

NOTE

Isolation and Characterization of Two Amino Acid-activating Domains of Peptide Synthetase Gene from *Bacillus subtilis* 713

LEE, SOON-YOUL, SANG-BAE YOU, JI-WAN LEE, TAE-YOUNG KIM, SUNG-UK KIM¹, SONG-HAE BOK¹, AND JOO-WON SUH*

Department of Biological Science, Myong Ji University, Kyonggi-Do 449-728, Korea

¹Biomolecule Research Division, Korea Research Institute of Bioscience and Biotechnology, Taejon 305-333, Korea

Received: February 5, 1998

Abstract From the sequence alignment of various non-ribosomal peptide synthetases, several motifs of highly conserved sequences have been identified within each domain of peptide synthetases. We designed PCR primers based on the highly conserved nucleotide sequences to amplify and isolate a ~7.2-kb DNA fragment of the *Bacillus subtilis* 713 which was isolated and reported to produce an antifungal peptide compound. Nucleotide sequence analysis of 4.8 kb of the predicted amino acids revealed significant homology to various peptide synthetases over the whole sequence and also revealed two amino acid-activating domains with highly conserved Core 1 to Core 6 and spacer motif. This suggests that the isolated DNA fragment is part of a peptide synthetase gene for antifungal peptide.

Key words: Peptide antibiotics, PCR, *Bacillus subtilis* 713, nucleotide sequence

In order to decrease environmental pollution and solve the problem of widespread appearance of drug resistant pathogens, the development of new antibiotics is urgently needed. In this sense, peptide antibiotics have been a source of attention since it is possible to program a noble antibiotics by gene manipulation of modular domains of peptide synthetase genes as described by Stachelhaus *et al.* [30]. The stability of peptide antibiotics to most peptidases and proteases has made them useful in the pharmaceutical and agricultural industries [19, 20, 29]. Bacilli are the extensively studied bacteria known to produce peptide antibiotics and to be regarded as relatively safe to humans [10, 11, 14, 29]. Peptide antibiotics are also produced by many microorganisms in addition to Bacilli [20], and exhibit diversity with respect to

chemical structure and biological activity. Although structurally diverse, peptide antibiotics are largely divided into two subgroups based on biosynthetic mechanisms, namely, ribosomal biosynthesis [9, 28] and nonribosomal biosynthesis.

Many of peptides are not gene-encoded, but are synthesized nonribosomally through a series of reactions catalyzed by large, often multisubunit, enzyme complexes called peptide synthetases by a thio-template mechanism [17, 29, 31, 32]. Each subunit of peptide synthetases has independent enzymatic activity. Isolation, sequencing, and characterization of several genes encoding multifunctional peptide synthetases of bacterial and fungal origin required for nonribosomal synthesis of peptide antibiotics have confirmed that the gene is composed of several domains for the corresponding subunits. The occurrence and specific order of domains of synthetase genes dictate the number and sequence of the amino acids incorporated into the peptide product [31]. These findings have made it possible to develop noble antibiotics. For example, Stachelhaus *et al.* [30] developed noble antibiotics by genetic manipulation of domains of the peptide synthetase genes.

Sequence alignment has also revealed six highly conserved motifs (Core 1 to Core 6) and a spacer motif in peptide synthetase genes [15, 16, 17, 31]. Cores 2 to 5 are believed to be involved in ATP-binding and hydrolysis. The Core 6 motif seems to be involved in the catalysis of the formation of thioester bonds. However, the function of the Core 1 motif is not yet known. Among the conserved regions, Core 1 and Core 2 are highly conserved even in nucleotide sequences (Fig. 1a) [3].

In this study, we designed PCR primers based on the highly conserved nucleotide sequences of Core 1 and Core 2 to isolate and characterize the peptide antibiotics synthetase genes from *Bacillus subtilis* 713 strain [12, 13]. *Bacillus subtilis* 713 strain was isolated from Korean soil and known to produce antifungal peptide

*Corresponding author

Phone: 82-335-30-6190; Fax: 82-335-35-8249;
E-mail: jwsuh@wh.myongji.ac.kr

compounds [12, 13]. Using the isolated DNA fragment as a probe, we isolated an approximately 7.2-kb DNA fragment and determined 4.8-kb of its nucleotide sequence. The predicted amino acid sequence based on the nucleotide sequence of the DNA fragment showed significant homology to various peptide synthetase genes. From the alignment with various synthetase genes, two amino acid-activating domains of peptide synthetase were detected in the 4.8-kb DNA fragment and these domains were well matched with the previous chemical

analysis of the amino acid components of the antifungal peptide compound, suggesting that the DNA fragment isolated in this study is a part of the peptide synthetase gene of *Bacillus subtilis* 713.

We used *Escherichia coli* DH5 α F' (Gibco BRL, Gaithersburg, U.S.A.) as the host for all plasmid constructions. Luria broth (LB) from Davis *et al.* [4] was used for routine growth of *Escherichia coli* strains and *Bacillus subtilis*. Penassay broth (Antibiotic medium 3, Difco Co., Detroit, U.S.A.) was used for growth of

A.

CORE I		CORE II	
TTI AAA GCA GGC GGI GCI TAT GTG CCG ATC GAC CC		ATI TAC ACI TCT GGI ACI ACA GGI AAG CCA AAA GG	
G G T	T T	T C	
<i>grsA</i> (328) TTA AAA GCA GGT GGA GCA TAT GTT CCG ATT GAT ATT (363)	199	(562) ATT TAT ACT TCT GGT ACA ACA GGC AAT CCA AAA GGT (597)	
L K A G G A Y V P I D I		I Y T S G T T G N P K G	
<i>grsB1</i> (1615) TTA AAA GCT GGT GGA GCT TTT GTG CCT ATT GAT CCT (1650)	184	(1834) ATT TAT ACA TCA GGA ACA ACA GGT AAA CCA AAA GGG (1869)	
L K A G G A F V P I D P		I Y T S G T T G K P K G	
<i>grsB2</i> (4732) CTT AAA GCA GGG GGA GCT TAT TTA CCT CTT GAT CCG (4767)	190	(4957) ATG TAC ACT TCT GGT TCT ACA GGA AAG CCT AAA GGT (4992)	
L K A G G A Y L P L D P		M Y T S G S T G K P K G	
<i>grsB3</i> (7840) TTA AAA GCA GGA GGA GCA TAT GTG CCT ATC GAT ATA (7875)	190	(8065) ATT TAC ACT TCT GGT ACA ACC GGA AAG CCA AAA GGT (8100)	
L K A G G A Y V P I D I		I Y T S G T T G K P K G	
<i>grsB4</i> (10981) TTG AAA GCA GGA GGA GCA TAT GTT CCA ATT GAT CCA (11016)	190	(11206) ATT TAT ACA TCC GGT ACA ACC GGA AAG CCT AAA GGG (11241)	
L K A G G A Y V P I D P		I Y T S G T T G K P K G	
<i>tycA</i> (292) CTT AAA GCA GGC GGA GCC TAT GTG CCC ATC GAC ATC (327)	199	(526) ATT TAC ACC TCA GGC ACG ACA GGC AAG CCA AAA GGC (561)	
L K A G G A Y V P I D I		I Y T S G T T G K P K G	
<i>tycB1</i> (1618) TTG AAG GCA GGC GGC GCA TTT GTG CCG ATC GAC CCG (1653)	190	(1843) ATC TAT ACG TCC GGA ACG ACA GGG CAA CCG AAA GGC (1878)	
L K A G G A F V P I D P		I Y T S G T T G Q P K G	
<i>tycB2</i> (4723) CTC AAA GCG GGC GGG GCG TAT GTG CCG ATC GAT CCG (4758)	187	(4945) ATC TAC ACA TCC GGT ACG ACA GGC AGG CCA AAA GGC (4980)	
L K A G G A Y V P I D P		I Y T S G T T G R P K G	
<i>pcbAB1</i> (1087) TGG AAA TCC GGC GCC GCC TAC GTG CCC ATC GCA CCC (1123)	228	(1351) ACT TAC ACC TCT GGG ACC ACT GGC TTC CCA AAG GGC (1386)	
W K S G A A Y V P I D P		T Y T S G T T G F P K G	
<i>pcbAB2</i> (4369) TGG AAG AGC GGC GGT GCC TAT GTC CCC ATT GAC CCT (4404)	208	(4612) ATC TAT ACC TCT GGA ACG ACA GST CCG CCC AAG GCC (4647)	
W K S G G A Y V P I D P		I Y T S G T T G R P K G	
<i>pcbAB3</i> (7618) TGG AAG GCT GGT GCA GCA TAC GTG CCC TTG GAT CCG (7653)	205	(7858) ATC TTT ACT TCA GGC ACT TCC GGT AAG CCA AAG GGA (7893)	
W K A G A A Y V P L D P		I F T S G T S G K P K G	

B.

Core I : TTI AAA(G) GCA(G) GGC GGI GCI TAT GTG CCG ATC(T) GAC(T) CC

P1

Core II : ATI TAC(T) ACI TCT(C) GGI ACI ACA GGT AAG CCA AAA GG

P2

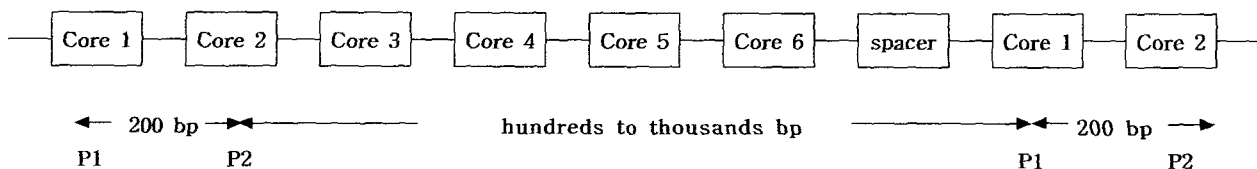


Fig. 1. Sequence of the primers used for polymerase chain reaction.

A. Multiple alignment of Core 1 and Core 2 region of various peptide synthetase domains. B. Primers 1 and 2 were designed based on the conserved sequence of Core 1 and the Core 2 sequence [31]. Primer 1: 5'-GGG(A) TCG(A) ATC CCG ACA TA-3' and primer 2: 5'-GGI ACI ACA GGI AAG CCA AAA GG-3'. PCR with primers 1 and 2 were supposed to give several bands ranging hundreds to thousands bp in length, because they would amplify the region from Core 2 of one domain to Core 1 of the following domain.

Bacillus subtilis for chromosomal isolation. For all standard recombinant DNA techniques, the suggested procedures of Davis *et al.* [4] and Sambrook *et al.* [27] were used. Most enzymes including restriction enzymes, ligase, Klenow, and *Taq* polymerase were purchased from Takara (Shiga, Japan). DNA sequencing was done by the dideoxynucleotide chain termination method for double stranded and single stranded plasmid using α -³²P-dATP (Amersham, Arlington Heights, U.S.A.) and the Sequenase version 2.0 enzyme as suggested by the manufacturer (US Biochemical, Cleveland, U.S.A.). Priming for DNA sequencing was done with custom-made oligonucleotide primers (Takara, Shiga, Japan and Bioneer, Chungwon, Korea) or with standard forward and reverse primers (US Biochemical, Cleveland, U.S.A.). Reaction conditions were as previously described [23]. Chromosomal DNA from the *Bacillus subtilis* was

isolated by a modified version of the method of Ferrari *et al.* [5].

Bacillus subtilis 713 was isolated from Korean soil and was reported to produce antifungal compounds which seem to comprise Arg, Val, Thr, Ala, Asp (or Asn), Pro, and so on [26]. However, the exact structure of the antifungal compound has not been determined. In order to isolate the biosynthetic gene for peptide antibiotics, primers based on the conserved regions — Core 1 and Core 2 — were used for polymerase chain reaction. A primer set was primer 1: 5'-GGG(A) TCG(A) ATC CGC ACA TA-3' and primer 2: 5'-GGI ACI ACA GGI AAG CCA AAA GG-3'. Polymerase chain reaction condition was performed as described by Lee *et al.* [22]. Amplified products were separated by agarose gel electrophoresis. The desired band was eluted from the gel using the Qiaex II Agarose Gel Extract system

<p>1 CAACAGCTCCATGGTCTGTGAGACACAAACCGCCACACACAGAGCAGCGTGTACCTCGAAGCAGCACTGAGTATGCG 90 1 Q T I H G L F E T Q T A H T P E Q P R C T F E P E Q L S Y R 30</p> <p>91 GAACCTAATGAACCTGCGAATCGTTGGCTGCTACTGCGCAGCAAGGTGTGAGCAAGATCGTTTGGTATGATGACCGAAGC 180 31 E L V E R A N R L A R T L R S Q G V T K D R L V G L M T E R 60</p> <p>181 TCGATAGATATGATCGTGGTATGTTAGGCATTTTGAAGCCGAGGTGCATATGTCCTCCATCGACCGAAGTACCAGAGGAGTATC 270 61 S I D M T I V G M L G I L K A G G A Y V P I D P N V P R G R I 90</p> <p>271 CCCTATGTTGGAGCAGCAGAGCTGTTGAGCAGAACCTCTGGTGGCAAGTACGTTTGGCCGTCACATCGCTGGT 360 91 P Y M L L D S G T E L L L T Q S H L V D K V A F D G H M L V 120</p> <p>361 CTGGATGGACCAAGTGTGTATCAGCGAGGTGTCCTCAATCTGGAGTGTGTCGGTGGCAAGCATTCGGCTATGTGATCTATACA 450 121 L D G A Q S V Y H E D G S N L E S L S G P N D L A Y V I V T 150</p> <p>451 TCGGGACACCGGGCAGCCGAAAGGCGTTATCTGGAGATCACCGGCTATGCAATCTGAAAAGCTTTGAAITGGAAITCAAGATC 540 151 S G T I G Q P K G V M L E H H G L C N L K T Y F E F E K I 180</p> <p>541 AGTACTAGATCATATGCTGTTTGGCAGCTCTGTTGATGGCGCTCTGGTGGAAATCTTCCAGCTCTATTCTCGCGAGCCAGC 630 181 S T L D H M L L F A S Y S F D A A C W E I F Q A L F C G A T 210</p> <p>631 TTTGACTCGCAAGCTCAGAGACGATTTGGACTACGAGTGTGAAACATATATGGCGATCACCACATAGGTTGGCAGTTGGCC 720 211 L Y V P T S E T I L D Y R F E T Q Y M A D H H I T V A A L P 240</p> <p>721 CCTACTATGCGGTGTATCTGGAGCCAGCGGATGCAATCTGGCTATCTCTGCTCGCGGTTCAAGCCGCTCAACAGACTGGT 810 241 P T Y A V Y L E P Q R M P N L R L V T A G S A A S T E L V 270</p> <p>811 TACAATGGAAAGTACAGTGGCTACTCAAGTACTGTTGCAACCGAAGCTGGTAGGACATCCATCTGGCCGATTCGCAAGT 900 271 Y K W K D Q V A Y Y N G V Y T E N S V A T S I W P V S K 300</p> <p>901 GAGAGCAGGACAGCTGATTTCCATGGACGCTGCTGCGCAATCATCGGTTATACATGGATGATGACACGGCCATTTGGCTCCGATA 990 301 E R A G Q L L I S I G R P V P N H R V Y M D D V H G H L A P I 330</p> <p>991 GCGCTAGCGGGTACTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCTGCTGGCT 1080 331 G V A G E L C V S G P G L A R G Y L D R P E L T A E K F V P 360</p> <p>1081 AATCGTGTCCGCTGGCGAAGCTGGATGAGGGTACCGCAGCGGCACTGGCAAGATGATGCAAGTAACTGCAAGTAACTGCAAGTAACT 1170 361 N P P P A G E A G Y E R Y R T G D L A R W M P D G N I E Y L 390</p> <p>1171 GCGCGGATGACCAAGTAAAGTCCGCGGTTACCGGTTGAGCTGGCGAAGTGGAGCAAAATATGAAAGTGGAGGACGTACA 1260 391 G R I D H Q V K I R G Y R I E L G E V E A Q I L K V E D V Q 420</p> <p>1261 GAGGCTACTGCTGGCCAGCGGACAGGACAGGACAAATCACTGCTGGCTACTAGCTCCGCAAGAGAGTAAAGTCTGGTGA 1350 421 E V I V L A Q A D E Q Q A D E Q Q N Q L V Y A Y V A E R E V S A G E 450</p> <p>1351 CTGCGGATTTACTGGTGGAGACTGCCCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440 451 L R S L L G E L P N Y M V P S F L V Q L E Q M P R S P A G 480</p> <p>1441 AAAATCGACCAGGCTACTACAGCAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1530 481 K I D R K A L P A P E G S L Q T V A D Y V A P R T W V E V K 510</p> <p>1531 CTGGCCAAATTTGGCAGGATGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1620 511 L A Q I W Q D V L G L T Q V G V K E N F F E I G G H S L R A 540</p> <p>1621 ACAGCTGGCTCGAAGTACAGAGCACTCAACAAGCCCTCCCTCGCCGATTTTTGAGCCACCAAGGATGGACCACTGGCA 1710 541 T L A S K I H K E L N K P L P L R S I F E A P T I E Q L A 570</p> <p>1711 GTGCTAGGCACTGGATCAAGTACATATGCTCCATCCAGTACGGCAAGAGAGCAAGCAAGCTTTCTTCCCTGTCTCTATGCGCA 1800 571 V L E H W I K I R G Y R I E L G E V E A Q I L K V E D V Q 600</p> <p>1801 AACGACTGTATGCTGCTGATCAGTTCGATCAGTATGATGACTACCAAGTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1890 601 N D C M C L H Q F D P H D V N Y N M P S V L Q V T V R W T L 630</p> <p>1891 AACATGAGGCTATCCGCTCACTGATTCGCTGATTCGCTGATTCGCTGATTCGCTGATTCGCTGATTCGCTGATTCGCTGATTCGCTGATTCGCT 1980 631 N M R I R Q L I A R H A I V R T R F D L V D S E P V Q S I 660</p> <p>1981 GAGGACAGCTACCTTGGAGTGAATTTAGCAAGTACAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2070 661 E D T V P P F E V E F T K V Q A E G A T T D T D T R V N K E V 690</p> <p>2071 CAAGAGCTGGTGGTCTTGTGCGCTGTTGATCTCAAAAGTCAACAGCTGCTGCGGCTAGGCTGATAGGCTGATAGGCTGATAGGCTGATAGGCT 2160 691 Q E L V S C F V R P F D L K A C A P L L R V G L L D L G V S G 720</p> <p>2161 CGAGGCGAGCAACCGCAAGTACTACTGCTCGACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2250 721 A G Q E P Q H V L M L D M H H I V S D E V S M G V L T E E F 750</p> <p>2251 GTCCCTGTATGAGCGCAAGCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCTAGCCT 2340 751 V R L Y D D G E L A P L R I Q Y K D L Y A V W Q Q S E V H K 780</p> <p>2341 TGGATCGAGCGAGGAGCCTACTGCTGACACTCTCGGCGGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2430 781 W M Q R O E A Y Y L D D F R G E L P V L D L P T D F A R P S 810</p>	<p>2431 ATACGAGTACCGCAGCGGATACCGTGGTGTGGCTGGACCGGAAGTGGCAACCGCTGAAGGAACCGCGCACACAGCGCTCT 2520 811 I R S T A G D T V V F G L E R E V S E R L K E L A A H T G S 840</p> <p>2521 ACGCTGATATGGTCTGCTGGCGGATACCGCACTCTGCATCAATATGTCGCAAGGACGATCGTTGTCGGCAGCGGATAAAC 2610 841 T L Y W V L L A A Y T A L L H Q Y S G Q E D I V V G T P I V 870</p> <p style="text-align: center;">SphI</p> <p>2611 GGAGCCCGCAGCATGCGGATTTGGAGCAATCTCGGGATGTCCTGGTACCTGGCATTCGCAACTCCGACAGCGGAAACATACC 2700 871 G R P Q H A D L E P I L G M F V G T L A L R N Y P T A E H T 900</p> <p>2701 TTCGCGACTACGTTGGAAGTGAAGTGGCTGCTGCAAGCTACGCAACATCAGGACTATCCATTTGAGGAATGGTGGAAAGCTG 2790 901 F R S Y V E E V K V R A L Q A Y E H Q D Y P F E L V E K L 930</p> <p>2791 AATGTAAGCGGATATGATGCCATCCGCTGACGATACGATTTACCTCCAAACGAGCGGAGGCGGATGCAATGCGGACG 2880 931 A V K R D M S R H P L N D T M F T L Q T T E E G E L Q L A D 960</p> <p>2881 CTGAGCTCCCTCCATCGGGTGGAGCAACCGCGGAAATTTGAGGAAGCGAGCTGGCAGACTGAGCTCCCTCCATCCGGTGG 2970 961 L S F R P Y G L E Q T P A K F F E E G E L A D L S F R P Y G L 990</p> <p>2971 GAGCAACCGCGGAAATTTGACTTCCGCTGGAAGCAGAGGAGGAGCGGCGGATCCGATGCTGATGCAATATGCACTGCGTGT 3060 991 E Q T P A K F D F T L E A R E E E A G I Q F V I A I C T A F 1020</p> <p>3061 GTTCACCGGAACCTGACCGGATGACGACTCCCTTATTCCTGTTAGCCAGCCGTCGCCCAACCGGAGGGAAGGGGGTGG 3150 1021 V S T G V D R M T R P P I R L A E A R P P K P G R E G G E 1050</p> <p>3151 CTGGAATTTAGCAGGAGGACCGAAGGATTTGAGGGGTTCAATGACCGAGCAGAGCTGGTGGCAATCCAGGTT 3240 1051 L E L I T T E G T E Q I L K G F N E P A H R W W N P R F 1080</p> <p>3241 CTGACCAATTTGTCAGCAAGTGTGAGGGCCAGGAAACCGCGCTGGAGGGAATTTGGCGTGGATTCGCAATATGCACTG 3330 1081 V S N I C Q R F E G P G K N R P E G N T L L E S G N M T L 1110</p> <p>3331 ACCTACGCTGAGTGGATGAGCAGCGGCAACCGGTTGGCCGTTGCTACCGGCAAGGTTGTTGAGCAGATCAATCGTAGGCGTACTG 3420 1111 T Y A E L D E Q A N R L A R V L R A K G V E A D Q S V G I L 1140</p> <p>3421 CTGGAAGCTCTCGCGGATCTGGTGGATCTGGGGGACTGAGCGGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3510 1141 L E R S A D L L V S I L G V I K A G G A H A P M D T M Y P Q 1170</p> <p>3511 GAGCGTACTGACTATGCTCAGACAGTGGGCAAGCTGGTGTATACCGACAGAGCCGCGAGCTTCCGCTGAGTCCGCACTCCCA 3600 1171 E R I D Y M L Q D S G A K V I I T S R A A S L S L S L P S T 1200</p> <p>3601 GTACAAGTATGCTTGGATGAGCGGCTCAAGCTCAATGGGAGCAGAGGATGAGCAGCCTGATCCGCTGCGCGGCTGTG 3690 1201 V Q A I V L D D E G V Q A Q W G A E S S N L I P F A G L S 1230</p> <p>3691 CCGCGCTCAATTTGGCTATATGATGACAGCTAGGTAACCGGCTGAGCGGAGGGGTTATGTTGAGCAAGCAGCTGCGGCAAT 3780 1231 P A C I I G L Y D D T S G T I G Q P K G I M I E Q G S V G V 1260</p> <p>3781 CTGATGAGTCTGATGAGCGGATATACAGTGGTATGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3870 1261 L I D A L Y A I V S R Y D R Q P L H I A V L S E P A F D A S 1290</p> <p>3871 GTGAGCAAACTTTGCGAGCT 3960 1291 V K Q I F A S L L L G H T L H V R V R Y V R P E R R A L I A 1320</p> <p>3961 TACTACCGCAGCAGCGGATGATCTGCTGCGGCGCAGCAGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4050 1321 Y Y R T H R I D L S D G T P A H L H I L N E S V S T E A P 1350</p> <p>4051 GATGAGGACTTCTGATCGCGGCGGAGCCTATCCGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4140 1351 D V K H V L I G G E A L S V Q L V G F L H K W S G Y R P V 1380</p> <p>4141 ATTACGAGCTGATGCTGCGCAGGAGCGGACCTGATGCGGCT 4230 1381 I T N V Y G P T E A T V D A T A L T L E I V S L I V L L E 1410</p> <p>4231 AAATTTTACCAAGGATCCATCGGACCGGATAGCAGCAGGCT 4320 1411 K F Y H T V S I G T P I A V Y I L N T R Q Q L V V P N 1440</p> <p>4321 GGGATCGCGGCTGATCTATTCGAGGGGAGGATCCGACAGGATACCTCAAGCTCGGCAAGTCCGCTGAGGATTTGTTGCG 4410 1441 G I A G E L Y I G G A G I A R G Y I N L P E L T A E K F V 1470</p> <p style="text-align: center;">Core 3</p> <p>4411 AATCCGTTGCTGATCGCAGCGGCTGATCCAGCGGACAGCACTTCCGCAACCGGATGACCGGACCGGACCTGCGACCGTGGCT 4500 1471 N P P F A D A T A A D P A D S T Y A N R M Y R T G D L A R W L 1500</p> <p>4501 CCGGATGGCACTATCGGATTTAGCGGATCGACCTACAGTCAAAATTCGCGGATTCGATGCTGCGGAGGCTGGAGGCAAG 4590 1501 P D G S I E Y L G R I D H Q V K I R G Y R I E L G E V E A Q 1530</p> <p style="text-align: center;">Core 4</p> <p>4591 CTGCTGAGCTGAGCGGATCAGAAAGCGGCTGCTGAGGATGGGAAACGAGGAGGAGGATCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4680 1531 L L T V D G I Q K A V V T A F E N E D G H K D L C A V I P R 1560</p> <p style="text-align: center;">PsII</p> <p>4681 CATCACTGCTACCTTCTGCT 4770 1561 H Q C R T R L L L P V F L D T R I F V A S E S L S L P E L R 1590</p> <p>4771 AATGCTGCGAG 4782 1591 N V L Q 1594</p>
---	--

Fig. 2. Nucleotide sequence and their predicted amino acid sequence of DNA fragment isolated from *Bacillus subtilis* 713. Nucleotide sequences are read in both directions. The deduced amino acid sequence is given in single letter code. Conserved peptide synthetase motifs (Core 1 to Core 6 and spacer motif) are underlined. Some restriction sites are also underlined. The nucleotide sequence has been submitted to the GenBank with accession number AF 071566.

(Qiagen, Chatsworth, U.S.A.) and then ligated into pGEM-T Easy vector (Promega, Madison, U.S.A.).

Core 1 region, whose function is not yet clear, and Core 2 region, which functions as an ATP binding site, were used [31] for designing the PCR primers (primers 1 and 2). With these primers and with 30°C as the annealing temperature, we obtained several bands including one major band of 450 bp which was used for further characterization.

In order to confirm that *B. subtilis* 713 chromosome contains the same DNA as the 450 bp PCR product, Southern hybridization was performed under medium stringent conditions using the PCR product as a probe. The probe hybridized with only one fragment for each restriction enzyme digest (data not shown), indicating that this gene exists as a single copy in the *B. subtilis* 713 chromosome.

Alignment of the predicted amino acid sequence based on the nucleotide sequence revealed that the DNA fragment is a part of the peptide antibiotics synthetase gene [22]. In order to isolate the rest of the synthetase gene from the *B. subtilis* 713, we performed Southern hybridization using the PCR product as a probe. From a

mini-library containing 7–8-kb fragments of *Hind*III digested chromosomal DNA from *B. subtilis* 713, we isolated a 7.2-kb DNA fragment (21, Fig. 4) and characterized 4.8-kb of the fragment.

In order to determine the nucleotide sequence of the isolated DNA fragment, we subcloned the fragment into pUC18, 19 and M13mp18, 19 and used forward and reverse primers as sequencing primers. Custom-made primers were also used as sequencing primers (Fig. 4a). Once the nucleotide sequence was determined, we used the DNASIS program (Hitachi software engineering, Japan) to analyse the predicted amino acid sequence (Fig. 2) and then searched the homologous gene using the gapped advanced BLAST and original BLAST [1, 2] program through the Internet. The statistical significance of the predicted protein sequence was evaluated using the DNASIS (Hitachi software engineering, Japan) program and BLAST [1, 2] through a server system supported by NCBI. Highly related sequences usually have an optimized alignment score greater than 150 and E-value smaller than $5e^{-34}$. For multiple alignment to identify Core motifs between amino acids, the CLUSTAL V [7] program was used.

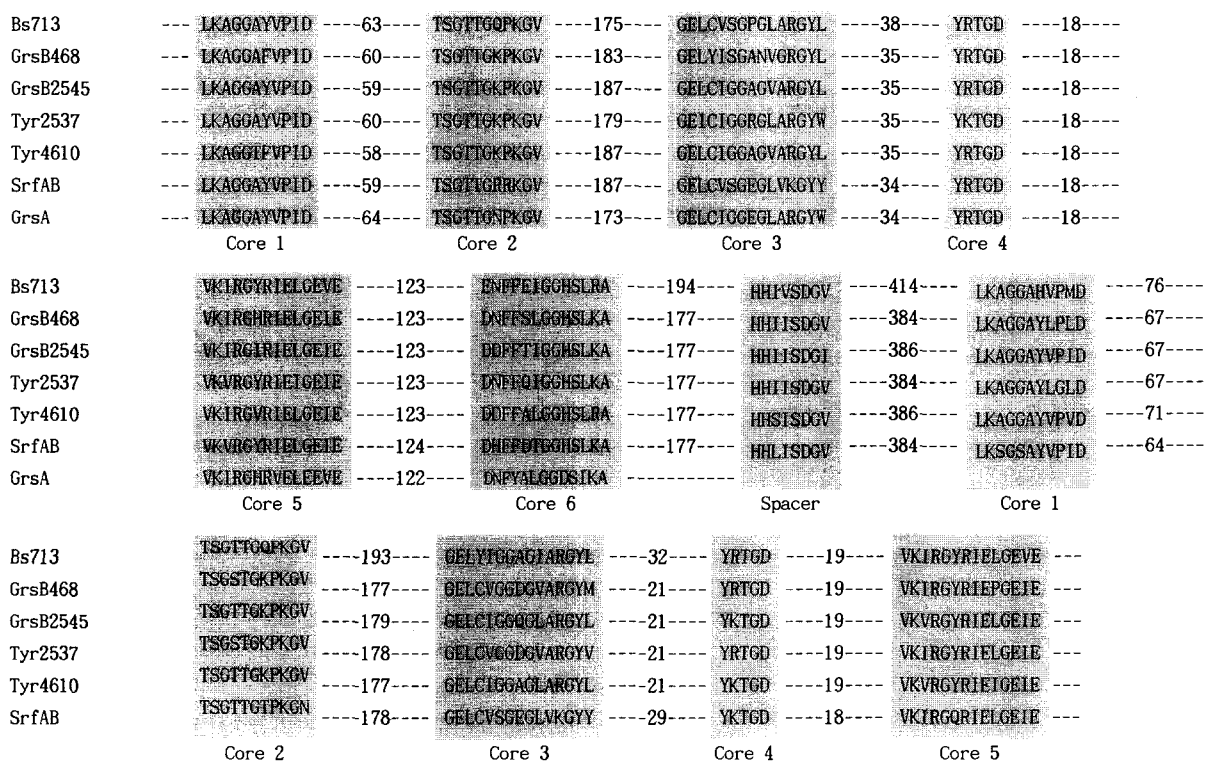


Fig. 3. Alignment of deduced amino acid sequences of DNA fragment from *Bacillus subtilis* 713 with other peptide synthetase domains.

Conserved regions including the 6 Core motifs and spacer motif of different peptide synthetases are presented (shaded and boxed). The distance between the conserved regions is indicated. Grs468 is from the amino acid sequence from aa 468 of gramicidine synthetase 2 including the proline and valine activating domains. GrsB2545 is from the amino acid sequence from aa 2545 of Gramicidine synthetase 2 including the ornithine and leucine activating domains. Tyr2537 is from the amino acid sequence from aa 2537 of tyrocidine synthetase including the tyrosine and valine activating domains. Tyr4610 is from the amino acid sequence from aa 4610 of tyrocidine synthetase including the ornithine and leucine activating domains.

The predicted amino acid sequence of the isolated DNA fragment showed high homology to various peptide antibiotics synthetase genes. The highest homology (41.9% identity over 1588 amino acids) was found with the Gramicidine synthetase gene [8, 18] of *Bacillus brevis*. The surfactin synthetase gene of *Bacillus subtilis* [6] and tyrocidine synthetase gene of *Bacillus brevis* also showed significant homology (39.5% identity over 1581 amino acids and 37.3% identity over 877 amino acids, respectively). These homologies are so remarkable that we might conclude that the DNA fragment we isolated was part of the peptide synthetase gene. Among the peptide synthetase genes, the regions of the proline and valine activating domains showed the highest homology. The fact that proline and valine seem to be the constituents of the peptide produced by *Bacillus subtilis* 713 [26] is very interesting in this sense. That is, the two domains identified in this study are likely to be the domains for the activation of proline and valine for peptide synthesis. This finding also supports the idea that the DNA fragment isolated in this study is a fragment of the biosynthetic gene for antifungal peptide produced by *B. subtilis* 713. However, this should be proven by further genetic and biochemical characterization.

By comparison of the deduced amino acid sequence with databases, two peptide synthetase domains containing

well-conserved peptide synthetase motifs were detected (Figs. 2 and 3). We will refer to the two domains as domain I and II hereafter. Peptide antibiotics biosynthetic genes have been reported to contain the conserved regions in each domain which are involved in ATP binding (Core 2, Core 3, and Core 5), the regions involved in ATP hydrolysis (Core 4), and the region involved in 4'-phosphopanthine binding (Core 6; 15, 16).

We could locate 6 conserved Core motifs and a spacer motif in domain I, and 5 conserved Core motifs in domain II (Figs. 2 and 3). That is, two Core 1 motifs (LKAGGAYVPID and LKAGGAHVPM) were identified in the regions of aa 72 - aa 82 and aa 1155 - aa 1165. These sequences showed 11/11 and 9/11 identity to the consensus sequence of Core 1 (LKAGGAYVPID), respectively. Two Core 2 motifs (TSGTTGQPKG) were identified in the regions aa 150 - aa 160 and aa 1241 - aa 1251 to have 10/11 identity to consensus (YSGTTGxPKG). Two Core 3 motifs (GELCVSGPGLARGYL and GELYIGGAGIARGYL) were identified in the regions aa 334 - aa 348 and aa 1444 - aa 1458 to have 13/15 and 14/15 identity to consensus (GELCIGGxGxARGYL), respectively. Two Core 4 motifs (YRTGD) were identified in the regions aa 373 - aa 377 and aa 1491 - aa 1495 to have 5/5 identity to consensus (YxTGD). Two Core 5

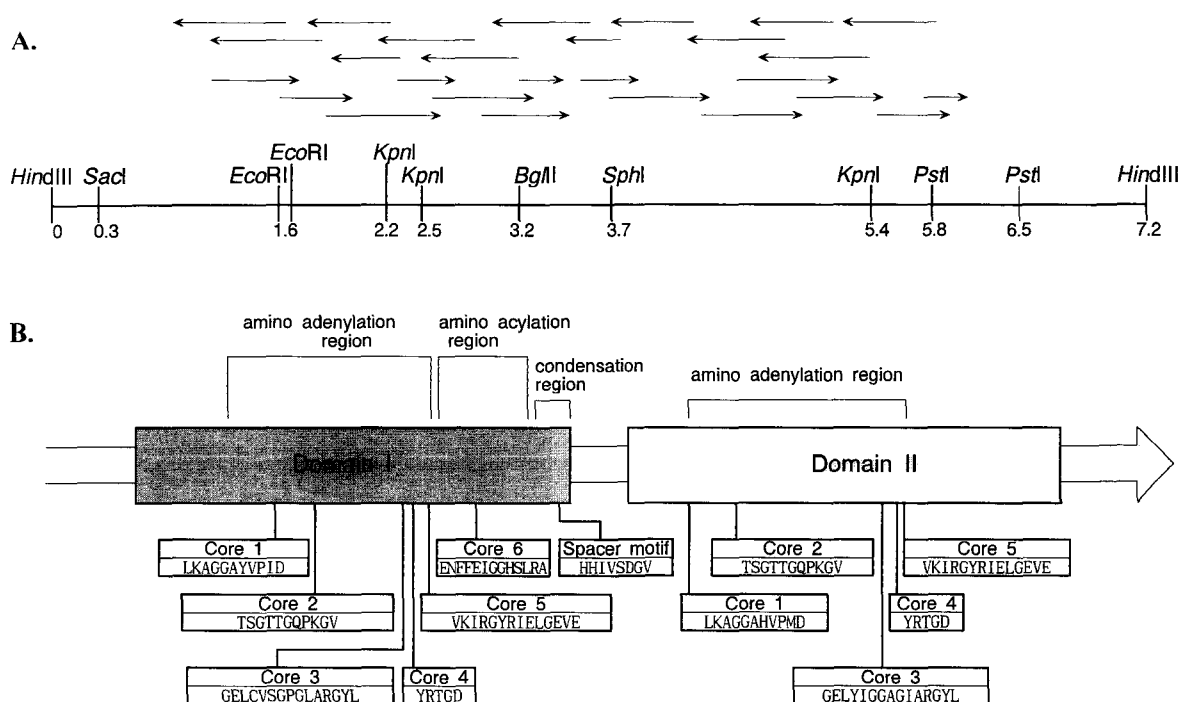


Fig. 4. Genetic organization of 7.2 Kb DNA fragment isolated from *Bacillus subtilis* 713.

A. Arrow indicates the direction and the length of sequence read by the sequencing reaction. Some restriction enzyme sites which were used for subcloning were indicated. The map is not on the scale. B. Schematic representation of two domains based on the sequence homology to various amino acid activating domains. Domain I and Domain II showed the highest homology to the proline activating domain and the valine activating domain, respectively.

motifs (VKIRGYRIELGEVE) were identified in the regions aa 397 - aa 410 and aa 1515 - aa 1528 to have 13/14 identity to consensus (VKIRGxRIELGEIE). One Core 6 motif (ENFFEIGGHSL) was identified in the region aa 528 - aa 538 to have 8/11 identity to consensus (DNFYxLGGHSL). The spacer motif (HHIVSDGV) was identified in the region aa 734 - aa 741 to have 6/8 identity to the consensus sequence (HHILxDGW; 31). In the fragment we characterized, however, Core 6 and spacer motif were not included in domain II. The preliminary result indicated that the spacer motif was located in the 7.2-kb DNA fragment we isolated.

In conclusion, we identified all six conserved Core motifs in Domain I and five Core motifs in Domain II including lysine in Core 2, aspartate in Core 4, and serine in Core 6 which have been known to play important roles [31]. Figure 4b shows the simplified scheme of the peptide synthetase gene fragment from *Bacillus subtilis* 713 based on the sequence homology to conserved core and spacer motifs.

In most peptide synthetase genes, the spacer motif is usually found at the end of each domain. It is also located at the N-terminal region of the domain if the domain is involved in the first step of peptide synthesis. Therefore, if the second domain has its own promoter to be the first domain for peptide synthesis, there should be two spacer motifs between the two domains. The fact that there is only one spacer motif between Domain I and Domain II in the DNA fragment we isolated suggests that these two domains are under the control of the same promoter since the spacer motif is missing at the N-terminal end of Domain II.

Interestingly, the region after the spacer motif showed high homology with *grsB* and *tyrC*, which do not have motifs for racemization motifs [24, 25, 31], which are found in the domain for racemized amino acids. These results indicate that the isolated peptide synthetase fragment is not involved in racemization. This finding should give some clue as to the structure of the peptide produced by *B. subtilis* 713 which has not yet been determined.

The isolation and characterization of domains for the activation of amino acid constituents of peptide synthetase genes can be a foundation for the development of noble antibiotics. Together with the bank of peptide synthetase genes already reported, the two amino acid-activating domains identified in this study can be used for designing noble and more efficient peptide antibiotics along with a bioassay against a wide spectrum of pathogens. In this sense, a recent study by the Stachelhaus group has proven that programming new antibiotics by such gene manipulation as switching, deleting, and adding the amino acid-activating domains could create noble peptides of amino acid constituents they designed [30].

Acknowledgments

This work was partially supported by the Genetic Engineering Program of the Ministry of Education in 1997 to J.-W. Suh and by the Postdoctoral Program of the Korea Research Foundation to S.-Y. Lee.

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acid. Res.* **25**: 3389–3402.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Borchert, S., S. Patil, and M. A. Marahiel. 1992. Identification of putative multifunctional peptide synthetase genes using highly conserved oligonucleotide sequences derived from known synthetases. *FEMS Microbiol. Lett.* **92**: 175–180.
- Davis, R. W., D. Botstein, and J. R. Roth. 1980. *Advanced Bacterial Genetics; A Manual for Genetic Engineering*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, U.S.A.
- Ferrari, F., D. Lang, E. Ferrari, and J. A. Hoch. 1982. Molecular cloning of the *spoOB* sporulation locus in bacteriophage lambda. *J. Bacteriol.* **152**: 809–814.
- Fuma, S., Y. Fujishima, N. Corbell, C. D'Souza, M. M. Nakano, P. Zuber, and K. Yamane. 1993. Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. *Nucl. Acid. Res.* **21**: 93–97.
- Higgins, D. G. 1994. CLUSTAL V: Multiple alignment of DNA and protein sequences. *Methods Mol. Biol.* **25**: 307–318.
- Hori, K., Y. Yamamoto, K. Tokita, F. Saito, T. Kurotsu, M. Kando, K. Okamura, J. Furuyama, and Y. Kurotsu. 1991. The nucleotide sequence for a proline-activating domain of gramicidine S synthetase gene from *Bacillus brevis*. *J. Biochem.* **110**: 111–119.
- Jung, G. 1991. Lantibiotica-ribosomal synthetisiete polypeptidwirkstoffe mit sulfidbrücken und α,β -dihydroamino sauren. *Angew. Chem.* **259**: 14158–14164.
- Kim, D.-O., K. Park, J.-W. Song, and J.-H. Seo. 1997. Enhanced production of maltotetraose-producing amylase by recombinant *Bacillus subtilis* LKS88 in fed-batch cultivation. *J. Microbiol. Biotechnol.* **7**: 417–422.
- Kim, H.-S., C.-H. Lee, H.-H. Suh, K.-H. Ahn, H.-M. Oh, G.-S. Kwon, J.-W. Yang, and B.-D. Yoon. 1997. A lipopeptide biosurfactant produced by *Bacillus subtilis* C9 selected through the oil film-collapsing assay. *J. Microbiol. Biotechnol.* **7**: 180–188.
- Kim, S.-W., K.-H. Sohn, S.-K. Kim, H.-W. Lee, N.-K. Lee, and S.-H. Bok. 1991. Development of noble secondary

- metabolites by microbial strains search (III). Research report of Ministry of Science and Technology, Korea.
13. Kim, S.-W., J.-W. Lee, S.-H. Lee, and S.-H. Bok. 1991. Isolation and characterization of new bacterial strains which produce antifungal compounds. *Kor. J. Appl. Microbiol. Biotechnol.* **19**: 337–342.
 14. Kim, Y.-S. and S.-D. Kim. 1994. Antifungal mechanism and properties of antibiotic substances produced by *Bacillus subtilis* YB-70 as a biological control agent. *J. Microbiol. Biotechnol.* **4**: 296–304.
 15. Kleinkauf, H. and H. V. Dören. 1996. A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.* **236**: 335–351.
 16. Kleinkauf, H. and H. V. Dören. 1990. *Biochemistry of Peptide Antibiotics*, 1st ed., pp. 103–127. Walter de Gruyter and Co., Berlin, Germany.
 17. Kleinkauf, H. and H. V. Dören. 1995. Linking peptide and polyketide biosynthesis. *J. Antibiot.* **48**: 563–567.
 18. Krätzschmar, J., M. Krause, and M. A. Marahiel. 1989. Gramicidine 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.
 19. Lancini, G., F. Parenti, and G. G. Gallo. 1995. *Antibiotics*. 2nd ed., Plenum Press, New York, U.S.A.
 20. Lee, H.-W., J.-W. Choi, H.-W. Kim, D.-P. Han, W.-S. Shin, and D.-H. Yi. 1997. A peptide antibiotic AMRSA1 active against multidrug-resistant *Staphylococcus aureus* produced by *Streptomyces* sp. HW-003. *J. Microbiol. Biotechnol.* **7**: 402–408.
 21. Lee, J.-W. 1995. Isolation and characterization of peptide antibiotics synthetase from *Bacillus subtilis* isolated from Korean soil. Master thesis. Myongji University.
 22. Lee, S.-Y., S.-K. Rhee, C.-H. Kim, and J.-W. Suh. Rapid and efficient isolation of biosynthetic genes for peptide antibiotics from Gram positive bacteria. *J. Microbiol. Biotechnol.* (submitted).
 23. Lee, S.-Y. and C. W. Price. 1993. MinCD locus of *Bacillus subtilis* lacks the *minE* determinant that provides topological specificity to cell division. *Mol. Microbiol.* **7**: 601–610.
 24. Marahiel, M. A., M. Krause, and H.-J. Skarpeid. 1985. Cloning of the tyrocidine synthetase 1 gene from *Bacillus brevis* and its expression in *E. coli*. *Mol. Gen. Genet.* **201**: 231–236.
 25. Mittenhuber, G., R. Weckermann, and M. A. Marahiel. 1989. Gene cluster containing the genes for tyrocidine synthetase 1 and 2 from *Bacillus brevis*: Evidence for an operon. *J. Bacteriol.* **171**: 4881–4887.
 26. Moon, S.-S., N.-S. Nahm, and S.-I. Lee. 1995. Structure determination of natural products from bacterial strain (II). Research report of Ministry of Science and Technology, Korea.
 27. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, U.S.A.
 28. Schnell, N., K.-D. Entan, U. Schneider, F. Gotz, H. Zahner, and G. Jung. 1988. Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature (London)* **333**: 276–278.
 29. Sonenshein, A. L., J. M. Hoch, and R. Losick. 1993. *Bacillus subtilis and other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*. pp. 897–916. American Society for Microbiology, Washington, D. C., U.S.A.
 30. Stachelhaus, T., A. Schneider, and M. A. Marahiel. 1995. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**: 69–72.
 31. Stachelhaus, T. and M. A. Marahiel. 1995. Modular structure of encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125**: 3–14.
 32. Strohl, W. R. ed. 1997. *Biotechnology of Antibiotics*. 2nd ed., Marcel Decker Inc., New York, U.S.A.