J. Microbiol. Biotechnol. (2005), **15**(3), 656–659

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Structure Prediction of the Peptide Synthesized with the Nonribosomal Peptide Synthetase Gene from *Bradyrhizobium japonicum*

JUNG, BO-RA, YUKYUNG LEE, YOONGHO LIM, AND JOONG-HOON AHN*

Bio/Molecular Informatics Center, Department of Molecular Biotechnology, Konkuk University, Seoul 143-701, Korea

Received: September 21, 2004
Accepted: December 13, 2004

Abstract Small peptides synthesized by nonribosomal peptide synthetases (NRPSs) genes are found in bacteria and fungi. While some microbial taxa have few, others make a large number and variety. However, biochemical characterization of the products synthesized by NPRS demands a great deal of efforts. Since the completion of genome projects of numerous microorganisms, the numbers of available NRPSs genes are being expanded. Prediction of the peptides encoded by NRPS could save time and efforts. We chose the NRPS gene from *Bradyrhizobium japonicum* as a model to predict the peptide structure encoded by NRPS genes. Using computational analyses, the domain structure of this gene was defined, and the structure of a peptide synthesized by this NRPS was deduced. It was found that it encoded a tripeptide consisting of proline-serine-phenylalanine. This method would be helpful to predict the structure of small peptides with various NPRS genes from the genome sequence.

Key words: Nonribosomal peptide synthetase, domain analysis, *Bradyrhizobium japonicum*

Proteins are generally synthesized in a ribosome. However, in some cases, protein synthesis occurs without ribosome, which is known as nonribosomal peptide synthesis. In bacteria and fungi, and even in plant, biologically active peptides are synthesized by nonribosomal peptide synthetase (NRPS), and they range from antibiotics (e.g., penicillin and gramicidin) to toxins (HC-toxin and syringomycin) that comprise one of the largest and most important groups of microbial secondary metabolites [17]. The basic units of NRPS are an adenylation (A) domain, a thiolation (T) domain, and a condensation (C) domain [10]. The A domain recognizes the amino acid substrate and adenylates it at the carboxylic acid group. The T domain contains 4'-phosphopanthotheine (4'-PP), whose thiol group takes over the amino acid through the formation of a thioester bond after aminoacyl-adenylate formation by the A domain. The C domain is located between A/T modules and catalyzes peptide bond formation. Since the elucidation of the crystalline structure of the Phe-specific A domain of gramicidin synthetase (*GraS*) from *Bacillus brevis* [4], it has been well established that the amino acids lining the substrate-binding pocket mediate amino acid specificity. The eight or nine amino acids occupying pockets in the substrate-binding site was proposed as nonribosomal codes to predict amino acid substrate specificity of A domains. Nonribosomal codes sequence can be predicted by comparing 110 amino acid sequence beginning just after conserved motif A3 and ending just after conserved motif A5 [2, 10, 14].

As a result of genome projects on many microorganisms, a great deal of information has been accumulated about NRPS, and functional characterization of individual genes from various organisms would be a task to be undertaken. On the other hand, biochemical characterization of the peptide synthesized by NRPS has a limitation, because microorganisms sometimes produce those secondary metabolites under a particular condition and of a minute amount [17]. Before the genome sequencing project was launched, gene cloning was a time-consuming task, especially in the nonribosomal peptide synthetase genes which are relatively large. Individual amino acid activation domains were subcloned and expressed to know the amino acids [7]. Now, however, anyone can find the nonribosmal pepetide synthetase gene from the public database. With the help and efforts of a few groups, it is now possible to define the domains of NRPS, identify the residues that interact with the specific amino acids to produce peptide, and predict the possible structure before biochemical characterization is undertaken. In this study, we chose *B. japonicum* USDA 110, whose genome sequences were completely elucidated, and NRPSs were selected and dissected to define their possible peptide structure.

*Corresponding author
Phone: 82-2-450-3764; Fax: 82-2-456-7183;
E-mail: jhahn@konkuk.ac.kr

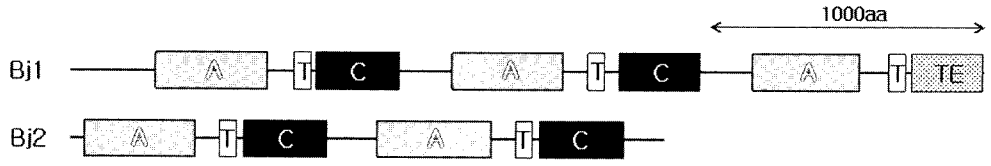

Fig. 1. Domain organization of nonribosomal peptide synthetases from *Bradyrhizobium japonicum.*
A, Adenylation domain; T, Thiolation domain; C, Condensation domain; TE, Thioeasterase domain.

A search of *B. japonicum* USDA110 genome revealed two NRPS genes that were named Bj1 and Bj2. Bj1 consisted of 3,310 amino acids, and Bj2 had 2,154 amino acids. BLAST result showed that Bj1 had a similarity with *Ncp* A from *Nostoc* sp. ATCC 53789 [1]. *NcpA* and *B* encode peptide synthetases that eventually mediate the synthesis of the heptapeptide consisting of Tyr-Gly-Gln-Ile-Ser-Pro-Leu/Phe. Bj2 had a homology with syringomycin synthetase from *Pseudomonas syringae* pv. *syringae* [5]. The peptide module of syringomycin comprised Ser-D-Ser-D-Dab-Dab-Arg-Phe-Dhb-(3-OH)Asp-(4-Cl)Thr (Dab, 2,4-diaminobutyric acid; D-Dab, D-isomer of Dab; Dhb, 2,3-dehydroaminobutyric acid). Domain analysis showed that Bj1 contains three A domains, three T domains, two C domains, and one TE domains, and that Bj2 contains two As, three Ts, and two Cs (Fig. 1). The D-form of amino acids and other unusual forms of amino acids are usually found in peptides synthesized by NRPS [17]. Bj1 and Bj2 did not contain an epimerase domain that converts the L-form of amino acid to D-form, indicating that peptides from these NRPSs do not contain the D-form of amino acid. However, we did not exclude the possibility that epimerase might be located in an other part of the genome, as found in fungi. The T domain has been well-defined in bacterial NRPSs and contains the 4'-phosphopanthotheine (4'-PP) group binding motif, DXFFXXLGG(H/D)S(L/I) [12]. Among amino acids comprising this motif, the serine residue is conserved, since the 4'-PP is directly bound to the serine residue. Alignments of all five T domains from Bj1 and Bj2 showed that the 4'-PP binding motif was well-conserved in addition to the conserved serine residue (Fig. 2). Also, the core condensation motif (HHXXXDG) [8] was found in all the C domains.

To determine the amino acid that is recognized by each A domain, each of the A domain sequences from Bj1 and Bj2 was aligned with the phenylalanine activation domain of *GraS*. Since the domain was divided into ten conserved core motifs (from A1 to A10; [10]), and it is known that the amino acid substrate-binding pocket is located with an about 110 amino acid sequence just after conserved motif A3 [AY(V/I)(L/I)FTSGSTGXPKG] and ending just after conserved motif A5[N(G/A)(Y/W)GP(T/A)E] [14, 16], this segment was extracted from the whole A domain (Fig. 3). Then, the eight amino acids that are known to form the substrate pocket in *GrsA* from *B. brevis* [3, 14] were located. Since the peptide has an N-terminal amino group that is positively charged, it is generally known in bacterial NRPS that N-terminal amino acid interacts with aspartic acid to neutralize the positive charge. Thus, the first position among eight amino acids is usually occupied by aspartic acid. Analysis of five A domains showed that three of them had an aspartic acid in the first position and two of them had a different amino acid (Bj2-1 had valine and Bj2-2 ala.). Thus, further analysis of Bj2 was ruled

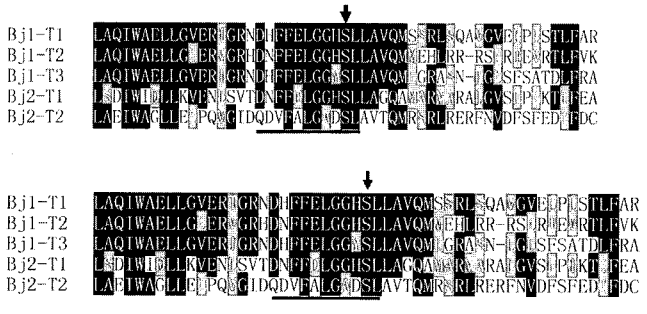

Fig. 2. Alignments of thiolation domain from Bj1 and Bj2.
Underlined is the conserved 4'-phosphopanthotheine binding motif. Arrows indicate the conserved serine residue.

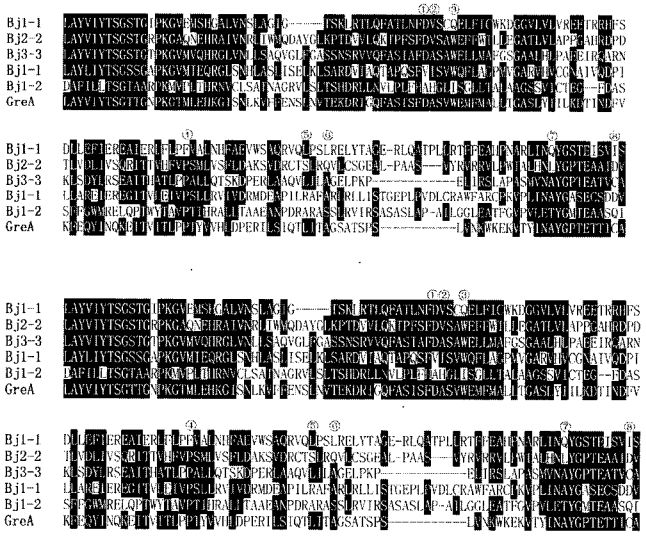

Fig. 3. Alignments of core motifs from A4 to A5.
Numbers indicate the amino acids positioning in the pocket. *GrsA* is a phenylalanine activation domain of gramicidin S synthetase.

Table 1. Residues in the substrate-binding pocket of NRPS from *B. japonicum* and the predicted amino acid substrates.

	Residue								
A-domain	① 235	② 236	③ 239	④ 278	⑤ 299	⑥ 301	⑦ 322	⑧ 330	Substrate
Bj2-1	D	V	Q	F	P	L	Q	V	Pro
Bj2-2	D	V	W	H	T	L	L	I	Ser
Bj2-3	D	A	W	T	I	A	A	V	Phe

out. Comparison of eight amino acid nonribosomal codes from three A domains from Bj1 showed that Bj1-1 has homology with the first A domain of peptide synthetase ORF4 from *B. subtilis* [15], recognizing a proline; Bj1-2 with pyoverdin synthetase from *Pesudomonas aeruginosa* [11], coding a serine; and Bj1-3 with gramicidin synthetase [6, 9] for phenylalanine. This result indicates that Bj1 predicts to produce a tripeptide of which the sequence is proline-serine-phenylalanine. With this primary peptide sequence, the structure of peptide was predicted. There is a TE domain at the end of Bj1, which mediates the release of a peptide from the enzyme. Therefore, the first possible structure would be a linear proline-serine-phenylalanine that is stabilized by the interresidue hydrogen bond (2.4 Å) between the carbonyl group of proline and the amino group of phenylalanine and the intraresidue hydrogen bond (1.9 Å) between the carbonyl group of serine and the hydroxyl group of serine side chain. However, the TE domain is present in systems producing not only linear peptide such as ACV but also branched via ester bond (such as surfactin) or branched through amide bond as found in bacitracin and cyclic peptides (such as gramicidin S and tyrocidine) [10]. All the nonlinear forms seem to be synthesized after the release of peptide. Thus, two cyclic forms could be imagined: The first one could be through the formation of imine bond between the nitrogen group of the proline and carboxy group of the serine. However, the formation of imine bond is not feasible, since the nitrogen in the proline is secondary. The other would be a dimer that contains two units of Pro-Ser-Phe. This form is unlikely since it has 75.6 kcal/mol higher energy due to the steric hindrance of the side chain. Based on these considerations, the predicted structure from this sequence is as shown in Fig. 4, and its structure is stabilized by two hydrogen bonds.

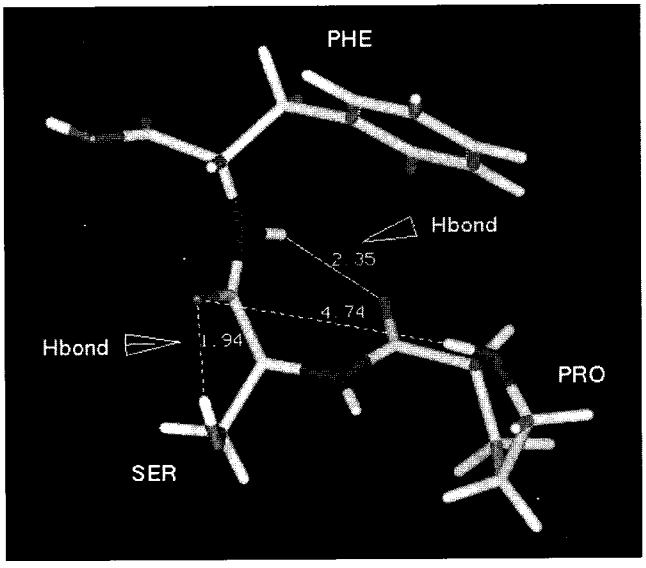

Fig. 4. Predicted peptide structure synthesized by Bj1.

Elucidation of the peptide structure from NPRS genes with computational analyses has been carried out on coelichin from *Streptomyces coelicolor* [3], bleomycin from *S. verticillus* ATCC 15003 [13], and nostocyclopeptide from *Nostoc* sp. [1]. Coelichin was predicted to have an internal cyclization to release the final peptide due to the lack of TE domain. In the cases of bleomycin and nostocyclopeptide, the predicted amino acid recognition domains were experimentally confirmed. As shown in the studies of bleomycin and nostocyclopeptide, the prediction method could provide at least approximate information about the backbone of the peptide synthesized by NRPSs.

By the time this manuscript was prepared, 158 bacterial genome projects had been accomplished. Search of NRPS genes with A domain showed the presence of at least 100 NPRS genes, and only a few of them have been characterized. Prediction of substrates of each of the amino acid activation domain would facilitate the biochemical characterization of peptides encoding NPRS genes and would be useful to engineer a natural peptide for the development of new compounds.

Acknowledgments

This work was supported by the R&D Program for Fusion Strategy of Advanced Technologies (Korea Ministry of Science and Technology) and KRF2004-F00019 (KRF).

References

1. Becker, J. E., R. E. Moore, and B. S. Moore. 2004. Cloning, sequencing, and biochemical characterization of

the nostocyclopeptide biosynthetic gene cluster: Molecular basis for imine macrocyclization. *Gene* **325:** 35–42.
2. Challis, G. L., J. Ravel, and C. A. Townsend. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7:** 211–224.
3. Challis, G. L. and J. Ravel. 2000. Coelichin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: Structure prediction from the sequence of its non-risobosomal peptide synthetase. *FEMS Microb. Lett.* **187:** 111–114.
4. Conti, E., T. Stachelhaus, M. A. Marahiel, and P. Brick. 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16:** 4174–4183.
5. Guenzi, E., G. Galli, I. Grgurina, D. C. Gross, and G. Grandi. 1998. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273:** 32857–32863.
6. Hori, K., Y. Yamamoto, T. Minetoki, T. Kurotsu, M. Kanda, S. Miura, K. Okamura, J. Furuyama, and Y. Saito. 1989. Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. *J. Biochem. Tokyo* **106:** 639–645
7. Kim, K.-R., I. H. Lee, and J.-W. Suh. 2001. A putative peptide synthetase from *Bacillus subtilis* 713 recognizing L-lysine, L-tryptophan, and L-glutaminc acid. *J. Microbiol. Biotechnol.* **11:** 798–803
8. Konz, D. and M. A. Marahiel. 1999. How do peptide synthetases generate structural diversity? *Chem. Biol.* **6:** R39–R48.
9. Krätzschmar, J., M. Krause, and M. A. Marahiel. 1989. Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J. Bacteriol.* **171:** 5422–5429
10. Marahiel, M. A., T. Stachelhaus, and H. D. Mootz. 1997. Modular peptide synthetase involved in nonribosomal peptide synthesis. *Chem. Rev.* **97:** 2651–2673.
11. McMorran, B. J., M. E. Merriman, I. T. Rombel, and I. L. Lamont. 1996. Characterisation of the *pvdE* gene which is required for pyoverdine synthesis in *Pseudomonas aeruginosa*. *Gene* **176:** 55–59.
12. Schlumbohm, W., T. Stein, C. Ullrich, J. Vater, M. Krause, M. A. Marahiel, V. Kruft, and B. Wittmann-Liebold. 1991. An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem.* **266:** 23135–23141.
13. Shen, B., D. L. Sancherz, C. Edwards, D. J. Chen, and M. J. M. Murrell. 2002. Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC 15003. *J. Nat. Prod.* **65:** 422–431.
14. Stachelhaus, T., H. D. Mootz, and M. A. Marahiel. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetase. *Chem. Biol.* **6:** 493–505.
15. Tognoni, A., E. Franchi, C. Magistrelli, E. Colombo, P. Cosmina, and G. Grandi. 1995. A putative new peptide synthase operon in *Bacillus subtilis*: Partial characterization. *Microbiology* **141:** 645–648.
16. Turgay, K., M. Krause, and M. A. Marahiel. 1992. Four homologous domains in the primary structure of *GrsB* are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6:** 529–546.
17. Walton, J. D., D. G. Panaccione, and H. E. Hallen. 2004. Peptide synthesis without ribosomes, pp. 127–162. *In* J. Tkacz and L. Lange (eds.), *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine.* Kluwer Academic Publishers, New York, U.S.A.