products from an anaerobe, are constructed from unusual trihydroxybenzoic acid monomers. Additionally, the chemical species **3** and **4** possess a likely reactive exocyclic olefin which appears to be produced by a noncanonical pathway. Heterologous expression results confirmed that the genes responsible for maturation of intermediates **5** and **6** into compounds **1**-**4** are found outside of *C. roseum* BGC 14a. The congeners **3** and **4** further demonstrate remarkable chromism. Preliminary data for the other clostyrylpyrones **1** and **2** suggest possible antioxidant activity of these natural products; the consequences of this on ABE fermentation remain to be determined. In **Chapter 4**, a broader profiling of extracts from nine clostridia yielded detectable but weak antibiotic activity from many of the species, although further work will be necessary to ascribe this activity to the molecules which mediate this activity.

The successful discovery of both the *N*-acyl dipeptidyl alcohol metabolite from *C. saccharoperbutylacetonicum* and the clostyrylpyrones from *C. roseum* was accomplished by application of a reliable BGC profiling and genome mining approach. The workflow of discovery was dependent on genomic analyses to identify organisms with high secondary metabolism potential, followed by transcriptional profiling by reverse transcriptase PCR or RNA sequencing. Expressed BGCs were candidates for reverse genetics characterization to enable identification of their associated natural products by comparative metabolomics. Here, genetic tools were developed as necessary to facilitate gene delivery, genomic manipulation, and screening. New compounds were purified from bulk cultures for structural elucidation by NMR. This enabled retrobiosynthetic analysis and proposal of the biosynthetic pathway of the compounds, as well as efforts for functional characterization of the compounds. Heterologous expression in either *E. coli*

or a more closely related *Clostridium* species provided additional evidence to confirm the proposed biosynthetic pathways. For *nrps3*, it also allowed probing of substrate specificity and *in vitro* reconstitution for more rigorous characterizations (**Chapter 2**). Other methods such as isotopically labeled substrate feeding, commonly used for metabolic flux measurements in ABE fermenters, are powerful tools to probe biosynthesis, as demonstrated in **Chapter 3** for the products of BGC 1. Overall, workflow was applied to gain in-depth characterizations of secondary metabolism in two promising species of ABE fermenting clostridia for natural product discovery.

From the next-generation ABE fermentation development standpoint, the genetic tools developed in this work could catalyze discovery of new biology and strain engineering efforts in *C. saccharoperbutylacetonicum* and *C. roseum*. Methods of CRISPR-Cas9 based gene deletion in *C. saccharoperbutylacetonicum* have been independently developed and applied by other groups,[216,217] showcasing the prevalence of research interest in ABE strain engineering. This strain has been improved by genetic engineering in several recent studies.[70,112,218,219] Likewise, the optimized method for reliable conjugation and screening of *C. roseum* will allow implementation of traditional metabolic engineering strategies. The key developments were the identification of critical media inputs including yeast-extract free solid medium for gene delivery and identification of a detergent treatment to improve the resolution of colony-forming units to a single-cell level. Further expansion of the genetic toolkit will include development of more selectable markers, reporters, and inducible regulatory elements. The development of recombineering in clostridia is another promising step forward.[220] This, in combination with CRISPR selection[221] will enable more advanced and multiplexed engineering strategies to improve the throughput and sophistication of design-build-test cycles in synthetic biology for engineering clostridia.[222]

This work expands the body of knowledge of anaerobe-derived natural products and lends new opportunities to further investigate the identity, distribution, and function of natural products isolated from clostridia. To date, only three of thirty-nine identified BGCs from the ten strains of **Chapter 4** have associated chemical structures. The results of high-throughput elicitor screening in *C. beijerinckii* B-598 and *C. saccharoperbutylacetonicum* produced several leads representing BGC activation events which merit follow-up discovery efforts. In *C. roseum*, five promising orphan BGCs remain to be linked to their respective natural products. Three of these (BGCs 9, 14b, and 84) have proven challenging to isolate by screening and may require a selection-based mutagenesis in the manner of BGC 14a. The remaining two BGCs were successfully knocked out (BGC 2 and BGC 24) and are poised for metabolite identification. Meanwhile, BGC 1 has been linked to its associated compounds and their chemical structures remain to be determined. The BGC 14a associated clostyrylpyrones could benefit from rigorous characterization of their biological activity as well as the mechanism of their solvatochromism/thermochromism.

As clostridia are gifted with the most PKS/NRPS BGCs, they are a natural starting point for this work. However, growth of available genomic data[223] easily outpaces discovery efforts. One way to simplify discovery pipelines is to filter for strains with reported transcriptomic data and available genetic tools, two major bottlenecks of discovery in the current work. Genetics procedures could be standardized by consolidating discovery efforts into a set of heterologous expression strains in the CRAGE method.[224] This is especially applicable in anaerobic natural product discovery due to the focus on one taxon, the clostridia. A panel of chassis clostridia could speed up discovery and change the rate limiting step to the construction of molecular assemblies. Another way to alleviate current bottlenecks may be a discovery method that can bypass the application of genetic tools in favor of multi-omic approaches.[225] In addition, as the anaerobe-derived natural product discovery field matures, more genomic data and more advanced

bioinformatic profiling could identify new promising clades for study. For example, this work focuses on bacteria rather than archaea or eukaryotes. However, several instances of non-bacterial thiotemplate assembly line BGC expression can be gleaned from transcription-level studies, including an example from the methanogenic archaeon *Methanobrevibacter ruminantium*,[226] and examples from several species of Neocallimastigomycota.[227,228] The Neocallimastigomycota are intriguing because all of the known genomes appear to possess large thiotemplated assembly line BGCs, none of which have been characterized. On the other hand, only a handful of genomes are available, none of them are completely assembled, and no species has reported genetic tools, suggesting they are not ready for systematic genome mining efforts in the native organism. Improvements in bioinformatic detection of BGCs, for example the noncanonical thiotemplated BGC of closthioamide,[35] could represent additional avenues for natural product discovery within and without the clostridia. In the nearer term, the confluence of natural product biosynthetic potential and global demand for a return to sustainable bioprocessing will continue to motivate investigations of these remarkable anaerobic bacteria.

References

1. O'Brien, J. & Wright, G. D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **22**, 552–558 (2011).

2. Raaijmakers, J. M. & Mazzola, M. Diversity and Natural Functions of Antibiotics Produced by Beneficial and Plant Pathogenic Bacteria. *Annu. Rev. Phytopathol.* **50**, 403–424 (2012).

3. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products* vol. 79 629–661 (2016).

4. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes-a review. *Natural Product Reports* vol. 33 988–1005 (2016).

5. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann. N. Y. Acad. Sci.* **1354**, 82–97 (2015).

6. Sturgen, N. O. & Casida, L. E. Antibiotic production by anaerobic bacteria. *Appl. Environ. Microbiol.* **10**, 55–59 (1962).

7. Letzel, A. C., Pidot, S. J. & Hertweck, C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat. Prod. Rep.* **30**, 392–428 (2013).

8. Letzel, A. C., Pidot, S. J. & Hertweck, C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics* **15**, 983 (2014).

9. Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* **44**, 573–588 (2017).

10. Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).

11. Baltz, R. H. Synthetic biology, genome mining, and combinatorial biosynthesis of NRPS-derived antibiotics: a perspective. *J. Ind. Microbiol. Biotechnol.* **45**, 635–649 (2018).

12. Behnken, S. & Hertweck, C. Anaerobic bacteria as producers of antibiotics. *Applied Microbiology and Biotechnology* vol. 96 61–67 (2012).

13. Gebhart, D. *et al.* Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. *J. Bacteriol.* **194**, 6240–6247 (2012).

14. Mahony, D. E. & Butler, M. E. Bacteriocins of *Clostridium perfringens*. 1. Isolation and preliminary studies. *Can. J. Microbiol.* **17**, 1–6 (1971).

15. Timbermont, L. *et al.* Perfrin, a novel bacteriocin associated with netB positive *Clostridium perfringens* strains from broilers with necrotic enteritis. *Vet. Res.* **45**, 1–10 (2014).

16. Kemperman, R. *et al.* Identification and characterization of two novel clostridial

bacteriocins, circularin A and closticin 574. *Appl. Environ. Microbiol.* **69**, 1589–1597 (2003).

17. Martinez, F. A. C., Balciunas, E. M., Converti, A., Cotter, P. D. & De Souza Oliveira, R. P. Bacteriocin production by *Bifidobacterium* spp. A review. *Biotechnology Advances* vol. 31 482–488 (2013).

18. Dabard, J. *et al.* Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. *Appl. Environ. Microbiol.* **67**, 4111–8 (2001).

19. Gonzalez, D. J. *et al.* Clostridiolysin S, a post-translationally modified biotoxin from *Clostridium botulinum*. *J. Biol. Chem.* **285**, 28220–28228 (2010).

20. Guttenberger, N., Blankenfeldt, W. & Breinbauer, R. Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorganic Med. Chem.* **25**, 6149–6166 (2017).

21. Abken, H. J. *et al.* Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of *Methanosarcina mazei* Go1. *J. Bacteriol.* **180**, 2027–2032 (1998).

22. Beifuss, U., Tietze, M., Bäumer, S. & Deppenmeier, U. Methanophenazine: Structure, total synthesis, and function of a new cofactor from methanogenic archaea. *Angew. Chemie Int. Ed.* **39**, 2470–2472 (2000).

23. Beifuss, U. & Tietze, M. Methanophenazine and other natural biologically active phenazines. in *Natural Products Synthesis II. Topics in Current Chemistry* (ed. Mulzer, J.) vol. 244 77–113 (Springer, Berlin, Heidelberg, 2005).

24. Price-Whelan, A., Dietrich, L. E. P. & Newman, D. K. Rethinking "secondary" metabolism: Physiological roles for phenazine antibiotics. *Nat. Chem. Biol.* **2**, 71–78 (2006).

25. Blankenfeldt, W. & Parsons, J. F. The structural biology of phenazine biosynthesis. *Curr. Opin. Struct. Biol.* **29**, 26–33 (2014).

26. Ezaki, M., Muramatsu, H., Takase, S., Hashimoto, M. & Nagai, K. Naphthalecin, a novel antibiotic produced by the anaerobic bacterium, *Sporotalea colonica* sp. nov. *J. Antibiot. (Tokyo).* **61**, 207–212 (2008).

27. Ding, C. *et al.* An unusual stress metabolite from a hydrothermal vent fungus *Aspergillus* sp. WU 243 induced by cobalt. *Molecules* **21**, 105 (2016).

28. Shi, Y. *et al.* Stress-driven discovery of a cryptic antibiotic produced by *Streptomyces* sp. WU20 from Kueishantao hydrothermal vent with an integrated metabolomics strategy. *Appl. Microbiol. Biotechnol.* **101**, 1395–1408 (2017).

29. Vetriani, C., Speck, M. D., Ellor, S. V., Lutz, R. A. & Starovoytor, V. *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **54**, 175–181 (2004).

30. Andrianasolo, E. H. *et al.* Ammonificins A and B, hydroxyethylamine chroman

derivatives from a cultured marine hydrothermal vent bacterium, *Thermovibrio ammonificans*. *J. Nat. Prod.* **72**, 1216–1219 (2009).

31. Andrianasolo, E. H. *et al.* Ammonificins C and D, hydroxyethylamine chromene derivatives from a cultured marine hydrothermal vent bacterium, *Thermovibrio ammonificans*. *Mar. Drugs* **10**, 2300–2311 (2012).

32. Paredes, C. J., Alsaker, K. V. & Papoutsakis, E. T. A comparative genomic view of clostridial sporulation and physiology. *Nature Reviews Microbiology* vol. 3 969–978 (2005).

33. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: An unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chemie Int. Ed.* **49**, 2011–2013 (2010).

34. Behnken, S., Lincke, T., Kloss, F., Ishida, K. & Hertweck, C. Antiterminator-mediated unveiling of cryptic polythioamides in an anaerobic bacterium. *Angew. Chemie Int. Ed.* **51**, 2425–2428 (2012).

35. Dunbar, K. L. *et al.* Genome editing reveals novel thiotemplated assembly of polythioamide antibiotics in anaerobic bacteria. *Angew. Chemie Int. Ed.* **57**, 14080–14084 (2018).

36. Miari, V. F., Solanki, P., Hleba, Y., Stabler, R. A. & Heap, J. T. In vitro susceptibility to closthioamide among clinical and reference strains of *Neisseria gonorrhoeae*. *Antimicrob. Agents Chemother.* **61**, e00929-17 (2017).

37. Chiriac, A. I. *et al.* Mode of action of closthioamide: the first member of the polythioamide class of bacterial DNA gyrase inhibitors. *J. Antimicrob. Chemother.* **70**, 2576–2588 (2015).

38. Dunbar, K. L., Dell, M., Gude, F. & Hertweck, C. Reconstitution of polythioamide antibiotic backbone formation reveals unusual thiotemplated assembly strategy. *Proc. Natl. Acad. Sci.* **117**, 8850–8858 (2020).

39. Pidot, S., Ishida, K., Cyrulies, M. & Hertweck, C. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Angew. Chemie - Int. Ed.* **53**, 7856–7859 (2014).

40. Shabuer, G. *et al.* Plant pathogenic anaerobic bacteria use aromatic polyketides to access aerobic territory. *Science (80-.).* **350**, 670–674 (2015).

41. Fajardo, A. & Martínez, J. L. Antibiotics as signals that trigger specific bacterial responses. *Current Opinion in Microbiology* vol. 11 161–167 (2008).

42. Schieferdecker, S., Shabuer, G., Knuepfer, U. & Hertweck, C. Clostrindolin is an antimycobacterial pyrone alkaloid from *Clostridium beijerinckii*. *Org. Biomol. Chem.* **17**, 6119–6121 (2019).

43. Schieferdecker, S. *et al.* Biosynthesis of diverse antimicrobial and antiproliferative acyloins in anaerobic bacteria. *ACS Chem. Biol.* **14**, 1490–1497 (2019).

44. Zhu, B., Wang, X. & Li, L. Human gut microbiome: The second genome of human body. *Protein and Cell* vol. 1 718–725 (2010).

45. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).

46. Guo, C. J. *et al.* Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* **168**, 517-526.e18 (2017).

47. Otto, H. H. & Schirmeister, T. Cysteine proteases and their inhibitors. *Chem. Rev.* **97**, 133–171 (1997).

48. Westerik, J. O. & Wolfenden, R. Aldehydes as inhibitors of papain. *J. Biol.* **247**, 8105–8197 (1972).

49. Woo, J.-T. *et al.* Peptidyl aldehyde derivatives as potent and selective inhibitors of cathepsin L. *Bioorganic Med. Chem. Lett.* **5**, 1501–1504 (1995).

50. Schneider, B. A. & Balskus, E. P. Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway. *Tetrahedron* **74**, 3215–3230 (2018).

51. Sugimoto, Y. *et al.* A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science (80-.).* **366**, (2019).

52. Rischer, M. *et al.* Biosynthesis, synthesis, and activities of barnesin A, a NRPS-PKS hybrid produced by an anaerobic epsilonproteobacterium. *ACS Chem. Biol.* **13**, 1990–1995 (2018).

53. Herman, N. A. *et al.* The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation. *Nat. Commun.* **8**, 1514 (2017).

54. Lütke-Eversloh, T. & Bahl, H. Metabolic engineering of *Clostridium acetobutylicum*: Recent advances to improve butanol production. *Current Opinion in Biotechnology* vol. 22 634–647 (2011).

55. Moon, H. G. *et al.* One hundred years of clostridial butanol fermentation. *FEMS Microbiol. Lett.* **363**, fnw001 (2016).

56. Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiol. Rev.* **50**, 484–524 (1986).

57. Guo, M., Song, W. & Buhain, J. Bioenergy and biofuels: History, status, and perspective. *Renew. Sustain. Energy Rev.* **42**, 712–725 (2015).

58. Cheng, C., Bao, T. & Yang, S. T. Engineering *Clostridium* for improved solvent production: recent progress and perspective. *Applied Microbiology and Biotechnology* vol. 103 5549–5566 (2019).

59. Yoo, M., Nguyen, N. P. T. & Soucaille, P. Trends in systems biology for the analysis and engineering of *Clostridium acetobutylicum* metabolism. *Trends Microbiol.* **28**, 118–140 (2020).

60. Venkataramanan, K. P. *et al.* The *Clostridium* small RNome that responds to stress: The paradigm and importance of toxic metabolite stress in *C. acetobutylicum*. *BMC Genomics* **14**, 849 (2013).

61. Senger, R. S. & Papoutsakis, E. T. Genome-scale model for *Clostridium acetobutylicum* : Part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* **101**, 1036–1052 (2008).

62. Wu, C. *et al.* A quantitative lens on anaerobic life: leveraging the state-of-the-art fluxomics approach to explore clostridial metabolism. *Current Opinion in Biotechnology* vol. 64 47–54 (2020).

63. Patakova, P., Linhova, M., Rychtera, M., Paulova, L. & Melzoch, K. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. *Biotechnol. Adv.* **31**, 58–67 (2013).

64. Poehlein, A. *et al.* Microbial solvent formation revisited by comparative genome analysis. *Biotechnol. Biofuels* **10**, 58 (2017).

65. Zhao, Y. *et al.* Expression of a cloned cyclopropane fatty acid synthase gene reduces solvent formation in *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **69**, 2831–2841 (2003).

66. Motoyoshi, H. Process for producing butanol by fermentation. US patent 2945786A. (1960).

67. Sauer, M. Industrial production of acetone and butanol by fermentation—100 years later. *FEMS Microbiol. Lett.* **363**, fnw134 (2016).

68. Green, E. M. Fermentative production of butanol-the industrial perspective. *Curr. Opin. Biotechnol.* **22**, 337–343 (2011).

69. Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C. & Papoutsakis, E. T. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr. Opin. Biotechnol.* **23**, 364–381 (2012).

70. Zhang, J. *et al.* Enhancement of sucrose metabolism in *Clostridium saccharoperbutylacetonicum* N1-4 through metabolic engineering for improved acetone–butanol–ethanol (ABE) fermentation. *Bioresour. Technol.* **270**, 430–438 (2018).

71. Al-Shorgani, N. K. N., Kalil, M. S. & Yusoff, W. M. W. Biobutanol production from rice bran and de-oiled rice bran by *Clostridium saccharoperbutylacetonicum* N1-4. *Bioprocess Biosyst. Eng.* **35**, 817–826 (2012).

72. Richter, H. *et al.* Prolonged conversion of n-butyrate to n-butanol with *Clostridium saccharoperbutylacetonicum* in a two-stage continuous culture with in-situ product removal. *Biotechnol. Bioeng.* **109**, 913–921 (2012).

73. Shukor, H. *et al.* Production of butanol by *Clostridium saccharoperbutylacetonicum* N1-4 from palm kernel cake in acetone-butanol-ethanol fermentation using an empirical model.

Bioresour. Technol. **170**, 565–573 (2014).

74. Tashiro, Y. *et al.* High butanol production by *Clostridium saccharoperbutylacetonicum* N1-4 in fed-batch culture with pH-stat continuous butyric acid and glucose feeding method. *J. Biosci. Bioeng.* **98**, 263–268 (2004).

75. Thang, V. H., Kanda, K. & Kobayashi, G. Production of Acetone-Butanol-Ethanol (ABE) in direct fermentation of cassava by *Clostridium saccharoperbutylacetonicum* N1-4. *Appl. Biochem. Biotechnol.* **161**, 157–170 (2010).

76. Yao, D. *et al.* Robustness of *Clostridium saccharoperbutylacetonicum* for acetone-butanol-ethanol production: Effects of lignocellulosic sugars and inhibitors. *Fuel* **208**, 549–557 (2017).

77. Avci, A., Kiliç, N. K., Dönmez, G. & Dönmez, S. Evaluation of hydrogen production by *Clostridium* strains on beet molasses. *Environ. Technol. (United Kingdom)* **35**, 278–285 (2014).

78. Baba, S. ichi, Tashiro, Y., Shinto, H. & Sonomoto, K. Development of high-speed and highly efficient butanol production systems from butyric acid with high density of living cells of *Clostridium saccharoperbutylacetonicum*. *J. Biotechnol.* **157**, 605–612 (2012).

79. Dada, O., Yusoff, W. M. W. & Kalil, M. S. Biohydrogen production from ricebran using *Clostridium saccharoperbutylacetonicum* N1-4. *Int. J. Hydrogen Energy* **38**, 15063–15073 (2013).

80. Kiyoshi, K. *et al.* Butanol production from alkali-pretreated rice straw by co-culture of *Clostridium thermocellum* and *Clostridium saccharoperbutylacetonicum*. *Bioresour. Technol.* **186**, 325–328 (2015).

81. Kobayashi, G. *et al.* Utilization of excess sludge by acetone-butanol-ethanol fermentation employing *Clostridium saccharoperbutylacetonicum* N1-4 (ATCC 13564). *J. Biosci. Bioeng.* **99**, 517–519 (2005).

82. Nakayama, S., Kiyoshi, K., Kadokura, T. & Nakazato, A. Butanol production from crystalline cellulose by cocultured *Clostridium thermocellum* and *Clostridium saccharoperbutylacetonicum* N1-4. *Appl. Environ. Microbiol.* **77**, 6470–6475 (2011).

83. Noguchi, T. *et al.* Efficient butanol production without carbon catabolite repression from mixed sugars with *Clostridium saccharoperbutylacetonicum* N1-4. *J. Biosci. Bioeng.* **116**, 716–721 (2013).

84. Oshiro, M., Hanada, K., Tashiro, Y. & Sonomoto, K. Efficient conversion of lactic acid to butanol with pH-stat continuous lactic acid and glucose feeding method by *Clostridium saccharoperbutylacetonicum*. *Appl. Microbiol. Biotechnol.* **87**, 1177–1185 (2010).

85. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chemie - Int. Ed.* **48**, 4688–4716 (2009).

86. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135 (2016).

87. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11**, 639–648 (2015).

88. del Cerro, C. *et al.* Genome sequence of the butanol hyperproducer *Clostridium saccharoperbutylacetonicum* N1-4. *Genome Announc.* **1**, e0007013 (2013).

89. Poehlein, A., Krabben, P., Dürre, P. & Daniel, R. Complete genome sequence of the solvent producer *Clostridium saccharoperbutylacetonicum* strain DSM 14923. *Genome Announc.* **2**, e01056-14 (2014).

90. Li, J. S., Barber, C. C. & Zhang, W. Natural products from anaerobes. *J. Ind. Microbiol. Biotechnol.* **46**, 375–383 (2019).

91. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

92. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45**, W49–W54 (2017).

93. Medema, M. H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with multigeneblast. *Mol. Biol. Evol.* **30**, 1218–1223 (2013).

94. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).

95. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**, 734 (2012).

96. Rocha, D. J. P., Santos, C. S. & Pacheco, L. G. C. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **108**, 685–693 (2015).

97. Li, Q. *et al.* CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol. J.* **11**, 961–972 (2016).

98. Xu, T. *et al.* Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Appl. Environ. Microbiol.* **81**, 4423–4431 (2015).

99. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

100. Keis, S., Shaheen, R. & Jones, D. T. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int. J. Syst. Evol. Microbiol.* **51**, 2095–2103 (2001).

101. Hayashi, S. *et al.* Analysis of organic solvent tolerance in *Escherichia coli* using gene expression profiles from DNA microarrays. *J. Biosci. Bioeng.* **95**, 379–383 (2003).

102. Shimizu, K. *et al.* Discovery of glpC, an organic solvent tolerance-related gene in *Escherichia coli*, using gene expression profiles from DNA microarrays. *Appl. Environ.*

Microbiol. **71**, 1093–1096 (2005).

103. Alsaker, K. V., Paredes, C. & Papoutsakis, E. T. Metabolite stress and tolerance in the production of biofuels and chemicals: Gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. *Biotechnol. Bioeng.* **105**, 1131–1147 (2010).

104. Kell, D. B., Peck, M. W., Rodger, G. & Morris, J. G. On the permeability to weak acids and bases of the cytoplasmic membrane of Clostridium pasteurianum. *Biochem. Biophys. Res. Commun.* **99**, 81–88 (1981).

105. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science (80-.).* **291**, 1790–1792 (2001).

106. Wei, H. *et al.* Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-seq. *Front. Microbiol.* **5**, 142 (2014).

107. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (2010).

108. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

109. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

110. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

111. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

112. Herman, N. A. *et al.* Development of a high-efficiency transformation method and implementation of rational metabolic engineering for the industrial butanol hyperproducer *Clostridium saccharoperbutylacetonicum* strain N1-4. *Appl. Environ. Microbiol.* **83**, e02942-16 (2017).

113. Green, E. M. *et al.* Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology* **142**, 2079–2086 (1996).

114. Gowda, H. *et al.* Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* **86**, 6931–6939 (2014).

115. Zhang, W., Bolla, M. L., Kahne, D. & Walsh, C. T. A three enzyme pathway for 2-amino-3-hydroxycyclopent-2-enone formation and incorporation in natural product biosynthesis. *J. Am. Chem. Soc.* **132**, 6402–6411 (2010).

116. Tomas, C. A., Beamish, J. & Papoutsakis, E. T. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* **186**, 2006–2018 (2004).

117. McCoy, E. & McClung, L. S. Studies on anaerobic bacteria. VI. The nature and systematic position of a new chromogenic *Clostridium*. *Arch. Mikrobiol.* **6**, 230–238 (1935).

118. Hitzman, D. O., Halvorson, H. O. & Ukita, T. Requirements for production and germination of spores of anaerobic bacteria. *J. Bacteriol.* **74**, 1–7 (1957).

119. Wooley, B. C. & Collier, R. E. Changes in thermoresistance of *Clostridium roseum* as related to the intracellular content of calcium and dipicolinic acid. *Can. J. Microbiol.* **11**, 279–285 (1965).

120. Byrne, A. F., Burton, T. H. & Koch, R. B. Relation of dipicolinic acid content of anaerobic bacterial endospores to their heat resistance. *J. Bacteriol.* **80**, 139–140 (1960).

121. Zou, W., Ye, G. & Zhang, K. Diversity, function, and application of *Clostridium* in Chinese strong flavor baijiu ecosystem: a review. *J. Food Sci.* **83**, 1193–1199 (2018).

122. Bando, Y., Fujimoto, N., Suzuki, M. & Ohnishi, A. A microbiological study of biohydrogen production from beer lees. *Int. J. Hydrogen Energy* **38**, 2709–2718 (2013).

123. Rajhi, H., Conthe, M., Puyol, D., Díaz, E. & Sanz, J. L. Dark fermentation: isolation and characterization of hydrogen-producing strains from sludges. *Int. Microbiol.* **16**, 53–62 (2013).

124. Kongjan, P., O-Thong, S. & Angelidaki, I. Performance and microbial community analysis of two-stage process with extreme thermophilic hydrogen and thermophilic methane production from hydrolysate in UASB reactors. *Bioresour. Technol.* **102**, 4028–4035 (2011).

125. Chang, J. J. *et al.* Syntrophic co-culture of aerobic Bacillus and anaerobic Clostridium for bio-fuels and bio-hydrogen production. *Int. J. Hydrogen Energy* **33**, 5137–5146 (2008).

126. Abd-Alla, M. H., Zohri, A. N. A., El-Enany, A. W. E. & Ali, S. M. Acetone-butanol-ethanol production from substandard and surplus dates by Egyptian native *Clostridium* strains. *Anaerobe* **32**, 77–86 (2015).

127. Behnken, S. & Hertweck, C. Cryptic polyketide synthase genes in non-pathogenic *Clostridium* spp. *PLoS One* **7**, (2012).

128. Zerikly, M. & Challis, G. L. Strategies for the discovery of new natural products by genome mining. *ChemBioChem* **10**, 625–633 (2009).

129. Okuyama, H., Orikasa, Y., Nishida, T., Watanabe, K. & Morita, N. Bacterial genes responsible for the biosynthesis of eicosapentaenoic and docosahexaenoic acids and their heterologous expression. *Appl. Environ. Microbiol.* **73**, 665–670 (2007).

130. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology* vol. 13 509–523 (2015).

131. Hopwood, D. A. & Wright, H. M. Bacterial protoplast fusion: recombination in fused protoplasts of *Streptomyces coelicolor*. **162**, 317 (1978).

132. Scott, P. T. & Rood, J. I. Electroporation-mediated transformation of lysostaphin-treated *Clostridium perfringens*. *Gene* **82**, 327–333 (1989).

133. Reysset, G., Hubert, J., Podvin, L. & Sebald, M. Protoplast formation and regeneration of *Clostridium acetobutylicum* strain N1-4080. *Microbiology* **133**, 2595–2600 (1987).

134. Lin, Y. L. & Blaschek, H. P. Transformation of teat-treated *Clostridium acetobutylicum* protoplasts with pUB110 plasmid DNA. *Appl. Environ. Microbiol.* **48**, 737–42 (1984).

135. Reysset, G., Hubert, J., Podvin, L. & Sebald, M. Transfection and transformation of *Clostridium acetobutylicum* strain N1-4081 protoplasts. *Biotechnology* **2**, 199–204 (1988).

136. Tabor, C. W. Stabilization of protoplasts and spheroplasts by spermine and other polyamines. *J. Bacteriol.* **83**, 1101–11 (1962).

137. Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* **46**, 439–452 (2002).

138. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical streptomyces genetics (Vol. 291)*. (John Innes Foundation, 2000).

139. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78**, 79–85 (2009).

140. Blodgett, J. A. V. *et al.* Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nat. Chem. Biol.* **3**, 480–485 (2007).

141. Bradley, D. Specification of the conjugative pili and surface mating systems of *Pseudomonas* plasmids. *J. Gen. Microbiol.* **129**, 2545–2556 (1983).

142. Kuehne, S. A. & Minton, N. P. ClosTron-mediated engineering of *Clostridium*. *Bioengineered* **3**, 247–254 (2012).

143. Zhong, J., Karberg, M. & Lambowitz, A. M. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res.* **31**, 1656–1664 (2003).

144. Minton, N. P. *et al.* A roadmap for gene system development in *Clostridium*. *Anaerobe* **41**, 104–112 (2016).

145. Saitoh, H., Yamane, N., Miyagi, C. & Nakasone, I. Comparative evaluation of two different formulae of Middlebrook 7H9 broth in a fully automated mycobacteria culture system, MB/BacT; the effect of Tween 80. *Rinsho Biseibutshu. Jinsoku Shindan Kenkyukai Shi* **11**, 79–85 (2000).

146. Perutka, J., Wang, W., Goerlitz, D. & Lambowitz, A. M. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J. Mol. Biol.* **336**, 421–439 (2004).

147. Al-Hinai, M. A., Fast, A. G. & Papoutsakis, E. T. Novel system for efficient isolation of clostridium double-crossover allelic exchange mutants enabling markerless chromosomal

gene deletions and DNA integration. *Appl. Environ. Microbiol.* **78**, 8112–8121 (2012).

148. Reichardt, C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chem. Rev.* **94**, 2319–2358 (1994).

149. He, J. & Chen, J. The solvatochromic materials: A pProgress review. *Mater. Sci. Forum* **914**, 182–192 (2018).

150. Pluskal, T. *et al.* The biosynthetic origin of psychoactive kavalactones in kava. *Nat. Plants* **5**, 867–878 (2019).

151. Wu, K., Chung, L., Revill, W. P., Katz, L. & Reeves, C. D. The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251**, 81–90 (2000).

152. Walton, L. J., Corre, C. & Challis, G. L. Mechanisms for incorporation of glycerol-derived precursors into polyketide metabolites. *J. Ind. Microbiol. Biotechnol.* **33**, 105–120 (2006).

153. Wenzel, S. C. *et al.* On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of "glycolate" units into ansamitocin and soraphen A. *J. Am. Chem. Soc.* **128**, 14325–14336 (2006).

154. Sun, Y., Hong, H., Gillies, F., Spencer, J. B. & Leadlay, P. F. Glyceryl-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics. *ChemBioChem* **9**, 150–156 (2008).

155. Dorrestein, P. C. *et al.* The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *J. Am. Chem. Soc.* **128**, 10386–10387 (2006).

156. Kjaerulff, L. *et al.* Pyxipyrrolones: structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. *Angew. Chemie Int. Ed.* **56**, 9614–9618 (2017).

157. Schäberle, T. F. Biosynthesis of α-pyrones. *Beilstein J. Org. Chem.* **12**, 571–588 (2016).

158. Fernández-Moriano, C., Gómez-Serranillos, M. P. & Crespo, A. Antioxidant potential of lichen species and their secondary metabolites. A systematic review. *Pharm. Biol.* **54**, 1–17 (2016).

159. Quadri, L. E. N. Assembly of aryl-capped siderophores by modular peptide synthetases and polyketide synthases. *Mol. Microbiol.* **37**, 1–12 (2000).

160. Crosa, J. H. & Walsh, C. T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **66**, 223–249 (2002).

161. Pfleger, B. F. *et al.* Structural and functional analysis of AsbF: Origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17133–17138 (2008).

162. Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **19**, 70–99 (2002).

163. Muir, R. M. *et al.* Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). *Plant Mol. Biol.* **75**, 555–565 (2011).

164. Demydchuk, Y. *et al.* Analysis of the tetronomycin gene cluster: insights into the biosynthesis of a polyether tetronate antibiotic. *ChemBioChem* **9**, 1136–1145 (2008).

165. Tao, W., Zhu, M., Deng, Z. & Sun, Y. Biosynthesis of tetronate antibiotics: A growing family of natural products with broad biological activities. *Science China Chemistry* vol. 56 1364–1371 (2013).

166. Kanchanabanca, C. *et al.* Unusual acetylation-elimination in the formation of tetronate antibiotics. *Angew. Chemie Int. Ed.* **52**, 5785–5788 (2013).

167. Brewer, M. S. Natural antioxidants: Sources, compounds, mechanisms of action, and potential applications. *Compr. Rev. Food Sci. Food Saf.* **10**, 221–247 (2011).

168. Sylvester, P. W. Optimization of the tetrazolium dye (MTT) colorimetric assay for cellular growth and viability. in *Methods in Molecular Biology* (ed. Satyanarayanajois, S.) vol. 716 157–168 (Humana Press, 2011).

169. Hafner, M., Niepel, M., Chung, M. & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* **13**, 521–527 (2016).

170. Liu, Y. & Nair, M. G. An efficient and economical MTT assay for determining the antioxidant activity of plant natural product extracts and pure compounds. *J. Nat. Prod.* **73**, 1193–1195 (2010).

171. O'Brien, R. W. & Morris, J. G. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. *J. Gen. Microbiol.* **68**, 307–318 (1971).

172. Kolek, J., Sedlar, K., Provaznik, I. & Patakova, P. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnol. Biofuels* **9**, 14 (2016).

173. Korbie, D. J. & Mattick, J. S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protoc.* **3**, 1452–1456 (2008).

174. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

175. Hillmann, F., Fischer, R.-J., Saint-Prix, F., Girbal, L. & Bahl, H. PerR acts as a switch for oxygen tolerance in the strict anaerobe *Clostridium acetobutylicum*. *Mol. Microbiol.* **68**, 848–860 (2008).

176. King, K. Y., Horenstein, J. A. & Caparon, M. G. Aerotolerance and peroxide resistance in peroxidase and PerR mutants of *Streptococcus pyogenes*. *J. Bacteriol.* **182**, 5290–5299 (2000).

177. Wright, G. D. Antibiotic adjuvants: rescuing antibiotics from resistance. *Trends in Microbiology* vol. 24 862–871 (2016).

178. Robinson, T. P. *et al.* Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg.* **110**, 377–380 (2016).

179. Chevrette, M. G. & Currie, C. R. Emerging evolutionary paradigms in antibiotic discovery. *J. Ind. Microbiol. Biotechnol.* **46**, 257–271 (2019).

180. Rossiter, S. E., Fletcher, M. H. & Wuest, W. M. Natural products as platforms to overcome antibiotic resistance. *Chem. Rev.* **117**, 12415–12474 (2017).

181. Olga Genilloud. Actinomycetes: Still a source of novel antibiotics. *Natural Product Reports* vol. 34 1203–1232 (2017).

182. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).

183. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai).* **47**, 231–243 (2015).

184. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4**, 1020–1029 (2015).

185. Zhang, M. M. *et al.* CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* **13**, 607–609 (2017).

186. Guo, F. *et al.* Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab. Eng.* **28**, 134–142 (2015).

187. Li, P., Guo, Z., Tang, W. & Chen, Y. Activation of three natural product biosynthetic gene clusters from *Streptomyces lavendulae* CGMCC 4.1386 by a reporter-guided strategy. *Synth. Syst. Biotechnol.* **3**, 254–260 (2018).

188. Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7266–7271 (2014).

189. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **3**, 619 (2002).

190. Donadio, S., Monciardini, P. & Sosio, M. Polyketide synthases and nonribosomal peptide synthetases: The emerging view from bacterial genomics. *Natural Product Reports* vol. 24 1073–1079 (2007).

191. Yin, Y., Zhang, Y., Karakashev, D. B., Wang, J. & Angelidaki, I. Biological caproate production by Clostridium kluyveri from ethanol and acetate as carbon sources. *Bioresour. Technol.* **241**, 638–644 (2017).

192. Mao, D., Okada, B. K., Wu, Y., Xu, F. & Seyedsayamdost, M. R. Recent advances in activating silent biosynthetic gene clusters in bacteria. *Current Opinion in Microbiology* vol. 45 156–163 (2018).

193. Xu, F., Nazari, B., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. Discovery of a

cryptic antifungal Compound from *Streptomyces albus* J1074 using high-throughput slicitor screens. *J. Am. Chem. Soc* **139**, 9203–9212 (2017).

194. Lin, H., Deng, E.-Z., Ding, H., Chen, W. & Chou, K.-C. iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition. *Nucleic Acids Res.* **42**, 12961–12972 (2014).

195. Solovyev, V. & Salamov, A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (ed. Li, R. W.) 61–78 (Nova Science Publishers, 2011).

196. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).

197. Cho, D. H., Lee, Y. J., Um, Y., Sang, B. I. & Kim, Y. H. Detoxification of model phenolic compounds in lignocellulosic hydrolysates with peroxidase for butanol production from *Clostridium beijerinckii*. *Appl. Microbiol. Biotechnol.* **83**, 1035–1043 (2009).

198. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. (2005) doi:10.1038/NMETH819.

199. Sample, V., Newman, R. H. & Zhang, J. The structure and function of fluorescent proteins. *Chem. Soc. Rev.* **38**, 2852–2864 (2009).

200. Drepper, T., Gensch, T. & Pohl, M. Advanced in vivo applications of blue light photoreceptors as alternative fluorescent proteins. *Photochemical and Photobiological Sciences* vol. 12 1125–1134 (2013).

201. Dolgosheina, E. V. *et al.* RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **9**, 2412–2420 (2014).

202. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).

203. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).

204. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).

205. Drepper, T. *et al.* Reporter proteins for *in vivo* fluorescence without oxygen. *Nat. Biotechnol.* **25**, 443–445 (2007).

206. Mukherjee, A. *et al.* Engineering and characterization of new LOV-based fluorescent proteins from *Chlamydomonas Reinhardtii* and *Vaucheria frigida*. *ACS Synth. Biol.* **4**, 371–377 (2015).

207. Higgins, S. A., Ouonkap, S. V. Y. & Savage, D. F. Rapid and programmable protein mutagenesis using plasmid recombineering. *ACS Synth. Biol.* **6**, 1825–1833 (2017).

208. Chapman, S. *et al.* The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20038–20043 (2008).

209. Patridge, E., Gareiss, P., Kinch, M. S. & Hoyer, D. An analysis of FDA-approved drugs: Natural products and their derivatives. *Drug Discovery Today* vol. 21 204–207 (2016).

210. Yang, X., Xu, M. & Yang, S. T. Metabolic and process engineering of *Clostridium cellulovorans* for biofuel production from cellulose. *Metab. Eng.* **32**, 39–48 (2015).

211. Ramachandran, U., Wrana, N., Cicek, N., Sparling, R. & Levin, D. B. Hydrogen production and end-product synthesis patterns by *Clostridium termitidis* strain CT1112 in batch fermentation cultures with cellobiose or α-cellulose. *Int. J. Hydrogen Energy* **33**, 7006–7012 (2008).

212. Tamburini, E., Daly, S., Steiner, U., Vandini, C. & Mastromei, G. *Clostridium felsineum* and *Clostridium acetobutylicum* are two distinct species that are phylogenetically closely related. *Int. J. Syst. Evol. Microbiol.* **51**, 963–966 (2001).

213. Lund, B. M. & Brocklehurst, T. F. Pectic enzymes of pigmented strains of *Clostridium*. *J. Gen. Microbiol.* **104**, 59–66 (1978).

214. Vogl, C. *et al.* Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J. Bacteriol.* **189**, 7367–7375 (2007).

215. Hongo, M., Murata, A., Kono, K. & Kato, F. Lysogeny and bacteriocinogeny in strains of *Clostridium* species. *Agric. Biol. Chem.* **32**, 27–33 (1968).

216. Atmadjaja, A. N. *et al.* CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT). *FEMS Microbiol. Lett.* **366**, fnz059 (2019).

217. Wang, S., Dong, S., Wang, P., Tao, Y. & Wang, Y. Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **83**, e00233-17 (2017).

218. Wang, S., Dong, S. & Wang, Y. Enhancement of solvent production by overexpressing key genes of the acetone-butanol-ethanol fermentation pathway in *Clostridium saccharoperbutylacetonicum* N1-4. *Bioresour. Technol.* **245**, 426–433 (2017).

219. Gu, Y. *et al.* Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1-4 (HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism. *Fuel* **236**, 1559–1566 (2019).

220. Dong, H., Tao, W., Gong, F., Li, Y. & Zhang, Y. A functional *recT* gene for recombineering of *Clostridium*. *J. Biotechnol.* **173**, 65–67 (2014).

221. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).

222. Charubin, K., Bennett, R. K., Fast, A. G. & Papoutsakis, E. T. Engineering *Clostridium*

organisms as microbial cell-factories: challenges & opportunities. *Metabolic Engineering* vol. 50 173–191 (2018).

223. Sayers, E. W. *et al.* GenBank. *Nucleic Acids Res.* **47**, D94–D99 (2019).

224. Wang, G. *et al.* CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* **4**, 2498–2510 (2019).

225. Li, A. *et al.* Multi-omic analyses provide links between low-dose antibiotic treatment and induction of secondary metabolism in *Burkholderia thailandensis*. *MBio* **11**, e03210-19 (2020).

226. Zhou, X. *et al.* The transcriptome response of the ruminal methanogen *Methanobrevibacter ruminantium* strain M1 to the inhibitor lauric acid. *BMC Res. Notes* **11**, 135 (2018).

227. Solomon, K. V. *et al.* Early-branching gut fungi possess large, comprehensive array of biomass-degrading enzymes. *Science (80-.).* **351**, 1192–1195 (2016).

228. Couger, M. B., Youssef, N. H., Struchtemeyer, C. G., Liggenstoffer, A. S. & Elshahed, M. S. Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* sp. strain C1A. *Biotechnol. Biofuels* **8**, 208 (2015).

229. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

230. Lee, S. Y., Bennett, G. N. & Papoutsakis, E. T. Construction of *Escherichia coli-Clostridium acetobutylicum* shuttle vectors and transformation of *Clostridium acetobutylicum* strains. *Biotechnol. Lett.* **14**, 427–432 (1992).

231. Zhu, X., Liu, J. & Zhang, W. *De novo* biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115–120 (2014).

232. Pesenti, C. *et al.* Total synthesis of a pepstatin analog incorporating two trifluoromethyl hydroxymethylene isosteres (Tfm-GABOB) and evaluation of Tfm-GABOB containing peptides as inhibitors of HIV-1 protease and MMP-9. *Tetrahedron* **57**, 6511–6522 (2001).

233. Rengasamy, R. *et al.* New building block for polyhydroxylated piperidine: Total synthesis of 1,6-dideoxynojirimycin. *J. Org. Chem.* **73**, 2898–2901 (2008).

234. Steiner, E., Scott, J., Minton, N. P. & Winzer, K. An *agr* quorum sensing system that regulates granulose formation and sporulation in *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **78**, 1113–1122 (2012).

Appendix A.

Supplementary Methods

Plasmid construction. All relevant synthetic oligonucleotide sequences are listed in **Table 2-2** (**Chapter 2**). The CRISPR-mediated gene knockout of CSPA_RS10760 in *Csa* was pNK52. Intermediate plasmid pNK45 was double-digested using XbaI and SpeI to prepare the 9.2 kb plasmid backbone. The 20-nt chimeric single guide RNA was performed by adding appropriate overhangs to two PCR products: the 0.37 kb C. beijerinckii NCIMB 8052 sRNA promoter P_{cbei_5830} (amplified with Pcbei_F and Pcbei_R_nrps3) and (amplified with primers gRNA_F_nrps3 and gRNA_R). Two additional target-specific PCRs were performed to yield the 1-2 kb upstream and downstream homology regions (UHR_nrps3_Fo and UHR_nrps3_Ro and DHR_nrps3_Fo & DHR_nrps3_Ro, respectively). Gibson assembly was performed using the five generated fragments: pNK45 backbone, retargeted P_{sCbei_5830}, retargeted gRNA, UHR, and DHR. All vectors were confirmed by test digestion and Sanger sequencing.

For the precursor plasmid pNK45, the *S. pyogenes* cas9 gene (4.1 kb) was amplified from vector pMJ806[229] (purchased from Addgene) using primers p45_cas9n_Fo & p45_cas9n_Ro to install the D10A mutation to generate *cas9* 'nickase'. For assembly of pNK45, double digested plasmid pNK44 (BsrGI and PmeI) (7.6 kb) was ligated with double digested *cas9* nickase PCR product (BsrGI and PmeI).

The precursor plasmid pNK44 was constructed from a modified pIMP1 vector backbone[230] with the *cas9* gene inserted (under control of P_{bdh}). The pWIS vector backbone was amplified from a pIMP, (pWIS_bk E, 2.8 kb, primers p43_bk_F & p43_bk_Ro; pWIS_bk F, 2.0 kb, primers p43_bk_F2 & p43_bk_R2), which feature a 24-bp overlap and addition of XbaI and SpeI restriction sites at the overlap site. The *bdh* promoter (0.3 kb) was amplified from *Csa* genomic DNA with primers p43_Pbdh_Fo & p43_Pbdh_R. The *S. pyogenes cas9* gene (4.1 kb) was amplified from vector pMJ806 with primers p43_cas9_Fo & p43_cas9_Ro, which also added a BsrGI restriction site before the start codon of *cas9*. For Gibson assembly of pNK43, PCR amplified pWIS_bk E, pWIS_bk_F, P_{bdh}, and *cas9* were combined. For construction of pNK44, a we obtained a synthesized chimeric guide RNA (gRNA) driven by the *C. beijerinckii* NCIMB 8052 sRNA promoter $P_{sCbei\_5830195}$ and featuring a 20-bp sequence targeting the *Csa* pyrF gene (CSPA_RS05335) based on an "NGG" PAM sequence (481-bp total). The entire sequence was arranged as follows (5' to 3'): pWIS_bk E overhang (24-bp), XbaI site (6-bp), sCbei_5803 promoter sequence (314-bp), XhoI site (6-bp), pyrF targeting region (20-bp), Cas9 binding region (96-bp), terminator (7-bp), and NotI site (8-bp). The upstream homologous region (UHR, 1.0 kb) and downstream homologous region (DHR, 1.0 kb) of the pyrF gene were amplified from *Csa* genomic DNA using primer sets p44_UHR_pyrF_Fo & p44_UHR_pyrF_R and p44_DHR_pyrF_Fo & p44_DHR_pyrF_Ro, respectively. For Gibson assembly of pNK44, double digested plasmid pNK43 (XbaI and SpeI) (9.2 kb) was combined with the PCR amplified *pyrF* UHR/DHR and the synthesized gRNA/promoter.

The pXZ247 construct was assembled by a three-fragment Gibson assembly. CSPA_RS10760 was amplified from *Csa* genomic DNA in two parts (using pXZ247_F_BamHI + pXZ247_R1, and pXZ247_F1 + pXZ247_R_pstI). The vector was prepared by linearizing a pETDuet-1 (EMD Millipore) based intermediate plasmid, pXZ34,[231] using BamHI and PstI.

Purification of compound 1. Batch fermentations of *Csa* wild-type were initiated from 10 ml overnight cultures in PL7 at 37°C. This was subcultured at 10% inoculum into 30 ml PL7S (PL7G with 65 g/L sucrose instead of glucose, and additionally supplemented wih 300 mg/L adenine. After 3-4 hours, the cultures in mid-exponential phase (OD_{600} of 0.6) were used to inoculate two aliquots of 1.5 L PL7S (with adenine and Antifoam 204 (10 µl each) in Ultra Yield™ Flasks (Thomson Instrument Company, Oceanside, CA). The stagnant cultures were incubated at 30°C for 4 days before harvesting.

The cells were pelleted out via centrifugation (3,500×g, 15 min) at room temperature, and the cell-free supernatant was extracted multiple times using 1 volume ethyl acetate. After phase separation, the solvent was collected and removed under vacuum and the residue was re-dissolved in dichloromethane. The crude extract was fractionated using a size-exclusion column packed with Sephadex LH-20 (Sigma-Aldrich) and manually fractionated in a 1:1 dichloromethane:methanol mobile phase. Fractions were screened by LC-MS and fractions containing compound **1** were consolidated and dried under vacuum before further purification using reverse-phase HPLC (using a Agilent 1260 HPLC with DAD). The material was dissolved in methanol and purified on a semi-preparative C18 column (5 µm, 10×250 mm, 300 Å, Vydac) using a linear gradient of 10-40% CH_3CN (v/v) in H_2O over 20 min, at a flow rate of 3 ml/min. The mobile phase contained 0.1% (v/v) formic acid to improve resolution. The separation was monitored at 210 nm and fractions containing the compound were collected, consolidated, dried under vacuum, and purified a final time using an analytical C18 column (5 µm, 250 x 4.6 mm, 100 Å, Inertsil ODS-4 column) using a 40 min linear gradient of 30-40% CH_3CN (v/v) in H_2O, again containing 0.1% formic acid, and a flow rate of 1 ml/min.

Preparation of chemical standard. The four-step synthetic scheme of compound **1** is presented in **Figure A-4**, based on a reported peptide synthesis. All reagents were obtained from MilliporeSigma (St. Louis, MO) unless otherwise specified. All steps were carried out at room temperature. After each step, the purified products were detected using an Agilent Technologies 6120 Quadrupole LC-MS (with DAD) instrument fitted with an Agilent Eclipse Plus C18 column (4.6 × 100 mm) and a 12 min gradient of 2-98% CH_3CN in H_2O with 0.1% formic acid.

The initial peptide coupling step was carried out with 5.5 mmol of **2** and 1 molar equivalent of **3** in 250 ml dichloromethane. Equimolar hydroxybenzotriazole (HOBt) and *N*,*N*'-dicyclohexylcarbodiimide (DCC) were added and the reaction stirred for 4 h. The solvent was removed under vacuum in a rotary evaporator and the reaction was separated using silica flash column chromatography (60 Å, 220-440 mesh) using an ethyl acetate and hexanes gradient. Thin layer chromatography was used to monitor the major peaks, and representative fractions were assessed for purity with LC-MS. The product was identified, and pure fractions were consolidated. Deprotection of the *N*-terminal L-Val was carried out in 2.5 ml trifluoroacetic acid (TFA) for 1 h. The solvent was removed under vacuum and the product **5** was washed over a filter disc with diethyl ether. The *N*-butyryl head group was attached by the dropwise addition of **6** to a 6.7 ml volume of tetrahydrofuran (THF) containing **5** and 1.1 ml of 2 M *N*,*N*-diisopropylethylamine (DIPEA) dissolved in *N*-methyl-2-pyrrolidone.[232] The reaction was dried under vacuum, dissolved in dichloromethane, and **7** was obtained, again using flash chromatography. C-terminal deprotection of **7** was carried out in 4:4:1 THF:diethyl ether:methanol using $LiBH_4$.[233] The reaction was quenched with dilute acetic acid in H_2O, extracted 1:1 in dichloromethane, and the product **1** was purified on an Agilent 1260 HPLC instrument with DAD, using a Hichrom Ltd (Theale, UK) Alltima C18 column (5 µm, 10×150 mm) and an isocratic eluent of 30% acetonitrile in H_2O, with

0.03% TFA. The overall yield was 0.7%. Final yield from two rounds of synthesis was 1.5 mg product.

Other phenotype assays. To assess differences in swimming motility, colonies of *Csa* wild-type and Δ*nrps3* were picked by pipette tips and stabbed into the center of fresh soft agar (0.5% rather than 1.5%) 2xYTG plates. Plates were incubated anaerobically for 48 h at 37°C before imaging.

Granulose accumulation assays were performed via iodine staining.[234] Stationary phase cultures (OD_{600} ~1-2) liquid cultures in 2xYTG were spotted in 10 µl droplets onto 2xYTG containing extra glucose (50 g/liter) to induce granulose accumulation. Plates were incubated for 2 days at 37°C before staining by exposing the agar to a bed of iodine crystals for 10 min. The plates were destained for 10 min before imaging.

Cathepsin B inhibition assays. In vitro protease activity was assayed using a method modified from Guo *et al*.[46] Cathepsin B and a chromogenic substrate, Z-Arg-Arg-pNA (both obtained from Sigma) were assayed in a reaction buffer containing 50 mM MES pH 5.5 and 5 mM DTT. Reaction mixture containing 47 µl buffer, 2 µl cathepsin B stock solution (5 µg/ml), and 1 µl of 1 in DMSO solution (0-3.5 mM) were incubated on ice for 20 min to allow any inhibiting interactions to occur. The reaction was initiated by addition of 1 µl Z-Arg-Arg-pNA stock solution (10 mM), and incubated at 37°C for 20 min. Reactions were quenched by addition of 85 µl quenching buffer (100 mM sodium monochloroacetate; 30 mM sodium acetate, 70 mM acetic acid). Reactions were performed in triplicate. 100 µl samples were quantified at 405 nm.

Table A-1. Bioinformatic summary o fBGCs in *Csa*

BGC	Predicted Amino Acid Monomers	Putative Non-Assembly-line Biosynthetic Genes	Putative homologs detected in:
nrps1	Leu-Leu	aspartate kinase, type I glutamate-ammonia lyase, glutamate synthase subunits, CPBP family intramembrane metalloprotease	*Clostridium* sp. DL-VIII
nrps2	Gly	4-phosphopantetheinyl transferase, radical SAM protein, a type-2 serine-tRNA ligase, an acyl-CoA dehydrogenase, the subunits of a FixB family electron transfer flavoprotein, NlpC/p60-like peptidase	none
nrps3	Leu-Leu	alpha/beta hydrolase, alcohol dehydrogenase, cysteine hydrolase	several ABE producers (see Fig. S4)
nrps4	Ser	Asp/Glu racemase superfamily protein, a M20/M25/M40 family metallo-hydrolase, a bifunctional acetaldehyde-CoA/alcohol dehydrogenase, GNAT-family N-acetyltransferase	none
hyb1	Ser	FMN-binding protein, an aminotransferase, iron-containing alcohol dehydrogenase	many Firmicutes (see Fig. S2)
hyb2	Val	none	*Paenibacillus* sp. KS1 (see Fig. S3)
hyb3	Gly	none	none

Table A-2. BGC expression profiles of *Csa*. Expression of the housekeeping gene *gyrB* is presented as an expression benchmark. Rows representing core biosynthetic genes are bolded

locus_tag	mean fpkm	standard deviation	BGC	putative gene product
CSPA_RS00030	267.6329	14.11877	housekeeping	gyrB
CSPA_RS21500	210.1266	54.37009	nrps1	hypothetical protein
CSPA_RS21505	88.58421	19.51843	nrps1	CPBP family intramembrane metalloprotease
CSPA_RS21510	359.5554	10.8462	nrps1	glutamate synthase subunit beta
CSPA_RS21515	3087.186	25.7374	nrps1	glutamate synthase large subunit
CSPA_RS21520	201.1556	11.06056	nrps1	ANTAR domain-containing protein
CSPA_RS21525	965.2242	28.55944	nrps1	type I glutamate-ammonia ligase
CSPA_RS21530	467.5652	18.87791	nrps1	aspartate kinase
CSPA_RS21535	**12.61314**	**1.772826**	**nrps1**	**C-A-T-E-C-A-T-TE**
CSPA_RS21540	30.51192	6.200816	nrps1	ABC transporter permease
CSPA_RS21545	5.238758	0.445454	nrps1	ABC transporter ATP-binding protein
CSPA_RS21550	6.314486	1.219476	nrps1	HlyD family efflux transporter periplasmic adaptor subunit
CSPA_RS01490	190.2536	25.06424	nrps2	AraC family transcriptional regulator
CSPA_RS01495	**2.210576**	**0.81502**	**nrps2**	**T**
CSPA_RS01500	4.739398	0.83288	nrps2	radical SAM protein
CSPA_RS01505	4.022174	0.934397	nrps2	type-2 serine--tRNA ligase SerS
CSPA_RS01510	0.597263	0.140008	nrps2	acyl-CoA dehydrogenase AcdA
CSPA_RS01515	0.405062	0.183929	nrps2	electron transfer flavoprotein subunit beta/FixA family protein
CSPA_RS01520	0.885385	0.26691	nrps2	electron transfer flavoprotein subunit alpha/FixB family protein
CSPA_RS01525	0.913504	0.232282	nrps2	ABC transporter permease
CSPA_RS01530	0.654118	0.127537	nrps2	hypothetical
CSPA_RS01535	3.059525	0.703062	nrps2	ATP-binding cassette domain-containing protein
CSPA_RS01540	2.055899	0.861045	nrps2	NlpC/p60-like peptidase
CSPA_RS01545	**3.505414**	**0.428673**	**nrps2**	**A-T-C**
CSPA_RS01550	2.435709	0.381814	nrps2	hypothetical
CSPA_RS01555	**1.919357**	**0.088705**	**nrps2**	**CAL**
CSPA_RS01560	**0.322086**	**0.202006**	**nrps2**	**PPT**
CSPA_RS01565	0.914014	0.14631	nrps2	hypothetical
CSPA_RS01570	3.892353	0.932641	nrps2	hypothetical protein
CSPA_RS01575	3.10295	0.61672	nrps2	hypothetical

Table A-2 (continued)**.**

locus_tag	mean fpkm	standard deviation	BGC	putative gene product
CSPA_RS10735	9.085743	1.706978	nrps3	alpha/beta hydrolase
CSPA_RS10740	34.71271	5.042884	nrps3	hemerythrin
CSPA_RS10745	4.540408	1.419	nrps3	hypothetical
CSPA_RS10750	407.3492	26.47701	nrps3	sigma-54-dependent Fis family transcriptional regulator
CSPA_RS10755	233.9419	9.247106	nrps3	2,3-butanediol dehydrogenase
CSPA_RS10760	**229.4108**	**22.99488**	**nrps3**	**C-A-T-C-A-T-R**
CSPA_RS10765	18.5522	4.16586	nrps3	methyl-accepting chemotaxis protein
CSPA_RS10770	27.52749	5.902638	nrps3	carboxylate/amino acid/amine transporter
CSPA_RS10775	40.13739	6.735017	nrps3	MFS transporter
CSPA_RS10780	45.41252	8.320144	nrps3	MarR family transcriptional regulator
CSPA_RS10785	17.96741	4.295844	nrps3	cysteine hydrolase
CSPA_RS10790	7.764638	1.200729	nrps3	Lrp/AsnC family transcriptional regulator
CSPA_RS10795	17.75605	4.069637	nrps3	hypothetical protein
CSPA_RS29230	17.61954	2.426804	nrps4	hypothetical protein
CSPA_RS18005	68.18586	12.94828	nrps4	GNAT family *N*-acetyltransferase
CSPA_RS18010	72.03751	13.09406	nrps4	YbhB/YbcL family Raf kinase inhibitor-like protein
CSPA_RS18015	234.1632	21.90018	nrps4	transcriptional regulator
CSPA_RS18020	98.35442	10.30323	nrps4	M20/M25/M40 family metallo-hydrolase
CSPA_RS18025	8.2804	1.59987	nrps4	amino acid racemase
CSPA_RS18030	19.04656	3.359851	nrps4	hypothetical protein
CSPA_RS18035	**258.6808**	**20.38416**	**nrps4**	**C-A-T-TE**
CSPA_RS18040	49.91332	5.094538	nrps4	hypothetical protein
CSPA_RS18045	753.2196	89.05311	nrps4	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
CSPA_RS14200	0.317376	0.13257	hyb1	hypothetical protein
CSPA_RS14205	11.11022	1.370519	hyb1	MFS transporter
CSPA_RS14210	14.30864	2.116373	hyb1	diguanylate cyclase
CSPA_RS14215	170.7348	21.18217	hyb1	methyl-accepting chemotaxis protein
CSPA_RS14220	**91.50774**	**23.37009**	**hyb1**	**TE**
CSPA_RS14225	**90.32083**	**15.70039**	**hyb1**	**PPT**
CSPA_RS14230	23.90792	5.302599	hyb1	helix-turn-helix domain-containing protein
CSPA_RS14235	**10.52487**	**0.613979**	**hyb1**	**T**
CSPA_RS14240	**93.28391**	**13.13009**	**hyb1**	**ER**
CSPA_RS14245	**211.1631**	**15.00449**	**hyb1**	**KS**

Table A-2 (continued)**.**

locus_tag	mean fpkm	standard deviation	BGC	putative gene product
CSPA_RS14250	12.91596	2.756376	hyb1	ACC biotin carrier protein
CSPA_RS14255	23.75766	1.497834	hyb1	ACC biotin carrier subunit
CSPA_RS14260	23.43882	3.519317	hyb1	ACC biotin carboxylase carboxyltransferase beta
CSPA_RS14265	24.41017	2.848952	hyb1	ACC biotin carboxylase carboxyltransferase alpha
CSPA_RS14270	**112.892**	**5.220678**	**hyb1**	**NRPS**
CSPA_RS14275	21.14646	2.006801	hyb1	MFS transporter
CSPA_RS14280	**236.2568**	**7.524213**	**hyb1**	**type I PKS**
CSPA_RS14285	14.23115	1.540054	hyb1	FMN-binding protein
CSPA_RS14290	7.47785	1.586153	hyb1	hypothetical protein
CSPA_RS14295	14.04237	2.392217	hyb1	LuxR family transcriptional regulator
CSPA_RS14300	2.6376	0.708296	hyb1	aminotransferase
CSPA_RS14305	0.227081	0.0918	hyb1	iron-containing alcohol dehydrogenase
CSPA_RS14310	40.72599	7.596511	hyb1	MerR family transcriptional regulator
CSPA_RS14315	7.167298	0.402726	hyb1	MFS transporter
CSPA_RS01865	4.901949	1.238258	hyb2	ABC transporter permease
CSPA_RS01870	51.3486	7.196234	hyb2	ABC transporter ATP-binding protein
CSPA_RS01875	18.34057	3.632487	hyb2	HlyD family efflux transporter periplasmic adaptor subunit
CSPA_RS01880	**4.645637**	**1.000377**	**hyb2**	**cMT-T**
CSPA_RS01885	**2.750513**	**0.306148**	**hyb2**	**ER**
CSPA_RS01890	**0.852303**	**0.156712**	**hyb2**	**TE**
CSPA_RS01895	**9.224148**	**1.350089**	**hyb2**	**A-T-C**
CSPA_RS01900	**23.07977**	**1.979215**	**hyb2**	**A-KR-T-TD**
CSPA_RS01905	**0.202625**	**0.083326**	**hyb2**	**T**
CSPA_RS01910	**3.101142**	**0.169807**	**hyb2**	**KS-AT-T**
CSPA_RS01915	**85.7177**	**10.50383**	**hyb2**	**A-KR-T**
CSPA_RS01920	**5.098486**	**0.426529**	**hyb2**	**C-T**
CSPA_RS01925	**11.85087**	**0.890903**	**hyb2**	**KS-T-C**
CSPA_RS01930	**6.229069**	**1.288215**	**hyb2**	**PPT**
CSPA_RS01840	259.6133	15.31642	hyb3	MATE family efflux transporter
CSPA_RS01845	**39.90932**	**1.201051**	**hyb3**	**CAL-KS-DOC-ACP-KS-DOC-KR-T-KS-DOC-ACP**
CSPA_RS01850	**56.23673**	**4.056765**	**hyb3**	**GNAT-KS-DOC-KR-ACP-KS-DOC-ACP**
CSPA_RS01855	**13.77323**	**0.861658**	**hyb3**	**C-A-T-KS-DH-KR-ACP-ACP-KS-DOC-TE**
CSPA_RS01860	**3.070295**	**0.175708**	**hyb3**	**AT**

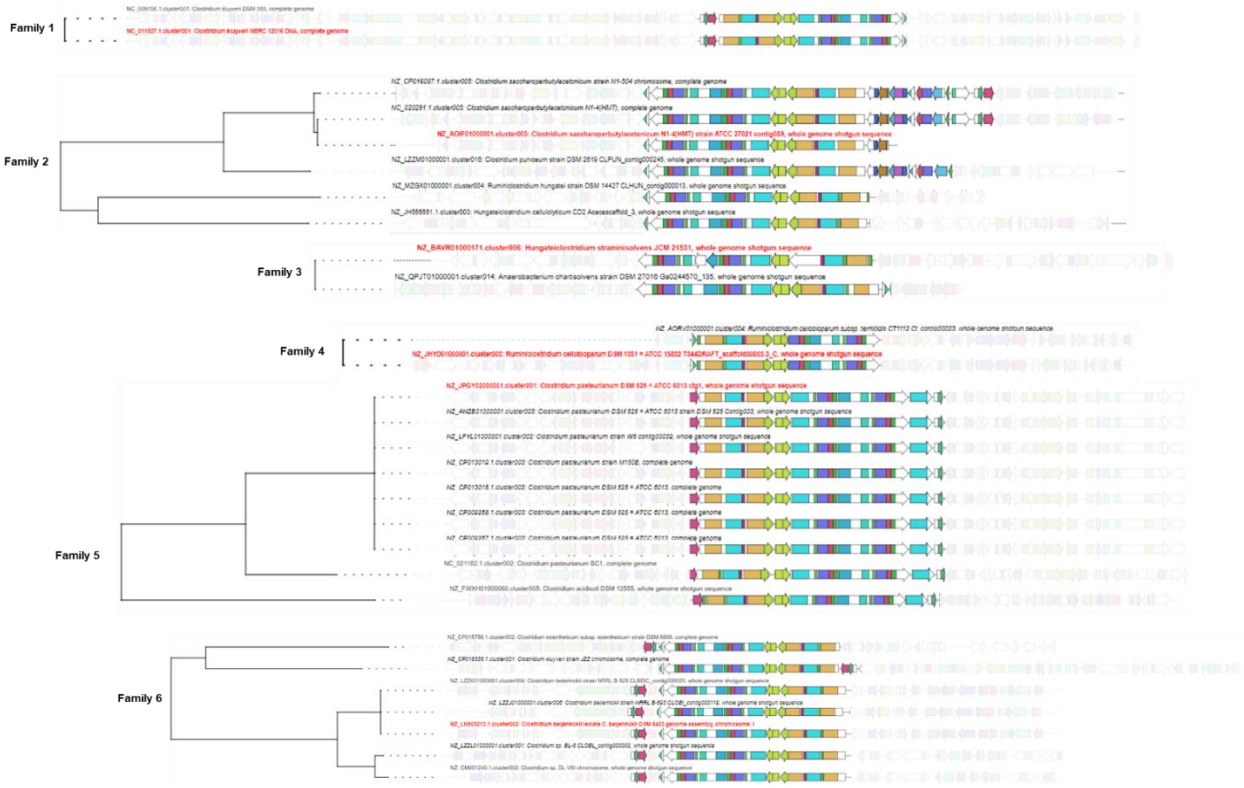


Figure A-1. Gene cluster families of *hyb1* homologs from clostridia. One genome was selected for each species-level classification. Families were identified using BiG-SCAPE.[94] Equivalent pfam predictions are colored the same. The genes encoding the conserved biosynthetic pfams are highlighted to denote the predicted limits of the BGC. These include the two hybrid PKS/NRPS genes with a nested transporter gene, and flanking genes encoding a FMN-binding protein, thioesterase, and 4'-phosphopantetheinyl transferase superfamily protein. The closest homologous cluster to *hyb1* outside of *C. saccharoperbutylacetonicum* is found in *C. puniceum*. These members of Family 2 include a biosynthetic operon predicted to generate malonyl-CoA, one of the expected extender units for the BGC. The Family 5 BGCs bear an extra acyl-CoA ligase domain

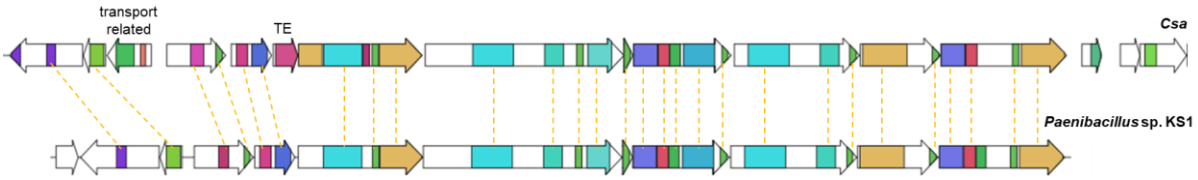


Figure A-2. Homolog of *hyb2* identified in *Paenibacillus* sp. KS1. Yellow dotted lines denote homologous pfam domains. Two genes in the cluster are unique to *Csa*; CSPA_RS01875 encodes a putative HlyD family efflux transporter periplasmic adaptor subunit, and CSPA_RS01890 encodes a putative TE

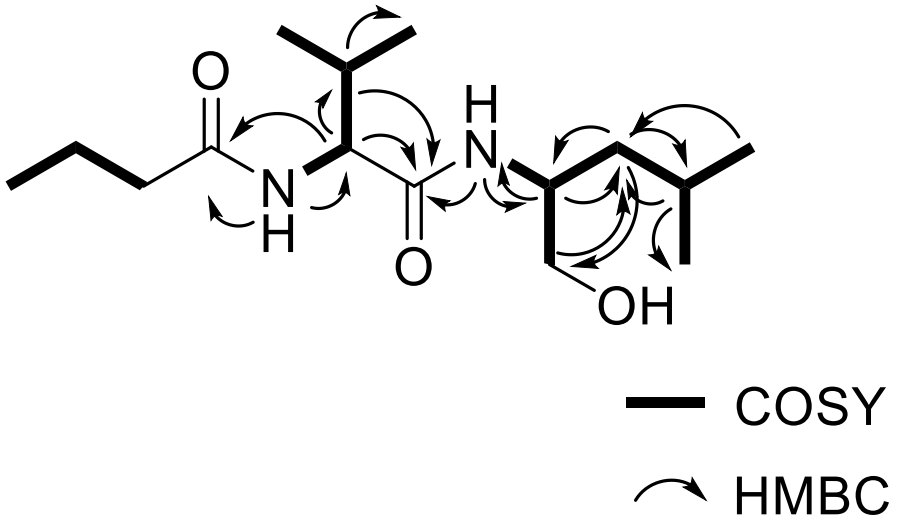


Figure A-3. NMR and optical rotation characterization of compound 1. a $^{1}H,^{1}H$-COSY and selected $^{1}H,^{13}C$-HMBC correlations of compound 1.

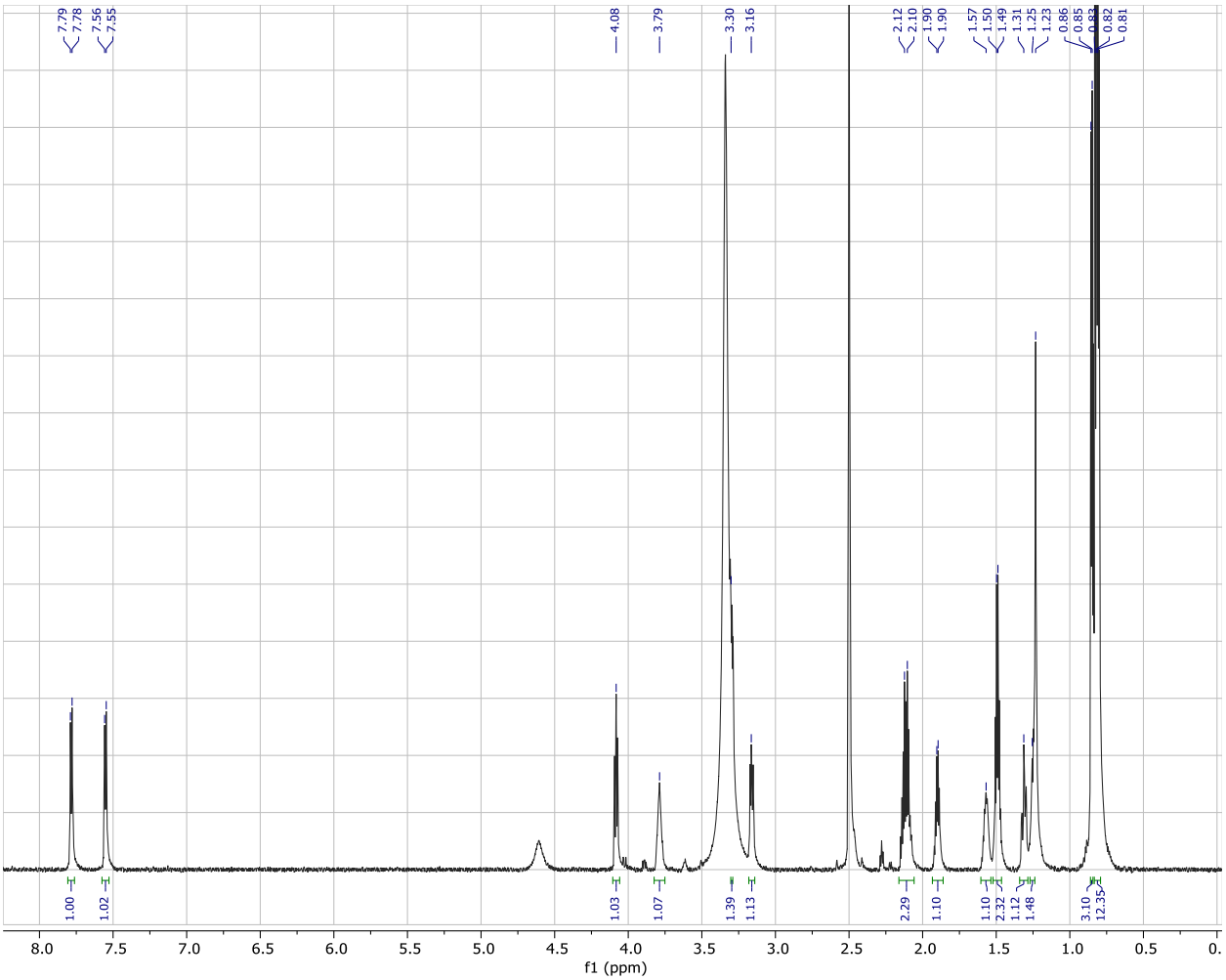


Figure A-3 (continued)**.** ^{1}H NMR spectrum (DMSO-d_6, 800 MHz) of compound **1**

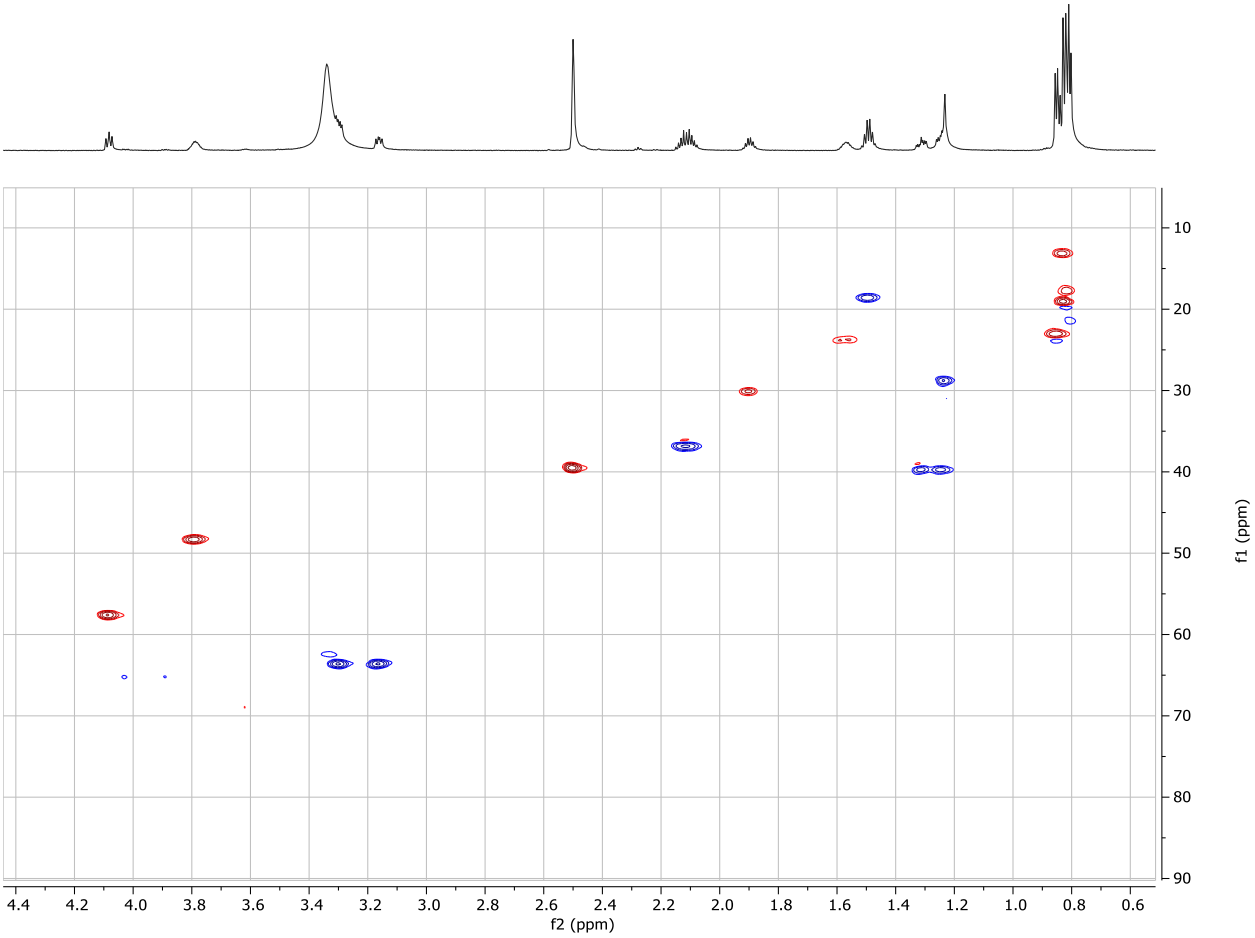


Figure A-3 (continued)**.** HSQC spectrum (DMSO-d_6, 800 MHz) of compound **1**

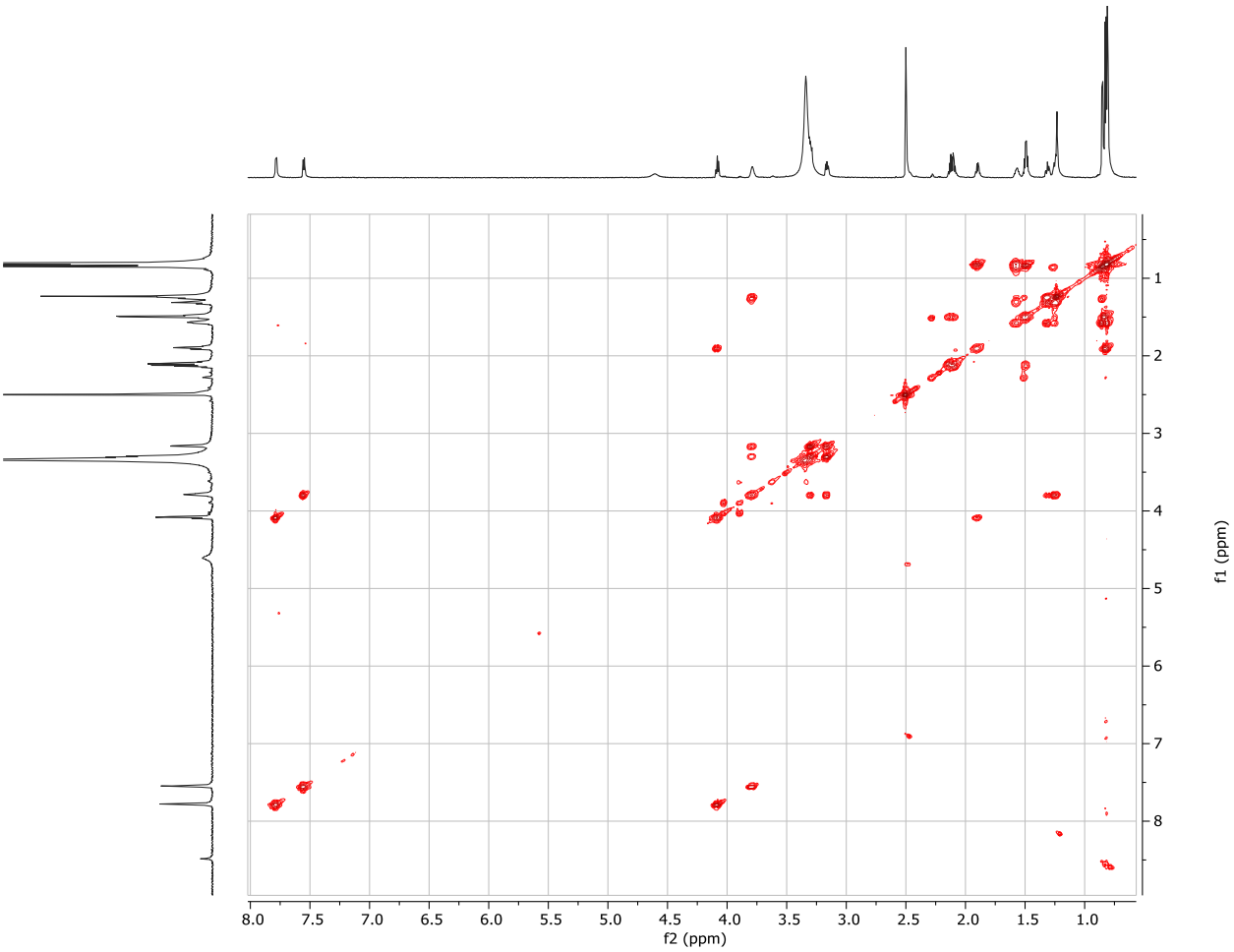


Figure A-3 (continued). ^{1}H,^{1}H-COSY spectrum (DMSO-d_6, 800 MHz) of compound **1**

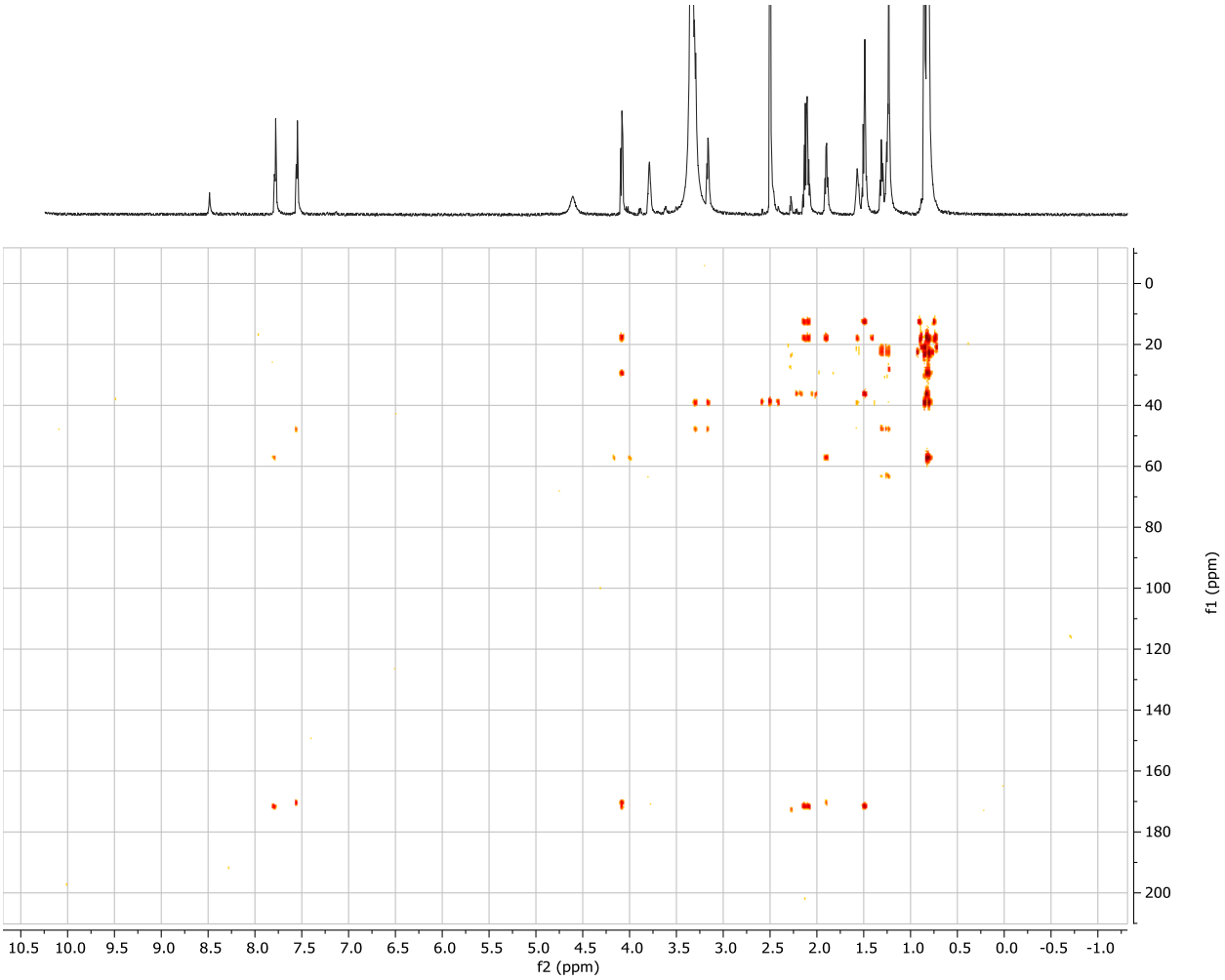


Figure A-3 (continued)**.** HMBC spectrum (DMSO-d_6, 800 MHz) of compound **1**

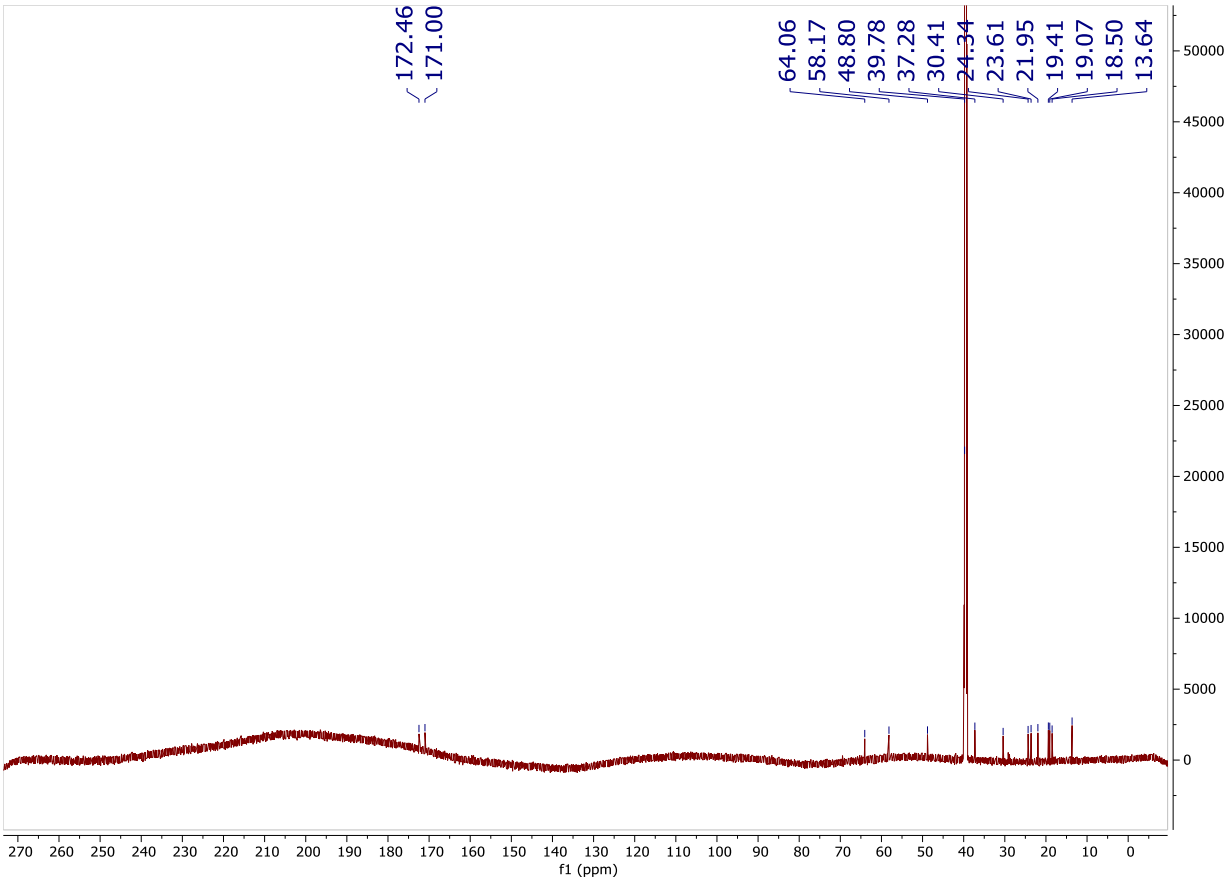
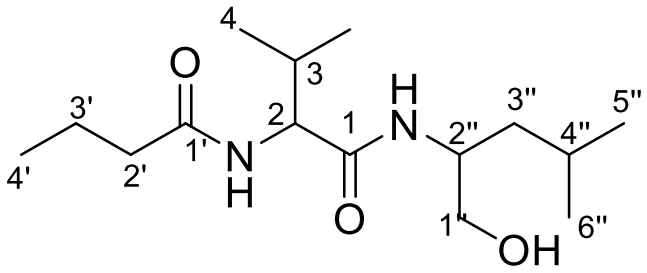


Figure A-3 (continued)**.** ^{13}C spectrum (DMSO-d_6, 900 MHz) of compound **1**



2-butyramido-*N*-(1-hydroxy-4-methylpentan-2-yl)-3-methylbutanamide

No.	δ_C	δ_H	Type	COSY	HMBC
1	170.32		C=O		
2	57.60	4.08, m	CH	2-NH, 3	1, 3, 1’
3	30.08	1.90, m	CH	2, 4	1, 4
4	18.20	0.82, m (overlap)	CH_3	3	2, 3
1-NH		7.55, d (8.8)	NH	2’’	1, 2’’
2-NH		7.78, d (9.5)	NH	1’	2, 1’
1’	171.92		C=O		
2’	36.84	2.11, m	CH_2	3’	1’, 3’, 4’
3’	18.59	1.49, m	CH_2	2’, 4’	2’, 4’
4’	13.16	0.82, m (overlap)	CH_3	3’	2’, 3’
1’’	63.31	3.16, dd (10.2, 6.5) 3.30, dd (10.2, 4.9)	CH_2	2’’	2’’, 3’’
2’’	48.32	3.79, m	CH	1’’, 3’’	3’’
3’’	39.71	1.25, m 1.31, m	CH_2	2’’, 4’’	1’’, 2’’, 4’’
4’’	23.74	1.57, m	CH	3’’, 5’’, 6’’	3’’, 6’’
5’’	19.04	0.82, m (overlap)	CH_3	4’’	3’’
6’’	23.03	0.85, d (6.7)	CH_3	4’’	3’’, 4’’

Figure A-3 (continued)**.** ^{1}H (800 MHz), ^{1}H,^{1}H-COSY and HMBC NMR data for compound **1**

$[\alpha]_{D}^{20} = -20.0°$

Figure A-3 (continued)**.** Optical rotation for **1**, in MeOH

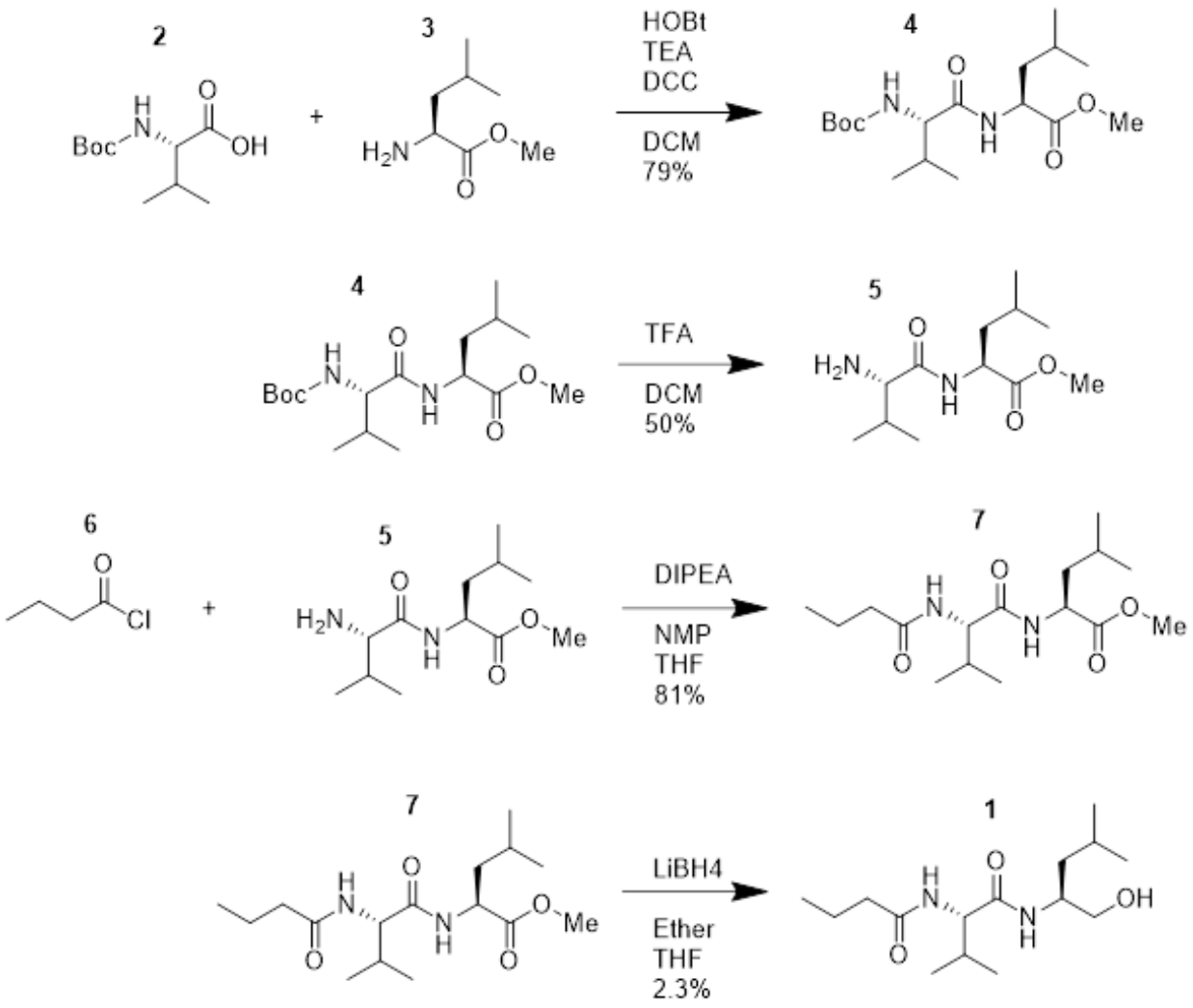


Figure A-4. Chemical synthesis of compound **1**

Appendix B.

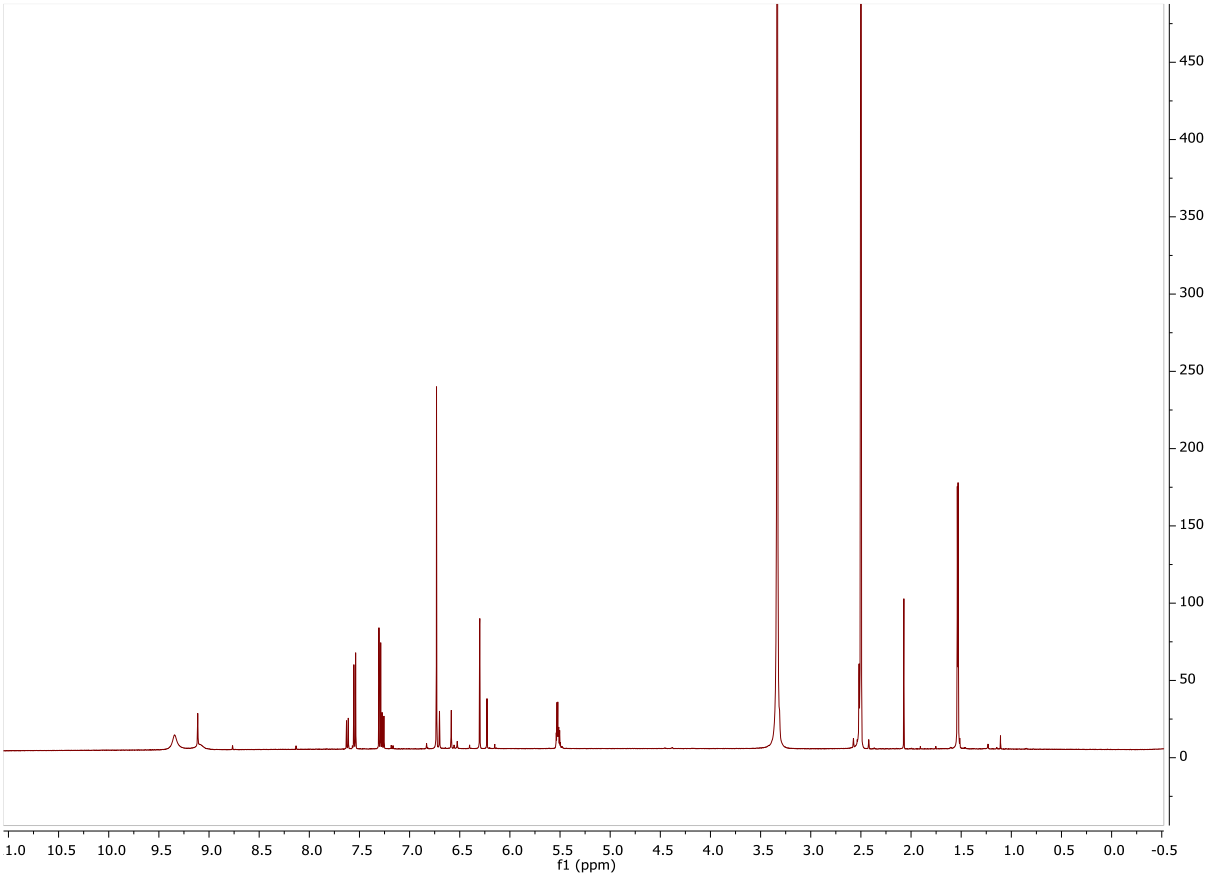


Figure B-1. ^{1}H NMR spectrum of compounds **1**, **2** in DMSO-d_6. Continued on the following pages

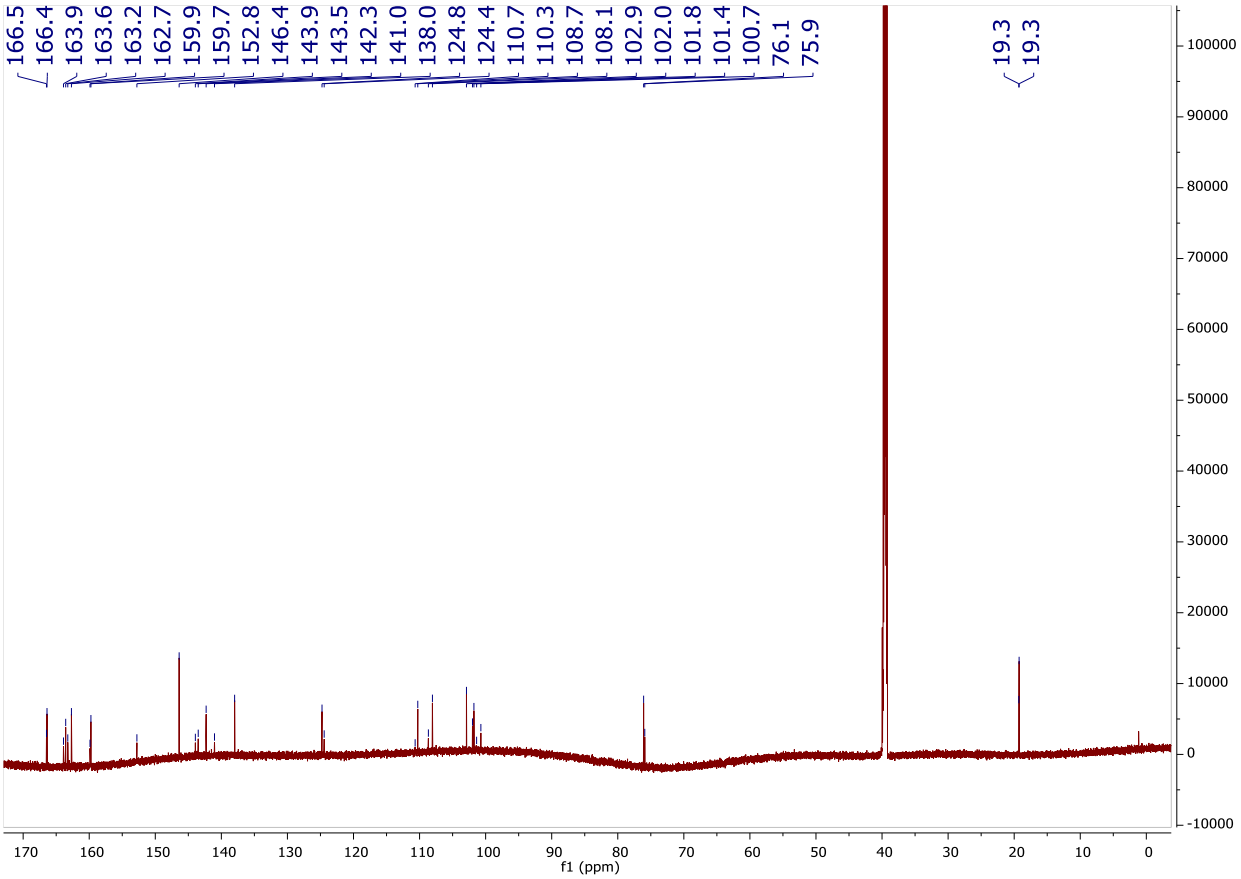


Figure B-1 (continued)**.** ^{13}C NMR spectrum of compounds **1**, **2** in DMSO-d_6

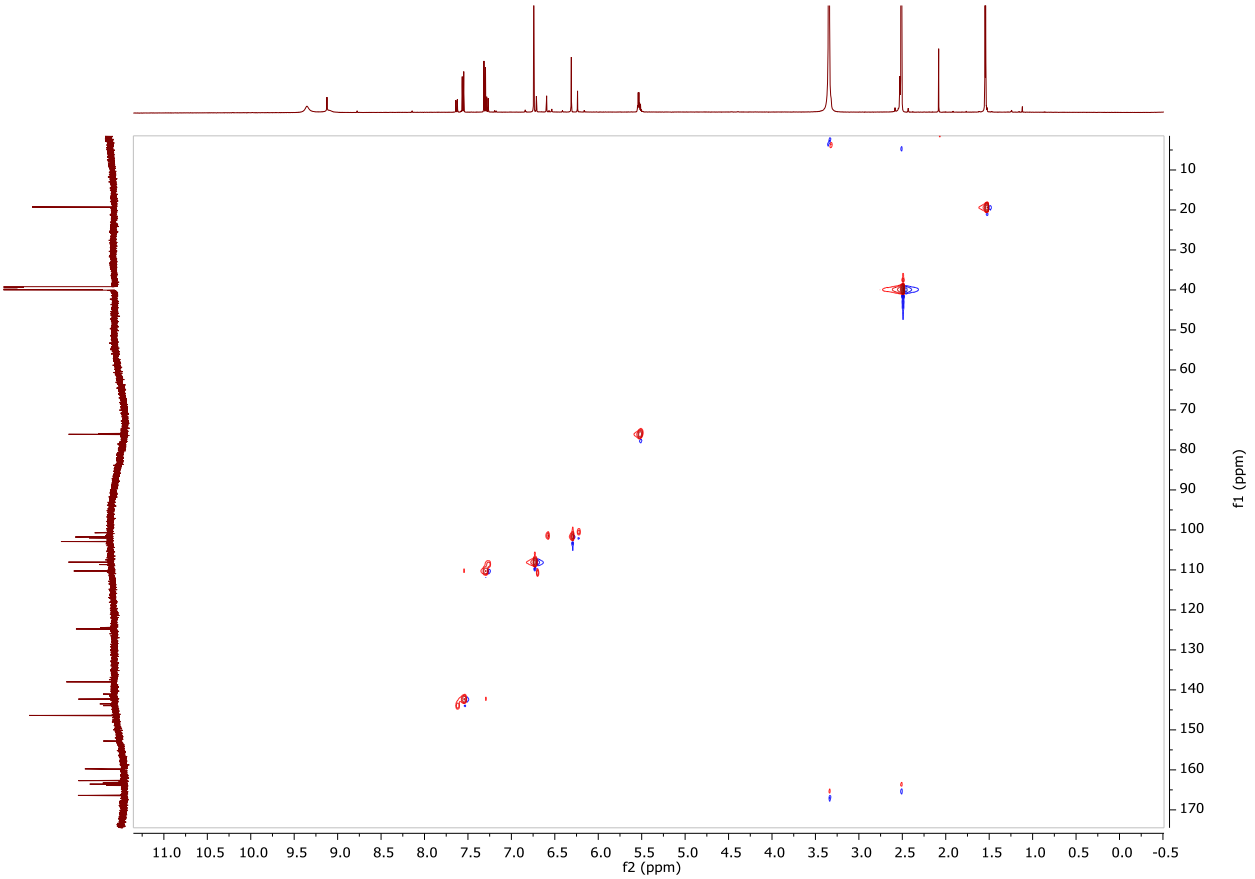


Figure B-1 (continued)**.** gHSQC spectrum of compounds **1**, **2** in DMSO-d_6

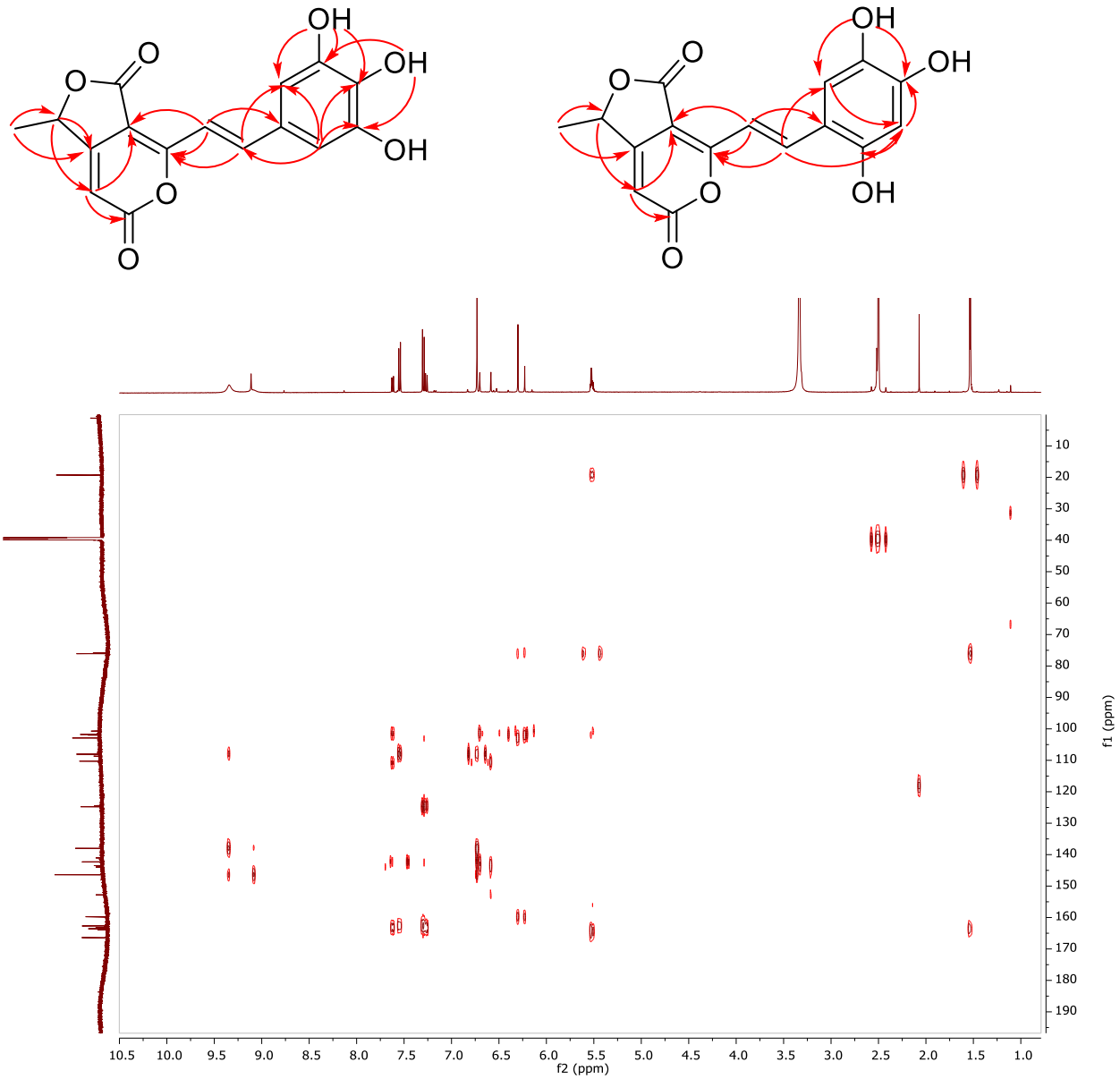


Figure B-1 (continued)**.** gHMBC correlations of compounds **1**, **2** in DMSO-d_6

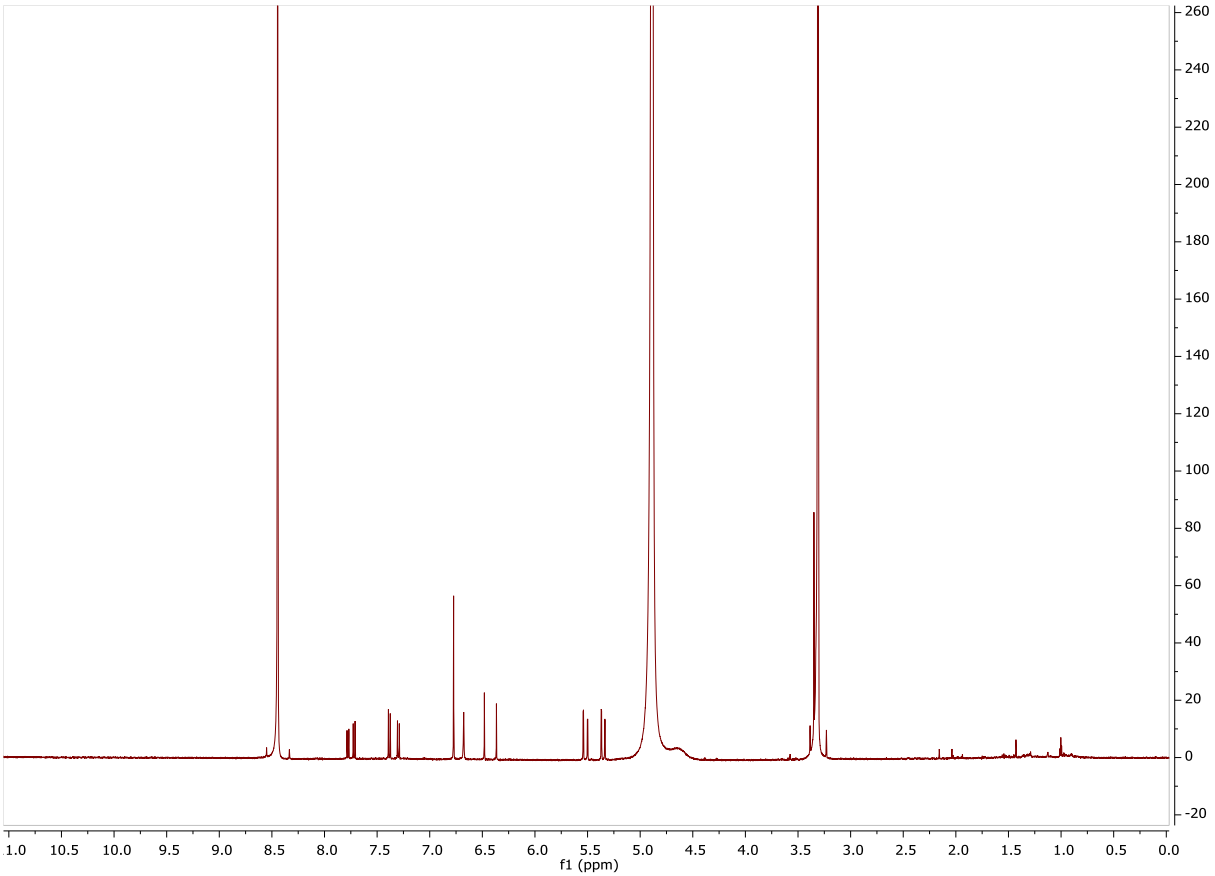


Figure B-1 (continued)**.** ^{1}H NMR spectrum of compounds **3**, **4** in CD_3OD

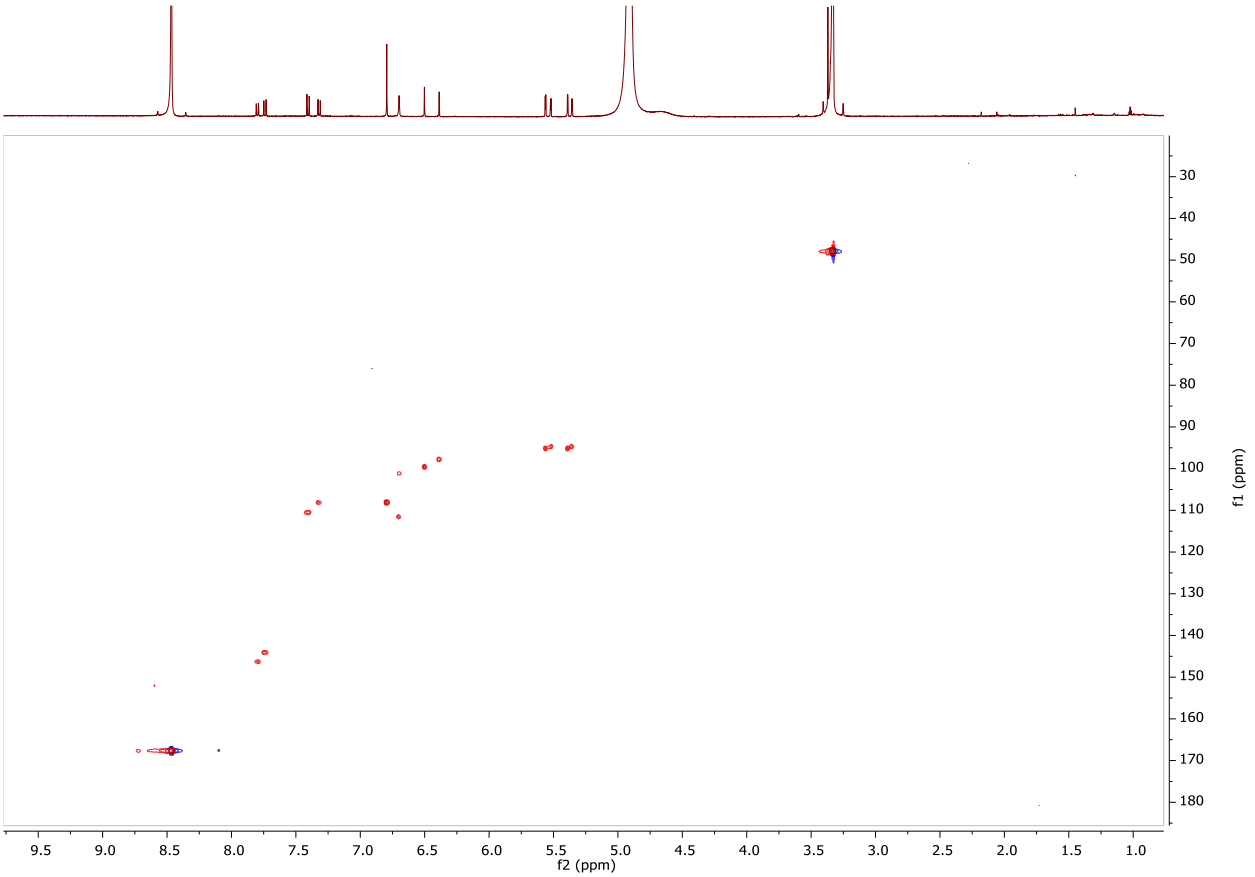


Figure B-1 (continued)**.** gHSQC spectrum of compounds **3**, **4** in CD_3OD

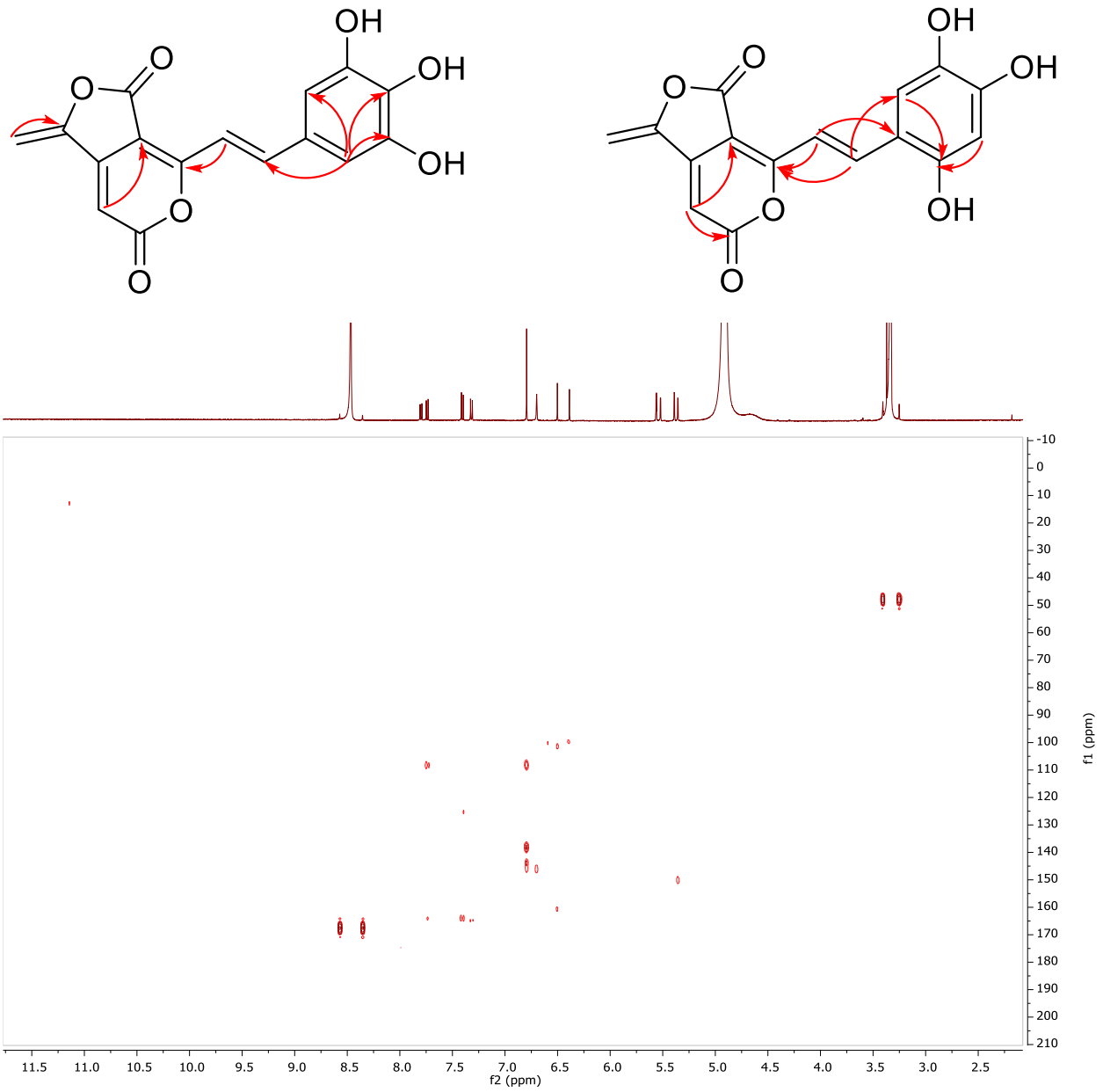


Figure B-1 (continued)**.** gHMBC correlations of compounds **3**, **4** in CD_3OD