The Linear Pentadecapeptide Gramicidin Is Assembled by Four Multimodular Nonribosomal Peptide Synthetases That Comprise 16 Modules with 56 Catalytic Domains*  

Received for publication, September 2, 2003, and in revised form, December 1, 2003  
Published, JBC Papers in Press, December 11, 2003, DOI 10.1074/jbc.M309658200

Nadine Kessler‡, Holger Schuhmann§, Sabrina Morneweg§, Uwe Linne, and Mohamed A. Marahiel¶  
From the FB Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, 35032 Marburg, Germany

Linear gramicidin is a membrane channel forming pentadecapeptide that is produced via the nonribosomal pathway. It consists of 15 hydrophobic amino acids with alternating L- and D-configuration forming a β-helix-like structure. It has an N-formylated valine and a C-terminal ethanolamine. Here we report cloning and sequencing of the entire biosynthetic gene cluster as well as initial biochemical analysis of a new reductase domain. The biosynthetic gene cluster was identified on two nonoverlapping fosmids and a 13-kilobase pair fragment covering a region of 74 kbp. Four very large open reading frames, lgrA, lgrB, lgrC, and lgrD with 6.8, 15.5, 23.3, and 15.3 kbp, were identified and shown to encode nonribosomal peptide synthetases with two, four, six, and four modules, respectively. Within the 16 modules identified, seven epimerization domains in alternating positions were detected as well as a putative formylation domain fused to the first module LgrA and a putative reductase domain attached to the C-terminal module of LgrD. Analysis of the substrate specificity by phylogenetic studies using the residues of the substrate-binding pockets of all 16 adenylation domains revealed a good agreement of the substrate-binding pockets of all 16 adenylation domains. Additional biochemical analysis of the three adenylation domains in modules 1, 2, and 3 confirmed the colinearity of this nonribosomal peptide synthetase assembly line. Module 16 was predicted to activate glycine, which would then, being the C-terminal residue of the peptide chain, be reduced by the adjacent reductase domain to give ethanolamine, thereby releasing the final product N-formyl-pentadecapeptide-ethanolamine. However, initial biochemical analysis of this reductase showed only a one-step reduction yielding the corresponding aldehyde in vitro.

Gramicidin is a pentadecapeptide antibiotic produced by Bacillus brevis ATCC 8185 during its sporulation phase (1). The primary structure of gramicidin A was determined as formyl-val-gly-alan-val-val-val-trp-val-trp-val-trp-ethanolamine. Gramicidins A, B, and C refer to the naturally occurring isoforms, gramicidin B and C, having either phenylalanine or tyrosine replacing tryptophan at position 11, respectively. Gramicidin D refers to the naturally produced mixture of gramicidins A, B, and C of ~80% A, 5% B, and 15% C (3). In all three gramicidin isoforms, an isoleucine residue instead of a valine one at position 1 has been observed (~5% (2)). Several facts about the sequence of gramicidin are striking. First, the amino acid sequence contains solely hydrophobic residues. Second, the N terminus is blocked by N-formylation of the first residue (valine), and third, the C terminus is blocked with ethanolamine. These three features provide the reason for the high insolubility in water but very good solubility in various organic solvents as gramicidin is unable to adopt a net charge or form a zwitterion at any pH. The last important feature is the alternating L- and D-amino acid composition except for position 2 (glycine). Through this, the molecule forms a helix where all side chains point outwards, resulting in the formation of a β-helix-like channel. Usually two molecules of gramicidin dimerize, giving either a double helix ("pore" form) or a helical dimer ("channel" form) (for more details, see Ref. 4). Once inserted into a membrane, it is specific for the transport of monovalent cations across the bilipid layer and thus collapses the transmembrane ion potentials. The above given features about the sequence of gramicidin have let to the assumption that it must be synthesized via the nonribosomal pathway.

The nonribosomal peptide synthetases (NRPSs) have been researched intensively over the past decades, and their mechanism is well understood by now. NRPSs are large multifunctional enzymes, carrying out several reactions in a specific and coordinated manner. Their products are a large and diverse group of bioactive natural products. These products typically consist of 3–22 carboxyl or amino acids including the nonproteinogenic ones, meaning that the generated structure is of an extremely broad diversity. As these products often contain hydroxy-, N-methylated, glycosylated, or D-amino acids that contribute to their biological activity, they have many applications in medicine (e.g. penicillin, vancomycin, and cyclosporine), agriculture, and biochemical research.

NRPSs are composed of modules where each module possesses all catalytic units to activate, covalently bind an amino (or carboxyl) acid, and perform a condensation reaction through

* This work was in part funded by the Deutsche Forschungsgemeinschaft. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Recipient of Ph.D. fellowships from the Graduiertenkolleg "Proteinfunktion auf atomarer Ebene" and from the Studienstiftung des deutschen Volkes.

‡ Both authors contributed equally to this work.

¶ To whom correspondence should be addressed. Tel.: 0049-6421-282-5722; Fax: 0049-6421-282-2191; E-mail: marahiel@chemie.uni-marburg.de.

The abbreviations used are: NRPS, nonribosomal peptide synthetases; kbp, kilobase pairs; ORF, open reading frame; PCP, peptidyl carrier proteins; HPLC, high pressure liquid chromatography; RBS, ribosome binding site; aa, amino acids; A, adenylation; C, condensation; E, epimerization; F, formylation; M, methylation; R, reductase; Cy, cyclization; Te, termination; formyl-THF, N₅-formyl-tetrahydrofolate.

This paper is available on line at http://www.jbc.org
a peptide bond formation. Each module can be divided further into domains where each domain carries out a specific catalytic reaction repeatedly. Three core domains of NRPSSs have been identified as being the minimal requirement needed, namely adenylay (A), thiolation (PCP), and condensation (C) domains (5). The A domain selects the cognate amino acid specifically and activates it as aminoacyl adenylate at the expense of ATP (6, 7). Next, the activated amino acid is transferred onto the thiol moiety of the downstream PCP domain, giving an energy-rich thioester bond (8). The C domain is located between two adjacent A-PCP domain pairs. It catalyzes the condensation of the thioester-bound intermediates, thereby elongating the peptide chain by 1 amino acid (9). The chain that is then attached to the downstream PCP domain is subsequently used in the condensation reaction of the next C domain and by consequent elongation handed on until it reaches the terminal PCP domain of the NRPSS. The peptide chain is then released by either a termination (Te) or a reductase (R) domain. A Te domain gives cyclic, branched cyclic, or hydrolyzed products (10), whereas an R domain performs a reduction step forming an aldehyde or alcohol at the C terminus (11). Optional domains have been found to modify the bound substrate through an epimerization (E (12, 13)), N-methylation (M (14)), or cyclization (Cy (15)) reaction, enlarging the above mentioned structural diversity of the product. A first example for a putative formylation (F) domain has been found in Anabaena strain 9114 (16). We describe here the cloning and sequencing of the gramicicin biosynthetic gene cluster encoding four large NRPSSs and provide biochemical data for the colinearity of this gigantic assembly line.

**General Methods—Antibiotics** were used in the following concentrations: 50 μg/ml kanamycin; 100 μg/ml ampicillin; 12.5 μg/ml chloramphenicol. All oligonucleotides used in this study were purchased from MWG Biotech or Qiagen Operon. Cells were usually grown in LB medium at 37 °C overnight for plasmid and fosmid preparations. Plasmid and fosmid preparations were carried out using the QiAprep spin Miniprep kit (Qiagen). All restriction and modification enzymes were purchased from New England Biolabs. Blast searches were carried out using either the BLASTX function or the BLASTP function at the NCBI homepage (www.ncbi.nlm.nih.gov/blast/Blast.cgi (17)). Multiple sequence alignments were made using MegAlign (DNASTar, GATC Biotech). Kinetic data were analyzed using Sigma Plot 8.0 with the Enzyme Kinetics Module 1.1 (SPSS Inc.).

**Isolation of Genomic DNA and PCR with Degenerate Primers—**In this study, all genomic DNA from *B. brevis* ATCC 8185 was isolated and purified using the Genomic-tip 100/G kit (Qiagen) as described in the manufacturer’s manual. From a DNA sequence alignment of the conserved cores E5 and E6 of E domains, the following primers were deduced and successfully used in a PCR: NK 5, 5'-aagc cagc tata ccca (tgcc ga) atc tgtg-3', and NK 6, 5'-cagc tata ccca (tgcc ga) atc tgtg-3'. The E domains used in the alignment were from *Bacillus subtillis* ATCC 21332 (srfa), *B. brevis* ATCC 9999 (grsa), and *B. brevis* ATCC 8185 (tyca, tybc). The PCR was carried out using the Roche Applied Science PCR kit using the following parameters for annealing: 5 cycles, 60 °C/60 °C; 5 cycles, 57.5–0.4 °C/Cycle; and 5 cycles, 57 °C. Elongation time was set to 5 min and 20 s. Only two of the DNA bands observed were dependent upon the addition of both primers to the PCR, having the size of ~7 and ~10.5 kb, the latter one being the corresponding fragment from the tyc operon. The 7-kbp fragment was extracted from the gel using Qiagen II (Qiagen) and cloned into pCR-XL-TOPO (Invitrogen) according to the manufacturer’s manual, and one of the few resulting clones carried the right insert as confirmed by sequencing. Sequencing reactions were carried out by the chain termination method (18) with dye-labeled deoxy terminators from a PRISM ready reaction dyeoxy terminator cycle sequencing kit with AmpliTaq FS polymerase (Applied Biosystems) according to the manufacturer’s protocol and analyzed on an ABI 310 genetic analyzer.

**Generating a Fosmid Library—**The fosmid library was generated using the CopyControl fosmid library production kit (Epiconter) according to the manual except that four times the recommended amount of DNA was used. Shotgun sequencing of the selected fosmids was carried out in publication quality by Qiagen with a coverage of 13- and 16-fold of fosmids 3 and 5, respectively.

**Cloning, Overproduction, and Purification of FAT, ATE, CAT, and PCP—**All constructs were cloned using the pBAD directional TOPO kit (Invitrogen) as described in the manual. The corresponding gene fragments were: FAT from bp 11,945 to 14,244 (2,298 bp), ATE from bp 15,521 to 17,873 (3,342 bp), CAT from bp 18,795 to 21,899 (3,104 bp), and PCP-R from bp 71,410 to 72,858 (1,448 bp). The corresponding vectors were transformed into Escherichia coli BL21 (DE3), and the desired proteins were overproduced as described in the manual. The recombinant proteins carried an N-terminal thioredoxin-fusion protein and a C-terminal V-epitope with His6 tag and were purified using Ni2⁺-nitrioltriacetic acid affinity chromatography as described before (20). Purified proteins were controlled via SDS-PAGE (21) and concentrated using Vivapen 20-kDa concentrators (membrane 50,000 MWCO polyethersulfone (PES); purchased from Vivascience/Sartorius) to a concentration of about 3–5 mg/ml. 

**Artificial Models—**Kinetic Studies of FAT, ATE, CAT, and PCP—Expression Assays (16) of the recombinant proteins were overproduced and purified by a 1100 series, Agilent) coupled with a 1100MSD-A ESI-Quadrupol mass spectrometer (Agilent). Samples (95 μl of each) were applied to a 125/2-Nucleosur-C18-Gravity column (Macherey-Nagel) with a particle diameter of 3 μm. The gradient of solvent A (water, 0.05% formic acid) and solvent B (acetonitrile) used was as follows: linear from 30% B to 55% B within 10 min, increasing to 95% B within 2 min, and holding 95% B for additional 3 min at a flow rate of 0.3 ml/min and a column temperature of 45 °C.

UV detection was carried out at 215 nm. Mass-sensitive detector (HPLC (1100 series, Agilent) coupled with a 1100MSD-A ESI-Quadrupol mass spectrometer (Agilent). Samples (95 μl of each) were applied to a 125/2-Nucleosur-C18-Gravity column (Macherey-Nagel) with a particle diameter of 3 μm. The gradient of solvent A (water, 0.05% formic acid) and solvent B (acetonitrile) used was as follows: linear from 30% B to 60% B within 10 min, increasing to 95% B within 2 min, and holding 95% B for additional 3 min at a flow rate of 0.3 ml/min and a column temperature of 45 °C.
gramicidin contains 6 D-amino acids, we postulated that the corresponding NRPS should be of the modular structure F-A-T-C-A-PCP-(C-A-PCP-C-A-PCP-E)_{5}C-A-PCP-C-A-PCP-R. The PCR strategy using degenerate primers as well as the following inverse PCR was successful (see “Experimental Procedures”). Sequencing and analysis of the amplified DNA fragment revealed an NRPS with the domain structure ‘A-PCP-C-A-PCP-E’ of a so far undetermined origin. According to the specificity conferring code (23), the substrate specificity of the identified four A domains was determined to be Ala, Val, Val, and Val, respectively, meaning that the structural element of 1-Ala-Val-1-Val-Val should be present in the product. As this is the case in gramicidin, we then generated a fosmid library. Using probes from the DNA region encoding the first A and the last E domain of the 13-kbp fragment in a Southern blot screen, we identified and sequenced two nonoverlapping fosmids covering the whole up- and downstream region of the fragment, yielding a total of 74 kbp of genetic information (accession number AJ566197). The overall G + C content of the region sequenced is 55.43%, which is slightly higher than the G + C content of the tyc operon (20) and higher than that reported for other bacilli.

The lgr Synthetases Are Encoded by Four Large ORFs Termed lgrA, lgrB, lgrC, and lgrD—Sequence analysis revealed 11 significant open reading frames as judged by blast searches (17) (Fig. 1), and among them, revealed four very large ORFs spanning almost 61 kbp of the region sequenced with a G + C content of 56.3%. The first of these four ORFs, lgrA (6,822 bp), begins with a GTG start codon at position 11,945 bp of the region sequenced, preceded by a putative RBS. The gene product, LgrA (2273 aa and 257,822 Da), has high similarity to reductases from polyketide synthases and other NADPH-dependent reductases such as MxcG (11) and SafA (25). We expect the four A domains to activate Trp, Leu, Trp, and Gly according to the product, strongly suggesting that ethanolamine is the product of a reduced glycine residue. The prediction is again unclear for the Trp-activating A domains 1 and 3 but clearly Leu for module 2 and Val for module 4. According to the product, we expect the specificity to be Ala, Leu, Ala, and Val, respectively. However, module 1 has the highest similarities regarding the binding pocket with the second module of LgrA (Gly). lgrC (23,271 bp) starts with a ATG 29 bp downstream of the stop codon of lgrA and is preceded by 7 bp by a putative RBS. LgrB (5162 aa and 577,862 Da) harbors four modules with the domain structure (C-A-PCP-C-A-PCP-E)_{2}. The specificity of the A domain specificity is difficult for modules 1 and 3 but clearly Leu for module 2 and Val for module 4. According to the product, we expect the specificity to be Ala, Leu, Ala, and Val, respectively. However, module 1 has the highest similarities regarding the binding pocket with the second module of LgrA (Gly). lgrC (23,271 bp) starts with an ATG 23 bp upstream of the stop codon of lgrB and is also preceded by a putative RBS. The corresponding gene product LgrC (7756 aa and 866,306 Da) is composed of six modules bearing the domain structure (C-A-PCP-C-A-PCP-E)_{3}. The specificity is expected to be Val-Val-Trp-Leu-(Trp/Tyr/Phe)-Leu according to the product. The prediction confirms this for the Val- and Leu-activating domains 1, 2, 4, and 6 but remains unclear for modules 3 and 5. The last large ORF, lgrD (15,258 bp), starts 76 bp downstream of the stop codon of lgrC and is also preceded by a putative RBS. LgrD (5085 aa and 567,448 Da) consists of four modules with the domain structure C-A-PCP-C-A-PCP-E-C-A-PCP-C-A-PCP-R. A putative termination loop (35 bp) starts 7 bp after the stop codon. The putative reductase domain at the C terminus of LgrD has high similarity to reductases from polyketide synthases and other NADPH-dependent reductases such as MxcG (11) and SafA (25).
specify the A domains shows clearly that the A domains that activate similar or identical amino acids cluster (Fig. 2).

**Amplification, Expression, and Biochemical Investigation of the First Three Internal A Domains**—DNA fragments encoding module 1 (FAT), module 2 lacking the C domain (ATE), and the complete module 3 (CAT) were amplified from the corresponding fosmid and cloned into pBAD vectors as described under Experimental Procedures. The enzymes were overproduced in E. coli BL21 (DE3) as His\(^6\)-tagged proteins and purified using Ni\(^{2+}\)-nitrilotriacetic acid affinity chromatography. All proteins were obtained in soluble form in good yield and purity. Determination of the A domain specificity (Fig. 3) was carried out as described above. We found module 1, FAT, to activate L-Val (100%), L-Ile (49%), and to a minor extent, L-Leu (12%; the highest activity was for Ile); the second substrate, L-Leu (50%), and to a minor extent, L-Pro (49%); and L-Val (5%). The apparent Km values for Gly and Ala were determined to be 2.2 and 1.5 mM, respectively, showing a preference of the A domain for Ala, which is solely found in the C-terminal region. Taking a closer look at this F domain of ApdA from Anabaena Strain 90 (size ~550 aa), we found a stretch of about 270 aa at its C terminus, which clearly belongs to the C-terminal domain of a poorly conserved C domain.

**E Domain of LgrA**—As the alignment of all seven lgr-E domains shows (Fig. 4), the first E domain is not as well conserved as the others. Core motif 7 is missing, and motifs 1, 2, and 4 are only poorly conserved. The catalytically essential His (H\(^\text{HXSIDG(WV)}\)) in core 2 (26) has been mutated to a Gln residue, suggesting this E domain to be inactive. This is supported by the fact that the cognate amino acid is Gly, which is achiral.

**Formylation Domain**—The first 200 aa of LgrA show high similarity to methionin-tRNA-formyltransferases from other bacteria such as Bacillus anthracis (33% identity, 52% similarity), Fusobacterium nucleatum (34% identity, 58% similarity), or Clostridium tetani (35% identity, 56% similarity; Fig. 5). However, the overall similarity with the putative F domain from Anabaena strain 90 (16) is rather low (only 17.2% identity). However, we found 35% identity and 50% similarity when comparing only the amino acids 70–170 of both proteins (N-terminal region). Taking a closer look at this F domain of ApdA from Anabaena Strain 90 (size ~550 aa), we found a stretch of about 270 aa at its C terminus, which clearly belongs to the C terminus of a poorly conserved C domain.

From sequence analysis, we propose N\(\text{\text{H}}_{10}\)-formyl-tetrahydrofolate (formyl-THF) to be the C1 carrier that donates the formyl group for the N-formylation of the first amino acid Val. As shown in the alignment (Fig. 5), the formyl-THF-binding motif SLLP is present, located within the conserved N-terminal part as is the case for other formyltransferases (27). From the align-
For myxochelin, it has been shown that further reduction of the aldehyde, resulting in myxochelin A (reduction to alcohol), is also carried out by MxcG itself (11); alternatively, MxlC, an aldehyde aminotransferase, accepts the aldehyde substrate and transaminates it to myxochelin B (29).

To prove whether the putative reductase of LgrD harbors a reduction activity, we cloned, overproduced, and purified the last two domains of LgrD (PCP-R) as described under "Experimental Procedures." For the preliminary biochemical investigations described here, we used as substrate mimic Ac-D-Leu-L-Trp-L-Leu-L-Trp-Gly-S-CoA for the following reason. The 4′-phosphopantethenyl-transferase Sfp from B. subtilis is able to use peptidyl-CoA substrates when performing the apo to holo modification of peptidyl carrier proteins in vitro (30). When using peptidyl-CoA under assay conditions, the result is a holo enzyme already modified with a designated peptide chain, which serves as a substrate for downstream enzyme(s) (31).

Using this pentapeptidyl-CoA substrate, we have strong evidence that the Ac-D-Leu-L-Trp-L-Leu-L-Trp-Gly-4′-phosphopantethenothione was covalently attached to the PCP adjacent to the reductase, and we tested this construct for reductive release. Upon incubation of Ac-D-Leu-L-Trp-L-Leu-L-Trp-Gly-S-CoA, the enzyme (PCP-R), Sfp, and NADPH at four different pH values ranging from 5 to 8 and subsequent analysis with HPLC-mass spectrometry, we clearly identified the product corresponding to a one-step reduction of the substrate to Ac-D-Leu-L-Trp-D-Leu-L-Trp-aminoethanol (M + H+, 700.3 atomic mass units; M + Na+, 722.3 atomic mass units; data not shown). The product did not occur in a negative control without the enzyme. When omitting Sfp, we observed minimal product formation (aldehyde) at pH 6, which was slightly higher at pH 7 and 8 (data not shown), probably due to the fact that the enzyme is able to utilize the peptidyl-CoA substrate present in the solution to some minor extent. In control samples lacking only NADPH, we observed aldehyde formation in the range of about 2–3% (data not shown), which is probably due to a small amount of NADPH that was still attached to or trapped inside the enzyme after purification and dialysis.

A Putative Type II Thioesterase Is Located Upstream of lgrA—The ORF termed lgrT (792 bp) starts with an ATG and is preceded by a putative RBS, located 464 bp upstream of lgrA. LgrT (263 aa and 30,102 Da) has high similarity to GrsT from B. brevis ATCC 9999 (42% identity, 62% similarity), TycF from B. brevis ATCC 8185 (34% identity, 53% similarity), BacT from Bacillus licheniformis (34% identity, 53% similarity), and other thioesterases associated with NRPS. Their role has recently been elucidated to be crucial for regeneration of misprimed NRPSs (32).

Two ABC Transporter Genes Are Located Directly Up- and Downstream of the lgr Genes—Directly upstream of lgrA, we found an enzyme encoding the ATP-binding subunit of an ABC transporter with high similarity to other permeases (35–43% identity and 55–65% similarity). Located 513 bp downstream of lgrD, we found another gene encoding an ABC transporter. As the middle of this gene is the 3′ region of the fosmid sequenced, we do not have full sequence data of this gene yet. We are currently sequencing a third fosmid obtained after a screen with the partial ABC transporter gene to gather more information about this transporter and the possible role of these two proteins in gramicidin self-resistance.

**Discussion**

We cloned and sequenced a DNA locus of B. brevis ATCC 8185 of about 74 kbp. 61 kbp of this region contain the entire gramicidin biosynthesis operon consisting of four large genes, lgrA, lgrB, lgrC, and lgrD. The corresponding proteins LgrA (258 kDa), LgrB (578 kDa), LgrC (866 kDa), and LgrD (567 kDa)
Gramicidin Biosynthesis Genes

**FIG. 4.** Shown are the seven conserved cores of all Lgr-E domains extracted from a multiple sequence alignment. The catalytically essential His in core 2 has been shaded in gray. The alignment of the cores shows that the E domain from LgrA2 is not as well conserved as the other 6 ones.

| Core E1 | Core E2 | Core E3 | Core E4 |
|---------|---------|---------|---------|
| P1QWFF-16 | P2QWFF-20 | P3QWFF-17 | P4QWFF-19 |
| HLWGDQSNV-185 | HLWGDQSNV-154 | HLWGDQSNV-155 | HLWGDQSNV-156 |
| TVQWFF-32 | TVQWFF-35 | TVQWFF-37 | TVQWFF-38 |
| DVLWALALE-262 | DVLWALALE-261 | DVLWALALE-261 | DVLWALALE-261 |
| EGRHRE-284 | EGRHRE-285 | EGRHRE-286 | EGRHRE-287 |

LgrA2: LgrA2_E2
LgrB2: LgrB2_E2
LgrC2: LgrC2_E2
LgrD2: LgrD2_E2

**FIG. 5.** Alignment of the two putative F domains from *B. brevis* ATCC 8185 and *Anabaena* strain 90 with six met-tRNA-formyltransferases from various bacteria. Shaded in gray is the most conserved region located in the N-terminal region containing the formyl-THF-binding site. Identical residues are marked with a dot, and highly conserved residues are marked with a star.

kDa) encode two, four, six, and four modules of nonribosomal peptide synthetases, respectively. We found a putative formylation domain upstream of the first module and directly joined to it, seven epimerization domains, and a putative reductase domain attached downstream to module 16. This modular arrangement is defined as linear NRPS or type A (22).

In 1977, Akashi et al. (33) purified an enzyme that activated Val and Gly, forming formylglycine; this enzyme was believed to be the initiation complex for gramicidin biosynthesis. This finding matches with LgrA, as this first dimodule was verified in this study by sequence analysis as well as biochemical investigations to activate Val and Gly. In the same year, Akers et al. (34) found two enzymes responsible for the synthesis of the initial portion of gramicidin. The first enzyme is again identical to LgrA. The second one should be LgrB, except that the authors postulated it to activate 5 amino acids and therefore to harbor five modules. They based this mainly on the size of the enzymes determined by sucrose gradient centrifugation and comparison of amino acid activation assays. The authors clearly underestimated the sizes of the tyrocidin synthetases (used as a reference in their study) and therefore for habor five modules. They based this mainly on the size of the two found enzymes, leading to wrong conclusions about the correct size of LgrA as well as wrong size and modular structure of LgrB. However, the biochemical data published from both groups correspond roughly to our own findings and support our assumption that we found the gramicidin biosynthesis genes. Module 1 activates Val (100%), Ile (48%), and Leu (12%). Val is mainly found in the product at position 1 (95%), whereas Leu has not been described at this position before (2), which is also consistent with the kinetic data. We believe that the C domain of LgrA plays an important role in substrate selectivity (35), too, and presumably, it has a donor site specifically for formylvaline. We propose that the formylation takes place when Val has been activated and bound as a thioester to the PCP before chain elongation occurs. We found module 2 to activate Gly only to some minor extent, whereas Ala appears to be the only amino acid activated to this extent by the A domain because of its very small size. Discrimination between Ala and Gly should be rather difficult; thus, Gly activation is not surprising. Our findings demonstrate the need to biochemically confirm all predictions made and to enlarge the data bases.

For each t-amino acid found in the product, we were able to localize an E domain in the corresponding module. Surprisingly, we found a seventh E domain in module 2. For the reasons given above (see “Results”), we do not believe this E domain to be functional in a catalytic manner. In accordance with a recent publication (36), we suggest a role in protein-protein interaction between LgrA and LgrB. This is supported further by the fact that the first three lgr synthetases carry an E domain at their C termini that is thought to interact with the N-terminal C domain of the following enzyme.

Because *B. brevis* ATCC 8185 is not genetically accessible, we are unable to give final proof by constructing a knockout mutant. However, from all the results obtained so far about the lgr-NRPSs, there should be no doubt about the identity of the four peptide synthetases.

The reductase domain localized at the C terminus of LgrD clearly showed activity with the shortened peptidylpeptide substrate in our *in vitro* assay, reducing this peptidylpeptide in an NADPH-dependent manner to the corresponding aldehyde. This reaction was absolutely dependent upon the presence of the reductase in the assay. We were also able to show that the enzyme accepts the peptidyl-CoA substrate to some minor extent in contrast to its preferred substrate, the peptide bound to the PCP. However, in mature gramicidine, the alcohol (ethanolamine) is exclusively found. Therefore, future biochemical
studies have to prove how or under which conditions the ethanalamine is formed. One possibility is that the shortened pentapeptide we utilized as substrate for the R domain leads to an altered reductase reactivity. On the other hand, a putative aldo/keto-reductase family protein encoded by a 1692-bp ORF located about 2.2 kbp upstream of lgrA was identified. In conclusion, we postulate that this putative oxidoreductase may be responsible for the second reduction of the gramicidin-aldehyde, which is formed first to the final product, gramicidin.

Our motivation to find, clone, and sequence the lgr synthetases was the expectation to find a formylation and reductase domain according to the well defined product, gramicidin. The putative F domain at the N terminus of LgrA shows significant similarity to other formyltransferases and has been produced and purified as soluble protein for further biochemical studies. From the alignment (Fig. 5), we propose the core motif IN(V)HXSSLPXXRG for formyl-THF-dependent formyltransferases and F domains. We want to investigate this interesting NRPS prototype enzyme to shed light on its mode of action. The F domain could be helpful for in vivo hybrid peptide synthetase approaches. After previous in vitro approaches (37, 38) producing unblocked peptides, the F domain could be the key to stabilize the formed peptides and prevent digestion with peptidases within the cell by formylation of the first amino acid. In addition, the reductase domain at the C terminus of LgrD, which was shown to be active in this study, is a promising enzyme, too, with regard to in vitro analysis using artificial substrates as well as its use in hybrid peptide synthetases to block the C terminus by reduction of the terminal amino acid for in vivo studies.

Acknowledgments—We thank Antje Schäfer for excellent technical assistance and Martina Pfeiff, Dr. Dirk Schwarzer, and Dr. Torsten Stachelhaus for helpful discussion. We also thank Robert Finkling and Stephan Sieber for commenting on the manuscript. Stephan Sieber kindly synthesized the peptidyl-CoA substrate. Dr. Michael Kube from the Max Planck Institute Berlin gave good tips for construction of the fosmid library.

REFERENCES
1. Hotchkiss, R. D., and Dubois, R. J. (1940) J. Biol. Chem. 132, 791–792
2. Sarges, R., and Wikop, B. (1965) J. Am. Chem. Soc. 87, 2011–2020
3. Weinstein, S., Wallace, B. A., Morrow, J. S., and Yeatich, W. R. (1986) J. Mol. Biol. 143, 1–19
4. Wallace, B. A. (1998) J. Struct. Biol. 121, 123–141
5. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Chem. Rev. 97, 2651–2674
6. Dieckmann, R., Lee, Y. O., van Liempt, H., von Dohren, H., and Kleinkauf, H. (1998) FEBS Lett. 427, 212–216
7. Stachelhaus, T., and Marahiel, M. A. (1995) J. Biol. Chem. 270, 6163–6169
8. Weber, T., and Marahiel, M. A. (2001) Structure 9, I3–I9
9. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) J. Biol. Chem. 273, 22773–22781
10. Stachelhaus, T., and Marahiel, M. A. (1995) FEMS Microbiol. Lett. 125, 3–14
11. Gaitatzis, N., Kunze, B., and Müller, R. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 11136–11141
12. Stein, T., Kluge, B., Vater, J., Franke, P., Otto, A., and Wittmann-Liebold, B. (1995) Biochemistry 34, 4633–4642
13. Linne, U., Doehel, S., and Marahiel, M. A. (2001) Biochemistry 40, 15824–15834
14. Haase, A., Schwab, M., Herrmann, M., and Zöcher, R. (1993) Mol. Microbiol. 7, 905–914
15. Koen, D., Kleno, A., Schorgendorfer, K., and Marahiel, M. A. (1997) Chem. Biol. 4, 927–937
16. Roushani, L., Paulin, L., Suomalainen, S., Hyytiainen, H., Buikema, W., Haselkor, B., and Sivenen, K. (2000) Mol. Microbiol. 37, 156–167
17. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3839–3842
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
19. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
20. Mootz, H. D., and Marahiel, M. A. (1997) J. Bacteriol. 179, 6843–6850
21. Laemmli, U. K. (1970) Nature 227, 680–685
22. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002) ChemBioChem 3, 490–504
23. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) Chem. Biol. 6, 493–505
24. Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Chem. Biol. 7, 211–224
25. 25 Pospiech, A., Bietenhader, J., and Schupp, T. (1996) FEBS Lett. 384, 505–510
26. Stachelhaus, T., and Walsh, C. T. (2000) Biochemistry 39, 5775–5787
27. Schmitz, E., Blanquet, S., and Mochulsky, Y. (1996) EMBO J. 15, 4749–4758
28. Schöfers, S., and Shimura, K. (1959) Nature 184, Suppl. 22, 1719–1719
29. Siłakowski, B., Kunze, B., Nordseik, G., Blocher, H., Hofle, G., and Müller, R. (2000) Eur. J. Biochem. 267, 6476–6485
30. Sieber, S. A., Walsh, C. T., and Marahiel, M. A. (2003) J. Am. Chem. Soc. 125, 10962–10966
31. Clugston, S. L., Sieber, S. A., Marahiel, M. A., and Walsh, C. T. (2003) Biochemistry 42, 12095–12104
32. Schwarzer, D., Mootz, H. D., Linne, U., and Marahiel, M. A. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 14083–14088
33. Akashi, K., Kubota, K., and Kurohashi, K. (1977) J. Biochem. (Tokyo) 81, 269–272
34. Aker, H. A., Lee, S. G., and Lipman, F. (1977) Biochemistry 16, 5722–5729
35. Kesting, T. A., Marshall, C. G., Walsh, C. T., and Kesting, A. E. (2002) Nat. Struct. Biol. 9, 526–526
36. Linne, U., Stein, D. B., Mootz, H. D., and Marahiel, M. A. (2003) Biochemistry 42, 5114–5124
37. Deckel, S., and Marahiel, M. A. (2000) Chem. Biol. 7, 373–384
38. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 5848–5853
39. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) EMBO J. 16, 4174–4183