A peptide ligase and the ribosome cooperate to synthesize the peptide pheganomycin.

Motoyoshi Noike,[1] Takashi Matsui,[2] Koichi Ooya,[1] Ikuo Sasaki,[1] Shouta Ohtaki,[2] Yoshimitsu Hamano,[3] Chitose Maruyama,[3] Jun Ishikawa,[4] Yasuharu Satoh,[1] Hajime Ito,[1] Hiroyuki Morita[*2] & Tohru Dairi[*1]

[1] Graduate School of Engineering, Hokkaido University, Hokkaido 060-8628, Japan,

[2] Institute of Natural Medicine, University of Toyama, 2630-Sugitani, Toyama 930-0194, Japan,

[3] Department of Bioscience, Fukui Prefectural University, Fukui 910-1195, Japan,

[4] National Institute of Infectious Diseases, Tokyo 162-8640, Japan.

[*]**Contact Information.** Address correspondence to: Tohru Dairi, Graduate School of Engineering, Hokkaido University, Hokkaido 060-8628, Japan. Tel: +81-11-706-7815; Fax: +81-11-706-7118; E-mail: dairi@eng.hokudai.ac.jp; Hiroyuki Morita, Institute of Natural Medicine, University of Toyama, 2630-Sugitani, Toyama 930-0194, Japan. Tel: +81-76-434-7625; Fax: +81-76-434-5059; E-mail: hmortia@inm.u-toyama.ac.jp.

Abstract

Peptide antibiotics are typically biosynthesized by one of two distinct machineries in a ribosome-dependent or -independent manner. Pheganomycin (PGM) (**1**) and related analogues consist of the nonproteinogenic amino acid (*S*)-2-(3,5-dihydroxy-4-hydroxymethyl)phenyl-2-guanidinoacetic acid (**2**) and a proteinogenic core peptide, making their origin uncertain. We report the identification of the biosynthetic gene cluster from *Streptomyces cirratus* responsible for PGM production. Surprisingly, the cluster contains a gene encoding multiple precursor peptides along with several genes plausibly encoding enzymes for the synthesis of amino acid **2**. We identified *pgm1*, which has an ATP-grasp domain, as potentially capable of linking the precursor peptides with **2**, and validate this hypothesis using deletion mutants and in vitro reconstitution. We document PGM1's substrate permissivity, which could be rationalized by a large binding pocket as confirmed via structural and mutagenesis experiments. This is the first example of cooperative peptide synthesis achieved by ribosomes and peptide ligases using a peptide nucleophile.

Introduction

Peptide antibiotics form a large subclass of natural products, including clinically important pharmaceuticals such as vancomycin and daptomycin [1]. Recent progress in genome sequencing techniques has uncovered the presence of novel genes responsible for the biosynthesis of this category of compounds. We often categorize naturally occurring peptides by their biosynthetic machineries: either ribosomally synthesized and post-translationally modified peptides (RiPPs) [2] or nonribosomal peptides (NRPs). Ribosomal peptide and NRP synthetases (NRPSs) using thiotemplate assembly logic produce RiPPs and NRPs, respectively [3]. In addition to these two machineries for peptide/amide-bond formation, 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) have been shown to catalyze amide-bond formations [4-6].

Streptomyces cirratus produces structurally unique pheganomycins (PGMs), which show anti-mycobacteria activity [7]. PGM (**1**) contains two nonproteinogenic amino acids [(*S*)-2-(3,5-dihydroxy-4-hydroxymethyl)phenyl-2-guanidinoacetic acid (**2**) and (*S*)-2-amino-3,3-dimethylbutanoic acid (L-*tert*-Leu)] and two proteinogenic L-amino acids (Asn and Lys) in the order of (*N*-terminus)-**2**-Asn-L-*tert*-Leu-Lys-(*C*-terminus) (Fig. 1a). PGM-D, PGM-DR and PGM-DGPT have the additional amino acids Asp, Asp-Arg and Asp-Gly-Pro-Thr, respectively, at the *C*-terminus of **1**. Though the L-*tert*-Leu residue is nominally non-proteinogenic, it could be formed via methylation of the Val residue by a radical *S*-adenosyl methionine (SAM) enzyme, as shown in bottromycin biosynthesis [8, 9], suggesting that **2** is the only true nonproteinogenic component. Vancomycin, which is known to be biosynthesized by NRPS machineries [10], contains (*S*)-2-amino-2-(3,5-dihydroxyphenyl)acetic acid (**3**); because PGMs also contain the structurally-related **2**, PGMs could logically be synthesized by NRPS. However, the peptide chain length differences and the type of *C*-terminal amino acid sequences in PGMs defy this expectation, because the substrate specificities of the A-domain of NRPSs are strict and there are few examples of this type of diversity in peptides synthesized by NRPS. However, the ribosomally synthesized

cyanobactins provide one example of *C*-terminal amino acid diversity [11]. In sum, the origins of the PGMs is unclear.

To provide insights into this question, we report the biosynthetic machinery responsible for the unique structural features of PGMs. We identified the biosynthetic gene cluster responsible for **1** synthesis, which contained a gene encoding a precursor peptide including two core peptides, several genes plausibly encoding enzymes for **2** biosynthesis and tailoring reactions. We also identify *pgm1*, containing an ATP-grasp domain, as necessary for PGM productivity in cells, while recombinant PGM1 catalyzed the ATP-dependent peptide bond formation between **2** and representative peptide substrates. PGM1 accepted a variety of peptides as the nucleophile and we rationalize this flexibility via the crystal structure and mutational analyses of PGM1. This is the first example of cooperative peptide synthesis achieved by ribosomes and peptide ligases using a peptide as a nucleophile.

Results

Identification of the biosynthetic gene cluster for 1

PGMs have the **2** moiety, probably derived from **3,** at the *N*-terminus. In vancomycin and teicoplanin biosynthesis, four enzymes, DpgA-DpgD catalyze the formation of 2-(3,5-dihydroxyphenyl)-2-oxoacetic acid, the deamination product of **3** [12]. We therefore searched for orthologs of Dpg(s) in a draft genome database. Using BLAST search, we identified three ORFs (PGM11, 10 and 9), which have high similarities to DpgA, B and C, respectively. In the flanking regions of the genes, we identified 18 additional genes (a total of 21 genes) that should form a cluster (Fig. 1b and Supplementary Results, Supplementary Table 1) based on the good agreement between the predicted functions of some genes and the hypothetical biosynthetic pathway based on the PGM structures. We anticipate PGM12, PGM8, PGM7 and PGM4—which have high similarities to amidinotransferases, aminotransferases, cytochrome P450s and methyl transferases—participate in the modifications of **3**. PGM3, similar to radical SAM enzymes, may be responsible for the methylation of the Val moiety to yield L-*tert*-Leu. However, we could not find a gene that could encode a large NRPS with enough multi-modules to condense the eight amino acids,

although three genes encoding stand-alone NRPSs, PGM21 with an A-domain, PGM20 with a T-domain, and PGM19 with a C-domain, were present. An ORF with an ATP-grasp domain (PGM1) (Supplementary Fig. 1) also existed in the cluster. In the upstream and downstream regions of *pgm1*, there are two genes encoding an aldose 1-epimerase and a dehalogenase/phosphatase, though the utility of these genes in the construction of PGM is not obvious. We therefore examined the possibility that ribosome supplied the proteinogenic amino acids in PGMs because a putative peptidase gene (*pgm14*, 1159 amino acids), which is usually present in the biosynthetic gene clusters of ribosomally synthesized peptide antibiotics, exists in the cluster. We carefully searched for a small ORF that could encode a core peptide with an NVKDGPT and/or NVKDR sequence and identified an ORF (PGM2) composed of 38 amino acids (MEREIVWTEIEESDLAAVVSAS**NVKDGPT**VSSS**NVKDR)** in the cluster (Fig. 1b). This potential precursor peptide would clearly explain the *C*-terminal diversity of PGMs including differences in the length and kinds of amino acids (underlined in the above sequences). Based on these results, we hypothesized that NRPS (PGM19 to 21) or PGM1 catalyzes the peptide bond formation between **2** and the peptides, although there are no reports demonstrating that these machineries synthesize a peptide backbone.

Gene disruptions

To test the involvement of these genes in PGM biosynthesis, we first deleted the *pgm2* gene encoding the precursor peptide by a gene replacement technique [13]. We successfully obtained *pgm2*-deletion mutants (Supplementary Fig. 2). After cultivation of the mutants in liquid medium, we monitored the production of PGMs by LC-ESI-MS analysis and confirmed that the mutants produced no PGMs (Supplementary Fig. 2c), supporting our proposal that ribosome-dependent machinery supplied the proteinogenic precursor peptides.

We then examined whether the NRPSs participates in the biosynthesis of **1**. We disrupted the *pgm19* gene by the same method used for *pgm2* disruption and examined the production of **1**. As shown in Supplementary Fig. 3c, the mutant maintained the ability to produce **1**, indicating that

PGM19 has no connection with **1** biosynthesis.

Finally, we disrupted *pgm1* by the method described above. The mutant did not produce **1** (Supplementary Fig. 4c), showing that PGM1 is essential for **1** biosynthesis.

PGM1 catalyzed the peptide bond formation

The ATP-grasp domain is commonly found in enzymes from the ATP-dependent (ADP-forming) carboxylate-amine/thiol ligase superfamily [14, 15, 16, 17, 18]. Within this group, aminoacyl phosphate is synthesized as a reaction intermediate in the presence of ATP and Mg^{2+}. Although there are no examples of the use of a peptide as the nucleophile in the reaction, we examined this possibility through *in vitro* experiments.

We prepared a recombinant *C*-terminal His-tagged PGM1 protein (Supplementary Fig. 5). Gel filtration analysis indicated a molecular mass of 99.8 kDa, suggesting that PGM1 forms a homodimer structure (Supplementary Fig. 5b). The homodimer structure was also supported by a crystal structure (see below). We used PGM1 to catalyze the reaction of the substrates, (*S*)-2-guanidino-2-phenylacetic acid (**4**) (Fig. 2a) and NVKDR. In the presence of ATP and Mg^{2+}, a product was specifically detected by HPLC analysis (Fig. 2b), and its molecular mass was in good agreement with the estimated product by LC-ESI-MS analysis (Supplementary Fig. 6). We purified the product (**5**) and determined its structure to be NVKDR with **4** at the *N*-terminus (Fig. 2c, and Supplementary Figs. 7–14) by NMR analysis. This result confirmed that PGM1 catalyzed peptide bond formation between a nonproteinogenic amino acid and a core peptide.

To examine the activation mechanism of the substrate, we attempted to detect phosphorylated **4** by LC-ESI-MS analysis, but all trials failed despite employing several different reaction and analytical conditions (data not shown). As an alternative approach, we tried detecting the byproduct, ADP, in the reaction mixture containing substrate (**4**) as the sole substrate. As shown in the Supplementary Fig. 15a, we could detect ADP. However, less ADP also formed in the control experiment without **4** probably because of a contamination of a phosphatase in recombinant PGM1. The calculated concentration of ADP specifically formed by PGM1 was 0.048 mM, which

corresponds to almost half concentration of recombinant PGM1 (0.083 mM) used in the experiment. We also quantitatively measured the amounts of consumed substrate (ATP) and formed products (**5**, ADP and Pi) and confirmed these to be stoichiometrically equivalent (Supplementary Fig. 15b). Furthermore, we confirmed no formation of AMP (Supplementary Fig. 15c). Taken together, these data suggest, but do not definitively show, that PGM1 activates its substrate by phosphorylation.

Enzymatic properties of PGM1

We examined the substrate specificity of PGM1 with the compounds shown in Fig. 2a as *N*-terminus substrates. In addition to **4**, PGM1 accepted its enantiomer (*R*)-2-guanidino-2-phenylacetic acid as well as 2-guanidinoacetic acid, and 2-(1-methylguanidino)acetic acid (creatine), all of which possess a guanidino group, with NVKDR as the peptide cosubstrate (Supplementary Figs. 16–18). PGM1 was similarly able to use a variety of peptides as nucleophiles for modification of the *C*-terminus (Fig. 3a, Supplementary Figs. 19–29 and Supplementary Table 2), although PGM1 could not accept Gly or any L-amino acids including Asn (data not shown). Besides the natural substrate peptide NVKDGPT, PGM1 was also able to use the dipeptide NV, the pentapeptide NVKDG and the hexapeptide NVKDGP (Supplementary Figs. 19–22). Considering that the tripeptide NVK and the tetrapeptide NVKD would be the substrates of PGM1 for **1** and PGM-D biosynthesis, we can conclude that PGM1 accepts peptides with one to five amino acids deleted from the *C*-terminus of NVKDGPT. Peptides in which one to three amino acids in NVKDR were replaced with other specific residues were also acceptable (Supplementary Figs. 23–26). Furthermore, PGM1 accepted two peptides composed of amino acids different from those of the natural substrate: aspartame (Asp-Phe methyl ester) and the tetrapeptide MRFA (a commercially available peptide with Met at the *N*-terminus) (Supplementary Figs. 27 and 28, respectively). More unexpectedly and importantly, PGM1 accepted the antibacterial peptide, apidaecin [19] (GNNRPVYIPQPRPPHPRL, 18 aa) (Fig. 3b) and its derivative (RVRRPVYIPQPRPPHPRL, 18 aa) (Supplementary Fig. 29) as substrates. These results unambiguously show that PGM1 can accept a variety of peptides as nucleophiles.

Finally, we investigated the biochemical properties of PGM1 using **4** and NVKDR as the

substrates. Product formation was optimal at 45°C and pH 8.0 (Supplementary Figs. 30 and 31, respectively), and the PGM1 reaction followed Michaelis–Menten kinetics (Supplementary Figs. 32 and 33). Using Hanes–Woolf plots, the *Km* values were 97 ± 4 μM for **4**, and 61 ± 6 μM for NVKDR. The *k*cat value of the reaction was 26 ± 0.5 s^{-1} (Supplementary Table 3).

Structure of PGM1

To understand the structural basis for the remarkable substrate tolerance we observed, we solved the crystal structures of PGM1-AMP complex (Fig. 4a and Supplementary Fig. 34) and apo form (Supplementary Fig. 34) at a resolution of 2.18 Å and 1.96 Å, respectively. For the complex, we used adenylyl imidodiphosphate (AMPPNP) as the substrate but the electron densities of the β- and γ-phosphates of AMPPNP in monomers were weak or disordered. The PGM1-AMP complex may therefore either reflect contamination of AMP in AMPPNP (5.5 %) or decomposition of AMPPNP to AMP during the crystallization. Although we tried obtaining complex structures of PGM1 with *N*-terminus and peptide substrates, as well as magnesium and ATP, these conditions were not successful.

PGM1 consisted of four domains N, A, B and C, in which the N, A and C domains clustered to form a central core structure, while the B domain solely located outward to form the upper structure of the central core in a manner similar to that of the ATP-grasp superfamily enzymes [14]. The structures consisted of four monomers A, B, C and D, in which A/B and C/D form two sets of biologically active, symmetric dimers (Supplementary Fig. 34). A structure-based similarity search [20] revealed that the overall structures of PGM1 exhibited root mean square deviation (r.m.s.d.) values of 3.7 Å, 3.8 Å, 4.2 Å and 4.7 Å from those of glycinamide ribonucleotide synthetase (PurD) of *Aquifex aeolicus* (PDB entry 2YW2) [21], N^5-carboxyaminoimidazole ribonucleotide synthetase (N^5-CAIR) of *Aspergillus clavatus* (PDB entry 3K5H) [22], glycinamide ribonucleotide transformilase (PurT) of *Escherichia coli* (PDB entry 1KJ8) [23], and L-amino acid ligase (LAL) of *Bacillus licheniformis* (PDB entry 3VOT) [24], respectively. In the PurD forms, the closed form represents the initial phosphorylation and second amido-bond formation stages of the enzyme reaction.

Comparisons of the PGM1 structures with the opened form of PurD of *E. coli* (PDB entry 1GSO) [25] and the closed form of PurD of *A. aeolicus* (PDB entry 2YW2) [21] revealed that the structure of PGM1 obtained here was in a closed conformation (Supplementary Fig. 35).

The central core and upper structures also formed a large cavity at the center of the PGM1 structure and a long cleft connecting from the large cavity lying at the boundary of the N and B domains (Fig. 4b). The ATP-binding site was located deep inside the large cavity, and Arg335, corresponding to the putative catalytic Arg in PurDs [21], sat at the intersection of the ATP-binding and the *N*-terminus substrate-binding sites by protruding its side-chain toward the ATP-binding site. PGM1 possessed most of the conserved residues lining the ATP-binding site, including Arg335, in a location and orientation similar to those of PurD enzymes (Supplementary Fig. 36). However, the large cavity and the cleft created an unusual active site in PGM1 (Supplementary Figs. 37 and 38).

Model building and structure-based mutagenesis of PGM1

To further clarify the intimate structural details of substrate and product specificities of PGM1, we carried out model building studies and investigated the effect of the mutagenesis of key amino acids on enzyme activity using **4** as the *N*-terminus substrate, and the NVKDR and NVKDGPT peptides as the cosubstrates (Supplementary Table 4). A PGM1-ATP-**2** ternary complex model structure predicted that the carboxyl group of **2** would be located near the γ-phosphate of ATP and the side chain amidino group of the putative catalytic residue Arg335 (Fig. 4c). Moreover, the amidino group of **2** occupied near the side chains of Glu255, Asp318, Ser316, Thr337 and Ser339. Indeed, the R335Q, R335N and R335M substitutions completed loss of enzyme ability. In addition to these mutants, the E255D, E255S, S316D and S339A substitutions at the *N*-terminus substrate-binding site resulted in loss of enzyme activity. Thus, we suspect PGM1 initiates the reaction by a nucleophilic attack of the carboxylic anion of the *N*-terminus substrate on the γ-phosphate of ATP.

Conversely, a structure model of PGM1-ADP-phosphorylated **2**-NVKDR peptide quaternary complex predicted that the *N*-terminal Asn of the peptide places near the phosphate moiety of the phosphorylated **2** (Fig. 4d). To bind here, Ser190 and Gly191-located at the entrance to

the cleft-would be well placed to recognize the side chain of the *N*-terminal Asn of the peptide. In contrast, the *C*-terminal Arg of the peptide would be located near Ala14, Glu15 and Thr17 on the cleft. A similar case was found in the quaternary model structure of the PGM1-ADP-phosphorylated **2**-NVKDGPT peptide complex, in which the *C*-terminal Thr located near Ser190 and Gln186 (Fig. 4e). In line with these models, S190L and S190A substitutions showed a 77% and 43% decrease in the **5**-forming activity, respectively, depending on the bulkiness of the mutated residues. Furthermore, when we used NVKDGPT as the counter substrate instead of NVKDR, the S190L and S190A substitutions were more effective (81% and 45% decrease). More importantly, the mutagenesis studies also revealed that the small-to-large A14R and A14Q substitutions near the end of the cleft showed a 49% and 47% decrease in the **5**-forming activity and a 45% and 46% decrease in the **4**-primed NVKDGPT forming activity, respectively. In contrast, the A14L substitution, using a less bulky residue than Arg and Gln, had almost no effect on the enzyme's activity. Thus, the steric contraction at this position significantly interrupted enzyme activity.

Discussion

In this study, we revealed that two distinct machineries-a new type of peptide ligase and ribosomes-cooperated to synthesize **1**. PGM1 uses an ATP-grasp domain to catalyze peptide bond formation via activation of the nonproteinogenic phenylglycine derivatives, likely by phosphorylation and successive nucleophilic attack by the ribosomally produced peptides. Although we can find many enzymes with an ATP-grasp domain in genomic databases by searching conserved domains such as Pfam [26], most of their functions remain unknown. Among the enzymes identified, approximately 20 catalyze an ATP-dependent phosphorylation of a carboxyl group of one substrate and a nucleophilic attack by an amino or imino group nitrogen of another substrate. In these reactions, the first substrate is a protein, a short peptide, an amino acid, or an organic/inorganic acid, but an amino acid, a thiol group (CoA), a biotinylated protein, a ribonucleotide derivative, an inositol derivative, or an ammonium ion is the second substrate [14]. Therefore, PGM1 is the first example in which a peptide is used as the nucleophile and, moreover, of ribosomes and peptide

ligases cooperatively synthesizing a single peptide backbone (Fig. 5). To date, there are other examples of machinery that combine two proteins/peptides such as ubiquitin ligase found in the ubiquitin-proteasome complex [27], sortase [28], which participates in peptidoglycan biosynthesis, and intein, which is active in protein splicing [29]. However, PGM1 uses a completely distinct mechanism from these enzymes, and thus defines a new type of machinery.

The crystal structure of PGM1 demonstrated that PGM1 has an unusually large active site consisting of a large cavity and a long cleft. This type of large active site has never been reported in the ATP-grasp superfamily enzymes. The site-directed mutagenesis studies coupled with docking simulations suggested that both the amidino and terminal carboxyl groups of the guanidinoacetic acid scaffold of the substrate molecules might be responsible for the binding to PGM1. The findings would explain the *N*-terminus substrate specificities of PGM1, which no longer accepted **3**, 3-guanidinopropanoic acid, 4-guanidinobutanoic acids, and L-amino acids, because of the loss of both interactions between the amidino or terminal carboxyl group of the substrate and PGM1.

The most interesting functional feature of PGM1 is the remarkable second substrate tolerance. The PGM1 structure and mutagenesis studies suggested that the large cavity at the center of the PGM1 structure and the long cleft connecting the large cavity lying at the boundary of the N and B domains enabled to accept many peptides composed of different lengths and kinds of amino acids as the nucleophiles.

The broad substrate specificity of PGM1 enabled us to synthesize various derivatives of **1,** which may have higher activity against mycobacteria as well as other biological activities. Moreover, as exemplified by apidaecin, the attachment of nonproteinogenic amino acids at the *N*-terminus of various antibiotic peptides by PGM1 may improve the proteolytic stability of short peptides, which are usually unstable in physiological environments because of their susceptibility to degradation. Furthermore, PGM1 may be useful for modifying the *N*-terminus of proteins of interest, since there are many reports demonstrating such trials by chemical and enzyme methods for a diverse range of aims [30-33].

PGM1 orthologs exist in some actinobacteria such as SGM_5893 in *Streptomyces griseoaurantiacus* M045 (50% amino acid sequence identity), SSFG_06970 in *Streptomyces ghanaensis* ATCC 14672 (51%), MCAG_05280 in *Micromonospora* sp. ATCC 39149 (44%), Sare_2923 in *Salinispora arenicola* CNS-205 (43%), Strop_1717 in *Salinispora* sp. CNB-440 (43%), SSMG_07232 in *Streptomyces* sp. AA4 (36%), Francci3_2066 in *Frankia* sp. CcI3 (36%), and SPW_2896 in *Streptomyces* sp. W007 (35%). Of these, the first two strains possess clustered orthologs of *pgm7* to *pgm12* (SGM_5887 to SGM_5892 and SSFG_06973 to SSFG_06978) in the regions near the *pgm1* orthologs. These facts strongly suggest that both strains produce a peptide compound with derivatives of **2** at the *N*-terminus. In contrast, other strains may produce peptides with an amino acid or its derivative at the *N*-terminus, because no orthologs of *pgm7* to *pgm12* exist in the flanking regions of the *pgm1* ortholog. Using these PGM1 orthologs, we may develop a simple synthetic method to produce more diverse peptides.

Methods

General. Sequence analysis of PCR fragments was performed by the dideoxy chain termination method with an automatic DNA sequencer (Li-Cor, model 4000L, Lincoln, NE, USA). Cell disruption was performed with an ultrasonic disruptor (TOMY, UD-200, Tokyo, Japan). Analysis of the samples during protein purification was performed using SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and proteins were visualized by Coomassie brilliant blue staining. Protein concentration was determined by the kit (Sigma-Aldrich, B6916) with bovine serum albumin as a standard. Compound **4** was chemically synthesized from phenylglycines and formamidinesulfinic acid [34]. All the synthetic peptides were obtained from Life Technologies Japan (Tokyo, Japan). The dipeptides and the tetrapeptide (MRFA) were purchased from AnaSpec (Fremont, CA, USA) and Sigma-Aldrich (St. Louis, MO, USA), respectively. Apidaecin and its derivative were kindly provided by Prof. S. Taguchi (Hokkaido University, Sapporo, Japan). Other chemicals were purchased from Sigma-Aldrich and Wako Pure Chemical Industry (Osaka, Japan). Purity of these compounds was examined by HPLC.

Strain and plasmids. *Streptomyces cirratus* strain MD227-A9 was used for the production of PGM and the preparation of genomic DNA.

Draft genome sequences. Draft genome sequences of the *S. cirratus* MD227-A9 genome were determined by Takara Bio Inc. (Shiga, Japan) using the Genome Analyzer IIx system (Illumina, San Diego, CA, USA). Genomic libraries (a few hundred bp of inserted DNA) were constructed with NEBNext DNA Sample Prep Reagent Set 1 (NEB, Ipswich, MA, USA). Using 75 bp paired-end reads, a total of 3,032 million reads were obtained. Using Velvet (version 1.0.18) [35], these were assembled and 181 contigs were constructed (8.8 million base reads and the average length of the contigs was 49,102). Gene prediction was carried out by MetaGeneAnnotator [36].

Gene disruption. The *pgm1*, *pgm2* and *pgm19* genes were disrupted by double-crossover homologous recombination. To construct the gene disruption plasmids, two 2-kb DNA fragments carrying the upstream (*Eco*RI-*Xba*I fragment) and downstream (*Xba*I-*Hin*dIII fragment) regions of the target genes were amplified by PCR with the primer set listed in Supplementary Table 5. After sequence confirmation, the fragments were inserted into the *Eco*RI/*Xba*I and *Xba*I/*Hin*dIII sites of pUC18 in the same direction as in the genomic region. The inserted DNAs were then digested with *Eco*RI and *Hin*dIII, and inserted into the same sites of pWHM3 [37]. The constructed plasmids were used to transform the **1** producer and the thiostrepton-resistant colonies were selected by the same method used for *S. coelicolor* A3(2) [38]. After protoplasting and regeneration of the transformants, thiostrepton-sensitive colonies were collected. Among them, a deletion mutant was selected by PCR analysis and the production of **1** was analyzed by LC-ESI-MS.

Cloning, overexpression and purification of PGM1. *pgm1* was amplified by PCR using gene-specific primers (Supplementary Table 5). After subcloning and sequence confirmation, the DNA fragments obtained by digestion with *Nde*I and *Xho*I were ligated into the same sites of pET21b. *E. coli* BL21 (DE3) harboring the recombinant expression vector was grown in L-broth supplemented with 100 μg/mL ampicillin. The strain was grown at 37°C until the OD600 reached 0.5 and then isopropyl β-D-thiogalactopyranoside (0.5 mM) was added to the culture, followed by additional cultivation at 25°C for 18 h. Recombinant PGM1 was purified according to the

manufacturer's protocol. After the purity of the recombinant PGM1 was checked by SDS-PAGE (Supplementary Fig. 5a), the enzyme was used for the *in vitro* assay. For subunit structure determination, the purified PGM1 was loaded onto a HiLoad 26/60 Superdex 75 pg gel-filtration column (Amersham Biosciences, Piscataway, NJ, USA) and eluted with a buffer containing 50 mM Tris-HCl (pH 8.0) and 150 mM NaCl. The retention volume of PGM1 was compared with those of the marker proteins.

PGM1 *in vitro* assay. The standard assay mixture for PGM1 contained in a final volume of 100 μL, 1 mM of **4**, 1 mM of NVKDR, 50 mM Tris-HCl (pH 8.0), 10 mM ATP, 10 mM Mg^{2+}, and a suitable amount of PGM1. The mixture was incubated at 30°C overnight and the reaction was stopped by the addition of 100 μL of methanol. The products were analyzed by HPLC. The analytical conditions were as follows: Merck Mightisil RP-18GP Aqua column (250 mm × 4.6 mm) (Kanto Chemicals, Tokyo, Japan); mobile phase of acetonitrile in water containing 0.05% trifluoroacetic acid (0–5 min, 5% acetonitrile; 5–20 min, 5–50%); flow rate, 1.0 mL/min; detection, 210 nm.

The steady-state kinetic parameters were determined by fitting to the Michaelis–Menten equation. The 30-min assay was linear up to a protein amount of 10 μg, and no substrate inhibition was observed with **4** or NVKDR up to 1.0 mM of each substrate. The assay mixture for determination of the kinetic parameters of **4** contained in a final volume of 100 μL, 50 mM Tris-HCl, (pH 8.0), 0.5 mM NVKDR, 10 μM to 1 mM **4**, 5 mM ATP, 5 mM Mg^{2+} and 4 μg of PGM1. This mixture was incubated at 30°C for 5 min. Alternatively, when the concentration of **4** was fixed at 0.5 mM, the concentration of NVKDR was varied from 10 μM to 1 mM. Triplicate sets of enzyme assays were performed at each substrate concentration.

To determine the optimum temperature, the assay was conducted from 10 to 60°C. The assay mixture contained in a final volume of 100 μL was 50 mM Tris-HCl, pH 8.0, 50 μM NVKDR, 50 μM **4**, 5 mM ATP, 5 mM Mg^{2+}, and 4 μg of PGM1. The mixture was incubated at 30°C for 5 min. For optimum pH analysis, the assay was conducted from pH 6.5 to 10 using 50 mM MES (pH 6.5 and 7), 50 mM Tris-HCl (pH 7.5 to 9), and 50 mM glycine-NaOH (pH 9 to 10). The assay mixture in each of the buffer contained in a final volume of 100 μL was 50 μM NVKDR, 50 μM **4**, 5 mM ATP,

5 mM Mg^{2+}, and 4 μg of PGM1. The mixture was incubated at 30°C for 5 min. Average value obtained by two independent experiments was shown.

To measure the amounts of consumed substrate (ATP) and formed products (**5**, ADP and Pi) quantitatively, 9 μg of recombinant PGM1 were incubated with 1 mM **4**, 1 mM NVKDR, 50 mM Tris-HCl (pH 8.0), 1 mM ATP and 5 mM Mg^{2+} in a final volume of 100 μL for 30 min at 30°C. After 30 min, the amounts of ATP consumed and ADP formed were measured by HPLC. Triplicate sets of enzyme assays were performed. To detect the byproduct (ADP), 200μg (0.083 mM) of recombinant PGM1 were incubated with 1 mM **4** as the sole substrate. Average value obtained by two independent experiments was calculated. The analytical conditions were as follows: Merck Mightisil RP-18GP Aqua column (250 mm × 4.6 mm); mobile phase of 5% acetonitrile in 0.1 M phosphate buffer, pH 6.8, 10 mM tetrabutylammonium bromide; flow rate, 1.0 mL/min; detection, 254 nm. The released Pi was determined using BIOMOL GREEN, a reagent for phosphate detection (Cosmo Bio Co. Ltd, Tokyo, Japan), according to the manufacturer's protocol.

To determine the conversion ratios, 18 μg of recombinant PGM1 were incubated overnight at 30°C with 1 mM **4**, (*R*)-2-guanidino-2-phenylacetic acid, 2-guanidinoacetic acid, or creatine with 1 mM NVKDR in a final volume of 100 μL buffer containing 50 mM Tris-HCl (pH 8.0), 1 mM ATP and 5 mM Mg^{2+}. The consumed amounts of **4** or (*R*)-2-guanidino-2-phenylacetic acid were measured by HPLC. When 2-guanidinoacetic acid and creatine were used as *N*-terminus substrates, the consumed amounts of NVKDR were measured. Average value obtained by two independent experiments was calculated. The analytical conditions were as follows: Merck Mightisil RP-18GP Aqua column (250 mm × 4.6 mm); mobile phase of 4% acetonitrile in water containing 0.05% trifluoroacetic acid; flow rate, 1.0 mL/min; detection, 210 nm.

Structural analysis of the reaction product. The products formed during the *in vitro* reactions were analyzed by LC-ESI-MS analysis (Waters ACQUITY UPLC equipped with SQD2). The analytical conditions were as follows: InertSustain C18 column (150 mm × 2.1 mm) (GL Sciences, Tokyo, Japan); column temperature, 40°C; detection, positive mode; mobile phase, 0.05% trifluoroacetic acid:methanol 95:5 for 10 min, and a linear gradient of 15:85 for an additional 30

min; flow rate, 0.2 ml/min; cone voltage, 30 V in total and selected ion chromatograms.

When the peptides with internal Lys and Arg, the side chains of which can also be utilized as nucleophile, were used as substrates, the enzyme reaction mixtures were analyzed by LC-ESI-MS/MS (Esquire 4000, Bruker Japan, Yokohama, Japan) using a reversed-phase column (Sunniest RP-AQUA, 3 μm, 100 × 2.0 mm; ChromaNik Technologies, Osaka, Japan) at 30°C, at a flow rate of 0.2 ml/min, with a gradient of acetonitrile in water containing 0.05% (v/v) n-heptafluorobutyric acid and 0.05% (v/v) formic acid run over 20 min [5% (v/v) acetonitrile for 2 min, 5–95% (v/v) acetonitrile for 13 min, and 95% (v/v) acetonitrile for 5 min].

When apidaecin and its derivative were used as the substrates, the assay mixture contained in a final volume of 100 μL, 50 mM Tris-HCl (pH 8.0), 1 mM **4**, 2.5 μM apidaecin or its derivative, 10 mM ATP, 10 mM Mg^{2+} and 4 μg of PGM1. After the reaction was incubated at 30°C overnight, the following products were analyzed by MALDI-TOF-MS analysis (Ultraflex III, Bruker Japan) with alpha-cyano-4-hydroxycinnamic acid as the matrix: apidaecin, $[M+H]^+$ (calculated: 2108.163, observed: 2108.155); the reaction product from **4** and apidaecin, $[M+H]^+$ (calculated: 2283.237, observed: 2283.37); the apidaecin derivative RVR, $[M+H]^+$ (calculated: 2234.326, observed: 2234.324); the reaction product from **4** and the apidaecin derivative RVR, $[M+H]^+$ (calculated: 2409.400, observed: 2409.470).

The reaction products obtained from **4** and the peptide NVKDR as the substrates were fractionated using preparative HPLC. ^{1}H NMR (500 MHz, D_2O, δ): 0.74 (d, *J* = 6.8 Hz, 3H), 0.79 (d, *J* = 6.8 Hz, 3H), 1.42 (m, 2H), 1.61 (m, 2H), 1.67 (m, 2H), 1.73 (m, 1H), 1.75 (m, 1H), 1.81 (m, 1H), 1.91 (m, 1H), 1.93 (m, 1H), 2.74 (dd, *J* = 7.5, 15.7 Hz, 1H), 2.80 (dd, *J* = 8.4, 16.9 Hz, 1H), 2.83 (dd, *J* = 6.6, 15.7 Hz, 1H), 2.92 (dd, *J* = 5.2, 16.9 Hz, 1H), 2.98 (t, *J* = 7.6 Hz, 2H), 3.19 (br t, *J* = 7.6 Hz, 2H), 3.97 (d, *J* = 7.0 Hz, 1H), 4.29 (dd, *J* = 5.9, 8.8 Hz, 1H), 4.31 (dd, *J* = 5.0, 8.7 Hz, 1H), 4.70 (dd, *J* = 5.3, 8.4 Hz, 1H), 4.77 (m, 1H), 5.34 (s, 1H), 7.45–7.49 (m, 5H). ^{13}C NMR (125 MHz, D_2O, δ): 18.21 (*C*H_3), 19.15 (*C*H_3), 22.90 (*C*H_2), 25.12 (*C*H_2), 27.01 (*C*H_2), 28.84 (*C*H_2), 30.80 (*C*H), 30.99 (*C*H_2), 36.46 (*C*H_2), 40.00 (*C*H_2), 41.39 (*C*H_2), 51.14 (*C*H), 51.35 (*C*H), 53.95 (*C*H), 54.41 (*C*H), 59.35 (*C*H), 60.41 (*C*H), 127.73 (*C*H), 130.45 (*C*H), 135.52 (*C*), 157.20 (*C*), 157.59 (*C*), 170.97 (*C*),

172.63 (*C*), 172.75 (*C*), 174.05 (*C*), 174.27 (*C*), 174.98 (*C*), 175.20 (*C*), 176.62 (*C*) (Supplementary Figs. 7–14). $[\alpha]^{22.9}_{D}$ -11.3 (c 0.40, H_2O). HRMS-ESI (m/z): $[M+H]^+$ calculated for $C_{34}H_{56}N_{13}O_{10}$, 806.4268; observed, 806.4281.

Preparation of non- and SeMet-labelled purified PGM1s for crystallization. The non-labelled PGM1 was expressed by the same way as described above. *E. coli* BL21 (DE3) harboring the PGM1 expression plasmid was grown at 37°C in M9 medium containing 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1× BME vitamins solution (Sigma-Aldrich), 0.4% (w/v) glucose, and 100 μg/L ampicillin until the optical density reached 0.6. 25 mg/L selenomethionine, 100 mg/L each of L-Lys, L-Thr, L-Ile, L-Leu, L-Val, and L-Phe, and 0.5 mM isopropyl β-D-thiogalactopyranoside was then added to express SeMet-labeled and *C*-terminal hexahistidin-fused PGM1 and growth was continued for 12 h at 25˚C.

The each cells were independently harvested and resuspended in lysis buffer [50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol, 0.5 mg/mL lysozyme and 0.1 U/ml Benzonase Nuclease (Merck Japan, Tokyo)]. The cell lysate was sonicated and then centrifuged at 10,000 × *g* for 20 min. The supernatant was loaded onto a Ni Sepharose 6 Fast Flow column and PGM1 was eluted from the column with elution buffer (50 mM HEPES-NaOH, pH 7.0, 200 mM NaCl, 5% glycerol and 600 mM imidazole). The elution was diluted fivefold with buffer A (50 mM HEPES-NaOH, pH 7.0, and 2 mM DTT) and purified by the Resource Q column (GE Healthcare) with a linear gradient of buffer A to buffer B (50 mM HEPES-NaOH, pH 7.0, 1 M NaCl and 2 mM DTT). Finally, non- and SeMet-labelled PGM1s were concentrated to 20 mg/mL by Macrocep 10 k centrifuge device (PALL Life sciences, MI, USA) and used for crystallization, respectively.

Crystallization, data collection and structure refinement. Well-diffracted PGM apo crystals were obtained using 100 mM Tris-HCl, pH 8.3, containing 1.2 M lithium sulfate and 10 mM L-Met at 20°C with the sitting-drop vapor-diffusion method by mixing 2 μL of the SeMet-labelled, purified PGM1 with an equal volume of reservoir solution and equilibrating the mixture against 500 μL reservoir solution. Diffraction quality crystals of PGM1 complexed with AMP were obtained from the SeMet-labelled, purified PGM1 by the same method with the PGM1 apo crystal at 20°C in 100 mM Tris-HCl, pH 8.1, containing 1.2 M lithium sulfate and 10 mM AMPPNP. Both crystals

appeared within a few days. The PGM1 apo and PGM1-AMP complex crystals were transferred into a cryoprotectant solution, consisting of each crystallization solution with 10% glycerol. After a few seconds, the crystals were picked up in a nylon loop and then flash-cooled at −173°C by a nitrogen-gas stream.

Single-wavelength anomalous diffraction (SAD) data of PGM1 apo and PGM1-AMP complex crystals were collected on beamline BL17A and BL5A at Photon Factory, respectively, under cryogenic conditions at −173°C. Wavelengths of 0.97939 Å at BL17A, and 0.97901 Å at BL5A were used for data collection based on the fluorescence spectrum of the Se *K* absorption edge [39]. The diffraction data were processed and scaled using *XDS* [40]. Se sites were determined and refined, and the initial phase of PGM1 apo was calculated with *AutoSol* in *PHENIX* [41, 42]. The initial phase of PGM1-AMP complex was then calculated by the molecular replacement (MR)-SAD method with *AutoSol* in *PHENIX* [41] using PGM1 apo as the search model. The structures were rebuilt using *AutoBuild* in *PHENIX* [41], modified manually with *Coot* [43] and refined with *PHENIX* [41]. The detailed data collection and refinement statistics are summarized in Supplementary Table 6. The quality of the final models was assessed with MolProbity [44]. A total of 98.2% of the residues in the PGM1 apo structure were in the most favored regions of the Ramachandran plot, and 1.7% were in the allowed regions, while a total of 97.2% of the residues in the PGM1-AMP complex structure were within the most favored regions of the Ramachandran plot, and 2.5% were in the allowed regions.

A structural similarity search was performed using the program [20]. All crystallographic figures were prepared with PyMOL (DeLano Scientific, http://www.pymol.org).

Modeling and docking studies. Three-dimension models of the *N*-terminus substrate and phosphorylated **2** were generated by the Chem3D Ultra 10 program (CambridgeSoft, MA, USA), and their geometries were optimized using *elbow* in *PHENIX* [43]. The peptide models were generated and optimized using *CNS 1.2* [45].

The ADP- and ATP-bound PGM1 models were built by swapping the ADP and ATP molecules with AMP, followed by placing the β-phosphate of the ADP and ATP models on the weak electron density neighboring α-phosphate of the AMP molecule in the PGM1-AMP crystal structure with

Coot. The *N*-terminus substrate and phosphorylated **2** were then docked into the ATP- and ADP-bound PGM1 model structures, respectively, by *Autodock vina 1.0.2* [46]. The PGM1 quaternary complex structures with the respective NVKDR, or NVKDGPT were built from the tertiary complex of PGM1-ADP-phospholyrated substrate **2**, by docking the peptide into the model structure. The docking parameters were kept to the default values. The size of the docking grid of each simulation is summarized in Supplementary Table 7.

Mutagenesis analyses. All the mutant enzymes were constructed with the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA), according to the manufacturer's protocol with the primers listed in Supplementary Table 8. The mutant enzymes were expressed and purified by the same procedure used for the wild-type PGM1 (Supplementary Fig. 39).

Accession numbers. The nucleotide sequences reported in this paper have been submitted to the DDBJ/GenBankTM/EBI Data Bank with the following accession number: AB896796. The coordinates and structure factors reported in this paper have been deposited in the Protein Data Bank, www.pdb.org with the following accession numbers: 3WVQ for the PGM1 apo structure and 3WVR for the PGM1-AMP complex structure.

References

1. Humphries, R.M., Pollett, S. & Sakoulas G. A current perspective on daptomycin for the clinical microbiologist. *Clin. Microbiol. Rev*. **26**, 759-780 (2013).
2. Arnison, P.G. *et al*. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep*. **30**, 108-160 (2013).
3. Koglin, A. & Walsh, C.T. Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep*. **26**, 987-1000 (2009).
4. Giessen, T.W. & Marahiel, M.A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett*. **586**, 2065-2075 (2012).
5. Maruyama, C. et al. A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat. Chem. Biol*. **8**, 791-797 (2012).
6. Kadi, N., Oves-Costales, D., Barona-Gomez, F. & Challis, G.L. A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. *Nat. Chem. Biol*. **3**, 652-656 (2007).
7. Suzukake-Tsuchiya, K., Hori, M., Shimada, N. & Hamada, M. Mode of action of deoxypheganomycin D on *Mycobacterium smegmatis* ATCC 607. *J. Antibiot. (Tokyo)* **41**, 675-683 (1988).
8. Crone, W.J.K., Leeper, F.J. & Truman, A.W. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science* **3**, 3516-3521 (2012).

9. Huo, L., Rachid, S., Stadler, M., Wenzel, S.C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **19**, 1278-1287 (2012).

10. Hubbard, B.K. & Walsh, C.T. Vancomycin assembly: nature's way. *Angew. Chem. Int. Ed Engl.* **42**, 730-765 (2003).

11. Donia, M.S., Ravel, J. & Schmidt, E.W. A global assembly line for cyanobactins. *Nat. Chem. Biol.* **4**, 341-343 (2008).

12. Chen, H., Tseng, C.C., Hubbard, B.K. & Walsh, C.T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc. Natl. Acad. Sci. USA* **98**, 14901-14906 (2001).

13. Anzai, H. *et al*. Replacement of *Streptomyces hygroscopicus* genomic segments with in vitro altered DNA sequences. *J. Antibiot. (Tokyo)* **41**, 226-233 (1988).

14. Fawaz, M.V., Topper, M.E. & Firestine, S.M. The ATP-grasp enzymes. *Bioorg. Chem.* **39**, 185-191 (2011).

15. Yamaguchi, H. *et al*. Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 A resolution. *J. Mol. Biol.* **229**, 1083-1100 (1993).

16. Fan, C., Moews, P.C., Shi, Y., Walsh, C.T. & Knox, J.R. A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine:D-alanine ligase of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **92**, 1172-1176 (1995).

17. Tabata, K., Ikeda, H. & Hashimoto, S. *ywfE* in *Bacillus subtilis* codes for a novel enzyme, L-amino acid ligase. *J. Bacteriol*. **187**, 5195-5202 (2005).

18. Kino, K., Arai, T. & Tateiwa, D. A novel L-amino acid ligase from *Bacillus subtilis* NBRC3134 catalyzed oligopeptide synthesis. *Biosci. Biotechnol. Biochem.* **74**, 129-134 (2010).

19. Taguchi, S., Mita, K., Ichinohe, K. & Hashimoto, S. Targeted engineering of the antibacterial peptide apidaecin, based on an in vivo monitoring assay system. *Appl. Environ. Microbiol.* **75**, 1460-1464 (2009).

20. Holm, L. & Sander, C. Dali: A network tool for protein structure comparison. *Trends*

Biochem. Sci. **20**, 478-480 (1995).

21. Sampei, G. *et al*. Crystal structure of glycinamide ribonucleotide synthetase, PurD, from thermophilic eubacteria. *J. Biochem.* **148**, 429-438 (2010).

22. Thoden, J.B., Holden, H.M., Paritala, H. & Firestine, S.M. Structural and functional studies of *Aspergillus clavatus* N^5-carboxyaminoimidazole ribonucleotide synthetase. *Biochemistry* **49**, 752-760 (2010).

23. Thoden, J.B., Firestine, S.M., Benkovic, S.J. & Holden, H.M. PurT-encoded glycinamide ribonucleotide transformylase. *J. Biol. Chem.* **277**, 23898-23908 (2002).

24. Suzuki, M. *et al.* The structure of L-amino-acid ligase from *Bacillus licheniformis*. *Acta Crystallogr. D* **68**, 1535-1540 (2012).

25. Wang, W., Kappock, T.J., Stubbe, J. & Ealick, S.E. X-ray crystal structure of glycinamide ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* **37**, 15647-15662 (1998).

26. Bateman, A. *et al*. The Pfam protein families database. *Nucleic Acids Res.* **28**, 263-266 (2000).

27. Teixeira, L.K. & Reed, S.I. Ubiquitin ligases and cell cycle control. *Annu. Rev. Biochem.* **82**, 387-414 (2013).

28. Mazmanian, S.K., Liu, G., Ton-That, H. & Schneewind, O. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760-763 (1999).

29. Kane, P.M. *et al*. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**, 651-657 (1990).

30. Chan, A.O. *et al*. Modification of *N*-terminal alpha-amino groups of peptides and proteins using ketenes. *J. Am. Chem. Soc.* **134**, 2589-2598 (2012).

31. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252-262 (2005).

32. Wagner, A.M. *et al*. *N*-Terminal Protein Modification Using Simple Aminoacyl Transferase Substrates. *J. Am. Chem. Soc.* **133**, 15139-15147 (2011).

33. Pan, Y. *et al*. *N*-terminal labeling of peptides by trypsin-catalyzed ligation for quantitative

proteomics. *Angew. Chem. Int. Ed Engl.* **52**, 9205-9209 (2013).

34. Jursic, B.S., Neumann, D. & McPherson, A. Preparation of *N*-formamidinylamino acids from amino and formamidinesulfinic acids. *Synthesis.* **32**, 1656-1658 (2000).

35. Zerbino, D.R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829 (2000).

36. Noguchi, H., Taniguchi, T., & Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res*. **15**, 387-396 (2008).

37. Vara, J., Lewandowska-Skarbek, M., Wang, Y.G., Donadio, S. & Hutchinson, C.R. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J. Bacteriol.* **171**, 5872-5881 (1989).

38. Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. & and Schremp, H. Gene manipulation of *Streptomyces*, a laboratory manual. The John Innes Foundation, Norwich, UK (1985).

39. Rice, L.M., Earnest, T.N. & Brunger, A.T. Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr D* **56**, 1413-1420 (2000).

40. Kabsch, W. XDS. *Acta Crystallogr D* **66**, 125-132 (2010).

41. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221 (2010).

42. Afonine, P.V. *et al.* Towars automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr. D* **68**, 352-367 (2012).

43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallogr. D* **60**, 2126-2132 (2004).

44. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12-21 (2010).

45. Brunger A. T. Version 1.2 of the crystallography and NMR system, *Nature Protocols* **2**,

2728-2733 (2007).

46. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455-461 (2010).

Figure Legends

Figure 1. Chemical structures of pheganomycins and its biosynthetic gene cluster identified in *S. cirratus*. (**a**) Chemical structures of pheganomycins isolated as natural products are indicated. **(b)** A schematic illustration of the pheganomycin biosynthetic gene cluster is shown together with the genes located at boundary regions. The genes mentioned in the text were highlighted by solid arrows.

Figure 2. Compounds used for *N*-terminus substrate and analysis of the reaction product formed from 4 and the peptide. (**a**) Chemical structures used for *N*-terminus substrate and the natural substrate (**2**) are shown. (**b**) HPLC traces of the reaction products formed from **4** and peptide NVKDR with (upper) and without (lower) PGM1 are indicated. (**c**) Structure of the reaction product (**5**) determined by NMR analyses is shown

Figure 3. Peptides used for nucleophile and analysis of the reaction product formed from 4 and apidaecin. (**a**) Amino acid sequences of peptides used for *C*-terminus substrate are shown. The substituted amino acids in the peptides are highlighted in bold. (**b**) MALDI-TOF-MS analyses of the reaction products formed from **4** and apidaecin with (upper) and without (bottom) PGM1 are indicated.

Figure 4. Overall structure and docking models of PGM1. (**a**) Crystal structure of PGM1 AMP complex is shown in cartoon model. The N, A, B and C domains are colored in blue, yellow, red and cyan, respectively. The AMP molecule and Arg335 are represented with green and light brown stick

models, respectively. (**b**) Close-up view of the active site of the PGM1-AMP complex structure is shown in a surface model. Each domain and AMP molecule are indicated as in (**a**). The bottom of the large cavity and the cleft are highlighted with blue and orange, respectively. The position of Arg335 is indicated with light brown. (**c**) Model structure of PGM1-ATP-**2** is shown. The ATP and **2** are shown in green and cyan stick models, respectively. (**d** and **e**) Model structures of PGM1-ADP-phosphorylated **2**-NVKDR peptide (**d**) and -NVKDGPT peptide (**e**) are shown. The ADP, phosphorylated **2**, and the two peptide substrates are shown in green, cyan and light blue stick models, respectively.

Figure 5. Hypothetical biosynthetic machinery of PGM. A hypothetical biosynthetic pathway of PGM is shown. Compound **2** plausibly biosynthesized by Type III polyketide synthase (PGM11) and modified by tailoring enzymes (PGM3, 4, 7–10 and 12) is phosphorylated with ATP by PGM1, and then the peptide (NVKDR or NVKDPGT) produced by ribosomes and probably processed by a peptidase (PGM14) attacks as the nucleophile.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research (23108101 to T. Dairi and 2510870 to H. Morita) from the Japan Society for the Promotion of Science. We thank Dr. Masayuki Igarashi from the Institute of Microbial Chemistry and Prof. Seiichi Taguchi from Hokkaido University for providing us with *S. cirratus* and apidaecin, respectively

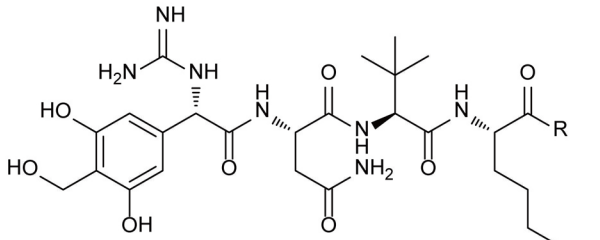
Author contributions

M.N., H.M. and T.D. designed the research. M.N., T.M., and K.O. performed in vitro experiments. T.M, S.O. and H.M. carried out crystallography. I.S and H.I synthesized substrates and performed NMR studies. M.N., Y.H., C.M. and Y.S analyzed products by LC/MS. J.I. and T.D. identified the gene cluster. M.N., H.M. and T.D. analyzed data and wrote the manuscript.

Competing financial interests

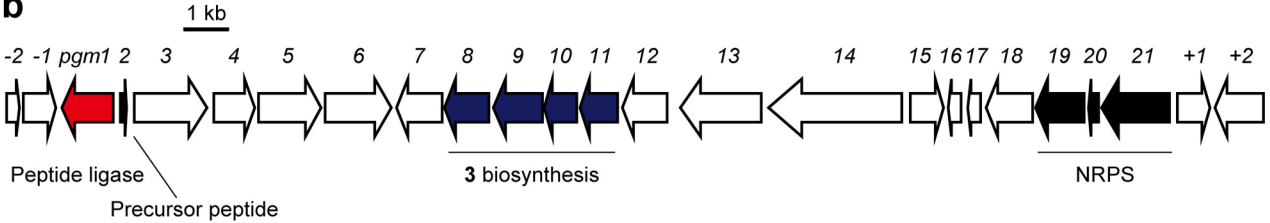
The authors declare no competing financial interests.

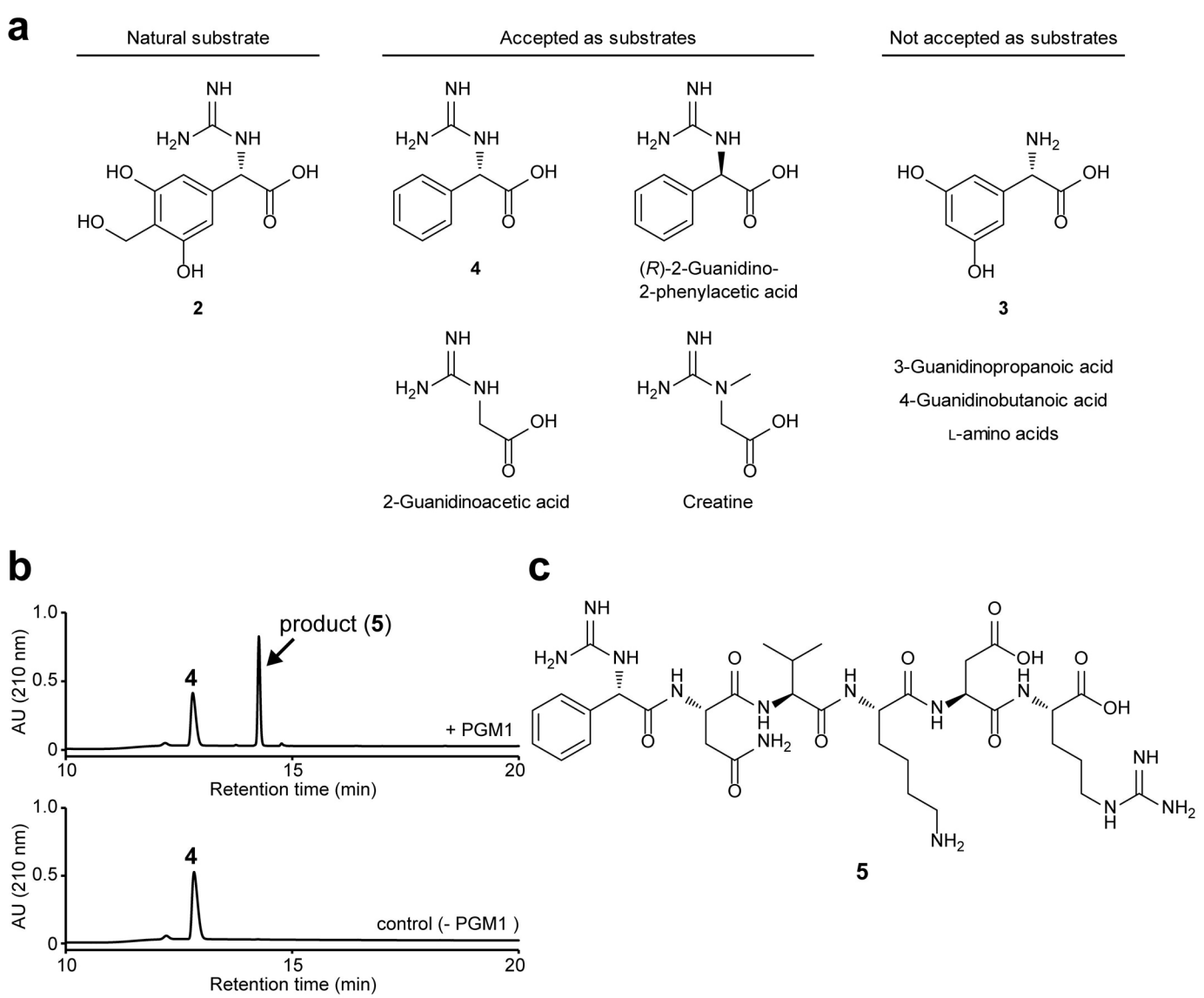
a



PGM (**1**) : R = H
PGM-D : R = Asp
PGM-DR : R = Asp-Arg
PGM-DGPT: R = Asp-Gly-Pro-Thr

b

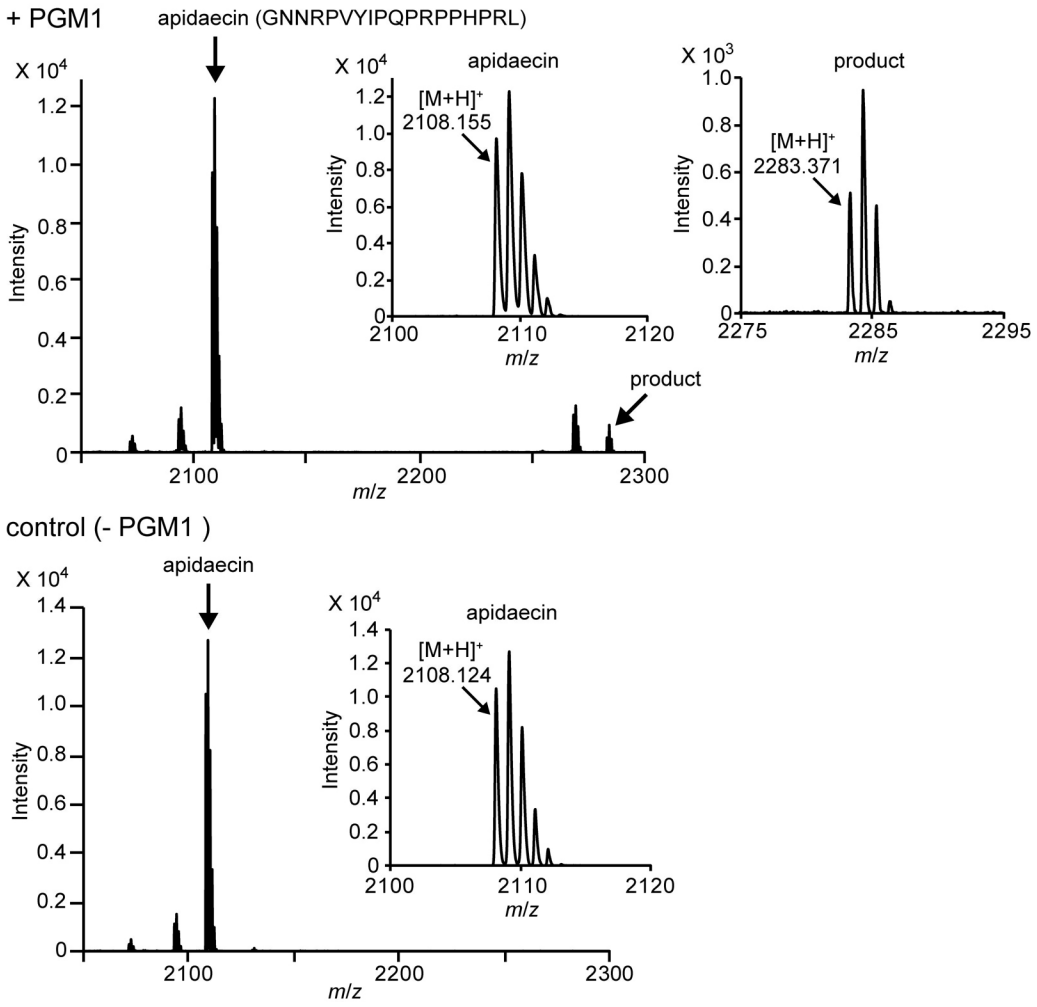


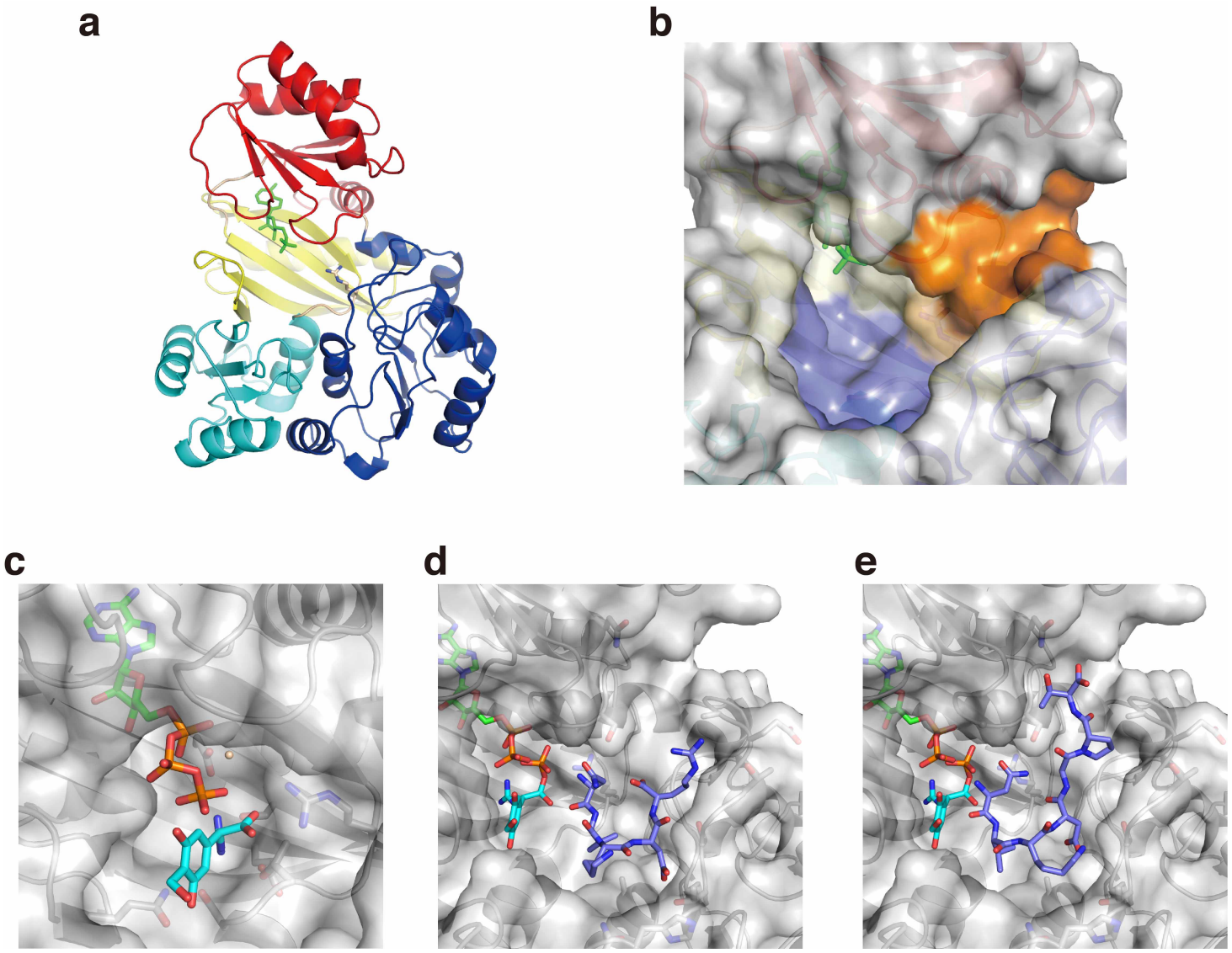

a
Natural substrate
Accepted as substrates
Not accepted as substrates
2
4
(R)-2-Guanidino-
2-phenylacetic acid
3
3-Guanidinopropanoic acid
4-Guanidinobutanoic acid
L-amino acids
2-Guanidinoacetic acid
Creatine
b
product (5)
4
+ PGM1
AU (210 nm)
Retention time (min)
control (- PGM1)
c
5

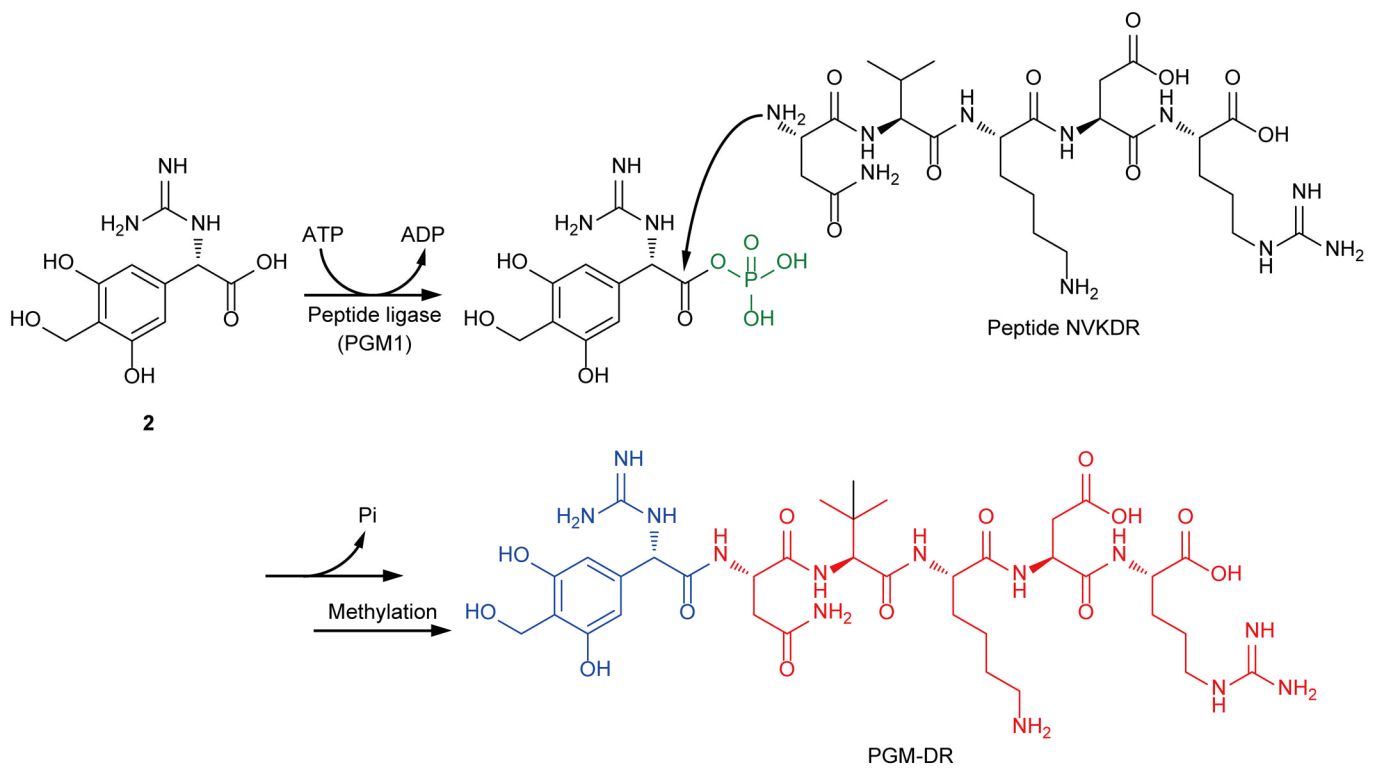
a

Natural substrates	Accepted as substrates			Not accepted as substrates
NVKDR	NVKD**A**GP	**A**VKDR	Aspartame (Asp-Phe methylester)	L-amino acids
NVKDGPT	**Q**VKDR	N**A**KDR	MRFA (tetra peptide)	
NV	N**L**KDR	NV**A**DR	Apidaecin	
NVKDG	NV**R**DR	**AA**KDR	(GNNRPVYIPQPRPPHPRL)	
NVKDGP	**D**VKDR	**AAA**DR	Apidaecin derivative	
	FVKDR		(RVRRPVYIPQPRPPHPRL)	

b




a
b
c
d
e


ATP
ADP
Peptide ligase
(PGM1)
Peptide NVKDR
2
Pi
Methylation
PGM-DR