Volume 307, number 1, 40–43 FEBS 11171 July 1992

Minireview

Multidomain enzymes involved in peptide synthesis

Mohamed A. Marahiel

Biochemie/FB Chemie, Hans-Meerwein-Straße, Philipps Universität Marburg, W-3550 Marburg, Germany

Received 3 April 1992

Biosynthesis of peptides in non-ribosomal systems is catalyzed by multifunctional enzymes that employ the thio-template mechanism. Recent studies on the analysis of the primary structure of several peptide synthetases have revealed that they are organized in highly conserved and repeated functional domains. The aligned domains provide the template for peptide synthesis, and their order determines the sequence of the peptide product.

Multidomain enzyme; Peptide synthetase; Thio-template; Core sequence

1. INTRODUCTION

Bioactive peptides with linear and cyclic structures are produced non-ribosomally by several bacterial and fungal species with the aid of multi-enzyme complexes [1]. Over 20 years ago Fritz Lipmann [2,3] proposed that the multi-enzyme peptide synthetases employ the thio-template mechanism in the synthesis of these peptides. The constituents of the peptides, amino- and hydroxy acids, that have to be sequentially connected by the corresponding peptide synthetase are first activated as acyl adenylates where ATP serves as the energy source. This activation is similar to that catalyzed by the aminoacyl-tRNA synthetases [4], however, no tRNA intermediates are formed, but instead peptide synthetases covalently link the activated amino acid as a carboxy thioester [5–7].

For peptide synthetases activating more than one amino acid residue individual sites catalyzing the activation reactions were proposed to be attached on the multi-enzyme complex in the same order in which their activated residues are incorporated into the growing peptide chain. These distinct sites are referred to as 'domains'. Thus, the domains of the multi-enzyme provide the template for the pre-defined peptide sequence.

The elongation reaction is catalyzed by the enzyme bound cofactor 4'-phosphopantetheine [8,9]. The activated amino acid bound as carboxy thioester to the synthetase domain is then transferred to the thiol group of the cofactor which acts as an internal transport system. In repeated trans-peptidation and trans-thiolation reactions the peptide chain is completed and then released from the multi-enzyme either by cyclization or by the action of a specific thioesterase. This polymerization mechanism links non-ribosomal peptide synthesis to that of fatty acid and polyketide synthesis. The protein-thio-template pathway has been shown to direct the synthesis of gramicidin S, tyrocidine, surfactin, bacitracin, enniatin, actinomycin, cyclosporin, and also the first step in β-lactam biosynthesis, the condensation of the tripeptide δ-(L-α-aminoadipoyl)-L-cysteinyl-D-valine (ACV) (for review see [1,10–12]).

Correspondence address: Biochemie/FB Chemie, Hans-Meerwein-Straße, Philipps Universität Marburg, W-3550 Marburg, Germany. Fax: (49) (6421) 285 547.

2. GENES ENCODING MULTIFUNCTIONAL PEPTIDE SYNTHETASES

Genes encoding multifunctional peptide synthetases of prokaryotes and lower eukaryotes have been recently isolated and characterized. The primary structure and gene organization of the operons encoding peptide synthetases for the *Bacillus* cyclic peptide antibiotics gramicidin S (*grs*), tyrocidine (*tyc*) and surfactin (*srf*), have been analysed in more detail [13–17]. The entire nucleotide sequence of the *grs* operon, which comprises the three genes *grsT*, *grsA* and *grsB* on a 19 kb DNA fragment, has been determined [18]. The gramicidin S synthetase 1 (GrsA), which activates and racemises the first amino acid, Phe, is a protein composed of 1,098 amino acid residues with a calculated molecular weight of 126,663 Da. The synthetase 2 (GrsB) is a single polypeptide chain composed of 4,452 amino acids in length (510,287 Da) and activates the four amino acids, Pro, Val, Orn and Leu. The role of the *grsT* gene product in peptide synthesis is unknown. However, the *grsT*-encoded 29,000 Da protein shows striking homology to fatty acid thioesterases of mammalian origin [14]. In addition, the thioesterase active site (GHSXG) was also

Published by Elsevier Science Publishers B.V.

found within acyltransferases, which may suggest for GrsT a thioesterase- or an acyltransferase-like function in the non-ribosomal peptide synthesis. Support for this can also be inferred from studies on the *bap* gene cluster and ACV synthetase. The *bap* gene cluster encodes enzymes involved in the synthesis of the tripeptide antibiotic, bialaphos, in *Streptomyces hygroscopicus* [19]. Two open reading frames of the *bap* operon encode two putative thioesterases homologous to GrsT (30% identity). Furthermore, the ACV synthetase contains at its carboxy terminus an integrated domain homologous to the above mentioned thioesterases [20]. Also, recent studies on the purified GrsT protein show that it stimulates gramicidin S synthesis in vitro, however, GrsT at high concentration slightly inhibits synthesis (T. Stachelhaus and M.A. Marahiel, unpublished results). Therefore, it is tempting to speculate that GrsT may have a thioesterase function that is involved in correcting errant amino acylation in peptide synthesis.

Analysis of the primary structures of GrsA and GrsB derived from the nucleotide sequences revealed the presence of five distinct homologous domains, each about 600 amino acid residues in length. The domains show about 45–50% identity and are separated by non-homologous regions of about 500 amino acids [18]. Four domains are located within the GrsB sequence and in the same order as the constituent amino acids Pro, Val, Orn and Leu, which are incorporated into the growing peptide chain (Fig. 1A). This is in agreement with the biochemical data obtained from proteins encoded by the *grsB* domains and from the proteolytic analysis of cleaved GrsB fragments [21–24].

For tyrocidine synthesis three multi-enzymes are needed: TycA, TycB and TycC [25]. TycA has a high degree of homology to GrsA and also activates and racemises phenylalanine. The multi-enzymes, TycB and TycC, synthesize the tripeptide, Pro-Phe-D-Phe, and the hexapeptide, Asn-Gln-Tyr-Val-Orn-Leu, respectively. Based on biochemical studies and analysis of the primary structure it can be concluded that the mechanism of tyrocidine synthesis is the same as that employed by gramicidin S peptide synthetases. The *tyc* genes contain homologous domains and are organized in a large operon in the order *tycA-tycB-tycC* [15]. The operon is regulated at the transcriptional level from a promotor located at the 5′-end of *tycA*, which is under the control of the *spoOA-abrB* system [26]. AbrB, a 10.7 kDa protein, acts as a negative regulator by direct interaction with sequences upstream and downstream of the *tyc* promotor, whereas SpoOA, a 29 kDa protein, is an ambivalent transcriptional regulator, negatively affecting *abrB* transcription at the onset of stationary growth [27,28].

Genetic studies in *B. subtilis* on the production of the lipopeptide antibiotic, surfactin, identified three loci (*sfp, srfA, srfB*) essential for its production [29]. The *sfp* gene encodes a 244 amino acid protein showing a high degree of homology (35% identity) to the product of an open reading frame (*orfx*) associated with the *grs* operon [30]. The function of *sfp* or *orfx* products is not known. In contrast, *srfA*, which covers a DNA region of about 25 kb, contains 4 *orfs* which encode three putative multi-enzymes and a thioesterase-like protein. Grandi and co-workers [31] determined the entire nucleotide sequence of the *srf* operon and discovered seven homologous and repeated domains similar to those identified in the *grs* and *tyc* operons. These domains are located in three *orfs* which are responsible for the activation of three, three and one amino acid, respectively. The third locus, *srfB*, was found to be identical with an early competence gene of *B. subtilis, comA*. *comA* encodes a response regulator protein which, when activated by the gene product of *comP*, a histidine protein kinase, acts as a positive transcriptional regulator of *srfA*. The gene products of *comP* and *comA* are thought to be two component regulatory partners of the sensor–regulator type [32].

A multifunctional peptide synthetase similar to those identified in *Bacillus* species was also discovered in several β-lactam producers of fungal and bacterial origin. The first step in the biosynthesis of β-lactam antibiotics, such as cephamycin, cephalosporin and penicillin, is catalyzed by the formation of the tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV). The gene encoding ACV synthetase, *pcbAB*, has been isolated from the Gram-positive bacteriae species, *Flavobacterium, Streptomyces clavoligerus* and *Nocardia lactamdurans* [33,34], as well as from the fungi, *Penicillium chrysogenum* [20], *Cephalosporium acremonium* [35,36], and *Aspergillus nidulans* [37]. The ACV synthetase from β-lactam producers of prokaryotic and eukaryotic origin is encoded by a single gene (*pcbAB*) located within a gene cluster encoding other enzymes needed in β-lactam synthesis. The *pcbAB* gene encodes a large protein of more than 3,700 amino acid residues with a similar sequence and organization as the bacterial peptide synthetases involved in gramicidin S and tyrocidine synthesis. Three homologous and repeated domains are located within the ACV synthetase sequence and are thought to be involved in the condensation of aminoadipate, L-cysteine and D-valine to produce the tripeptide precursor for penicillin G.

3. A CONSERVED FUNCTIONAL DOMAIN

The domains of peptide synthetases seem to represent the functional units and are believed to be the sites of amino acid activation and thioester formation. Extensive homology (35–50% identity) is observed between domains of different peptide synthetases, irrespective of their origin, over a region of about 600 amino acids [18]. In addition, in all domains several highly conserved and potentially important motifs (Core sequences, Fig. 1) appear in a defined order. Among these core regions

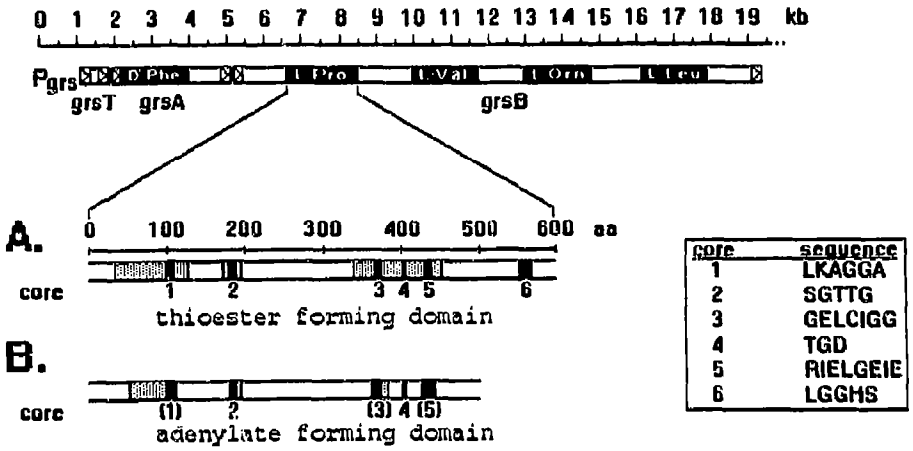

Fig. 1. The diagram depicts the organization of the homologous amino acid-activating domains within the *grs* operon. Also shown are (A) the relative locations of the conserved core sequences (numbered dark boxes) within the thioester forming domain of GrsB-Pro, which are also found in all other peptide synthetase domains, and (B) in comparison, the location of the corresponding core sequences within an adenylate-forming domain. Shaded regions in both domains represent significant degrees of sequence similarities, and core sequences given in brackets (within the adenylate-forming domain) have a low degree of conservation to those in the corresponding thioester-forming domain (for more details see [18]).

Vater and co-workers defined the LGGHS sequence (Core 6, Fig. 1) as the site of thioester formation in GrsB-Val and GrsB-Leu domains [38]. This sequence, which is found at the carboxy-terminal end of each domain, has a serine instead of cysteine residue. However, it resembles the pantetheine-binding site of acyl carrier proteins in fatty acid- and polyketide-synthases [38]. This would suggest the presence of multiple pantetheine binding sites in multidomain peptide synthetases, and link the thioester formation to a thiol group of a pantetheine rather than to a cysteine residue.

Core 2, SGTTG (Fig. 1), is another potentially important motif. It was found to be highly conserved, not only among all domains of peptide synthetases, but also in a family of enzymes involved in carboxyl-group activation utilizing ATP. Among those enzymes are 4-coumarate-CoA ligase, firefly luciferase, long-chain acyl CoA synthetase, acetyl CoA synthetase and the *E. coli* enterobactin synthetases, EntE and EntF [18]. In all these enzymes and peptide synthetases the activation of the carboxyl group occurs with concomitant ATP hydrolysis. With the exception of EntF all adenylate-forming enzymes contain a shorter domain of about 500 amino acids (Fig. 1B), which is homologous to that of peptide synthetases. As shown in Fig. 1B it contains the adenylation core sequence, SGTTG, and the putative ATPase motif, TGD. However, the core sequences, 1, 3 and 5, present in all domains of peptide synthetases, are not strictly conserved in the adenylate-forming domain. More interestingly the site of thioester formation, LGGHS (Core 6, Fig. 1A), is completely missing in all adenylate-forming domains, which is in agreement with their inability to form thioesters [18]. In contrast EntF, which activates L-serine, also forms acyl and carboxy thioester intermediates similar to peptide synthetases. It also contains the cofactor 4′-phosphopantetheine, a thioesterase active-site motif, and a putative thioester forming core sequence (LGGHS) at its carboxy terminal end [39].

In conclusion, the sequence similarities between peptide synthetases and adenylate-forming enzymes suggest the existence of a gene family that may have a common ancestry.

4. OUTLOOK

The recent studies on the biochemistry and molecular biology of antibiotic production in microorganisms have not only led to a better understanding of the structure–function relationship of peptide synthetases, but also to a general mechanism for non-ribosomal peptide syntheses. The findings support a model of integrated enzyme activities, in which compact structural domains, joined in a defined order on the multifunctional peptide synthetase, determine the sequence of the peptide product. Of particular interest in the future are the following studies:

(i) the exact structural and functional definition of peptide synthetase domains;

(ii) the elucidation of the function of the putative core sequences believed to be involved in adenylation and thioester formation using site-directed mutagenesis;

(iii) investigating the real role played by thioesterases;

(iv) obtaining evidence for the hypothesis suggesting the existence and involvement of multiple pantetheine binding sites in peptide synthesis.

Additional exciting studies would involve joining

heterologous amino acid-activating domains by in vitro gene fusion or in vivo recombination to synthesize novel peptides of desired activity.

Acknowledgements: The author wishes to thank T. Stachelhaus for Fig. 1 and helpful discussion, and S.S. Patil, S. Borchert and K. Turgay for critical reading of the manuscript. Work in the author's laboratory was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

[1] Kleinkauf, H. and Von Döhren, H. (1990) Eur. J. Biochem. 192, 1–15.
[2] Lipmann, F., Gevers, W., Kleinkauf, H. and Roskoski, R. (1971) Adv. Enzymol. 35, 1–33.
[3] Lipmann, F. (1980) Adv. Microbial Phys. 21, 227–260.
[4] Schimmel, P.R. (1987) Annu. Rev. Biochem. 56, 125–158.
[5] Laland, S.G. and Zimmer, T.-L. (1973) Essays Biochem. 9, 31–57.
[6] Kurahashi, K. (1974) Annu. Rev. Biochem. 43, 445–459.
[7] Katz, E. and Demain, A.L. (1977) Bacteriol. Rev. 41, 449–474.
[8] Kleinkauf, H., Roskoski Jr., R. and Lipmann, F. (1971) Proc. Natl. Acad. Sci. USA 68, 2069–2072.
[9] Lipmann, F. (1973) Acc. Chem. Res. 6, 361–367.
[10] Kleinkauf, H. and Von Döhren, H. (1987) Annu. Rev. Microbiol. 41, 259–289.
[11] Kleinkauf, H. and Von Döhren, H. (1988) Crit. Rev. Biotechnol. 8, 1–32.
[12] Nakano, M.M. and Zuber, P. (1990) Crit. Rev. Biotechnol. 10, 223–240.
[13] Weckermann, R., Fürbaß, R. and Marahiel, M.A. (1988) Nucleic Acids Res. 16, 11841.
[14] Krätzschmar, J., Krause, M. and Marahiel, M.A. (1989) J. Bacteriol. 171, 5422–5429.
[15] Mittenhuber, G., Weckermann, R. and Marahiel, M.A. (1989) J. Bacteriol. 171, 4881–4887.
[16] Nakano, M.M., Marahiel, M.A. and Zuber, P. (1988) J. Bacteriol. 171, 5347–5353.
[17] Nakano, M.M., Magnuson, R., Meyers, A., Curry, J., Grossman, A.D. and Zuber, P. (1991) J. Bacteriol. 173, 1770–1778.
[18] Turgay, K., Krause, M. and Marahiel, M.A. (1992) Mol. Microbiol. 6, 529–546.
[19] Raibaud, A., Zalasan, M., Holt, G. and Thompson, C.J. (1991) J. Bacteriol. 173, 4454–4463.
[20] Gutierrez, S., Diez, B., Montenegro, E. and Martin, J.F. (1991) J. Bacteriol. 173, 2354–2365.
[21] Krause, M., Marahiel, M.A., Von Döhren, H. and Kleinkauf, H. (1985) J. Bacteriol. 162, 1120–1125.
[22] Hori, K., Yamamoto, Y., Tokita, K., Saito, F., Kurotsu, T., Kanda, M., Okamura, K., Furuyama, J. and Saito, Y. (1991) J. Biochem. 110, 111–119.
[23] Skarpeid, H.-J., Zimmer, T.-L., Shen, B. and Von Döhren, H. (1990) Eur. J. Biochem. 187, 627–633.
[24] Skarpeid, H.-J., Zimmer, T.-L. and Von Döhren, H. (1990) Eur. J. Biochem. 189, 517–522.
[25] Roskoski, R., Gevers, W., Kleinkauf, H. and Lipmann, F. (1970) Biochemistry 9, 4839–4845.
[26] Marahiel, M.A., Zuber, P., Czekay, G. and Losick, R. (1987) J. Bacteriol. 169, 2215–2222.
[27] Robertson, J.R., Gocht, M., Marahiel, M.A. and Zuber, P. (1989) Proc. Natl. Acad. Sci. USA 86, 8457–8461.
[28] Fürbaß, R., Gocht, M., Zuber, P. and Marahiel, M.A. (1991) Mol. Gen. Genet. 225, 347–354.
[29] Nakano, M.M. and Zuber, P. (1990) in: Genetics and Biotechnology of Bacilli (Hoch, J. and Ganesan, A.T. eds.) pp. 397–405, Academic Press, NY.
[30] Nakano, M.M., Corbell, N., Besson, I. and Zuber, P. (1992) Mol. Gen. Genet. 232, 313–321.
[31] Grandi, G., Cosmina, P., Rodriguez, F., Henke, M., Drees, M. and de Ferra, F., Perego, M., van Sinderen, D. and Venema, G. (1992) submitted.
[32] Nakano, M.M., Xia, L. and Zuber, P. (1991) J. Bacteriol. 173, 5487–5493.
[33] Tobin, M.B., Kovacevic, S., Madduri, K., Hoskins, J.A., Skatrud, P.L., Vining, L.C., Stutterd, C. and Miller, R.J. (1991) J. Bacteriol. 173, 6223–6229.
[34] Coque, J.-J.R., Liras, P., Laiz, L. and Martin, J.F. (1991) J. Bacteriol. 173, 6258–6264.
[35] Diez, B.S., Gutierrez, J.L., Barredo, P., Van Solingen, P., Van der Vorrt, L.H.M. and Martin, J.F. (1990) J. Biol. Chem. 265, 16358–16365.
[36] Smith, D.J., Earl, A.F. and Turner, G. (1990) EMBO J. 9, 2743–2750.
[37] McCabe, A.P., Van Liempt, H., Palissa, H., Unkles, S.E., Riach, M.B.R., Pfeifer, E., Von Döhren, H. and Kinghorn, H. (1991) J. Biol. Chem. 266, 12646–12654.
[38] Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M.A., Kruft, V. and Wittmann-Liebold (1991) J. Biol. Chem. 266, 23135–23141.
[39] Rusnak, F., Sakaitani, M., Drueckhammer, D., Reichert, J. and Walsh, C.T. (1991) Biochemistry 30, 2916–2927.