Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis

Mohamed A. Marahiel,* Torsten Stachelhaus, and Henning D. Mootz

Biochemie/Fachbereich Chemie, Hans-Meerwein-Strasse, Philipps-Universität Marburg, 35032 Marburg, Germany

Received February 24, 1997 (Revised Manuscript Received June 16, 1997)

Contents

I. Introduction

A. Ribosomal and Nonribosomal Peptide Synthesis

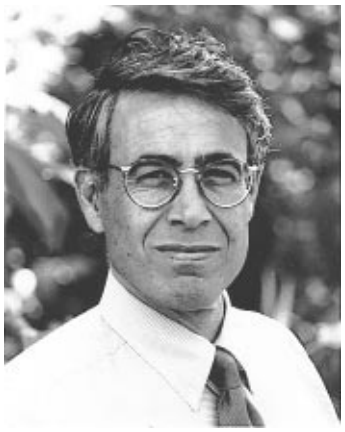
Significant progress has been made in the past four decades toward understanding the structures and synthesis of bioactive peptides produced by microorganisms through the ribosomal and the nonribosomal mechanisms.[1−8] Some peptides, such as the lantibiotics (which contain the thioether amino acid lanthionine) belong to a group of highly stable multicyclic peptide antibiotics that are of ribosomal origin. They are synthesized through proteolytic processing of gene-encoded precursors that have undergone several posttranslational modification events such as dehydration and addition of neighboring sulfhydryl groups to form thioethers.[3,4,9] Prototype peptides of this group are nisin, subtilin, and epidermin. They act primarily on Gram-positive bacteria and most serve as food preservatives. A vast array of other natural peptides with remarkable structural diversity (Figure 1) produced by microorganisms living in different habitats, spread from aquatic to terrestrial environments, are not gene encoded but are synthesized nonribosomally on large multifunctional enzymes called peptide synthetases.[1,2,5−8] The component moieties of these special metabolites are activated in the form of adenylate, acylphosphorylate, or coenzyme A derivatives, before they are linked together to form the final products. It is now accepted that this nonribosomal peptide synthetic route is an alternative means of manufacturing highly specialized polypeptides.

B. Template-Directed Peptide Synthesis

Some microorganisms contain multienzyme complexes that build specific protein templates for a nucleic acid-independent biosynthesis of low molecular weight peptides of diverse structures and broad spectrum of biological activities.[2] In this nonribosomal mechanism of peptide synthesis, compounds such as lipopeptides, depsipeptides, and peptidolactones are assembled from an exceedingly diverse group of precursors (to date more than 300 are known[10]) including pseudo, nonproteinogenic, hydroxy, *N*-methylated, and D-amino acids (Table 1). In contrast, the nucleic acid-dependent ribosomal synthesis of peptides and proteins is restricted to the incorporation of only 21 proteinogenic amino acids (including selenocysteine[11−13]). Nonribosomal protein template-directed synthesis of peptides is only limited by the length of the peptide chain formed, which has been found to range from 2 to 48 residues.[2,6] However, the peptide backbone of these short bioactive peptides can be composed of linear, cyclic, or cyclic branched structures that can be further modified by acylation, glycosylation, or heterocyclic ring formation (Table 2). These structurally diverse compounds are endowed with a broad spectrum of biological properties including antimicrobial, antiviral or antitumor activities.[2,9,14] Others express immunosuppressive or enzyme-inhibiting activity. Thus, members of this important class of peptide secondary metabolites have found widespread use in medicine, agriculture, and biological research. On the other hand, their physiological role in the metabolism of the source organisms has been the subject of considerable speculation.[14] These range from being signal mol-

* Author to whom correspondence should be addressed. Fax: 49-6421-28 21 91. E-mail: marahiel@ps1515.chemie.uni-marburg.de.

S0009-2665(96)00029-5 CCC: $28.00



Mohamed A. Marahiel was born 1949 in Gaza, Palestine. He studied chemistry at the Universities of Cairo and Göttingen, where he received his Ph.D. degree in Microbiology/Biochemistry with Professor Dr. Hans Künzel and Professor Dr. Fritz Cramer in 1977 at the Max-Planck-Institut für experimentelle Medizin, Göttingen, Germany. Subsequently he received a research assistant position at the Technical University of Berlin by Professor Dr. Horst Kleinkauf, where he was promoted 1985 to an assistant professor. In 1987 and 1986 he spent his sabbaticals at the John Innes Institute, Norwich, with Professor David Hopwood, and at the Harvard University with Professor Richard Losick, respectively. He obtained his "Habilitation" 1987 in Biochemistry at the institute of biochemistry and molecular biology of the Technical University Berlin. Three years later he moved to the Philipps University of Marburg as a professor of biochemistry at the chemistry department. His research focuses on the structure and function of genes encoding modular peptide synthetases involved in the nonribosomal synthesis of peptide antibiotics and on the rational design of recombinant enzymes for the synthesis of novel bioactive peptides. His group is also involved in studying the function and regulation of the major cold shock proteins in soil bacteria and other stress-induced proteins, specially those involved in protein folding and RNA binding.

Torsten Stachelhaus was born 1964 in Berlin, Germany, and has studied chemistry at the Technical University of Berlin. During his graduate thesis, he became acquainted with the field of nonribosomal peptide biosynthesis in the lab of Dr. Marahiel, were he received his Diploma in 1991. During his Ph.D. studies in the group of Dr. Marahiel, now at the Philipps University of Marburg, Germany, he worked on elucidation of the modular organization of multifunctional peptide synthetases and the targeted reprogramming of the biosynthetic template of the lipopeptide antibiotic surfactin. In 1994, he obtained his Ph.D. in biochemistry with a honor degree for the best thesis from the Chemistry Department. In his postdoctoral studies in the Marahiel lab, he has been and is presently focused on the biochemical and structural characterization of dissected domains from peptide synthetases, as well as on prospects for the engineering of novel peptide antibiotics. In Autumn 1997, he will move to Boston, to the Harvard Medical School, as an EMBO fellow, to work on the stereoselectivity of enzymatic reactions in the lab of Dr. Christopher T. Walsh. When not worrying about things like cloning and protein purification, Torsten likes to spend time playing tennis and squash as well as skiing and biking.

Henning D. Mootz was born in Braunschweig, Germany, in 1970. He has studied chemistry at the Universities of Wuppertal and Marburg, Germany, and at the University of Kent at Canterbury, England. Fascinated by the complexity of the chemical processes in living organisms he joined the biochemistry group of Professor Dr. M. A. Marahiel, where he obtained the Diploma in 1996. During his undergraduate studies he was a fellow of the Studienstiftung des Deutschen Volkes. In his Ph.D. he is working on the tyrocidine biosynthesis system and general aspects in peptide antibiotic assembly. He is a recipient of a Ph.D. fellowship of the Fonds der Chemischen Industrie. In his time off, he likes meeting friends, reading, and playing squash.

ecules for coordination of growth and differentiation in the producers (most of them are spore-forming soil inhabitants),[15−18] evolutionary relics or breakdown products of cellular metabolism,[19] to defense weapons that kill other competitor microorganisms.[9]

Although structurally diverse, most of these biologically active peptides share a common mode of synthesis, the multienzyme thiotemplate mechanism (Figure 2).[5−8,20,21] According to this model, peptide bond formation takes place on multienzymes designated peptide synthetases, on which amino acid substrates are first activated by ATP hydrolysis to the corresponding adenylate. This unstable intermediate is subsequently transferred to another site of the multienzyme where it is bound as a thioester to the cysteamine group of an enzyme-bound 4′-phosphopantetheinyl (4′-PP) cofactor.[7,22−24]

Recently, it has been shown that peptide synthetases, like fatty acid synthases and polyketide synthases, require posttranslational modification to become catalytically active.[24−26] The inactive apoproteins are converted to their active holoforms by posttranslational transfer of the 4′-PP moiety of coenzyme A to the side chain of a highly conserved serine residue located in peptide synthetases at the C-terminal region of each substrate activating unit, a region recently defined as the acylation or thiolation domain (see IV).

At this stage, the thiol-activated substrates can undergo modifications such as epimerization or *N*-methylation.[6,27,28] Thioesterified substrate amino acids are then integrated into the peptide product through a step-by-step elongation by a series of transpeptidation reactions.[6−8,23] These occur by transfer of the thioester-activated carboxyl group of one residue to the adjacent amino group of the next amino acid, thus effecting N to C stepwise assembly of the peptide product. During this condensation process all intermediates are covalently attached to the multienzyme complex. In conclusion, as shown

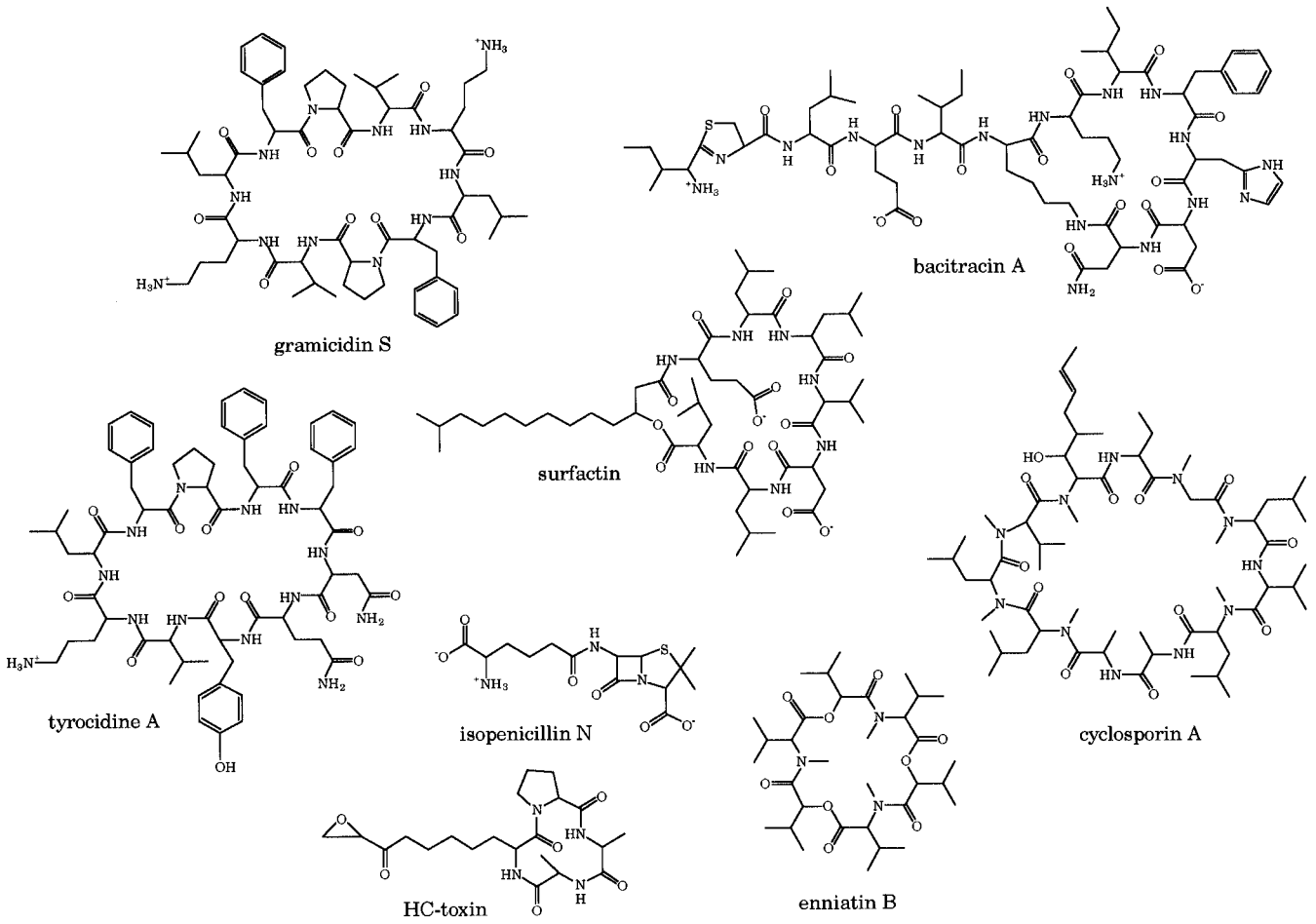


Figure 1. Chemical structure of some bacterial (gramicidin S, surfactin, bacitracin A, and tyrocidine A) and fungal (HC-toxin, enniatin, cyclosporin, and isopenicillin) peptide antibiotic whose peptide-bound backbones are synthesized by the nonribosomal thiotemplate mechanism. Genes encoding the involved peptide synthetases are shown in Figure 3.

in Figure 2, the 4′-PP cofactors facilitate the ordered transfer of the carboxy-activated thioester substrates between the active units that constitute the peptide synthetases, resulting in the formation of a peptide of defined sequence.

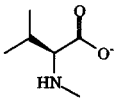
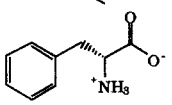
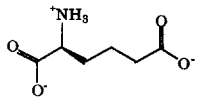
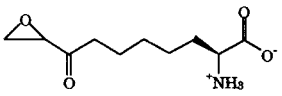
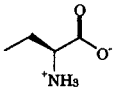
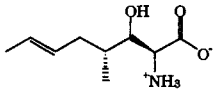
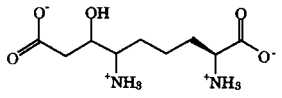
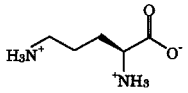
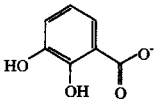
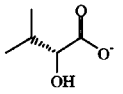
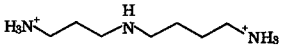
Protein chemical studies and the recent progress in cloning and sequencing of genes encoding peptide synthetases of bacterial and fungal origin provide valuable insights into the molecular architecture of these enzymes.[29−42] A modular structure for these multienzyme complexes has emerged, in which the substrate activating/modifying units are aligned in a sequence that is colinear with the amino acid sequence of the assembled peptide.[5−7,23] These units have been designated as modules according to a definition originally applied by L. Katz and coworkers to the arrangement of genes encoding type I polyketide synthases.[43] On the basis of comparison of DNA sequences encoding several peptide synthetases and recent studies on heterologous expression of DNA fragments[28,44−49] that encode proteins that activate individual amino acids, modules were defined as semiautonomous units within peptide synthetases that carry all information needed for recognition, activation, and modification of one substrate. This means that the number of modules and their order within a peptide synthetase define the sequence and the length of the synthesized peptide (Figure 2A, B). The modules, although proposed to act independently of each other, have to work in concert during peptide elongation. This template-based mode of action, in which different 4′-PP prosthetic groups (one for each module) are involved in peptide and depsipeptide bond formation, has been designated the multiple carrier thiotemplate mechanism. It is now an universally accepted model for nonribosomal peptide synthesis.[6−8,23]

II. Genes Encoding Modular Peptide Synthetases

A large number of bacterial operons and fungal genes encoding peptide synthetases have recently been cloned, sequenced, and partially characterized.[29−42,50] Different cloning strategies were used, including probing of expression libraries by antibodies raised against peptide synthetases, complementation of deficient mutants, and the use of designed oligonucleotides derived from amino acid sequences of peptide synthetase fragments.[44,51,52] Recently, utilization of the polymerase chain reaction (PCR) technology to amplify specific sequences from genomic DNA, by using degenerate oligonucleotides corresponding to highly conserved motifs in peptide synthetases (see section IV.A), established a convenient general approach for the identification and cloning of putative genes encoding these multienzymes.[51,52]

The complete DNA sequences of several bacterial operons, including *grs*,[32,53] *srfA*,[33,37] *tyc*,[41,54] and *bac*[42] for the biosynthesis of the cyclic peptide antibiotics

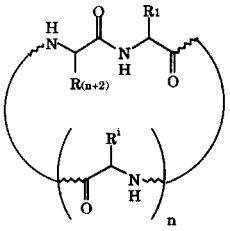
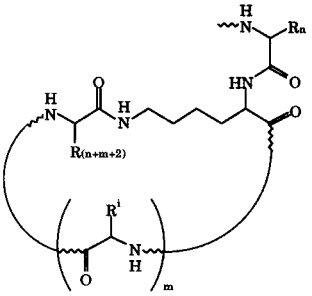
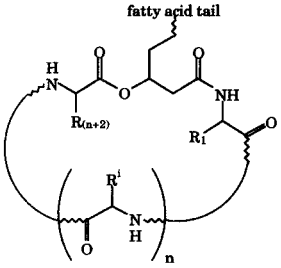
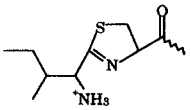
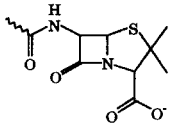
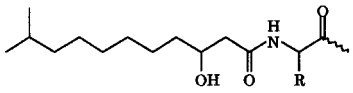
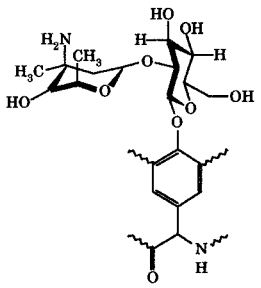
Table 1. Nonproteinogenic Constituents of Peptide Antibiotics (Examples for Some Unusual Moieties)

Name	Structure	Abbreviation	System(s)	Organism(s)
modified, proteinogenic amino acids				
N-methyl aa (e.g. *N*-methyl valine)		MeVal	cyclosporin enniatin	*Tolypocladium niveum* *Fusarium scirpi*
D-aa (e.g. D-phenylalanine)		D-Phe	e.g. bacitracin gramicidin S tyrocidine	*Bacillus licheniformis* *Bacillus brevis* *Bacillus brevis*
non-proteinogenic amino acids				
δ-(L-α-amino adipic acid)		Aad	ACV-tripeptide (precursor of penicillin and cephalosporin)	*Penicillium chrysogenum* *Aspergillus nidulans* *Streptomyces clavuligerus*
2-amino-9,10-epoxy-8-oxodecanoic acid		Aeo	HC-toxin	*Cochliobolus carbonum*
L-α-amino butyric acid		Abu	cyclosporin	*Tolypocladium niveum*
(4R)-4[(E)-2-butenyl-4-methyl-L-threonine]		Bmt	cyclosporin	*Tolypocladium niveum*
2,6-diamino-7-hydroxy-azealic acid		Dha	edeine	*Bacillus brevis*
ornithine		Orn	e.g. bacitracin gramicidin S tyrocidine	*Bacillus licheniformis* *Bacillus brevis* *Bacillus brevis*
carboxy acids				
2,3-dihydroxy benzoic acid		Dhb	enterobactin	*Escherichia coli*
D-α-hydroxyisovaleric acid		Hiv	enniatin B	*Fusarium scirpi*
amines				
spermidine		Sperm	edeine	*Bacillus brevis*

gramicidin S, surfactin, tyrocidine, and bacitracin, respectively, have been determined. These operons span regions of 18−45 kb (Figure 3) and encode several peptide synthetases comprising one to six modules, respectively. In the bacterial system the encoded multienzymes range in size between 126 kDa for one module enzymes (GrsA, TycA) to over 700 kDa for the six modules of tyrocidine synthetase 3 (TycC). A minimal module for substrate adenylation and thiolation contains two distinct domains:[24,32,49] the adenylation domain (Figure 3, red region, about 550 residues) and the thiolation domain (green region, about 100 residues). Several gene fragment-encoding adenylation domains of different modules have been amplified by PCR and expressed in heterologous systems. The overproduced proteins were shown to be active in substrate recognition and adenylation, but devoid of thiolation activity.[28,41,49,55] The thiolation domain of TycA, designated PCP for peptidyl carrier protein (Figure 3c, green region, see section IV.B), was also independently expressed and shown to be active in acylation reaction after posttranslational modification with the cofactor 4′-PP.[24,56]

In the fungal systems, exemplified by *acvA*,[29−31,57,58] *hts1*,[36] *esyn1*,[34] and *cssA*[35] genes, which encode the templates that direct the synthesis of the tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (LLD-ACV, the precursor molecule of isopenicillin N), HC-toxin of the maize pathogen *Cochliobolus carbonum*, the depsipeptide enniatin of *Fusarium scirpi*, and the immunosuppressive cyclic peptide cyclosporin A of *Tolypocladium niveum* (Figures 1 and 3e−h), respec-

Table 2. Structural Details of Peptide Antibiotics (Examples for Some Modifications at the Peptide Main Chain)

Name	Structure	System(s)	Organism(s)
ring formation			
cyclization		e.g. cyclosporin gramicidin S tyrocidine	*Tolypocladium niveum* *Bacillus brevis* *Bacillus brevis*
branching (amide bond)		bacitracin	*Bacillus licheniformis*
branching (ester bond)		surfactin	*Bacillus subtilis*
thiazoline ring formation (between Ile and Cys)		bacitracin	*Bacillus licheniformis*
b-lactam ring formation (between Cys and D-Val)		ACV-tripeptide (precursor of penicillin and cephalosporin)	*Penicillium chrysogenum* *Aspergillus nidulans* *Streptomyces clavuligeri*
additions (fatty acids and carbohydrates)			
acylation		surfactin	*Bacillus subtilis*
glycosylation		vancomycin	*Streptomyces orientalis*

tively, the peptide synthetases involved, are without exception, integrated multienzymes. All modules are aligned on a single polypeptide chain (Figure 3). They range in size from 350 kDa for a two module enzyme like Esyn1 to a molecular mass over 1600 kDa for the 11 modules containing cyclosporin A

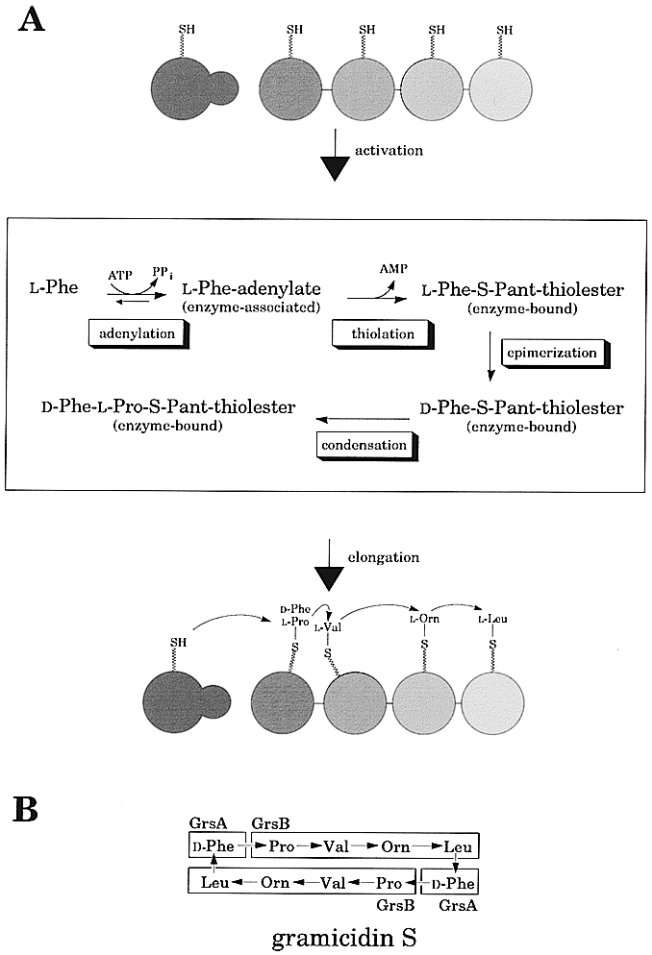


Figure 2. A simplified scheme displaying the principles of the thiotemplate-directed nonribosomal peptide synthesis. (A) For example, the synthesis of the cyclic decapeptide gramicidin S on the multifunctional enzymes GrsA (one amino acid-activating module plus epimerization domain) and GrsB (four amino acid-activating modules) is shown. Each module (symbolized by a circle) activates the cognate amino acid by ATP hydrolysis as amino acyl adenylate. This relatively instable intermediate is stabilized by thioesterification on the cofactor 4′-PP. Thioesterified substrates are then integrated into the growing peptide through a step-by-step condensation. The amino acid-activating modules are arranged in the order that corresponds to the amino acid sequence of the peptide. The arrows indicate the direction of polymerization. (B) Structure of the peptide antibiotic gramicidin S. The cyclic decapeptide was obtained by a head-to-tail condensation of two identical pentapeptides synthesized by the protein template described above.

synthetase. The *cssA* gene is over 45 kb in length, represents the largest gene known and encodes a single polypeptide chain of over 15000 amino acid residues.[35] In general the modular arrangement and the domain structure of fungal peptide synthetases is very similar to that of bacterial enzymes (Figure 3).

As mentioned above, a minimal module contains an adenylation and a thiolation domain and comprises about 650 amino acid residues. This minimal size is increased when additional functional domains, e.g., for epimerization (Figure 3, blue regions) or *N*-methylation (yellow regions) are integrated. While the epimerization domains (about 430 amino acid residues) in bacterial and fungal peptide synthetases were found to be contiguously integrated downstream of the thiolation domains,[29−31,33,37,41,42,53] the *N*-methylation domains (about 420 amino acid residues), which were only found in the fungal enniatin and cyclosporin A synthetases, are located between the adenylation and thiolation domains (Figure 3g,h).[34,35] The significance of this domain arrangement within the different modules and its influence on the synthesis of peptides are matters of speculation. Additional biochemical studies on dissected domains and the analysis of their activities and interactions *in vitro* may shed light on this multidomain arrangement.

Between the different modules constituting the bacterial and fungal templates for the generation of a defined peptide, one would expect specific intra- and/or intermolecular interactions. These interactions are not only needed for acyl and peptidyl transfer reactions but also for the correct channeling of the peptide product. Putative regions within modular peptide synthetases, designated condensation domains (Figure 3, white regions; see section IV.C) are believed to be the sites of such specific communication.[5−7,59] These domains are located upstream of most internal adenylation domains (red regions) and seem to be associated with the peptide elongation reaction since their occurrence corresponds to the number of peptide bonds in the derived peptide product. Moreover, these domains are absent from peptide synthetase modules that are involved in initiation reactions, such as the gramicidin S synthetase 1 (GrsA)[53] and the tyrocidine synthetase 1 (TycA).[54] In both of the latter cases, no putative condensation regions are present upstream of the adenylation domains (Figure 3a, c). Nothing is known about the exact mechanism of peptide elongation in the nonribosomal system, nor is it known how modules interact and how this interaction may affect the direction of polymerization.[59,60] However, all peptide intermediates remain covalently attached to the protein template during the elongation reaction.[2,5−7,21,60] Termination of nonribosomal peptide synthesis is either initiated by the action of thioesterases (see section IV.D), the transfer of the peptide chain to another functional group, or by cyclization.[2,6,21,61]

III. Genes Associated with Nonribosomal Peptide Synthesis

Other associated genes, whose products affect nonribosomal peptide synthesis, have been identified within bacterial operons encoding peptide synthetases. Those located inside the operons were found to encode proteins that show significant homology to fatty acid thioesterases of type II. The genes encoding these thioesterase-like proteins are either located at the 5′-end or the 3′-end of the biosynthetic operon (Figure 3a−c, light pink regions).[33,41,53] The encoded proteins (25−29 kDa) show over 30% identity and are believed to be important but not essential for the synthesis of the corresponding peptide (see section IV.D).[62]

Another class of genes (*gsp*, *sfp*, and *bli*, Figure 3a, b, and d, gray regions), associated with but not integrated in the bacterial operons of *grs*,[63] *srfA*,[64,65]

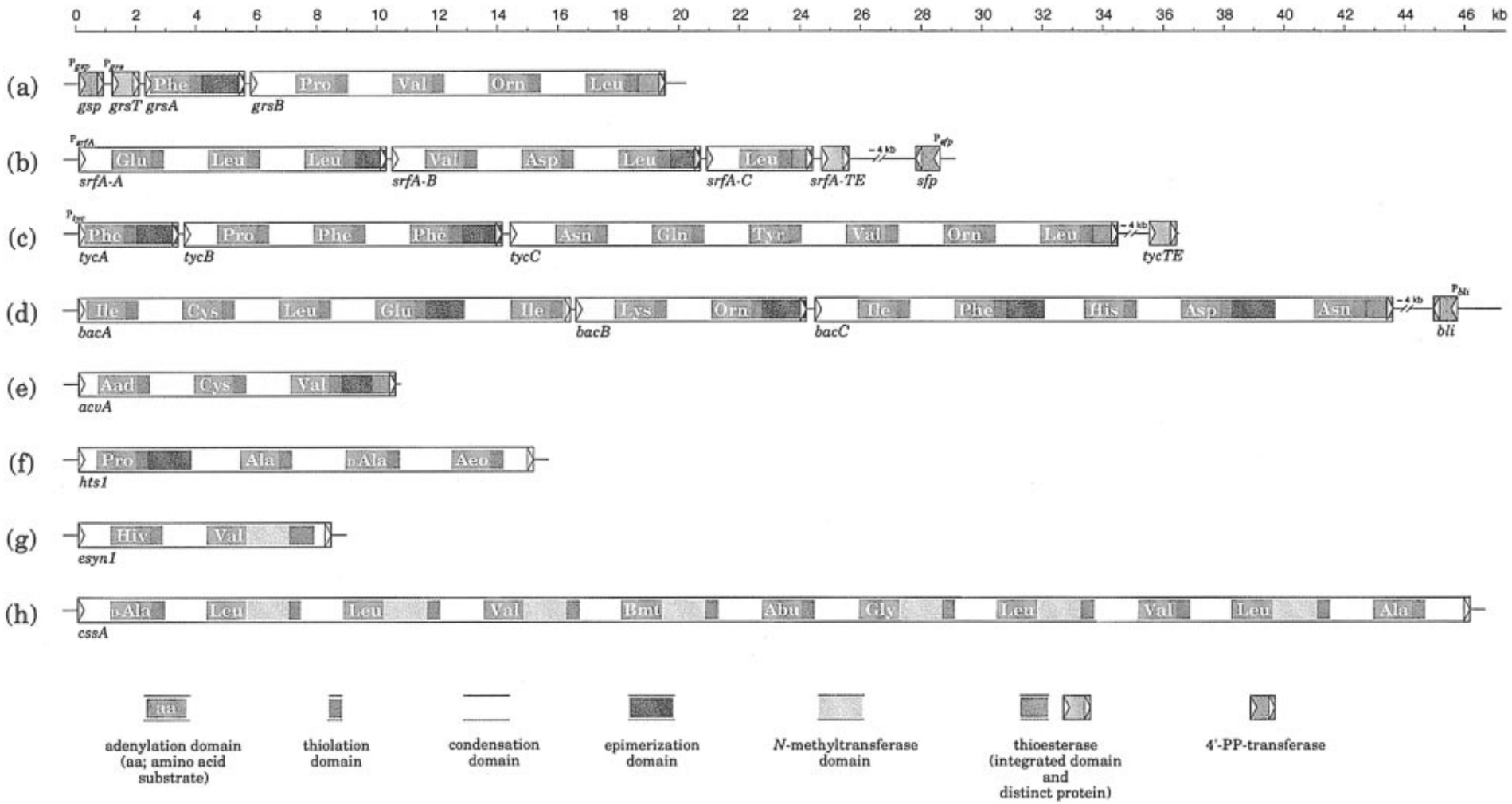


Figure 3. Schematic diagram showing the modular organization of peptide synthetases encoded by the bacterial operons *grs* (gramicidin S; row a), *srfA* (surfactin; row b), *tyc* (tyrocidine; row c) and *bac* (bacitracin; row d) as well as the fungal genes *acvA* (δ-(L-α-aminoadipyl)-L-Cys-D-Val; row e), *hts1* (HC-toxin; row f), *esyn1* (enniatin; row g) and *cssA* (cyclosporin A; row h). Red boxes indicate the amino acid adenylation domains, and green strips show the location of the thiolation domain, the site of cofactor 4′-PP-binding. Modules associated with the activation and incorporation of D-amino acids (GrsA, TycA, SrfA-A-Leu, SrfA-B-Leu, TycB-Phe, BacA-Glu, BacB-Orn, BacC-Phe and BacC-Asp, AcvA-Val, and Hts1-Pro) have an epimerization domain (blue regions) located downstream of the thiolation domain. The first module of the *cssA* gene (D-Ala) and the third one in the *hts1* gene (D-Ala) have no epimerization modules; however, it incorporates D isomers that are supplied by external racemases. A fourth type of domain (yellow regions) is only found in peptide synthetases of fungal origin (*esyn1* and *cssA*) that are associated with the activation and incorporation of *N*-methylated amino acids. These putative *N*-methyltransferases have a cosubstrate SAM-binding site and are inserted between the adenylation and the thiolation domains. The intermodular regions shown as white boxes represent putative elongation domains. Also shown is the location of gene-(fragment)s encoding the thioesterase-like proteins (pale pink) and thioesterase-like domains (pink) found associated with terminal modules in bacterial operons (*grs*, *srfA*, *tyc*, and *bac*). Genes that encode the 4′-PP transferase, located outside the operons *grs*, *srfA*, and *bac*, are also displayed (gray regions).

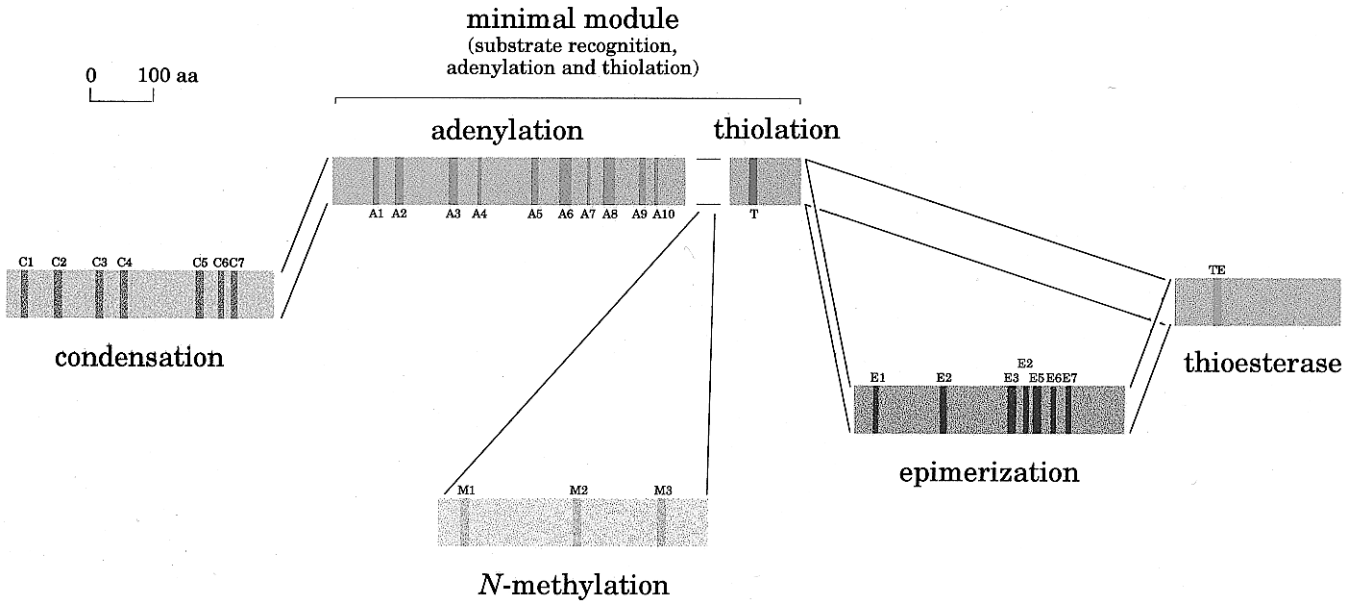


Figure 4. Schematic diagram showing the building up of a peptide synthetase module on functional domains. The particular composition of a module depends on the given requirements in regard of substrate activation, elongation and modification. (See Figure 3.) With exception of the adenylation (A domain, red) and the thiolation (T domain, green) domains, which were both biochemically characterized, all other domains, such as condensation (C domain, gray), *N*-methylation (M domain, yellow), epimerization (E domain, blue) and the thioesterase (TE domain, pink), were predicted from sequence alignments. Highly conserved signature sequences, which are also conserved in their relative locations, are shown. These core sequences are indicated in Table 3.

and *bac*[42,66] were found to be essential for nonribosomal peptide synthesis.[25] For example, disruption of the *sfp* gene, which is located about 4 kb downstream of the 3′-end of *srfA* operon, caused complete inhibition of the production of the lipopeptide antibiotic surfactin, although the expression of the peptide synthetases (SrfA-A/-B/-C) was not affected.[64,65] Surprisingly, the *gsp* gene of the *grs* operon, which encodes a 28 kDa protein that shows about 34% identity to Sfp, complements *in trans* the *sfp*-null mutation, indicating that Sfp and Gsp have similar functions in nonribosomal peptide synthesis.[63] Recently, it has been also shown that *bli*, located downstream of the bacitracin biosynthetic operon (Figure 3d) and homologous to Sfp and Gsp, restores surfactin production in the *sfp*-null mutant to a normal level.[42,66] These genetic studies clearly indicate that Gsp, Sfp, Bli, and the EntD protein of *Escherichia coli* (needed for synthesis of the siderophore enterobactin) are members of a new protein family that is associated with the synthesis of secondary metabolites. Recently, Lambalot and coworkers have shown that these proteins display 4′-PP transferase activity and are responsible for the posttranslational modification of the corresponding peptide synthetases.[25] These studies led to the discovery of a superfamily of such 4′-PP transferases (Table 5, see section IV.B) that are involved in the specific modification of 4′-PP-requiring enzymes, including fatty acid and polyketide synthases as well as several peptide synthetases from different species.

IV. The Functional Domains of Modular Peptide Synthetases

For a better understanding of the structure–function relationship of the building blocks of peptide synthetases, the modules, and how their functional domains are arranged, biochemical dissection studies and sequence alignments were undertaken. The structural features of putative domains and potentially important residues that might be involved in substrate specific adenylation (A domain), thiolation (T domain), epimerization (E domain), *N*-methylation (M domain), elongation/condensation (C domain), and release of the thioester bound peptide chain (TE domain) are detailed below (Figure 4 and Table 3).

A. Adenylation Domain

The adenylation domains (A domain) represent the central points of action in multifunctional peptide synthetases. For each incorporated amino acid in the peptide product a specific adenylation domain exists, whose location also dictates the primary structure of the peptide product (see above and Figures 1 and 3). Hence, investigations on peptide synthetases have notably focused on the A domain in recent years.

1. Activation Reaction

In order to incorporate an amino acid residue into a peptide through the protein template a two-step mechanism (Figure 2) for substrate activation is required.[2,5−7,20,21] First, the cognate amino acid is activated as aminoacyl-adenylate at the expense of Mg^{2+}-ATP (Figure 5). Second, the enzyme-attached thiol moiety 4′-phosphopantetheine (4′-PP) attacks the aminoacyl adenylate to yield the aminoacyl thioester and AMP as leaving group. The second step of the reaction requires the presence of the thiolation domain (T domain), which will be discussed below.

The way in which the amino acid residues are activated resembles that catalyzed by aminoacyl-tRNA synthetases in the ribosomal system of peptide synthesis.[5,6,21,67−69] There, the cognate amino acid is also activated as aminoacyl adenylate and then becomes esterified onto the 2′- or 3′-OH of the

Table 3. Highly Conserved Core Motifs of the Catalytic Domains of Peptide Synthetases

domain[a]	core(s)[b]	consensus sequence
adenylation	A1	L(TS)YxEL
	A2 (core 1)	LKAGxAYL(VL)P(LI)D
	A3 (core 2)	LAYxxYTSG(ST)TGxPKG
	A4	FDxS
	A5	NxYGPTE
	A6 (core 3)	GELxIxGxG(VL)ARGYL
	A7 (core 4)	Y(RK)TGDL
	A8 (core 5)	GRxDxQVKIRGxRIELGEIE
	A9	LPxYM(IV)P
	A10	NGK(VL)DR
thiolation	T (core 6)	DxFFxxLGG(HD)S(LI)
condensation	C1	SxAQxR(LM)(WY)xL
	C2	RHExLRTxF
	C3 (His)	MHHxISDG(WV)S
	C4	YxD(FY)AVW
	C5	(IV)GxFVNT(QL)(CA)xR
	C6	(HN)QD(YV)PFE
	C7	RDxSRNPL
thioesterase	TE	G(HY)SxG
epimerization	E1	PIQxWF
	E2 (His)	HHxISDG(WV)S
	E3 (race A)	DxLLxAxG
	E4 (race B)	EGHGRE
	E5 (race C)	RTVGWFTxxYP(YV)PFE
	E6	PxxGxGYG
	E7 (race D)	FNYLG(QR)
N-methylation	M1 (SAM)	VL(DE)GxGxG
	M2	NELSxYRYxAV
	M3	VExSxARQxGxLD

[a] See Figure 4. [b] Former nomenclature is given in brackets.

3′-nucleotide of the corresponding tRNA, which acts as the carrier of the activated amino acid. Despite these similarities shared by the ribosomal and nonribosomal system during amino acid activation, the enzymes involved have no similarity in primary and 3D structures.[68,70−73] Strikingly, there exist two classes of aminoacyl-tRNA synthetases, catalyzing the formation of aminoacyl adenylate, with fundamentally different folding topologies.[69,74] The recently solved crystal structure of an adenylation domain of a peptide synthetase reveals that Nature has invented a third fold for the same reaction.[72,73,75]

2. Adenylation Domains of Peptide Synthetases Are Members of a Superfamily of Adenylate-Forming Enzymes

The conserved region of the A domain was identified by comparing several genes encoding peptide synthetases.[32] The highly conserved A domains were found as repetitive blocks, the number of which coincides with the number of amino acids activated by the corresponding synthetase. These blocks, connected by regions now designated the condensation domains (see section IV.C),[59] represent what we call the minimal module, containing the A and the T domain. The A domain is about 550 amino acids in length.[49] It shares significant homology with the family of acyl-CoA synthases and luciferases, which are about the same size. Since all these enzymes catalyze an analogous reaction, the adenylation of their carboxy substrates, they constitute a superfamily of adenylate forming enzymes[32] A T domain connected to an A domain is exclusively found in peptide synthetases and is involved in the second part of the amino acid activation, the thiolation reaction (see section IV.B).[23] Taken together, these two domains exhibit a specific set of conserved motifs, the fingerprint of peptide synthetases, which has enabled the detection of previously unidentified genes encoding peptide synthetases by PCR.[51,52] Deletion studies on gramicidin S synthetase 1 (GrsA) defined the boundaries of the adenylation and thiolation activities associated with the A and T domains (Figure 4).[49]

3. The A Domains Are Enzymatically and Structurally Independent

While the A domain of peptide synthetases is an integrated part of a multifunctional enzyme, the homologous acyl-CoA synthases are distinct proteins.[32,72] Following the idea of the modular architecture of peptide synthetases, the question arose whether such a single A domain could function independently from the adjacent domains. The first insights were obtained through biochemical characterization of proteolytic and gene encoded fragments, expressed in *E. coli*, of the multimodular gramicidin S and tyrocidine synthetases, which exhibited activation of one specific amino acid residue only.[28,44−48,76−78] The catalytic independence of the integrated A domain itself was demonstrated for the first time by deletion studies on the starter synthetases GrsA and TycA (Figure 3).[49,55] The A domains, located in the N-terminal region, were proven to catalyze amino acid activation (Figure 5) with the same specificity as the wild-type enzymes. Very recently, it has also been shown that internal A domains of the multimodular enzymes TycB and TycC can be expressed in *E. coli* as soluble, functional proteins.[41] These findings reinforce the idea of multimodular peptide synthetases being an assembly of structurally and functionally independent domains on a polypeptide chain that act in concert with respect to their order on this giant template. By swapping single domains within the template, novel peptide products may be produced (see section VI).[79] The technique of investigating distinctly expressed internal domains opens up a way to decode the primary structure of a peptide product, for which the genes for the unknown template have been determined.[80,81]

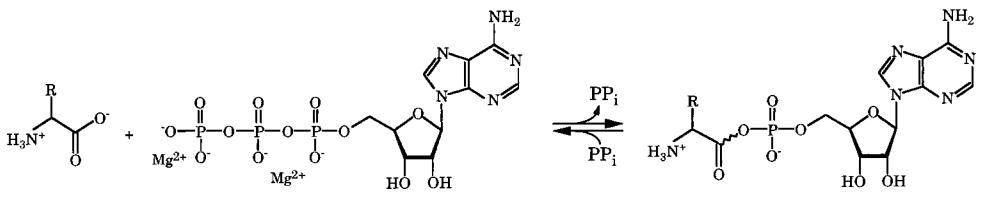


Figure 5. Amino acid adenylation in peptide synthesis

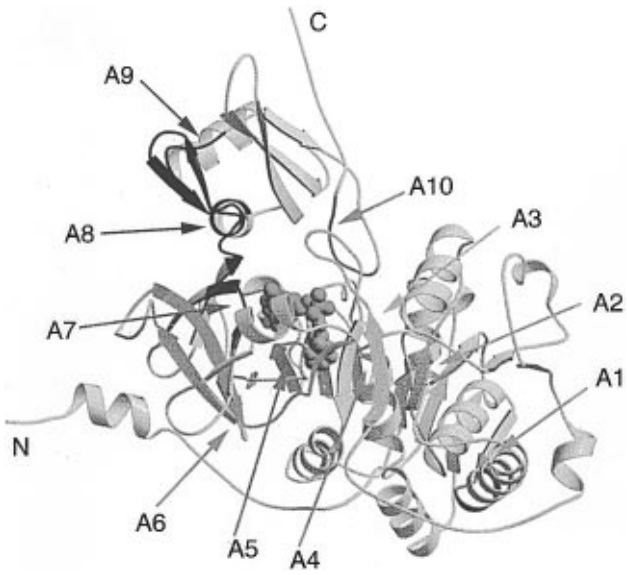


Figure 6. Ribbon diagram of the adenylation domain PheA showing the large N-terminal domain and the small C-terminal domain. The substrates, AMP (red) and phenylalanine (orange), are drawn using a space-filling representation. The locations of the highly conserved core motifs (A1−A10) in the superfamily of adenylate-forming enzymes within the PheA structure are indicated.

4. The Crystal Structure of the Phe-Activating Adenylation Domain of Gramicidin S Synthetase 1

Very recently, the first peptide synthetase fragment, the A domain of GrsA obtained through C-terminal deletion and expression in *E. coli* (designated PheA), has been crystallized and the 3D structure been solved at 1.9 Å (Figure 6).[49,75] The overall topology is highly similar to the structure of firefly luciferase,[72,73] although the two proteins are only 16% identical in their primary sequence. Thus, it can be assumed that the other members of the homologous superfamily of adenylate-forming enzymes, i.e., the A domains of peptide synthetases, have a very similar structure.

The A domain of GrsA is folded into two compact subdomains (to describe the crystal structure, the term domain means a stable tertiary fold), a large N-terminal and a smaller C-terminal portion, that are connected with a short hinge. Strikingly, the small C-terminal domain is rotated relatively to the N-terminal domain about 94° with respect to the structure of firefly luciferase. It remains unknown whether this rotation represents different stages of the catalytic mechanism, as the crystals of PheA contain the bound substrates L-phenylalanine, AMP, and Mg^{2+} in contrast to firefly luciferase, whose structure was determined without substrates.[73] The smaller C-terminal domain is indispensable for the activity of the protein, since its deletion results in a complete loss of activity.[82]

The core motifs of the A domain are the best conserved short amino acid sequences throughout the superfamily of adenylate-forming enzymes (see Figure 4 and Table 3).[32] They have been the subject of extensive investigations in recent years. Due to their ubiquitous presence they were thought to be involved in the common reactions, i.e., ATP binding, hydrolysis and adenylation of the carboxylate moiety of the substrate (Figure 5). Their location in the structure of PheA and their interactions with the substrates will now be discussed and compared with the available biochemical data (Figure 6).

Almost all core motifs are positioned around the active site where the substrates are bound (Figure 6, A1−10). Instead of the phenylalanine adenylate, the free amino acid and AMP are found in the structure: the adenylate has been hydrolyzed, and the pyrophosphate is missing. Most of the residues involved in substrate recognition are contributed by the larger N-terminal domain. However, a strictly conserved lysine residue (Lys517, Figure 6 and Table 3, A10) of the C-terminal domain is involved in key interactions (see below).

A signature sequence of the superfamily of adenylate-forming enzymes,[83] TSGTTGKPKG (motif A3, see Table 3), is mostly disordered in the structure. However, its orientation and distance to the AMP suggest an interaction with the pyrophosphate leaving group. The three G residues of the motif A3 of tyrocidine synthetase 1 (TycA) were mutated to A and the P to V without significant effect on adenylation activity in any mutant.[84] Introduction of a negative charge by replacing the first G to D in the second adenylation domain of gramicidin S synthetase 2 (GrsB) led to complete inactivation of the enzyme however.[85] Mutagenesis of the second K resulted in drastic reduction of activity, K to Q 61%,[86] K to R 90%, and K to T 99.5%,[84] whereas a K to Q mutant of the first lysine had no significant effect on the activities of the valine-activating domain of surfactin synthetase B (SrfA-B) and TycA, respectively.[85] The side chain of the second lysine is poorly ordered in the structure of PheA and projects into the solvent, while the first threonine of the motif A3 interacts with the α-phosphate.

The highly conserved core motif A7, Y(RK)TGDL (see Table 3 and Figure 6), which is observed in various ATPases, has also been investigated by means of mutagenesis.[84,87] A mutation of D to N reduced activity to 78% while a D to S substitution retained only 12% of wild-type activity.[84] In the structure of PheA the side chain of the aspartic residue, which is strictly invariant in all known members of the superfamily of adenylate-forming enzymes, interacts via hydrogen bonds with the oxygen atoms of the nucleotide ribose moiety.

Further contacts with the nucleotide base are formed by the main-chain carbonyl atom of A322 (numbering with respect to GrsA) and the side-chain carbonyl oxygen of N321 to the amino group of adenine, the latter N321 being well conserved and part of the core motif A5, NxYGPTE (Figure 6 and Table 3). A mutant strain of *Bacillus brevis*, deficient in gramicidin S production, was traced back to have a G to D substitution in this motif of the valine-activating domain GrsB-Val.[85] The main-chain carbonyl atom of this glycine points toward the α-amino group of the substrate phenylalanine. Other residues interacting with the amino group are I330, via the main chain carbonyl oxygen, and the strictly conserved D235, a part of core motif A4, FDxS (Figure 6). Consequently, D235 is only conserved in peptide synthetases that activate amino acids, in contrast to

luciferases and acyl-CoA synthases, the substrates of which do not have an α-amino group.[1,6,32]

The lysine residue of core motif A10, NGK (Figure 6), binds to the α-carboxylate group of the substrate phenylalanine as well as the ribose oxygens O-4′ and O-5′. The strictly invariant lysine is thus involved in two key polar interactions with both the adenosine and the amino acid, presumably fixing their position in the active site and clamping the C-terminal domain in a certain orientation. The key role of this lysine is also confirmed by a K to Q mutation in the valine-activating domain of surfactin synthetase, which caused a reduction in activity of >90%,[86,88] and by its specific labeling with fluorescein 5′-isothiocyanate.[88] In contrast to the other interactions discussed above, the NGK motif is located in the smaller C-terminal domain of PheA. Another well-conserved sequence in this folding domain, GRxxxQVKIRGx-RIELGEIE (motif A8, see Figure 6), was shown to be essential for adenylation. Mutation of the second G to various residues led to a loss of activity in the proline-activating domain of GrsB.[78] Additionally, labeling studies with fluorescein 5′-isothiocyanate and the photolabel 2-azido-ATP suggested the participation of at least a part of this motif in the adenylation reaction.[88−90] The first arginine of this motif is another possible candidate to interact with the pyrophosphate.

The core motifs A2, especially conserved in peptide synthetases, and A1 are probably only conserved for structural reasons, as they are far away from the active site. Motif A1 is part of a large helix, which significantly contributes to the fold of the N-terminal domain. The special functions of the core motifs A6 and A9, both of which were labeled with 2-azido-ATP and therefore thought to be involved in adenylation,[89] remain unclear. Nevertheless, they are also in proximity to the active site.

For further details about the location of conserved residues the reader is referred to the paper dealing with the structure of PheA.[75] It can be concluded that the significance of most of the core motifs has been confirmed by the structure. A rotation of the small C-terminal domain might be the key to understanding the course of the reaction, in particular with respect to the second part of the activation reaction, the formation of the thioester linkage. From the structure it is not clear in which direction the polypeptide chain will continue, i.e., the relative location of the following thiolation domain (see T domain). It is conceivable that highly conserved motifs that do not directly bind any of the substrates in PheA are important for interactions with the incoming 4′-phosphopantetheine arm.

5. Specificity of Peptide Synthetases and the Amino Acid Binding Pocket of PheA

Peptide synthetases are known to be of moderate substrate specificity compared to the aminoacyl-tRNA synthases.[91−93] In contrast to the high fidelity required of the latter,[68,69,94] there is no significant evolutionary pressure for accurate substrate recognition by peptide synthetases to be expected. A dependence of amino acid incorporation on amino acids added to the growth media can be observed in many cases, e.g., cyclosporin A, enniatin, surfactin, and tyrocidine synthesis.[91−93,95−100] Some adenylation domains exhibit a higher specificity than others,[41] a finding that may reflect positions of special importance in the peptide products with regard to their mode of action, or simply the relative difficulties in discriminating against other amino acids. In *in vitro* studies, when the respective amino acid concentrations can arbitrarily be chosen, peptide synthetases can be forced to synthesize an even wider range of products, as demonstrated in the case of cyclosporin A, enniatin, and other peptide antibiotics.

The residues forming the binding pocket for the amino acid substrate, and thus determining the specificity of an adenylation domain, have been of particular interest ever since the determination of peptide synthetases primary structures. Because of its relative inhomogenity, the region between core motifs A3 and A6 (see Figure 4) was thought to accommodate the binding pocket.[33] When sequences for this region were aligned, slight tendencies toward a clustering of A domains with the same specificity were observed.[32,33] Nevertheless, a stronger effect was the superposition of the origin of the respective domains, i.e., the organism from which they were taken.[32]

The residues interacting with the α-amino and the α-carboxylate group of phenylalanine in the structure of PheA have already been described above. As determined from the crystal structure, all residues involved in building the hydrophobic pocket are located between the core motifs A3 and A6 (data not shown). The pocket is closed at the bottom by the indole ring of W239, on one side by A236, I330, and C331, and on the opposite side by A322, A301, and T278. At one side of the pocket there is a water-filled channel that connects with the solvent. Considering the remarkably high homology in the 3D structure between firefly luciferase and PheA,[72,73,75] it can be expected that in other A domains of peptide synthetases (which show between 30−80% identity)[32] the binding pocket is formed from the equivalent residues. Examination of these residues in multiple alignments of A domains (data not shown) reveals a correlation between the polarity of either the substrate and the residues forming the pocket. The field is now open for attempts to alter substrate specificity and thereby the structure of the peptide antibiotic by means of site-directed mutagenesis (see section VI).

B. Thiolation Domain and Its Posttranslational Modification

The thiolation domain (T domain) of peptide synthetases, also called peptidyl carrier protein (PCP), is the site of 4′-PP cofactor binding and substrate acylation.[22−24,27,49,101] In analogy to the acyl carrier proteins (ACP) of modular fatty acid and polyketide synthases, the T domain of peptide synthetases is an integral part of these multi enzymes.[1,6,7,24,102−105] This functional unit of about 100 amino acid residues, to which aminoacyl substrates are bound as carboxy thioesters, is located in peptide synthetases directly downstream of the adenylation domains (A domains). An exception to this arrangement was found in the fungal modules activating *N*-methylated amino acids

Table 4. Enzyme Superfamily of Acyl/Peptidyl Carrier Proteins (ACPs/PCPs): Sequence Alignment around the Highly Conserved Cofactor 4′-PP Binding Site

Enzyme	Organism	Position (aa)	Sequence #
A) Peptide synthetases			
TycA	*Bacillus brevis*	553	D N F Y S L G G H **S** I Q A I Q V
GrsB	*Bacillus brevis*	2033	D N F F E L G G H **S** L R A M T M
SrfA-B	*Bacillus subtilis*	990	D N F F M I G G H **S** L K A M M M
AcvA	*Penicillium chrysogenum*	3049	D D L F K L G G D **S** I T S L H L
Hts1	*Cochliobolus carbonum*	2405	S D F F S S G G N **S** M A A I A L
CssA	*Tolypocladium niveum*	13645	D N F F E L G G H **S** L L A T K L
B) Acyl carrier proteins			
Polyketide synthases			
Act-ACP	*Streptomyces coelicolor*	33	L R F E D I G Y D **S** L A L M E T
Gra-ACP	*Streptomyces violaceoruber*	33	I T F E E L G Y D **S** L A L M E S
Fatty acid synthases			
FAS-ACP	*Escherichia coli*	28	S F V E D L G A D **S** L D T V E L
FAS-ACP	*Saccharomyces cervisiae*	73	Q F H K D L G L D **S** L D T V E L
		consensus	L G x(HD)S L

Sequence data are derived from: TycA [54], GrsB [32], SrfA-B [33]; AcvA [29], CssA [35], Hts [36], Act-ACP [104], Gra-ACP [105], FAS-ACP from *E. coli* [103], FAS-ACP from yeast [102].

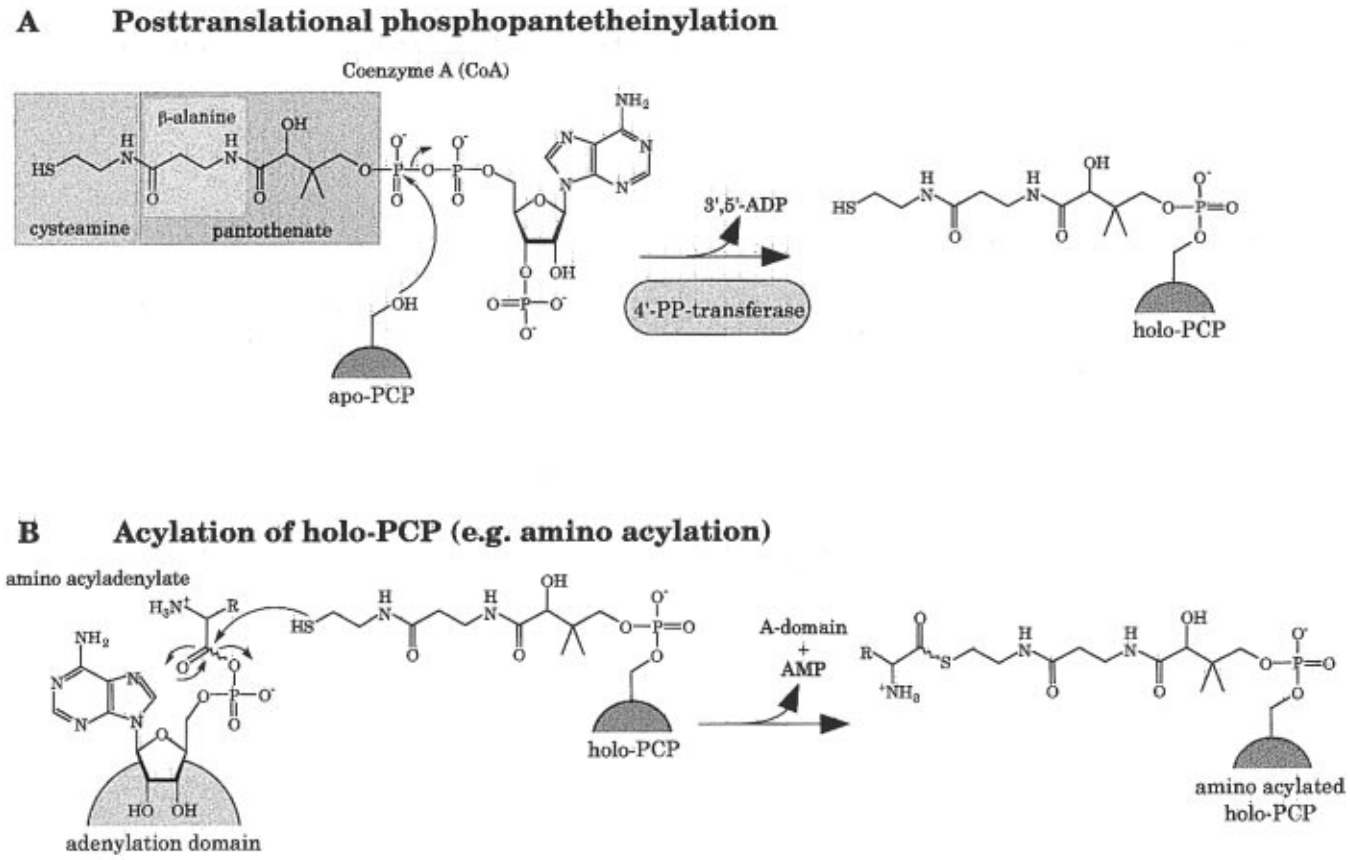


Figure 7. A scheme showing (A) the conversion of the thiolation domain (PCP) from apo to holo protein, through the action of a 4′-PP transferase, which directs the nucleophilic attack of the hydroxyl group of the highly conserved PCP-serine to the β-phosphate of CoA allowing the transfer of the 4′-PP moiety onto PCP. (B) Acylation of holo-PCP by an amino adenylated substrate attached to the adenylation domain. (C) Amide (and ester) bond formation between two amino acid residues (between an amino acid and a carboxy acid, respectively) that are activated as acyl-S-Pant thioesters on two adjacent thiolation (PCP) domains. Shown is the attack of nitrogen (or oxygen) nucleophiles to yield an amide (ester) bond in peptide (and depsipeptide) biosynthesis.

(see Figures 3 and 4), in which the A and T domains are separated by the *N*-methylation domain.[34,35] Although ACP and PCP proteins are functionally similar, they show only a limited degree of overall homologies except around the site of cofactor binding within the signature sequence LGx(HD)SL (Table 4).

In integrated peptide synthetases, the activated amino acyladenylate substrates on the A domain are transferred to the terminal cysteamine thiol group of the 4′-PP cofactor (Figures 2 and 7B), which is covalently attached to the side chain of the conserved serine residue within the signature sequence.[22−24,27,101] The essential role of this serine residue in cofactor binding and in linking the activated amino acid substrates as carboxyl thioester to the 4′-PP prosthetic group has been demonstrated by numerous investigations, including site-directed mutagenesis and affinity labeling studies.[22,23,27,84,101] Recently the thiolation domain of the peptide synthetase TycA has been biochemically characterized.[24] A region of about 100 amino acid residues surrounding the site of cofactor binding was overproduced in *E. coli* and partially posttranslationally modified from apo to the active holo form. The modification was assisted by a 4′-PP transferase (EntD, see below) that utilizes CoA and the T-domain as substrates.[25] It catalyzes the nucleophilic attack of the β-hydroxy side chain of the conserved serine on the pyrophosphate linkage of CoA, resulting in the transfer of the 4′-PP moiety onto the attacking serine (Figure 7A). This recombinant PCP protein fragment was active in amino acylation in the presence of an adenylation domain and radiolabeled cognate amino acid (Figure 7B). The detection of radiolabeled amino acid covalently attached to the nonintegrated, separately expressed PCP domain of TycA clearly indicated that PCP can be acylated *in vitro* by the A domain.[24] These results are a strong evidence for the functional integrity of these domains and for the multiple carrier model of nonribosomal synthesis (Figure 7C).

Further studies on posttranslational modification of PCP and ACP proteins *in vitro* using radiolabeled CoA led to the discovery of a superfamily of proteins that catalyze the conversion of apoproteins to their holo forms.[25] Among this group of 4′-PP transferases are the gene products encoded by *sfp*, *gsp*, and *bli*, which all are associated with bacterial operons encoding peptide synthetases. They utilize CoA as a common substrate, and appear to attain specificity through protein/protein interactions. For example, it has been shown that the *E. coli* apo ACP is the mutual substrate of the *E. coli* encoded ACPS (ACP synthase, a specific 4′-PP transferase for ACP) and not a substrate of the EntD protein, a second 4′-PP transferase present in *E. coli*, which was shown to be specific for the EntF protein involved in enterobactin synthesis.[25,56] The apo PCP protein (the T domain of TycA) was found to be a poor substrate for ACPS of *E. coli*, but an excellent substrate for Sfp protein, which is the 4′-PP transferase associated with the surfactin biosynthesis operon of *Bacillus subtilis*. These findings argue for the presence of 4′-PP transferases that show a specific protein partnership, when converting an apoprotein to its holo form. Therefore, one would expect that there are additional as yet unidentified 4′-PP transferases specific for each biosynthetic system. Moreover, since Sfp, Gsp, and other transferases associated with template-directed synthesis are not essential proteins for survival of the host, one has to predict the presence of homologous transferases that are specific for the modification of the essential ACP proteins of fatty acid synthesis and other proteins that require the prosthetic 4′-PP group.[25,56] Through refined sequence comparisons, which indicated low level similarity with the primary structure of ACPS, Lambalot and co-workers recognized two conserved sequence motifs shared among a group of enzymes, whose genes were previously shown to be associated with peptide antibiotic production, anabolic pathways (e.g., fatty acid synthesis) and cellular division (Table 5).[25] The overall similarity of these proposed 4′-PP transferases, e.g. Sfp, Gsp, and Bli, with ACPS is only about 12−22%, whereas those 4′-PP transferases associated with bacterial peptide antibiotic production (Sfp, Gsp and Bli) show more than 30% identity.[25,63−66,106]

Table 5. The 4′-Phosophopantetheinyl Transferases

protein[a]	pathway	organism(s)	size (aa)
	antibiotic production		
Gsp	gramicidinS	*B. brevis*	273
Bli	bacitracin	*B. licheniformis*	225
EntD	enterobactin	*E. coli*	209
		S. typhimurium	232
		S. austin	232
		S. flexneri	209
Lpa-14	iturin	*B subtilis*	224
Psrf-1	surfactin	*B. pumilis*	233
Sfp	surfactin	*B. subtilis*	224
	anabolic pathways		
ACPS	fatty acids	*E. coli*	126
FAS2	fatty acids	*S. cerevisiae*	1894
		C albicans	1885
		P. patulum	1857
		S pombe	1842
		A. nidulans	1559
HI0152	fatty acids	*H. influenzae*	235
LYS5	lysine	*S. cerevisiae*	272
	cellular division		
HetI	differentiation	*Anabaena* sp.	237
		Synechchocystis sp.	246

[a] Sequences are available from the GenBank, SwissProt, or EMBL databases.

C. Condensation Domain

In contrast to the A and T domains catalyzing amino acid activation and thioesterification, virtually no biochemical data are available to date about the part of modular peptide synthetases referred to as the condensation domain (C domain, see Figure 4 and Table 3). Its actual function therefore remains putative. Nevertheless, the accumulating sequence information of different peptide synthetase systems suggests that the C domain is responsible for the condensation of two amino acids activated on adjacent modules, i.e., catalyzes elongation of the growing peptide chain.[6,23,59]

The C domains are inserted between each consecutive pair of activating units (which may include additional tailoring domains like epimerization and *N*-methylation) within the polypeptide chains of

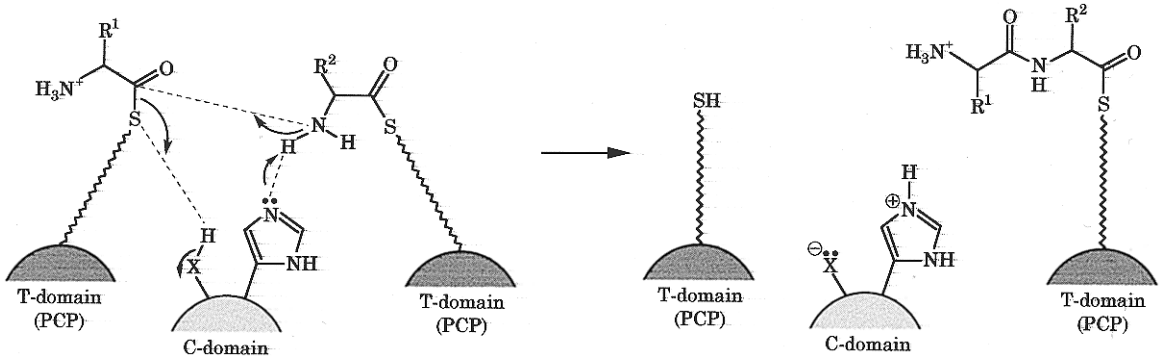


Figure 8. Suggested mechanism for the condensation/elongation reaction in peptide synthesis. Two amino acid residues attached as thioesters to adjacent thiolation domains (T-domain or PCP) via the cofactor 4′-PP and the second histidine residue conserved within motif C3 (HHxxxDG, see Table 6) of the condensation domain (C-domain) are shown. A nucleophilic attack of the incoming amino group on the thioester activated carboxyl group of the preceding amino acid is proposed.

peptide synthetases. This setup corresponds to the basic chemical requirements for the sequential linkage of activated amino acids to yield a linear peptide. Consequently, the number of C domains found in bacterial peptide synthetase systems coincides with the number of peptide bonds of the linear intermediate (see Figure 3). Since the functions of the A and T domains in amino acid activation have largely been elucidated,[24,49] the remaining C domains are the ideal candidates to catalyze peptide bond formation. It is unlikely that the A and T domains could also take charge of the elongation reaction, because separate, equivalent proteins in other systems (acyl-CoA-synthases, ACPs) are known to catalyze only the reactions shown to be attributed to the A and T domains, respectively. Moreover, as no other essential chemical reaction except amino acid activation and peptide bond formation is required to build up a linear peptide chain, no basically different catalytic function would be conceivable for the ubiquitous C domain, considering that a role as a simple spacer between functional domains is unlikely due to its size (Figures 3 and 4).

The condensation domain is about 450 amino acids in length. Database searches of this region have not revealed any related enzymes that might have had a common ancestor with a similar enzymatic activity. The occurrence of this domain seems to be restricted to the superfamily of peptide synthetases. Within this group, the C domains show moderate homology to each other. Their distribution in multifunctional peptide synthetases seems to follow two simple rules (Figure 3): (I) A C domain is always present between two adjacent activating units located on the same polypeptide (intramolecular amino acid transfer; e.g., GrsB, TycB, and TycC, SrfA-A and SrfA-B, AcvA, Hts, and CssA).[30−36,41,42] (II) When the two consecutive A domains are not located on the same enzyme and thus peptide bond formation has to be achieved between amino acids activated on two synthetases, the C domain is found at the N-terminus of the amino acid-accepting synthetase (intermolecular amino acid transfer; e.g., TycA to TycB, SrfA-A to SrfA-B, BacA to BacB).[32,33,41,42]

The C domains located at the N-terminus of accepting synthetases are less conserved than the internal ones, and the core sequences given in Table 3 are better conserved for internal C domains.[59] Therefore, it is intriguing to speculate that the N-terminal C domains are also necessary for the accurate recognition of the preceding synthetase (protein/protein interaction); their observed sequence variations may be an indication of such a specialized recognition reaction.

Recently, de Crécy-Lagard and co-workers have pointed out a possible relationship of the C domain to chloramphenicol acetyltransferases (CAT) and dihydrolipoyl transacetylases (E2p), a part of the pyruvate dehydrogenase multienzyme complex (and other 2-oxo acid dehydrogenase complexes).[59] These enzymes catalyze the transfer of an acetyl group, activated as acetyl-CoA, onto a hydroxy moiety of chloramphenicol and an acetyl group bound as a thioester on dihydrolipoamide onto CoA, respectively. These reactions resemble the fate of amino acyl or peptidyl intermediates transferred from their thioester linkage on one 4′-phosphopantetheinyl moiety of peptide synthetases to the 4′-phosphopantetheinyl group of the following module (Figure 8). Although the C domains show no overall homology to CAT and E2p, the best conserved core motif C3 (HHxxxDG, see Tables 3 and 6) is a common feature. The crystal structures of CAT and the catalytic domain E2p have been solved and exhibit virtually identical topology.[113,114] In both cases, the second histidine of the HHxxxDG motif (the first is not conserved in E2p) is thought to act as the general base promoting nucleophilic attack of the hydroxy moiety of chloramphenicol and of the thiol group of CoA on the carbonyl carbon atom of the acetyl thioester.[113,114] The histidine is found in both structures to have an unusual conformation with regard to its dihedral angles. This conformation allows a hydrogen bond between the imidazole nitrogen N1 and the carbonyl oxygen of the same amino acid, which has been related to its suggested properties as a base.[113,114] The conservation of the HHxxxDG motif and the similar reaction catalyzed in peptide bond formation may suggest an analogous function of the second histidine in nonribosomal peptide synthesis (Figure 8, see also the chapter on the epimerization domain in which the HHxxxDG motif is also found and a base is needed for the chemical reaction).[59] Studies on the pH optimum of nonribosomal peptide synthesis suggested the possibility of a catalytic histidine residue.[26] However, although the role of histidine seems to be

Table 6. Comparison of the Highly Conserved Core Motif C3 Found in the Putative Condensation Domains as Well as Dihydrolipoyl Transacetylases and Chloramphenicol Acetyltransferases

Enzyme	Organism	Position (aa)	Sequence #
A) Peptide synthetases			
GrsB	*Bacillus brevis*	143	W S F **H H** I L M **D G** W C F N I I
SrfA-A	*Bacillus subtilis*	1183	F D M **H H** I I S **D G** V S I G I M
CssA	*Tolypocladium niveum*	1239	I V M **H H** I V S **D G** W S L D L L
EntF	*Escherichia coli*	134	Q R Y **H H** L L V **D G** F S F P A I
SnbC	*Streptomyces pristinaespiralis*	1165	L L M **H H** V A G **D G** W S L R P L
PvdD	*Pseudomonas aeruginosa*	170	L V Q **H H** I V S **D G** W S M Q V M
RapP	*Streptomyces hygroscopicus*	164	L T V **H H** I A G **D G** W S L A I L
B) Dihydrolipoyl transacetylases			
Component E2	*Bacillus subtilis*	391	L S L D **H** R V L **D G** L V C G R F
Component E2	*Homo sapiens*	448	W S A D **H** R V I **D G** A T M S R F
C) Chloramphenicol acetyltransferases			
CAT	*Escherichia coli*	185	V Q V **H H** A V C **D G** F H A A R F
CAT	*Clostridium perfringens*	182	I Q V **H H** A V C **D G** F H I C R F
		consensus	H x x x D G

Sequence data are derived from: GrsB [32], SrfA-A [33], CssA [35], EntF [107], SnbC [38], PvdD [39], RapP [108], E2 from from *B. subtilis* [109], E2 from *Homo sapiens* [110], Cat from *E. coli* [111] and CAT from *C. perfringens* [112].

consistent, its specific assignment to the process of elongation requires further evaluation. A mutation in which the Asp in the C3 motif (Table 3) was converted to Ala within the Val module of surfactin synthetase 2 (SrfA-B) resulted in loss of product formation, underlining its crucial role in nonribosomal peptide synthesis.[115]

The presence of a waiting position for the incoming amino acid within the C domain has been postulated as an alternative to the direct transfer from one 4′-PP cofactor to the next.[38] The mechanism for the first peptide bond formation in pristinamycin I, actinomycin D, and enterobactin is in obvious conflict with the model outlined in Figure 2, since in these cases the first module (SnbA, AcmS I, and EntE) lacks the thiolation domain (the binding site for 4′-phosphopantetheine).[38,116−118] The concept put forward was extended as the general elongation process in peptide synthesis, following the corresponding mechanism in polyketide and fatty acid synthesis.

More biochemical studies, ideally focused on a single elongation event, will be required to understand the general principles of peptide bond formation catalyzed by peptide synthetases and the as yet undiscovered and still putative role of the condensation domain within this process.

D. Thioesterase as Integrated Domain and Distinct Protein

A region of about 250 amino acid residues located to the C-terminal end of bacterial modules that are involved in adding the last amino acid to the linear peptides (Figure 3 and 4, pink color in following modules: ACV-Val, GrsB-Leu, SrfA-C-Leu, TycC-Leu, and BacC-Asn) exhibits homology to thioesterases.[29,31−33,41,42,119,120] This region is referred to as the thioesterase domain (TE domain). It has been found in the same location in the bacterial operons encoding multifunctional enzymes for the synthesis of the ACV tripeptide,[29,31,119,120] bacitracin,[42] enterobactin,[107] gramicidin S,[32] pyoverdine,[39] surfactin,[33] and tyrocidine[41] which are of bacterial and fungal origin. Due to its location, it is tempting to speculate that the TE domain is involved in hydrolytic cleavage of the linear peptide products, i.e., termination of nonribosomal peptide biosynthesis.

Such a spatial arrangement of TE domains in modular polyketide (e.g., erythromycin systems)[121] and fatty acid synthases (integrated systems type I)[122,123] has been shown to be responsible for product release. However, things seem to be more complicated in the polypeptide systems: the TE domain is present in systems producing linear (ACV), branched via ester bond (surfactin), branched through amide bond (bacitracin), and cyclic peptides (gramicidin S, tyrocidine). Strictly speaking, a thioesterase function would only be required in the case of linear products. The cleavage of the thioester linkage of the peptide chain attached to the 4′-PP cofactor of the last module should be achievable by intramolecular attack of a side chain to build branched products (bacitracin), or of the amino group of the first amino acid incorporated for manufacturing cyclic products (gramicidin S, tyrocidine). The apparent explanation of the sequence data would be to postulate a cleaved linear intermediate in all cases, which could then be cyclized in a special fashion or not. However, no such intermediate has ever been described for peptide synthetases. Perhaps other, as yet unidentified, proteins are responsible for the final shape of the product. Alternatively, the TE domain could serve another function during product formation. In this respect, it is noteworthy that thioesterases and acyltransferases share a similar catalytic center (Table 7, signature sequence GxSxG). Thus the TE domain might actually be an acyltransferase domain. Cyclization or branching could then be the result of an intramolecular acyl transfer of the linear peptide chain.

Eukaryotic peptide synthetases of which the primary structure is known (Hts1,[36] Esyn1,[34] and CssA;[35] see Figure 3) lack the TE domain. It is striking that all their products are cyclic, i.e., cleavage of the enzyme might be achieved by intramolecular attack of the linear peptide intermediate.

In analogy to the catalytic triad of thioesterases, a conserved aspartatic residue is present within the TE domain, whereas a conserved histidine can only be found when large gaps in the amino acid alignments are allowed (not shown). An *in-frame* deletion of the TE domain of the surfactin synthetase 3 (SrfA-C) resulted in blocking of surfactin synthesis.[33] G. Turner and co-workers have recently mutated the conserved serine residue of the signature sequence (GxSxG) to alanine and also deleted the entire TE domain of ACV-synthetase of *Penicillium chrysoge-*

Table 7. Putative Thioesterases and Acyltransferases: Sequence Alignment around the Highly Conserved Core Motif TE

	Enzyme	Organism	Position (aa)	Sequence
A) Integrated thioesterases				
Peptide synthetases (putative)	GrsB	*Bacillus brevis*	4265	V L I **G** Y **S** S **G** G N L
	SrfA-C	*Bacillus subtilis*	1114	T L F **G** Y **S** A **G** C S L
	AcvA	*Penicillium chrysogenum*	3626	H F I **G** W **S** F **G** G I L
	EntF	*Escherichia coli*	1133	Y L L **G** Y **S** L **G** G T L
	PvdD	*Pseudomonas aeruginosa*	2448	N L A **G** W **S** L **G** G N L
Polyketide synthases	Ery3	*Saccharopolyspora erythraea*	3034	V V A **G** H **S** A **G** A L M
Fatty acid synthases	FAS	*Rattus norvegicus*	2297	R V A **G** Y **S** F **G** A C V
	FAS	*Homo sapiens*	2297	R V A **G** Y **S** Y **G** A C V
B) Distinct thioesterases				
Peptide synthetases (putative)	GrsT	*Bacillus brevis*	90	A F L **G** H **S** M **G** A L I
	SrfA-TE	*Bacillus subtilis*	81	V L F **G** H **S** M **G** G M I
Fatty acid synthases	TE mc	*Rattus norvegicus*	96	A F F **G** H **S** F **G** S Y I
	TE mc	*Anas platyrhynchos*	85	A L F **G** H **S** F **G** S F V
C) Acyltransferases				
Polyketide synthases	Act1	*Streptomyces coelicolor*	342	S M V **G** H **S** L **G** A I G
	Rap3	*Streptomyces hygroscopicus*	616	A V I **G** H **S** V **G** E L A
Fatty acid synthases	FAS	*Saccharomyces cerevisiae*	269	G A T **G** H **S** Q **G** L V T
	FAS	*Candida albicans*	256	W S T **G** H **S** Q **G** L V T
			consensus	G x S x G

[#] Sequence data are derived from: GrsB [32], SrfA-C [33], AcvA [29], EntF [107], PvdD [39], Ery3 [124], rat-FAS [125], human-FAS [126], GrsT [53], SrfA-TE [33], rat-medium chain(mc)-TE [127], chicken-mc-TE [128], Act1 [105], RapP [108], yeast-FAS [129] and FAS from Candida albicans [130].

num to analyze their role in nonribosomal peptide synthesis.[120] The drastic reduction of product formation observed in both cases underlines the importance of the TE domain.

Distinct genes encoding thioesterases have been detected within almost all bacterial peptide synthetase coding operons (Figure 3). The gene products are about 220−340 amino acid residues in length and show clear homology to thioesterases involved in fatty acid biosynthesis in mammalian cells (see Figure 3 and Table 7 for the GxSxG signature motif, also found in the TE domain).[33,41,53] As is the case for the integrated thioesterases (TE domain), the actual function of the operon associated thioesterases GrsT, SrfA-TE, and Tyc-TE (see Figure 3; light pink) remains unknown. There is some evidence that these thioesterases copurify with peptide synthetases. GrsT, the thioesterase of the gramicidin S biosynthesis operon, stimulates gramicidin S production *in vitro* to a certain extent. However it has an inhibitory effect at higher concentrations.[1,131] A knockout mutant of Srf-TE of the surfactin biosynthesis operon results in a 6-fold reduction of surfactin production.[62] Therefore, it can be speculated that these enzymes liberate mischarged peptide synthetases, which are blocked by an unspecific thioesterification of their 4′-PP cofactor.

V. Modifying Domains

In addition to the incorporation of a wide variety of amino and hydroxy acids for which no ribosome-recognizing amino acyl-tRNAs in nature exist, peptide synthetases can also carry out numerous modifications including *N*-acylations of β-hydroxy fatty acids, *N*-methylations, and site-specific epimerizations (Figure 1 and Table 1).[1,2,6,7] While *N*-acylations depend on the action of a nonintegrated acyltransferase, the particular domains of peptide synthetases catalyzing substrate epimerization and *N*-methylation are marked by signature sequence motifs neighboring the adenylation and thiolation domains (Figures 3 and 4; Table 3). These modifying domains in peptide synthetases dramatically increase the versatility and biological activity of nonribosomally synthesized peptides.[2] However, although a great deal of work has been done to elucidate the mode of substrate activation (adenylation and thiolation), the enzymatic reactions of substrate alteration are not completely understood.

A. Epimerization Domain

As concluded from initial work on GrsA and TycA, the phenylalanine racemases of *Bacillus brevis* (Figure 3),[22,27,49,84,132] substrate epimerization has been shown to occur at the thioester stage, with the amino acyl-S-Pant enzyme-bound substrate. It was found that initiation of D-Phe-L-Pro dipeptide formation takes place exclusively with D-Phe, and therefore the L-Phe substrate should be epimerized prior to condensation. By contrast, all attempts to detect D-Val as an intermediate in biosynthesis of the tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine failed.[60,108] Similar results were obtained while investigating the epimerization reaction of D-Val during the biosynthesis of actinomycin D.[116,133] Therefore, it has been assumed that in the latter two cases racemization takes place at the peptidyl rather than the amino acyl stage. A third example for introduction of D-amino acids using a protein template has been found in the fungal peptides cyclosporin A and HC-toxin (Figures 1 and 3). In both peptides D-Ala residues were found to be incorporated by amino acid-activating modules that are devoid of an epimerase domain;[35,36] substrates are provided in the D configuration, which is brought about through the action of nonintegrated racemases.[134]

With the exception of the *bac* operon that encodes three bacitracin synthetases,[42] all epimerization domains of bacterial operons encoding peptide synthetases were found to be localized at the C-terminal end of the corresponding peptide synthetase (Figure 3).[33,53,54] The recent sequence data from the *bac* operon of *Bacillus licheniformis*[42] unveiled the first bacterial examples of internal epimerase domains

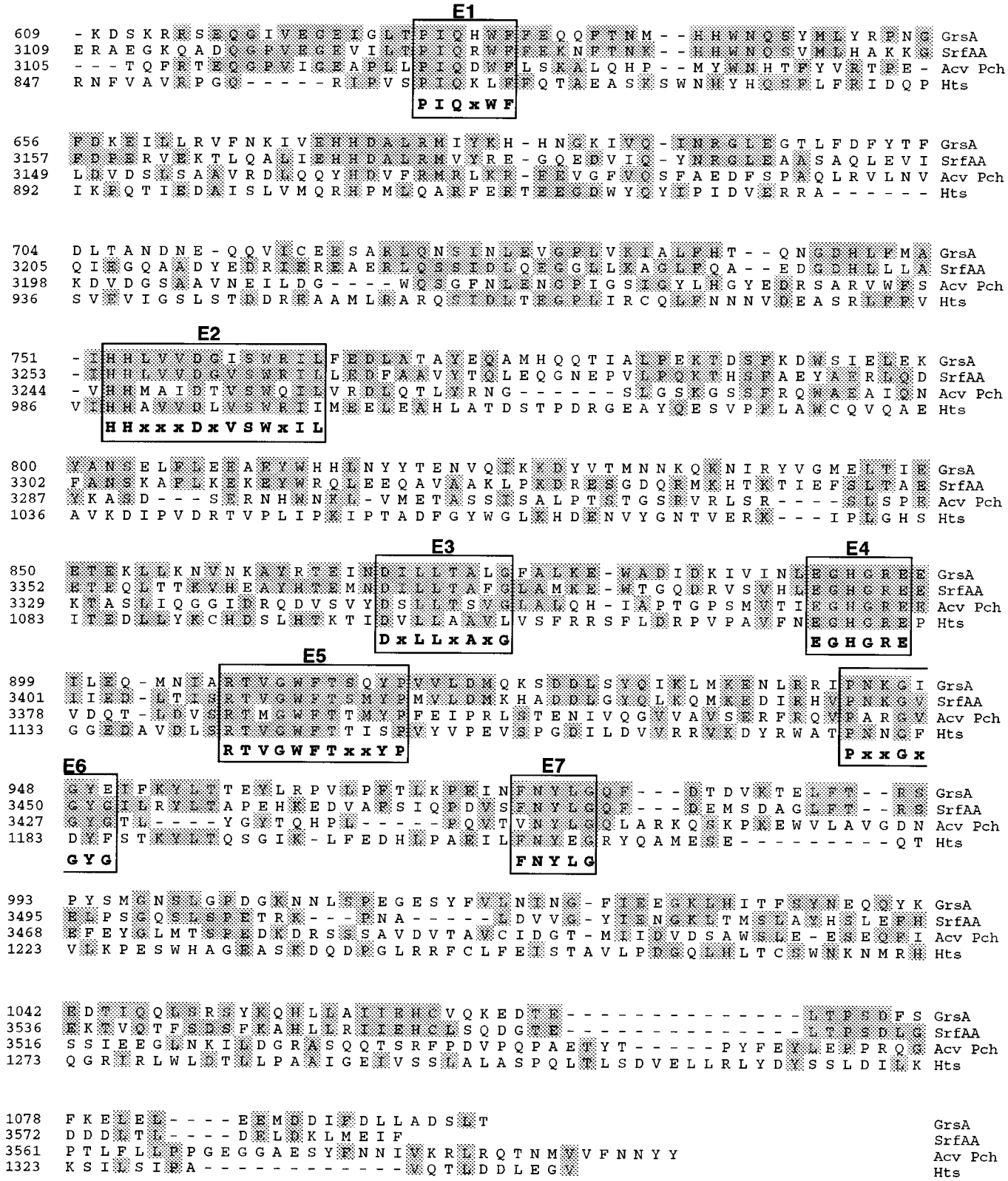


Figure 9. Amino acid sequence alignment of epimerase domains (E domains) from gramicidin synthetase 1 (GrsA),[53] surfactin synthetase A (SrfA-A),[34,38] ACV-synthetase[30] of *Penicillium chrysogenum* and HC-toxin synthetase.[37] The locations of the conserved core sequences E1 to E7 are marked by the boxes.

within the bacitracin synthetase 1 (BacA-Glu) and bacitracin synthetase 3 (BacC-Phe and BacC-Asp). On the basis of the substrates utilized and the location, one could describe four structurally homologous types of epimerase domains that are present within modular peptide synthetases (see Figure 3): (1) C-terminal located amino acyl-epimerases (e.g., GrsA[53]); (2) C-terminal located peptidyl epimerases (e.g., ACV synthetase[29−31]); (3) internal amino acyl epimerases (e.g., Hts1[36]); and (4) internal peptidyl epimerases (e.g., BacA and BacC[42]).

Apart from the above-mentioned classification, sequence comparisons of epimerization domains of peptide synthetases revealed no significant homologies either to known amino acid epimerases or to *N*-acyl racemases (Figure 9 and Table 3). Therefore,

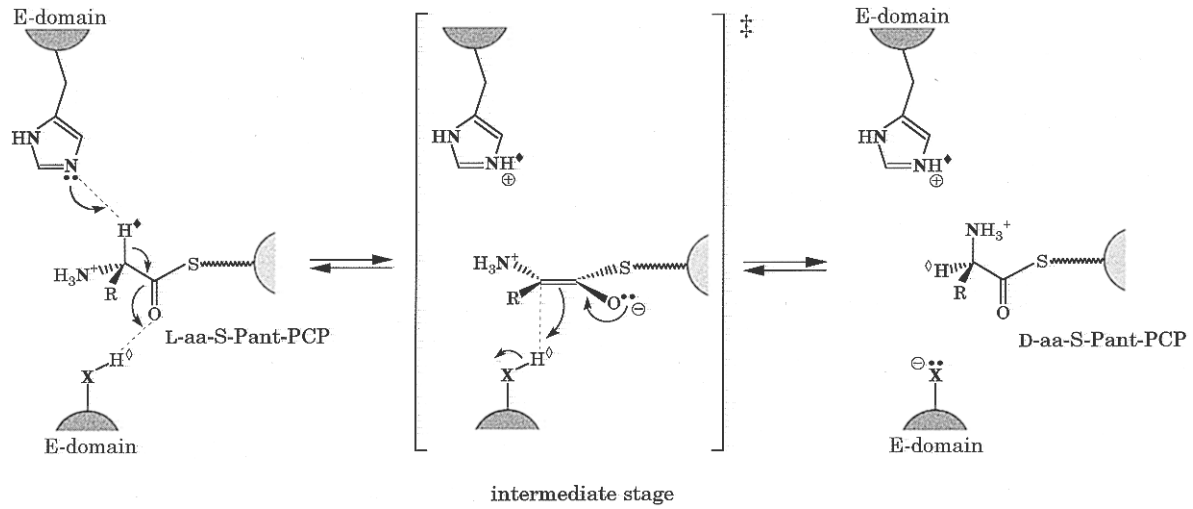


Figure 10. Proposed two-base mechanism of substrate epimerization during nonribosomal peptide biosynthesis. Carboxy thioester-activated amino acid attached to the thiolation domain (PCP), the location of the second histidine residue of motif E2 (see Figure 9), and the conjugate acid of a second enzymic base (indicated by XH) are shown.

one may infer that epimerase domains of peptide synthetases may represent a novel class (Figure 8) distinct from the well-known pyridoxal phosphate (PLP)-dependent racemases (e.g., alanine racemases, essential for providing D-Ala for the bacterial cell wall biosynthesis)[135−137] and the PLP-independent racemases (e.g., glutamate and proline racemases).[138−141] The latter two classes racemize free amino acids exclusively, rather than amino acyl or peptidyl-S-Pant enzyme-bound substrates.

Sequence analysis and alignment studies of several epimerization domains highlight at least seven signature sequence motifs within a region of about 450 amino acid residues (Table 3 and Figure 9, E1-E7). According to a suggested reaction mechanism for epimerization, in which one of these core sequences (E2) may be involved, de Crécy-Lagard and coworkers have implicated motif E2 (HHxxxDxVSW) as a signature sequence for a superfamily of enzymes involved in acyl transfer and epimerization (Figure 10).[59] This group of enzymes may share a similar catalytic mechanism based on the acid/base properties of the second histidine residue in E2.[59,113,114] In fact, this motif is also conserved within the proposed condensation domain (see condensation domain and Table 3, C3 motif), whose action requires a nucleophilic attack of the incoming acyl N-terminus on the activated carbonyl of the preceding amino acyl thioester (Figure 8). In analogy, epimerization involves a proton abstraction and readdition of the Cα proton of the amino acyl or peptidyl moiety linked to the cofactor 4′-PP (Figure 10).[59] The observed dependency of template-directed peptide synthesis on pH indicates the possible involvement of a histidine residue.[26] Although this seems to be consistent with the proposed mechanism, it remains to be confirmed if a histidine residue would be required for such a racemization reaction.

B. *N*-Methyltransferase Domain

N-Methylation is another modification of nonribosomally synthesized peptides that significantly contributes to their biological activity and to peptide bond stabilization against proteolytic cleavage (Table 1). Recently, sequencing of the entire fungal genes of cyclosporin A synthetase (*cssA*)[35] and enniatin synthetase (*esyn1*)[34] confirmed that the *N*-methyltransferase activity is associated with integral parts of the respective multifunctional peptide synthetases (Figures 3 and 4). The sequence data revealed a novel type of module possessing an insertion of about 420 amino acids (M domain) between the A and T domains. The occurrence of these insertions within the amino acid activating modules coincides with the number of *N*-methylated residues in the corresponding peptide product (e.g., seven *N*-methylated residues are present in cyclosporin A and one in enniatin; Figures 1 and 3). The insertion contains at least three signature core motifs (M1−M3; Figure 4 and Table 3), including a glycine-rich sequence M1 (VL(ED)xGxGxG), that exhibits significant similarity to the common *S*-adenosylmethionine (SAM) binding site of a heterologous class of cosubstrate-dependent methyltransferases (Table 8).

Some characteristics of *N*-methyltransferase domains have been analyzed by overproducing functional fragments of enniatin synthetase (Figure 3g).[28,96,150] Analysis of the recombinant proteins revealed that L-methylvaline activation can be assigned to a C-terminal 158 kDa fragment of the second module. This protein encompasses the A, M, and T domains and can be affinity labeled with [^{14}C]SAM, verifying the presence of the methyltransferase domain. Further N- and C-terminal deletions led to a 65 kDa protein forming the 420 amino acid insertion mentioned above.[28,150] UV-induced photoaffinity labeling of this deletion mutant indicated the localization of the methyltransferase activity in this region.[151] These studies also revealed that *N*-methylation occurs at the thioester stage prior to peptide bond formation.[28,96,150,151]

Cosubstrate dependence of the methylation reaction in general and SAM charging of the methylation domain in particular have been confirmed by the use of potent inhibitors like sinefungin and *S*-adenosylhomocysteine.[151] Sinefungin acts as a competitive inhibitor and totally prevents photolabeling with [^{14}C]SAM, while the noncompetitive inhibitor S-

Table 8. SAM-Dependent *N*-Methyltransferases: Sequence Alignment around the Highly Conserved Cofactor-Binding Motif M1

Enzyme	Organism	Position (aa)	Sequence [#]
A) Peptide synthetases			
Esyn	*Fusarium scirpi*	2083	R D **V L E** I **G** T **G** S **G** M I L F N L
CssA	*Tolypocladium niveum*	2100	G H **V L E** I **G** T **G** T **G** M V L F N L
		3592	G H **V L E** V **G** T **G** T **G** M V L F N L
		5083	G R **V L E** V **G** T **G** T **G** M I M F N L
		6577	G H **V L E** I **G** T **G** T **G** M V M F N L
		9133	G H I **L E** I **G** A **G** T **G** M I L S N L
		10625	R P C **A E** I **G** T **G** T **G** M V L F N L
		13190	G K **V L E** I **G** T **G** T **G** M V L F N L
B) diverse SAM-dependent Methyltransferases			
DNA (Adenine, *Hin*dIII)	*Haemophilus influenzae*	255	Q I **V L D** P F A **G** S **G** T T L L A A
DNA (Adenine, *Ban*III)	*Bacillus aneurinolyticus*	47	I R G **L D** P S C **G** D **G** E L L L S L
DNA (Cytosine, *Bam*HI)	*Bacillus amyloliquefaciens*	320	D V **V L D** P F G **G** S **G** T T F A V S
rRNA (Adenine)	*Bacillus anthracis*	46	D T **V L E** L **G** A **G** K **G** A L T T V L
tRNA (Uracil)	*Escherichia coli*	212	G D L **L E** L Y C **G** N **G** N F S L A L
EryG	*Saccharopolyspora erythraea*	83	D E **V L D** V **G** F **G** L **G** A Q D F F W
RapM	*Streptomyces hygroscopicus*	104	R T **V L E** V **G** C **G** M **G** E G L N F L
		consensus	V L(DE) x G x G x G

[#] Sequence data are derived from: Esyn [34], CssA [35], DNA-MTR from *H. influenzae* [142], DNA-MTR from *B. aneurinolyticus* [143], DNA-MTR from *B. amyloliquefaciens* [144], rRNA-MTR from *B. anthracis* [145], rRNA-MTR from *E. coli* [146], EryG [124] and RapM [147-149].

adenosylhomocysteine only lowers the apparent affinity for the cosubstrate, causing a reduced SAM charging even at high concentrations of the inhibitor. In any case, a dramatically reduced synthesis of a non-methylated peptide product could be observed indicating the SAM dependence of *N*-methylation during nonribosomal synthesis of modified peptide antibiotics.

VI. Prospects for the Construction of Hybrid Antibiotics

We have discussed the modular organization of multifunctional peptide synthetases, the large enzyme complexes representing the protein templates for the biosynthesis of defined peptide products, and have shown that they are assembled from multifunctional building blocks (domains).[1,2,5,6] Localization and enzymatic properties of these disparate building blocks were originally postulated on the basis of sequence comparison with enzymes having known functions or have been revealed by biochemical means. Accordingly, dissection of particular modules and biochemical analysis of the separated domains shed light on the molecular bricks used for assembling the conveyor belt (template) required for the biosynthesis of a defined peptide product.[24,28,49,55,150] It has been established that the nonribosomal synthesis of a bioactive peptide is brought about by such a protein template that contains the appropriate number and correct order of activating units.[6,7,23] These advances will (and already have) enable the development of techniques for the rational design of bioactive peptides[79] and for exploring the potential of protein templates in combinatorial synthesis for the generation of structural diversity.[24]

As a first attempt, we have recently described the reprogramming of a given protein template (Figure 11a).[79] A programmed alteration within the primary structure of a peptide antibiotic could be accomplished by the substitution of an amino acid-activating module at the genetic level (Figure 11a). According to this two-step recombination method, the chromosomal target site of a desired biosynthesis gene has been marked through a specific double cross-over event with a selectable marker. Subsequently, the disrupted gene was reconstituted by a replacement plasmid that delivered an engineered hybrid gene into the marked chromosome through a second marker exchange reaction. The introduced hybrid gene encodes a peptide synthetase with an altered substrate specificity that targets amino acid substitution into the corresponding position of the peptide product.

Initially, this recombination method was set up for reprogramming the surfactin synthetase 3, which integrates L-leucine at position 7 in the cyclic lipopeptide antibiotic surfactin (SrfA-C; Figure 1 and 3). An integration vector was constructed that contains the flanking region of the leucine-activating minimal module, strictly speaking the coding fragments of the N-terminal condensation domain and the C-terminal thioesterase (TE-domain) domain. An *in-frame* integration of coding regions of various A-T modules of bacterial and fungal origin between the linkers led to the construction of hybrid genes, encoding heterologous SrfA-C derivatives (C-[A-T]-TE) with altered substrate specificities, defined by the incoming adenylation domains. After delivering the hybrid gene(s) into the marked chromosome by homologous recombination, the surfactin derivatives produced by the various *B. subtilis* strains were extracted from the cultured broth and analyzed by infrared spectroscopy as well as mass spectrometry. These studies clearly confirmed the identity of the novel, engineered lipopeptides derived by targeted domain replacement. In order to investigate the influence of amino acid substitutions on surfactin hemolytic activity, the derivatives of surfactin were investigated for their ability to lyse erythrocytes. It was found that disrupting the operon resulted in a complete loss of activity, whereas the hybrid biosurfactants restored that activity. Until now, numerous domain swaps (see Figure 3) have been accomplished,[7,152] indicating that the stage of rational design for bioactive peptides is no longer an illusion.

Although the method described above represents a general comprehensive approach for specific modification of a desired peptide, there are a few limitations one has to recognize that further investigations

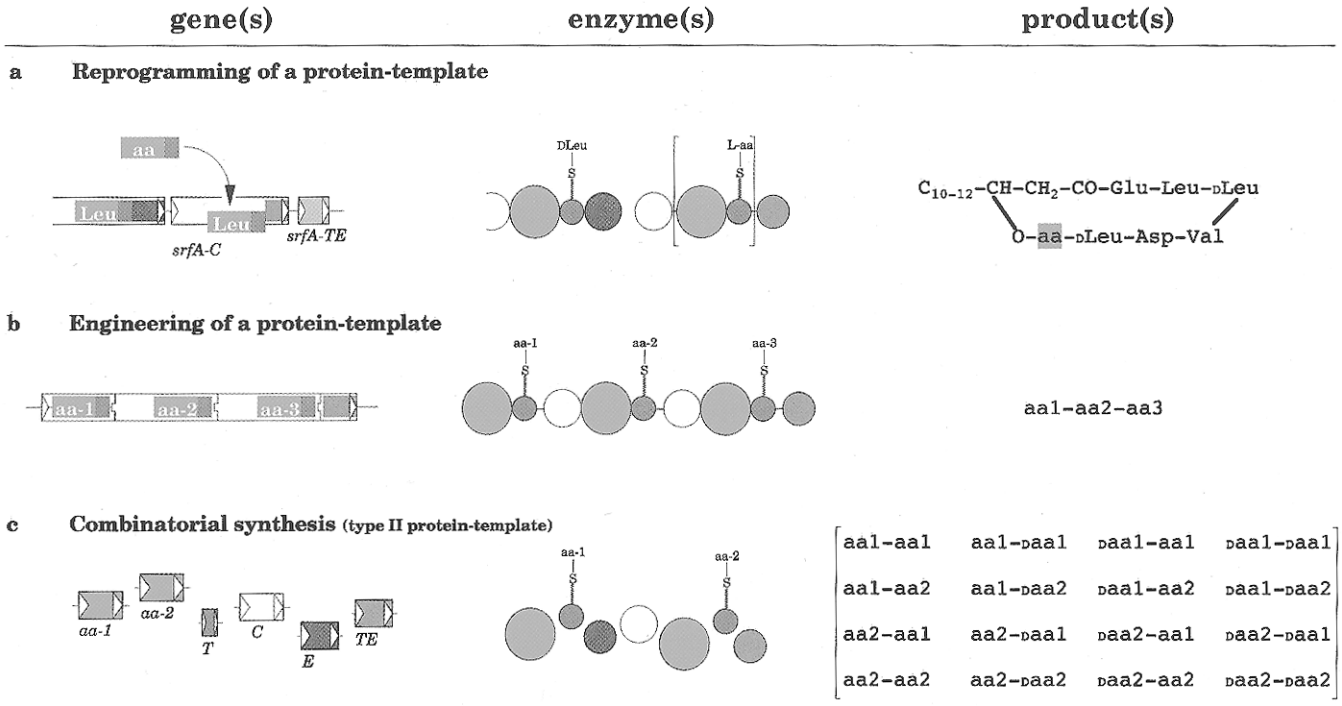


Figure 11. Proposed strategies of gene manipulations for the production of hybrid (a and b) and type II (c) peptide synthetases to generate specific alteration (a and b) and diversity (c) in peptide synthesis using the multienzyme thiotemplate mechanism.

have to conquer. On the one hand, the method requires well-defined sequence information about the biosynthetic system to be manipulated. Unfortunately, only a limited number of biosynthetic systems are currently characterized at the primary level. However, progress in the identification and sequencing of peptide synthetase genes from various organisms has been recently made by taking advantage of the strong conservation of signature core motifs in the domain structure (see Figure 4 and Table 3). These fingerprint regions have allowed the detection of peptide synthetase-encoding genes using the PCR method.[51,52]

However, the structure−function relationships of the engineered secondary metabolites are difficult to predict. The success of novel peptides depends mainly on their biological activity, and it seems unlikely that random changes would improve properties. Structural modeling tools like computer-aided drug design could be a prospect to overcome this problem and to stimulate the development of peptide-based drugs.[153−155] The emerging structural and functional capabilities of nonribosomally synthesized peptides will allow us to evaluate whether these sources can be used to create new products of biological significance.

A direct approach for a targeted reprogramming of peptide synthetases could be available by defining the structural basis of substrate specificity. Determination of the crystal structure of the first adenylation domain PheA (see section IV.A) accommodates the knowledge about the moieties involved in substrate binding and in constituting the specificity pocket.[75] Combined with the available sequence data for several A domains,[29−39,41,42] this may provide a foundation for understanding the structural basis of substrate specificity in modular peptide synthetases. These results may permit the alteration of certain residues within the adenylation domains by site-directed mutagenesis to modify the substrate specificity. Thus, programmed modifications in the primary structure of a given peptide antibiotic might be achieved by the substitution of few particular amino acid residues (within its adenylation unit) instead of exchanging entire modules or domains.

In general, the concepts presented have focused on the modification of natural secondary metabolites that are known to possess biological activities. In the near future, the field of study may be moving toward a complete engineering of a biosynthetic system (conveyor belt; Figure 11b).[7]

In addition to the reprogramming and possible *de novo* generation of protein templates, a third possibility for the production of hybrid bioactive peptides would be combinatorial peptide synthesis accomplished by nonintegrated peptide synthetases (Figure 11c).[24] As mentioned above, fatty acid synthases (FAS), polyketide synthases (PKS), and peptide synthetases share a similar mode of product assembly and possess a modular arrangement.[43,156−158] Nevertheless, a distinction can be made between large modular enzymes and enzyme complexes composed of sets of freely dissociable proteins, which are classified as type I and type II enzymes, respectively. The architecture of discrete proteins (type II) is only appropriate for systems involving multiple repeated reaction cycles (e.g. FAS, PKS), whereas the biosynthesis of a defined peptide essentially depends on the presence of a protein template containing the correct order and appropriate amount of activating modules.[1,2,5−7] All peptide synthetases studied so far are exclusively modular enzymes of type I. However, we have recently demonstrated that catalytic domains of peptide synthetases are also able to act as

individual enzymes and have shown *in trans* a productive interaction with other distinct domains.[24] Such a complex corresponds to a functional type II FAS or PKS and may provide a tool for some combinatorial approaches as previously shown for PKS.[43,159−161] Construction of artificial type II synthetases would increase the ability to generate manifold peptides with diverse structures (Figure 11c).

The increasing number of pathogenic organisms that are becoming more and more resistant to traditional therapy necessarily requires innovative concepts to generate novel pharmaceutically useful drugs.[162−165] Because of their enormous structural and functional diversity, nonribosomally synthesized peptides seem to be privileged to meet these demands.[7] The (near) future will show how far the various approaches will go toward the engineering of novel peptides of therapeutic use.

VII. Conclusions

Much has been learned about the modular structure of peptide synthetases, the multienzymes needed for nonribosomal peptides synthesis in bacteria and filamentous fungi. These protein templates catalyze the successive condensation of amino acids through adenylation, thiolation, and transpeptidation reactions on specific activation units designated modules, whose spatial organization defines the order of the incorporated residues in the final product. The modules, depending on their role in activation and modification of the substrate are comprised of several domains. Most of these domains seem to act as independent catalytic units, although located on a single polypeptide chain. For example, the adenylation domain, the heart of each module, acts independently at the level of substrate recognition and activation. For peptide condensation and substrate modification, however, specific contacts with other domains are necessary. The crystal structure of the adenylation domain PheA may act as a prototype for other adenylation domains within this superfamily of enzymes. From these structural studies, detailed information is expected to emerge that may allow specific alterations to modify the substrate specificity. Biochemical data on the integrity of the thiolation domain and its posttranslational modification with the cofactor has been recently obtained, strengthening the model of the multidomain arrangement of peptide synthetases. Other modifying domains involved in epimerization and *N*-methylation have been mapped on the basis of limited biochemical studies or sequence alignments with enzymes of known function. Finally, knowledge of the modular structure and domain organization of these multifunctional enzymes has been used successfully at the genetic level to alter the protein template.

Although a large body of information is now accumulating on the structure−function relationships of this highly interesting family of multienzymes, much of their basic aspects still remain unclear. It is not known how transpeptidation occurs and which role the condensation domains and the cofactor 4′-PP play during this process. Accordingly, it is unclear how this interplay precisely controls the direction of peptide chain growth. Little is known also about the details of epimerization, *N*-methylation and cyclization reactions and exactly how the substrate specificity within the adenylation domain is determined.

VIII. Acknowledgments

The authors thank all members of the Marahiel group for helpful discussion, Dr. Milton Stubbs for comments on the manuscript and Dr. Peter Brick, Imperial College London, for providing Figure 6. H.D.M. wishes to acknowledge his Ph.D. fellowship of the Fonds der Chemischen Industrie. Work carried out in the MAM laboratory was supported by the Deutsche Forschungsgemeinschaft, EC-project (Contract-No. BI04-CT95-01769), Biochemie Gesellschaft m.b.H. and the Fonds der Chemischen Industrie.

IX. References

(1) Marahiel, M. A. *FEBS Lett.* **1992**, *307*, 40−43.
(2) Zuber, P.; Nakano, M. M.; Marahiel, M. A. In *Bacillus subtilis and other Gram-positive bacteria*; Sonenshein, A. L., Hoch, J. A., Losick, R., Eds.; Am. Soc. Microbiol.: Washington, DC, 1993.
(3) Hansen, N. J. *Annu. Rev. Microbiol.* **1993**, *47*, 535−564.
(4) Sahl, H.-G.; Jack, R. W.; Bierbaum, G. *Eur. J. Biochem.* **1995**, *230*, 827−853.
(5) Stachelhaus, T.; Marahiel, M. A. *FEMS Microbiol. Lett.* **1995**, *125*, 3−14.
(6) Kleinkauf, H.; von Döhren, H. *Eur. J. Biochem.* **1996**, *236*, 335−351.
(7) Stachelhaus, T.; Schneider, A.; Marahiel, M. A. *Biochem. Pharmacol.* **1996**, *52*, 177−186.
(8) Zocher, R.; Keller, U. *Adv. Microbiol. Physiol.* **1997**, *38*, 85−131.
(9) Vining, L. C. *Annu. Rev. Microbiol.* **1990**, *44*, 395−427.
(10) Kleinkauf, H.; von Döhren, H. *Eur. J. Biochem.* **1990**, *192*, 1−15.
(11) Heider, J.; Baron, C.; Bock, A. *EMBO J.* **1992**, *11*, 3759−3766.
(12) Berry, M. J.; Harney, J. W.; Ohama, T.; Hatfield, D. L. *Nucleic Acids Res.* **1994**, *22*, 3753−3759.
(13) Baron, C.; Böck, A. In *tRNA: Structure, biosynthesis and function*; Söll, D., RajBhandary, U., Eds.; Am. Soc. Microbiol.: Washington, DC, 1995.
(14) Demain, A. L. In *Secondary metabolites: their function and evolution*; Symposium, C. F., Ed.; John Wiley & Sons: New York, 1992; Vol. 171.
(15) Schaeffer, P. *Bacteriol. Rev.* **1969**, *33*, 48−71.
(16) Marahiel, M. A.; Danders, W.; Krause, M.; Kleinkauf, H. *Eur. J. Biochem.* **1979**, *99*, 49−55.
(17) Horinouchi, S.; Beppu, T. *Crit. Rev. Biotechnol.* **1990**, *10*, 191−204.
(18) Hahn, J.; Dubnau, D. *J. Bacteriol.* **1991**, *173*, 7275−7282.
(19) Davies, J. *Mol. Microbiol.* **1990**, *4*, 1227−1232.
(20) Lipmann, F.; Gevers, W.; Kleinkauf, H.; Roskoski, R. *Adv. Enzymol.* **1971**, *35*, 1−33.
(21) Lipmann, F. *Adv. Microbiol. Physiol.* **1980**, *21*, 227−266.
(22) Stein, T.; Vater, J.; Kruft, V.; Wittmann-Liebold, B.; Franke, P.; Panico, M.; McDowell, R.; Morris, H. R. *FEBS Lett.* **1994**, *340*, 39−44.
(23) Stein, T.; Vater, J.; Kruft, V.; Wittmann-Liebold, B.; Franke, P.; Panico, M.; McDowell, R.; Morris, H. R. *J. Biol. Chem.* **1996**, *271*, 15428−15435.
(24) Stachelhaus, T.; Hüser, A.; Marahiel, M. A. *Chem. Biol.* **1996**, *3*, 913−921.
(25) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923−936.
(26) Pfeifer, E.; Pavela-Vrancic, M.; von Döhren, H.; Kleinkauf, H. *Biochemistry* **1995**, *34*, 7450−7459.
(27) Stein, T.; Kluge, B.; Vater, J.; Franke, P.; Otto, A.; Wittmann-Liebold, B. *Biochemistry* **1995**, *34*, 4633−4642.
(28) Haese, A.; Pieper, R.; von Ostrowski, T.; Zocher, R. *J. Mol. Biol.* **1994**, *243*, 116−122.
(29) Smith, D. J.; Alison, J. E.; Turner, G. *EMBO J.* **1990**, *9*, 2743-2750.
(30) Gutiérrez, S.; Diez, B.; Montenegro, E.; Martín, J. F. *J. Bacteriol.* **1991**, *173*, 2354−2365.
(31) Coque, J. R. R.; Martin, J. F.; Liras, P. *Mol. Microbiol.* **1991**, *5*, 1125−1133.
(32) Turgay, K.; Krause, M.; Marahiel, M. A. *Mol. Microbiol.* **1992**, *6*, 529−546.

(33) Cosmina, P.; Rodriguez, F.; de Ferra, F.; Perego, M.; Venema, G.; van Sinderen, D. *Mol. Microbiol.* **1993**, *8*, 821−831.
(34) Haese, A.; Schubert, M.; Herrmann, M.; Zocher, R. *Mol. Microbiol.* **1993**, *7*, 905−914.
(35) Weber, G.; Schörgendorfer, K.; Schneider-Scherzer, E.; Leitner, E. *Curr. Genet.* **1994**, *26*, 120−125.
(36) Scott-Craig, J. S.; Panaccione, D. G.; Pocard, J.-A.; Walton, J. D. *J. Biol. Chem.* **1992**, *267*, 26044−26049.
(37) Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M. M.; Zuber, P.; Yamane, K. *Nucleic Acids Res.* **1993**, *21*, 93−97.
(38) de Crécy-Lagard, V.; Blanc, V.; Gil, P.; Naudin, L.; Laurenzon, S.; Famechon, A.; Bamas-Jacques, N.; Crouzet, J.; Thibaut, D. *J. Bacteriol.* **1997**, *179*, 705−713.
(39) Merriman, T. R.; Merriman, M. E.; Lamont, I. L. *J. Bacteriol.* **1995**, *177*, 252−258.
(40) Zhang, J.-H.; Quigley, N. B.; Gross, D. C. *J. Bacteriol.* **1995**, *177*, 4009−4020.
(41) Mootz, H. D.; Marahiel, M. A. *J. Bacteriol.* **1997**, in press.
(42) Konz, D.; Marahiel, M. A. Manuscript in preparation, 1997.
(43) Katz, L.; Donadio, S. *Annu. Rev. Microbiol.* **1993**, *47*, 875−912.
(44) Krause, M.; Marahiel, M. A.; von Döhren, H.; Kleinkauf, H. *J. Bacteriol.* **1985**, *162*, 1120−1125.
(45) Kurotsu, T.; Hori, K.; Kanda, M.; Saito, Y. *J. Biochem.* **1991**, *109*, 763−769.
(46) Hori, K.; Yamamoto, Y.; Tokita, K.; Saito, F.; Kurotsu, T.; Kanda, M.; Okamura, K.; Furuyama, J.; Saito, Y. *J. Biochem.* **1991**, *110*, 111−119.
(47) Skarpeid, H. J.; Zimmer, T.-L.; Shen, B.; Döhren, H. v. *Eur. J. Biochem.* **1990**, *187*, 627−633.
(48) Skarpeid, H. J.; Zimmer, T.-L.; Döhren, H. v. *Eur. J. Biochem.* **1990**, *189*, 517−522.
(49) Stachelhaus, T.; Marahiel, M. A. *J. Biol. Chem.* **1995**, *270*, 6163−6169.
(50) Pospiech, A.; Bietenhader, J.; Schupp, T. *Microbiology* **1996**, *142*, 741−746.
(51) Borchert, S.; Patil, S. S.; Marahiel, M. A. *FEMS Microbiol. Lett.* **1992**, *92*, 175−180.
(52) Turgay, K.; Marahiel, M. A. *Peptide Res.* **1994**, *7*, 238−241.
(53) Krätzschmar, J.; Krause, M.; Marahiel, M. A. *J. Bacteriol.* **1989**, *171*, 5422−5429.
(54) Weckermann, R.; Fürbass, R.; Marahiel, M. A. *Nucleic Acids Res.* **1988**, *16*, 11841.
(55) Dieckmann, R.; Lee, Y.-O.; van Liempt, H.; von Döhren, H.; Kleinkauf, H. *FEBS Lett.* **1995**, *357*, 212−216.
(56) Lambalot, R. H.; Walsh, C. T. *J. Biol. Chem.* **1995**, *270*, 24658−24661.
(57) Aharonowitz, Y.; Cohen, G.; Martin, J. F. *Annu. Rev. Microbiol.* **1992**, *46*, 461−495.
(58) Aharonowitz, Y.; Bergmeyer, J.; Cantoral, J. M.; Cohen, G.; Demain, A. L.; Fink, U.; Kinghorn, J.; Kleinkauf, H.; MacCabe, A.; Palissa, H.; Pfeifer, E.; Schwecke, T.; Liempt, H. v.; Döhren, H. v.; Wolfe, S.; Zhang, J. *Bio/Technology* **1993**, *11*, 807−810.
(59) de Crécy-Lagard, V.; Marlière, P.; Saurin, W. *Life Sci.* **1995**, *318*, 927−936.
(60) Shiau, C. Y.; Baldwin, J. E.; Byford, M. F.; Sobey, W. J.; Schofield, C. J. *FEBS Lett.* **1995**, *358*, 97−100.
(61) Kallow, W.; von Döhren, H.; Kennedy, J.; Turner, G. In the proceeding book of the symposium of the special research group *Integrated Enzyme Systems*: Enzymology of biosynthesis of natural products, Berlin, 1996.
(62) Schneider, A.; Marahiel, M. A. Personal communication.
(63) Borchert, S.; Stachelhaus, T.; Marahiel, M. A. *J. Bacteriol.* **1994**, *176*, 2458−2462.
(64) Nakano, M. M.; Marahiel, M. A.; Zuber, P. *J. Bacteriol.* **1988**, *170*, 5662−5668.
(65) Grossman, T. H.; Tuckman, M.; Ellestad, S.; Osburne, M. S. *J. Bacteriol.* **1993**, *175*, 6203−6211.
(66) Gaidenko, T. A.; Belitsky, B. R.; Haykinson, M. J. *Biotechnologia* **1992**, 13−19.
(67) Rapaport, E.; Remy, P.; Kleinkauf, H.; Vater, J.; Zamecnik, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7891−7895.
(68) Delarue, M. *J. Mol. Evol.* **1995**, *41*, 703−711.
(69) Carter, C. W. J. *Annu. Rev. Biochem.* **1993**, *62*, 715−748.
(70) Onesti, S.; Miller, A. D.; Brick, P. *Structure* **1995**, *3*, 163−176.
(71) Cusack, S.; Yaremchuk, A.; Tukalo, M. *EMBO J.* **1996**, *15*, 2834−2842.
(72) Baldwin, T. O. *Structure* **1996**, *4*, 223−228.
(73) Conti, E.; Franks, N. P.; Brick, P. *Structure* **1996**, *4*, 287−298.
(74) Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D. *Nature* **1990**, *347*, 203−206.
(75) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997**, *16*, 4174−4183.
(76) Lee, S. G.; Lipmann, F. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 607−611.
(77) Lee, S. G.; Roskowski, R. J.; Bauer, K.; Lipmann, F. *Biochemistry* **1973**, *12*, 398−405.
(78) Tokita, K.; Hori, K.; Kurotsu, T.; Kanda, M.; Saito, Y. *J. Biochem.* **1993**, *114*, 522−527.
(79) Stachelhaus, T.; Schneider, A.; Marahiel, M. A. *Science* **1995**, *269*, 69−72.
(80) Scotti, C.; Piatti, M.; Cuzzoni, A.; Perani, P.; Tognoni, A.; Grandi, G.; Galizzi, A.; Albertini, A. M. *Gene* **1993**, *130*, 65−71.
(81) Albertini, A. M.; Caramori, T.; Scoffone, F.; Scotti, C.; Galizzi, A. *Microbiology* **1995**, *141*, 299−309.
(82) Mootz, M.; Marahiel, M. A. Personal communication.
(83) Walker, J. E.; Saraste, M.; Runswick, M. J.; Gay, N. J. *EMBO J.* **1982**, *1*, 945−951.
(84) Gocht, M.; Marahiel, M. A. *J. Bacteriol.* **1994**, *176*, 2654−2662.
(85) Saito, M.; Hori, K.; Kurotsu, T.; Kanda, M.; Saito, Y. *J. Biochem.* **1995**, *117*, 276−282.
(86) Hamoen, L. W.; Eshuis, H.; Jongbloed, J.; Venema, G.; van Sinderen, D. *Mol. Microbiol.* **1994**, *15*, 55−63.
(87) Serrano, R.; Kielland-Brandt, M. C.; Fink, G. R. *Nature* **1986**, *319*, 689−693.
(88) Pavela-Vrancic, M.; Pfeifer, E.; Schröder, W.; von Döhren, H.; Kleinkauf, H. *J. Biol. Chem.* **1994**, *269*, 14962−14966.
(89) Pavela-Vrancic, M.; Pfeifer, E.; van Liempt, H.; Schäfer, H.-J.; von Döhren, H.; Kleinkauf, H. *Biochemistry* **1994**, *33*, 6276−6283.
(90) Pavela-Vrancic, M.; Liempt, H. v.; Pfeifer, E.; Freist, W.; Döhren, H. v. *Eur. J. Biochem.* **1994**, *220*, 535−542.
(91) Peypoux, F.; Bonmatin, J. M.; Labbé, H.; Das, B. C.; Ptak, M.; Michel, G. *Eur. J. Biochem.* **1991**, *202*, 101−106.
(92) Peypoux, F.; Bonmatin, J.-M.; Labbé, H.; Grangemard, I.; Das, B. C.; Ptak, M.; Wallach, J.; Michel, G. *Eur. J. Biochem.* **1994**, *1994*, 89−96.
(93) Lawen, A.; Traber, R. *J. Biol. Chem.* **1993**, *268*, 20452−20465.
(94) Schmidt, E.; Schimmel, P. *Biochemistry* **1995**, *34*, 11204−11210.
(95) Kleinkauf, H. *J. Med. Plant Res.* **1979**, *35*, 1−18.
(96) Zocher, R. In the proceeding book of the symposium of the special research group *Integrated Enzyme Systems*: Enzymology of biosynthesis of natural products, Berlin, 1996.
(97) Ruttenberg, M. A.; Mach, B. *Biochemistry* **1966**, *5*, 2864−2869.
(98) Kleinkauf, H.; Dittmann, J.; Lawen, A. *Biomed. Biochim. Acta* **1991**, 219−224.
(99) Pieper, R.; Kleinkauf, H.; Zocher, R. *J. Antibiot.* **1992**, *45*, 1273−1277.
(100) Galli, G.; Rodriguez, F.; Cosmina, P.; Pratesi, C.; R., N.; de Ferra, F.; Grandi, G. *Biochim. Biophys. Acta* **1994**, *1205*, 19−28.
(101) Schlumbohm, W.; Stein, T.; Ullrich, C.; Vater, J.; Krause, M.; Marahiel, M. A.; Kruft, V.; Wittmann-Liebold, B. *J. Biol. Chem.* **1991**, *266*, 23135−23141.
(102) Cheret, G.; Mattheakis, L. C.; Sor, F. *Yeast* **1993**, *9*, 661−667.
(103) Rawlings, M.; Cronan, J. E. J. *J. Biol. Chem.* **1992**, *267*, 5751−5754.
(104) Sherman, D. H.; Malpartida, F.; Bibb, M. J.; Kieser, H. M.; Bibb, M. J.; Hopwood, D. A. *EMBO J.* **1989**, *8*, 2717−2725.
(105) Fernadez-Moreno, M. A.; Martinez, E.; Boto, L.; Hopwood, D. A.; Malpartida, F. *J. Biol. Chem.* **1992**, *267*, 19278−19290.
(106) Nakano, M. M.; Corbell, N.; Besson, J.; Zuber, P. *Mol. Gen. Genet.* **1992**, *232*, 313−321.
(107) Rusnak, F.; Sakaitani, M.; Drueckhammer, D.; Reichert, J.; Walsh, C. T. *Biochemistry* **1991**, *30*, 2916−2927.
(108) Schwecke, T.; Aharonowitz, Y.; Palissa, H.; von Döhren, H.; Kleinkauf, H.; van Liempt, H. *Eur. J. Biochem.* **1992**, *205*, 687−694.
(109) Wang, G. F.; Kuriki, T.; Roy, K. L.; Kaneda, T. *Eur. J. Biochem.* **1993**, *213*, 1091−1099.
(110) Lau, K. S.; Chuang, J. L.; Herring, W. J.; Danner, D. J.; Cox, R. P.; Chuang, D. T. *Biochim. Biophys. Acta* **1992**, *1132*, 319−321.
(111) Murray, I. A.; Martinez-Suarez, J. V.; Close, T. J.; Shaw, W. V. *Biochem. J.* **1990**, *272*, 505−510.
(112) Steffel, G.; Matzura, H. *Gene* **1989**, *75*, 349−354.
(113) Mattevi, A.; Obmolova, G.; Schulze, E.; Kalk, K. H.; Westphal, A. H.; de Kok, A.; Hol, W. G. J. *Science* **1992**, *255*, 1544−1550.
(114) Leslie, A. G. W. *J. Mol. Biol.* **1990**, *213*, 167−186.
(115) Vater, J.; Vollenbroich, D.; Zuber, P. In the proceeding book of the symposium of the special research group *Integrated Enzyme Systems*: Enzymology of biosynthesis of natural products, Berlin, 1996.
(116) Stindl, A.; Keller, U. *J. Biol. Chem.* **1993**, *268*, 10612−10620.
(117) Keller, U.; Schlumbohm, W. *J. Biol. Chem.* **1992**, *267*, 11745−11752.
(118) Keller, U. In the proceeding book of the symposium of the special research group *Integrated Enzyme Systems*: Enzymology of biosynthesis of natural products, Berlin, 1996.
(119) Diez, B.; Gutierrez, S.; Barredo, J. L.; van Solingen, P.; van der Voort, L. H. M.; Martin, J. F. *J. Biol. Chem.* **1990**, *265*, 16358−16365.
(120) Turner, G.; Kennedy, J.; Arezi, B.; Kallow, W. In the proceeding book of the symposium of the special research group *Integrated Enzyme Systems*: Enzymology of biosynthesis of natural products, Berlin, 1996.
(121) Cortes, J.; Wiesmann, E. H.; Roberts, G. A.; Brown, M. J. B.; Staunton, J.; Leadlay, P. F. *Science* **1995**, *268*, 1487−1489.
(122) Pazirandeh, M.; Chirala, S. S.; Huang, W.-Y.; Wakil, S. J. *J. Biol. Chem.* **1989**, *264*, 18195−18201.
(123) Pazirandeh, M.; Chirala, S. S.; Wakil, S. J. *J. Biol. Chem.* **1991**, *266*, 20946−20952.

(124) Haydock, S. F.; Dowson, J. A.; Dhillon, N.; Roberts, G. A.; Cortes, J.; Leadlay, P. F. *Mol. Gen. Genet.* **1991**, *230*, 120−128.
(125) Beck, K. F.; Schreglmann, R.; Stathopulos, I.; Klein, H.; Hoch, J.; Schweizer, M. *DNA Sequence* **1992**, *2*, 359−386.
(126) Jayakumar, A.; Tai, M. H.; Huang, W. Y.; al-Feel, W.; Hsu, M.; Abu-Elheiga, L.; Chirala, S. S.; Wakil, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8695−8699.
(127) Randhawa, Z. I.; Smith, S. *Biochemistry* **1987**, *26*, 1356−1373.
(128) Sasaki, G. C.; Cheesbrouhh, V.; Kolattukudy, P. E. *DNA* **1988**, *7*, 449−457.
(129) Schweizer, M.; Roberts, L. M.; Holtke, H. J.; Takabayashi, K.; Hollerer, E.; Hoffmann, B.; Muller, G.; Kottig, H.; Schweizer, E. *Mol. Gen. Genet.* **1986**, *203*, 479−486.
(130) Zhao, X. J.; Cihlar, R. L. *Gene* **1994**, *147*, 119−124.
(131) Stachelhaus, T.; Marahiel, M. A. Personal communication.
(132) Vater, J.; Kleinkauf, H. *Biochim. Biophys. Acta* **1976**, *429*, 1062−1072.
(133) Stindl, A.; Keller, U. *Biochemistry* **1994**, *33*, 9358−9364.
(134) Hoffmann, K.; Schneider-Scherzer, E.; Kleinkauf, H.; Zocher, R. *J. Biol. Chem.* **1994**, *269*, 12710−12714.
(135) Faraci, W. S.; Walsh, C. T. *Biochemistry* **1988**, *27*, 3267−3276.
(136) Galakatos, N. G.; Walsh, C. T. *Biochemistry* **1987**, *26*, 8475−8480.
(137) Badet, B.; Inagaki, K.; Soda, K.; Walsh, C. T. *Biochemistry* **1986**, *25*, 3275−3282.
(138) Gallo, K. A.; Knowles, J. R. *Biochemistry* **1993**, *32*, 3981−3990.
(139) Gallo, K. A.; Tanner, M. E.; Knowles, J. R. *Biochemistry* **1993**, *32*, 3991−3997.
(140) Tanner, M. E.; Gallo, K. A.; Knowles, J. R. *Biochemistry* **1993**, *32*, 3998−4006.
(141) Cardinale, G. J.; Abeles, R. H. *Biochemistry* **1968**, *7*, 3970−3978.
(142) Nwankwo, D. O.; Moran, L. S.; Slatko, B. E.; Waite-Rees, P. A.; Dorner, L. F.; Benner, J. S.; Wilson, G. G. *Gene* **1994**, *150*, 75−80.
(143) Kawakami, B.; Sasaki, A.; Oka, M.; Maekawa, Y. *Agric. Biol. Chem.* **1990**, *54*, 3227−3233.
(144) Brooks, J. E.; Nathan, P. D.; Landry, D.; Sznyter, L. A.; Waite-Rees, P.; Ives, C. L.; Moran, L. S.; Slatko, B. E.; Benner, J. S. *Nucleic Acids Res.* **1991**, *19*, 841−850.
(145) Kim, H. S.; Choi, E. C.; Kim, B. K. *J. Gen. Microbiol.* **1993**, *139*, 601−607.
(146) Blattner, F. R.; Burland, V.; Plunkett, G. I.; Sofia, H. J.; Daniels, D. L. *Nucleic Acids. Res.* **1993**, *21*, 5408−5417.
(147) Molnár, I.; Aparicio, J. F.; Haydock, S. F.; Khaw, L. E.; Schwecke, T.; König, A.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 1−7.
(148) Schwecke, T.; Aparico, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B.; Böhm, G. A.; Staunton, J.; Leadlay, P. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839−7843.
(149) Aparicio, J. F.; Molnár, I.; Schwecke, T.; König, A.; Haydock, S. F.; Khaw, L. E.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 9−16.
(150) Pieper, R.; Haese, A.; Schröder, W.; Zocher, R. *Eur. J. Biochem.* **1995**, *230*, 119−126.
(151) Billich, A.; Zocher, R. *Biochemistry* **1987**, *26*, 8417−8423.
(152) Schneider, A.; Stachelhaus, T.; Marahiel, M. A. *Mol. Gen. Genet.* **1997**, in press.
(153) Ringe, D. *J. Nucleic Med* **1995**, *36*, 28−30.
(154) Marshall, G. R. *Annu Rev. Pharmacol. Toxicol.* **1987**, *27*, 193−213.
(155) Ward, D. J.; Chen, Y.; Platt, E.; Robson, B. *J. Theor. Biol.* **1991**, *148*, 193−227.
(156) Wakil, S. J.; Stoops, J. K.; Joshi, V. C. *Annu. Rev. Biochem.* **1983**, *52*.
(157) Smith, S. *FASEB J.* **1994**, *8*, 1248−1259.
(158) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675−679.
(159) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546−1550.
(160) Khosla, C.; McDaniel, R.; Ebert-Khosla, S.; Torres, R.; Sherman, D. H.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1993**, *175*, 2197−2204.
(161) Tsoi, J. C.; Khosla, C. *Chem. Biol.* **1995**, *2*, 355−362.
(162) Neu, H. C. *Science* **1992**, *257*, 1064−1073.
(163) Bloom, B. R.; Murray, C. J. L. *Science* **1992**, *257*, 1055−1064.
(164) Service, R. F. *Science* **1995**, *270*, 724−727.
(165) Gibbons, A. *Science* **1992**, *257*, 1036−1038.

CR960029E