Novel biosynthesis pathway for aniline derivatives

Shunsuke MASUO

Faculty of Life and Environmental Sciences, University of Tsukuba

Research aims

Aniline derivatives are important compounds that can be used as raw materials for pharmaceuticals and chemicals. Recently, we discovered a novel pyrazine biosynthesis gene cluster in *Pseudomonas fluorescens*.[1)] This gene cluster contains the 4-aminophenylalanine (4APhe) biosynthesis gene *papABC* and the pyrazine-ring biosynthesis gene *papDEF*, which is involved in the production of 4-aminobenzyl-substituted pyrazines. Using the genome mining approach, we discovered a gene cluster containing a putative non-ribosomal peptide synthase (NRPS) gene and *papABC*-like genes in *Haliangium ochraceum* (Fig. 1A). Here, we tried to understand the function of this gene cluster and to reveal the biosynthesis mechanism of novel aniline derivatives.

Methods

1. *Plasmid construction*

Gene fragments were amplified by PCR using KOD One (Toyobo), appropriate primers, and the genomic DNA of *H. ochraceum* JCM 11303, and were cloned into expression vectors using the NEB Builder HiFi DNA Assembly (NEB). Sequences were confirmed by next-generation sequencing and Sanger sequencing.

2. *LC-MS analysis*

The *E. coli* culture supernatant and ethyl acetate extraction fractions were subjected to LC-MS analysis using LCMS 8045 (Shimadzu). Metabolites were separated using an ACQUITY UPLC BEH Shield RP18 Column (Waters).

Results

1. *Domain prediction of the novel non-ribosomal peptide synthase*

In silico analysis using AntiSmash 5.1 predicted that Hoch1367 codes an NRPS consisting of five modules (Fig. 1B). Based on comparison of the amino acid residues involved in substrate recognition that were conserved in the adenylation (A) domain, modules 1, 2, 4, and 5 were predicted to

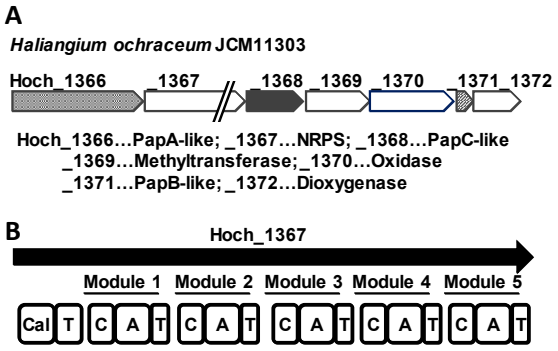

Figure 2. Gene cluster(A)and NRPS (B)

recognize serine, serine, proline, and glycine, respectively. The N-terminal acyl-CoA ligase (CAL) domain could be involved in the acylation of serine in module 1.[2)] On the contrary, none of the residues in module 3 showed high homology with known substrate recognition sequences, suggesting that module 3 recognizes a previously unknown substrate. As Hoch_1367 clusters with putative 4-aminophenylalanine (4APhe) biosynthesis genes, it is assumed that module 3 recognizes 4APhe or its derivatives.

2. *Hoch_1366, _1368, and _1370 encode PapA, PapB, and PapC, respectively to produce 4APhe*

Hoch_1366, Hoch_1368, and Hoch_1370 were heterologously expressed in *Escherichia coli* BL21 (DE3). HPLC analysis showed that *E. coli* expressing Hoch_1366, _1368, and _1370 accumulated 4APhe in the culture supernatant, whereas *E. coli* expressing only Hoch_1366 or Hoch_1368 and 1370 did not. These results indicated that Hoch_1366, Hoch_1368, and Hoch_1370 function as PapA, PapB, and PapC, respectively, to produce 4APhe.

3. *Functional analysis of the novel gene cluster*

To express active NRPS, we generated *E. coli* expressing phosphopantetheinyl transferase (PPTase), which is involved in the post-translational modification of NRPS. The *mtaA* gene, which encodes PPTase with broad substrate specificity derived from the myxobacteria *Stigmatella aurantiaca*, was fused with the T7lac promoter. This fragment was then introduced into the *entD* locus of *E. coli* BL21 (DE3), HMS (DE3), and BL21 Star (DE3) to generate NRPS expression hosts BLmtaA, HMSmtaA, and BLSmtaA, respectively. A 27 kb DNA fragment containing cluster genes from Hoch_1366 to Hoch_1372 was cloned into the T7lac promoter-driven expression plasmids (Fig. 2A), and introduced into each NRPS expression host. The resulting strains were then cultured in various media, temperatures, and IPTG concentrations, and metabolites after IPTG induction, and then subjected to LC-MS analysis. Unfortunately, no specific metabolites were identified in the strain harboring the expression plasmid compared to the control strains harboring the empty plasmid. We constructed another expression plasmid, in which each cluster gene was independently cloned under the

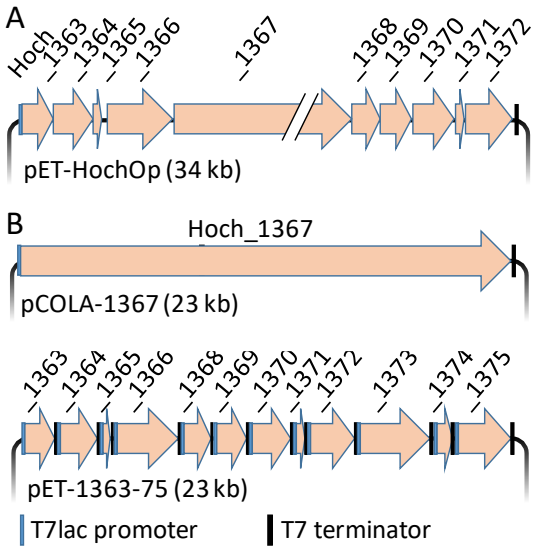

Figure 2. Expression plasmids constructed in this study.

T7lac promoter (Fig. 2B) and introduced into the NRPS expression host. We are currently performing LC-MS analysis to detect the metabolites produced by the heterologously expressed gene cluster from *H. ochraceum*.

4. *Substrate recognition by the A domain of module 3*

To analyze the function of module 3, which consists of the condensation (C), A, and PCP (T) domains (Fig. 2B), we attempted to prepare the recombinant N-terminal or C-terminal His-tagged module 3. Although the target protein was expressed as inclusion bodies in most *E. coli* hosts and culture conditions, small amounts of soluble expression were observed when cultured at low temperatures in ArcticExpress (DE3) RP. Recombinant module 3 was then prepared by affinity purification using Ni-NTA beads. We will next measure the adenylation activities of module 3 against 4APhe and its derivatives.

Conclusion

In this study, we discovered a novel gene cluster containing the NRPS-like protein that is considered to recognize 4APhe derivatives. Hoch_1366, _1368, and _1370 in the cluster encode PapA, PapB, and PapC, respectively, and were identified as 4APhe biosynthetic genes. This research focused on heterologous expression in *E. coli*, but in future, we would like to utilize other expression hosts such as actinomycetes and myxobacteria. Further, we would like to construct gene deletion strains of *H. ochraceum*. Our study will thus reveal the functions of novel NRPS and related enzymes and identify novel aniline-structure-containing peptide compounds.

References

1) Masuo, S., Tsuda, Y., Namai, T., Minakawa, H., Shigemoto, R., and Takaya, N. (2020) Enzymatic Cascade in Pseudomonas that Produces Pyrazine from α-amino acids. *ChemBioChem* **21:** 353-359.
2) Baars, O., Zhang, X., Gibson, M.I., Stone, A.T., Morel, F.M.M., and Seyedsayamdost, M.R. (2018) Crochelins: Siderophores with an Unprecedented Iron-Chelating Moiety from the Nitrogen-Fixing Bacterium *Azotobacter chroococcum*. *Angew Chem Int Ed Engl.* **57:** 536-541.