Amino Acids (1996) 10: 201–227

Amino
Acids

Amino acid activation and polymerization at modular multienzymes in nonribosomal peptide biosynthesis

Review Article

T. Stein and **J. Vater**

Institut für Biochemie und Molekulare Biologie, Technische Universität Berlin, Berlin-Charlottenburg, Federal Republic of Germany

Accepted November 20, 1995

Summary. The biosynthesis of microbial bioactive peptides is accomplished nonribosomally by large multifunctional enzymes consisting of linearly arranged building blocks of 1,000–1,500 amino acid residues. Each of these units acts as an independent enzyme which catalyzes the selection, activation, and in some cases modification (epimerization, N-methylation) of its cognate amino acid, as well as the elongation of the peptide product. The specific linkage of amino acid activating modules upon such polyenzymes defines the sequence of the peptide product. A series of functional domains could be identified upon an amino acid activating module which are involved in the sequential reactions in nonribosomal peptide biosynthesis.

Keywords: Amino acids – Nonribosomal peptide biosynthesis – Peptide antibiotics – Amino acid activation – Aminoacyl–adenylation – Multienzymes – Modular structure – Multiple 4′-phosphopantetheine cofactors

Abbrevations: aaRS, aminoacyl tRNA synthetase; GS1, gramicidin S synthetase 1 (phenylalanine racemase); GS2, gramicidin S synthetase 2; TY1 and 2, tyrocidine synthetase 1 and 2; ACV, [δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine]; FITC, fluorescein 5′-isothiocyanate; FAB-, ESI-MS; fast atom bombardment-, electrospray ionization-mass spectrometry; Pan, 4′-phosphopantetheine; NMR, nuclear magnetic resonance; ACP, acyl carrier protein; SAM, S-adenosyl-L-methionine; CM, carboxy-methyl; NES, N-ethylsuccinimido.

1. Introduction

Microbial organisms produce a variety of low molecular weight bioactive peptides, depsipeptides, peptidolactones, and lipopeptides with linear, branched, or cyclic structures (Kleinkauf and von Döhren, 1990a). Many of

these secondary metabolites exhibit valuable properties qualifying them for biotechnological and medical uses as antibiotic-, antifungal-, antiviral-, antitumor-, insecticide-, and nemadotical agents, as well as biosurfactants, for example. In general the biosynthesis of these compounds is accomplished nonribosomally by large multifunctional and multisubunit enzymes with molecular weights in the range from 100 to 1,700 kDa. Each of these peptide synthetases function as a cellular factory representing the protein template for the construction of one defined peptide structure. Such multifunctional enzymes catalyze all essential processes necessary for peptide biosynthesis including selection, activation, and modification of the cognate amino acids or related carboxy-containing substrates (N-methylation or epimerization), specific step-by-step linkage of the activated substrates, as well as in many cases cyclization of the product by amidation, lactonization, and piperazinedione formation, e.g. The chain length potential of multienzymatic peptide forming systems ranges from 2 (bacilysin) up to 19 (alamethicin) amino residues. Whereas the number of constituent amino acids in ribosomally formed peptides or proteins is restricted to the 20 naturally occuring "proteinogenic" amino acids by the number of transfer RNA and aminoacyl-tRNA synthetase (aaRS) molecules, at least 300 different amino acids and related carboxy-containing compounds are known to be incorporated into enzymatically synthesized peptides including non-proteinogenic-, N-methylated-, and D-amino acids, as well as compounds containing no amino group like hydroxy- and fatty acids, as well as chromophores (Kleinkauf and von Döhren, 1990a).

More than 20 years ago the multienzymatic thiotemplate model for nonribosomal peptide biosynthesis was proposed by several groups (Kleinkauf et al., 1971; Lipmann, 1973; Laland and Zimmer, 1973; Kurahashi, 1974). According to this mechanism amino acid activation occurs in a two step process including 1) aminoacyl adenylation similar to the amino acid activation process catalyzed by tRNA synthetases, and 2) aminoacyl thioesterification at specific reactive thiol groups of the multienzyme (thiotemplates), a process characteristical for nonribosomal peptide biosynthesis. In addition, it was assumed that a central thiol-group of an intrinsic 4′-phosphopantetheine carrier molecule interacts with the thioesterified amino acid substrates managing a step-by-step elongation of the peptide product in a series of transpeptidation and transthiolation reactions.

A second group of nonribosomal peptide synthetases activates their amino acid substrates and peptide intermediates non-covalently as aminoacyl-phosphates, driven by the hydrolysis of an β-γ-linkage and release of ADP. This class includes enzymes as glutamine- (Liaw and Eisenberg, 1994), γ-glutamyl-L-cysteine-(Huang et al., 1993), and glutathione synthetase (Meister, 1988), D-Ala:D-Ala ligase (Fan et al., 1995) and the MurC, -D, and -E bacterial peptidoglycan synthetic enzymes (Walsh et al., 1989).

2. Multifunctional peptide synthetases

Table 1 summarizes representative examples of well characterized peptide forming systems following the thiotemplate pathway. The biosynthesis of the

cyclo-decapeptide gramicidin S in *Bacillus brevis* occurs by the interaction of two multifunctional proteins. Gramicidin S synthetase 1 (GS1, phenylalanine recemase, EC 5.1. 1.11) activates and racemizes phenylalanine and transfers D-Phe to the condensing enzyme gramicidin S synthetase 2 (GS2) which activates L-Pro, L-Val, L-Orn (L-ornithine), and L-Leu and catalyses the synthesis of a penta-peptide. Cyclization is performed by two head to tail condensations between two pentapeptides in an as yet unknown mechanism (for reviews see Vater, 1990; Kleinkauf and van Döhren, 1990). The biosynthesis of tyrocidine is accomplished by three multienzymes, tyrocidine synthetase 1, 2, and 3 (TY1, TY2, and TY3). TY1, activates and racemizes phenylalanine, TY2 elongates the peptide chain until the tetrapeptide, the final decapeptide and cyclization is catalyzed by TY3 (Lee and Lipmann, 1974). In the biosynthesis of the cyclo-lipoheptapeptide surfactin at least four enzymes cooperate. An acyl transferase, surfactin synthetase 1 (SRFS1), initiates surfactin biosynthesis by transferring a β-hydroxy fatty acid moiety to the start amino acid L-Glu which is bound at the first module of surfactin synthetase 2 (SRFS2). This enzyme elongates the lipopeptide chain to the lipotripeptide. Surfactin synthetase 3 (SRFS3) accepts the elongated intermediate and forms the lipohexapeptide. The terminal L-Leu residue is added by surfactin synthetase 4 (SRFS4) which catalyses the cyclization of the product by lactone formation (Ullrich et al., 1991; Menkhaus et al., 1993; Galli et al., 1994).

As is apparent from Table 1 the biosynthesis of the fungal peptide ACV [δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine], the biosynthetic key-precursor of the antibiotics penicillin and cephalosporine in certain filamentous fungi (Aharonowitz et al., 1993), enniatin B (Pieper et al., 1995), HC-toxin (Walton, 1987), and cyclosporin A (Dittmann et al., 1994) is accomplished by single polypeptide chains. Cyclosporin synthetase, encoded by the *sim*A gene from *Tolyplocadium niveum* represents the largest multienzyme protein chain so far described (Weber et al., 1994) catalyzing at least 40 different reactions to accomplish cyclosporin synthesis (Lawen and Zocher, 1990). This mechanism represents an impressive example of product channeling on a single multidomain enzyme. To avoid dilution in the surrounding cytosol, all intermediates of the cyclosporinA biosynthetic pathway are kept covalently bound at a giant multienzyme until the completed product can be released. Examples for such a channeling has been reviewed for a series of biochemical pathways by Srere et al. (1987) and more recently Hawkins and Lamb (1995).

3. Modular structure of peptide synthetases

An important milestone concerning the comprehension of the structure-function relationships of peptide synthetases has been achieved by sequencing the first multienzyme encoding genes *tyc*A (Weckermann et al., 1988) and *grs*A (Krätzschmar et al., 1989; Hori et al., 1989) encoding TY1 and GS1 in *B. brevis*. Both enzymes are structurally closely related (80% homology in their amino acid sequence) and activate phenylalanine. During the next 3 years the

Table 1. Representative nonribosomal peptide antibiotics and the peptide forming multienzymes produced by Gram-positive bacteria (**A**) and filamentous fungi (**B**)

Peptide [a]	Organism	Genes	Enzymes	No. of activated amino acids	MW/kDa [b]	Properties	References
A							
Gramicidin S	*Bacillus brevis* ATCC 9999 [Nagano]	*grs*A *grs*B	GS1 GS2	1 4	126.7 [123.5] 510.3 [508.7]	Antibiotic	Krätzschmar et al. (1989); [Hori et al. (1989)] Turgay et al. (1992); [Saito et al. (1994)]
Tyrocidine	*Bacillus brevis* ATCC 8185	*tyc*A *tyc*B *tyc*C	TY1 TY2 TY3	1 3 6	122.6 230 460	Antibiotic	Weckermann et al. (1988) Lee & Lipmann (1974)
Surfactin	*Bacillus subtilis* ATCC 21332 (OKB 105)	*srf*AA1 *srf*AB1 *srf*AC	SRFS2 SRFS3 SRFS4	3 3 1	402 401 144	Surfactant, antibiotic	Ullrich et al. (1991); Fuma et al. (1993); Cosmina et al. (1993); Menkhaus et al. (1993) Galli et al. (1994)
B							
ACV	*Penicillium chrysogenum* *Acremonium chrysogenum* *Nocardia lactamdurans* *Aspergillus nidulans*	*pcb*A *pcb*A *pcb*A *acv*A	ACVS ACVS ACVS ACVS	3 3 3 3	404.1 426.0 404.1 422.5	Biosynthetic precursor of the antibiotics penicillin and cephalosporin	Díez et al. (1990); Smith et al. (1990) Gutiérrez et al. (1991) Coque et al. (1991) MacCabe et al. (1991); Aharonowitz et al. (1993)
Cyclosporin A	*Tolypocladium niveum*	*sim*A	Cysyn	11	1689.2	Immunomodulator	Lawen & Zocher (1990); Weber et al. (1994)
Enniatin B	*Fusarium scirpi*	esyn1	Esyn	2	346.9	Antifungal	Haese et al. (1993); Pieper et al. (1995)
HC-Toxin	*Cochlibolus carbonum*	*hts*1	HTS1	4	574.6	Phytotoxin	Walton (1987); Scott-Craig et al. (1992)

[a] The structures of gramicidin S, surfactin, ACV [δ-(L-α-aminoadipinyl)-L-Cys-D-Val], and cyclosporin A are given Fig. 1. Tyrocidine, cyclo(D-Phe-L-Pro-L-Phe-D-Phe-L-Asn-L-Gln-L-Tyr-L-Val-L-Orn[1]-L-Leu); enniatin B, cyclo(D-HOiVl[2]-Me-L-Val)$_3$; HC-toxin, cyclo[D-Pro-L-Ala-D-Ala-(2-amino-9, 10-epoxy-8-oxodecanoic acid)]. [b] The molecular masses were deduced from the DNA sequence. The MW of TY2 and TY3 were determined by sucrose density gradient centrifugation. [1] Orn, ornithine; [2] HOiVl, 2-hydroxyisovaleric acid.

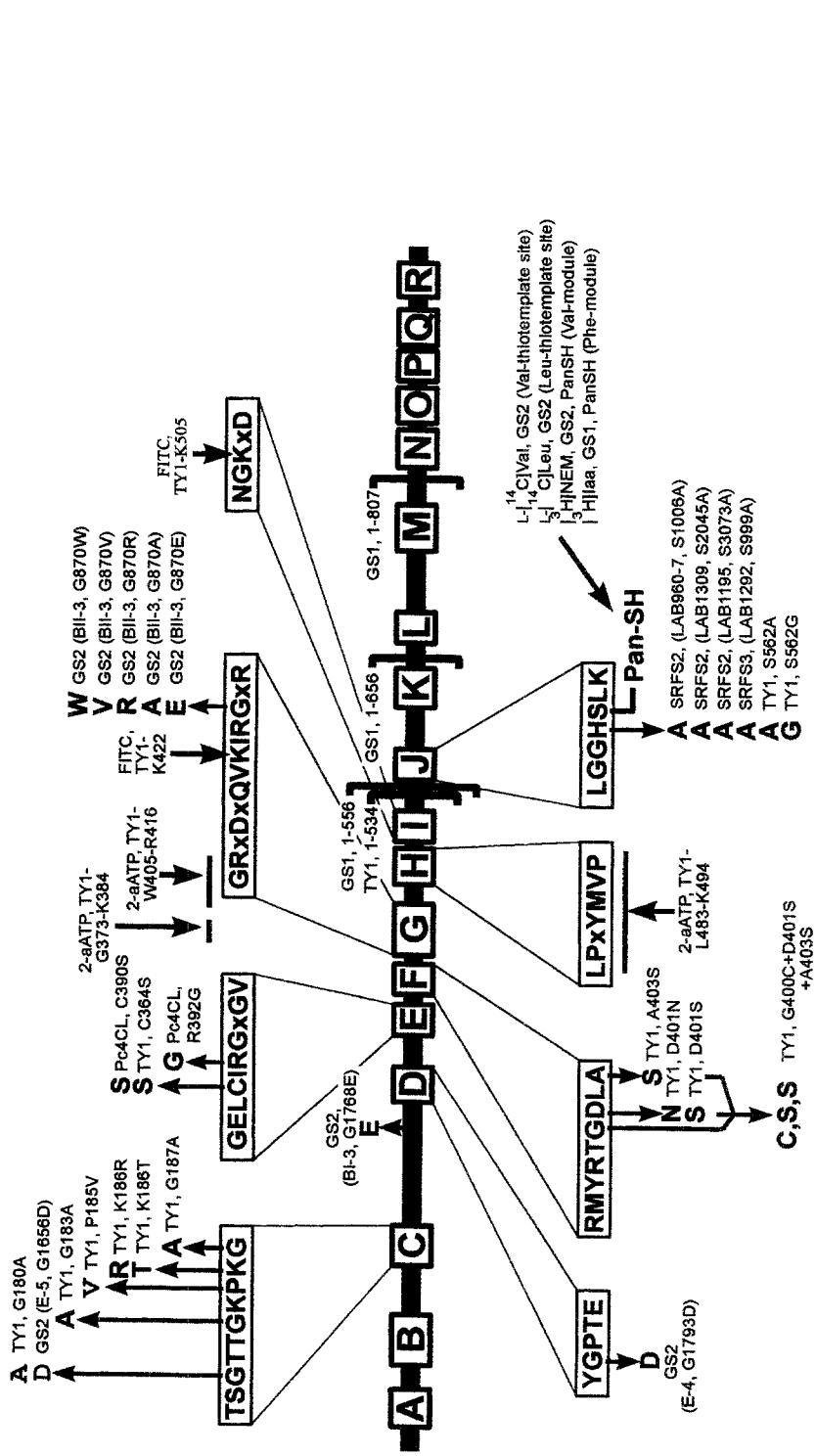


Fig. 1. Schematic presentation of an amino acid activating module of a peptide synthetase which summarizes the experiments leading to the comprehension of structure-function relationships of highly conserved sequence motifs. The protein chain of the amino acid activating module is symbolized by the central thick line containing boxes which indicate highly conserved sequence motifs (big letters). Their core-sequence is given in the one letter code in boxes above or underneath the protein chain[a]. Arrows pointing away from the sequence motifs symbolize site directed mutagenesis experiments indicating the peptide synthetase as well as the exact position of the exchanged amino acid. Affinity labeling experiments are shown by arrows pointing towards the sequence motifs. In addition, the exact position of specific dissection experiments of gramicidin S- and tyrocidine synthetase 1 are symbolized by square brackets within the protein chain. [a]The sequences of motifs *A*, *B*, as well as *K-R* are not shown in this figure. Motif *A*, LTYxELNxxAN, and *B*, ILAVLKAGGAYPVIDF are observed in all amino acid activating modules. The motifs *K*, YPVSSAQKRMY; *L*, LxxRHEALRTxF; *M*, HHIIxDGxSxxIL; *N*, TLYxVLxxxYxVL; *O*, DIIVGTPxAGRxxP; *P*, VGMFVNTLxLR; *Q*, VKxxLxAFxx-QDYPF; *R*, SRHPLLFxxxF are found within non-epimerizing modules. In epimerizing modules the motifs *K-R* are specifically modified as follows: *K*, GExLTPIQxWFF; *L*, LxxHHDALRMxY; *M*, HHLVVDGVSWxIL; *N*, ExNDILLTAxGLAL; *O*, LEGHGREII; *P*, SRTVGWFTSMTP; *Q*, VPxKGVGYGILxY; *R*, PxxxFNYLGQF

nucleotide sequences of the *pcb*AB and *acv*A genes from *Penicillium chrysogenum* (Smith et al., 1990; Díez et al., 1990), *Acremonium chrysogenum* (Gutiérrez et al., 1991), *Nocardia lactamdurans* (Coque et al., 1991), and *Aspergillus nidulans* (MacCabe et al., 1991) were determined. These genes encode the ACV synthetases, multifunctional enzymes that activate three amino acid compounds. By sequence analysis a multimodular structure has been proposed for all these peptide forming systems (Mc Cabe et al., 1991; Marahiel, 1992; Zuber et al., 1993; Stachelhaus and Marahiel, 1995a). This hypothesis has been supported by comparison of their primary structures with the constantly increasing number of known sequences of peptide synthetases, as gramicidin S synthetase 2 (Turgay et al., 1992; Saito et al., 1994), HC-toxin synthetase (Scott-Craig et al., 1992), surfactin synthetase (Fuma et al., 1993; Cosmina et al., 1993), enniatin synthetase (Haese et al., 1993), as well as cyclosporin synthetase (Weber et al., 1994).

Thorough sequence alignments of multifunctional peptide synthetases revealed that these enzymes are composed of homologous building blocks which comprise 1,000–1,500 amino acid residues corresponding to a molecular mass in the range of 120–180 kDa. Each of these modules is able to act as an independent enzyme catalyzing the selection and activation of its cognate substrates as demonstrated by limited proteolysis of gramicidin S synthetase. 100–120 kDa fragments of this enzyme have been observed to catalyze adenylation of Pro, Val, Orn and Leu (Skarpeid et al., 1990a/b; Kurotsu et al., 1991) as well as proline thioesterification (Kurotsu et al., 1991). These findings are in agreement with the polyenzyme model proposed by Fritz Lipmann (1954) in the pre-ribosomal area of protein/peptide biosynthesis research. The specific linkage of amino acid activation modules along the peptide synthetase protein chains defines the sequence of the amino acids in the peptide product. Each of the modules is distinguished by a linear array of strictly conserved sequence motifs which are thought to be involved in adenylate- and peptide formation. As is reviewed in the following chapters, some of the consensus motifs could be attributed to specific functions by site directed mutagenesis (Vollenbroich et al., 1993, 1994; Gocht and Marahiel, 1994), analysis of adenylation lacking mutants (Tokita et al., 1993; Saito et al., 1995), fragments of recombinant amino acid activating modules after specific dissection of the synthetase encoding genes (Stachelhaus and Marahiel, 1995b; Dieckmann et al., 1995), and affinity labeling (Schlumbohm et al., 1991; Stein et al., 1994; Pavela-Vrancic et al., 1994 a/b; Stein et al., 1995). A summary of these experiments are given in Fig. 1.

4. Characterization of functional domains within an amino acid activating module

4.1 Aminoacyl-adenylation domain

Amino acid sequence alignments of peptide forming multienzymes revealed that the N-terminal part of each amino acid activation module belongs to a super-family of adenylate-forming enzymes, the carboxyl:adenylate ligases

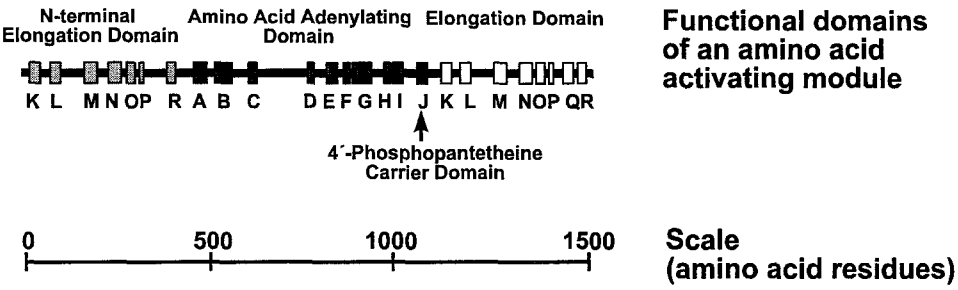
(AMP-binding proteins; MacCabe et al., 1991; Turgay et al., 1992; Fulda et al., 1994). Prominent members of this enzyme class represent all peptide synthetases so far sequenced, enterobactin synthetase E (Staab et al., 1989) and F (Rusnak et al., 1991), bialaphos (alanyl-alanyl-phosphinotricine tripeptide) synthetase (Wohlleben et al., 1992), long chain acyl-CoA (EC 6.2.1.3) and acetyl-CoA (EC 6.2.1.1) synthetases, insect luciferases (EC 1.13.12.7), plant 4-coumaroyl CoA synthetases (EC 6.2.1.12), as well as microbial 4-chlorobenzoate-CoA ligases (see Fulda et al., 1994 and references herein). As indicated in Fig. 2, shared among this class is a domain of 400–500 amino acids of 20–80% homology that is distinguished by a linear array of at least 9 (Turgay et al., 1992; Pavela-Vrancic et al., 1994c; Pfeifer et al., 1995) strictly conserved sequence motifs.

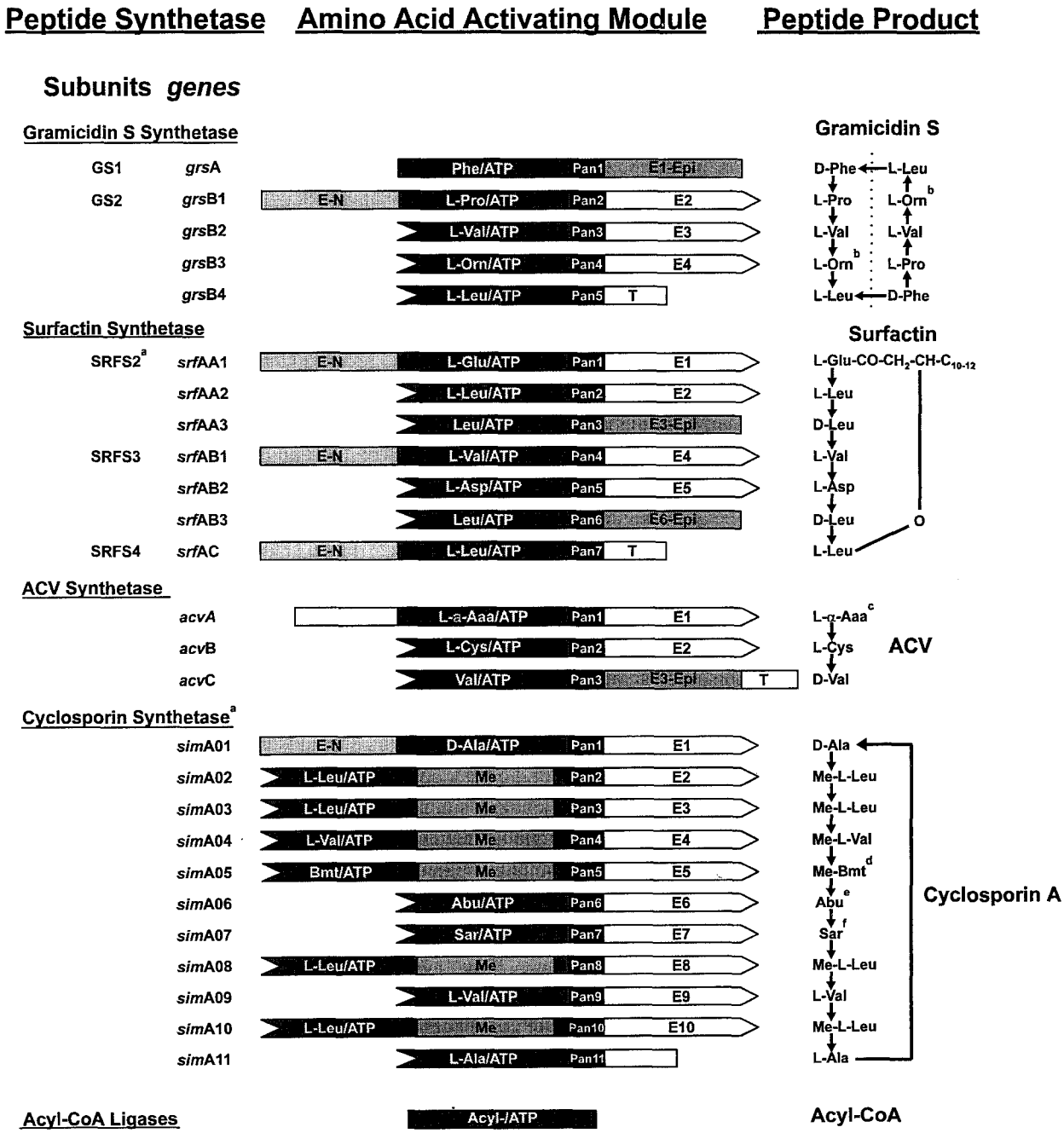
4.1.1 Motifs A and B

The functions of both have not been attributed so far. It is tempting to speculate that a strictly conserved basic lysine residue within motif B may be involved in forming a hydrogen bond to the carboxyl-group of the amino- or hydroxy acid substrate.

4.1.2 Motif C

The highly conserved motif C (SGTTGxPKG) of an amino acid activating module resembles the motif of the phosphate-binding loop (P-loop or Walker A motif GxxxxGK; Walker et al., 1982) which is found in all guanosine and some adenosine binding proteins (Higgins et al., 1986; Taylor and Green, 1989; Saraste et al., 1990; Traut, 1994). The highest similarity can be found in comparison with adenylate kinases, a class of enzymes which contain a glycine residue following the invariant lysine, a feature that distinguishes adenylate kinases from other members of the P-loop containing family (Hanks et al., 1988; Saraste et al., 1990), and the adenylosuccinate synthetase from *E. coli* (Poland et al., 1993). X-ray analyses of the crystal structures of certain members of the P-loop superfamily revealed that the P-loop connects a β-strand to an α-helix and wraps around the β-phosphate of the bound nucleotide (for reviews see Wittinghofer and Pai, 1991; Traut, 1994). The main chain amide nitrogens point towards the β- and γ-phosphate oxygens, neutralizing their charge. This structure is stabilized by the invariant lysine residue of the P-loop which forms hydrogen bonds to both, the main chain oxygens, as well as the bound NTP. The residues behind the essential Lys are involved in coordination of the Mg^{2+}-cation (Saraste et al., 1990; Wittinghofer and Pai, 1991). An α-helical segment which has been observed for such enzymes behind the invariant lysine of the P-loop has been predicted by secondary structure analysis of the area behind motif C, the putative P-loop of enzymes of the adenylate forming-family using the Garnier and GGBSM programs (Stein, T. and Vater, J., unpublished). This finding supports the proposal that motif C is part of a P-loop-like structure and that a α-helical structure behind this loop


N-terminal Elongation Domain
Amino Acid Adenylating Domain
Elongation Domain
K L MNOP R A B C DEFGHI J K L M NOP QR
4´-Phosphopantetheine Carrier Domain
0
500
1000
1500
Functional domains of an amino acid activating module
Scale (amino acid residues)


Peptide Synthetase
Amino Acid Activating Module
Peptide Product
Subunits genes
Gramicidin S Synthetase
GS1 grsA
GS2 grsB1
grsB2
grsB3
grsB4
Phe/ATP Pan1 E1-Epi
E-N L-Pro/ATP Pan2 E2
L-Val/ATP Pan3 E3
L-Orn/ATP Pan4 E4
L-Leu/ATP Pan5 T
Gramicidin S
D-Phe L-Leu
L-Pro L-Orn
L-Val L-Val
L-Orn L-Pro
L-Leu D-Phe
Surfactin Synthetase
SRFS2 srfAA1
srfAA2
srfAA3
SRFS3 srfAB1
srfAB2
srfAB3
SRFS4 srfAC
E-N L-Glu/ATP Pan1 E1
L-Leu/ATP Pan2 E2
Leu/ATP Pan3 E3-Epi
E-N L-Val/ATP Pan4 E4
L-Asp/ATP Pan5 E5
Leu/ATP Pan6 E6-Epi
E-N L-Leu/ATP Pan7 T
Surfactin
L-Glu-CO-CH2-CH-C10-12
L-Leu
D-Leu
L-Val
L-Asp
D-Leu
L-Leu
O
ACV Synthetase
acvA
acvB
acvC
L-α-Aaa/ATP Pan1 E1
L-Cys/ATP Pan2 E2
Val/ATP Pan3 E3-Epi T
L-α-Aaa
L-Cys
D-Val
ACV
Cyclosporin Synthetase
simA01
simA02
simA03
simA04
simA05
simA06
simA07
simA08
simA09
simA10
simA11
E-N D-Ala/ATP Pan1 E1
L-Leu/ATP Me Pan2 E2
L-Leu/ATP Me Pan3 E3
L-Val/ATP Me Pan4 E4
Bmt/ATP Me Pan5 E5
Abu/ATP Pan6 E6
Sar/ATP Pan7 E7
L-Leu/ATP Me Pan8 E8
L-Val/ATP Pan9 E9
L-Leu/ATP Me Pan10 E10
L-Ala/ATP Pan11
D-Ala
Me-L-Leu
Me-L-Leu
Me-L-Val
Me-Bmt
Abu
Sar
Me-L-Leu
L-Val
Me-L-Leu
L-Ala
Cyclosporin A
Acyl-CoA Ligases
Acyl-/ATP
Acyl-CoA

is involved in binding of the nucleotide also in the adenylate forming family. Detailed information is lacking about the catalytic carboxylate-group which acts as a general base to activate a water molecule for the attack of the γ-phosphate-group of ATP (Yoshida et al., 1995). Two highly conserved Glu or Asp residues 40–50 amino acids behind the P-loop motif of peptide synthetases are attractive candidates to catalyze this process but their participation in ATP hydrolysis has to be elucidated.

The importance of the strictly conserved lysine as well as the glycine residue as constituents of a slightly modified P-loop within amino acid activating modules of peptide synthetases was confirmed by site directed mutagenesis experiments which are indicated in Fig. 1. Both Lys to Arg as well as Lys to Thr substitutions of TY1 (K186R, K186T; the nomenclature of the mutations is explained in Fig. 1) show a significant or total loss of the adenylation activity, respectively (Gocht and Marahiel, 1994). The exchange of the highly conserved glycine to a structurally closely related alanine residue in TY1 (G180A) resulted in a decrease of the ATP/PP_i-exchange to 82% (Gocht and Marahiel, 1994), whereas the exchange of this Gly to a negatively charged amino acid within the Val-activating domain of GS2 (mutant gene from *Bacillus brevis* Nagano E-5, G1656D) induced a total loss of the adenylation reaction (Saito et al., 1995).

4.1.3 Motif D

Saito et al. (1995) observed a total loss of the adenylation activity for two mutans of the Val activating module of GS2 bearing a Gly→Glu replacement

Fig. 2. Modular architecture of some prominent peptide forming multienzymes. Peptide synthetases can be divided into highly homologous building blocks of 1000–1500 amino acid residues. Each of these modules is organized into specific functional domains for amino acyl adenylation [500–600 amino acid residues including motifs A-I, dark grey], thioester binding [80–100 amino acid residues around the 4'-phosphopantetheine binding motif J, black], and elongation of the peptide product [E_i, 300–400 amino acids including motifs K–R, white]. Epimerizing/racemizing modules as GS1 contain an elongation domain with significant modifications in comparison to non-epimerization modules (E_i–*Epi*, 300–400 amino acids including the modified elongation motifs K–R, grey]. For interacting protein components of peptide forming multienzymes as GS1 and GS2, for example, a specific domain at the N-terminus of the acceptor enzyme is observed showing significant homologies to the elongation domains [E_N including the N-terminal motifs K-R (without Q), light grey]. At the C-terminal end of each peptide forming multienzyme system like GS2, for example, a thioesterase II like sequence motif (T) is found instead of the elongation domain which is observed in all other modules. At the right side the structure of the peptide products are shown. [a] In the cases of SRFS2 and cyclosporin A synthetase also N-terminal elongation domains are observed. Presumably this fact reflects a necessary interaction of SRFS2 with SRFS1, an enzyme which exhibits the characteristics of an acyl-transferase, the initiating step of surfactin biosynthesis (Menkhaus et al., 1993; Galli et al., 1994). Because an alanine racemase acts as the key enzyme to provide D-Ala for cyclosporin A biosynthesis (Hoffmann et al., 1995), also in this case the interaction of the racemase with the N-terminal region of cyclosporin synthetase can be postulated. [b] *Orn*, ornithine; [c] *L-α-Aaa*, L-α-aminoadipic acid; [d] *MeBmt*, (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine; [e] *Abu*, α–amino butyric acid; [f] *Sar*, sarcosine; *Me*, N-methylated amino acid

in the region between motif C and D (mutant gene from *B. brevis* Nagano BI-3, G1768E) or a Gly→Asp exchange within motif D (E-4, G1793D). Secondary structure analysis of the region between motif C and D revealed that the G1768E mutation results in the disappearance of a β-turn which is a highly conserved structural element among adenylate-forming enzymes, whereas the G1793D mutation should not result in any change of the 3D-structure (Saito et al., 1995). Most probably the G1768 between motif C and D is an important constituent for the maintenance of the 3D-structure, while G1793 should be part of an active site functional in adenylate formation.

4.1.4 Motif E

Replacement of the cysteine residue which is highly conserved in bacterial peptide synthetases with a serine in TY1 (C364S) resulted only in a slight effect on the Phe dependent ATP/PP_i-exchange, indicating that this residue is not essentially involved in this reaction (Gocht and Marahiel, 1994). Both Cys→Ser and Arg→Gly site directed mutagenesis in coumarate CoA ligase rendered the protein completely unstable after expression in *E. coli* cells (Becker-Andre et al., 1991). Much more information is needed to clarify, whether this motif is necessary for stability of the tertiary structure of the enzyme, binding of the substrates, or catalytic activities.

4.1.5 Motif F

The highly conserved TGD-core of motif F is also observed in a family of cation pumping ATPases forming a phosphorylated intermediate (Taylor and Green, 1989). Affinity labeling with 2-azido-ATP (Davis et al., 1990) and site directed mutagenesis of the yeast plasma ATPase within the TGD-core (Portillo and Serrano, 1988) revealed that this motif is involved in ATP-binding. Similar results are obtained by site directed mutagenesis and affinity labeling of the peptide synthetase TY1. Replacement of Asp→Asn or Ser (D401N, D401S) resulted in a decrease of the adenylation reaction to 78% or 12%, respectively. For the triple mutant of TY1 (G400C, D401S, A403S) most probably changes in the conformation of the active site are responsible for the loss of adenylation activity (Gocht and Marahiel, 1994). After affinity labeling of TY1 with the ATP-derivative 2-azido-ATP followed by tryptic proteolysis of the resulting complex, a labeled active site peptide fragment could be identified as G373-K384 supporting the proposal that motif F (R394-A403) and its surrounding is involved in ATP binding (Pavela-Vrancic et al., 1994a).

4.1.6 Motif G

The five mutants with Gly→Glu/Ala/Arg/Val/Trp exchanges in position 870 of the proline activating domain of GS2 (*B. brevis* Nagano BII-3 mutants) show a significant or total loss of the L-Pro adenylation activity in comparison

to the wild type enzyme (Tokita et al., 1993). Secondary structure analysis of this enzymic region revealed that this glycine residue is a constituent of a coil or β-sheet structure between two α-helices. For all these replacements a lack in this loop-structure was predicted (Tokita et al., 1993). Affinity labeling experiments support the hypothesis that this motif is involved in ATP-binding: An active site peptide fragment of 2-azido-ATP labeled TY1 could be identified as the W405-R416, a part of the consensus motif G (Pavela-Vrancic et al., 1994a). FITC labeling of TY1 resulted in labeling of the highly conserved lysine residue (K422) of motif G (Pavela-Vrancic et al., 1994b).

4.1.7 Motif H and I

The affinity labeling experiments of TY1 described above (Pavela-Vrancic et al., 1994a/b) resulted in another 2-azido-ATP labeled active site peptide fragment (L483-K494, motif H) as well as a FITC-labeled fragment with FITC bound to position K505 (motif I) demonstrating that both motifs are involved in nucleotide binding.

4.1.8 Summary of the structure-function relationships within an aminoacyl-adenylation domain

In conclusion, the ATP recognition region of an amino acid activating module includes the phosphate binding area (motif C) and the binding regions for the adenine moiety of ATP (motifs G, H and I), whereas motif F is involved in the ATPase activity. Specific dissection of the *grs*A and *tyc*A genes and heterologous expression in *E. coli* resulted in catalytically active fragments, Specific dissection of the gramicidin S- (Stachelhaus and Marahiel, 1995b) and tyrocidine synthetase (Dieckmann et al., 1995) encoding genes between motif I and J and heterologous expression of the related proteins (see Fig. 1; GS1, 1–556 and TY1, 1–534) resulted in catalytically active fragments of amino acid activating modules which are similar in size and equipment of consensus motifs (motif A–I) like other members of the adenylate-forming superfamily. These fragments lacked in thioester incorporation activity, but were active in the adenylation of their cognate amino acid (Stachelhaus and Marahiel, 1995b; Dieckmann et al., 1995).

Little is known about the enzymic region for recognition of the carboxylic substrate. Because all members of the adenylation super-family have to recognize specifically the amino- or hydroxy acid substrate which has to be adenylated, there might be a strictly conserved positively charged residue whithin one of the functionally not yet characterized consensus motifs A and B. The less conserved region in close proximity to the phosphate binding loop between motif C and D as well as D and E has been speculated to contain the recognition elements for the carboxylic substrate side chain (Cosmina et al., 1993). However, comparison of this region, putatively involved in Val or Leu recognition, for example, between the bacterial surfactin synthetase and the fungal ACV- or cyclosporin synthetase did not reveal strong sequence simi-

larities (Cosmina et al., 1993; Stein, T. and Vater, J., unpublished), a fact that might reflect a lower degree of conservation between enzymes of procaryotic and eucaryotic origin.

4.1.9 Aminoacyl-adenylation of tRNA synthetases

In ribosomal peptide biosynthesis a family of 20 aminoacyl-tRNA synthetases (aaRS) catalyze the activation of their cognate substrates in a two step process including 1) the specific aminoacyl-adenylation similar to the nonribosomal system, and 2) aminoacyl-esterification of the substrate to the 2′ or 3′-OH of its cognate transfer RNA ($tRNA^{AA}$) (Schimmel, 1987). Peptide bond formation is subsequently managed by the ribosomal machinery.

Amino acid sequence analysis of the 20 aaRS (Eriani et al., 1990) together with the X-ray crystal-structures of TyrRS, Met RS, GlnRS, SerRS, and AspRS of several organisms revealed that these enzymes can be subdivided into two different classes of 10 each, termed Class I and Class II (for recent reviews see Schimmel, 1991; Burbaum and Schimmel, 1991; Mirande, 1991; Moras, 1992; Carter, Jr., 1993). Two different active site architectures for ATP dependent acyl group activation have been evolved and conserved throughout all members of each class. The structure shared among Class I synthetases is a central core domain which is built around a parallel β-sheet surrounded by α-helices with a topology first discovered in dehydrogenases by Rossmann et al. (1974) and therefore known as the Rossmann- or dinucleotide-binding-fold. The two highly conserved motifs HxGH and KMSK can be used as signature sequences (Eriani et al., 1990). Although their position and distance is not conserved within the primary structure of all members of Class I synthetases, they have the same site of location within the tertiary structure of the Rossmann-fold. The two histidine residues of the HxGH-motif are thought to interact with specific phosphates of bound ATP. The KMSK-sequence participates also in ATP binding and possibly in docking the 3′ -end of the tRNA into the nucleotide fold. The structure of the nucleotide binding element of Class II synthetases is distinguished by an eight-stranded antiparallel β-sheet structure. Three signature sequences of degenerated sequence similarity were identified within this region (Eriani et al., 1990). It is remarkable that none of the Class I or II aaRS signature sequences is found within the primary sequence of peptide synthetases. Obviously the enzymatic structures catalyzing the fundamental process of amino acid activation as aminoacyl-adenylates within both the ribosomal and nonribosomal peptide/protein forming systems, have evolved independently. In contradiction to the high fidelity of the amino acid recognition of aaRS which includes hydrolysis of misactivated aminoacyl-adenylates as well as acylated tRNA (for a review see Schimmel and Schmidt, 1995), the aminoacyl-adenylating domains of peptide synthetases also catalyze the activation of a series of substrate analogues which were incorporated into the peptide product in vivo and in vitro (Kleinkauf and von Döhren, 1990b; Vater, 1990).

Detailed information about the process of substrate binding and adenylation can be obtained from the tertiary structure of the tyrosine-

tRNA-synthetase. At the moment no tertiary structure of peptide forming multienzymes is available and it is tempting to speculate that a similar structure (convergent evolution) can be observed for peptide synthetases. The structure of tyrosine-RS from E. coli, a member of Class I, is known in the presence of tyrosine (Brick and Blow, 1987) as well as tyrosyl-adenylate (Brick et al., 1989). Tyrosine binding of this enzyme occurs by fixing its amino group by tetrahedrically arranged hydrogen bonds to Asp78, Tyr169, and Gln173 which leave little flexibility for positioning the carboxyl-group. The hydroxyl-group of tyrosine is hydrogen-bonded to the side chains of Asp176 and Tyr34. The contribution to enzyme specificity of Asp176 for tyrosine versus phenylalanine cannot be estimated by site directed mutagenesis, a method which is used for other interactions in the active site, because mutation of this residue renders the enzyme totally inactive. Investigation of this enzyme altered at Asn123 and Trp126, both residues forming hydrogen bonds with the essential Asp176, revealed that the specificity of these mutants for tyrosine and phenylalanine can be decreased (N123A) or increased (Q173E). Surprisingly the wild type enzyme has not reached the maximum limit of discrimination between both aromatic compounds (de Prat Gay, 1993). At least 7–8 amino acid residues of TyrRS distributed within the Rossmann fold are involved in ATP binding. However, the α-phosphate group of the nucleotide is almost devoid of specific interactions with the synthetase with exception of a hydrogen bond to the amide NH of Asp 38. Obviously, these structural properties give the opportunity for nucleophilic attack of the carboxyl-group of the tyrosine to the α-phosphate-group of the nucleotide resulting in the highly reactive tyrosyl-adenylate intermediate. This compound is stabilized by additional, specific interactions to the protein chain (Brick et al., 1989) which prevent the hydrolysis or the release of the intermediate before acyl-transfer to the enzymatically bound cognate tRNA can occur.

4.2 Thioester binding domain

4.2.1 Detection of 4′-phosphopantetheine as the thioester binding site

The consensus motif J is strictly conserved within each amino acid activating module of peptide synthetases whose sequences are known so far, enzymes which thioesterify their cognate substrates at intrinsic thiol-groups (Schlumbohm et al., 1991; Stein et al., 1995). This motif is not found within other members of the adenylate-forming superfamily that transfer their activated substrates to the thiol-group of an external CoA molecule (McCabe et al., 1991; Turgay et al., 1992; Dieckmann et al., 1995). The research in our laboratory is focused on the characterization of the thioester binding sites of gramicidin S synthetase on the molecular level. For this purpose GS1 and GS2 were affinity labeled either with radioactive substrate amino acids (Schlumbohm et al., 1991) or the thiol inhibitors N-ethylmaleimide (Stein et

al., 1994), and iodoacetic acid (Stein et al., 1995). The enzyme-substrate/inhibitor complexes were digested with cyanogen bromide or proteases (trypsin, *Staphylococcus aureus* V8 protease, and pepsin) and the radioactively labeled peptide fragments of the thioester binding reaction centers were isolated in pure form by multistep reversed phase methodology (Schlumbohm et al., 1991; Stein et al., 1994; T. Stein and J. Vater, unpublished). Their amino acid sequences contain the highly conserved thioester binding motif LGG(H/D)S(L/I), which does not contain a reactive cysteine as postulated by the original thiotemplate hypothesis (Kleinkauf et al., 1971; Lipmann, 1973; Laland and Zimmer, 1973; Kurahashi, 1974). However, the (H/D)SL/I-core of motif J resembles the 4′-phosphopantetheine binding sites of acylcarrier proteins or domains of fatty acid and polyketide synthases with the serine residue as binding site for the Pan-cofactor (Schlumbohm et al., 1991) implying that each module of gramicidin S synthetase is equipped with a separate Pan-prosthetic group. To obtain experimental evidence for the attachment of such a cofactor to the thiotemplate sites of gramicidin S synthetase the active site peptides were investigated by FAB- and ESI-MS). The molecular masses of the thiolation site peptide fragments of gramicidin S synthetase are appreciably higher than the values calculated from the grsA and grsB genes which encode GS1 and GS2, respectively (Table 2). Each of the observed mass differences indicate a covalent substitution of the peptide moiety with a 4′-phosphopantetheine cofactor which is alkylated with the radiolabeled thiol inhibitor. The nature, covalent linkage and site of location of the Pan substituent attached to the serine residue of the thiolation motif was proven by interpretation of the fragmentation data obtained from FAB-MS as well as collision induced dissociation ESI-MS. The nature of the substituent was supported by amino acid analysis (Stein et al., 1994). One mole of each active site peptide fragment of gramicidin S synthetase indicated in Table 2 contained approx. 1 mole β-alanine, which is a constituent of 4′-phosphopantetheine. These results give evidence that each amino acid activating module of gramicidin S synthetase is equipped with a separate 4′-phosphopantetheine prosthetic group esterified to the active site serine of the (H/D)SL-core motif and are strong support for a "Multiple Carrier Model" of nonribosomal peptide biosynthesis at multienzymatic templates. The essential role of the serine residue for binding of the Pan-cofactor was supported by site directed mutagenesis of surfactin synthetase and TY1 and by analysing of fragments of the GS1 and TY1 modules. Specific replacement of the serine residue by alanine in the modules of surfactin synthetase 2 (SrfAA1, S1006A; SrfAA2, S2045A; SrfAA3, S3073A), the first module of surfactin synthetase 3 (SrfAB1, S999A; Vollenbroich et al., 1993 and 1994), and TY1 (S562A; Gocht and Marahiel, 1994) resulted in a loss of the thioester binding capacity of the mutagenized enzymes, whereas usually the substrate dependent ATP/PP_i-exchange activities were not significantly affected. Taken together, all these experiments are strong indications for a 4′-phosphopantetheine attachment of peptide synthetases with multiple cofactors, one within each amino acid activating module.

Table 2. **A** Summary of the structure analysis of the active site peptides of the thiolation centers for phenylalanine and L-valine of gramicidin S synthetase
B 4′-Phosphopantetheine carrier domains of peptide synthetases, polyketide-, and fatty acid synthases, as well as NodF gene products

A

Enzyme	Thiotemplate site	Structure of the radiolabeled active site peptide fragments	Molecular mass (Dalton) calculated from gene[a]	Pan-adduct[b]	measured by ESI-MS[c]
GS1	Phenylalanine	DNFYALGGDSIK └Pan-[^{3}H]CM	1299.4	1698.1	1697.5
GS2	L-Valine	LGGHSLR └Pan-[^{3}H]NES	739.0	1204.3	1204.0

B

```
                motif I                                  Pan-binding motif J
grsA      ..LPEPDLTFGMRVDYEAPRNEIEETLVTIWQDVLG----IEKIGI---KDNFYALGGDSIKAIQVAARLHSY-QLKLETKDLLKYPTIDQLVH-YIKDSKRRSQEGIVEG..
grsB1     ..LPNLEGIVNTNAKVVPTNELEEKLAKIWEEVLG----ISQIGI---QDNFFSLGGHSLKAITLISRMNKECNVDIPLRLLFEAPTIQELSN-YINGAKKESYVAIQPV..
grsB2     ..LPKPDGEFGTATEYVAPSSDIEMKLAEIWHNVLG----VNKIGV---LDNFFELGGHSLKAMTMISQVHKEFDVELPLKVLFETPTISALAQ-YIADGQKGMYLAIQPV..
grsB3     ..LPEPDGSISIGTEYDRPRTMLEGKLEEIWKDVLG----LQRVGI---HDDFFTIGGHSLKAMAVISQVHKECQTEVPLRVLFETPTIQGLAK-YIEETDTEQYMAIQPV..
grsB4     ..LPEPQTIGLMAREYVAPRNEIEAQLVLIWQEVLG----IELIGI---TDNFFELGGHSLKATLIVAKIYEYMQIEMPLNVVFKHSTIMKIAE-YITHQESENNVH-QPI..
srfAA1    ..LFALEVKAVSGTAYTAPRNETEKAIAAIWQDVLN----VEKAGI---FDNFFETGGDSLKAMTLLTKIHKETGIEIPLQFLFEHPTITALAE-EADHRESKAFAVIEPA..
srfAA2    ..LPIPDANVSRGVSYVAPRNGTEQKVADIWAQVLQ----AEQVGA---YDHFFDIGGHSLAGMKMLALVHQELGVELSLKDLFQSPTVEGLAQ-VIASAEKGTAASISPA..
srfAA3    ..LPEPDIEAG-SGEYKAPTTDMEELLAGIWQDVLG----MSEVGV---TDNFFSLGGDSIKGIQMASRLNQH-GWKLEMKDLFQHPTIEELTQ-YVERAEGKQAD-QGPV..
srfAB1    ..LPAPQSE-AVQPEYAAPKTESEKKLAEIWEGILG----VKAGV----TDNFFMIGGHSLKAMMMTAKIQEHFHKEVPIKVLFEKPTIQELAL-YLEEIESKEEQTFEPI..
srfAB2    ..LPKPNAAQSGGKALAAPETALEESLCRIWQKTLG----IEAIGI---DDNFFDLGGHSLKGMMLIANIQAELEKSVPLKALFEQPTVCQLAV-YMEASAVSGGH-QVLK..
srfAB3    ..LPEPDIEAG-SGEYKAPTTDMEELLAGIWQDVLG----MSEVGV---TDNFFSLGGDSIKGIQMASRLNQH-GWKLEMKDLFQHPTIEELTQ-YVERAEGKQAD-QGPP..
srfAC     ..LPKPDQDQ-LAEEWIGPRNEMEETIAQIWSEVLG----RKQIGI---HDDFFALGGHSLKAMTAVPHQQEL-GIDLPVKLLFEAPTIAGISA-YLKNGGSDGL--QDVT..
acvA      ..LPSVDLIQPKVSSCEL-TDEVEIALGKIWADVLG----AHHLSI-SRKDNFFRLGGHSITCIQLIARIRQQLGVIISIEDVFSSRTLERMAE-LLRSKESNGTPDERAR..
acvB      ..LPTAEEKGAMN--VLAPRNEIESILCGISAGLLD----ISAQTI---GSDFFTLGGDSLKSTKISFKIHEVFGRTISVSALFRHRTIESL-AHLIMNNVGD-IQEITPV..
acvC      ..LPDIGNPQHQIS-YNPPRDVLEADLCRLWASALG----TERCGI---DDDLFRLGGDSITALHIAAQIHHQIGRKVTVRDIFDHPTIRGIHDNVMVKLVP-HVPQFQA..
                          XXXXXXXXX\\\\          \\\\\\\\\//XXXXXXXXX/////  XXXXXXX XXXXXXXXX
ACP E.coli[d]          STIEERVKKIIGEQLG----VKQEEVTNNASFVEDLGADSLDTVELVMALEEEFDTEIPDEEAEEKITTVQAAIDYINGHQA
ACP yeast           MDRKEIFE-RIEQVLAEQLG----IPAEQITEEADLREDLGMDSLDLVELVSALEDEVGMRVEQSQLEGIETVGHVME-LTLDLVARLATASAADK..
ACP rape            AAKQETVE-KVSEIVKKQLS---LKDDQQVV-AETKFVDLGADSLDTVEIVMGLEEEFGIQMAEEKAQKIATVEQAAE-LIEELMQAKK
ACP spinach          AKKETID-KVSDIVKEKLALG---ADVVVT-ADSEFSKLGADSLDTVEIVMGLEEEFGINVDEDKAQDISTIQQAAD-VIE-LLEKKA
ACP chicken          SEGGSQR-DLVEAVAHILG----VRDVSSLNAESSLADLGLDSLMGVEVRQTIERDYDIVMTMREIRLL-TINKLRE-LSSKTGTAEELKPSQVL
ACP rat           EKKAVAHGDGEAQR-DLVKAVAHILG---IRDLAGINL-DSSLADLGLDSLMGVEVRQILEREHDLVLPIREVRQL-TLRKLQE-MSSKAGSDTEL
ACP DEBS2N           EQQENLLE-LVANAVAEVLG---HES-AAEIN-VRAFSELGLDSLNAMALRKRLSASTGLRLPASLVFDHPTVTALAQHLRARLVGDAD--QAA..
ACP DEBS3N           REREHLAH-LIRAEVAAVLG---HGDDAAIDR-DRAFRDLGFDSMTAVDLRNRLAAVTGVREAATVVFDHPTITRLADHYLERLVGAAEAEQAP..
ACP ARS              MATLLTTD-DLRRALVECAGETDGTDLSGDFL-DLRFEDIGYDSLALMETAARLESRYGVSIPDDVAGRVDTPRELLD-LINGALAEAA..
ACP GRANS             MARLTLDGLRTIIVA-CAGEDDGVDLSGDIL-DITFEELGYDSLALMESASRIERELGVALADGDINEELTPRVLLD-LVNGAQAEAA..
ACP TCMS             MPQIGLPR--LVEIIRECAGDPDERDLDGDIL-DVTYQDLGYDSIALLEISAKLEQDLGVSIPGEELK---TPRHTLH-LVNTETAGEVA..
NodF R.meliloti      MVDQLESEIIGIIKNRVESEGGDGETALIVGDLTAATELTALGVDSLGLADIIWDVEQAYGIR------IEMNTAEAWSD-LQNVGDIVGAIRGLLTKGA
NodF R.trifolii      MADQLTVEIIAAIKNLAQSENG-GRIPAAIGDITADRQLTSLGIDSLALADVLWDLEQAYGIR------IEMNTADAWSN-LKNIGDVVEAVRGLIAKEA
NodF R.legumin.      MADQLTLEIISAINKLVKAENG-ERTSVALGEITTDTELTSLGIDSLGLADVLWDLEQLYGIK------IEMNTADAWSN-LNNIGDVVEAVRGLLTKEV
```

A [a]The molecular masses of the peptides were calculated from the gene derived sequence. [b]The molecular masses of the peptide-adducts were calculated as the sum of the molecular masses of the peptide moiety, the 4′-phosphopantetheine substituent which is covalently attached to the serine residue, and the radioactively labeled tracer bound to the reactive thiol group of the Pan cofactor (*CM*, carboxy-methyl; *NES*, N-ethylsuccinimido). [c]The molecular masses of the active site peptide fragments were measured by electrospray mass spectrometry (ESI-MS; Stein et al., 1994).

B The gene derived sequences of the thioester binding domain of some peptide synthetases are compared with acyl carrier proteins or domains of fatty acid-, and polyketide synthases, as well as NodF gene products which contain the Pan binding motif. *Grs*A and *grs*B1–4, gramicidin S synthetase 1 and domain 1–4 of gramicidin S synthetase 2 from *Bacillus brevis* (Turgay et al., 1992); *srf*AA1–3, *srf*AB1–3, domains 1–3 of surfactin synthetase 2 and 3, as well as *srf*AC, surfactin synthetase 4 from *Bacillus subtilis* (Cosmina et al., 1993); *acv*A-C, δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) synthetase from *Aspergillus nidulans* (MacCabe et al., 1991); ACP, acyl carrier protein from *Escherichia coli* (Vanaman et al., 1968), *Saccharopolyspora erythraea* (Hale et al., 1987), rape (Safford et al., 1988), spinach (Kuo et al., 1984), chicken (Yuan et al., 1988), rat (Witkowski et al., 1987), 6-deoxyerythronolide-B synthase 2 and 3 from *Saccharopolyspora erythraea* (Bevitt et al., 1992) actinorhodin synthase from *Streptomyces coelicolor* A3 (Fernández-Moreno et al., 1992), granaticin synthase from *Streptomyces violaceoruber* (Sherman et al., 1989), tetracenomycin synthase from *Streptomyces lividans* (Bibb et al., 1982); NodF gene products of *Rhizobium meliloti* (Debellé and Sharma, 1986), *trifolii* (Schofield et al., 1986), and *leguminosarum* (Shearman et al., 1986). [d]The amino acid sequence of the acyl carrier protein of *E. coli* is shown in bold types. Its three-dimensional structure in solution was determined by NMR (Holak et al. 1988a/b). Results of the structure analysis are summarized in the line above: *X*, α-helix region; /, right, and \, left handed turn; the residues Val^{29} and Ala^{34} are underlined. The short contact between $Val^{29}C^{\alpha}H$ and $Ala^{34}NH$ of 0.23 nm is characteristic for the 3D-structure of the acyl carrier protein.

4.2.2 The 4′phosphopantetheine binding site is part of an acyl carrier domain

Sequence alignment of a region comprising 80–100 amino acid residues around the Pan binding motif of peptide synthetases with acyl carrier proteins or domains of fatty acid and polyketide synthases, as well as NodF gene products indicates certain sequence similarities as demonstrated in Table 2. A secondary structure analysis of this part of peptide synthetase modules was performed using the GARNIER and GGBSM programs (T. Stein and J. Vater, unpublished). In each case three prominent α-helical stretches were predicted. This structure was compared with the structure analysis of the acyl carrier protein (ACP) of *E. coli* in solution determined by NMR (Holak et al., 1988a/b). The N-terminal part of the ACP of *E. coli* is distinguished by a short left handed β-turn changing its direction in the core region of the Pan binding motif, followed by an α-helical segment extending 15 amino acid residues from the active site serine (α-helix 2, residues 37-51). It corresponds to an α-helical region predicted for the amino acid activating modules comprising 10–15 amino acid residues in connection to the active site serine. The two other α-helical regions within the Pan-binding domains of peptide synthetase modules were predicted 20 amino acid residues upstream, and 25 amino acid residues downstream the Pan binding motif, respectively. Also these helices can be related to α-helices of the *E. coli* ACP (helix 1, residues 3–15 and helix 4, residues 65–75; Holak et al., 1988a/b). Obviously each amino acid activating module of peptide forming multienzymes is equipped with a separate 4′-phosphopantetheine carrier domain with a size of 8–10kDa which is integrated between the adenylation and elongation domain of a module. Its three-dimensional structure is homologous to acyl carrier proteins of fatty acid and polyketide synthases.

4.2.3 The prebiotic history of 4′-phosphopantetheine

All components of 4′-phosphopantetheine, the inorganic phosphate (Osterberg and Orgel, 1972), as well as the organic constituents pantoic acid (pantoyl lactone; Miller and Schlesinger, 1993b), β-alanine (Miller, 1953; Miller and Urey, 1959), cysteamine (Choughuley and Lemmon, 1966; Miller and Schlesinger, 1993a), and pantothenic acid (the condensation product of pantoic acid with β-alanine, Miller and Schlesinger, 1993b) are probably prebiotic compounds. A recent report of a synthesis of pantetheine (the condensation product of pantothenic acid with cysteamine) under prebiotic conditions supports the proposal that pantetheine could have been a component of the prebiotic soup (Keefe et al., 1995), and that this compound could have played an important role in chemical evolution, in particular, as the precursor of coenzymeA functioning as an acyl carrier and activator in very early metabolic systems (Lipmann, 1965; de Duve, 1987, 1991). The utilization of multiple 4′-phosphopantetheine cofactors in peptide forming multienzyme systems for aminoacyl-thioesterification of the substrate amino acids and transport of the peptide intermediate is consistent with the view that our

present DNA/RNA-world is preserving reactive structures that might have evolved in a preceding archaic thioester world (Lipmann, 1965; de Duve, 1987, 1991).

4.3 Elongation domain

The C-terminal region of the amino acid activating module with a size of 300–400 amino acids is less conserved than the adenylation or the Pan-carrier domain. Within this segment a series of 8 conserved sequences have been identified (motifs K-R). Their site of location behind the Pan-carrier domain and their appearance only in modules catalyzing elongation processes led to the proposal that some of these motifs are involved in peptide bond formation and domain interaction (Fuma et al., 1993; Pfeifer et al., 1995; De Crécy-Lagard et al., 1995). Their position is strictly conserved within amino acid activation modules, but their specific functions still need to be elucidated. De Crécy-Lagard et al. (1995) reported that motif M (HHxxxDG) is putatively involved in acyl- and peptidyl transfer reactions. This hypothesis is supported by the observation of a similar motif in other enzymes affecting acyl transfer such as chloramphenicol acatyltransferase and dihydrolipoamide acyltransferase.

4.4 Epimerizing domain

Because of the significant differences of the consensus sequences N-R between epimerizing and non-epimerizing modules it has been hypothesized that some of these structures play a crucial role in the epimerization process (Fuma et al., 1993; Stachelhaus and Marahiel, 1995b; Stein et al., 1995; Pfeifer et al., 1995). Truncation of gramicidin S synthetase 1 between motif M and N results in a fragment which is catalytically active in adenylate- and thioester formation, but devoid of the racemization activity of phenylalanine (Fig. 1; GS1, 1–807; Stachelhaus and Marahiel, 1995b).

The wild-type gramicidin S synthetase 1, catalyzes the racemization of phenylalanine in the thioesterified stage exhibiting the characteristics of a racemase using a single base for the proton transfer process (Kanda et al., 1989). This enzyme does not show any sequence homologies to other amino acid racemases that follow the one-base pathway, like the pyridoxal-dependent alanine racemases (for a review see Walsh, 1989). Therefore, gramicidin S synthetase 1 can be classified as a prototype of amino acid racemases containing the cofactor 4′-phosphopantetheine (Stein et al., 1995) which epimerize their thioesterified amino acid portion either in the stage of the single amino acid (GS1, TY1; Stein et al., 1995; Stachelhaus and Marahiel, 1995b) or in the stage of the peptide intermediate (ACV-, surfactin-, actinomycin synthetase; Stindl and Keller, 1994; Shiau et al., 1995). Presumably one of the highly conserved basic amino acid residues within the motifs N-R represents the reactive element involved in the proton transfer during the epimerization process (Stein et al., 1995) or is a constituent of an acid-base charge transfer

system which stabilizes the negative charge of the enolic intermediate in a similar manner as observed for mandelate racemase (Neidhart et al., 1991; Gerlt and Gassmann, 1993; Landro et al., 1994; Babbitt et al., 1995).

4.5 N-terminal elongation domain

In the case of the interaction of individual proteins of peptide forming multienzymes as GS1 and GS2 or the components of surfactin synthetase a specific domain is observed at the N-terminus of the acceptor enzyme showing homology to the elongation domains (Fuma et al., 1993; Stein and Vater, unpublished). Presumably this structural element plays an important role in the protein-protein interaction of the consecutive enzymes.

4.6 N-methylation domain

Certain bioactive peptides, in particular of eucaryotic origin, as the depsipeptide enniatin and the lipopeptide cyclosporin A contain several N-methylated amino acid residues. The related peptide synthetases catalyze the N-methylation reaction of these residues using S-adenosyl methionine as the methyl donor at the stage of the thioesterified amino acid (Billich and Zocher, 1987; Lawen and Zocher, 1990). Modules containing the N-methylation activity can be considered as hybrides between normal modules and N-methyltransferases (Haese et al., 1993). They have been classified as type II modules (Stachelhaus and Marahiel, 1995a). Here a segment of approximately 450 amino acid residues is inserted between motif G and H of the adenylation domain (Haese et al., 1993; Weber et al., 1994). This segment, especially the glycine-rich motif (E/D)x(G/F)xG, shows similarities to SAM-dependent methyltransferases specific for proteins, rRNAs and DNA (Haese et al., 1993). The N-methylation domain is integrated into the last fourth of the adenylation domain between motif G an H, both motifs are involved in nucleotide binding (Tokita et al., 1993; Pavela-Vrancic et al., 1994a/b). It has to be elucidated if the N-methylation of the substrates already occurs at the stage of the adenylated substrate.

4.7 Thioesterase domain/protein

At the C-terminal end of bacterial peptide forming systems, like gramicidin S or surfactin synthetase and in the case of the ACV synthetase a segment is located replacing the elongation domain which shows similarities to thioesterases (Alvarez et al., 1993; Cosmina et al., 1993; van Liempt et al., 1993) and contains a GxSxG-core sequence in its reaction center. In addition, at the 5′-end of the grs operon (Krätzschmar et al., 1989), at the 3′-end of the *srf*A operon (Cosmina et al., 1993), and associated with the gene cluster encoding the proteins involved in bialaphos biosynthesis (Raibaud et al., 1991) structural genes are located which encode proteins with molecular

masses of 29, 25, and 27kDa resp., showing about 30% homology among themselves and to vertebrate thioesterase II ezymes involved in fatty acid biosynthesis (Krätzschmar et al., 1989). These proteins contain the GxSxG-core, the "lipase or esterase"-consensus sequence (Brenner et al., 1988) which has been observed in different classes of hydrolytic enzymes and are involved in many different pathways like thioesterases and acyltransferases in the biosynthesis of fatty acids (Babbitt et al., 1992), various polyketides (Donadio et al., 1991; Cortes et al., 1990), and penicillin (Alvarez et al., 1993). The function of both the thioesterase-like sequence within the last amino acid activating module of peptide forming multienzyme systems as well as the thioesterase-like enzymes associated with the peptide synthetase genes in the same operons is not known. The location of such a domain at the end of a multienzyme implies that it could catalyze a termination reaction like the release and/or cyclization of the thioesterified peptide product (van Liempt et al., 1993). Further biochemical and genetic analysis is neccesary to attribute specific functions to the thioesterase domains or proteins in peptide biosynthesis.

5. Summary and prospect

Thorough analysis of the primary structures of peptide forming multienzymes revealed a highly conserved and ordered modular architecture. Each of the modules is organized into specific functional domains for aminoacyl adenylation, thioesterification, and in some cases modification (N-methylation, epimerization) of its amino- or hydroxy acid substrate as well as elongation of the peptide product. The adenylation domain comprising 500–600 amino acid residues is conserved within a superfamily of adenylate forming enzymes and has been demonstrated to contain specific sites for substrate (ATP and amino acid) recognition and activation. An acyl-carrier domain (80–100 amino acid residues) has been identified to contain the LGG(H/D)-S(L/I)-motif as a 4′-phosphopantetheine binding site. The thiol-group of the cofactor represents the binding site (thiotemplate) for covalent incorporation of the substrate amino acid and functions as an internal carrier system for the transport of the peptide intermediate. A third domain of 300–400 amino acid residues has been postulated to be involved in the elongation of the peptide product. In the case of modules catalyzing the epimerization of their amino acid portion slightly modified elongation domains have been observed. The functional characterization of the highly conserved consensus motifs in the hypothesized elongation domain will provide the clue to understand the process of peptide bond formation. The existence of specific recognition elements for the growing peptide chain within this part of a module is a possible explanation for the unidirectional assembly of the peptide product, presumably in combination with conformational changes induced during the transpeptidation reactions. Each module of peptide synthetases catalyzing the SAM-dependent N-methylation of their cognate substrate contains a methyltransferase domain inserted within the last fourth of the adenylation

Table 3. Schematic presentation of the Multiple Carrier Model of nonribosomal peptide biosynthesis at multienzymatic templates

1)	**Aminoacyl-adenylation**	$E - Pan_n SH + AA_n + ATP$	$\rightleftharpoons$	$AA_n - AMP \cdot E - Pan_n SH + PP_i$
2)	**Aminoacyl-thiolation**	$AA_n - AMP \cdot E - Pan_n SH$	$\rightleftharpoons$	$E - Pan_n S - AA_n + AMP$
3)	**Elongation**	$E - Pan_n S - AA_n - .... - AA_1 + E - Pan_{n+1} S - AA_{n+1}$	$\longrightarrow$	$E - Pan_{n+1} S - AA_{n+1} - AA_n - .... - AA_1 + E - Pan_n SH$
4)	**Modification**			
	a) **Epimerization**	$E - Pan_n S - L\text{-}AA_n - .... - AA_1$	$\rightleftharpoons$	$E - Pan_n S - D\text{-}AA_n - .... - AA_1$
	b) **N-Methylation**	$E - Pan_n S - AA_n - .... - AA_1 + AdoMet$	$\rightleftharpoons$	$E - Pan_n S - AA_n (Met) - .... - AA_1$

1) Aminoacyl-adenylation. The n^{th} amino acid activating module of a peptide synthetase E_n selects its cognate amino acid (AA_n) and activates it as an aminoacyl-adenylate within the adenylation domain. The energy for this process is provided by hydrolysis of an ATP α-β-linkage. This activation reaction is similar to the amino acid adenylation process of aminoacyl-tRNA synthetases in the ribosomal system.
2) Aminoacyl-thiolation. Each amino acid activating module E_n is equipped with a separate 4′-phosphopantetheine cofactor (Pan_n). The substrate amino acid is thioesterified to the sulfhydryl group of the Pan carrier ($Pan_n SH$). Such an enzyme-substrate complex is characteristic for the non-ribosomal peptide forming system.
3) Elongation. The elongation of the growing peptide chain is performed in a series of transpeptidation steps by the interplay of the Pan cofactors which act as an internal transport system for the growing peptide intermediate.
4) Modification. Possible modifications of the substrates, like *a*) epimerization or *b*) N-methylation are performed in the thioester-bound stage of the peptide intermediate.

domain. The schematic presentation for all possible reactions catalyzed by an amino acid activating module of peptide synthetases considering the "Multiple Carrier Model" of nonribosomal peptide biosynthesis is given in Table 3.

Microorganisms have evolved a vast number of gene structures coding for a great variety of homologous amino- and hydroxy acid activating modules representing the genetic code for nonribosomal peptide biosynthesis. The recently reported rational design and the production of peptide antibiotics by targeted replacement of bacterial and fungal amino acid activating modules (Stachelhaus et al., 1995) demonstrate that these functional units can act as independent enzymes and are interchangable between different organisms. Also the biosynthesis of several novel polyketides by recombinant assembly of multifunctional polyketide synthase subunits has been reported (McDaniel et al., 1993, 1995; Cortes et al., 1995). An attractive challenge for the future will be the extensive programmed recombination of such cellular factories for the biotechnological production of bioactive molecules with novel and potentially useful properties.

Acknowledgements

For the work in the authors laboratory the financial support by the Deutsche Forschungsgemeinschaft (grant Va 63/3-2) and by the European Community (Concerted Action Bio 2-CT 94-2051) is gratefully acknowledged.

References

Aharonowitz Y, Bergmeyer J, Cantoral JM, Cohen G, Demain AL, Fink U, Kinghorn J, Kleinkauf H, MacCabe A, Palissa H, Pfeifer E, Schwecke T, van Liempt H, von Döhren H, Wolfe S, Zhang J (1993) δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine synthetase, the multienzyme integrating the four primary reactions in β-lactam biosynthsis, as a model peptide synthetase. Biotechnology 11: 807–810

Alvarez E, Meesschaert B, Montenegro E, Gutiérrez S, Díez B, Barredo JL, Martín JF (1993) The isopenicillin-N acyltransferase of *Penicillium chrysogenum* has isopenicillin-N-amidohydrolase, 6-aminopenicillanic acid acyltransferase and penicillin amidase activities, all of which encoded by the single *pen*DE gene. Eur J Biochem 215: 323–332

Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang K-H, Liang P-H, Dunaway-Mariano D (1992) Ancestry of the 4-chlorobenzoate dehalogenase: analysis of amino acid sequence identities among families of acyl:adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases. Biochemistry 31: 5594–5604

Babbitt PC, Mrachko GT, Hasson MS, Huisman GW, Kolter R, Ringe D, Petsko GA, Kenyon GL, Gerlt JA (1995) A functionally diverse enzyme superfamily that abstracts the α-protons of carboxylic acids. Science 267: 1159–1161

Becker-André M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato. J Biol Chem 266: 8551–8559

Bevitt DJ, Cortes J, Haydock SF, Leadlay PF (1992) 6-Deoxyerythronolide-B synthetase 2 from *Saccharopolyspora erythraea*. Eur J Biochem 204: 39–49

Bibb MJ, Cohen SN (1982) Gene expression in streptomyces: construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. Mol Gen Genet 187: 265–277

Billich A, Zocher R (1987) N-Methyltransferase function of the multifunctional enzyme enniatin synthetase. Biochemistry 26: 8417–8423

Brick P, Blow DM (1987) Crystal structure of a deletion mutant of a tyrosine-tRNA synthetase complexed with tyrosine. J Mol Biol 194: 407–17

Brick P, Bhat TH, Blow DM (1989) Structure of tyrosyl-tRNA synthetase refined at 2.3. Å resolution. Interaction of the enzyme with the tyrosyl adenylate. J Mol Biol 208: 83–98

Brenner S (1988) The molecular evolution of genes and proteins: a tale of two serines. Nature 334: 528–530

Burbaum JJ, Schimmel P (1991) Structural relationships and the classification of aminoacyl-tRNA synthetases. J Biol Chem 266: 16965–16968

Carter CW Jr (1993) Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu Rev Biochem 62: 715–748

Choughuley ASU, Lemmon RM (1966) Production of cysteic acid, taurine and cysteamine under primitive earth conditions. Nature 210: 628–629

Coque JJR, Martín JF, Calzada JG, Liras P (1991) The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptide synthetase, and *pcb*C of *Nocardia lactamdurans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. Mol Microbiol 5: 1125–1133

Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadley PF (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348: 176–178

Cortes J, Wiesmann KEH, Roberts GA, Brown MJB, Staunton J, Leadley PF (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. Science 268: 1487–1489

Cosmina P, Rodriguez F, Ferra FD, Grandi G, Perego M, Venema G, van Sinderen D (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol Microbiol 8: 821–831

Davis CB, Smith KE, Campbell BN Jr, Hammes GG (1990) The ATP binding site of the Yeast plasma membrane proton-translocating ATPase. J Biol Chem 265: 1300–1305
Debellé F, Sharma SB (1986) Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. Nucleic Acids Res 14: 7453–7472
De Crécy-Lagard V, Marlière P, Saurin W (1995) Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. C R Acad Sci Paris. Life Sci 318: 927–936
de Duve C (1987) Selection by differential molecular survival: a possible mechanism of early chemical evolution. Proc Natl Acad Sci USA 84: 8253–8256
de Duve C (1991) Blueprint for a cell: the nature and origin of life. N Patterson Publishers, Burlington New York
de Prat Gay G, Duckworth HW, Fersht AR (1993) Modification of the amino acid specificity of tyrosyl-tRNA synthetase by protein engineering. FEBS Lett 318: 167–171
Dieckmann R, Lee Y-O, van Liempt H, von Döhren H, Kleinkauf H (1995) Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. FEBS Lett 357: 212–216
Díez B, Gutiérrez S, Barredo JL, van Solingen P, van der Voort LHM, Martín JF (1990) The cluster of penicillin biosynthetic genes. Identification and characterization of the *pcb*AB gene encoding the α-aminoadipyl-cysteinyl-valine synthetase and linkage to the *pcb*C and *pen*DE genes. J Biol Chem 265: 16358–16365
Dittmann J, Wenger RM, Kleinkauf H, Lawen A (1994) Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor. J Biol Chem 269: 2841–2846
Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252: 675–679
Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive ests of sequence motifs. Nature 347: 203–206
Fan C, Moews PC, Shi Y, Walsh CT, Knox JR (1995) A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine:D-alanine ligase of *Escherichia coli*. Proc Natl Acad Sci USA 92: 1172–1176
Fernández-Moreno MA, Martínez E, Boto L, Hopwood DA, Malpartida F (1992) Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A32 including the polyketide synthase for the antibiotic actinorhodin. J Biol Chem 267: 19278–19290
Fulda M, Heinz E, Wolter FP (1994) The *fad*D gene of *Escherichia coli* K12 is located close to *rnd* at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family. Mol Gen Genet 242: 241–249
Fuma S, Fujishima Y, Corbell N, D'Souza C, Nakano MM, Zuber P, Yamane K (1993) Nucleotide sequence of 5′ portion of srfA that contains the region required for competence establishement in *Bacillus subtilis*. Nucleic Acids Res 21: 93–97
Galli G, Rodriguez F, Cosmina P, Pratesi C, Nogarotto R, de Ferra F, Grandi G (1994) Characterization of the surfactin synthetase multi-enzyme complex. Biochim Biophys Acta 1205: 19–28
Gerlt JA, Gassman PG (1993) Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement of phosphodiesters. Biochemistry 32: 11943–11952
Gocht M, Marahiel MA (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site directed mutagenesis. J Bacteriol 176: 2654–2662
Gutiérrez S, Díez B, Montenegro E, Martín JF (1991) Characterization of the *Cephalosporium acremonium pcb*AB gene encoding α-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the *pcb*C gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. J Bacteriol 173: 2354–2365

Haese A, Schubert M, Herrmann M, Zocher R (1993) Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. Mol Microbiol 7: 905–914

Hale RS, Jordan KN, Leadlay PF (1987) A small, discrete acyl carrier protein is involved in de novo fatty acid biosynthesis in *Streptomyces erythraeus*. FEBS Lett 224: 133–136

Hanks SH, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241: 42–52

Hawkins AR, Lamb HK (1995) The molecular biology of multidomain proteins. Eur J Biochem 232: 7–18

Higgins CF, Hiles ID, Salmond GPC, Gill DR, A Downie J, Evans IJ, Holland IB, Gray L, Buckel SD, Bell AW, Hermodson MA (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature 323: 448–450

Hoffmann K, Schneider-Scherzer E, Kleinkauf H, Zocher R (1994) Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. J Biol Chem 269: 12710–12714

Holak TA, Kearsley SK, Kim Y, Prestegard JH (1988) Three-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations. Biochemistry 27: 6135–6142

Holak T, Nilges M, Prestegard JH, Gronenborn AM, Clore GM (1988) Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic. resonance and the combined use of dynamical simulated annealing and distance geometry. Eur J Biochem 175: 9–15

Hori K, Yamamoto Y, Minetoki T, Kurotsu T, Kanda M, Miura S, Okamura K, Furuyama J, Saito Y (1989) Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. J Biochem (Tokyo) 106: 639–645

Huang C-S, Anderson ME, Meister A (1993) Amino acid sequence and function of the light subunit of rat kidney γ-glutamylcysteine synthetase. J Biol Chem 268: 20578–20583

Kanda M, Hori K, Kurotsu T, Miura S, Saito Y (1989) Reaction mechanism of gramicidin S synthetase 1, phenylalanine racemase, of *Bacillus brevis*. J Biochem (Tokyo) 105: 653–659

Keefe AD, Newton GL, Miller SL (1995) A possible prebiotic synthesis of pantetheine, a pecursor to coenzyme A. Nature 373: 683–685

Kleinkauf H, von Döhren H (1990a) Nonribosomal biosynthesis of peptide antibiotics. Eur J Biochem 192: 1–15

Kleinkauf H, von Döhren H (1990b) Bioactive peptide analogs: in vivo and in vitro production. Progr Drug Res 34: 287–317

Kleinkauf H, Roskoski jr R, Lipmann F (1971) Pantetheine-linked peptide intermediates in gramicidin S and tyrocidine biosynthesis. Proc Natl Acad Sci USA 68: 2069–2072

Krätzschmar J, Krause M, Marahiel MA (1989) Gramicidin S biosynthesis operon containing the structural genes *grs*A and *grs*B has an open reading frame encoding a protein homologous to fatty acid thioesterases. J Bacteriol 171: 5422–5429

Kuo TM, Ohlrogge JB (1984) The primary structure of spinach acylcarrier protein. Arch Biochem Biophys 234: 290–296

Kurahashi K (1974) Biosynthesis of small peptides. Annu Rev Biochem 43: 445–459

Kurotsu T, Hori K, Saito Y (1991) Characterization and location of the L-proline activating fragment from the multifunctional gramicidin S synthetase 2. J Biochem (Tokyo) 109: 763–769

Laland SG, Zimmer T-L (1973) The protein thiotemplate mechanism of synthesis for the peptide antibiotics produced by *Bacillus brevis*. Essays Biochem 9: 31–57

Landro JA, Gerlt JA, Kozarich KW, Koo CW, Shah VJ, Kenyon GL, Neidhart DJ, Fujita S, Petsko GA (1994) The role of lysine 166 in the mechanism of mandelate racemase

from Pseudomonas putida: mechanistic and crystallographic evidence for stereospecific alkylation by (R)-α-phenylglycidate. Biochemistry 33: 635–643
Lawen A, Zocher R (1990) Cyclosporin synthetase: the most complex synthesizing multienzyme polypeptide so far described. J Biol Chem 265: 11355–11360
Lee SG, Lipmann F (1974) Isolation of a peptidyl-pantetheine-protein from tyrocidine-synthesizing polyenzyme. Proc Natl Acad Sci USA 71: 607–611
Liaw S-H, Eisenberg D (1994) Structural model for the reaction mechanism of glutamine synthetase, based on five crystal structures of enzyme-substrate complexes. Biochemistry 33: 675–681
Lipmann F (1954) In: Mc Elroy WD, Glass B (eds) The mechanism of enzyme action. John Hopkins University Press, Baltimore, p 599
Lipmann F (1965) Projecting backward from the present stage of evolution of biosynthesis. In: Fox SW (ed) The origins of prebiological systems. Academic Press, New York, pp 259–273
Lipmann F (1973) Nonribosomal polypeptide synthesis on polyenzyme templates. Accts Chem Res 6: 361–367
MacCabe AP, van Liempt H, Palissa H, Unkles SE, Riach MBR, Pfeifer E, von Döhren H, Kinghorn J (1991) δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*. Molecular characterization of the acvA gene encoding the first enzyme of the penicillin biosynthetic pathway. J Biol Chem 266: 12646–12654
Marahiel MA (1992) Multidomain enzymes involved in peptide synthesis. FEBS Lett 307: 40–43
McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C (1993) Engineered biosynthesis of novel polyketides. Science 262: 1546–1550
McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. Nature 375: 549–554
Meister A (1988) Glutathione metabolism and its selective modification. J Biol Chem 263: 17205–17208
Menkhaus M, Ullrich C, Kluge B, Vater J, Vollenbroich D, Kamp RM (1993) Structural and functional organization of the surfactin synthetase multienzyme system. J Biol Chem 268: 7678–7684
Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117: 528–529
Miller SL, Schlesinger G (1993a) Prebiotic syntheses of vitamin coenzymes: I. Cysteamine and 2-mercaptoethanesulfonic acid (coenzyme M). J Mol Evol 36: 302–307
Miller SL, Schlesinger G (1993b) Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. J Mol Evol 36: 308–314
Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. Science 130: 245–251
Mirande M (1991) Aminoacyl-tRNA synthetase family from procaryotes and eucaryotes: structural domains and their implications. Prog Nucleic Acid Res 40: 95–142
Moras D (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. TIBS 17: 159–164
Neidhart DJ, Howell PL, Petsko GA, Powers VM, Li R, Kenyon GL, Gerlt JA (1991) Mechanism of the reaction catalyzed by mandelate racemase. 2. Crystal structure of mandelate racemase at 2.5-Å resolution: identification of the active site and possible catalytic residues. Biochemistry 30: 9264–9273
Osterberg R, Orgel LE (1972) Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. J Molec Evol 1: 241–248
Pavela-Vrancic M, Pfeifer E, Schröder W, von Döhren H, Kleinkauf H (1994a) Identification of the ATP binding site in tyrocidine synthetase 1 by selective modification with fluorescein 5′-isothiocyanate. J Biol Chem 269: 14962–14966
Pavela-Vrancic M, Pfeifer E, van Liempt H, Schäfer H-J, von Döhren H, Kleinkauf H (1994b) ATP binding in peptide synthetases: determination of contact sites of the

adenine moiety by photoaffinity labeling of tyrocidine synthetase 1 with 2-azidoadenosine triphosphate Biochemistry 33: 6276–6283
Pavela-Vrancic M, van Liempt H, Pfeifer E, Freist W, von Döhren H (1994c) Nucleotide binding by multienzyme peptide synthetases. Eur J Biochem 220: 535–542
Pfeifer E, Pavela-Vrancic M, von Döhren H, Kleinkauf H (1995) Characterization of tyrocidine synthetase 1 (TY1): requirement of post-translational modification for peptide biosynthesis. Biochemistry 34: 7450–7459
Pieper R, Haese A, Schröder W, Zocher R (1995) Arrangement of catalytic sites in the multifunctional enzyme enniatin synthetase. Eur J Biochem 230: 119–126
Poland BW, Silva MM, Serra MA, Cho Y, Kim KH, Harris EMS, Honzatko RB (1993) Crystal structure of adenylosuccinate synthetase from *Escherichia coli*. Evidence for convergent evolution of GTP-binding domains. J Biol Chem 268: 25334–25342
Portillo F, Serrano R (1988) Dissection of functional domains of the yeast proton-pumping ATPase by site directed mutagenesis. EMBO J 7: 1793–1798
Raibaud A, Zalacain M, Holt TG, Tizard R, Thompson CJ (1991) Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of *Streptomyces hygroscopicus*. J Bacteriol 173: 4454–4462
Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide-binding protein. Nature 250: 194–199
Rusnak F, Sakaitani M, Drueckhammer D, Reichert J, Walsh CT (1991) Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *ent*F gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. Biochemistry 30: 2916–2927
Safford R, Windust JHC, Lucas C, De Silva J, James CM, Hellyer A, Smith CG, Slabas AR, Hughes SG (1988) Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. Eur J Biochem 174: 287–295
Saito F, Hori K, Kanda M, Kurotsu T, Saito Y (1994) Entire nucleotide sequence for *Bacillus brevis* Nagano Grs2 gene encoding gramicidin S synthetase 2: a multifunctional peptide synthetase. J Biochem (Tokyo) 116: 357–367
Saito M, Hori K, Kurotsu T, Kanda M, Saito Y (1995) Three conserved glycine residues in valine activation of gramicidin S synthetase 2 from *Bacillus brevis*. J Biochem (Tokyo) 117: 276–282
Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop- a common motif in ATP- and GTP-binding proteins. TIBS 15: 430–434
Schimmel P (1987) Aminoacyl-tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. Annu Rev Biochem 56: 125–158
Schimmel P (1991) Classes of aminoacyl-tRNA synthetases and the establishment of the genetic code. Trends Biochem Sci 16: 1–3
Schimmel P, Schmidt E (1995) Making connections: RNA dependent amino acid recognition. Trends Biochem Sci 20: 1–2
Schlumbohm W, Stein T, Ullrich C, Vater J, Krause M, Marahiel MA, Kruft V, Wittmann-Liebold B (1991) An active site serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. J Biol Chem 266: 23135–23141
Schofield PR, Watson JM (1986) DNA Sequence of *Rhizobium trifolii* nodulation genes reveales a reiterated and potentially regulatory sequence preceding *nod*ABC and *nod*FE. Nucleic Acids Res 14: 2891–2903
Scott-Craig JS, Panaccione DG, Pocard J-A, Walton JD (1992) The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. J Biol Chem 267: 26044–26049
Shearman CA, Rossen L, Johnston AWB, Downie JA (1986) The *Rhizobium leguminosarum* nodulation gene *nod*F encodes a polypeptide similar to acyl carrier protein and is regulated by *nod*D plus a factor in pea root exudate. EMBO J 5: 647–652

Sherman DH, Malpratida F, Bibb J, Kieser HM, Bibb MJ, Hopwood DA (1989) Structure and deduced function of the granaticin producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. EMBO J 8: 2717–2725

Shiau CY, Baldwin JE, Byford MF, Sobey WJ, Schofield CJ (1995) δ-L-(α-Aminoadipoyl)-L-cysteinyl-D-valine synthetase: the order of peptide bond formation and timing of the epimerisation reaction. FEBS Lett 358: 97–100

Skarpeid H-J, Zimmer T-L, Shen B, von Döhren H (1990a) The proline-activating activity of the multienzyme gramicidin S synthetase 2 can be recovered on a 115-kDa tryptic fragment. Eur J Biochem 187: 627–633

Skarpeid H-J, Zimmer T-L, von Döhren H (1990b) On the domain structure of the multienzyme gramicidin S synthetase 2. Isolation of domains activating valine and leucine. Eur J Biochem 189: 517–522

Smith DJ, Burnham MKR, Bull JH, Hodgson JE, Ward JM, Browne P, Brown J, Barton B, Earl AJ, Turner G (1990) β-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. EMBO J 9: 741–747

Srere PA (1987) Complexes of sequential metabolic enzymes. Annu Rev Biochem 56: 89–124

Staab JF, Elkins MF, Earhart CF (1989) Nucleotide sequence of the *Escherichia coli ent*E gene. FEMS Microbiol Lett 59: 15–20

Stachelhaus T, Marahiel MA (1995a) Modular structure of genes encoding multifunctional peptide synthetases required for nonribosomal peptide biosynthesis. FEMS Microbiology Lett 125: 3–14

Stachelhaus T, Marahiel MA (1995b) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem 270: 6163–6169

Stachelhaus T, Schneider A, Marahiel MA (1995) Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science 269: 69–72

Stein T, Vater J, Kruft V, Wittmann-Liebold B, Franke P, Panico M, Dowell RM, Morris HR (1994) Detection of 4′-phosphopantetheine at the thioester binding site for L-valine of gramicidin S synthetase 2. FEBS Lett 340: 39–44

Stein T, Kluge B, Vater J, Franke P, Otto A, Wittmann-Liebold B (1995) Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4′-phosphopantetheine. Biochemistry 34: 4633–4642

Taylor W, Green NM (1989) The predicted secondary structures of the nucleotide-binding sites of six cation-transporting ATPases lead to a probable tertiary fold. Eur J Biochem 179: 241–248

Tokita K, Hori K, Kurotsu T, Kanda M, Saito Y (1993) Effect of single base substitutions at glycine-870 codon of gramicidin S synthetase 2 gene on proline activation. J Biochem (Tokyo) 114: 522–527

Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 222: 9–19

Turgay K, Krause M, Marahiel MA (1992) Four homologous domains in the primary structure of *Grs*B are related to domains in a superfamily of adenylate-forming enzymes. Mol Microbiol 6: 529–546

Ullrich C, Kluge B, Palacz Z, Vater J (1991) Cell-free biosynthesis of surfactin, a cyclic lipopeptide produced by *Bacillus subtilis*. Biochemistry 30: 6503–6508

Vanaman TC, Wakil SJ, Hill RL (1968) The complete amino acid sequence of the acyl carrier protein of *Escherichia coli*. J Biol Chem 243: 6420–6431

van Liempt H, Palissa H, Pfeifer E, Schwecke T, Weckermann R, von Döhren H, Kleinkauf H (1993) ACV-synthetase: implications of the amino-acid sequence data for the thiotemplate mechanism of peptide biosynthesis. In: Kleinkauf H, von Döhren H (eds) 50 years of penicillin application-history and trends. Publica, Prague, pp 136–144

Vater J (1990) Gramicidin S synthetase. In: Kleinkauf H, von Döhren H (eds) Biochemistry of peptide antibiotics. de Gruyter, Berlin, pp 33–55

Vollenbroich D, Kluge B, D'Souza C, Zuber P, Vater J (1993) Analysis of a mutant amino acid-activating domain of surfactin synthetase bearing a serine-to-alanine substitution at the site of carboxylthioester formation. FEBS Lett 325: 220–224

Vollenbroich D, Mehta N, Zuber P, Vater J, Kamp RM (1994) Analysis of surfactin synthetase subunits in srfA mutants of *Bacillus subtilis* OKB105. J Bacteriol 176: 395–400

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the α- and β-subunits of ATP-synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1: 945–951

Walsh CT (1989) Enzymes in the D-alanine branch of bacterial cell wall peptidoglycan assembly. J Biol Chem 264: 2393–2396

Walton JD (1987) Two enzymes involved in biosynthesis of the host-selective phytotoxin HC-toxin. Proc Natl Acad Sci USA 84: 8444–8447

Weber G, Schörgendorfer K, Schneider-Scherzer E, Leitner E (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8 kilobase open reading frame. Current Genetics 26: 120–125

Weckermann R, Fürbaß R, Marahiel MA (1988) Complete nucleotide sequence of the tycA gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. Nucleic Acids Res 16: 11841

Witkowski A, Naggert J, Mikkelsen J, Smith S (1987) Molecular cloning and sequencing of a cDNA encoding the acyl carrier protein and its flanking domains in the mammalian fatty acid synthetase. Eur J Biochem 165: 601–606

Wittinghofer A, Pai E (1991) The structure of Ras protein: a model for a universal switch. Trends Biochem Sci 16: 382–391

Wohlleben W, Alijah R, Dorendorf J, Hillemann D, Nussbaumer B, Pelzer S (1992) Identification and characterization of phosphinotricin-tripeptide biosynthetic genes in *Streptomyces viridochromogenes*. Gene 115: 127–132

Yoshida M, Amano T (1995) A common topology of proteins catalyzing ATP-triggered reactions. FEBS Lett 359: 1–5

Yuan Z, Liu W, Hammes GG (1988) Molecular cloning and sequencing of DNA complementary to chicken liver fatty acid synthase mRNA. Proc Natl Acad Sci USA 85: 6328–6331

Zuber P, Nakano MM, Marahiel MA (1993) Peptide antibiotics. In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, pp 897–916

Authors' address: Priv-Doz. Dr. J.Vater, Institut für Biochemie und Molekulare Biologie, Technische Universität Berlin, Franklinstrasse 29, D-10587 Berlin-Charlottenburg, Federal Republic of Germany.

Received November 1, 1995