

NOTE

In silico identification of lysocin biosynthetic gene cluster from *Lysobacter* sp. RH2180-5

Suresh Panthee¹, Hiroshi Hamamoto¹, Yutaka Suzuki² and Kazuhisa Sekimizu^{1,3}

The Journal of Antibiotics (2017) 70, 204–207; doi:10.1038/ja.2016.102; published online 24 August 2016

Despite continual advances in antimicrobial therapies, infectious diseases pose a serious threat to global health owing to the ability of microbes to acquire resistance against antibiotics and transform themselves into superbugs.¹ Infectious diseases are an even more serious threat to health during immunocompromised conditions such as HIV/AIDS, diabetes and old age.² The recent increase in multidrug-resistant pathogens emphasizes the urgent need for the discovery of therapeutically active novel antimicrobial agents.¹

Lysobacter species are widely distributed in nature and characterized by their ability to lyse other organisms by producing a broad range of proteases and antibiotics.^{3,4} *Lysobacter* sp. RH2180-5 was isolated from soil in Japan while screening for novel therapeutically active antibiotics using the silkworm infection model. Lysocin E, a therapeutically active novel cyclic peptide produced by *Lysobacter* sp. RH2180-5, has antimicrobial activity against clinical isolates of methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*, including *S. simulans*, *S. haemolyticus*, *S. pseudintermedius*, *Bacillus subtilis*, *B. cereus* and *Listeria monocytogenes* with MICs ranging from 1–4 µg ml⁻¹.⁵ Lysocin E exhibits a novel mode of action that involves its binding to menaquinone in the cell membrane. In addition to lysocin E, *Lysobacter* sp. RH2180-5 produces at least eight more derivatives, most of which have antimicrobial activity. To further improve the production of lysocin E, analyze the derivatives and prepare novel derivatives with improved activity, the genetic analysis of the lysocin biosynthetic gene cluster is required.

The critical feature of the lysocin E chemical structure involves 12 amino acids including one *N*-methylphenylalanine and one fatty acid side chain. On the basis of this structure, we speculated that lysocin E could be synthesized via non-ribosomal peptide synthetases (NRPSs). In this work, we sequenced the whole genome of RH2180-5 and analyzed the putative lysocin biosynthetic gene cluster. The 71 kb DNA contained two large multimodular NRPSs named lysocin E synthetase (Les) A and LesB with a 1.7 MDa core peptide. A total of 12 modules and 43 domains were detected where LesA and LesB contained a loading condensation domain and terminal thioesterase domain, respectively. The specificities of the adenylation

domains suggested a linear mode of lysocin biosynthesis, and confirmed the association between the cluster and lysocin biosynthesis.

RESULTS AND DISCUSSION

Genome sequence and mining of lysocin biosynthetic gene cluster

Genomic DNA was extracted using a Qiagen DNA-blood Mini Kit (Qiagen, Hilden, Germany) using lysozyme for bacterial lysis. DNA libraries were prepared using the Illumina TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) for paired-end reads, and the quality and quantity of the constructed libraries were confirmed using a BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Reads were generated using the Illumina HiSeq2000 platform to obtain 4 million paired-end reads. Subsequent data processing was performed with CLC Genomics Workbench ver 8 (CLC bio, Aarhus, Denmark), resulting in 224 contigs with a combined length of 5.8 Mb and a G+C content of 70%. The longest contig was 289 kb, and the N₅₀ length, defined as the largest length *L* such that 50% of all nucleotides are contained in contigs with a size of at least *L*,⁶ was 76.6 kb. Genome annotation using MiGAP⁷ revealed a total of 4814 putative protein-coding sequences, and genes for 1 ribosomal RNA, 53 transfer RNAs and 38 miscellaneous RNAs. Genome analysis using antiSMASH⁸ revealed the presence of multiple secondary metabolite gene clusters encoding for non-ribosomal peptide, polyketide, arylpolene, bacteriocin and lantipeptide.

We detected 20 regions containing NRPS genes in the genome. On the basis of the module and domain organization of each NRPS, we speculated that 12 NRPS modules and 43 domains in LesA and LesB were involved in lysocin E biosynthesis, and focused our analysis on the 71 kb region of the genome (Figure 1a; Supplementary Table 1). LesA and LesB were estimated to be 998 kDa and 705 kDa, respectively, with a core peptide of ~1.7 MDa—one of the biggest NRPSs reported to date. We found 7 modules and 25 domains in LesA, and 5 modules and 18 domains in LesB (Figure 1b), and genes associated with the transport and regulation were detected in proximity to LesA and LesB. An MbtH-like protein was found downstream of LesB. Although MbtH-like proteins do not directly act as catalysts, biochemical studies have revealed that these proteins

¹Teikyo University Institute of Medical Mycology, Tokyo, Japan; ²Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan and ³Genome Pharmaceutical Institute, Tokyo, Japan
Correspondence: Professor K Sekimizu, Teikyo University Institute of Medical Mycology, 359 Otsuka, Hachioji, Tokyo 192-0395, Japan.
E-mail: sekimizu@main.teikyo-u.ac.jp

Received 1 June 2016; accepted 9 July 2016; published online 24 August 2016

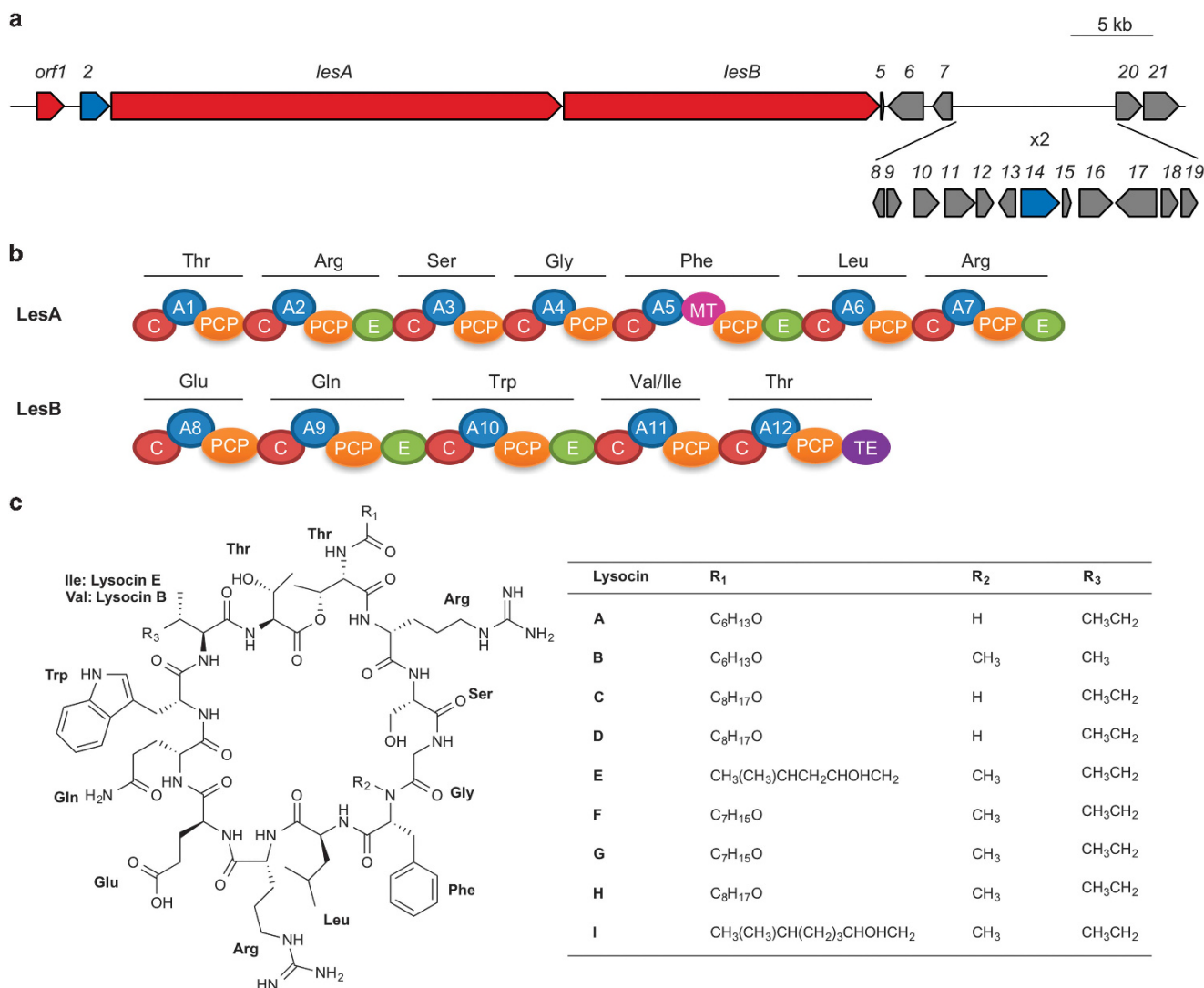


Figure 1 NRPS involved in lysocin biosynthesis. **(a)** Genetic organization of the lysocin biosynthetic gene cluster from *Lyso bacter* sp. RH2180-5. The annotation of the cluster is shown in Supplementary Table 1. **(b)** Deduced module and domain organization of LesA and LesB. Bars indicate the position of modules within the protein and each circle represents individual domains: A, adenylation domain; C, condensation domain; E, epimerization domain; MT, methyltransferase domain; PCP, peptidyl carrier protein domain; TE, thioesterase domain. **(c)** Chemical structure of nine lysocin derivatives identified so far. Les, lysocin E synthetase; NRPS, non-ribosomal peptide synthetase.

participate in tight binding to NRPS proteins containing an adenylation domain where they stimulate an adenylation reaction.⁹ The final cyclization process for formation of the lactone ring is carried out by the LesB thioesterase domain, which, in general, accounts for the majority of cyclization reactions within NRPSs.¹⁰

Specificities of adenylation domains in LesA and LesB

The adenylation domain is responsible for specifically attaching amino acids to the growing peptide chain.¹¹ The adenylation domain sequences on both LesA and LesB were extracted using the available bioinformatics tools.^{8,12} The specificity of the 12 domains toward the activating amino acid was predicted on the basis of sequence alignments of the signature code of pocket residues, as defined by Stachelhaus *et al.*,¹¹ and the nearest neighbor predicted from NRPSpredictor2 (Table 1).¹² Consistent with the lysocin structure, A1, A2, A3, A4, A5, A6, A7, A9 and A12 were predicted to activate Thr, Arg, Ser, Gly, Phe, Leu, Arg, Gln and Thr, respectively. A11 was predicted to activate Val, which is present in lysocin B, whereas lysocin

E harbors Ile (Figure 1c). This suggests the possibility that A11 activates both the amino acids. The lysocin A8 and A10 domains were predicted to activate Ser and Val, respectively, but no lysocin derivatives with Ser and Val at these positions have been isolated so far, requiring further analysis of these domains. The alignment of lysocin A8 and A10 with the characterized gene clusters showed that these domains shared a high degree of similarity with the A8 and A10 of the WAP-8294A, a cyclic depsipeptide containing 12 amino-acid, biosynthetic gene cluster,¹³ respectively, with nearly identical specificity-conferring codes. Because A8 and A10 from the WAP-8294A biosynthetic gene cluster activate Glu and Trp, respectively,¹³ we speculated that lysocin A8 and A10 would have similar affinity.

Domains conferring stereochemistry of amino acids

During NRPS biosynthesis, the elongation reaction of the peptidyl chain tethered to the phosphopantetheinyl arm of the upstream peptidyl carrier protein (PCP) domain to the amino acid bound to

Table 1 Analysis of adenylation and condensation domains in LesA and LesB

Module	A domain		C domain	Amino acid in lysocins
	Stachelhaus code	Nearest neighbor		
1	DFWNVGMVH-	Thr	^F C _L	L-Thr
2	DTEDVGAVD-	Arg	^L C _L	D-Arg
3	DVWHVSLVD-	Ser	^D C _L	L-Ser
4	DILQLGLIW-	Gly	^L C _L	Gly
5	DAWTIAAVC-	Phe	^L C _L	D-Phe
6	DAMLIGAIC-	Leu	^D C _L	L-Leu
7	DTEDVGCVD-	Arg	^L C _L	D-Arg
8	DVWHVGSIG-	Ser ^a	^D C _L	L-Glu
9	DAWQFGLIQ-	Gln	^L C _L	D-Gln
10	DAAIIGEIV-	Val ^a	^D C _L	D-Trp
11	DAYFWGVTF-	Val	^D C _L	L-Ile: lysocin E Val: lysocin B L-Thr
12	DFWNVGMVH-	Thr	^L C _L	

Abbreviation: Les, lysocin E synthetase.

^aIndicates amino acids not detected in lysocin derivatives.

the downstream PCP domain is catalyzed by the condensation (C) domains.¹⁴ The C domains that couple long-chain fatty acids to the PCP-tethered aminoacyl acceptor substrate are referred as the starter¹⁴ or ^FC_L.^{15,16} These ^FC_Ls catalyze the lipoinitiation process during lipopeptide biosynthesis. A ^DC_L domain catalyzes the condensation reaction between the D-residue in the upstream peptidyl donor and the L-residue in the downstream aminoacyl acceptor, whereas an ^LC_L is responsible for the condensation between two L-residues.¹⁴ Before a ^DC_L can perform its action, the chirality of the last amino acid in the growing peptide must be flipped, which is carried out by the epimerization domain. LesA and LesB harbor a total of 12 condensation domains and 5 epimerization domains. Lysocins harbor a relatively long β-hydroxyl fatty acid side chains attached to the peptide ring, and therefore, we expected that the first C domain of LesA would clade with known ^FC_L domains. To decipher the activity of the lysocin C domains, we first prepared a phylogenetic tree by using the reported sequences,¹⁴⁻¹⁷ and identified that the 12 C domains could be categorized into three domain clades, ^FC_L, ^DC_L and ^LC_L, each consisting of 1, 5 and 6 domains, respectively (Supplementary Figure 1). Each class of condensation domains differs in the amino-acid sequences present in the conserved residue of motifs



Figure 2 Amino-acid sequence alignment of motif C1–C7¹⁴ of three subtypes of condensation domains found in the lysocin biosynthetic gene cluster. The ^FC_L domain was aligned with ^FC_Ls from daptomycin¹⁵ and A54145¹⁷ gene clusters.

1–7.¹⁴ Sequence alignment showed a clear distinction between these three classes of condensation domains (Figure 2). The presence of epimerization domains upstream of these five ^DC_L domains suggested the presence of five D-amino acids in lysocin E, which was in complete agreement with the number and position of the ^DC_L domains in the lysocin cluster (Table 1).

Methylation and lipoinitiation of lysocin

N- or C-methylation of amino-acid residues makes the peptide less susceptible to proteolytic breakdown.¹⁸ The presence of D-N-methylphenylalanine in the lysocin E structure led us to search for the methyltransferase domain. In addition, of the nine isolated lysocin derivatives, six were methylated at Phe₅ (Figure 1c). Within the whole-lysocin biosynthetic gene cluster, we found an N-methyltransferase domain in module 5 of the *lesA* gene. As the A domain of module 5 encodes for Phe, the presence of methylated phenylalanine in lysocin E was in accordance with this finding.

The presence of a lipid side chain attached to the first amino acid, Thr, in lysocin E adds to the diversity of the lysocin derivatives. Biochemical analysis of lipoinitiation during the biosynthesis of daptomycin is reported. DptE, an acyl ligase (AL) activates the fatty acid, which is then transferred to DptF, an acyl carrier protein (ACP) before condensation by an ^FC_L of DptA.^{16,19} This AL-ACP mode of activation is predicted for A54145, where fused AL-ACP (LptEF) carries out both steps, and homologs of DptE and DptF are observed in multiple other gene clusters.¹⁶ In the absence of the AL-ACP mechanism, activation of the fatty acid for lipoinitiation is carried out by acyl-CoA ligase (ACL), where the fatty acid is converted to fatty acyl-CoA thioester and incorporated into the ^FC_L as in surfactin²⁰ or WAP-8492 A.²¹ ACLs also activate fatty acids for polyketide biosynthesis.^{22,23} Our analysis of the open reading frames of the predicted gene cluster did not reveal the genes encoding for either the AL-ACP or ACL mode of activation, suggesting the involvement of a mechanism outside the gene cluster.

In summary, we performed a genome sequence analysis of *Lysobacter* sp. RH2180-5, identified the lysocin E biosynthetic gene cluster and proposed a linear logic of biosynthesis. Our bioinformatic analysis will facilitate further genetic and biochemical analyses of lysocin biosynthesis, which will enhance our understanding of the molecular mechanisms of lysocin biosynthesis, thus paving the path for the generation of new analogs with better antimicrobial properties. Although *Lysobacter* species have recently drawn attention as the producer of bioactive secondary metabolites,^{3,4} the RH2180-5 genome sequence provides the opportunity to elucidate the metabolic potential of this species. As an enormous range of medicinally important compounds is synthesized by bacterial NRPSs, our findings will facilitate genetic manipulation of the gene clusters involved in the biosynthesis of lysocin and other NRPSs for the generation of novel (un)natural analogs with improved bioactivity.

Nucleotide sequence accession number

The lysocin E biosynthetic gene cluster has been deposited at DDBJ under the accession number LC128664.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by MEXT KAKENHI (No. 221S0002) and a Grant-in-Aid for young scientists (A) (JP24689008) to HH; and in part by Grant-in-Aid for scientific research (S) (JP15H05783) and Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED to KS.

- 1 WHO. *Antimicrobial Resistance: Global Report on Surveillance* (World Health Organization, Geneva, 2014).
- 2 Elliott, I. & Venkatesan, P. in *eLS* (John Wiley & Sons, Chichester, UK, 2015).
- 3 Xie, Y., Wright, S., Shen, Y. & Du, L. Bioactive natural products from *Lysobacter*. *Nat. Prod. Rep.* **29**, 1277–1287 (2012).
- 4 Pidot, S. J., Coyne, S., Kloss, F. & Hertweck, C. Antibiotics from neglected bacterial sources. *Int. J. Med. Microbiol.* **304**, 14–22 (2014).
- 5 Hamamoto, H. *et al.* Lysocin E is a new antibiotic that targets menaquinone in the bacterial membrane. *Nat. Chem. Biol.* **11**, 127–133 (2015).
- 6 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- 7 Sugawara, H., Ohyama, A., Mori, H. & Kurokawa, K. Microbial genome annotation pipeline (MiGAP) for diverse users. In *The 20th International Conference on Genome Informatics (GIW2009)*. 1–2 (2009).
- 8 Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339–W346 (2011).
- 9 Baltz, R. H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747–1760 (2011).
- 10 Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
- 11 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
- 12 Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).
- 13 Zhang, W. *et al.* Identification and characterization of the anti-methicillin-resistant *Staphylococcus aureus* WAP-8294A2 biosynthetic gene cluster from *Lysobacter enzymogenes* OH11. *Antimicrob. Agents Chemother.* **55**, 5581–5589 (2011).
- 14 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 1–15 (2007).
- 15 Miao, V. *et al.* Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507–1523 (2005).
- 16 Baltz, R. H. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).
- 17 Miao, V. *et al.* The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J. Ind. Microbiol. Biotechnol.* **33**, 129–140 (2006).
- 18 Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453–488 (2004).
- 19 Wittmann, M., Linne, U., Pohlmann, V. & Marahiel, M. A. Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS J.* **275**, 5343–5354 (2008).
- 20 Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M. & Marahiel, M. A. Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. *Chem. Biol.* **17**, 872–880 (2010).
- 21 Chen, H., Olson, A. S., Su, W., Dussault, P. H. & Du, L. Fatty acyl incorporation in the biosynthesis of WAP-8294A, a group of potent anti-MRSA cyclic lipodepsipeptides. *RSC Adv.* **5**, 105753–105759 (2015).
- 22 Miyazawa, T. *et al.* Identification of middle chain fatty acyl-CoA ligase responsible for the biosynthesis of 2-alkylmalonyl-CoAs for polyketide extender unit. *J. Biol. Chem.* **290**, 26994–27011 (2015).
- 23 Takahashi, S. *et al.* Reveromycin A biosynthesis uses RevG and RevJ for stereospecific spiroacetal formation. *Nat. Chem. Biol.* **7**, 461–468 (2011).

Supplementary Information accompanies the paper on The Journal of Antibiotics website (<http://www.nature.com/ja>)