How do peptide synthetases generate structural diversity?

Dirk Konz and Mohamed A Marahiel

Many low-molecular-weight peptides of microbial origin are synthesized nonribosomally on large multifunctional proteins, termed peptide synthetases. These enzymes contain repeated building blocks in which several defined domains catalyze specific reactions of peptide synthesis. The order of these domains within the enzyme determines the sequence and structure of the peptide product.

Address: Philipps-Universität Marburg, Fachbereich Chemie/Biochemie, Hans-Meerwein-Strasse, D-35032 Marburg, Germany.

Correspondence: Mohamed A Marahiel
E-mail: marahiel@chemie.uni-marburg.de

Chemistry & Biology February 1999, **6**:R39–R48
http://biomednet.com/elecref/10745521006R0039

 ISSN 1074-5521

Introduction

Nonribosomal peptides of microbial origin 2–48 residues in length usually have an outstanding structural diversity that can, in part, be attributed to the incorporation of many unusual, nonproteinogenic residues. To date, more than 300 such residues, including D-configured and *N*-methylated amino acids or a variety of hydroxy acids, have been identified. Variations within the peptide backbone resulting in linear, cyclic or branched cyclic molecules, which can be further modified by acylation, glycosylation or heterocyclic ring formation, also contribute to the enormous variety of structures within this class of substances. Many nonribosomal peptides show interesting physicochemical or pharmacological characteristics, including biosurfactant, siderophore, antibiotic, antiviral, cytostatic, anticancer and immunosuppressive properties [1–3].

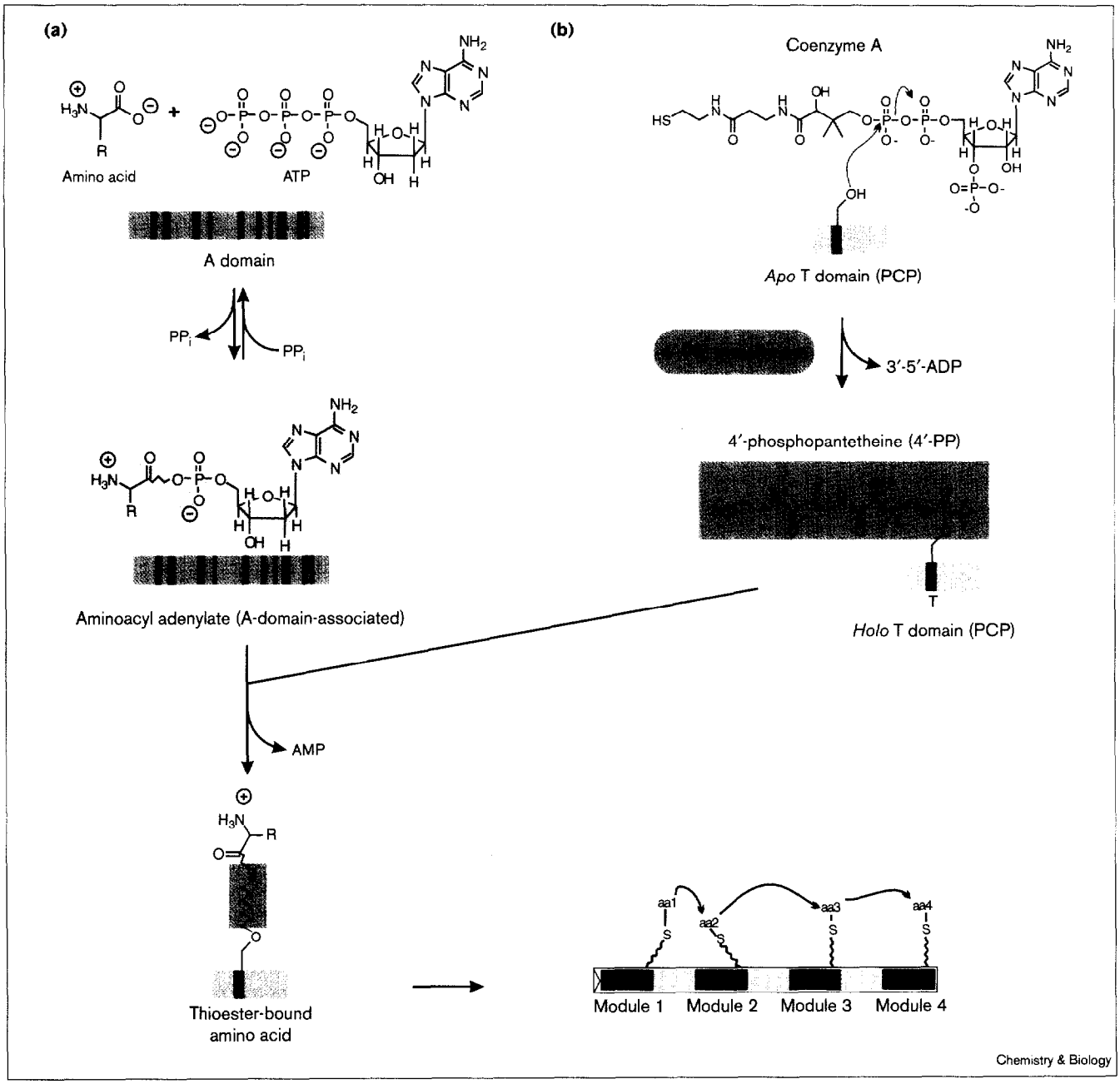
Structural diversity and a common mode of biosynthesis

In contrast with their diverse structures, most nonribosomally synthesized peptides share a uniform mode of biosynthesis — the multiple carrier thiotemplate mechanism (Figure 1a) [4–6]. Large multifunctional enzymes, termed peptide synthetases (PPS), activate their acyl substrates, by ATP-hydrolysis, as acyl adenylates. These unstable intermediates are subsequently tethered to covalently enzyme bound 4′-phosphopantetheinyl (4′-PP) cofactors as thioesters. For each incorporated residue, PPSs contain specific active sites, termed modules (Figure 2; colored regions), which represent semiautonomous regions within these enzymes that carry all the information needed for recognition, activation, thiolation and, in some cases, modification (epimerization or *N*-methylation) of a single substrate. These modules interact in an ordered fashion to generate the peptide product by the stepwise incorporation of the thioesterified residues in a series of amino- to carboxy-terminal directed transpeptidations. The order and nature of biosynthetic modules, from amino to carboxyl termini, in each PPS directs the sequence and structure of the formed product. PPSs therefore act as protein templates for the nonribosomal assembly of peptides [7–9].

Functional domains represent the 'toolbox' of PPS

Within the last decade, an increasing number of genes encoding PPSs of bacterial and fungal origin have been identified, cloned and sequenced (for examples, see Table 1, and for a more comprehensive list see the Supplementary material available with the online version). In bacterial systems, the genes coding for several PPSs involved in the synthesis of a specific peptide are typically organized in operons that can span regions of 6–45 kilobases (kb)

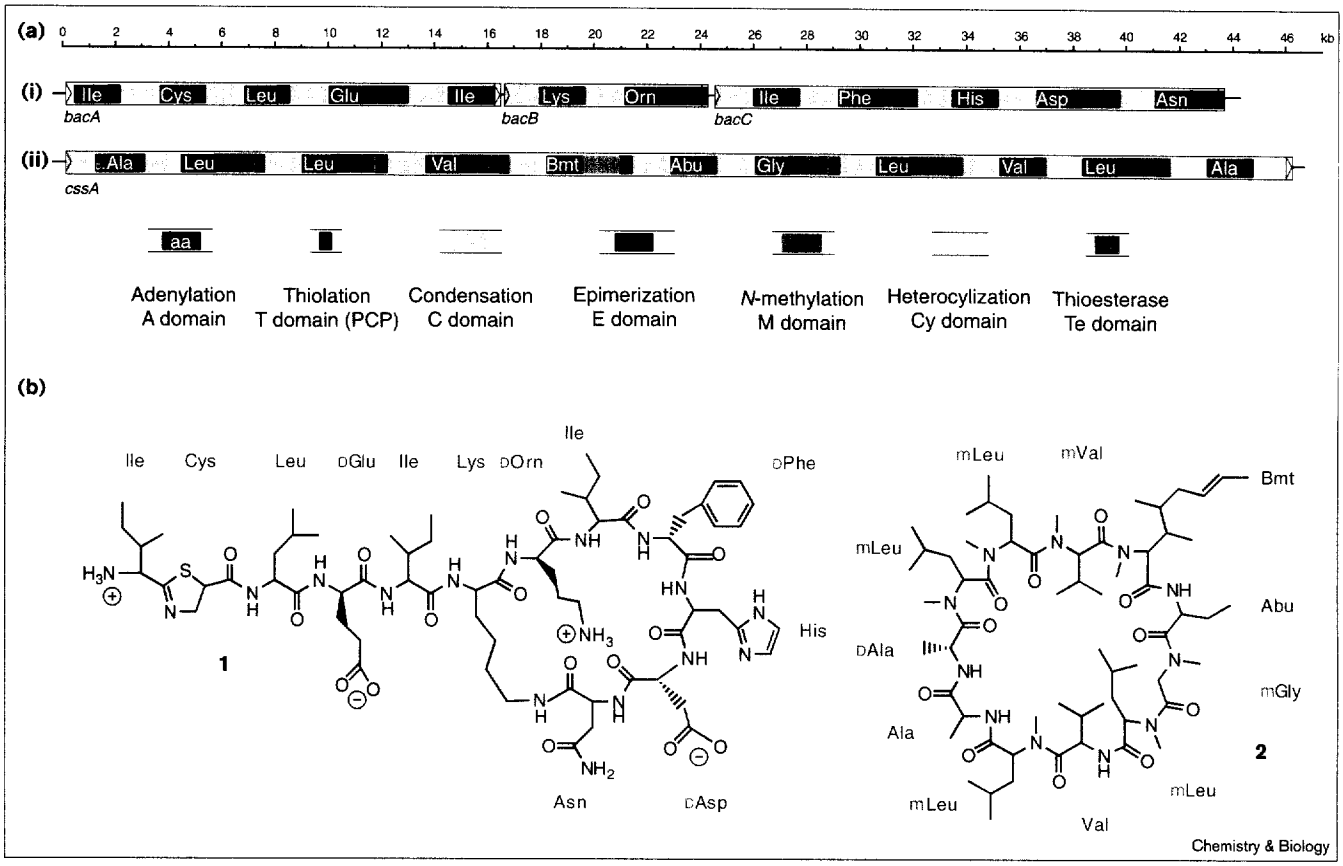
Figure 1



(a) The multiple carrier thiotemplate mechanism. The cognate amino acid of each PPS module is first activated as an enzyme associated acyl adenylate by the action of an adenylation A domain. This unstable intermediate is subsequently transferred to the T domain (PCP) of each module, where it is bound as a thioester to the cysteamine group of a covalently enzyme bound 4′-phosphopantetheine (4′-PP) cofactor. The thioesterified amino acids are then integrated into the peptide product through a stepwise elongation by a series of transpeptidation reactions. These occur by transfer of the thioester-activated carboxyl group of one residue to the adjacent amino group of the next amino acid. During this condensation all intermediates stay covalently linked to the PPS. **(b)** Post-translational conversion of the T domain (PCP) from its *apo* to *holo* form, through the action of a 4′-PP-transferase, which directs the nucleophilic attack of the hydroxy group of a highly conserved serine residue to the β-phosphate of CoA, mediating the transfer of the 4′PP moiety onto the T-domain. aa, amino acid.

[1,2]. The single PPSs in these systems can be comprised of 1–8 modules. As an example for such an organization, the genes encoding the three bacitracin synthetases BA1 (5 modules; 598 kDa), BA2 (2 modules; 297 kDa), and BA3 (5 modules; 723 kDa) of *Bacillus licheniformis* are shown in Figure 2 [10]. In contrast with the bacterial systems, the fungal protein templates for nonribosomal peptide synthesis are encoded by large, single genes. One of the

Figure 2



(a) The modular organization of PPS encoded by (i) the bacterial bacitracin operon *bacA-C* from *Bacillus licheniformis* and (ii) the fungal cyclosporin synthetase gene *cssA*. Red regions indicate the position and substrate specificity of A domains, and green stripes show the location of T domains (PCPs), the site of 4′-PP cofactor binding. Grey regions between the single modules mark the position of C domains, whereas yellow regions indicate Cy domains. Modules involved in the incorporation of D-amino acids have an E domain (blue region) located downstream of the T domain. The first module of the *cssA* gene (D-Ala) does not contain an E domain but specifically incorporates D-Ala supplied by an external epimerase [38]. M domains are shown as dark yellow boxes inserted between an A and a T domain, and the amino-terminal Te domain of the *bac*-operon is shown in pink. **(b)** The chemical structures of the peptide antibiotics bacitracin A **(1)** and cyclosporin A **(2)** assembled on the protein templates encoded by the genes *bacA–C* (i) and *cssA* (ii). Orn, ornithine; Bmt, (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine; aa, amino acid.

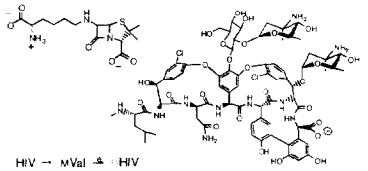
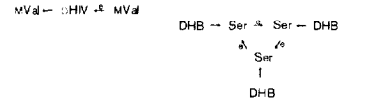
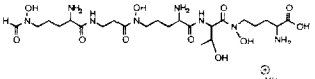
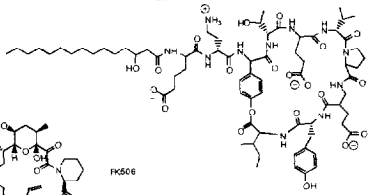
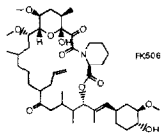
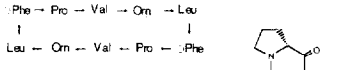
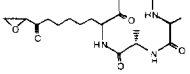
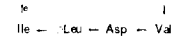
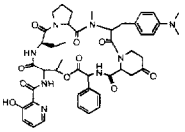
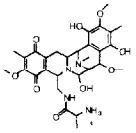
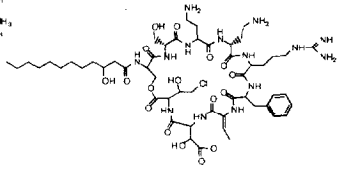
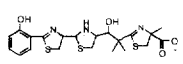
most impressive examples is the giant cyclosporin synthetase gene *cssA* from *Tolypocladium niveum* (Figure 2), which is the largest (46 kb) known open reading frame identified so far, encoding a huge PPS (1600 kDa) with 11 modules [11].

Thorough sequence and structure–function analysis of PPS genes have confirmed the modular architecture of this enzyme class at the molecular level, and have also revealed that each module is comprised of several defined domains catalyzing specific reactions within the sequence of nonribosomal peptide synthesis [1,2,12]. These domains can therefore be described as the 'toolbox' of PPS. In most cases, the single domains can be detected easily in sequence alignments by the presence of highly conserved signature or 'core' sequences that impart each domain type a characteristic fingerprint, irrespective of the enzyme origin (Figure 3 and Table 2). It has been demonstrated, using site-directed mutagenesis and photoaffinity labeling, that these core sequences contain important residues directly involved in reaction catalysis [13–17].

Probably the most important domain of each module is the adenylation (A) domain (Figures 2,3 and Table 2; red regions; 550 amino acids in length), which catalyzes the specific recognition and activation of a cognate carboxy acid as acyl adenylate by the hydrolysis of ATP (Figure 1a) [4,5,12]. According to sequence homology and the type of reaction it catalyzes, the A domain belongs to the large superfamily of adenylate-forming enzymes that includes firefly luciferase and acetyl-CoA synthetases [18,19]. In contrast to the A domains, which are integrative parts of the PPS, the homologous acetyl-CoA synthetases are single, independently acting enzymes. But in several

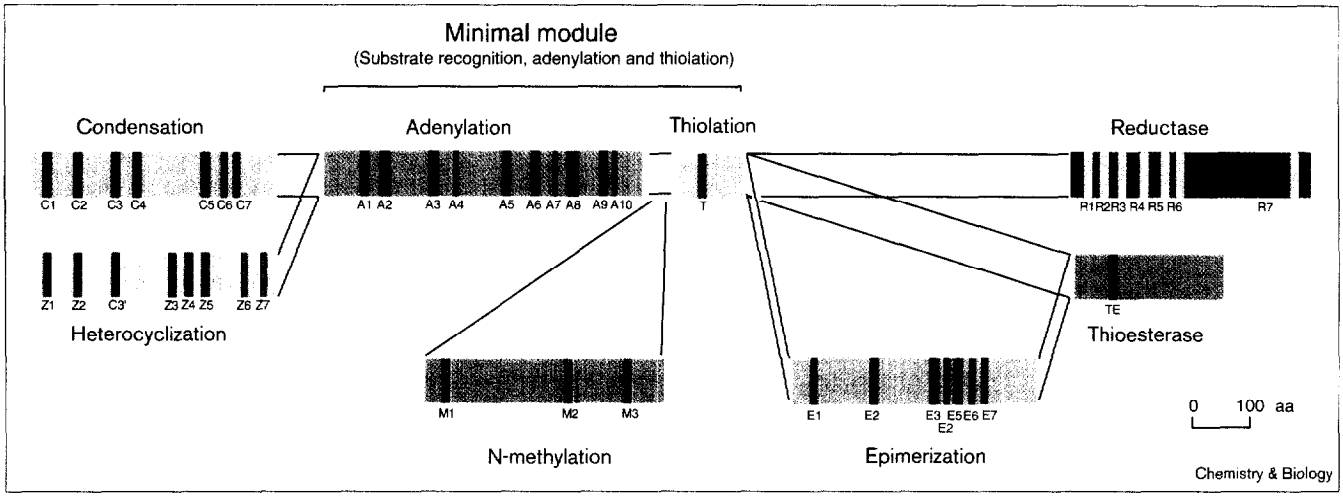
Table 1

Deduced domain structure of PPS genes and primary structure of the corresponding peptide product.

Product	Organism	Gene	Organisation	Accession PID	Literature	Primary structure of product
2,3-Dihydroxy-benzoylglycine*	*Bacillus subtilis*	*dhbE*	DHB	U26444 *g837332*	[56,57]	
		dhbB	isochrismatase	*g837833*		
		dhbF	Gly	Z99120 *g2635963*		
ACV†	*Penicillium chrysogenum*	*pcbAB*	Aad Cys Val	M54296 *g3118*	[58]	
Chloroeremo-mycin†	*Amycolatopsis orientalis*	*CepA*	DLeu Tyr Asn	AJ223999 *g2894188*	[59]	
		CepB	HPG HPG Tyr	*g2894189*		
		CepC	DHPG	*g2894190*		
Enniatin†	*Fusarium scirpi*	*esyn1*	DHIV MVal	Z18755 *g2730*	[33]	
Enterobactin*	*E. coli*	*entE*	DHB	X15058 *g41346*	[53,60,61]	
		entB	ischorismatase	*g522182*		
		entF	Ser	J05325 *g145843*		
Exochelin*	*Mycobacterium smegmatis*	*fxbB*	Orn bAla	AF027770 *g3560506*	[54]	
		fxbC	Orn Thr Orn ?	*g3560507*		
Fengycin†	*Bacillus subtilis*	*fenC*	Glu Orn	AF087452 *g3643187*		
		fenE	Glu Val	AF023465 *g2522214*	[62]	
		fenA	Pro Glu Tyr	AF023464 *g2522212*		
		fenB	Ile	L42523 *g840624*		
FK506†	*Streptomyces sp. MA6548*	*fkbP*	Pip	AF082100 *g3798625*		
Gramicidin S†	*Bacillus brevis*	*grsA*	Phe	X15577 *g39369*	[18,63]	
		grsB	Pro Val Orn Leu	X61658 *g39372*		
HC-toxin§	*Cochliobulus carbonum*	*hts1*	Pro Ala DAla Aeo	A45086 *g167219*	[64]	
Lichenysin D†	*Bacillus licheniformis*	*licA*	Gln Leu Leu	U95370 *g3080742*	[21]	
		licB	Val Asp Leu	*g3080743*		
		licC	Leu	*g3080744*		
Pistinamycin I†	*Streptomyces pristinaespiralis*	*snbA*	3hPic	X98515 *g1835254*	[65]	
		snbC	Thr Abu	Y11548 *g2052248*		
		snbDE	Pro MDMAP 4oPip PGly	*g2052249*		
Saframycin Mx1†	*Myxococcus Xanthus*	*safB*	Ala Gly	U24657 *g1171128*	[48]	
		safA	Tyr Tyr	*g117129*		
Syringomycin†§	*Pseudomonas syringae*	*syrE*	Ser DSer Dab DDab Arg Phe	AF047828 *g3510629*	[55]	
			Dht Asp		[66]	
		syrB	Thr	U88574 *g837256*		
Yersiniabactin*	*Yersinia enterolitica*	*irp5*	Sal	Y12527 *g2765198*	[67,68]	
		irp2	Cy Cy Cys	L18881 *g408802*		
		irp1	KS AT M KR T Cy	Y12527 *g2765195*		

*Siderophore; †antibiotic; ‡immunosuppressant; §phytotoxin. A (red), adenylation domain; T (green), thiolation domain or peptidyl carrier protein (PCP); C (grey), condensation domain; M (dark green), *N*-methylation domain; E (blue), epimerization domain; Cy (yellow), heterocyclization domain; Te (pink), thioesterase domain; Red (purple), reductase domain. Domains with homology to polyketide synthases: T, acyl carrier protein (ACP); AT, acyltransferase; KR, ketoreductase, KS, ketosynthase; M, methyltransferase. Modified amino acids are indicated by the use of prefixes: D, D configuration, M, *N*-methylation; xh, hydroxylation at position x relative to the α-carbon atom. Amino or hydroxy acids coupled by nontypical peptide bonds are marked: e, ester bond. Abbreviations for nonproteinogenic amino and hydroxy acids: 3hPic, 3-hydroxy-a-picolinic acid; Dab/4nAbu, 2,4-diamino-butyric acid; 4oPip, 4-oxo-L-pipecolinic acid; Abu = α-amino butyric acid; Aeo, 2-amino-8-oxo-9,10-epoxy decanoic acid; DHB, 2,3-dihydroxy benzoic acid; DHPG, 3,5-hydroxy-L-phenylglycine; Dht, dehydro-threonine, DMAP = *N*-methyl-(4)-dimethyl-amino-L-phenylalanine; FA, fatty acid; HIV, 2-hydroxy-isovaleric acid; HPG, 4-hydroxy-L-phenylglycine; Orn, L-ornithine; PGl, L-phenylglycine; Pip, L-pipecolinic acid; Sal, salicylate.

Figure 3



The organisation of domains in a structure of a PPS module. The particular composition of a module depends on the given requirements in regard of substrate activation, elongation and modification (see Figure 2). The location of highly conserved signature sequences within the particular domain types (A, red; T, green; C, grey; Cy, yellow; M, dark yellow; E, blue; Te, pink; and Red, violet) are indicated as stripes. The sequence of these core sequences is shown in Table 2.

recently reported studies it has been demonstrated that A domains of different origin, cloned and expressed in *Escherichia coli*, show catalytic activities comparable with those reported for wild-type PPS enzymes [20–22]. These findings clearly indicate that A domains operate as functionally independent units, as well as acting in concert with the surrounding domains of each particular PPS. Further support for this point of view comes from the observation that many PPS systems involved in the biosynthesis of catechol siderophores, for example, enterobactin from *E. coli* or yersiniabactin from *Yersinia pestis*, contain aryl-AMP-ligases. These enzymes, activating aryl carboxy acids as adenylates, are composed of a single, isolated A domain, supporting the finding that A domains are functionally independent.

The solution of the crystal structure of two members of the adenylate-forming enzymes, firefly luciferase of *Photinus pyralis* [23] and the A domain of the gramicidin S synthetase A (GrsA) from *Bacillus brevis* [24], revealed that these enzymes have an active fold that is different from those found in class I and class II amino acyl tRNA synthetases. Although the two enzymes share only 16% identity in their primary sequence, the overall topology of their three-dimensional structure is very similar. From the high degree of sequence identity (30–60%) of PPS A domains to each other it can be concluded that the GrsA structure, in particular, represents a prototype for all PPS adenylation domains [24].

In most PPS modules the A domain is followed by a thiolation (T) domain (Figures 2,3 and Table 2; green regions; 100 aa in length) that has a 4′-PP cofactor in its *holo* form. During peptide synthesis the acyl adenylates, associated with the A domains, are covalently tethered to the cystamine group of each corresponding T-domain-bound cofactor (Figure 1a) [6,15,25,26]. As the T domain shares high homology with the acyl carrier protein (ACP) of fatty acid and polyketide synthases, many authors also use the equivalent term peptidyl carrier protein (PCP) for this domain type [25]. It contains a core sequence with an invariant serine residue (Figure 3 and Table 2), which represents the 4′-PP-attachment site [16,27,28]. The specific 4′-PP modification of each T domain within PPSs is catalyzed by enzymes belonging to the superfamily of 4′-PP-transferases. These enzymes promote the nucleophilic attack of the invariant serine hydroxyl group to the pyrophosphate bridge of CoA, resulting in a transfer of the 4′-PP cofactor to the T domain and a liberation of 3′,5′-ADP (Figure 1b) [28,29].

Recently, it has been demonstrated by several investigators that heterologously expressed T domains are efficiently modified with 4′-PP *in vitro* by the action of purified 4′-PP-transferases [30–32]. In some cases, even an *in trans* aminoacylation of the *holo* T domains could be observed when incubated with the corresponding A domain, ATP and the cognate amino acid [25]. Similar conditions are found in natural PPS systems that synthesize catechol siderophores. Here, terminal-localized T domains become acylated with aryl carboxy acids *in trans* by the action of aryl-AMP-ligases. These data strongly support the idea of independently acting PPS domains.

Table 2

Highly conserved core motifs of the catalytic PPS domains.

Domain	Core*	Consensus sequence†
Adenylation	A1	L(TS)YxEL
	A2 (core 1)	LKAGxAYL(VL)P(LI)D
	A3 (core2)	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)(~)xR
	C6	(HN)QD(YV)PFE
	C7	RDxSRNPL
Heterocyclization	Z1	FPL(TS)xxQxAYxxGR
	Z2	RHx(IM)L(PAL)x(ND)GxQ
	C3′	(DNR)4xDxxS
	Z3	(LI)Pxx(PAL)x(LPF)P
	Z4	(TS)(PA)3x(LAF)6x(IVT)LxxW
	Z5	(GA)(DQN)FT
	Z6	P(IV)VF(TA)SxL
	Z7	QV(x(LI)Dx(QH)11xW(DYF)
N-methylation	MI (SAM)	VL(DE)GxGxG
	M2	NELSxYRYxAV
	M3	VExSxARQxGxLD
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)
Thioesterase	Te	G(HY)SxG
Reductase	R1 (NAD?H)	V(LF)(LV)TG(AV)(TN)G(YF)LG
	R2	V3xVRA
	R3	GDL
	R4	VYPYxxLRx(PL)NVxxT
	R5	GYxxSKWxxE
	R6	RPG
	R7	LExx(VI)GFLxxP

*Former nomenclature is in parentheses. †Single-letter amino-acid code is used for core sequences; alternative amino acids for a particular position are shown in parentheses; x, any amino acid; numbers indicate the spacing between conserved residues.

Modules responsible for the incorporation of D-configured or *N*-methylated amino acids have additional editing domains [1,2,9]. Modules incorporating *N*-methylated amino acids have an extra methylation (M) domain (Figures 2,3 and Tables 1,2; dark yellow regions; 420 aa in length) between the A domain and the T domain. The M domain contains at least three core motifs (Figure 3 and Table 2), including a glycine-rich sequence M1 that shows significant similarity to the common S-adenosylmethionine (SAM)-binding site of a heterologous class of cosubstrate-dependent methyltransferases. Biochemical studies revealed that this domain catalyzes the *N*-modification on the thioester prior to peptide-bond formation [33–35]. Similar conditions are found in D-amino-acid-utilizing modules that are extended at the carboxyl terminus of the T domain by an additional epimerization (E) domain (Figures 2,3 and Tables 1,2; blue regions, 400 aa in length), which catalyzes the L to D tranformation of the thioester-bound intermediates [22,36,37]. Some examples of modules are also known, however, that incorporate D amino acids into the product peptide but lack an E domain (e.g. chloroeremomycin, cyclosporin, HC-toxin and syringomycin; Table 1). In these examples the corresponding A domains specifically activate only the D-configured amino acid that is provided by an external epimerase, as has been demonstrated for the cyclosporin system [38].

The directed condensation of the thioesterified intermediates is catalyzed by condensation (C) domains (Figures 2,3 and Table 1 grey regions; ~. 450 aa in length), which are found as a part of the repetitive modules [39]. They coincide in frequency with the number of peptide bonds in the final linear peptide. C domains are conventionally fused to the amino-terminal end of modules accepting acyl groups from the preceding module, and they are absent in modules activating the first acyl constituent to be incorporated. In sequence alignments, no significant homology to other proteins that might have a common ancestor with similar catalytic properties has been identified. The C domains share a highly conserved core sequence C3 (Figure 3 and Table 2, His–His–X–X–Asp–Gly), however, with a class of well-studied acyl transferases, that includes dihydrolipoyl transacetylase and chloramphenicol acetyltransferase [1,39]. This signature motif has been identified, by mutational studies, to be critical for amide-bond formation. The second histidine is believed to serve as a base for deprotonation of the NH_3^+ moiety of the thioester-bound nucleophiles prior to amide-bond formation. Recently, it has been demonstrated that mutation of the second histidine to valine abolishes dipeptide formation *in vitro* in a gramicidin S/tyrocidine synthetase system, supporting the functional role of C domains in nonribosomal peptide-bond formation [40].

Besides the 'normal' positioning of the C domain between two modules mediating peptide-chain elongation, an extra C domain is found in several PPS systems, such as at the amino terminus of cyclosporin synthetase or at the carboxyl terminus of the enniatin, HC-toxin, rapamycin and FK506 systems (Table 1). According to this organization and the structure of the formed products, it can be concluded that these C domains are probably involved in peptide-chain termination and cyclization. Some of the

synthesized molecules are cyclized by the formation of an amide bond (e.g. cyclosporin and HC-toxin) and others by the formation of an ester bond (e.g. enniatin, rapamycin and FK506). C domains therefore must be able to catalyze two types of nucleophilic attack on the thioester carboxyl group (Figure 4): one by an amine leading to the formation of an amide bond and the other by a hydroxyl group leading to ester bonds or eventually hydrolysis (e.g. 2,3-dihydroxybenzoyl-glycine of *B. subtilis*).

A further domain type that shows similarity to C domains is found in modules involved in the formation of heterocyclic rings, such as oxazolines or thiazolines, within the peptide backbone. These domains, termed cyclization (Cy) domains (Figures 2,3 and Table 1 yellow regions; ~450 aa in length), can substitute C domains at the amino terminus of modules incorporating serine, threonine or cysteine residues and conduct heterocylization during peptide-bond formation [10,31,41,42]. Recently, the functional role of the Cy domain in condensation and heterocyclic ring formation has been demonstrated *in vitro* in a yersiniabactin synthetase system [43]. Little is known about the timing and the molecular mechanism of this reaction sequence, however. Residues conserved in C and Cy domains therefore might be good candidates for structure–function mutagenesis studies to attempt uncoupling of the condensation and heterocyclization processes.

Bacterial modules incorporating the last amino acid into a product peptide are often extended by a thioesterase (Te) domain (Figures 2,3 and Table 1 pink regions; ~250 aa in length) [1,2,7]. This domain shares sequence homology with thioesterases and bears a signature sequence Te (Figure 3 and Table 2, Gly–X–Ser–X–Gly) that is similar to the active-site motif of acyltransferases and thioesterases. It is thought that the full-length peptide bound to the last T domain is transferred to the hydroxyl group of the highly conserved serine residue within the Te domain to generate a transient acyl-O-enzyme intermediate [8]. This covalent species is then cleaved by an acyltransfer to water, resulting in a linear peptide (e.g. α-aminoadipyl-L-cysteinyl-D-valine (ACV) or chloroeremomycin) or to a functional group of a peptide sidechain liberating a cyclic (e.g. tyrocidine) or branched cyclic (e.g. bacitracin) product. In recent mutational studies it has been demonstrated that the Te domain of surfactin synthetase (SrfA–C) is essential for lipopeptide production [44] and is portable; when fused downstream of other SrfA–B modules, the Te domain directs the release of resultant intermediate peptide chains [45]. In addition, genes encoding 25–29 kDa proteins with significant homology to type II thioesterases have been found in many bacterial PPS operons [20,46,47]. These proteins have been shown to be important, but not essential, for peptide synthesis and their specific function is still unknown [44].

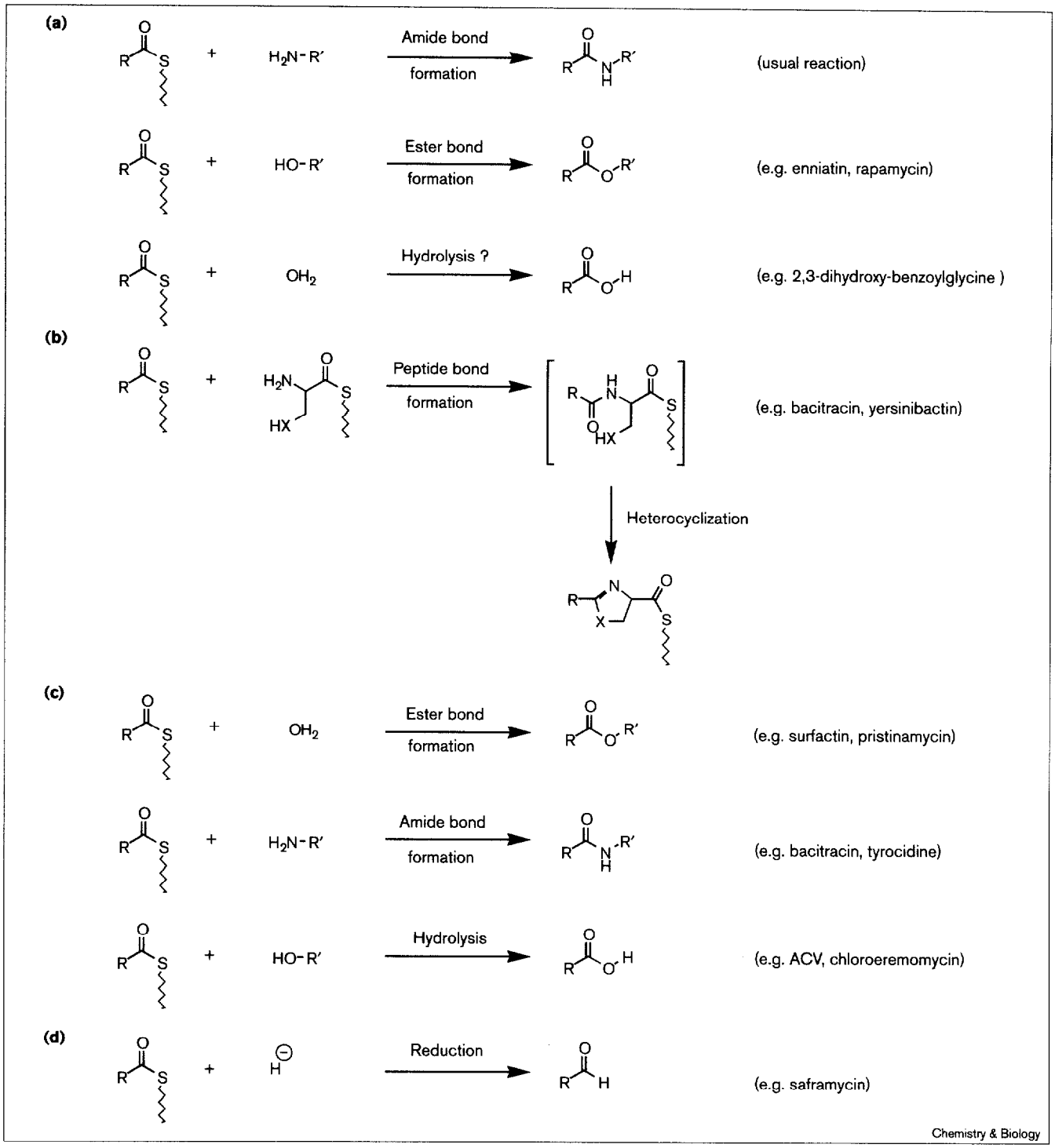
At the carboxy-terminal end of the aminoadipate reductase Lys2 of *Saccharomyces cerevisiae* and *Candida albicans* and the saframycin synthetase SafA of *Myxococcus xanthus* a reductase (Red) domain has been identified (Figures 2,3 and Table 2 dark violet regions; ~. 350 aa in length) containing a NAD(P)H-binding site R1 (Table 2) characteristic of a diverse group of reductive enzymes [48–50]. Sequence data and biochemical studies on Lys2 indicate that the Red domain catalyzes the reductive cleavage of the associated T-domain-tethered acyl group, releasing a linear aldehyde. In analogy to the reactions catalyzed by the C and Te domains, this reaction formally can be formulated as an acyl transfer of the T-domain-bound carboxy moiety to a hydride ion. The Red domain, as well as the C and Te domains, therefore might represent an alternative strategy for peptide-chain termination in nonribosomal peptide synthesis.

New PPS systems with an unusual domain organization

The increasing amount of sequence data on PPS genes in public databases continuously provides us with new interesting insights into the complexity of possible combinations of these modular systems. One of the most remarkable systems reported recently represents the biosynthetic cluster directing the nonribosomal assembling of yersiniabactin [41], a siderophore essentially involved in the pathogenesis of the bubonic-plague-causing organism *Yersinia pestis*. This system comprises three major enzymes (Table 1) — an aryl-AMP-ligase (YbtE), a PPS with an unusual domain structure (Irp2) and a mixed polyketide synthase (PKS)–PPS enzyme (Irp1) — and has a number of specific characteristics that are not found in other PPS systems. One of the most striking aspects of the system is that the Irp2 protein contains an A domain that is probably responsible for the aminoacylation of three different T domains (two at Irp2 and one at Irp1), two of which are not physically linked to the A domain. This finding raises the general question of how aminoacylation of PPS takes place. The common point of view is that the physical linkage of A and T domains within a PPS directs the intramolecular transfer of the activated amino acid from the A domain to the subsequent T domain of the same PPS. The other alternative could be that the PPS, in analogy to the type I PKS [51,52], possibly acts *in trans* as (homo-) dimers in which, for example, the A domain of one dimer aminoacylates the corresponding T domain of the other dimer. Such *in trans* activities, as mentioned above, are well known for aryl-AMP-ligases [43,53] and have also been demonstrated *in vitro* for recombinant PPS domains [25]. Future experiments will probably shed more light on the common interaction mechanisms of PPS.

Another recently reported siderophore biosynthetic cluster (*fxbB/C*, Table 2), responsible for the production of the

Figure 4



Acyl transfer reactions catalyzed by different PPS domains. **(a)** C domain, **(b)** Cy domain, **(c)** Te domain and **(d)** Red domain.

hydroxame siderophore exochelin of *Mycobacterium smegmatis*, contains further features that are to some extent 'unusual' for PPS [54]. Sequence analysis of the deduced amino acids of the *fxbB* and *fxbC* genes shows that there are a total of six modules, but the final secreted exochelin is a pentapeptide. To date it is unclear if an intermediate hexapeptide is formed that is later cleaved into the secreted exochelin or if the last module of FxbC is inactive and is

skipped during peptide synthesis. A further interesting aspect of this system is the distribution of C domains within the single PPS and the type of bonds formed by these domains. Only a carboxy-terminal C domain at FxbB is found at the junction of FxbB and FxbC. This finding might indicate that C domains are able to catalyze their specific reactions irrespective of their positioning relative to the protein terminus of a particular PPS. Three of the four bonds formed by the C domains are amide bonds not peptide bonds (Table 1), further extending the capable spectrum of reactions catalyzed by C domains.

The characterization of the genes encoding the syringomycin synthetases of *Pseudomonas syringae*, reported recently, revealed an additional uncommon architecture of PPS (Table 1) [55]. Here, the synthetase SyrE, responsible for the activation of the first eight residues of syringomycin, contains a CT domain at its carboxyl end inserted between the last module and the terminal Te domain. It is thought that, at this position, SyrE is aminoacylated *in trans* with the last constituent moiety threonine by the action of SyrB, a synthetase comprised of a single module. Support for this theory comes from the finding that similar domain structures (AT→CT→C/Te) are also found in the PPS systems of chloroeremomycin and ferrichrome (Table 1), although the functional meaning of this organization in these systems remains unclear.

These few examples of new PPS systems illustrate that we are just beginning to understand the rules and mechanisms directing the architecture of PPS and the complexity of reactions catalyzed by them. The discovery of new PPS genes and the biochemical potential of their encoded modules and domains will therefore provide us with new interesting insights into the molecular architecture of these fascinating enzymes.

Acknowledgements

The authors acknowledge generous funding from the Deutsche Forschungsgemeinschaft (DFG), the EC (project cell factories) and the Fonds der Chemischen Industrie.

References

1. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in non-ribosomal peptide synthesis. *Chem. Rev.* **97**, 2651-2673.
2. von Döhren, H., Keller, U., Vater, J. & Zocher, R. (1997). Multifunctional peptide synthetases. *Chem. Rev.* **97**, 2675-2705.
3. Kleinkauf, H. & von Dohren, H. (1990). Nonribosomal biosynthesis of peptide antibiotics. *Eur J Biochem* **192**, 1-15.
4. Lipmann, F., Gevers, W., Kleinkauf, H. & Roskoski, R., Jr. (1971). Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. *Adv. Enzymol. Relat. Areas Mol. Biol.* **35**, 1-34.
5. Lipmann, F. (1980). Bacterial production of antibiotic polypeptides by thiol-linked synthesis on protein templates. *Adv. Microb. Physiol.* **21**, 227-266.
6. Stein, T., *et al.* & Morris H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.* **271**, 15428-15435.
7. Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chem. Biol.* **4**, 561-567.
8. Cane, D.E., Walsh, C.T. & Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63-68.
9. Mootz, H.D. & Marahiel, M.A. (1997). Biosynthetic systems for nonribosomal peptide antibiotic assembly. *Curr. Opin. Chem. Biol.* **1**, 543-551.
10. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927-937.
11. Weber, G., Schörgendorfer, K., Schneider-Scherzer, E. & Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.* **26**, 120-125.
12. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125**, 3-14.
13. Gocht, M. & Marahiel, M.A. (1994). Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. *J. Bacteriol.* **176**, 2654-2662.
14. Saito, M., Hori, K., Kurotsu, T., Kanda, M. & Saito, Y. (1995). Three conserved glycine residues in valine activation of gramicidin S synthetase 2 from *Bacillus brevis*. *J. Biochem. (Tokyo)* **117**, 276-282.
15. Stein, T., *et al.* & Morris H.R. (1994). Detection of 4'-phosphopantetheine at the thioester binding site for L-valine of gramicidinS synthetase 2. *FEBS Lett.* **340**, 39-44.
16. Schlumbohm, W., *et al.*, & Wittmann-Liebold, B. (1991). An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem.* **266**, 23135-23141.
17. Billich, A., Zocher, R., Kleinkauf, H., Braun, D.G., Lavanchy, D. & Hochkeppel, H.K. (1987). Monoclonal antibodies to the multienzyme enniatin synthetase. Production and use in structural studies. *Biol Chem Hoppe Seyler* **368**, 521-529.
18. Turgay, K., Krause, M. & Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**, 529-546.
19. Baldwin, T.O. (1996). Firefly luciferase: the structure is known, but the mystery remains. *Structure* **4**, 223-228.
20. Mootz, H.D. & Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843-6850.
21. Konz, D., Doekel, S. & Marahiel, M.A. (1998). Molecular and biochemical characterization of the protein template controlling the biosynthesis of the lipopeptide lichenysin of *Bacillus licheniformis*. *J. Bacteriol.* **181**, 133-140.
22. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270**, 6163-6169.
23. Conti, E., Franks, N.P. & Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4**, 287-298.
24. Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174-4183.
25. Stachelhaus, T., Hüser, A. & Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.* **3**, 913-921.
26. Schlumbohm, W., Vater, J. & Kleinkauf, H. (1985). Reactive sulfhydryl groups involved in the aminoacyl adenylate activation reactions of the gramicidin S synthetase 2. *Biol. Chem. Hoppe Seyler* **366**, 925-930.
27. Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1996). Engineered biosynthesis of peptide antibiotics. *Biochem. Pharmacol.* **52**, 177-186.
28. Lambalot, R.H., *et al.*, & Walsh, C.T. (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923-936.
29. Walsh, C.T., Gehring, A.M., Weinreb, P.H., Luis, E.N. & Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthetases. *Curr. Opin. Chem. Biol.* **1**, 309-315.
30. Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. & Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595.
31. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. (1998). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631-645.

32. Weinreb, P.H., Cluadri, L.E., Walsh, CT. & Zuber, P. (1996). Stoichiometry and specificity of *in* vitro phosphopantetheinylation and aminoacylation of the vafine-activating module of surfactin synthetase. ***Biochemistry 37,* 1575-I 584.**

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

34. **Pieper,** R., Haese, A., **Schroder,** W. **&** Zocher, R. **(1995).** Arrangement of catalytic sites in the **multifunctional** enzyme enniatin synthetase. Eur. *J. Biochem. 230, 119-I* **26.**

35. Billich, A. **&** Zocher, R. (1987). Enzymatic synthesis of cyclosporin A. *J. Biol. Chem. 262,* **17258-I 7259.**

36. Stein, T., **Kluge,** B., Vater, J., Franke, P., Otto, A. **&** Wittmann-Liebold, B. **(1995).** Gramicidin **S** synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4'-phosphopantetheine. *Biochemistry 34, 4633-4842.*

37. Vater, J. **&** Kfeinkauf, H. (1978). Gramicidin S-synthetase. A further characterization of phenylalanine racemase. the light enzyme of gramicidin S-synthetaee. *Biochim. Biophys. Acta 429,* 1082-I 072.

38. Hoffmann, K., **Schneider-Scherzer,** E., Kleinkauf, H. **&** Zocher, R. (1994). Purification and characterization of eucaryotic **alanine** racemase acting as key enzyme in cyclosporin biosynthesis. *J. Biol. Chem.* 269, 12710.12714.

39. De Crecy-Lagard, V., Marliere, P. **&** Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing **peptide** elongation and epimerisation. C. R. *Acad. Sci. III* **318,** 927-936.

40. Stachelhaus, T., **Mootz,** H.D., Bergendahl, V. **&** Marahiel, MA. (1998). peptide-bond formation in **non-ribosomal peptide** biosynthesis: catalytic role of the condensation domain. *J. Biol.* Clrem. **273, 22773-22781.**

41. Gehring, A.M., et a/. **&** Walsh C.J. (1998). Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia *pestis. Chem. Biol. 5, 573-586.*

42. Reimmann, C., Serino, L., Beyeler, M. **&** Haas, D. (1998). Dihydroaeruginoic acid synthetase and pyochelin synthetase, products of the **pchEF** genes, are induced by extracellular pyochelin in *Pseudomonas aeruginosa. Microbiology* **144, 3135-3148.**

43. Gehring. A.M., Mori, **I.,** Perry, R.D. &Walsh, C.T. (1998). The nonribosomal **peptide** synthetase **HMWP2** forms a **thiazoline** ring during biogenesis of Yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis. Biochemistry 37,* 11637-I 1650.

44. Schneider, A. **&** Marahiel, **M.A.** (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with **peptide** synthetases, in non-ribosomal **peptide** biosynthesis in *Bacillus subtilis. Arch. Microbiol.* **169, 404-4** 10.

45. de Ferra, F., Rodriguez, F., **Tortora,** O., **Tosi,** C. **&** Grandi, G. (1997). Engineering of **peptide** synthetases. Key role of the thioesterase-like domain for efficient production of recombinant peptides. *J. Biol. Chem.* **272,** 25304-25309.

46. Cosmina, P., et *al.,* **&** van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis. Mol. Microbiol. 8,* **821-831.**

47. Kratzschmar, J., Krause, **M. &** Marahiel, M.A. (1989). Gramicidin **S** biosynthesis operon containing the structural genes **grsA** and **grsB** has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J Bacteriol 171, 5422-5429.*

48. Pospiech, A, Bietenhader, J. **&** Schupp, T. (1996). Two multifunctional **peptide** synthetases and an 0-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mxl from Myxococcus xanthus. *Microbiology 142,* **741-746.**

49. Morris, M.E. & Jinks-Robertson, **S.** (1991). Nucleotide sequence of the **LYS2** gene of Saccharomyces cerevisiae: homology to Bacillus brevis tyrocidine synthetase 1. *Gene 98,* **141-I 45.**

50. Suvarna, **K.,** Seah, L., Bhattacherjee, V. **&** Bhattacharjee, J.K. (1998). Molecular analysis of the **LYS2** *gene* of *Candida albicans:* homology to **peptide** antibiotic synthetases and the regulation of the **alpha**-aminoadipate reductase. *Curr. Genet. 33. 268-275.*

51. Kao, C.M., **Pieper,** R., Cane, D.E. **&** Khosla, C. (1996). Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. ***Biochemistry 35,*** 12363-I 2368.

52. Staunton, J., Caffrey, P., Aparicio, **J.F.,** Roberts, **G.A.,** Bethell, **S.S.** & Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. ***Nat. Struct. Biol. 3,* 188-I 92.**

53. Gehring, A.M., Bradley, K.A. &Walsh, C.T. (1997). Enterobactin biosynthesis in *Escherichia coli:* isochorismate lyase **(EntB)** is a bifunctional enzyme that is phosphopantetheinylated by **EntD** and then acylated by **EntE** using ATP and **2,3-dihydroxybenzoate.** *Biochemistry* **36, 8495-8503.**

54. Yu, **S.,** Fiss, E. &Jacobs, W.R., Jr. (1996). Analysis of the exochelin locus in *Mycobacferium smegmatis*: biosvnthesis oenes have homology-with genes of the **peptide** synthetase **family.** *J. Bacteriol.* **180, 4676-4685.**

55. Guenzi, E., **Galli,** G., Grgurina, I., Gross, DC. **&** Grandi, G. (1998). Characterization of the Syringomycin Synthetase Gene Cluster. A link between prokaryotic and eukaryotic **peptide** synthetases. *J. Biol.* Chem. **273, 32857-32863.**

58. Rowland, B.M., Grossman, T.H., Osburne, **M.S. & Taber,** H.W. (1996). Sequence and genetic organization of a *Bacillus subtilis* operon encoding **2,3-dihydroxybenzoate** biosynthetic enzymes. *Gene* 178, 119-123.

57. Kunst, F., et *a/.* (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtili*s. **Nature 390, 249-256.**

58. Smith, D.J., Earl, A.J. **&** Turner, G. (1990). The multifunctional **peptide** synthetase performing the first step of penicillin biosynthesis in *Penicillium chrysogenum* is a 421,073 dalton protein similar to *Bacillus brevis* **peptide** antibiotic synthetases. *EMBO J. 9. 2743-2750.*

59. van Wageningen, A.. et *a/.,* **&** Solenberg, P. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol. 5,* 155-I 62.

80. Rusnak, F., Faraci, W.S. &Walsh, C.T. (1989). Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2,3-dihydroxybenzoate-AMP ligase: demonstration of enzyme-bound (2,3-dihydroxybenzoyl)adenylate product. *Biochemistry 28, 6827-6835.*

61. Rusnak, F., Sakaitani, M., Drueckhammer, *D.,* Reichert, J. **&** Walsh, CT. **(1991).** Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *entF* gene, expression and purification of **EntF,** and analysis of covalent phosphopantetheine. *Biochemistry* **30, 2916-2927.**

62. Lin, G.H., Chen, CL.. Tschen, **J.S.,** Tsay, **S.S.,** Chang, Y.S. **&** Liu, **S.T.** (1998). Molecular cloning and characterization of fengycin synthetase **gene *fen8* from *Bacillus subtilis. J. Bacteriol.* 180, 1338-I 34 1.**

63. Krätzschmar, J., Krause, M. **&** Marahiel, M.A. (1989). Gramicidin **S** biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J. Bacteriol.* **171, 5422-5429.**

64. Scott-Craig, J.S., Panaccione, D.G., Pocard, J.A. & Walton, J.D. (1992). The cyclic **peptide** synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a **15.7-kilobase** open reading frame. *J. Biol. Chem. 267.* **26044-26049.**

65. de Crecy-Lagard, V., et *al.,* **&** Thibaut, D. (1997). Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural **peptide** synthetase genes. *J. Bacteriol.* **179, 705-7 13.**

66. Zhang, J.H., **Quigley,** N.B. **&** Gross, D.C. (1995). Analysis of the *syrB* and *syrC* genes of *Pseudomonas syringae pv. svrinoae* indicates that syringomycin is synthesized by a **thiotemplate** mechanism. *J. Bacteriol.* 177, 4009-4020.

67. Guilvout, I., **Mercereau-Pauijalon,** O., Bonnefoy, **S.,** Pugsley, A.P. **&** Carniel, E. (1993). High-Molecular-Weight Protein 2 of *Yersinia enterocolitica* is homologous to **AngR** of *Vibrio anguillarum* and belongs to a family of proteins in nonribosomal **peptide** synthesis. *J. Bacteriol.* **175, 5488-5504.**

68. Pelludat, C., Rakin, A., Jacobi, C., Schubert, **S. &** Heesemann, J. (1998). The yersiniabactin biosynthetic gene cluster of *Yersinia enterocolitica*: organization and siderophore-dependent regulation. *J. Bacteriol.* **180, 538-546.**