The EMBO Journal Vol.18 No.23 pp.6823–6831, 1999

Crystal structure of the surfactin synthetase-activating enzyme Sfp: a prototype of the 4′-phosphopantetheinyl transferase superfamily

Klaus Reuter, Mohammad R.Mofid[1], Mohamed A.Marahiel[1] and Ralf Ficner[2]

Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Emil-Mannkopff-Strasse 2, D-35037 Marburg and [1]Institut für Biochemie, Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, D-35043 Marburg, Germany

[2]Corresponding author
e-mail: ficner@imt.uni-marburg.de

The *Bacillus subtilis* Sfp protein activates the peptidyl carrier protein (PCP) domains of surfactin synthetase by transferring the 4′-phosphopantetheinyl moiety of coenzyme A (CoA) to a serine residue conserved in all PCPs. Its wide PCP substrate spectrum renders Sfp a biotechnologically valuable enzyme for use in combinatorial non-ribosomal peptide synthesis. The structure of the Sfp–CoA complex determined at 1.8 Å resolution reveals a novel α/β-fold exhibiting an unexpected intramolecular 2-fold pseudosymmetry. This suggests a similar fold and dimerization mode for the homodimeric phosphopantetheinyl transferases such as acyl carrier protein synthase. The active site of Sfp accommodates a magnesium ion, which is complexed by the CoA pyrophosphate, the side chains of three acidic amino acids and one water molecule. CoA is bound in a fashion that differs in many aspects from all known CoA–protein complex structures. The structure reveals regions likely to be involved in the interaction with the PCP substrate.

Keywords: coenzyme A/peptide synthetases/phosphopantetheinyl transferase/three-dimensional structure/X-ray crystallography

Introduction

A large number of bioactive oligopeptides are produced by bacteria and fungi via a unique non-ribosomal mechanism. The composition of these low molecular weight peptides is not restricted to the proteinogenic amino acids but covers many unusual residues such as α-hydroxy acids, N-methylated and D-amino acids. The resulting structural diversity is increased further by the occurrence of modifications downstream including acylation, heterocyclic ring formation, glycosylation, lipoylation and cyclization of the peptide backbone. It is not surprising that this class of secondary metabolites comprises versatile pharmacologically active agents exhibiting antibiotic, antiviral, antitumor, cytostatic or immunosuppressive effects. The biosynthesis of these low molecular weight peptides is catalyzed by large modular enzymes referred to as non-ribosomal peptide synthetases. Each module of a peptide synthetase is composed of different domains and is responsible for the incorporation and modification of one specific amino acid into the peptide. The sequence of modules is co-linear with the sequence of the peptide product (Stachelhaus and Marahiel, 1995; Kleinkauf and Von Dohren, 1996; Dieckmann *et al*., 1999; Konz and Marahiel, 1999). The minimal set necessary to build up a module comprises three domains. (i) The adenylation domain catalyzes the formation of an aminoacyl-adenosine monophosphate from the cognate amino acid and ATP, releasing pyrophosphate (Turgay *et al*., 1992; Dieckmann *et al*., 1995; Conti *et al*., 1997; Stachelhaus *et al*., 1999). (ii) The adjacent peptidyl carrier protein (PCP) domain contains a 4′-phosphopantetheinyl prosthetic group, to which the activated amino acid is then covalently bound in thioester linkage (Stachelhaus *et al*., 1996a; Stein *et al*., 1996). (iii) The condensation domain finally mediates the direct condensation of the thioesterified intermediate in the growing chain (Stachelhaus *et al*., 1998). It accepts the acyl group of the preceding module and is therefore missing in the module activating the first amino acid of the peptide. To become catalytically active, the PCP domains of peptide synthetases have to be converted from inactive apo- to cofactor-containing holo-forms by specific phosphopantetheinyl transferases (Lambalot *et al*., 1996). These enzymes transfer the 4′-phosphopantetheinyl moiety of coenzyme A (CoA) to the side chain hydroxyl of a serine residue invariant in all PCP domains (Walsh *et al*., 1997).

The production of the lipoheptapeptide antibiotic surfactin in *Bacillus subtilis* depends on the phosphopantetheinyl transferase Sfp (Nakano *et al*., 1992), which converts the inactive apo-forms of the seven PCP domains of surfactin synthetase (SfrA-ABC) to their active holo-forms (Lambalot *et al*., 1996). Sfp is likely to play an important role in biotechnology in the future. Encouraged by the strictly modular structure of non-ribosomal peptide synthetases, efforts are being made to combine subunits from naturally occurring enzymes in order to create artificial proteins capable of synthesizing completely novel clinically or agrochemically useful oligopeptide metabolites *in vivo* (Stachelhaus *et al*., 1995, 1996b; Cane *et al*., 1998; Mootz and Marahiel, 1999). The recombinant expression of combinatorial peptide synthetases composed of modules derived from various enzymes, which may moreover originate from different organisms, requires, however, the co-expression of a phosphopantetheinyl transferase with a PCP substrate recognition spectrum as broad as possible. Among all phosphopantetheinyl transferases characterized so far, Sfp is the one which best fulfills this criterion. Beyond its natural substrates, Sfp was shown to phosphopantetheinylate efficiently the apo-PCP domains from enterobactin synthetase (Lambalot *et al*., 1996), bacitracin, gramicidin S and tyrocidine synthetases (M.R.Mofid and M.A.Marahiel, unpublished) and even of the acyl carrier protein (ACP) domains or subunits from fatty acid

Table I. Crystallographic data

Data collection	Native	SeMet-λ1	SeMet-λ2
Crystal			
Resolution (Å)	25–1.8	100–2.1	100–2.1
Wavelength (Å)	0.9537	0.9777	0.9782
No. of observed reflections	96 195	141 818	96 772
No. of unique reflections	28 951	19 493	19 393
Completeness	93.2	99.2	99.1
R_{sym}[a]	0.067	0.074	0.074
Phasing			
PhP[b] centric (iso)		1.87	–
PhP acentric (iso/ano)		2.84/2.86	–/2.38
Mean FOM[c] (centric)	0.41		
Mean FOM (acentric)	0.57		
Refinement			
R-factor[d]	21.3%		
R_{free}[e]	27.0 %		
Deviations from ideal geometry			
Bond length (Å)	0.012		
Bond angles (°)	1.51		

[a]$R_{sym} = \Sigma |I - <I>|/\Sigma I$, where I is the observed intensity and $<I>$ is the average intensity for multiple measurements.
[b]PhP, phasing power is the root-mean-square ($|F_h|/E$), where $|F_h|$ is the heavy atom structure factor amplitude and E is the residual lack of closure error.
[c]FOM, Figure of merit.
[d]*R*-factor = $\Sigma|F_o - F_c|/ \Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[e]R_{free} is the cross-validation *R*-factor calculated for 5% of the reflections omitted in the refinement process.

synthases (Lambalot *et al*., 1996) and polyketide synthases (Kealey *et al*., 1998). This makes Sfp the leading candidate for co-expression with artificial peptide synthetases, and it is already widely used for biotechnological studies on these enzymes. We have determined the crystal structure of Sfp in complex with its substrate CoA. The structure presented reveals regions within Sfp that may act as PCP-binding sites. Hence it provides a framework for mutational and biochemical studies to investigate the interaction of this phosphopantetheinyl transferase with its protein substrates, which will help to understand its broad PCP specificity.

Beyond its potential biotechnological significance, Sfp belongs to an enzyme superfamily, whose members are not involved solely in the maturation of peptide synthetases. In addition, this superfamily subsumes phosphopantetheinyl transferases, which are necessary for heterocyst formation in cyanobacteria (Black and Wolk, 1994) and lysine biosynthesis in yeast (Ehmann *et al*., 1999), or which endow the ACP domains and subunits of fatty acid synthases and probably polyketide synthases with a 4′-phosphopantetheinyl prosthetic group (Lambalot and Walsh, 1995; Stuible *et al*., 1997, 1998). Here it serves in analogy to peptide synthetases as a swinging arm to hand the growing acyl chain from one catalytic center of these multienzyme systems to the next. Hence the structure presented serves as a prototype allowing conclusions to be made on the topology and substrate recognition of related enzymes, which are prerequisites for the biosyntheses of further biotechnologically important metabolites, namely polyketides, and for a biological process as fundamental as fatty acid biosynthesis.

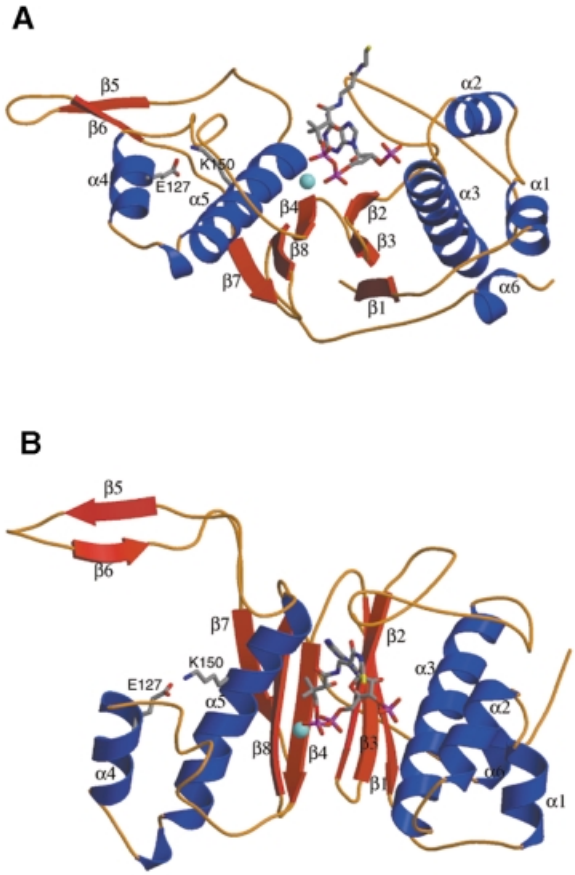


Fig. 1. Ribbon representation of Sfp in two perpendicular views. The protein displays a pseudo 2-fold symmetry dividing it into two halves of roughly identical size. The interface of both halves forms the binding site for CoA, which is depicted as a stick model. The conformation of the pantetheinyl moiety of CoA in this Figure is arbitrary, since only its two atoms nearest to the pyrophosphate are clearly visible in the electron density map. The bright blue sphere represents an Mg^{2+} ion present in the active site. The invariant Lys150 located on helix α5 forms a salt bridge to Glu127 on helix α4. The view of Sfp in (**B**) is generated from view (**A**) by a rotation of ~90° about a horizontal axis parallel to the plane of the paper.

Results and discussion

Structure determination

We solved the crystal structure of the monomeric Sfp in complex with CoA by means of multiple wavelength anomalous diffraction (MAD) using a selenomethionine derivative (Hendrickson *et al*., 1990). The final model was refined at 1.8 Å resolution with good stereochemistry to an *R*-factor of 21.3% and an R_{free} of 27.0% (Table I). The protein used for crystallization was recombinant full-length Sfp endowed with a C-terminal His_6 tag spaced by two amino acids (Arg–Ser) from the original protein (Mofid *et al*., 1999). While the four C-terminal histidines are flexible in the crystal and therefore not visible in the electron density map, the additional arginine is well defined and included in the model. For the serine residue as well as for the two subsequent histidine residues, only weak electron density is present. These residues were modeled as alanines.

Three-dimensional structure of Sfp

To the best of our knowledge, Sfp represents a completely new protein fold. A structural homology search using the

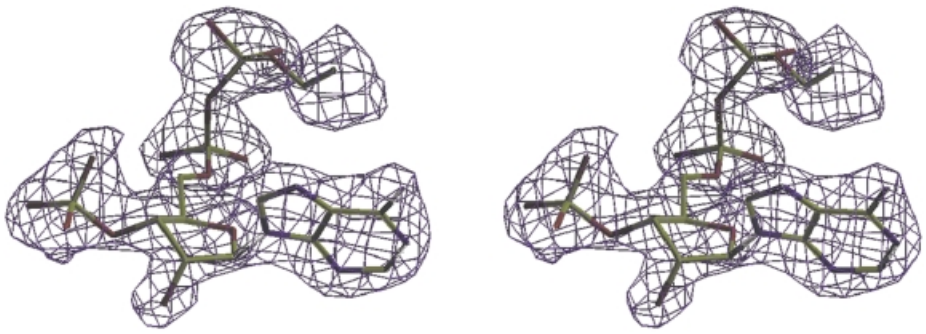


Fig. 2. Stereo view of the experimental MAD electron density map covering CoA after solvent flattening contoured at 1.7 σ. The ribose is clearly present in a 3′-endo conformation leading to the axial orientation of the 2′-hydroxyl and the horizontal orientation of the 3′-phosphate group. Electron density can only be attributed for the two atoms of the pantetheinyl moiety nearest to the pyrophosphate.

program DALI (Holm and Sander, 1993) did not reveal a protein structure with significant similarity to Sfp. The structure of Sfp shows a pseudo 2-fold symmetry dividing the molecule into two similarly folded halves of roughly identical size (Figure 1). The N-terminal half extends from Met1 to Pro103, the C-terminal half from Ile104 to Pro209. The remaining C-terminal amino acids from Asp210 to Leu224 have no counterpart in the N-terminal half. Each half consists of a three-stranded antiparallel β-sheet and a long α-helix diagonally packing against the β-sheet. In both halves, this α-helix is separated from the β-sheet by two extended loops, the N-terminal loop containing additional short α-helical elements. The β-sheets of both halves of Sfp are arranged in a barrel-like structure. The C-terminal residues 210–217, which belong to the stretch of amino acids without any counterpart in the N-terminal half, close a gap within this pseudo barrel. The most striking difference between both halves is found in the loops C-terminally following the long α-helices. While loop α3–β2 clings to the protein body, the corresponding loop α5–β7 in the C-terminal half, which includes a small two-stranded antiparallel β-sheet (β6 and β7), protrudes from the rest of the protein. Indeed this conformation of 'loop' α5–β7 may not represent the situation in solution, since in the crystal this region is kept in position through extensive interactions with the same loop of a symmetry-related Sfp molecule forming a four-stranded β-sheet. In solution, however, Sfp shows no tendency to dimerize but is present as a monomer (Mofid *et al*., 1999). It is fair to speculate, that the extensive protein–protein interaction caused by the β-strands of the 'loop' α5–β7 in the crystal may reflect its involvement in binding the substrate PCP domain *in vivo*.

The substrate CoA is bound in a pocket formed by the interface of the two Sfp halves (Figure 1). Amino acids framing the CoA-binding pocket may be involved in binding of the PCP substrate, which may occur in concert with the region between helix α5 and strand β7.

Implications of the Sfp structure for the structures of other phosphopantetheinyl transferases

The attachment of the phosphopantetheinyl residue derived from CoA is required not only for the PCP domains of peptide synthetases but also for the ACP domains or subunits of fatty acid synthases, polyketide synthases and further proteins (Lambalot *et al*., 1996). Accordingly, there exist a vast number of phosphopantetheinyl transferases. They comprise an enzyme superfamily, whose members, however, show a surprising variability in size and low sequence homology with each other. This superfamily can be divided into two families. The first family subsumes enzymes involved in cyanobacterial heterocyst differentiation (Black and Wolk, 1994) and fungal lysine biosynthesis (Ehmann *et al*., 1999), the gene product of *HI0152* probably modifying the ACP subunit of a fatty acid synthase (Lambalot *et al*., 1996), enzymes modifying the PCP subunits of peptide synthetases and enzymes whose exact functions have not yet been elucidated. They consist on average of ~230 amino acids. Like Sfp, they are probably all present in a monomeric form. In the following, we will refer to members of this family as 'Sfp-type' phosphopantetheinyl transferases. Most enzymes that endow the ACP domains or subunits of fatty acid synthases with a phosphopantetheinyl arm belong to the second family whose members are about half the size of Sfp-type enzymes, form homodimers (Lambalot and Walsh, 1995) and show weak sequence homology to the C-terminal half of Sfp. Since holo acyl carrier protein synthase (ACPS) from *Escherichia coli* was the first protein of this family to be discovered (Lambalot and Walsh, 1995), we will refer to its members as 'ACPS-type' phosphopantetheinyl transferases.

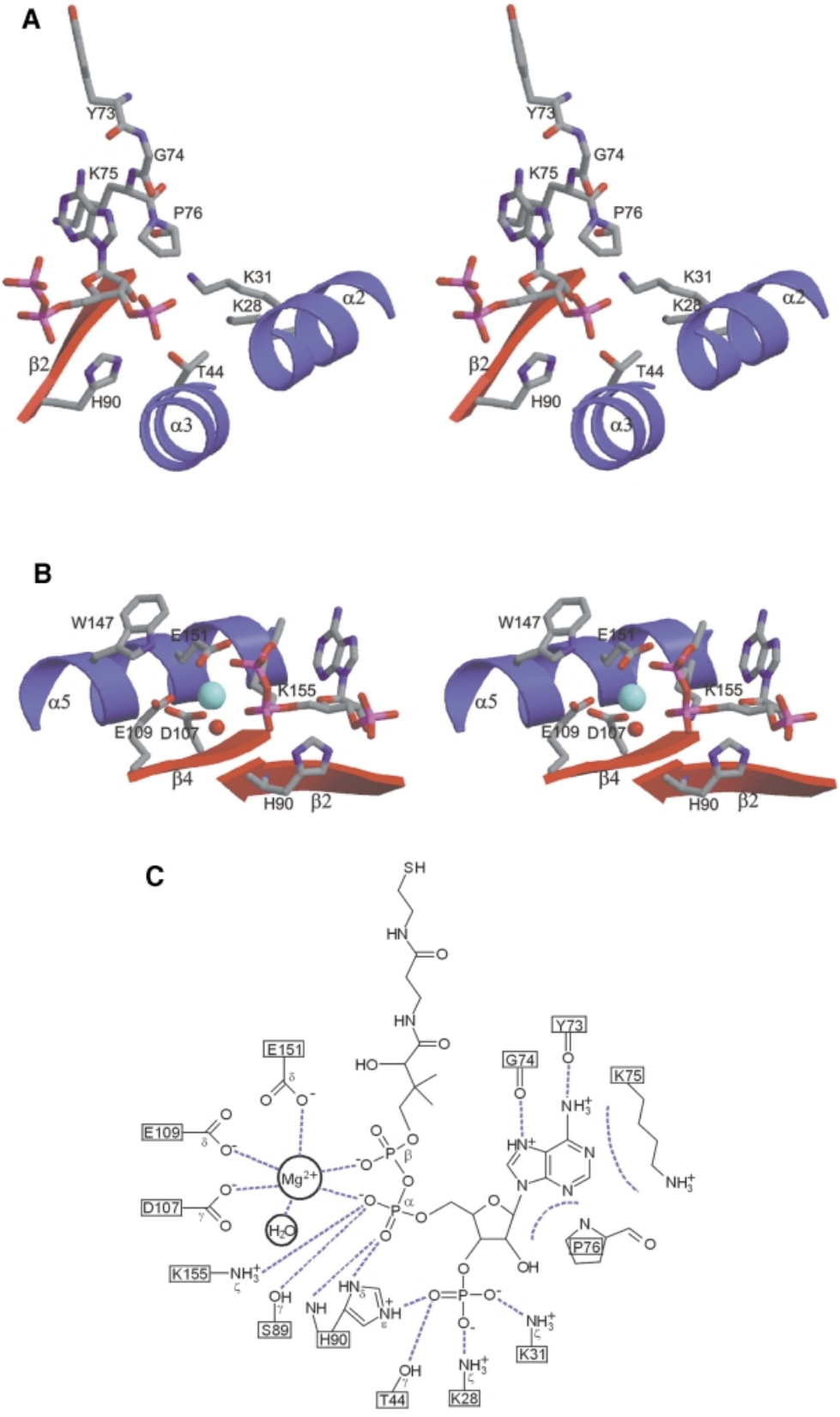
The pseudo 2-fold architecture of Sfp suggests a dimerization mode for the phosphopantetheinyl transferases of the ACPS type. It is likely that monomers of this family fold in a manner similar to one half of Sfp. Dimerization may occur via formation of a β-barrel-like structure similar to that present in Sfp, and the CoA-binding pocket may be formed at the intermolecular interface of the homodimer. In yeast cytoplasmic fatty acid synthase, the 4′-phosphopantetheinyl transferase is an integral domain of the fatty acid synthase 2 (FAS2) subunit. The size of this domain corresponds to one *E.coli* ACPS monomer (Lambalot *et al*., 1996). Since yeast fatty acid synthase has to form a homodimer to become functionally active, the structure of Sfp suggests that dimerization is necessary for the formation of an intact phosphopantetheinyl transferase.

Mode of CoA binding and structural analysis of the active center

CoA is bound by Sfp in a bent conformation within a pocket whose bottom is formed by residues from strands β2 and β4. The pocket is lined by the C-terminal part of

loop α3–β2, helix α2, the N-terminus of helix α3, the C-terminus of helix α5 and the N-terminal part of loop α5–β5 (Figure 1).

No defined conformation is observed for the pantetheinyl residue of CoA in the electron density map apart from the two atoms next to the pyrophosphate (Figure 2). The main part of the pantothenic acid as well as the β-mercaptoethylamine are flexible in the crystal, since they rise from the CoA-binding pocket into the solvent and make no interactions with Sfp. This agrees well with



the finding that the SH-group of the β-mercaptoethylamine moiety may be removed or substituted in various ways without affecting the capacity of the modified CoA to bind to Sfp (Quadri *et al*., 1998).

The 3′-phospho-5′-ADP moiety of CoA is well defined in the electron density map. The adenine base of CoA is held in place by a motif formed by amino acids Tyr73, Gly74, Lys75 and Pro76, where it stacks against the side chain of Lys75 (Figure 3A and C). Its amino group forms a hydrogen bond with the main chain carbonyl of Tyr73. The five rings of Pro76 fit well into the knee formed by the adenine base and the ribose moiety of CoA. As observed in all other protein–CoA complex structures determined so far, the adenine is bound to the ribose moiety with an *anti* glycosidic torsion angle (Engel and Wierenga, 1996; Modis and Wierenga, 1998; Clements *et al*., 1999; Hickman *et al*., 1999). The ribose is present in a 3′-endo conformation resulting in the axial orientation of the 2′-hydroxyl and the equatorial orientation of the 3′-phosphate group (Figure 2). This is in contrast to all other known structures of protein-bound CoA, where the ribose moiety is always found in a 2′-endo conformation, with the exception only of succinyl-CoA synthetase where it is present in a 1′-exo conformation (Fraser *et al*., 1999).

The 3′-phosphate is bound tightly to the protein through a number of interactions including a hydrogen bond to the side chain hydroxyl of Thr44 as well as salt bridges to the ζ-amino groups of Lys28 and Lys31 and to the ε-nitrogen of the imidazole group of His90 (Figure 3A and C). This is in good agreement with the finding that the presence of the 3′-phosphate group is indispensable for CoA binding (Quadri *et al*., 1998). In addition to the 3′-phosphate, His90 binds the α-phosphate of the pyrophosphate through a salt bridge via the δ-imidazole nitrogen and a hydrogen bond via the main chain amide. The involvement of the His90 imidazole in CoA binding probably contributes to the strict pH dependence of Sfp activity (Quadri *et al*., 1998). The α-phosphate of the pyrophosphate is salt bridged further by the ζ-amino group of Lys155 and hydrogen bonded by the side chain hydroxyl of Ser89 (Figure 3B and C). The orientation of the pyrophosphate leads to a sharp bend in the CoA molecule.

Both α- and β-phosphates of the pyrophosphate participate in the complexation of a metal ion whose additional ligands are the carboxylate groups of Asp107, Glu109 and Glu151 and one water molecule (Figure 3B and C). Asp107 was shown previously by mutagenesis experiments to be essential for catalysis (Quadri *et al*., 1998). The metal ion seen in this complex probably represents magnesium, since it was shown that phosphopantetheinyl transferase activity strictly depends on Mg^{2+} (Lambalot and Walsh, 1997; Stuible *et al*., 1997). It is tempting to speculate that this Mg^{2+} ion is involved in catalysis rather than merely in cofactor binding. It seems likely that during catalysis, the water molecule involved in Mg^{2+} complexation will be replaced by the hydroxyl group of the invariant serine residue of a substrate PCP. This hydroxyl group may have to be deprotonated to serve as a nucleophile for the β-phosphate of the CoA pyrophosphate. In order to prevent hydrolysis of the CoA pyrophosphate in the absence of a protein substrate, the deprotonating base may have to be moved to the catalytic center by a conformational change of Sfp upon PCP binding, or be provided by the PCP domain. Due to a lateral gap in the CoA-binding pocket, the β-phosphate is easily accessible for nucleophilic attack through the hydroxyl group of the invariant PCP serine (Figure 4). Without structural information on an Sfp–PCP complex, however, it is impossible to make specific predictions about the catalytic mechanism or the catalytic role of particular side chains. Efforts to prepare Sfp–PCP co-crystals are presently underway.

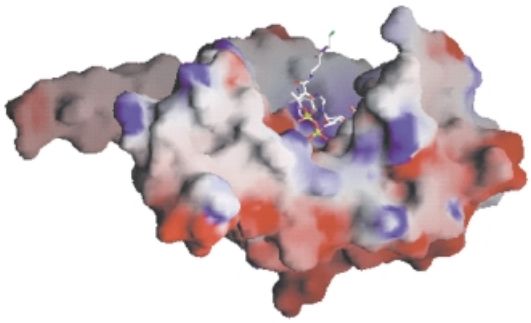


Fig. 4. Surface representation of SFP in a similar orientation to that in Figure 1A. Red represents a negative, and blue a positive electrostatic surface potential.

Sequences of phosphopantetheinyl transferases and the function of conserved residues

Figure 5A shows the sequence alignment of the five Sfp-type phosphopantetheinyl transferases for whom enzymatic activity has been proven directly or indirectly so far (Lambalot *et al*., 1996; Ehmann *et al*., 1999). The alignment shows that the region between helix α5 and strand β7, which we propose to be involved in substrate PCP binding, is largely missing in o195, an enzyme whose exact physiological role is still unclear. Although phosphopantetheinyl transferase activity could be shown for o195, all phosphopantetheinyl acceptor proteins that have been offered to this enzyme were transferred only inefficiently to their holo forms (Lambalot *et al*., 1996). Likewise, in EntD, a phosphopantetheinyl transferase with a reduced protein substrate spectrum compared with Sfp (Lambalot *et al*., 1996), this region is shortened. This suggests that the region between helix α5 and strand β7 may be responsible at least in part for the wide protein

Fig. 3. Sfp–CoA interactions. (**A**) Stereo view of contacts of the adenine base and 3′-phosphate of CoA to Sfp. CoA as well as selected residues are shown in stick representation. For the sake of clarity, the pantetheinyl moiety of CoA is omitted completely. Apart from the stretch of amino acids Tyr73–Pro76, which form the adenine base-binding motif, only the side chains of amino acids attached to the corresponding secondary structural elements are shown. (**B**) Stereo view of Sfp-bound CoA displaying interactions of the CoA pyrophosphate. The mode of representation is as in (A). Main chain atoms are shown only for His90, as its main chain amide hydrogen-bonds the CoA α-phosphate. The bright blue sphere represents an Mg^{2+} ion complexed by the side chain carboxylates of Asp107, Glu109, Glu151, one water molecule (represented by a red sphere) and the pyrophosphate of CoA. The pantetheinyl moiety of CoA has been largely omitted. Trp147 shields the active center from numerous hydrophobic amino acid side chains, which are not shown for the sake of clarity. (**C**) Schematic overview of Sfp–CoA contacts. Hydrophilic interactions are indicated by straight dotted lines. Van der Waals interactions are represented by curved dotted lines.

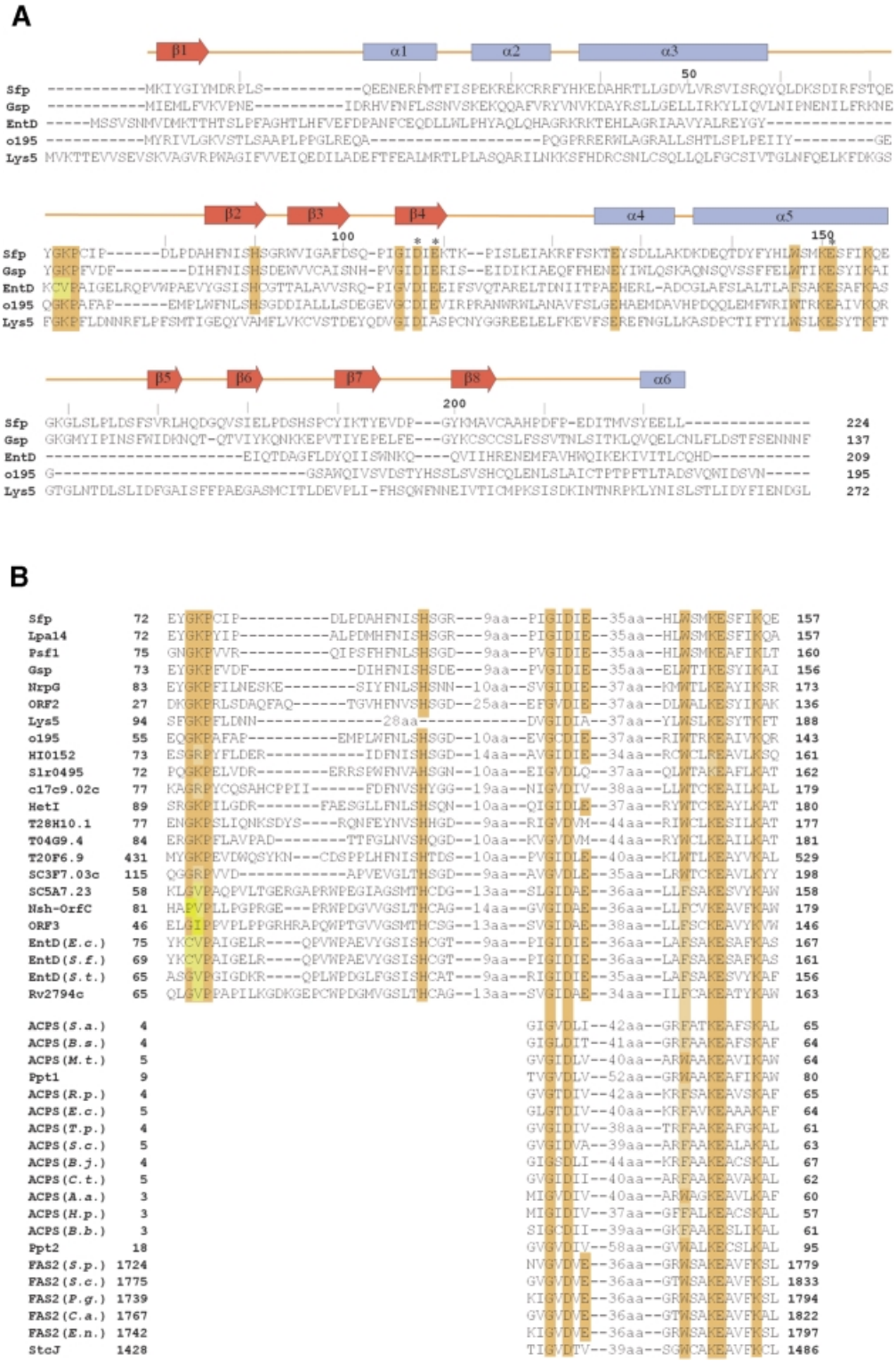


Fig. 5. (**A**) Sequence alignment of the five Sfp-type phosphopantetheinyl transferases for which enzymatic activity has been proven so far. Invariant residues as well as highly conserved residues are shaded in ocher. Residues occupying conserved positions but deviating from the consensus amino acid are shaded in brighter colors. The three acidic residues of Sfp involved in Mg^{2+} complexation are marked by an asterisk (*). The sequence alignment was performed using CLUSTALW (Higgins *et al*., 1996) and manually readjusted taking into account secondary structural features predicted for Gsp, EntD, o195 and Lys5 by the program PHDsec (Rost and Sander, 1994). Nevertheless, due to the low homology of the phosphopantetheinyl transferases shown, the alignment is only tentative at some positions. The Sfp secondary structure was assigned according to PROCHECK (Laskowski *et al*., 1993). (**B**) Partial sequence alignment of 45 putative and proven phosphopantetheinyl transferases of the Sfp-type (top) and ACPS-type (bottom). Invariant and conserved residues are indicated as in (A).

substrate spectrum of Sfp by recognizing more structural features of the phosphopantetheinyl acceptor rather than specific amino acid side chains.

Surprisingly, Figure 5A shows that only two of the three acidic amino acids involved in Mg^{2+} complexation are strictly conserved among the five enzymes compared in this alignment. The role of Glu109, which is missing in Lys5, must therefore be taken over by another residue or, although less likely, Mg^{2+} complexation may not take place in this enzyme. To determine, in view of this, whether Lys5 is an exception, we performed a homology search in the sequence databases using the five sequences of the alignment in Figure 5A plus the sequence of *E.coli* ACPS as probes. We found another 37 sequences of putative phosphopantetheinyl transferases, among them 16 that had already been identified in a previous search performed by Lambalot *et al*. (1996). Eighteen of the 37 sequences represented enzymes of the Sfp-type, while the remaining 19 sequences could be allocated to the ACPS-type family, including the recently characterized Ppt1 (Stuible *et al*., 1997) and Ppt2 (Stuible *et al*., 1998) proteins. Figure 5B shows a partial alignment of all these presently available sequences of potential and proven phosphopantetheinyl transferases. It demonstrates that Glu109 is missing in another four Sfp-like enzymes and in all members of the ACPS-type family except the subfamily of the FAS2-ACPS domains.

In addition to the Mg^{2+}-liganding residues Asp107 and Glu151 and the CoA α-phosphate-binding Lys155, two residues not involved in CoA binding, namely Gly105 and Lys150, are invariant in the alignment shown in Figure 5B (with the exception only of HetI, where Lys150 is replaced by Arg; see also Lambalot *et al*., 1996). While the residue in position 105 is confined to a glycine solely for spatial reasons, Lys150 salt-bridges via its ζ-amino group the side chain carboxylate of Glu127 (Figure 1), a residue obviously invariant among the five phosphopantetheinyl transferases whose sequences are aligned in Figure 5B. Due to the very low conservation of the amino acids flanking Glu127, however, it is impossible to determine whether this residue is invariant among all phosphopantetheinyl transferases (data not shown).

Trp147 is another residue for which no interaction with CoA is observed in the presented structure, but which is highly conserved. Apart from a tryptophan, the corresponding position can only be occupied by a phenylalanine in all sequences used for the alignment in Figure 5B. In Sfp, the Trp147 side chain, whose position and orientation within the structure are shown in Figure 3B, shields the active center from numerous hydrophobic residues by van der Waals interactions with the side chains of Ile118, Phe122, Phe123, Phe143, Tyr144 and the aliphatic part of the Lys150 side chain.

After the importance of His90 in CoA binding due to its interaction with both the 3′- and the α-phosphate had been elucidated, this residue was found, through manual readjustment of the original computer-generated sequence alignment, to be highly conserved among the phosphopantetheinyl transferases of the Sfp-type. Among all investigated sequences of this type, Lys5 was the only protein where this histidine was not conserved. Similarly, the sequence motif Gly74–Lys75–Pro76, which is involved in binding the adenine base of CoA (Figure 3A), was found to be highly conserved among the Sfp-type sequences. While the proline of this motif is invariant, the lysine residue, whose side chain stacks against the adenine base, may be replaced by an arginine, a valine or an isoleucine. The stacking interaction may well be taken over by the side chains of any of these amino acids. Since the ζ-amino group of Lys75 is not involved in the formation of a salt bridge but only forms hydrogen bonds to the main chain carbonyls of Ile104, Lys155 and Gln156, the presence of a basic residue in this position seems not to be mandatory, explaining the occurrence of valine or isoleucine. The glycine within this motif, which makes no direct interactions with the adenine base, is in rare cases replaced by a proline or a cysteine.

Conclusions

The structure of Sfp supports all genetic and biochemical data so far available for Sfp and other phosphopantetheinyl transferases. In particular, it explains the ability of Sfp to misload peptide synthetases with modified 4′-phosphopantetheinyl moieties of CoA analogs, a fact that recently allowed the investigation of condensation domain specificity in peptide synthetases (Belshaw *et al*., 1999). Interestingly, several residues of Sfp that are involved in CoA interactions are not conserved in other phosphopantetheinyl transferases, which suggests that the mode of CoA binding may vary somewhat between the members of this superfamily. The structure provides an explanation as to why *E.coli* ACPS and the ACPS domain of yeast fatty acid synthase have to form dimers in order to be functional. To understand the interactions of Sfp and the PCP substrate in detail and to trace back a catalytic mechanism, determination of the structure of an Sfp–PCP complex will be required. Nevertheless, the structure presented here already suggests regions within the protein that are candidates for mutational studies to address the question of how PCP recognition and specificity are achieved. The outcome of such mutagenesis experiments may help to optimize the efficiency of interaction between Sfp and non-cognate PCP domains of combinatorial peptide synthetases.

Materials and methods

Crystallization and X-ray data collection

Recombinant Sfp and its selenomethione derivative were purified and crystallized as described (Mofid *et al*., 1999). X-ray data of a selenomethionylated Sfp crystal were collected at two wavelengths (λ_1 = 0.9777 Å, λ_2 = 0.9782 Å) at beamline 5.2 R of the ELETTRA synchrotron (Trieste). A complete data set of a native crystal at 1.8 Å resolution was collected at the EMBL Hamburg outstation beamline BW7A at the DESY synchrotron. Data were processed using DENZO and SCALEPACK (Otwinowski and Minor, 1997). The crystals belong to space group $P4_32_12$ with cell constants $a = b = 64.8$ Å, $c = 150.4$ Å and $a = b = 65.3$ Å, $c = 150.3$ Å for the selenomethionine and native crystal, respectively, and contain one molecule in the asymmetric unit.

Structure determination and refinement

The expected six selenium sites in the asymmetric unit were found in the anomalous Patterson map using the program VERIFY (S.Roderick, unpublished). The selenium atom parameters were refined and phases calculated using SHARP (De La Fortelle and Bricogne, 1997) and improved by solvent flattening using SOLOMON (Abrahams and Leslie, 1996). Model building into the electron density map was done with the program O (Jones and Kjeldgaard, 1997) and the structure was refined using X-PLOR (Brünger, 1993). The model was improved manually and water molecules were built with the help of the program ARP (Lamzin and Wilson, 1997). The final model contains 228 amino acids and 272

water molecules, one molecule of CoA and one Mg^{2+} ion. The Sfp coordinates have been deposited with the Protein Data Bank and will be released upon publication (PDB ID code 1QR0). Figures 1, 3A and B were created using the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merrit and Bacon, 1997). Figure 2 was created using program O (Jones and Kjeldgaard, 1997) and Figure 4 using GRASP (Nicholls *et al*., 1991).

Sequence alignment of proven and putative phosphopantetheinyl transferases

The sequence alignments in Figure 5 were carried out as described in the Figure legends using the following sequences (DDBJ/EMBL/GenBank accession Nos in parentheses): Sfp-type, Sfp (X63158), Lpa14 (D21876), Psf 1 (SWISS-PROT: P55810), Gsp (X76434), NrpG (U46488), ORF2 (U73935), Lys5 (U32586), o195 (U00039), HI0152 (U32701), Slr0495 (D64001), c17c9.02c (Z73099), HetI (L22883), T28H10.1 (Z75551), T04G9.4 (U41247), T20F6.9 (AC002521), SC3F7.03c (AL021409), SC5A7.23 (AL031107), Nsh-OrfC (U75434), ORF3 (AB007189), *Shigella flexneri* EntD (U52684), *E.coli* EntD (D90700), *Salmonella typhimurium* EntD (U52686), Rv2794c (AL008967); ACPS-type, *E.coli* ACPS (AE000343), *Bradyrhizobium japonicum* ACPS (AF065159), *Rickettsia prowazekii* ACPS (AJ235272), *Lactobacillus plantarum* ACPS (Y08941), *Staphylococcus aureus* ACPS (Y16431); *Treponema pallidum* ACPS (AE001253), *Chlamydia trachomatis* ACPS (AE001284), *Aquifex aeolicus* ACPS (AE000708), *Streptomyces coelicolor* ACPS (AL031317), *Helicobacter pylori* ACPS (AE000592), *Bacillus subtilis* ACPS (Z99106), *Mycobacterium tuberculosis* ACPS (AL021185), *Borrelia burgdorferi* ACPS (AE001115), Ppt1 (Y15081), Ppt2 (Y16253), *Schizosaccharomyces pombe* FAS2 (D83412), *Saccharomyces cerevisiae* FAS2 (J03936), *Penicillium griseofulvum* FAS2 (M37461), *Candida albicans* FAS2 (L29063), *Emericella nidulans* FAS2 (U75347), StcJ (U34740).

Acknowledgements

We acknowledge the help of the staff from the ELETTRA synchrotron (Trieste) as well as from the EMBL outstation at DESY (Hamburg) during data collection. This work was supported by the Deutsche Forschungsgemeinschaft (R.F. SFB286-TPA11; M.A.M. MA811/14-1) and the Fonds der Chemischen Industrie (M.A.M.).

References

Abrahams,J.P. and Leslie,A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D*, **52**, 30–42.

Belshaw,P.J., Walsh,C.T. and Stachelhaus,T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284**, 486–489.

Black,T.A. and Wolk,C.P. (1994) Analysis of a Het$^-$ mutation in *Anabaena* sp. strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing. *J. Bacteriol.*, **176**, 2282–2292.

Brünger,A.T. (1993) *X-PLOR Version 3.1*. Yale University Press, New Haven, CT.

Cane,D.E., Walsh,C.T. and Khosla,C. (1998) Harnessing the biosynthetic code: combinations, permutations and mutations. *Science*, **282**, 63–68.

Clements,A., Rojas,J.R., Trievel,R.C., Wang,L., Berger,S.L. and Marmorstein,R. (1999) Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J.*, **18**, 3521–3532.

Conti,E., Stachelhaus,T., Marahiel,M.A. and Brick,P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.*, **16**, 4174–4183.

De La Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.*, **276**, 472–494.

Dieckmann,R., Lee,Y.O., van Liempt,H., von Dohren,H. and Kleinkauf,H. (1995) Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Lett*., **357**, 212–216.

Dieckmann,R., Pavela-Vrancic,M., von Dohren,H. and Kleinkauf,H. (1999) Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1. *J. Mol. Biol.*, **288**, 129–140.

Ehmann,D.E., Gehring,A.M. and Walsh,C.T. (1999) Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of α-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*, **38**, 6171–6177.

Engel,C. and Wierenga,R. (1996) The diverse world of coenzyme A binding proteins. *Curr. Opin. Struct. Biol*., **6**, 790–797.

Fraser,M.E., James,M.N.G., Bridger,W.A. and Wolodko,W.T. (1999) A detailed structural description of *Escherichia coli* succinyl-CoA synthetase. *J. Mol. Biol*., **285**, 1633–1653.

Hendrickson,W.A., Horton,J.R. and LeMaster,D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*., **9**, 1665–1672.

Hickman,A.B., Namboodiri,M.A., Klein,D.C. and Dyda,F. (1999) The structural basis of ordered substrate binding by serotonin *N*-acetyltransferase: enzyme complex at 1.8 Å resolution with a bisubstrate analog. *Cell*, **97**, 361–369.

Higgins,D.G., Thompson,J.D. and Gibson,T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383–402.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Jones,T.A. and Kjeldgaard,M. (1997) Electron-density map interpretation. *Methods Enzymol.*, **277**, 173–208.

Kealey,J.T., Liu,S., Santi,D.V., Betlach,M.C. and Barr,P.J. (1998) Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eucaryotic hosts. *Proc. Natl Acad. Sci. USA*, **95**, 505–509.

Kleinkauf,H. and Von Dohren,H. (1996) A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem*., **236**, 335–351.

Konz,D. and Marahiel,M.A. (1999) How do peptide synthetases generate structural diversity? *Chem. Biol.*, **6**, R39–R48.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Lambalot,R.H. and Walsh,C.T. (1995) Cloning, overproduction and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J. Biol. Chem.*, **270**, 24658–24661.

Lambalot,R.H. and Walsh,C.T. (1997) Holo-[acyl-carrier-protein] synthase of *Escherichia coli. Methods Enzymol.*, **279**, 254–262.

Lambalot,R.H., Gehring,A.M., Flugel,R.S., Zuber,P., LaCelle,M., Marahiel,M.A., Reid,R., Khosla,C. and Walsh,C.T. (1996) A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.*, **3**, 923–936.

Lamzin,V.S. and Wilson,K.S. (1997) Automated refinement for protein crystallography. *Methods Enzymol.*, **277**, 269–305.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Merrit,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

Modis,Y. and Wierenga,R. (1998). Two crystal structures of *N*-acetyltransferases reveal a new fold for CoA-dependent enzymes. *Structure*, **6**, 1345–1350.

Mofid,M.R., Marahiel,M.A., Ficner,R. and Reuter,K. (1999) Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases. *Acta Crystallogr. D*, **55**, 1098–1100.

Mootz,H.D. and Marahiel,M.A. (1999) Design and application of multimodular peptide synthetases. *Curr. Opin. Biotechnol*., **10**, 341–348.

Nakano,M.M., Corbell,N., Besson,J. and Zuber,P. (1992) Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis. Mol. Gen. Genet.*, **232**, 313–321.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Quadri,L.E., Weinreb,P.H., Lei,M., Nakano,M.M., Zuber,P. and Walsh,C.T. (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*, **37**, 1585–1595.

Rost,B. and Sander,C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19**, 55–72.

Stachelhaus,T. and Marahiel,M.A. (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.*, **270**, 6163–6139.

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

Stachelhaus,T., Huser,A. and Marahiel,M.A. (1996a) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.*, **3**, 913–921.

Stachelhaus,T., Schneider,A. and Marahiel,M.A. (1996b) Engineered biosynthesis of peptide antibiotics. *Biochem. Pharmacol*., **52**, 177–186.

Stachelhaus,T., Mootz,H.D., Bergendahl,V. and Marahiel,M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.*, **273**, 22773–22781.

Stachelhaus,T., Mootz,H.D. and Marahiel,M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, **6**, 493–505.

Stein,T., Vater,J., Kruft,V., Otto,A., Wittmann-Liebold,B., Franke,P., Panico,M., McDowell,R. and Morris,H.R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.*, **271**, 15428–15435.

Stuible,H.P., Meier,S. and Schweizer,E. (1997) Identification, isolation and biochemical characterization of a phosphopantetheine:protein transferase that activates the two type-I fatty acid synthases of *Brevibacterium ammoniagenes*. *Eur. J. Biochem*., **248**, 481–487.

Stuible,H.P., Meier,S., Wagner,C., Hannappel,E. and Schweizer,E. (1998) A novel phosphopantetheine:protein transferase activating yeast mitochondrial acyl carrier protein. *J. Biol. Chem.*, **273**, 22334–22339.

Turgay,K., Krause,M. and 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.

Walsh,C.T., Gehring,A.M., Weinreb,P.H., Quadri,L.E. and Flugel,R.S. (1997) Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.*, **1**, 309–315.

Received September 8, 1999; revised October 1, 1999; accepted October 4, 1999