Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S

Elena Conti^{1,2}, Torsten Stachelhaus³,
Mohamed A. Marahiel³ and Peter Brick^{1,4}

¹Biophysics Section, Blackett Laboratory, Imperial College, London SW7 2BZ, UK and ³Biochemie/Fachbereich Chemie, Philipps-Universität Marburg, D-35032 Marburg, Germany

²Present address: Laboratory of Molecular Biophysics, Rockefeller University, New York, NY10021, USA

⁴Corresponding author

The non-ribosomal synthesis of the cyclic peptide antibiotic gramicidin S is accomplished by two large multifunctional enzymes, the peptide synthetases 1 and 2. The enzyme complex contains five conserved subunits of ~60 kDa which carry out ATP-dependent activation of specific amino acids and share extensive regions of sequence similarity with adenylating enzymes such as firefly luciferases and acyl-CoA ligases. We have determined the crystal structure of the N-terminal adenylation subunit in a complex with AMP and L-phenylalanine to 1.9 Å resolution. The 556 amino acid residue fragment is folded into two domains with the active site situated at their interface. Each domain of the enzyme has a similar topology to the corresponding domain of unliganded firefly luciferase, but a remarkable relative domain rotation of 94° occurs. This conformation places the absolutely conserved Lys517 in a position to form electrostatic interactions with both ligands. The AMP is bound with the phosphate moiety interacting with Lys517 and the hydroxyl groups of the ribose forming hydrogen bonds with Asp413. The phenylalanine substrate binds in a hydrophobic pocket with the carboxylate group interacting with Lys517 and the α-amino group with Asp235. The structure reveals the role of the invariant residues within the superfamily of adenylate-forming enzymes and indicates a conserved mechanism of nucleotide binding and substrate activation.

Keywords: non-ribosomal peptide biosynthesis/peptide synthetases/X-ray crystallography

Introduction

A number of oligopeptides, some of which have important medical and biotechnological applications, are produced by fungi and bacteria via a non-ribosomal mechanism. Peptides such as the cyclic gramicidin S and cyclosporin A, the lactone actinomycin, the branched bacitracin and the linear precursor of both penicillin and cephalosporin, are synthesized by large multifunctional enzymes which act as protein templates for the growing polypeptide chain. Peptide synthetases catalyse the repetitive activation and condensation of the constituent amino acids to yield

the peptide product. Each amino acid is activated by adenylation of its carboxylate group with ATP and then transferred to the thiol group of an enzyme-bound phosphopantetheine cofactor for possible modification and the elongation reaction (Stachelhaus and Marahiel, 1995a; Kleinkauf and von Döhren, 1996).

The cloning and sequencing of several peptide synthetase genes have revealed a conserved and ordered modular organization. Each module encodes a functional building unit containing ~1000 amino acids, which specifically recognizes a single amino acid. Within such a protein template-directed peptide biosynthesis, the occurrence and specific order of the modules in the genomic DNA dictate the number and sequence of the amino acids to be incorporated into the resulting oligopeptide. The modular arrangement of peptide synthetases closely parallels the multienzyme complexes responsible for the biogenesis of fatty acids and of the polyketide family of natural products. Furthermore, peptide synthetases, fatty-acid synthetases and polyketide synthetases all use enzyme-bound phosphopantetheine cofactors as acyl carriers, in a thiotemplate mechanism first proposed by Lipmann more than 20 years ago (Lipmann, 1971) and revised recently (Stein *et al.*, 1996).

In particular, the synthesis of the cyclic antibiotic gramicidin S has been studied in detail. Gramicidin S is produced by the Gram-positive bacterium *Bacillus brevis* and consists of two identical pentapeptides joined head to tail. It is synthesized by the multienzyme complex gramicidin S synthetase, which is encoded by the 19 kb *grs* operon that includes the genes *grsA*, *grsB* and *grsT*. The *grsT* gene, which is located at the 5'-end of the *grs* operon, encodes a 29 kDa protein homologous to fatty-acid thioesterases. The *grsA* gene product, gramicidin S synthetase 1 (GrsA) is a protein composed of 1098 amino acids (Hori *et al.*, 1989; Krätzschmar *et al.*, 1989). GrsA activates L-phenylalanine to the corresponding acyl-adenylate and catalyses the inversion of configuration of the amino acid. D-phenylalanine is then transferred to the *grsB* gene product, gramicidin S synthetase 2 (GrsB), a 510 kDa polypeptide chain which sequentially activates proline, valine, ornithine and leucine and forms the peptide bonds in the elongation reaction, releasing the decapeptide (D-Phe-Pro-Val-Orn-Leu)₂ after cyclization.

Each of the five modules in which the *grs* operon is organized encodes for highly conserved functional subunits. The major one is a 60 kDa fragment which recognizes a specific amino acid and catalyses the adenylation of the amino acid carboxylate group with the α-phosphate of ATP. This adenylation subunit is conserved not only within all known peptide synthetases, but also shares extensive sequence similarity with firefly luciferases and acyl CoA ligases. Common to all these enzymes is the ATP-dependent activation of substrates as acyl adenylates.

On the other hand, the adenylation subunit shares no sequence homology with enzymes involved in the ribosomal synthesis of polypeptides, despite the fact that the formation of aminoacyl-adenylates is chemically analogous in the two systems. Indeed, the crystal structure of firefly luciferase has indicated a structural framework unrelated to those of both class I and class II aminoacyl-tRNA synthetases (Conti *et al.*, 1996).

We report here the crystal structure of the phenylalanine-activating subunit of gramicidin synthetase 1 (PheA) in a ternary complex with phenylalanine and AMP. The structure reveals the role of residues which are highly conserved in the superfamily of adenylate-forming enzymes. In addition, the presence of the substrate provides details of the amino acid specificity and allows a sequence-based comparison to be made with other peptide synthetases. A comparison of the structure with that of unliganded firefly luciferase reveals that both a domain rotation and a conformational change of a loop in the N-terminal domain must occur for luciferase to form an active complex with luciferin and ATP.

Results and discussion

Crystal structure determination

The crystal structure of PheA was determined by the multiple isomorphous replacement method, together with real-space non-crystallographic symmetry averaging and refined against 1.9 Å resolution diffraction data to a crystallographic *R*-factor and *R*-free (Brünger, 1992) of 21.4% and 24.6% respectively. The model for 512 residues has good stereochemistry and includes phenylalanine and AMP bound at the active site. No interpretable electron density is present for the 16 N-terminal residues, the 33 C-terminal residues, nor for a loop containing residues 192–196. The two copies of the molecule in the asymmetric unit have a very similar conformation: after superposition the r.m.s. difference in the position of the main chain atoms of residues 21–530 is 0.26 Å.

Description of the overall structure

The polypeptide chain folds into two compact domains (Figure 1). There are very few direct protein–protein interdomain contacts and instead the interactions between the structural domains are mediated by a network of hydrogen bonds between the side chains of the protein and a sandwiched layer of ordered water molecules. The much larger N-terminal domain comprising residues 17–428 contains three subdomains: a distorted β -barrel and two β -sheets which pack together to form a five-layered $\alpha\beta\alpha\beta\alpha$ tertiary structure (Figure 2). Subdomain A contains a six-stranded β -sheet and three helices formed by a single segment of the polypeptide chain (residues 91–203) while a seventh strand is formed by an insertion in the β -barrel subdomain (Figure 3). The β -sheet B contains eight strands, of which the first two (B1–B2) are formed by residues occurring before β -sheet A in the polypeptide chain, while the remaining six strands (B3–B8) and four helices form a contiguous polypeptide segment located before the β -barrel subdomain in the sequence. Strands 1–6 in the two β -sheets share a similar topology, with strands A1–A4 in sheet A corresponding to strands B3–

Fig. 1. Ribbon diagram of the PheA molecule with the large N-terminal domain shown in blue and the small C-terminal domain in green. The disordered loop (residues 192–196) near the active site is coloured violet. The AMP (red) and phenylalanine (orange) ligands are drawn using a space-filling representation. The side chain of Lys517 on the loop that projects down from the C-terminal domain is drawn in green using a ball-and-stick representation.

B6 in sheet B while strands A5–A6 correspond to strands B1–B2.

The C-terminal domain (residues 429–530) includes two helices which pack against one side of a three-stranded antiparallel β -sheet E as well as an additional small sheet containing two β -strands. The polypeptide chain at the C-terminus of the protein loops back towards the N-terminal domain and then packs against the remaining face of β sheet E (Figure 1). Residues at both the N- and C-termini of the polypeptide chain project out from the surface of the molecule and are relatively less well ordered.

Ligand binding

The crystal structure shows unambiguous electron density for the ligands bound at the active site (Figure 4). In spite of the presence of Mg-ATP and phenylalanine in the crystallization conditions, the electron density is not consistent with the product of the activation reaction: the phenylalanyl adenylate has been hydrolysed to the corresponding amino acid and AMP. Substrate recognition is accomplished by an extensive network of hydrogen bonds with a number of charged or polar amino acid residues. Most of the protein residues involved in substrate recognition are contributed by the large N-terminal domain. However, it is a charged residue of the C-terminal domain, the strictly invariant Lys517, which is involved in two key polar interactions with both the amino acid and the adenosine, fixing their position in the active site and clamping the C-terminal domain in a productive orientation. The lysine residue is located at the bottom of the large loop that projects down into the active site from the C-terminal domain (Figure 1). The key role played by this residue has been demonstrated by site-directed mutagenesis studies where the replacement of the corresponding lysine to a glutamine in the valine-activating domain of

Fig. 2. Stereo diagram of the large N-terminal domain of PheA showing the bound ligands coloured as in Figure 1. The β -sheet A is on the left-hand side, β -sheet B is on the right-hand side, and the β -barrel is at the top of the figure. The N-terminus of the protein is at the top of the figure. The disordered loop (residues 192–196) near the active site is coloured violet.

surfactin synthetase 1 results in the reduction of the reverse rate of adenylate formation by 94% (Hamoen *et al.*, 1995).

Adenylate binding

The adenylate is bound in a cleft present on the surface of the large N-terminal domain, between residues from β -sheet B and the β -barrel subdomains. A stereo diagram of the adenylate binding site is shown in Figure 5, while Figure 6 shows the hydrogen bonding interactions. The adenine moiety lies in a slot sandwiched between the side chains of Tyr323, Tyr425 and Ile348 on one side and the main chain atoms of residues 302–304 on the other. The binding of the base is mediated not only by the large area of hydrophobic and van der Waals interactions on the sides of the slot, but also by the hydrogen bonds of the N6 amino group with the main chain carbonyl oxygen of Ala322 and the side chain oxygen of Asn321. Hydrogen bonding to the exocyclic nitrogen of the adenine is the major specificity determinant by which the enzyme discriminates against guanine. No other ring nitrogen is in direct contact with the protein; of the possible hydrogen bonding interactions with the acceptor groups at positions 1, 3 and 7 of the purine ring, only N1 contacts a well-ordered water molecule ($B = 22 \text{ \AA}^2$) which is in turn at $3.0 \text{ \AA}$ from the side chain nitrogen of Asn321. This pattern of interactions accounts for the catalytic activity displayed by the peptide synthetase in the presence of ATP analogues such as 7-deaza-ATP (Pavela-Vrancic *et al.*, 1994).

The ribose moiety is held in the C3'-endo conformation. The two hydroxyls of the sugar are involved in hydrogen-bonding interactions with the carboxylate of Asp413, which is a strictly invariant residue within the superfamily of adenylate-forming enzymes. The 2' hydroxyl also forms a rather long ($3.2 \text{ \AA}$) hydrogen bond with the side chain of Tyr425 while Tyr323, which packs edge-on against the

adenine ring, also hydrogen bonds to Asp413. In site-directed mutagenesis studies on the highly homologous tyrocidine synthetase 1, Gocht and Marahiel (1994) find that the replacement of the invariant aspartate by an asparagine residue reduces the ATP-PP_i activity to 78% of that of the wild-type enzyme, while the substitution of a serine residue at this position reduces the activity to just 12%. Gramicidin synthetase 1 shows a much higher activity when ATP is replaced by 2'-deoxy-ATP (a 40% reduction) than when 3'-deoxy-ATP (85% reduction) is used in the ATP-PP_i exchange reaction (Pavela-Vrancic *et al.*, 1994). Although believed to be poor acceptors (Moodie *et al.*, 1996), both the ribose O-4' and O-5' are hydrogen-bonded to the invariant Lys517.

The α -phosphate of AMP has slightly weaker electron density, indicating that the binding site is more disordered. Interaction is with Thr326, a highly conserved residue in the superfamily of adenylate-forming enzymes, Thr190, also well conserved although replaced by a serine in the luciferases, together with the invariant Glu327. Glu327 points towards O1 of the phosphate and is bridged presumably by a magnesium ion, at $2.54 \text{ \AA}$ from the carboxylate and $2.26 \text{ \AA}$ from the phosphate.

Phenylalanine binding

The amino acid binding site of the peptide synthetase is a pocket with an entrance on the concave surface of the large domain near the intersection of the three subdomains. A stereo representation of the phenylalanine binding site of PheA is shown in Figure 7. The side chain of Asp235 and the main chain carbonyl oxygens of Gly324 and Ile330 are well placed to form hydrogen bonds with the α -amino group of the phenylalanine substrate. The aspartic acid is conserved in all peptide synthetases apart from the L- α -amino adipate activating domain of ACV synthetase.

Fig. 3. Topological arrangement of the secondary structural elements in PheA. The circles represent α -helices and the arrows β -strands. The strands have been numbered sequentially for each of the five β -sheets. Sheets A, B and the β -barrel C are in the N-terminal domain while sheets D and E are in the C-terminal domain.

In the PheA structure, Ile330 has dihedral angles ($\phi = 74^\circ$, $\psi = -64^\circ$) outside the allowed regions of a Ramachandran plot (Ramakrishnan and Ramachandran, 1965). As is commonly observed in protein structures, the energetically unfavourable main chain dihedral angle is associated with a region of the molecule having a functional role (Herzberg and Moulton, 1991). The α -carboxylate group of the substrate amino acid is stabilized by an electrostatic interaction with the invariant Lys517 from the C-terminal domain.

The specificity pocket for the phenylalanine side chain is surrounded by residues from the strands and a helix associated with β -sheet B (Figures 2 and 7). The pocket is lined at the bottom by the indole ring of Trp239, on one side by Ala236, Ile330 and Cys331 and on the opposite side by Ala322, Ala301 and Thr278. The two sides of the pocket are appropriately separated to accommodate an aromatic residue but at one end of the pocket (towards the viewer in Figure 7) there is a water-filled channel that connects with the solvent.

The phenylalanine binding pocket can accommodate both stereoisomers of the amino acid with no significant

Fig. 4. The difference electron density calculated with the AMP and phenylalanine ligands omitted from the model. The map was calculated using all data between 20.0 Å and 1.9 Å and contoured at 5 σ .

change in the protein conformation. In the 2.0 Å refined crystal structure of a ternary complex containing D-phenylalanine and AMP, the polar interactions between the ligand and the protein are identical to those in the L-phenylalanyl complex but the benzene ring of the side chain is rotated by 30° about an axis perpendicular to the plane of the ring. This rotation leads to a 1.3 Å displacement in the relative position of the C_β atoms, but the C_α atoms are within 0.5 Å of one another and the oxygen atoms that interact with the side chain of Lys517 are within 0.26 Å.

Over 50 sequences are now known for the amino acid activating modules of peptide synthetases and although the enzymes differ in their substrate specificity, they show extensive regions of sequence similarity to PheA. Of the modules listed in Table I, the percentage of identical residues ranges from 26% for module 3 of the HC-toxin synthetase to 56% for the phenylalanine-activating module of tyrocidine synthetase. With this level of sequence identity, the main chain conformation of the enzymes is likely to be very similar and the differing substrate specificities will be mainly determined by the nature of the amino acids lining the substrate binding pocket. Table I lists these amino acids for several different peptide synthetases. A number of enzymes contain charged residues near the binding pocket and for the activation of substrates with charged side chains such as ornithine, aspartate and glutamate, there are protein side chains at either position 239 or 278 with opposite charge. The first module of ACV synthetase adenylates the δ -carboxylate rather than the α -carboxylate of the L- α -amino adipate side chain, and it is possible that the α -amino and α -carboxylate groups of the substrate bind at the bottom of the pocket and interact with the arginine at position 239 and the

Fig. 5. Stereo diagram showing the interactions made by the phenylalanine substrate and AMP (both in green) with PheA. The position of a possible magnesium ion is shown in yellow. Potential hydrogen bonding interactions are indicated by dotted lines.

Table I. Identity of residues lining the substrate binding pocket in various peptide synthetases

Synthetase (module number)	Substrate	(%) Identity with PheA	Residue (PheA numbering)							
			236	239	278	299	301	322	330	331
Gramicidin 1: PheA	Phe		A	W	T	I	A	A	I	C
Tyrocidine 1	Phe	56	T	F	T	I	A	A	I	C
Surfactin 1 (2)	Leu	37	A	F	M	F	G	M	V	F
Surfactin 1 (3) & 2 (3)	Leu	36–37	A	W	F	T	G	N	V	V
Gramicidin 2 (4)	Leu	42	G	A	Y	T	G	E	V	V
Cyclosporin (2,3,8,10)	Leu	30–31	A	W	L	Y	G	A	V	M
Gramicidin 2 (2)	Val	40	A	F	W	I	G	G	T	F
ACV (3)	Val	28	F	E	S	T	A	A	V	Y
Cyclosporin (4,9)	Val	30–31	A	W	M	F	A	A	I/V	L
Surfactin 2 (1)	Val	38	A	F	W	I	G	G	T	F
ACV (2)	Cys	33	H	E	S	D	V	G	I	T
HC-toxin (1)	Pro	27	I	A	V	I	T	V	L	I
Gramicidin 2 (1)	Pro	41	C	K	S	I	A	H	V	V
Cyclosporin (1)	D-Ala	30	L	W	F	L	I	A	V	V
Cyclosporin (11)	Ala	30	V	F	I	Y	A	A	I	L
HC-toxin (2)	Ala	27	A	G	G	C	A	M	V	A
HC-toxin (3)	Ala	26	L	L	F	G	I	S	V	L
Cyclosporin (7)	Gly	31	I	Q	M	F	V	A	M	Q
ACV (1)	Aad	32	P	R	N	I	V	E	F	V
Gramicidin 2 (3)	Orn	42	V	G	E	I	G	S	L	I
Cyclosporin (5)	Bmt	31	A	W	T	Y	G	G	V	I
Cyclosporin (6)	Abu	30	A	W	F	H	A	V	A	Y
Surfactin 1 (1)	Glu	34	A	K	D	L	G	V	V	D
Surfactin 2 (2)	Asp	33	L	T	K	V	G	H	I	A

Unusual amino acids: Aad, L-2-amino adipic acid; Abu, α -amino butyric acid; Bmt, (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine. References for sequences: Tyrocidine synthetase 1 (Weckermann *et al.*, 1988), HC-toxin synthetase (Scott-Craig *et al.*, 1992), surfactin synthetase 1 (Fuma *et al.*, 1993), surfactin synthetase 2 (Cosmina *et al.*, 1993), ACV synthetase from *Penicillium chrysogenum* (Smith *et al.*, 1990), gramicidin synthetase 1 (Krättschmar *et al.*, 1989), gramicidin synthetase 2 (Hori *et al.*, 1989), cyclosporin synthetase (Weber *et al.*, 1994). Modules 2,3,4,5,7,8 and 10 of cyclosporin synthetase contain a 447 residue insertion between β -strand E2 and the following helix in the C-terminal domain which catalyses the N-methylation of the amino-acid.

glutamate at position 322. This mode of binding could explain the absence of an aspartate residue at position 235 to interact with the α -amino group of the amino acid substrate. It can also be seen from Table I that charged residues are sometimes present near the pocket in enzymes that activate amino acids with neutral side chains, while synthetase modules that are specific for the same substrate can have different amino acids around the substrate binding pocket.

Comparison with the structure of firefly luciferase

As was anticipated from the 16% sequence identity and the presence of short highly conserved amino acid motifs equidistantly separated in the two sequences (Figure 8), the topology of each of the structural domains of PheA is very similar to that observed in the crystal structure of the unliganded firefly luciferase (Conti *et al.*, 1996). PheA is slightly smaller, is missing an extra strand in sheet A, an α -helix near the C-terminus and, where the luciferase

enzyme has a β -strand at the N-terminus of the molecule, PheA has an α -helix. The most striking difference between the two crystal structures is the relative orientation of the N- and C-terminal domains. When compared with the luciferase structure, the C-terminal domain of PheA is rotated by 94° relative to the N-terminal domain and is 5 Å closer to it (Figure 9). The loops 436–440 and 524–528 (luciferase numbering) near the domain interface which are disordered in luciferase have well-defined electron density in the PheA crystal structure. Optimal superposition of the N-terminal domains of the two enzymes results in an r.m.s. separation of 1.5 Å for the 287 pairs of equivalent C_α atoms with a separation <3 Å (Figures 8 and 9), while superposition of the C-terminal domains gives an r.m.s. separation of 1.42 Å for the 77 C_α atoms within 3 Å. The structure of PheA in a binary complex with Mg-AMP-PNP in which the PNP moiety is disordered (data not shown) reveals a similar domain orientation to the one observed for the ternary complex, indicating that the presence of the amino acid is not

Fig. 6. Schematic representation of the hydrogen bonding between PheA and the phenylalanine and AMP ligands.

Fig. 7. Stereo diagram showing the side chains of PheA that line the specificity pocket for the phenylalanine substrate (in green). The unlabelled Cys331 is hidden behind the substrate. The main chain carbonyl groups of residues Ile334 and Gly324 have been included. Potential hydrogen bonding interactions are indicated by dotted lines.

Fig. 8. Structurally based amino acid sequence alignment of PheA (top row) with firefly luciferase (Conti *et al.*, 1996) (bottom row). Lower-case characters indicate residues omitted from the molecular models. Asterisks indicate pairs of residues for which the separation of the C_α atoms is <3 Å after separate optimal superposition of either the N-terminal or the C-terminal domains of the molecular structures. The position of secondary structural elements in PheA determined using the algorithm of Kabsch and Sander (1983) are indicated above the sequence. The β -strands are labelled as in Figure 3. Residues that are highly conserved within the superfamily of adenylate-forming enzymes are shaded.

Fig. 9. Comparison of the crystal structures of (A) PheA and (B) firefly luciferase. The larger N-terminal domains in the lower portion of the figures are viewed in a similar orientation. For each domain, the shaded regions of the structure corresponding to residues for which the separation of the C_{α} atoms is <3 Å after optimal superposition. The relative orientation of the smaller C-terminal domains are related by a 94° rotation about an axis represented by a dashed line in each figure.

required to obtain this productive conformation of the protein.

There is a major difference in the entrance to the cavity which in PheA is occupied by the amino acid ligand. Although most of the main chain atoms of the two enzymes superpose well around the active site, the loop containing residues 314–319 in firefly luciferase (residues 300–305 in PheA) has a different conformation and obstructs both the entrance to the large water-filled luciferin binding pocket and the binding of the adenine ring of the nucleotide. A significant conformational change of this loop must occur in the firefly luciferase molecule to accommodate the binding of the ligands. The mutation of the highly conserved glycine residue in this loop (Gly302 in PheA) to a glutamic acid in the valine activating domain of GrsB results in the complete loss of PP_i -ATP exchange activity (Saito *et al.*, 1995). The addition of a side chain at this position is likely to compromise the movement of the loop and cause a change in the conformation of the main chain as the dihedral angles for Gly302 ($\phi = 99^{\circ}$, $\psi = -41^{\circ}$) are unfavourable for non-glycine residues.

Signature sequence

The most highly conserved sequence of amino acids in the superfamily of adenylate-forming enzymes involves residues 190-TSGTTGNPKG-199 which in the PheA structure form a loop between β -strands 5 and 6 in subdomain A. The absence of significant electron density for residues 192–196 implies that the central residues of the loop have conformational flexibility. The corresponding loop is also disordered in the crystal structure of the firefly luciferase. These residues are not involved in the binding of the AMP moiety; their position with respect to the AMP binding site suggests that they are likely to interact with the pyrophosphate leaving group (Figure 5). Glycine-rich loops are often present in ATP- and GTP-binding proteins and are generally found to form an anion hole which accommodates the phosphate of the nucleotide (Pai *et al.*, 1989; Knighton *et al.*, 1991). As discussed

above, Thr190 at the beginning of the conserved peptide interacts with the α -phosphate, while the side chain of Lys198 is poorly ordered and projects into solvent. The importance of Lys198 in the PP_i -ATP exchange reaction has been demonstrated by site-directed mutagenesis of the corresponding lysine residue in tyrocidine synthetase 1 (Gocht and Marahiel, 1994). Both Lys198 and the invariant arginine at position 428 (Figure 5) are probably involved in coordinating the pyrophosphate group, while the lysine at position 517 is likely to stabilize the negatively charged pentavalent transition state in a manner analogous to that of the invariant arginine of class II aminoacyl-tRNA synthetases. The arginine residue forms an ion pair with the α -carboxylate moiety of the amino acid substrate but interacts with the α -phosphate when the aminoacyl-adenylate intermediate is formed (Onesti *et al.*, 1995).

The knowledge of the residues involved in forming the substrate specificity pocket provided by this study combined with the wealth of amino acid sequence information available for other adenylate-forming domains provides a structural basis for understanding the specificity of peptide synthetases. These findings should allow the manipulation of these enzymes for the synthesis of novel peptides with modified biological activity. Indeed, the introduction of alternate amino acid-activating modules by recombinational integration in the *Bacillus* system has already demonstrated the possibility of biosynthesis of different antibiotics (Stachelhaus *et al.*, 1995).

Materials and methods

Crystallization and data collection

The recombinant phenylalanine-activating domain of gramicidin synthetase 1 (PheA) from *B. brevis* ATCC 999 was expressed in overproducing *Escherichia coli* cells and purified as described previously (Stachelhaus and Marahiel, 1995b). The PheA construct consists of residues 1–557 and a C-terminal hexahistidine tag which was not cleaved after purification. The enzyme was co-crystallized with Mg-ATP, in the presence of either L- or D-phenylalanine, from the sparse matrix screening (Jancarik and Kim, 1991).

Table II. Statistics for X-ray data collection and phase determination

Data set	Native	UAc	HgAc	UAc/HgAc	ATP+LPhe	ATP+DPhe
Synchrotron	Hamburg	Daresbury	Daresbury	Hamburg	Hamburg	Hamburg
	BW7A	9.6	9.5	X31	BW7B	BW7B
Wavelength (Å)	0.90	0.87	0.97	0.98	0.88	0.88
Maximum resolution (Å)	2.80	2.80	2.85	3.10	1.90	2.00
No. of measurements	64 873	59 826	57 663	52 044	230 285	144 834
Independent reflections	29 020	29 278	26 596	22 004	91 144	80 438
Multiplicity ^a	2.2 (2.1)	2.0 (1.9)	2.2 (1.9)	2.4 (2.3)	2.5 (1.9)	1.8 (1.8)
Completeness (%) ^a	96.0 (93.0)	96.0 (95.8)	96.1 (91.6)	98.1 (98.4)	95.3 (88.3)	96.4 (93.5)
<i>R</i> _{merge} (%) ^a	7.8 (13.2)	6.8 (17.8)	4.4 (8.8)	6.8 (13.9)	5.1 (23.2)	4.4 (13.9)
<i>I</i> /σ ^a	6.5 (5.3)	8.1 (4.1)	6.3 (7.3)	9.3 (4.9)	7.3 (2.7)	7.6 (4.8)
Heavy atom concentration (mM)		10.0	0.1	10.0/0.1		
No. of sites		4	4	4/4		
<i>R</i> _{iso} (%)		18.1 (20.8) ^b	15.4 (17.6) ^b	16.5 (18.0) ^b		23.4 (33.5) ^c
Phasing power (%) ^b		1.34 (1.11)	1.13 (0.95)	1.46 (1.26)		
<i>R</i> _{Cullis} (%) ^b		0.72 (0.79)	0.75 (0.87)	0.66 (0.84)		

^aValues for the outermost resolution shell are given in parentheses.

^bPhasing statistics calculated to 3.2 Å resolution.

^c*R*_{iso} with the ATP+LPhe data set, with the outermost resolution shell given in parentheses.

*R*_{merge} = $100 \times \sum_h \sum_j |I_{hj} - I_h| / \sum_h \sum_j I_{hj}$, where *I*_h is the weighted mean intensity of the symmetry-related reflections *I*_h.

*R*_{iso} = $100 \times \sum_{hkl} |F_p - F_{ph}| / \sum_{hkl} F_p$, where *F*_p is the native amplitude and *F*_{ph} is the derivative amplitude.

*R*_{Cullis} = $100 \times \sum_{hkl} ||F_p \pm F_{ph} - F_{h(calc)}| / \sum_{hkl} |F_p \pm F_{ph}|$, where *F*_p is the native amplitude, *F*_{ph} is the derivative amplitude for centric reflections and *F*_{h(calc)} is the calculated heavy atom structure factor.

Phasing power: r.m.s. calculated heavy atom structure factor amplitudes divided by the residual lack of closure.

The crystals of PheA complexed with Mg-ATP and L-phenylalanine used for the structure determination were grown by vapour diffusion techniques. The protein was concentrated to 15 mg/ml in 10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 15% glycerol, 2 mM ATP, 2 mM L-phenylalanine and 4 mM MgCl₂. The hanging drops contained equal volumes of the protein solution and of a reservoir solution consisting of 28–32% (w/v) methoxy polyethylene glycol (MePEG) 5000, 200 mM ammonium sulfate and 100 mM N-(2-acetamido)-iminodiacetic acid (ADA) at pH 6.5. Equilibration at 18°C yielded crystals that grew as plates, reaching their maximum size (0.5×0.5×0.2 mm) in 2 or 3 weeks. They belong to the primitive monoclinic spacegroup *P*2₁, with cell dimensions *a* = 61.9 Å, *b* = 155.6 Å, *c* = 65.8 Å and β = 94°. The volume to mass ratio calculation (Matthews, 1968) suggests the presence of two molecules per asymmetric unit, resulting in a *V*_m of 2.5 Å³/Dalton and a solvent content of 50%.

All data collection was carried out at cryogenic temperatures. Crystals were frozen using standard techniques (Teng, 1990) in a stream of nitrogen gas at 100 K produced by an Oxford Cryosystem (Cosier and Glazer, 1986). Successful freezing of the PheA crystals was achieved by transferring them for <1 min into a cryoprotectant solution containing 20% MePEG 5000, 100 mM ADA at pH 6.5 and 30% glycerol. For heavy atom screening, the crystals were previously stabilized into a harvesting solution consisting of 20% MePEG 5000, 200 mM ammonium sulfate, 100 mM ADA at pH 6.5 and 15% glycerol. Screening for heavy atom derivatives was carried out to low resolution on a MarResearch (Hamburg, Germany) image plate detector, using CuK_α radiation from an Elliott GX21 rotating anode X-ray generator. Higher-resolution synchrotron data for the native and derivatives used for phasing were obtained at SRS (Daresbury, UK) and DESY (Hamburg, Germany). The 1.9 Å resolution data used in the refinement were measured on the wiggler BW7B line at DESY. Diffracted intensities were evaluated and integrated using a modified version of MOSFLM for processing image plate data (A.G.W.Leslie, personal communication) and the CCP4 suite was used for data reduction (CCP4, 1994). A summary of the data collection statistics is given in Table II.

Structure determination

The presence of two molecules in the asymmetric unit, suggested by the calculation of the *V*_m value, was confirmed by self-rotation function studies which were carried out with the program POLARRFN (CCP4, 1994). An unambiguous peak which was six times as high as the r.m.s. of the map was identified and corresponded to a rotation of 180° around an axis in the *ac* plane. Attempts to solve the structure by the molecular replacement method using the firefly luciferase coordinates were unsuccessful. Phases were determined by the method of multiple isomorphous replacement using a uranyl acetate derivative, a mercury acetate deriv-

ative and a uranyl/mercury double derivative. Initial SIR phases were calculated from the positions of two uranyl sites which had been determined with the real-space Patterson search program RSPS (CCP4, 1994). Additional heavy atom sites were located by difference Fourier methods. Refinement of heavy atom parameters and phase calculations were performed using the program MLPHARE (Otwinowski, 1991). The refinement of the heavy atom sites and occupancies of the double derivative was carried out independently to avoid artificial sharpening of the phase probability distributions. The final MIR phases had an overall figure of merit of 0.538 for data between 20 and 3.2 Å resolution. The phasing statistics are shown in Table II.

The MIR phases were improved by density modification procedures. After solvent flattening and solvent flipping (Abrahams and Leslie, 1996), the 3.2 Å phases were used to compute an electron density map which clearly showed the presence of two molecules in the asymmetric unit. The RAVE package (Kleywegt and Jones, 1994) was used for creating a protein envelope from the skeletonized electron density map for one molecule, and for subsequent mask manipulation and averaging protocols. The non-crystallographic symmetry (ncs) operator were derived from the heavy atom positions and agreed with the results of the self-rotation function calculations. The ncs operator was refined by real-space density correlation with RAVE, to give a final density correlation coefficient of 0.71. Ten cycles of iterative averaging yielded an electron density map with many recognizable features resembling the luciferase structure.

The backbone of the luciferase molecule was used as a guide when building the model into the 3.2 Å averaged electron density map. A polyaniline model for 451 amino acid residues was traced by fitting fragments from a database of highly refined structures using the graphics program O (Jones and Thirup, 1986; Jones *et al.*, 1991). The amino acid sequence could be fitted unambiguously to give an initial model which included 78% of the atoms.

Crystallographic refinement

The PheA model was refined with the program X-PLOR (Brünger *et al.*, 1987) against a 1.9 Å resolution data set which had not been used for MIR phasing due to lack of isomorphism with the derivative data sets. Low-resolution data to 20 Å were included and a bulk solvent correction was applied throughout the refinement procedure. A random sample containing 5% of the data was excluded from the refinement and the agreement between calculated and observed structure factors for these reflections (*R*-free) was used to monitor the course of the refinement (Brünger, 1992). The model was restrained with the Engh and Huber stereochemical parameters (Engh and Huber, 1991).

An initial round of rigid-body refinement was carried out to 3.2 Å resolution, with each PheA molecule treated as a separate rigid body,

giving a free *R*-factor of 39.3%. A new matrix corresponding to the ncs operators was derived and strict non-crystallographic 2-fold symmetry was enforced until the later stages of refinement. The model was subjected to cycles of positional refinement to 3.2 Å resolution, followed by rounds of refinement in which the maximum resolution of the data was gradually increased. Beyond 2.4 Å resolution the refinement of the atomic positions was alternated with the refinement of the atomic temperature factors. After simulated annealing, the resulting molecular model had an *R*-free of 32.9% and an *R*-factor of 29.7% at 1.9 Å resolution. Manual rebuilding into an averaged electron density map allowed most gaps in the polypeptide chain to be closed and the electron density corresponding to the substrates AMP and phenylalanine could be easily located in the active site of the enzyme. Water molecules were added at geometrically reasonable positions at which the electron density was $>4\sigma$ in the averaged difference map. Subsequent refinement with ncs constraints and rebuilding in the averaged electron density map was performed until the free *R*-factor had dropped to 29.9%. A round of rigid-body refinement with each of the two molecules subdivided into two domains decreased the *R*-free to 25.8% and subsequent positional and individual temperature factor refinement was carried out independently for the two molecules in the asymmetric unit. The last round of refinement was carried out including the 5% of the data previously omitted.

The refined model contains 512 amino acid residues and 284 water molecules per subunit. There is no interpretable electron density for the 16 N-terminal and the 33 C-terminal residues, as well as for the 192–196 loop. The crystallographic *R*-factor is 21.3% for all reflections with no σ cut-off between 20 Å and 1.9 Å resolution. The free *R*-factor is 24.6% using a random sample of 3656 reflections out of the total 91142. The r.m.s. deviation from ideality is 0.008 Å in bond lengths and 1.37° in bond angles. The main chain dihedral angles for the majority of the residues in the refined model (90.1%) lie in the most favoured regions of the Ramachandran plot (Ramakrishnan and Ramachandran, 1965; Laskowski *et al.*, 1993). Only two amino acid residues (Ile330 and Glu61) have energetically unfavourable dihedral angles, corresponding to disallowed regions in the Ramachandran plot. The average temperature factor for all protein atoms is 33 Å². The refined B-factors for the atoms of the substrates fall within the 17.6–30.7 Å² range, but the α -phosphate of AMP is modelled as only half occupied.

The refined coordinates of the ternary complex was used as a starting model in the refinement against a 2.0 Å resolution diffraction data set collected from a crystal grown in the presence of ATP and D-phenylalanine (Table II). The resulting molecular model has a crystallographic *R*-factor of 21.1% using data between 20.0 Å and 2.0 Å. The value of *R*-free using a random sample containing 5% of the diffraction data is 24.5%. The r.m.s. deviation in bond lengths and bond angles from their target values are 0.007 Å and 1.25° respectively. The D-phenylalanine substrate has low temperature factors (15–20 Å²) and the α -phosphate of AMP has been modelled as being fully occupied. The structure of the protein in the complex containing D-phenylalanine is very similar to that containing L-phenylalanine: after superposition the r.m.s. difference in the positions of the main chain atoms is 0.19 Å.

Acknowledgements

We are very grateful to Silvia Onesti and Nicholas Franks for helpful discussions. We thank the staff at both the Synchrotron Radiation Source, Daresbury Laboratory and at DESY, Hamburg. We also thank the Deutsche Forschungsgemeinschaft, the EC (contract Nr. BIO4-CT95-01769) and Fonds der chemischen Industrie for support. The European Union supported the work at EMBL Hamburg through the HCMP Access to Large Installations Project, Contract Number CHGE-CT93-0040. Figures 1, 2, 5, 7 and 9 were prepared using MOLSCRIPT (Kraulis, 1991).

References

- Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr.*, **D52**, 30–42.
- Brünger, A.T. (1992) The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic *R*-factor refinement by molecular-dynamics. *Science*, **235**, 458–460.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–767.
- Conti, E., Franks, N.P. and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure*, **4**, 287–298.
- Cosier, J. and Glazer, A.M. (1986) A nitrogen gas stream cryostat for general X-ray diffraction studies. *J. Appl. Crystallogr.*, **19**, 105–107.
- Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G. and van Sinderen, D. (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.*, **8**, 821–831.
- Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.
- Fuma, S., Fujishima, Y., Corbell, N., D'Souza, C., Nakano, M.M., Zuber, P. and Yamane, K. (1993) Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. *Nucleic Acids Res.*, **21**, 93–97.
- Gocht, M. and 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.
- Hamoen, L.W., Eshuis, H., Jongbloed, J., Venema, G. and van Sinderen, D. (1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol. Microbiol.*, **15**, 55–63.
- Herzberg, O. and Moulton, J. (1991) Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins*, **11**, 223–229.
- Hori, K., Yamamoto, Y., Minetoki, T., Kurotsu, T., Kanda, M., Miura, S., Okamura, K., Furuyama, J. and Saito, Y. (1989) Molecular cloning and nucleotide sequence of the gramicidin S synthetase I gene. *J. Biochem. Tokyo*, **106**, 639–645.
- Jancarik, J. and Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.*, **24**, 409–411.
- Jones, T.A. and Thirup, S. (1986) Using known substructures in protein model building and crystallography. *EMBO J.*, **5**, 819–822.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Kleinkauf, H. and von Döhren, H. (1996) A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.*, **236**, 335–351.
- Kleywegt, G.J. and Jones, T.A. (1994) Halloween ... masks and bones. In Bailey, S., Hubbard, R. and Waller, D. (eds), *From First Map to Final Model: Proceedings of the CCP4 Study Weekend 6–7 January 1994*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.
- Knighton, D.R., Zheng, J., Ten Eyck, L.F., Ashord, V.A., Xuong, N.-H., Taylor, S.S. and Sowadaski, J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, **253**, 407–414.
- Krättschmar, J., Krause, M. and 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.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lipmann, F. (1971) Attempts to map a process evolution of peptide biosynthesis. *Science*, **173**, 875–884.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Moodie, S.L., Mitchell, J.B.O. and Thornton, J.M. (1996) Protein recognition of adenylate – an example of a fuzzy recognition template. *J. Mol. Biol.*, **263**, 486–500.
- Onesti, S., Miller, A.D. and Brick, P. (1995) The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli*. *Structure*, **3**, 163–176.
- Otwinski, Z. (1991) Maximum likelihood refinement of heavy-atom parameters. In Wolf, W., Evans, P.R. and Leslie, A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend 25–26 January 1991*. SERC Daresbury Laboratory, Warrington, UK, pp. 80–86.

- Pai, E.F., Kabsch, W., Krengel, U., Holmes, K.C., John, J. and Wittinghofer, A. (1989) Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature*, **341**, 209–214.
- Pavela-Vrancic, M., Van Liempt, H., Pfeifer, E., Freist, W. and von Döhren, H. (1994) Nucleotide binding by multienzyme peptide synthetases. *Eur. J. Biochem.*, **220**, 535–542.
- Ramakrishnan, C. and Ramachandran, G.N. (1965) Stereochemical criteria for polypeptide chain conformations. Allowed conformations for a pair of peptide units. *Biophys. J.*, **5**, 909–933.
- Saito, M., Hori, K., Kurotsu, T., Kanda, M. and Saito, Y. (1995) Three conserved glycine residues in valine activation of gramicidin S synthetase 2 from *Bacillus brevis*. *J. Biochem.*, **117**, 276–282.
- Scott-Craig, J.S., Panaccione, D.G., Pocard, J.-A. and 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.
- Smith, D.J., Earl, A.J. and 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.
- Stachelhaus, T. and Marahiel, M.A. (1995a) Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.*, **125**, 3–14.
- Stachelhaus, T. and Marahiel, M.A. (1995b) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.*, **270**, 6163–6169.
- 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.
- 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.
- Teng, T.Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallogr.*, **23**, 387–391.
- Weber, G., Schörgendorfer, K., Schneider-Scherzer, E. and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporin production in *Tolypocladium niveum* is encoded by a giant 45.8 kilobase open reading frame. *Curr. Genet.*, **26**, 120–125.
- Weckermann, R., Fürbass, R. and Marahiel, M.A. (1988) Complete nucleotide sequence of the *tycA* gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. *Nucleic Acids Res.*, **16**, 11841.

Received on March 6, 1997; revised on April 24, 1997