Published in final edited form as:

Org Lett. 2017 October 06; 19(19): 5010–5013. doi:10.1021/acs.orglett.7b01611.

Activation of a cryptic gene cluster in *Lysobacter enzymogenes* reveals a module/domain portable mechanism of nonribosomal peptide synthetases in the biosynthesis of pyrrolopyrazines

Shanren Li^{†,‡}, Xiuli Wu^{‡,§}, Limei Zhang^{||}, Yuemao Shen^{†,*}, and Liangcheng Du^{†,‡,*}

[†]State Key Laboratory of Microbial Technology, College of Life Sciences, Shandong University, Jinan 250100, China

[‡]Departments of Chemistry, University of Nebraska-Lincoln, NE 68588, USA

[§]College of Pharmacy, Ningxia Medical University, Yinchuan 750004, China

^{||}Departments of Biochemistry, University of Nebraska-Lincoln, NE 68588, USA

Abstract

Lysobacter are considered “peptide specialists”. However, many of the nonribosomal peptide synthetase genes are silent. Three new compounds were identified from *L. enzymogenes* upon activating the six-module-containing *led* cluster by the strong promoter P_{HSAF}. Although *ledD* was the first gene under P_{HSAF} control, the second gene *ledE* was expressed the highest. Targeted gene inactivation showed that the two-module *LedE* and the one-module *LedF* were selectively used in pyrrolopyrazine biosynthesis, revealing a module/domain portable mechanism.

Graphical Abstract

Lysobacter is a genus of Gram-negative gliding bacteria that are emerged as a new source of bioactive natural products.^{1–3} Previous analyses of several *Lysobacter* genomes revealed that the species are “peptide production specialists” because of the predominance of gene clusters encoding nonribosomal peptide synthetases (NRPS).³ However, many of the NRPS genes are silent under standard laboratory conditions. Herein, we report application of the

*Corresponding Author: ldu3@unl.edu; yshen@sdu.edu.cn.

Notes

The authors declare no competing financial interest.

Supporting Information

Details of experimental procedures, construction of plasmids for promoter replacement, gene deletion, and point mutation, generation and verification of mutants, qRT-PCR, metabolite production and isolation, primer list, gene cluster organization, sequence analysis, and spectroscopic data (PDF). The Supporting Information is available free of charge on the ACS Publications website.

promoter replacement strategy to activate a cryptic gene cluster (*led*) that contains three NRPS genes (*ledD*, *ledE*, and *ledF*) coding for six NRPS modules in *L. enzymogenes* OH11. The results showed that new pyrrolopyrazines were synthesized by the NRPS using a “module/domain portable mechanism”.

L. enzymogenes is known to produce two types of antibiotics, the polycyclic tetramate macrolactam HSAF and the cyclic lipodepsipeptide WAP-8294A2.^{4, 5} HSAF is biosynthesized by a single-module polyketide synthases (PKS)-NRPS, and WAP-8294A2 is by a 12-module NRPS.^{6–8} Bioinformatics analysis of the draft genome sequence of *L. enzymogenes* OH11 revealed nine gene clusters coding for NRPS, PKS or hybrid PKS/NRPS. Most of these cryptic clusters have no significant similarity to the characterized pathways, indicating that there might still be novel natural products remained to be discovered in *L. enzymogenes* OH11. Consequently, we spent much effort to identify the putative products of these cryptic gene clusters through manipulation of regulatory and biosynthetic genes, test of potential stimuli for biosynthesis, and comparative metabolic analyses of cultures under various growth conditions. However, none of the effort panned out the products of the cryptic gene clusters.⁹ Finally, we took the promoter replacement strategy to activate the *led* gene cluster (Genbank accession number MF495862), which is located at a 25 kb locus that hosts 11 open reading frames, including three NRPS genes (*ledD*, *ledE* and *ledF*) (Figure 1A, Table S3). The NRPS genes are transcribed in the same direction and appear to belong to the same operon. Thus, we figured that insertion of a strong constitutive *Lysobacter* promoter in front of the first NRPS gene, *ledD*, could stimulate expression of all NRPS genes even at laboratory conditions (Figure 1B).

Analysis of the amino acid sequence of LedD revealed three NRPS modules. The first module consists of two domains, adenylation (A₁) and thiolation (T₁); the second module contains three domains, condensation (C₂), A₂ and T₂; the third module has four domains, C₃, A₃, T₃, and reductase (R₁). A₁, A₂, A₃ domains were predicted to activate L-Phe, L-Pro and L-Phe, respectively (Table S4). LedE consists of two modules, with a domain organization of A₄-T₄ for the first module and C₅-A₅-T₅ for the second module. A₄ and A₅ were predicted to activate L-Pro and L-Val/Leu/Ile, respectively. LedF only contains one module containing C₆-A₆-T₆-R₂ domains, and A₆ was predicted to activate L-Pro (Table S4). Notably, there are two R domains embedded at the C-terminus of both LedD and LedF. The two R domains show sequence similarities to the short-chain dehydrogenase/reductase (SDR) superfamily (Figure S1), which has the Rossmann-fold structure and typically catalyzes the NAD(P)H-dependent reductive release of the acyl chain as an aldehyde.¹⁰ Overall, the *led* cluster has no significant sequence similarity to any known NRPS gene clusters.

Based on the NRPS analysis, the predicted products of the *led* cluster would be a hexapeptide. However, no obvious product was detected or isolated under our culture conditions. We used strain HW as a “clean” starting strain, because the biosynthetic genes for both HSAF and WAP-8294A2 were inactivated in strain HW.¹¹ To test if any of the NRPS genes was expressed in the strains, we conducted quantitative reverse transcription PCR (qRT-PCR). The results showed that expression of the three NRPS genes was barely detectable in the starting strain (Figure 1C). This result confirmed that the *led* gene cluster

was silent under the laboratory conditions. We then replaced the *ledD* native promoter with the constitutive promoter P_{HSAF} by homologous recombination. The resulting recombinant strain $\text{led}P_{\text{HSAF}}$ was verified by diagnostic PCR and DNA sequencing (Figure S2). qRT-PCR showed that all three NRPS genes were clearly expressed in strain $\text{led}P_{\text{HSAF}}$ (Figure 1C). P_{HSAF} is the promoter controlling expression of the PKS-NRPS cluster for HSAF biosynthesis in *L. enzymogenes* OH11. It is the strongest promoter known so far in this species, due to the constitutively high production of HSAF under laboratory conditions.¹² Interestingly, the effect of P_{HSAF} on expression of *ledD*, *ledE*, and *ledF* was not the same. The expression of *ledE* increased by 34 fold, *ledF* by 14 fold, and *ledD* by 7 fold, although *ledD* was the first gene under P_{HSAF} control. The results seem to suggest some kind of unusual selective promotion of the genes expression.

Next, we cultivated the promoter replaced strain $\text{led}P_{\text{HSAF}}$ and the starting strain HW in 10% TSB and analyzed the metabolite profile by HPLC (Figure 1D). Three new peaks (**1–3**) were clearly produced in strain $\text{led}P_{\text{HSAF}}$ (Figure 1D). The compounds were isolated, and their structures were elucidated by spectroscopic analysis (Figure 2, Table S5, Figures S13–S34).

The molecular formula of **1** was determined as $\text{C}_{10}\text{H}_{14}\text{ON}_2$ by the (+)-HREIMS (Figure S14–S15). The ^1H NMR spectrum of **1** displayed resonances attributable to two secondary methyl δ_{H} 1.23 (6H, $J = 7.0$ Hz, H_3 -11,12), three methylenes [δ_{H} 4.13 (t, $J = 7.7$ Hz, H_2 -9), δ_{H} 3.08 (t, $J = 7.7$ Hz, H_2 -7) and 2.22 (quint, $J = 7.7$ Hz, H_2 -8)], and a methine δ_{H} 3.48 (m, H -10), and a trisubstituted double bond δ_{H} 7.28 (s, H -1). The ^{13}C NMR spectrum of **1** gave 10 carbon resonances that were attributed to two sp^3 methyl, three sp^3 methylene, a sp^3 methine, and four sp^2 carbons. The HSQC spectroscopic data analysis of **1** furnished assignments of the proton-bearing carbon and corresponding proton resonances in the NMR spectra (Table S5). In the ^1H - ^1H gCOSY spectrum of **1**, homonuclear coupling correlations for H_2 -7/ H_2 -8/ H_2 -9, and H -10/ H_3 -11,12 indicated structural units containing vicinal coupling protons in **1** (Figures S13, S16–S19). In the HMBC spectrum of **1**, long-range heteronuclear correlations were observed for H_3 -11,12/ C -3, and C -10, H -1/ C -3, C -6 and C -7, H_2 -7/ C -1, C -6, C -8, and C -9, H -10/ C -3, C -4, C -11, and C -12 (Figure S20). In combination with the chemical shifts of these protons and carbons, the structure of **1** was established to be 1-benzoyloxy-3-isopropyl-7,8-dihydro-6*H*-pyrrolo[1,2- α]pyrazine-4-one, and **1** was named Le-pyrrolopyrazine A.

The (+)-HREIMS data indicted compound **2** ($\text{C}_{11}\text{H}_{16}\text{ON}_2$) had one more CH_2 group than **1** (Figures S21, S22). Comparison of the NMR data of **2** and **1** (Table S5) indicated that **2** had the same pyrrolo[1,2- α]pyrazine-4-one moiety, but the side chain of **2** was replaced by an isobutyl group. On the basis of assignments of the proton-bearing carbon and corresponding proton resonances (Table S5) in the NMR spectra by gHSQC, gCOSY cross-peaks, and HMBC correlations (Figures S13, S23–S27), the structure of **2** was established to be 1-benzoyloxy-3-isobutyl-7,8-dihydro-6*H*-pyrrolo[1,2- α]pyrazine-4-one, which is consistent with the literature,¹³ and **2** was named Le-pyrrolopyrazine B.

Compound **3** ($\text{C}_{11}\text{H}_{16}\text{ON}_2$) exhibited spectroscopic features very similar to that of **2** (Table S5, Figures S28, S29), suggesting that it was an isomer of **2**. Detailed comparison of the ^1H

NMR data between **3** and **2** indicated that the methyl group, at C-11 in **2**, was at C-10 in **3**, which was confirmed by 2D NMR (Figures S13, S30–S34). The planar structure of **3** was identified to be 1-benzoyloxy-3(2)-butyl-7,8-dihydro-6H-pyrrolo[1,2- α]pyrazine-4-one. The absolute configuration of **3** was elucidated with the ORD data. Compound **3** exhibited the same right direction as (*S*)-2-butanol, and **3** was deduced to be (+)-10*S*-1-benzoyloxy-3(2)-butyl-7,8-dihydro-6*H*-pyrrolo[1,2- α]pyrazine-4-one, and it was named Le-pyrrolopyrazine C.

The production of Le-pyrrolopyrazines from strain ledP_{HSAF} was unexpected, because there were six NRPS modules in this activated *led* gene cluster. Le-pyrrolopyrazines apparently resulted from condensation of two amino acids (proline and valine/leucine/isoleucine), which would require only two NRPS modules. One possibility was that Le-pyrrolopyrazines might not be direct products of the *led* gene cluster, but rather resulted from a different pathway that was stimulated by the activated *led* gene expression. To test this possibility, we first in-frame deleted *ledE* in strain ledP_{HSAF}, the most highly expressed gene in the promoter replaced strain (Figures S4). The resulted mutant completely lost the ability of producing Le-pyrrolopyrazines (Figure 3). The similar result was obtained when *ledF* was in-frame deleted (Figure 3 and S5). However, when *ledD* was deleted from strain ledP_{HSAF}, the mutant still produced a similar level of Le-pyrrolopyrazines as strain ledP_{HSAF} (Figure 3 and S3). Furthermore, the double deletion of both *ledD* and *ledE* eliminated Le-pyrrolopyrazine production in the mutant; the double deletion of *ledE* and *ledF* also led to no production of the compounds in the mutant (Figure 3, Figure S3–S7). The results showed that only two (*ledE* and *ledF*) of the three NRPS genes are required for Le-pyrrolopyrazines production, and the data also support a direct relationship between the *led* genes and Le-pyrrolopyrazines production.

ledE encodes two NRPS modules containing A₄-T₄-C₅-A₅-T₅, and NRPSpredictor2 predicted that A₄ activates proline and A₅ activates valine/leucine/isoleucine (Table S4). The substrate specificity of ledE matches to the composition of amino acids in Le-pyrrolopyrazines. However, A₆ of the single-module NRPS (C₆-A₆-T₆-R₂) encoded by *ledF* was also predicted to activate proline. To determine the source of the proline residue in Le-pyrrolopyrazines, we point-mutated an active site of A₄ and A₆, in the well-conserved motif GRxxxQVKIRGxRIELGEIE. This conserved motif was shown to be essential for adenylation; a mutation of the second G (equivalent to G455 in A₄ and G963 in A₆) in the motif to various residues led to loss of activity in the proline-activating domain of GrsB in the biosynthesis of gramicidin.^{14, 15} We therefore generated G455E mutant of A₄ and G963E mutant of A₆ (Figure S8, S11), and found the G455E mutant was abolished with ~90% production of Le-pyrrolopyrazines and G963E mutant abolished with the total production (Figure 3). The results show that ledF A₆ is required for proline incorporation in Le-pyrrolopyrazine biosynthesis, while ledE A₄ also contributes significantly to the proline incorporation. Furthermore, two C-domains, C₅ in ledE and C₆ in ledF, could contribute to the condensation between the amino acids (proline and valine/leucine/isoleucine). To clarify the role of these C domains, we point-mutated the active site of C₅ and C₆ in the well-conserved motif HHxxxDGxSW, in which the second histidine residue is indispensable for C-domain's activity. The C₅ mutant H785V lost ~60% production of Le-pyrrolopyrazines,

while C₆ mutant H205V lost all production of Le-pyrrolopyrazines (Figure 3, Figure S9, S10). The results indicate that ledF C₆ is required for the peptide bond formation between proline and valine/leucine/isoleucine and ledE C₅ also plays an important role. Together, among the five domains of LedE, A₅-T₅ domains are required for Le-pyrrolopyrazines, as they are responsible for incorporation of valine/leucine/isoleucine. A₄-T₄-C₅ domains are not absolutely required, but they play a significant role in the yield of Le-pyrrolopyrazines.

Among the four domains of LedF, C₆-A₆-T₆ domains are required for Le-pyrrolopyrazine production. ledF also hosts the second reductase (R₂) domain, which is obviously required, considering R domain's function in product release and the requirement for reduction steps in forming pyrazine ring. To obtain experimental evidence, we generated another point-mutant, Tyr1357Phe. The tyrosine residue in the SDR family's Ser-Tyr-Lys catalytic triad is crucial for the reductase activity. We mutated Tyr1357 within LedF R₂ to Phe and found Le-pyrrolopyrazine production was abolished from the mutant (Figure 3, Figure S12). The result demonstrates the involvement of the LedF reductase domain in the biosynthesis. This also suggests that R₁ in LedD could not functionally complement the loss of R₂ in LedF.

Taken together, the results from the mutants are consistent with the results from qRT-PCR. Le-pyrrolopyrazine production in strain ledP_{HSAF} results from the most highly expressed NRPS genes, *ledE* and *ledF*. The biosynthesis follows a module/domain portable mechanism as illustrated in Figure 4, which exemplifies the remarkable flexibility and versatility of the modular assembly line of NRPS. Interestingly, we did not observe compounds other than compounds **1**, **2**, and **3** under the experimental condition. It seems that the putative diketopiperazine (DKP) intermediates that resulted from condensations between proline and valine/leucine/isoleucine were quickly reduced by the R₂ domain, followed by a dehydration and dehydrogenation (probably non-enzymatically) to generate the more stable pyrazines (Figure 4).

PKS domain/module skipping in bacteria has been reported in the biosynthesis of several polyketide natural products.¹⁶ NRPS domain/module skipping is relatively rare. One example was reported in bleomycin biosynthesis, in which one of the A domains involved in bithiazole ring formation was skipped due to a natural mutation.¹⁷ Another example was seen in the biosynthesis of myxobacterial lipoptides.¹⁸ A mutation in one of the PCP domains caused a module-skipping in the NRPS complex for myxochromide S biosynthesis. Under the same constitutive promoter, the first NRPS was expressed at a much lower level than the second and third NRPS genes. The reason for this result is unclear, and one possibility might be related to the many regulators flanking the NRPS genes in the cluster (Table S3). Further studies are needed to address this interesting possibility. It should be pointed out that we can't exclude the possibility that these NRPS genes may not be coupled at the enzyme level to make the same group of products, although they are clustered together and appear to belong to the same operon. Both *ledD* and *ledE* start with an A domain, and both *ledD* and *ledF* end with an R domain. These features seem suggest that the first three NRPS modules in LedD and the last three modules in LedE and LedF may belong to separate NRPS complexes. The putative products, if any, resulted from coupling of both enzyme complexes may involve some other machineries.

Although biosynthetic gene cluster activation has been reported in many microorganisms, the *led* cluster represents the first example in *Lysobacter* species that has been activated to produce new natural products. Our work provides a useful approach to access the large number of cryptic NRPS gene clusters in *Lysobacter* genomes that remain essentially unexplored. Le-pyrrolopyrazines A and C are new compounds, and Le-pyrrolopyrazine B has not been reported as a natural product. Although cyclodipeptides such as 2,5-diketopiperazines are very common in natural products, the reduced pyrazine is rare. Pyrrolopyrazine is an important moiety in a number of drugs, such as the anti-anxiety compound CSP-2503,¹⁹ and the anti-amnesic agent unifiram (DM-232).²⁰ The Le-pyrrolopyrazines produced in this engineered strain of *Lysobacter* may provide a new way to access this useful structural moiety.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the NSFC (31329005), NIH (R01AI097260), Nebraska Research Initiative, Faculty Seed Grant, and Taishan Scholars Program of Shandong Province. X.W. was partly supported by China Scholarship Council. The authors are grateful to Dr. Guoliang Qian at Nanjing Agricultural University for the gift of HW strain, Drs. Martha Morton, Ronald Cerny and Javier Seravalli at University of Nebraska-Lincoln for technical assistance.

References

1. Xie Y, Wright S, Shen Y, Du L. Nat Prod Rep. 2012; 29:1277–1287. [PubMed: 22898908]
2. Pidot SJ, Coyne S, Kloss F, Hertweck C. Int J Med Microbiol. 2014; 304:14–22. [PubMed: 24120363]
3. Panthee S, Hamamoto H, Paudel A, Sekimizu K. Arch Microbiol. 2016; 198:839–845. [PubMed: 27541998]
4. Yu F, Zaleta-Rivera K, Zhu X, Huffman J, Millet JC, Harris SD, Yuen G, Li XC, Du L. Antimicrob Agents Chemother. 2007; 51:64–72. [PubMed: 17074795]
5. Harada, K-i, Suzuki, M., Kato, A., Fujii, K., Oka, H., Ito, Y. J Chromatogr A. 2001; 932:75–81. [PubMed: 11695870]
6. Zhang W, Li Y, Qian G, Wang Y, Chen H, Li YZ, Liu F, Shen Y, Du L. Antimicrob Agents Chemother. 2011; 55:5581–5589. [PubMed: 21930890]
7. Li Y, Chen H, Ding Y, Xie Y, Wang H, Cerny RL, Shen Y, Du L. Angew Chem. 2014; 53:7524–7530. [PubMed: 24890524]
8. Lou L, Qian G, Xie Y, Hang J, Chen H, Zaleta-Rivera K, Li Y, Shen Y, Dussault PH, Liu F, Du L. J Am Chem Soc. 2011; 133:643–645. [PubMed: 21171605]
9. Zhang J, Du L, Liu F, Xu F, Hu B, Venturi V, Qian G. FEMS Microbiol Lett. 2014; 355:170–176. [PubMed: 24801439]
10. Liu X, Walsh CT. Biochem. 2009; 48:8746–8757. [PubMed: 19663400]
11. Wang Y, Zhao Y, Zhang J, Zhao Y, Shen Y, Su Z, Xu G, Du L, Huffman JM, Venturi V, Qian G, Liu F. Appl Microbiol Biotechnol. 2014; 98:9009–9020. [PubMed: 25236801]
12. Wang Y, Qian G, Liu F, Li YZ, Shen Y, Du L. ACS Synth Biol. 2013; 2:670–678. [PubMed: 23937053]
13. Jin S, Liebscher J. Adv Synth Catal. 1998; 340:390–392.
14. Marahiel MA, Stachelhaus T, Mootz HD. Chem Rev. 1997; 97:2651–2674. [PubMed: 11851476]

15. Tokita K, Hori K, Kurotsu T, Kanda M, Saito Y. J Biochem. 1993; 114:522–527. [PubMed: 8276762]
16. Chen H, Du L. Appl Microbiol Biotechnol. 2016; 100:541–557. [PubMed: 26549236]
17. Du L, Chen M, Zhang Y, Shen B. Biochem. 2003; 42:9731–9740. [PubMed: 12911315]
18. Wenzel SC, Meiser P, Binz TM, Mahmud T, Müller R. Angw Chem Int Ed. 2006; 45:2296–301.
19. Delgado M, Caicoya AG, Greciano V, Benhamu B, Lopez-Rodriguez ML, Fernandez-Alfonso MS, Pozo MA, Manzanares J, Fuentes JA. Eur J Pharmacol. 2005; 511:9–19. [PubMed: 15777774]
20. Romanelli MN, Galeotti N, Ghelardini C, Manetti D, Martini E, Gualtieri F. CNS Drug Rev. 2006; 12:39–52. [PubMed: 16834757]

Figure 1. Architecture of *led* gene cluster in *L. enzymogenes* and activation by promoter replacement. (A) Genetic organization of the *led* gene cluster in *L. enzymogenes* OH11 (strain HW); (B) Promoter P_{HSAF} replacement for the three NRPS genes (*ledD*, *ledE*, *ledF*) of the cluster (strain *ledP_{HSAF}*); (C) qRT-PCR analysis of the NRPS gene expression in strain HW and strain *ledP_{HSAF}*; (D) HPLC analysis of metabolites in strain HW and strain *ledP_{HSAF}*.

Figure 2.
Chemical structures of Le-pyrrolopyrazine A–C (**1** – **3**).

Figure 3.

HPLC analysis of metabolites of various *L. enzymogenes* strains: (i) $\text{ledP}_{\text{HSAF}}$, strain HW with P_{HSAF} being inserted upstream of *ledD*; (ii) $\text{ledP}_{\text{HSAF}}$ ΔledD , strain $\text{ledP}_{\text{HSAF}}$ with *ledD* being in-frame deleted; (iii) $\text{ledP}_{\text{HSAF}}$ ΔledE , strain $\text{ledP}_{\text{HSAF}}$ with *ledE* being in-frame deleted; (iv) $\text{ledP}_{\text{HSAF}}$ ΔledF , strain $\text{ledP}_{\text{HSAF}}$ with *ledF* being in-frame deleted; (v) $\text{ledP}_{\text{HSAF}}$ ΔledDE , strain $\text{ledP}_{\text{HSAF}}$ with both *ledD* & *ledE* being in-frame deleted; (vi) $\text{ledP}_{\text{HSAF}}$ ΔledEF , strain $\text{ledP}_{\text{HSAF}}$ with both *ledE* & *ledF* being in-frame deleted; (vii) $\text{ledP}_{\text{HSAF}}$ -*ledE*-G455E, strain $\text{ledP}_{\text{HSAF}}$ with *ledE* A4-domain active site being point-mutated, G455E; (viii) $\text{ledP}_{\text{HSAF}}$ -*ledE*-H785V, strain $\text{ledP}_{\text{HSAF}}$ with *ledE* C5-domain active site being point-mutated, H785V. (ix) $\text{ledP}_{\text{HSAF}}$ -*ledF*-G963E, strain $\text{ledP}_{\text{HSAF}}$ with *ledF* A6-domain active site being point-mutated, G963E. (x) $\text{ledP}_{\text{HSAF}}$ -*ledF*-H205V, strain $\text{ledP}_{\text{HSAF}}$ with *ledF* C6-domain active site being point-mutated, H205V. (xi) $\text{ledP}_{\text{HSAF}}$ -*ledF*-Y1357F, strain $\text{ledP}_{\text{HSAF}}$ with *ledF* R2-domain active site being point-mutated, Y1357F. The metabolites were detected at UV 315 nm.

Figure 4.

Proposed biosynthesis for Le-pyrrolopyrazine A–C (**1–3**) in *L. enzymogenes* through a portable mechanism of NRPS modules/domains, the expression of which was activated by swapping the native promoter with a strong constitutive promoter of the highly expressed HSAF biosynthetic pathway.¹²