Article

Unique *N*-phenylacetylation and NRPS with Substrate Promiscuity for Biosynthesis of Heptapeptide Variants, JBIR-78 and JBIR-95

Kunpei Takeda, Kohei Kemmoku, Yasuharu Satoh, Yasushi Ogasawara, Kazuo Shin-ya, and Tohru Dairi

ACS Chem. Biol., **Just Accepted Manuscript** • Publication Date (Web): 15 May 2017

Just Accepted

ACS Chemical Biology is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Unique *N*-phenylacetylation and NRPS with Substrate Promiscuity for Biosynthesis of Heptapeptide Variants, JBIR-78 and JBIR-95

Kunpei Takeda†, Kohei Kemmoku†, Yasuharu Satoh†, Yasushi Ogasawara†, Kazuo Shin-ya‡, and Tohru Dairi†*

†Graduate School of Engineering, Hokkaido University, N13-W8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan

‡Biomedicinal Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

*Corresponding author. E-mail: dairi@eng.hokudai.ac.jp, Tel: +81-11-706-7815

ABSTRUCT

JBIR-78 (**1**) and JBIR-95 (**2**), both of which are heptapeptide derivatives isolated from *Kibdelosporangium* sp. AK-AA56, have the same amino acid sequences except for the second amino acid; phenylacetic acid (Paa)–L-Val–D-Asp (**1**)/D-cysteic acid (**2**)–L-Ala–(3*S*)-3-hydroxy-D-Leu–Gly–D-Ala–L-Phe. Heterologous expression of the biosynthetic gene cluster including genes encoding non-ribosomal peptide synthetases (NRPS) and in vitro assays with recombinant Orf3, L-cysteic acid synthase homolog, suggested the single A-domain in module 2 activates both L-Asp and L-cysteic acid to yield **1** and **2**, respectively, although the substrate specificities of the A-domains of NRPSs are usually strict. Biosynthetic mechanism of introduction of *N*-terminal Paa was also investigated. Recombinant Orf1 and Orf2 similar to subunits of pyruvate dehydrogenase complex catalyzed the conversion of phenylpyruvate into phenylacetyl-CoA together with dihydrolipoyl dehydrogenase whose encoding gene is located at outside of the gene cluster. Moreover, we showed that phenylacetyl-CoA was directly condensed with L-Val, which was tethered to a peptidyl carrier protein, at the first condensation domain in the NRPS.

INTRODUCTION

Peptide antibiotics are a large family of natural products, including clinically important pharmaceuticals. In addition to isolated natural products, genome databases have revealed

the presence of many biosynthetic genes for this category of compounds. These peptides can be categorized by their biosynthetic machineries as either ribosomally synthesized and post-translationally modified peptides (RiPPs) or nonribosomal peptides (NRPs).[1,2] In the latter case, nonribosomal peptide synthetases (NRPSs) are representative biosynthetic enzymes. NRPSs are modular type large enzymes and a typical module consists of an adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation (C) domain. The A-domain activates the carboxylic acid of an amino acid with ATP and determines the amino acid to be selected and activated. Unlike ribosomes, NRPSs can utilize nonproteinogenic amino acids as building blocks. The C-domain catalyzes the peptide bond formation between the upstream peptidyl PCP and downstream amino acyl PCP. Besides NRPSs, amino acid ligases with ATP-grasp domains (ATP-grasp-ligases), tRNA-dependent aminoacyl transferase/cyclodipeptide synthases, stand-alone adenylation (A) domains and acyl-CoA synthetases (acyl-AMP-ligases) are also involved in NRP biosynthesis.[3–6]

1 and **2**, both of which are heptapeptide derivatives isolated from *Kibdelosporangium* sp. AK-AA56, have the same amino acid sequence except for the second amino acid (Figure 1).[7] We are interested in their biosynthetic machineries from the following points of view. One is the difference of the second amino acid. Considering the presence of D-amino acids in **1** and **2**, they could be biosynthesized by a NRPS. However, the substrate

specificities of the A-domains of NRPSs are strict and one A-domain usually activates one amino acid. Therefore, we hypothesized three possibilities: (i) the producer possesses separate NRPSs for **1** and **2** biosynthesis; (ii) one NRPS biosynthesizes **1**, which is then converted to **2** by tailoring reactions; (iii) L-cysteic acid is independently biosynthesized and a single A-domain is exceptionally able to activate both L-Asp and L-cysteic acid. The other interesting feature is the presence of the *N*-terminal phenylacetate. To date, a dozen phenylacetylated natural polyketides and a few *N*-terminal phenylacetylated peptides such as penicillin G, microcystin,[8] and JBIR-96[9] have been isolated. However, reports on the detailed phenylacetylation mechanism are limited. In the case of penicillin G, isopenicillin N acyltransferase was shown to convert isopenicillin N to penicillin G.[10] In microcystin biosynthesis, phenylpropanoids were suggested to be loaded onto the PCP-domain rather than phenylacetate by ATP–PPi exchange assays and mass spectrometry, although the mechanism by which one carbon is excised from the starter substrate remains unknown.[8] Recently, Fu et al. reported that a pyruvate dehydrogenase-like protein complex catalyzed the decarboxylation of phenylpyruvate to form a phenylacetyl-*S*-acyl carrier protein, which would be utilized in a subsequent polyketide biosynthetic assembly line, in ripostatin biosynthesis.[11] In this study, we investigated the phenylacetylation mechanism in JBIR-78 (**1**) and JBIR-95 (**2**) biosynthesis.

RESULTS AND DISCUSSION

Identification of 1 and 2 biosynthetic genes

To identify the **1** and **2** biosynthetic gene cluster(s), we generated a draft genome sequence for the producer and searched for *orf*s large enough to encode the seven A-domains essential for substrate activation in heptapeptide biosynthesis. From a BLAST search, we identified one candidate gene cluster (Figure 2, Table 1). The gene cluster consisted of eight *orf*s containing three *orf*s (*orf4* to *orf6*) encoding NRPSs. Orf4 and Orf5 had the same domain architectures; C-A-PCP-C-A-PCP-epimerization domain (E-domain). Orf6 was composed of C-A-PCP-C-A-PCP-E-C-A-PCP-thioesterase (TE) domains. The positions of the E-domains perfectly matched the locations of the D-amino acids in **1** and **2**. Moreover, the substrate specificities of each of the A-domains predicted by NRPSpredictor2 (http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob) were roughly identical to the amino acid sequences of **1** (Supplementary Table 1, 2, and 3). Two *orf*s (*orf1* and *orf2*), which encode two of the three component enzymes of pyruvate dehydrogenase complex, were found in the region upstream from the NRPS genes. Orf1 was composed of pyruvate dehydrogenase (EC 1.2.4.1, E1β unit) and dihydrolipoyl transacetylase (EC 2.3.1.12, E2 unit). Orf2 was similar to the E1α unit of pyruvate dehydrogenase. Although no homologs of dihydrolipoyl dehydrogenase (EC 1.8.1.4, E3 unit) existed in the gene cluster, we identified one homolog outside of the gene cluster.

Considering the reaction catalyzed by the pyruvate dehydrogenase complex, Orf1, Orf2, and the dihydrolipoyl dehydrogenase homolog perhaps catalyze the conversion of phenylpyruvate into phenylacetyl-CoA, which would be used for *N*-terminal phenylacetylation.

Orf3 showed similarity to the MA3297 protein, which is a pyridoxal 5′-phosphate dependent enzyme that catalyzes a β-replacement reaction converting L-phosphoserine and sulfite into L-cysteic acid and inorganic phosphate.[12] Therefore, Orf3 is plausibly responsible for the biosynthesis of **2**.

Orf8 was similar to cytochrome P450 and presumably participates in the hydroxylation of Leu to form 3-hydroxy-D-Leu. Additionally, Orf7 was similar to MbtH, an integral component of NRPSs that is essential for amino acid activation.[13] The identified gene cluster was the sole candidate and no paralogs of the identified genes existed in the draft genome database. Therefore, the production of both **1** and **2** was suggested to be governed by this gene cluster.

Heterologous expression of the gene cluster

Next, we performed a heterologous expression experiment to examine whether the gene cluster identified in this study contained all the genes responsible for **1** and **2** biosynthesis. A cosmid library of the producer was constructed with a shuttle cosmid vector,

pOJ446.[14] Positive cosmids carrying the gene cluster were obtained by PCR screening with appropriate primers and the cosmid pC35 was introduced into *Streptomyces lividans* TK23. After cultivation of the transformants, the products were analyzed by LC-ESI-MS. As shown in Figure 3 and Supplementary Figure 1, the production of both **1** and **2** was suggested. Although the cosmid had one additional gene (Figure 2) downstream of *orf8* gene, this gene was suggested to have no relation to **1**/**2** biosynthesis. Taken together, our results suggested that both products were biosynthesized by this gene cluster alone and that the A-domain responsible for activation of the second substrate exceptionally activated both L-Asp and L-cysteic acid.

Characterization of Orf1 and Orf2

To examine whether Orf1 and Orf2 catalyzed the conversion of phenylpyruvate to phenylacetyl-CoA, an *in vitro* assay with recombinant enzymes was carried out. A DNA fragment carrying both the *orf1* and *orf2* genes was inserted into the pET28a vector so that only Orf2 was expressed as a His-tagged recombinant enzyme because Orf1 was co-purified with Orf2 by Ni-NTA column chromatography (Supplementary Figure 2). A recombinant enzyme of the putative dihydrolipoyl dehydrogenase (E3 unit) was also prepared as a His-tagged recombinant enzyme (Supplementary Figure 2). After the recombinant enzymes were expressed and purified, they were mixed and incubated with

phenylpyruvate and CoA in the presence of known essential co-factors in the pyruvate dehydrogenase reaction. A new product whose molecular mass was consistent with that of phenylacetyl-CoA was specifically detected by LC-ESI-MS analysis (Figure 4). Its high resolution molecular mass also agreed with that of phenylacetyl-CoA ($[M+H]^+$ calcd. for $C29H43N7O17P3S^+$, 886.16435: found 886.16534).

Characterization of Orf3

Orf3 showed similarity to the pyridoxal 5′-phosphate dependent L-cysteic acid synthase. To confirm its expected activity, we tried an in vitro assay with recombinant enzymes. Although we could express Orf3 recombinant enzymes with several expression vectors such as pET vectors, pMAL-c5X and pCold-TF, all recombinant enzymes formed inclusion bodies. Therefore, we carried out heterologous expression of the gene cluster without the *orf3* gene. We constructed a cosmid lacking *orf3* gene (pC35Δorf3) from pC35. Unexpectedly, the transformant harboring pC35Δorf3 did not lose **2** productivity and still produced a small amount of **2** concomitant with enhanced production of **1** (Figure 3 and Supplementary Figure 1). A possible reason for this is that L-cysteic acid might be supplied by an alternative enzyme such as L-serine ammonia-lyase (EC 4.3.1.17) that catalyzes the elimination of water from L-serine to form L-dehydroalanine, to which sulfite attacks. Indeed, **2** productivity was dramatically recovered when L-cysteic acid was added into the

medium (Figure 3 and Supplementary Figure 1), suggesting that the Orf3 product supplies L-cysteic acid. Furthermore, the productivity of **1** was extremely reduced by the addition of L-cysteic acid, suggesting that L-cysteic acid competed with L-Asp in the incorporation into the peptide.

Previously, a few A-domains in NRPSs, which activated two different amino acids, were reported. Pelgipeptin A/C and pelgipeptin B/D possess L-Val and L-Ile at the second position, respectively. The substrate selectivity of A-domains (PlpE A1) responsible for activation of the second amino acid was examined by in vitro experiments and PlpE A1 was shown to activate both L-Val and L-Ile.[15] *C*-terminal diversity of lichenysins (L-Ile or L-Leu) was also derived by the substrate selectivity of A-domains (LicC-A).[16] In both cases, the A-domains recognized the structurally and chemically similar amino acids. For the same reason, the A-domain in module 2 might accept L-Asp and L-cysteic acid, both of which are acidic amino acids and have similar sizes.

We then tried denaturation and refolding of the inclusion body to obtain the active form of the recombinant enzyme. After extensive trials to find the best conditions, we successfully obtained active enzymes (Supplementary Figure 3). The refolded enzyme was incubated with L-phosphoserine and sodium sulfite for 15 min at 30°C and the reaction products were analyzed by LC-ESI-MS. As shown in Figure 5, a specific peak, which was eluted at the same retention time and had the same molecular mass as authentic L-cysteic

acid, was detected.

Characterization of Orf4

Through detailed analysis of each of the domains in Orf4, we found that a PCP-domain to anchor phenylpyruvate was absent. Moreover, Orf1 and Orf2 catalyzed the formation of phenylacetyl-CoA as mentioned above. These results suggested that phenylacetyl-CoA is directly condensed with L-Val at the first C-domain in module 1 of Orf4 in a similar manner to that of surfactin and amphi-enterobactin siderophore biosynthesis.[17,18] To examine this possibility, a truncated recombinant Orf4 (module 1) was prepared and used for in vitro assay. We first tried to prepare a recombinant enzyme possessing the first module with the C-, A-, and PCP-domains. However, we were unable to obtain recombinant proteins. We then prepared recombinant enzymes divided into C-A-domains and the PCP-domain. In this case, both recombinant enzymes were expressed as soluble forms (Supplementary Figure 4).

The purified recombinants were incubated with L-valine and phenylacetyl-CoA, which was prepared by the enzyme reaction described above, in the presence of ATP and $MgCl_2$. After the enzymes were precipitated by adding acetone, the thioester bonds were hydrolyzed with 0.1 M KOH. The resulting sample was analyzed by LC-ESI-MS and compared with authentic phenylacetyl valine prepared by chemical synthesis. As shown in

Figure 6, a specific peak with the same retention time and molecular mass as those of the authentic standard was clearly detected in the sample, showing that phenylacetyl-CoA was directly condensed with L-valine, which was activated at the A-domain and transferred to PCP in module 1.

Recently, a similar mechanism for the formation of phenylacetate from phenylpyruvate by a pyruvate dehydrogenase-like protein complex in ripostatin biosynthesis was reported.[11] In this case, however, the phenylacetyl group of *S*-phenylacetyldihydrolipoamide was directly transferred to an acyl carrier protein to form a phenylacetyl-*S*-acyl carrier protein without the formation of phenylacetyl-CoA.

In summary, we studied the biosynthetic machinery of **1** and **2**, which are both heptapeptide derivatives and have the same amino acid sequence except for the second amino acid. Through heterologous expression of the biosynthetic gene cluster and in vitro assay with the recombinant Orf3, the single A-domain in module 2 was suggested to activate both L-Asp and L-cysteic acid. We also showed that *N*-terminal phenylacetylation was catalyzed by the first C-domain in module 1 with L-Val-tethered PCP and phenylacetyl-CoA supplied by a phenylpyruvate dehydrogenase complex.

REFERENCES

(1) Fischbach, M. A., and Walsh, C. T. (2006) Assembly-line Enzymology for Polyketide

and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev. 106*, 3468–3496.

(2) Ortega, M. A., and van der Donk, W. A. (2016) New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell. Chem. Biol. 23*, 31–44.

(3) Giessen, T. W., and Marahiel, M. A. (2012) Ribosome-independent Biosynthesis of Biologically Active Peptides: Application of Synthetic Biology to Generate Structural Diversity. *FEBS Lett. 586*, 2065–2075.

(4) Kadi, N., Oves-Costales, D., Borona-Gomez, F., and Challis, G. L. (2007) A New Family of ATP-dependent Oligomerization-macrocyclization Biocatalysts. *Nat. Chem. Biol. 3*, 652–656.

(5) Maruyama, C., Toyoda, J., Kato, Y., Izumikawa, W., Takagi, M., Shin-ya, K., Katano, H., Utagawa, T., and Hamano, Y. (2012) A Stand-alone Adenylation Domain Forms Amide Bonds in Streptothricin Biosynthesis. *Nat. Chem. Biol. 8*, 791–797.

(6) Noike, M., Matsui, T., Ooya, K., Sasaki, I., Ohtaki, S., Hamano, Y., Maruyama, C., Ishikawa, J., Satoh, Y., Ito, H., Morita, H., and Dairi, T. (2015) A Peptide Ligase and the Ribosome Cooperate to Synthesize the Peptide pheganomycin. *Nat. Chem. Biol. 11*, 71–76.

(7) Izumikawa, M., Takagi, M., and Shin-ya, K. (2012) JBIR-78 and JBIR-95:

Phenylacetylated Peptides Isolated from *Kibdelosporangium* sp. AK-AA56. *J. Nat. Prod. 75*, 280–284.

(8) Hicks, L. M., Moffitt, M. C., Beer, L. L., Moore, B. S., and Kelleher, N. L. (2006) Structural Characterization of *in vitro* and *in vivo* Intermediates on the Loading Module of Microcystin Synthetase. *ACS Chem. Biol. 1*, 93–102.

(9) Ueda, J., Izumilawa, M., Kozone, I., Yamamura, H., Hayakawa, M., Takagi, M., and Shin-ya, K. (2011) A Phenylacetylated Peptide, JBIR-96, Isolated from *Streptomyces* sp. RI051-SDHV6. *J. Nat. Prod. 74*, 1344–1347.

(10) Tobin, M. B., Fleming. M. D., Skatrud, P. L., and Miller, J. R. (1990) Molecular Characterization of the Acyl-coenzyme A: Isopenicillin N Acyltransferase Gene (*penDE*) from *Penicillium chrysogenum* and *Aspergillus nidulans* and Activity of Recombinant Enzyme in *Escherichia coli*. *J. Bacteriol. 172*, 5908–5914.

(11) Fu, C., Auerbach, D., Li,Y., Scheid, U., Luxenburger, E., Garcia, R., Irschik, H., and Müller, R. (2017) Solving the Puzzle of One-carbon Loss in Ripostatin Biosynthesis. *Angew. Chem.*, *Int. Ed. 56*, 2192–2197.

(12) Graham, D. E., Taylor, S. M., Wolf, R. Z., and Namboori, S. C. (2009) Convergent Evolution of Coenzyme M Biosynthesis in the Methanosarcinales: Cysteate Synthase Evolved from an Ancestral Threonine Synthase. *Biochem. J. 424*, 467–478.

(13) Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D.

W., and Thomas, M. G. (2010) MbtH-like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry 49*, 8815–8817.

(14) Vara, J., Lewandowska-Skarbek, M., Wang, Y. G., Donadio, S., and Hutchinson, C. R. (1989) Cloning of Genes Governing the Deoxysugar Portion of the Erythromycin Biosynthesis Pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J. Bacteriol. 171*, 5872–5881.

(15) Qian, C. D., Liu, T. Z., Zhou, S. L., Ding, R., Zhao, W. P., Li, O., and Wu, X. C. (2012) Identification and Functional Analysis of Gene Cluster Involvement in Biosynthesis of the Cyclic Lipopeptide Antibiotic Pelgipeptin Produced by *Paenibacillus elgii*. *BMC Microbiology*, *12*, 197.

(16) Konz, D., Doekel, S., and Marahiel, M. A. (1999) Molecular and Biochemical Characterization of the Protein Template Controlling Biosynthesis of the Lipopeptide Lichenysin. *J. Bacteriol*. *181*, 133–140.

(17) Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M., and Marahiel, M. A. (2010) Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation. *Chem. Biol. 17*, 872–880.

(18) Zane, H. K., Naka, H., Rosconi, F., Snady, M., Haygoogd, M. G., and Butler, A. (2014) Biosynthesis of Amphi-enterobactin Siderophores by *Vibrio harveyi* BAA-1116: Identification of a Bifunctional Nonribosomal Peptide Synthetase

Condensation Domain. *J. Am. Chem. Soc. 136*, 5615–5618.

(19) Dairi, T., Nakano, T., Aisaka, K., Katsumata, R., and Hasegawa, M. (1995) Cloning and Nucleotide Sequence of the Gene Responsible for Chlorination of Tetracycline. *Biosci. Biotech. Biochem*. *59*, 1099–1106.

(20) Datsenko, K. A., and Wanner, B. L. (2000) One-step Inactivation of Chromosomal Genes in *Escherichia coli* K-12 Using PCR Products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640–6645.

(21) Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry 37*, 1585–1595.

METHODS

General

All chemicals were purchased from Sigma-Aldrich Japan K.K. (Japan), Tokyo Chemical Industry Co. Ltd. (Japan), or Wako Pure Chemical Industry (Japan). Oligonucleotides were obtained from FASMAC Co. Ltd. (Japan). Enzymes, molecular weight standards and kits for DNA manipulation were purchased from Takara Bio Inc. (Japan) or New England Biolabs Japan Inc. (Japan). PCR reactions were carried out using a GeneAmp PCR System 9700 thermal cycler (Thermo Fisher Scientific Inc.) with Tks Gflex DNA polymerase (Takara Bio). NMR spectra were obtained using a JEOL ECS-400 spectrometer (Japan).

Chemical synthesis of phenylacetyl valine

To a stirred solution of 1.03 g (8.8 mmol) of L-valine in 10 mL (20 mmol) of 2 M aq. NaOH at 0°C was added dropwise 1.05 mL (8 mmol) of phenylacetyl chloride. The reaction mixture was allowed to warm to room temperature and stirred. After 20 h of reaction, water (30 mL) was added and the aqueous layer containing the product was washed with Et_2O (50 mL) twice. The water layer was then acidified with 1 M HCl to obtain a white precipitate. The white crystalline product was washed thoroughly with water and then with Et_2O (606 mg, 29%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.24 (d, J = 8.6 Hz, 1H), 7.32–7.09 (m, 5H), 4.15 (dd, J = 8.6, 5.8 Hz, 1H), 3.56 (d, J = 13.8 Hz, 1H), 3.48 (d, J

= 13.8 Hz, 1H), 2.05 (m, 1H), 0.88 (d, J = 6.8 Hz, 3H), 0.85 (d, J = 6.8 Hz, 3H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 173.1, 170.4, 136.6, 129.0, 128.1, 126.3, 57.1, 41.9, 29.9, 19.1, 18.0.

Draft genome sequences

Draft sequences of the *Kibdelosporangium* sp. AK-AA56 genome were determined by a commercial company (Hokkaido System Science, Japan) using an Illumina HiSeq platform (Illumina). A genomic DNA library (350 bp insert) was constructed with the TruSeq Nano DNA LT Sample Prep Kit (Illumina). Paired end data (2 × 100 bp) were assembled using Velvet (version 1.0.18). Gene prediction and annotation was carried out using Microbial Genome Annotation Pipeline (MiGAP, http://www.migap.org/).

Cloning and heterologous expression of the gene cluster

To clone the gene cluster, a cosmid library of the producer was constructed with the shuttle cosmid vector pOJ446.[14] Genomic DNA of *Kibdelosporangium* sp. was partially digested with *Sau*3AI and ligated with the pOJ446 vector digested with *Bam*HI and *Hpa*I and treated bacterial alkaline phosphatase. This was packaged using Gigapack III XL Packaging Extract (Agilent Technologies Inc.) and transfected to the *E. coli* XL1-Blue MRF′ strain (Δ*(mcrA)183* Δ*(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*

[F′*proAB lacI*^{q}*ZΔM15 Tn10* (*Tet*^{r})]; Agilent Technologies). Positive cosmids carrying the gene cluster were screened by PCR based on the amplification of both the *orf1* and *8* genes with the primers 5′-GGCCGCGACGTGGGTGAGGAAGTTG-3′ and 5′-CGTTCCCGATCTCGGTGAAGGACTTGCTGAG-3′ for *orf1*, and 5′-CATGGGCGAGTAGCGGAACGAGAGGAACAAG-3′ and 5′-GGGCCGTAGGCGAATGTCAGGTGTTTGTTC-3′ for *orf8*.

The obtained cosmid, pC35, was introduced into *S. lividans* TK23 and then the transformants were cultured in 200 mL flasks with baffles containing 30 mL of SK#2 medium[19] containing thiostrepton (10 μg mL^{-1}) for 4 days at 30°C with agitation (200 rpm). The whole culture broths were directly analyzed by LC-ESI-MS as follows: Waters ACQUITY UPLC system equipped with an SQ Detector2 and ACQUITY PDA Detector (Japan); Mightysil RP-18GP Aqua column (150 mm L × 2.1 mm ID, 3 μm, KANTO CHEMICAL Co., Inc., Japan); flow rate, 0.2 mL min^{-1}; temperature, 40°C; mobile phase A, water containing 0.1% formic acid, mobile phase B, methanol containing 0.1% formic acid; gradient conditions, 15% B, 0–2 min; 15–85% B, 2–40 min; detection, ESI-negative mode; injection volume, 10 μL.

pC35Δorf3, in which the *orf3* gene was in-frame deleted, was constructed using Red/ET recombination.[20] In brief, DNA fragments containing a Km-resistance gene cassette (Gene Bridges GmbH, Germany) flanked with FRT sites and 50-bp homologous arms,

whose sequences were identical to the target regions, were amplified by PCR with the primers
5′-CCTGGTGTGGTGTGAATAGACCGCAGTCACGAAGGAGGTATGACCGGATGGA
ATTAACCCTCACTAAAGGGCGGC-3′ and
5′-CACCTGCCGCCTGTCCCGGGTGAGCTGTCATGTCACGCCCCACTCGGGGTAAT
ACGACTCACTATAGGGCTCG-3′. The amplified DNA fragments were used to transform *E. coli* XL1-Blue MRF′ harboring the pC35 cosmid and pRedET plasmid. Gene-disruption in Km-resistant colonies was checked by PCR using two primers,
5′-GCGTAATGCGAGGTTGGGTCTATCGAAGG-3′ and
5′-CGCATCGACTGGGCCGTGGATCTC-3′. The selection marker in the obtained cosmid was removed with FLP-recombinase and the deletion was confirmed by PCR with the primers used to check gene disruption as described above.

Preparation and enzyme assay of recombinant enzymes of Orf1, Orf2, and dihydrolipoyl dehydrogenase

A DNA fragment containing both the *orf1* and *2* genes was amplified by PCR using genomic DNA of *Kibdelosporangium* sp. as a template and the following primers;
5′-CCTATCGGCATATGGTGCTTGGCCGCCGATTCGAC-3′ and
5′-ATATAAGCTTCCGGCTACCCGACCATGATCAGACTGTC-3′. To amplify the

dihydrolipoyl dehydrogenase gene of *Kibdelosporangium* sp., the primers 5′-AAGTGACCATATGGCCACAGTTGACGCGCGCGTA-3′ and 5′-TATAAGCTTGGCTGACTTGGTGCCGACGGTCAG-3′ were used. Restriction sites (underlined) were introduced into the *N*- and *C*-terminal regions. The PCR products were cloned into the *Nde*I-*Hin*dIII sites of the pET28a vector (Merck KGaA, Germany). Each of the obtained plasmids was introduced into *E. coli* BL21(DE3) (F^-, *dcm*, *omp*T, *hsd*S(r_B^- m_B^-), *gal*, λ(DE3); NIPPON GENE CO., LTD, Japan). A liquid culture of each transformant in LB supplied with kanamycin (30 μg mL^{-1}) was induced by adding 0.5 mM IPTG when the optical density at 600 nm reached about 0.6. The cultivation was continued for an additional 16 h at 20°C. Purification of each recombinant protein was carried out as follows. After the cells were disrupted with an ultrasonic disruptor (TOMY, UD-200), the His-tag fused proteins were purified with Ni-NTA column chromatography (QIAGEN K.K., Japan) and their purities were analyzed by SDS-PAGE on 10% gels (Supplementary Figure 2). The proteins were visualized by Coomassie brilliant blue staining and the protein concentrations were determined by the Bradford method with bovine serum albumin as a standard.

Co-purified recombinant Orf1 (approximately 0.42 μM) and Orf2 (approximately 2.3 μM), and purified recombinant dihydrolipoyl dehydrogenase (1.9 μM) were incubated with CoA (1.0 mM), sodium phenylpyruvate (2.0 mM), $MgCl_2$ (1.0 mM), thiamine pyrophosphate (0.2 mM), NAD^+ (2.5 mM), and dithiothreitol (0.6 mM) in potassium

phosphate buffer (50 mM, pH 8.0) at 30°C for 16 h. The reaction product was directly analyzed by LC-ESI-MS as follows: Waters ACQUITY UPLC system equipped with an SQ Detector2 and ACQUITY PDA Detector; XBridge BEH C8 column (150 mm L × 2.1 mm ID, 2.5 μm, Waters); flow rate, 0.2 mL min^{-1}; temperature, 35°C; mobile phase A, 100 mM NH_4COOH containing 5 vol% methanol, mobile phase B, 100 mM NH_4COOH containing 50 vol% methanol; gradient conditions, 10% B, 0–5 min; 10–100% B, 5–40 min; detection, ESI-negative mode; injection volume, 10 μL. The fractionated product was also analyzed by high resolution-ESI-FT-MS (Exactive, Thermo Fisher Scientific Inc.).

Preparation and assay of recombinant L-cysteic acid synthase (Orf3) enzyme

To express maltose binding protein-fused L-cysteic acid synthase (MAL-CS), the primer pair pMal_CS_Fw: 5′-GTATGACCATATGTGCGCTCGGCGGCATTACTCGATC-3′ and pMal_CS_Rv: 5′-AGATAAGCTTCACGCCCCACTCGGGGTGGCTACGTAG-3′ was used. The PCR product was cloned into the *Nde*I-*Hin*dIII site of pMal-c5X (New England Biolabs) to construct pMal-CS. Culture and purification conditions were the same as those described above except for the use of amylose affinity column chromatography (New England Biolabs) for purification. The purified recombinant Orf3 was treated with Factor Xa protease (New England Biolabs) according to the manufacturer's protocol to remove MBP. To eliminate the MBP released, the reaction mixture was applied to an amylose

affinity column again. Then, the flow-through fraction containing the enzyme was mixed with 10 volumes of Tris·HCl buffer (20 mM, pH 8.0) containing 6 M urea and DTT (20 mM), and incubated at 37°C for 2 h. The solution was diluted ten times with refolding buffer; Tris·HCl buffer (50 mM, pH 8.0) containing KCl (1 M), glycerol (20 %), $MgCl_2$ (20 mM), and PLP (10 μM), and stirred at 4°C for 16 h. The enzyme solution was concentrated to 2 mg mL^{-1} with Amicon Ultra filter units (Merck).

The refolded recombinant Orf3 enzyme (8.8 μM) was incubated with PLP (10 μM), sodium sulfite (5 mM), *O*-phospho-L-serine (10 mM), and KCl (100 mM) in Tris·HCl buffer (100 mM, pH 7.0) at 30°C for 15 min. The reaction product was directly analyzed by LC-ESI-MS as follows: Waters ACQUITY UPLC system equipped with an SQ Detector2 and ACQUITY PDA Detector; Scherzo SM-C18 column (150 mm L × 2 mm ID, 3 μm, Imtakt Co., Japan); flow rate, 0.2 mL min^{-1}; temperature, 40°C; mobile phase A, water containing 0.1% formic acid, mobile phase B, acetonitrile containing 0.1% formic acid; 10% B isocratic conditions; detection, ESI-positive mode; injection volume, 10 μL.

Preparation and assay of recombinant Orf4 enzyme

The first C-A-domains and PCP-domain in module 1 of Orf4 were expressed separately. The former was amplified by PCR with two primers, 5′-AAAAAACATATGGACGCGGTCGCGGCCCA-3′ and

5′-TGGCGTAAGCTTTCAGCGGGAAGGACTGGT-3′. The latter was amplified with the primers 5′-AAAAAACATATGCCCGTCACCAGTCCTT-3′ and 5′-AAAAAGCTTGCCGACCAGCAGCGCGAT-3′. Each amplified fragment was cloned into the *Nde*I-*Hin*dIII site of pET28a to construct the plasmids pOrf4-CA and pOrf4-PCP, to express recombinant enzymes fused with an *N*-terminal His-tag. For activation of the PCP-domain, the plasmid (pACYC-Sfp) expressing phosphopantetheinyl transferase of *Bacillus subtilis* (Sfp) was constructed by digesting the pET-Sfp plasmid[21] with *Nde*I and *Xho*I and then cloning the *sfp* fragment into the *Nde*I-*Xho*I site of pACYCDuet-1 (Merck). Each plasmid was used to transform *E. coli* BL21(DE3) for protein overexpression. Purification procedures were the same as those described above.

The recombinant C-A-domains (2.9 μM) and recombinant PCP-domain (82 μM) were first incubated with L-valine (1 mM), ATP (1 mM), $MgCl_2$ (10 mM), and DTT (2 mM) in HEPES buffer (50 mM, pH 7.0) at 37°C for 30 min, before phenylacetyl-CoA (approximately 0.1 mM) prepared from the in vitro enzyme assay and additional recombinant C-A-domains (2.9 μM) were added and the reaction was incubated at 37°C for 30 min. The reaction product was precipitated by adding four volumes of acetone to the reaction and keeping it at −80°C for 1 h. After centrifugation, the precipitates were dissolved in 100 μL of 0.1 M KOH and incubated at 70°C for 10 min. After adding ethanol into the solution (final concentration 40%) and centrifugation, the product was recovered

and analyzed with a Waters ACQUITY UPLC system equipped with an SQ Detector 2 and ACQUITY PDA Detector. The analytical conditions were as follows; InertSustain C18 column (150 mm L × 2.1 mm ID, 3 μm GL Sciences Inc., Japan); flow rate, 0.2 mL min^{-1}; temperature, 40°C; mobile phase A, water containing 0.1% formic acid, mobile phase B, acetonitrile containing 0.1% formic acid; gradient conditions, 20% B, 0–2 min; 20–80% B, 2–35 min; detection, ESI-negative mode; injection volume, 10 μL.

Accession Codes

The nucleotide sequences reported here have been submitted to the DDBJ/GenBank/EBI Data Bank under accession nos. LC217607 (*orf1* to *orf8*) and LC223607 (dihydrolipoyl dehydrogenase gene).

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website

Acknowledgments

This study was supported by Grants-in-Aid for Research on Innovative Areas from MEXT, Japan (JSPS KAKENHI Grant Number 16H06452) and Grants-in-Aid for Scientific Research from JSPS (15H03110) to T. Dairi.

The authors declare no competing financial interest.

FIGURE LEGENDS

Figure 1. Chemical structures of JBIR-78 and -95.

JBIR-78 (**1**, upper) and -95 (**2**, lower).

Figure 2. JBIR-78 (1) and -95 (2) biosynthetic gene cluster.

The cloned DNA fragment (35,856 bp) contained 9 *orf*s. Orf4, 5, and 6 (yellow) were NRPSs and their domain architectures are shown. A, adenylation domain (red); C, condensation domain (blue); E, epimerization domain (green); PCP, peptidyl carrier protein domain (gray); TE, thioesterase domain (black). They activated and condensed, in order, L-Val, L-Asp (or L-cysteic acid), L-Ala, L-Leu, Gly, L-Ala, and L-Phe. The C-domain in module 1 condensed L-Val-PCP and phenylacetyl-CoA, which was supplied by Orf1 and 2 from phenylpyruvate, CoA, and NAD^+, to form phenylacetylated L-Val-PCP.

Figure 3. Heterologous expression of the gene cluster.

The broths of the transformants and authentic standards were analyzed by LC-MS monitoring at *m/z* 824 ($[M-H]^-$ of **1**, traces (a)−(e)) or 860 ($[M-H]^-$ of **2**, traces (f)−(j)). (a) Standard of **1**, (f) Standard of **2**, (b) and (g) recombinant cells harboring pOJ446 (empty

vector), (c) and (h) recombinant cells harboring pC35, (d) and (i) recombinant cells harboring pC35Δorf3 without L-cysteic acid, (e) and (j) recombinant cells harboring pC35Δorf3 with L-cysteic acid.

Figure 4. LC-MS analysis Orf1, Orf2, and the dihydrolipoyl dehydrogenase reactions. The LC-MS profile was monitored at *m/z* 884 ($[M-H]^-$ of phenylacetyl-CoA). (a) Reaction products formed with Orf1 and Orf2. (b) Reaction products formed with dihydrolipoyl dehydrogenase. (c) Reaction products formed with Orf1, Orf2, and dihydrolipoyl dehydrogenase.

Figure 5. LC-MS analysis of the Orf3 reaction. The LC-MS profile was monitored at *m/z* 170 ($[M+H]^+$ of cysteic acid). (a) Standard of L-cysteic acid. (b) Reaction products formed with boiled Orf3. (c) Reaction products formed with refolded Orf3.

Figure 6. LC-MS analysis of the Orf4 reaction. The LC-MS profile was monitored at *m/z* 234 ($[M-H]^-$ of phenylacetyl valine). (a) Standard of phenylacetyl valine. (b) Reaction products formed with the recombinant PCP-domain. (c) Reaction products formed with recombinant C-A- and PCP-domains.

Table 1. Deduced functions of the Orfs.

Orf	Amino acids (no.)	Proposed function
1	677	pyruvate dehydrogenase complex, E1β and E2 units
2	312	pyruvate dehydrogenase complex, E1α unit
3	425	cysteic acid synthase
4	2565	NRPS (C-A-PCP-C-A-PCP-E)
5	2540	NRPS (C-A-PCP-C-A-PCP-E)
6	3825	NRPS (C-A-PCP-C-A-PCP-E-C-A-PCP-TE)
7	69	MbtH domain protein
8	406	cytochrome P450
9	223	hypothetical protein

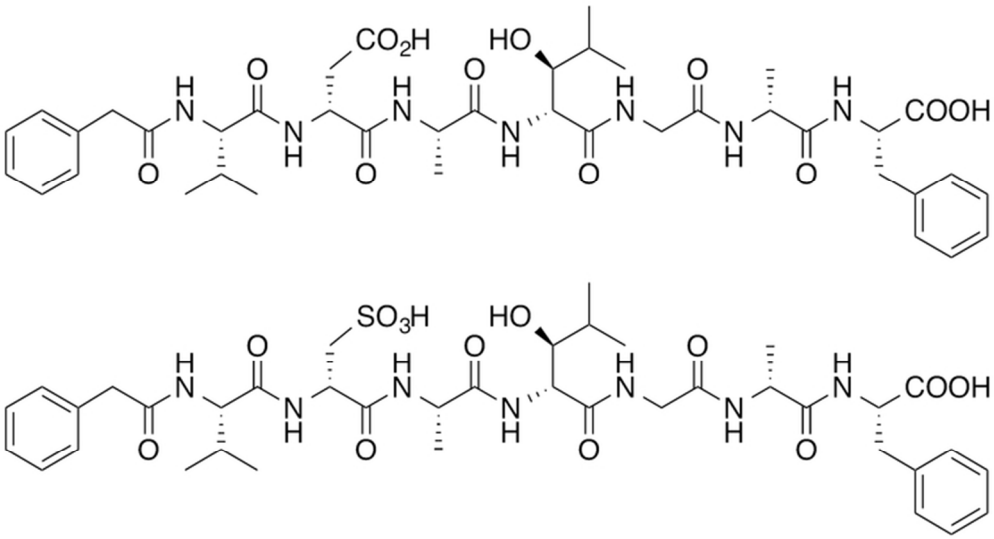


Figure 1

32x17mm (600 x 600 DPI)

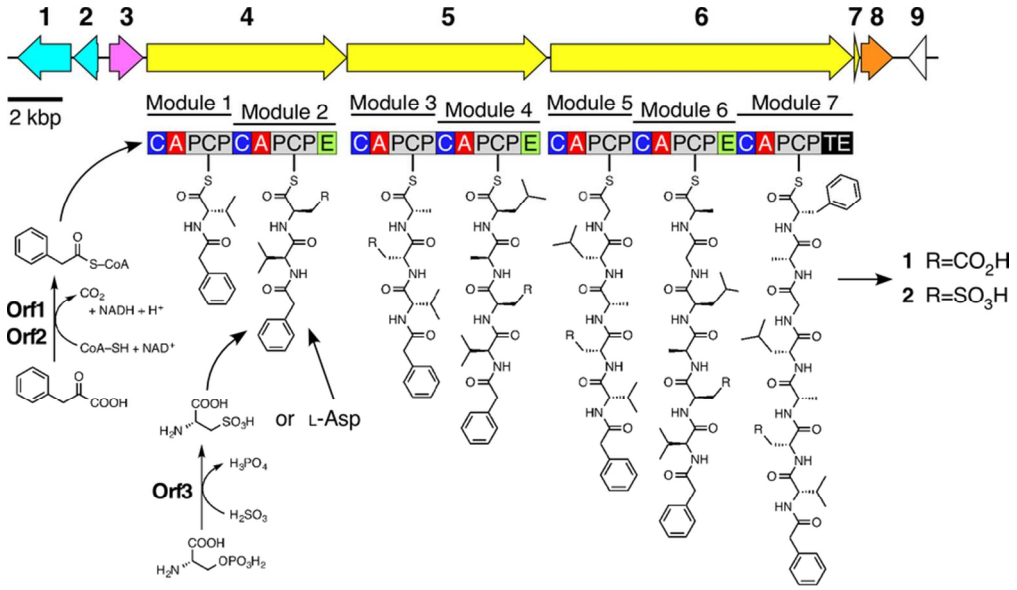


Figure 2

82x48mm (300 x 300 DPI)

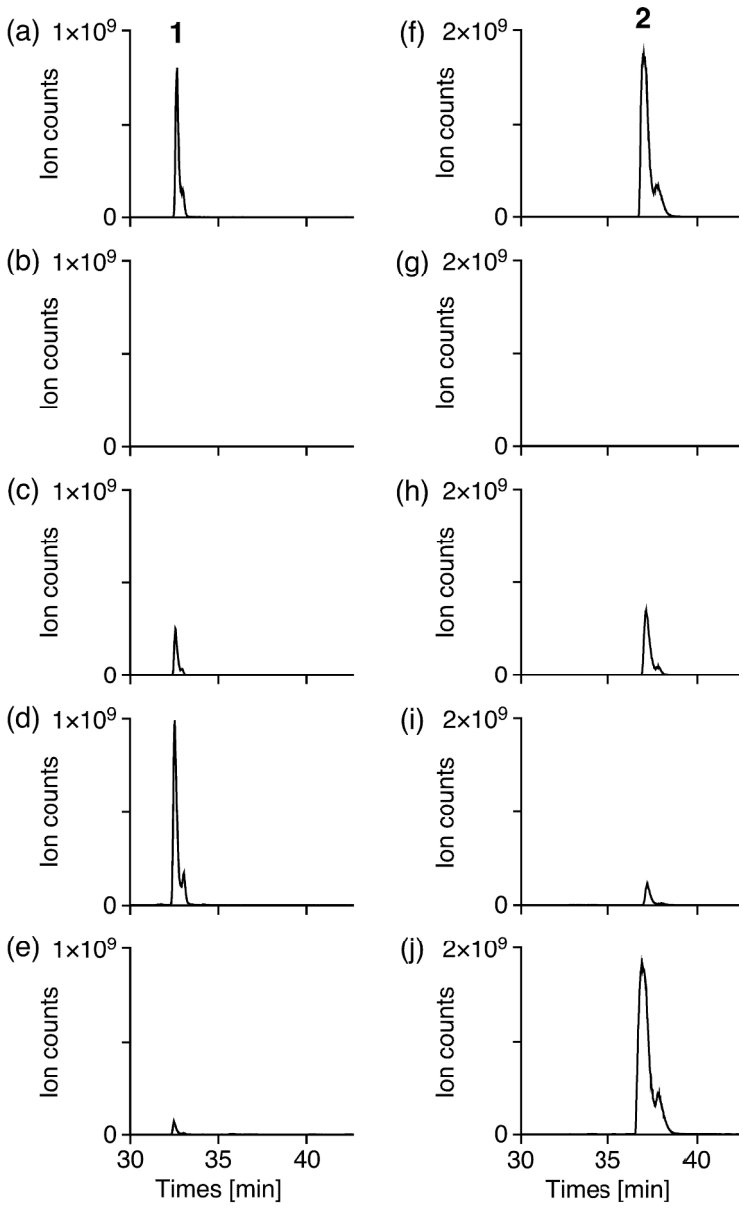


Figure 3

98x161mm (600 x 600 DPI)

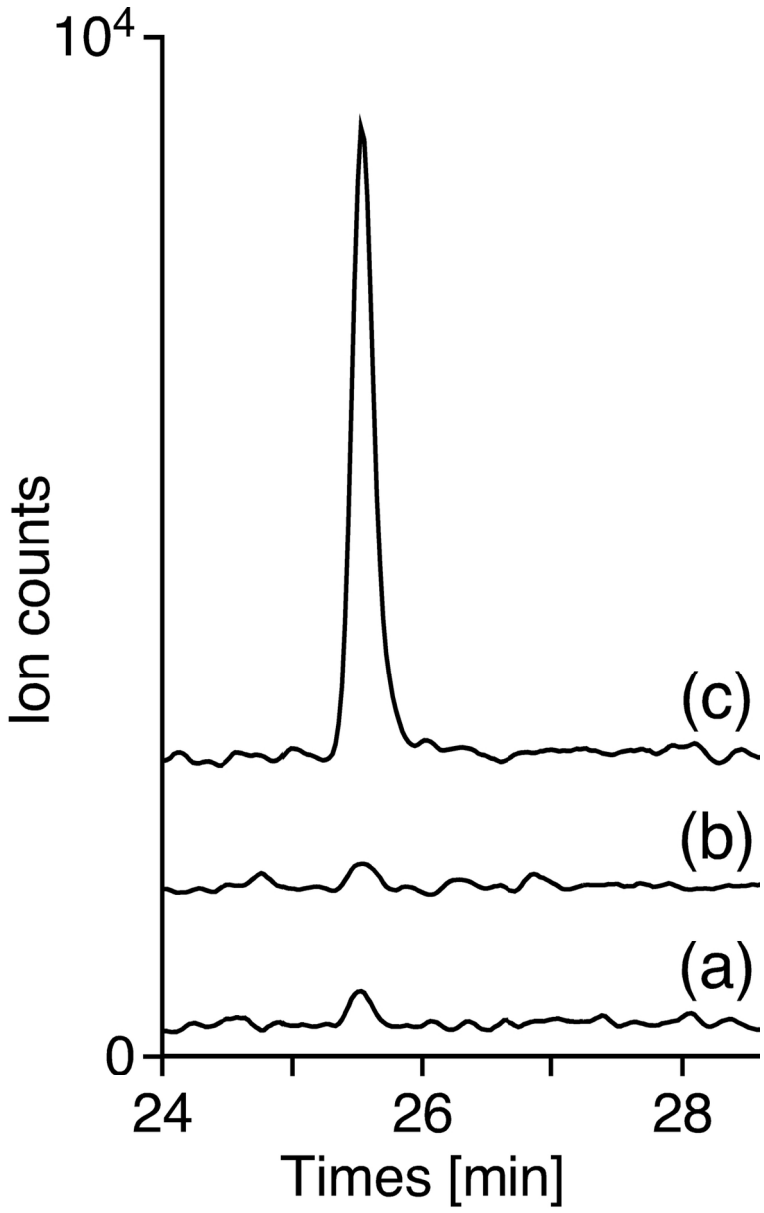


Figure 4

55x86mm (600 x 600 DPI)

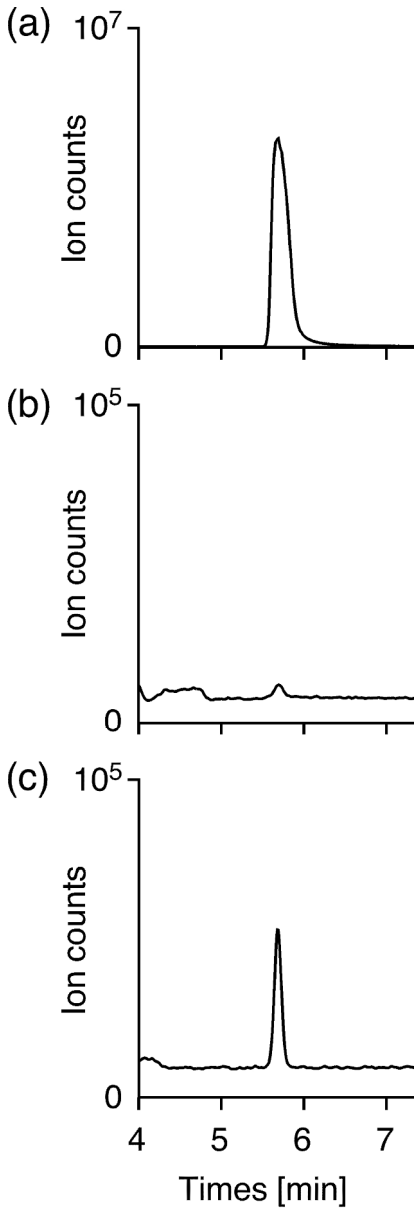


Figure 5

103x305mm (600 x 600 DPI)

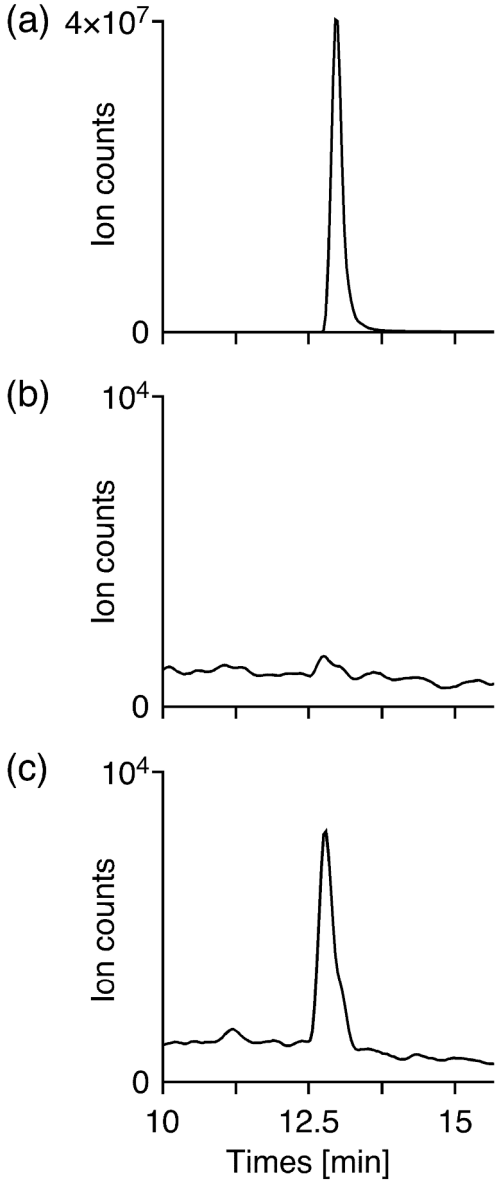
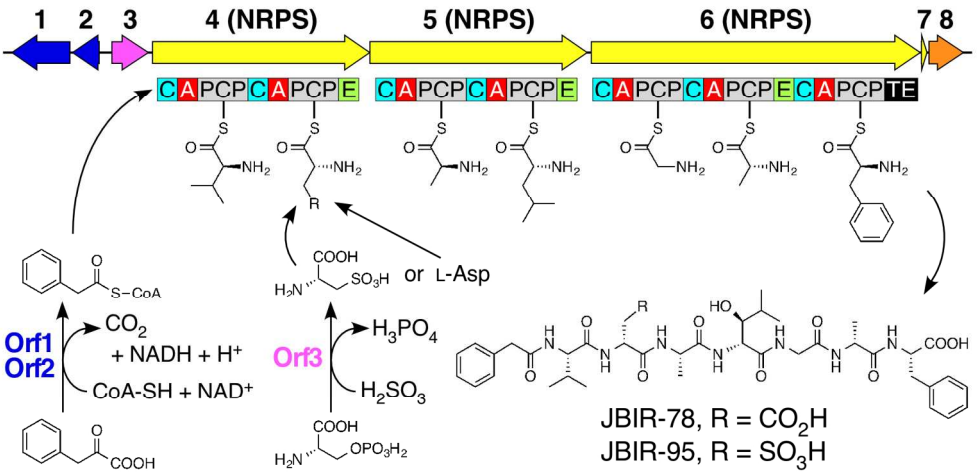


Figure 6

93x220mm (600 x 600 DPI)



TOC

160x80mm (300 x 300 DPI)