

## Review

Correspondence  
Sylvie Lautru  
S.Lautru@warwick.ac.uk

## Substrate recognition by nonribosomal peptide synthetase multi-enzymes

Sylvie Lautru and Gregory L. Challis

Department of Chemistry, The University of Warwick, Coventry CV4 7AL, UK

Nonribosomal peptide synthetases (NRPSs) are giant multi-domain enzymes that catalyse the biosynthesis of many commercially important peptides produced by bacteria and fungi. Several studies over the last decade have shown that many of the individual domains within NRPSs exhibit significant substrate selectivity, which impacts on our ability to engineer NRPSs to produce new bioactive microbial peptides. Adenylation domains appear to be the primary determinants of substrate selectivity in NRPSs. Much progress has been made towards an empirical understanding of substrate selection by these domains over the last 5 years, but the molecular basis of substrate selectivity in these domains is not yet well understood. Perhaps surprisingly, condensation domains have also been reported to exhibit moderate to high substrate selectivity, although the generality of this observation and its potential impact on engineered biosynthesis experiments has yet to be fully elucidated. The situation is less clear for the thioesterase domains, which seem in certain cases to be dedicated to the hydrolysis/cyclization of their natural substrate, whereas in other cases they are largely permissive.

### Background

Peptide-based natural products constitute a vast family of compounds mainly synthesized by micro-organisms. They possess a large range of biological activities, with functions ranging from chemical defence against other micro-organisms (vancomycin), to iron sequestration (yersinia-bactin) or components of the cell wall (glycopeptidolipids). Moreover, some of these peptides, such as the antibiotic vancomycin, the immunosuppressor cyclosporin or the antitumour agent bleomycin have proved to be of great therapeutic value. Although some peptides are synthesized by the ribosomal machinery (lantibiotics, for example), the vast majority are biosynthesized by gigantic multi-enzyme complexes known as nonribosomal peptide synthetases (NRPSs).

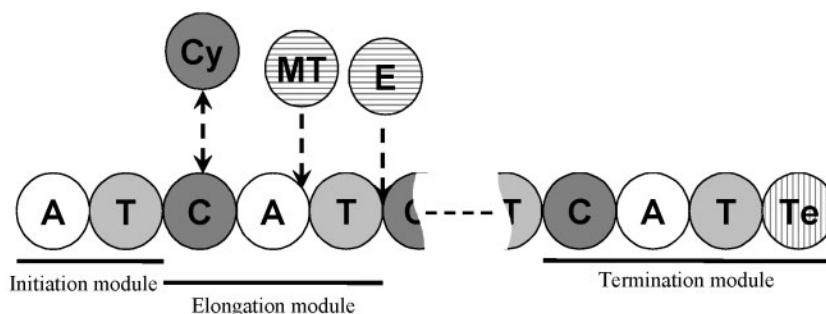
The elucidation of the organization of NRPSs has been made possible through structure–function analyses and sequencing of the genes encoding these enzymes. Such studies have revealed a modular architecture similar to that observed in polyketide synthases. Thus a peptide synthetase is formed by the repeat of basic units called modules which are responsible for the activation, covalent attachment as a thioester, and the optional modification of one amino acid (Fig. 1). The number and order of the modules within an NRPS usually correspond to the number of amino acids and the sequence of the peptide being synthesized, respectively. Each module can be divided into structurally independent domains, each fulfilling a specific function (Fig. 1). The adenylation (A) domain, the peptidyl carrier protein (PCP) domain and the

condensation (C) domain are the three core domains that define a minimal module (with the exception of the first module, which usually lacks a C domain). The A domain recognizes the amino acid substrate and activates it first through the formation of an aminoacyl adenylate and then via covalent binding of the activated amino acid as a thioester to the 4'-phosphopantetheinyl (4'Ppant) cofactor of the PCP domain (Fig. 2a). This cofactor is post-translationally attached to a conserved serine of PCP domains by a phosphopantetheinyl transferase. Elongation of the peptidyl chain is then carried out by the last essential domain, the C domain.

In addition to these three core domains, modules can contain extra or alternative domains which introduce a modification of the amino acid being incorporated, such as a change of the C $\alpha$  stereochemistry (epimerization domain), an N-methylation (N-methylation domain) of the  $\alpha$ -amino group, or the heterocyclization of Ser, Thr or Cys residues (heterocyclization domain).

The last step of nonribosomal peptide biosynthesis involves release of the assembled peptide from the NRPS into solution. In the majority of cases, this is accomplished by an extra domain located at the C-terminus of the last module of the NRPS, which can be a thioesterase (Te) domain, a variant of a C domain or a reductase domain.

There is much current interest in engineering NRPSs in order to create new peptides with potential biological activities. The extent to which this will be achievable depends critically on our understanding of how these systems can

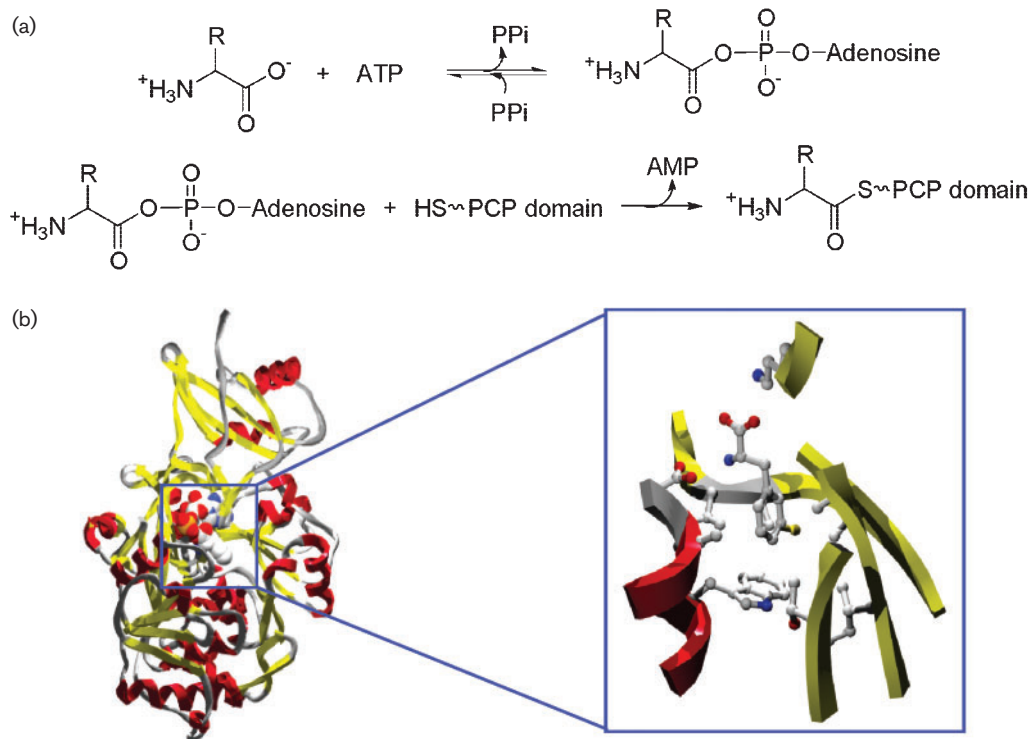


**Fig. 1.** Schematic diagram of the modular organization of NRPSs. Elongation modules typically contain three core domains: an adenylation (A) domain, a peptidyl carrier protein or thiolation (T) domain and a condensation (C) domain. The C domain is sometimes replaced by a heterocyclization (Cy) domain. Methyltransferase (MT) domains and epimerization (E) domains are optional domains introducing modifications on the elongating peptide chain. Initiation modules usually lack a C domain and termination modules usually possess a thioesterase (Te) domain.

discriminate between the wide variety of precursors known to be incorporated into nonribosomal peptides (including amino acids, hydroxy acids and aromatic acids). The degree of substrate discrimination exhibited by the different domains within NRPSs is the subject of this review.

### Substrate recognition by A domains

A domains activate and transfer amino acids to PCP domains (Fig. 2a) and are thus considered to be the primary determinant of substrate selectivity. The study of their specificity has been greatly facilitated by the determination



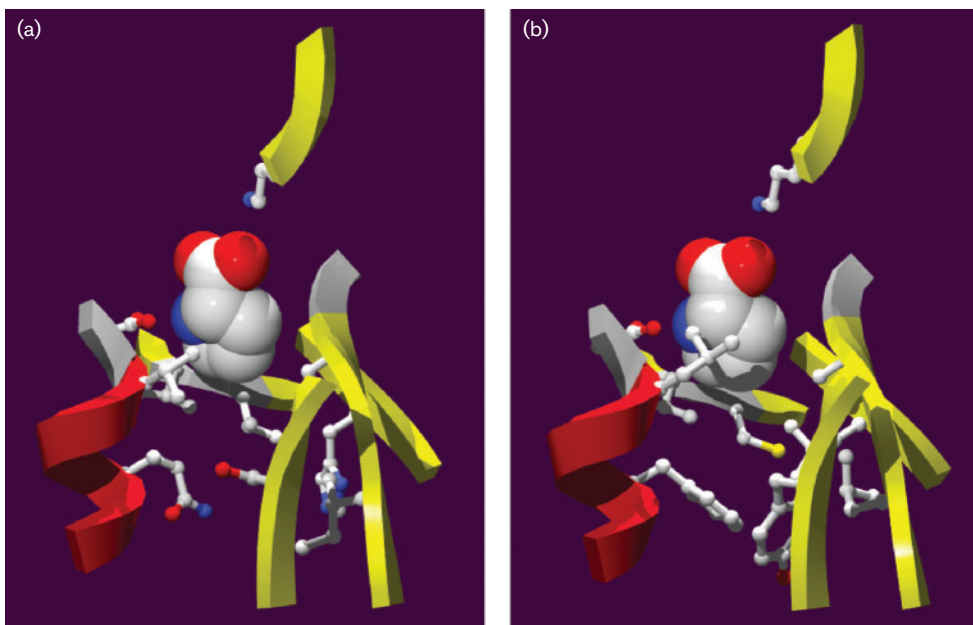
**Fig. 2.** A domains of NRPS. (a) Reactions catalysed by A domains: the amino acid substrate is activated through the formation of an aminoacyl adenylate followed by the covalent binding via a thioester bond to the 4'Ppant cofactor of the PCP domain. (b) Structure of the L-Phe-activating domain (PheA) of the gramicidin S synthetase GrsA, with an enlargement of the L-Phe-binding pocket showing the residues of PheA making contact with L-Phe. The side chains of these residues and the substrate L-Phe are rendered in ball and stick representation. The figure was created from the PDB file 1AMU using Swiss PDB viewer and rendered with POVray.

of the crystal structure of the A domain of the gramicidin S synthetase GrsA, the L-Phe-activating domain PheA, solved in 1997 by Conti and co-workers (Conti *et al.*, 1997). Cocrystallization of PheA with L-Phe and AMP allowed the identification of the hydrophobic L-Phe-binding pocket and of the residues making contact with Phe (Fig. 2b). Based on these results and the relatively high sequence similarity of A domains (between 26 and 56%), A domain sequence alignments in two independent studies (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999) showed an empirical correlation between the residues in each A domain corresponding to the 10 residues in direct contact with L-Phe in PheA and the substrate activated. In 2002, the structure of a second A domain was solved, the stand-alone 2,3-dihydroxybenzoic acid activating domain DhbE (May *et al.*, 2002). It reinforced the idea that A domains share a very similar fold and allowed the extension of the empirical correlations to some aryl acid activating A domains.

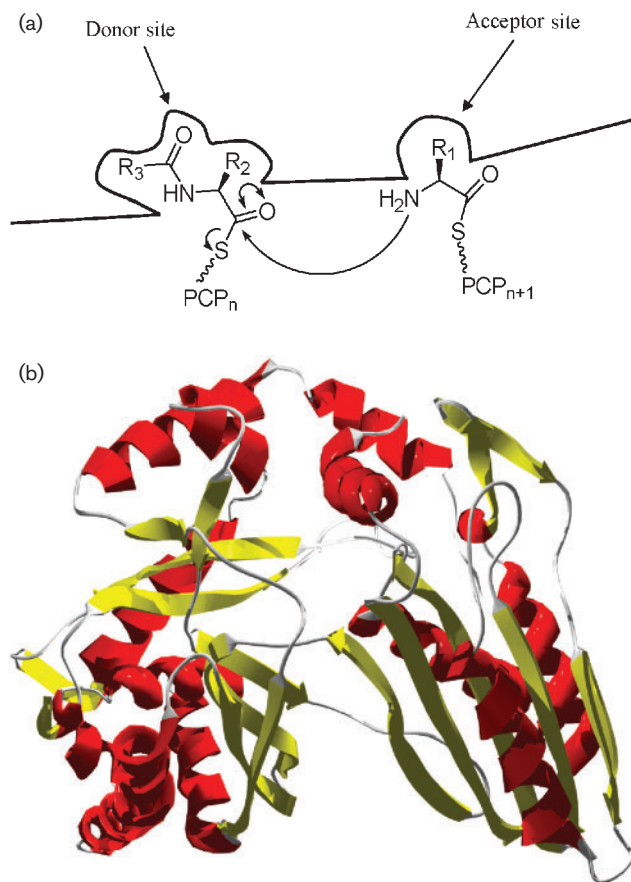
Since the empirical correlations were first published, sequence analyses of new NRPSs and biochemical characterization of A domains have permitted their extension and refinement. In a certain number of cases, it has been shown that the same amino acid can be activated by A domains with different predicted selectivity pocket residues, which suggests some degeneracy in the correlations. For example, the L-Pro-activating domains RedM, ProB and PltF have different predicted selectivity pocket residues from those first established for Pro-activating domains, e.g. from module 1 of TycB (Cerdeno *et al.*, 2001; Thomas *et al.*, 2002). This apparent degeneracy may originate from the fact that PheA, which selectively binds the relatively large side chain of L-Phe via seven to eight selectivity-determining residues, is likely to be a poor model for the selective binding of amino acids with relatively compact side chains such as L-Pro. Indeed, modelling of L-Pro binding to models of RedM and the A domain from module 1 of TycB based on the PheA structure (Fig. 3) clearly shows that only four to five residues at the top of the selectivity pocket are likely to be in direct contact with the L-Pro side chain and that the chemical nature of these residues is rather similar. The apparent degeneracy arises from the residues lower down in the selectivity pocket, which have very different chemical natures and are not in direct contact with the L-Pro side chain. A similar modelling approach by Lamont and coworkers suggests that the recognition of L-Thr may utilize only two selectivity pocket residues to make direct contacts with the substrate (Ackerley *et al.*, 2003). These studies demonstrate the naivety of attempting to identify selectivity pocket residues in A domains that activate substrates with side chains that differ significantly in size compared with the Phe side chain solely on the basis of alignment with the PheA selectivity pocket residues. In the absence of further A domain structures, molecular modelling approaches of the type described above may prove very useful in refining models for recognition of substrates with significantly smaller side chains than Phe, such as Pro, Gly, Ala, Thr, Ser and Val.

Two major applications have followed the development of models for A-domain selectivity. Firstly, these models have been used to predict the substrate selectivity of newly discovered NRPSs and consequently the structures of novel nonribosomal peptides. Thus in the genome of the soil bacterium *Streptomyces coelicolor* A3(2), several biosynthetic gene clusters including NRPSs have been identified, and the structure (coelichelin; Challis & Ravel, 2000) or partial structure (coelibactin; Bentley *et al.*, 2002) of two nonribosomal peptides has been suggested. Secondly, the selectivity models have been used to try to modify A domain substrate selectivity by mutating predicted selectivity-conferring residues (Eppelmann *et al.*, 2002; Stachelhaus *et al.*, 1999). L-Phe-, L-Glu- and L-Asp-activating domains have been mutated, using this method, to domains activating preferentially L-Leu, L-Gln and L-Asn, respectively, as evaluated by the ATP-pyrophosphate exchange assay. Because no kinetic parameters could be determined for the activation of the noncognate amino acid for both the wild-type and the mutant A domains (Asn activation by wild-type AspA domain, for example), it is difficult to evaluate the true extent of this switch of substrate specificity, which could simply arise as a result of abolition of activation of the cognate substrate. Catalytic efficiencies for the activation of the 'cognate' substrate for wild-type and mutant domains were comparable in two cases, but a 10-fold decrease was observed when the L-Asp-activating domain was mutated to an L-Asn-activating domain, even though a triple mutant had been constructed to perfectly match the Asn domain signature sequence. In this case particularly, but also in the two other cases, it would be useful to know if the catalytic efficiency of the mutant domain is comparable to the catalytic efficiency of wild-type A domains activating the same substrate.

The engineered changes in substrate selectivity reported thus far are relatively trivial, i.e. Asp to Asn, Glu to Gln and Phe to Leu. In most of these cases the substrates have side chains of very similar size and overall polarity, with only minor changes in charge or shape. The real challenges in this area, e.g. conversion of an A domain that recognizes a cognate substrate with a large side chain into one capable of selectively recognizing an amino acid with a small side chain, or engineering an A domain that recognizes a cognate substrate with a hydrophobic side chain into one that can selectively bind a substrate with a hydrophilic side chain, have yet to be addressed. Significant improvements in our understanding of substrate recognition by A domains will be required to successfully tackle these challenges. Such improvements are likely to come from quantitative application of the modelling approaches described above or from further structural studies of A domains that recognize substrates significantly different in structure to Phe. Arguably the pinnacle of achievement in this field would be to engineer an A domain with relaxed substrate selectivity that is capable of activating a wide range of substrates. Such an A domain would be of great utility in truly combinatorial biosynthesis of peptides – a significant goal



**Fig. 3.** Models of L-Pro-activating domains from (a) module 1 of the tyrocidine synthetase TycB and (b) the prodiginine synthetase RedM, illustrating that far fewer than eight residues are likely to form the selectivity pocket for small substrates such as L-Pro. Models were generated from the PDB file 1AMU by mutating the selectivity pocket residues to those predicted for the Pro-activating domains as described by Stachelhaus *et al.* (1999) and Challis *et al.* (2000).



that remains to be achieved in the engineered biosynthesis field – although the application of such an A domain to make combinatorial libraries may be severely limited by the substrate selectivity imposed by downstream domains, discussed below.

### Substrate recognition by C domains

C domains are responsible for the elongation of the peptidyl chain. They catalyse the condensation reaction between the peptidyl chain tethered to the phosphopantetheinyl arm of the upstream PCP domain and the amino acid bound to the downstream PCP domain (Fig. 4a). Recently, the structure of a first representative of C domains, VibH, was solved (Fig. 4b). VibH is a stand-alone C domain which catalyses the condensation of 2,3-dihydroxybenzoate (DHB), loaded on the aryl carrier protein (ArCP) of VibB, with norspermidine. It differs from ‘standard’ C domains as norspermidine is not covalently attached to a PCP domain but is free in solution. The determination of

**Fig. 4.** C domains of NRPS. (a) Schematic representation of the condensation reaction between the amino acid loaded onto the PCP domain of the  $(n+1)$  module (in the acceptor site) and the peptidyl chain on the PCP domain of the  $n$  module (in the donor site). (b) Structure of the stand-alone C domain VibH from the vibriobactin synthetase. The figure was created from the PDB file 1L5A using Swiss PDB viewer and rendered with POVray.

the crystal structure of VibH by Keating *et al.* (2002) revealed the existence of a solvent channel and allowed the modelling of a plausible position for VibB-S-DHB, with the 4'Ppant cofactor extending within the solvent channel. In a 'standard' C domain, it is expected that the 4'Ppant cofactor from the second PCP domain would enter the solvent channel from the opposite face of the C domain. However, since no structure of the enzyme in the presence of its substrates or of substrate analogues has been solved, it has not been possible to unravel potential rules governing C domain substrate specificity from this study.

Examining substrate specificity of C domains is not an easy task since the natural substrates of these domains are peptide chains or amino acids bound to the 4'Ppant arm of PCP domains. Studies aimed at determining substrate specificity of C domains have been facilitated by the use of substrate analogues, such as aminoacyl or peptidyl-*N*-acetyl cysteamine thioesters (SNAC). The SNAC moiety mimics the extremity of the 4'Ppant arm. At their acceptor (aminoacyl) site (Fig. 4a), C domains exhibit a strong stereoselectivity (L- or D-form of the amino acid) (Belshaw *et al.*, 1999; Ehmann *et al.*, 2000; Linne & Marahiel, 2000; Luo *et al.*, 2002) as well as some selectivity towards the side chain of the aminoacyl thioester (Ehmann *et al.*, 2000). Such selectivities imply the probable existence of a binding pocket of the amino acid at the acceptor site, and explain why elongation modules cannot initiate peptide chain synthesis and how directionality of nonribosomal peptide biosynthesis is controlled (Linne & Marahiel, 2000). However, condensation of noncognate amino acyl acceptors is possible, as variants of the main nonribosomal peptide can sometimes be observed *in vivo* [tyrocidine A, B, C and D (Ruttenberg & Mach, 1966) or Val7- and Leu7-surfactins (Peypoux & Michel, 1992)].

Broader substrate specificity is observed at the donor (peptidyl) site (Fig. 4a) of C domains, which has been shown to accept noncognate substrates, e.g. amino acyl instead of peptidyl moieties, or peptidyl chains with different size and amino acid composition (Belshaw *et al.*, 1999; Doekel & Marahiel, 2000; Linne *et al.*, 2003; Marshall *et al.*, 2001), but no kinetic data are yet available to quantify the catalytic efficiency of these condensations. The stereochemistry of the C-terminal amino acid of the peptidyl chain appears, however, to be an important element in substrate recognition, as observed for the acceptor site (Ehmann *et al.*, 2000; Luo *et al.*, 2002). It is interesting that C domains appear to be more selective at their acceptor site than at their donor site, because this means that the C domain exhibits greater selectivity towards the amino acid activated by the A domain within the same module compared with those amino acids activated by A domains in upstream modules. This observation suggests that module swapping is likely to be a more effective strategy for engineering NRPSs than A domain swapping. Furthermore, it may limit attempts to achieve truly combinatorial biosynthesis of peptides using a generic A domain possessing relaxed substrate

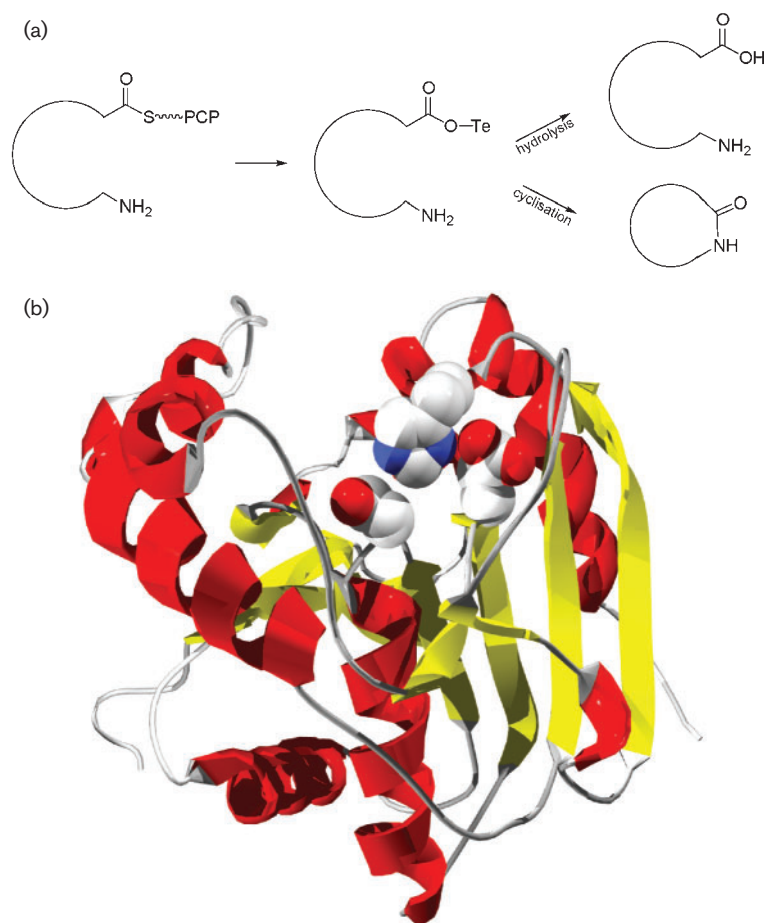
selectivity. From the evolutionary perspective this observation is also interesting because it suggests that new NRPS systems would probably evolve most effectively by recombination events that transfer both the A and C domains together between modules. The reason why C domain substrate selectivity has evolved is unclear since even moderate A domain selectivity would probably generate enough of a peptide containing the desired amino acid in the appropriate position to confer a selective advantage. Thus utilization of the C domain as a second selectivity mechanism to 'weed out' miscognate substrates loaded onto the PCP domain by the A domain would appear to be unnecessary. One possible explanation for the evolution of C domain selectivity is that it is required to ensure that the overall integrity of the peptide assembly process, in particular the order and timing of condensation reactions, is maintained.

With regard to the above discussion, it is important to bear in mind that studies of C domain specificity are still limited to a few examples. More detailed studies of a wider number of C domains will be required to better define substrate recognition at both the donor site and the acceptor site. In this respect, the use of systems incorporating A domains with relaxed substrate selectivity would eliminate the need to use substrate analogues such as the aminoacyl (peptidyl)-SNACs, for which the effective concentration is very likely to be different from the effective concentration of the natural PCP-bound aminoacyl (peptidyl) substrate. Nevertheless, it is already clear that C domains will not accept every substrate and that their substrate selectivity (especially stereoselectivity) is an important and potentially limiting factor in engineered biosynthesis experiments.

### Substrate recognition by TE domains

The Te domain is the last domain to play a role in the biosynthesis of a nonribosomal peptide, releasing the peptide into solution. It catalyses either the hydrolysis or the intramolecular cyclization (with or without oligomerization) of the peptidyl chain (Fig. 5a), leading to a variety of structures ranging from linear peptides (vancomycin) to cyclic peptides (tyrocidine A), oligomeric macrolactones (enterobactin) or branched-chain cyclic (lipo)peptides (surfactin A or bacitracin A). In 2002, the first structure of a Te domain (Fig. 5b), excised from SrfA-C, an NRPS involved in the biosynthesis of the lipopeptide antibiotic surfactin A, was solved (Bruner *et al.*, 2002). It showed that Te domains belong to the family of  $\alpha/\beta$ -hydrolases, with a catalytic triad defined by the residues Ser80, His207 and Asp107, which catalyses cleavage of the thioester bond between the peptide and the 4'Ppant cofactor to form a peptidyl-O-Te intermediate. This intermediate is then released in solution by the attack of a water molecule (hydrolysis) or of an intramolecular nucleophile (cyclization).

Cocrystallization of Srf-Te with substrate analogues or boronate inhibitors (Bruner *et al.*, 2002; Tseng *et al.*, 2002)



**Fig. 5.** T domains of NRPS. (a) Schematic representation of the hydrolysis or cyclization of the peptide catalysed by Te domains. (b) Structure of the surfactin Te domain, excised from SrfA-C. The residues forming the Ser-His-Asp catalytic triad are illustrated in space-filling representation. The figure was created from the PDB file 1JMK using Swiss PDB viewer and rendered with POVray.

has led to the identification of hydrophobic binding pockets for the two C-terminal amino acids (D-Leu6 and L-Leu7) of the substrate. Although no defined electron density was obtained for the remainder of the peptidyl chain, modelling of the substrate into the structure of the Te domain positioned it into a mainly hydrophobic cavity with the N-terminal extremity of the peptide lying adjacent to the active site Ser residue.

Substrate specificity for the cyclization reaction has been extensively studied for the excised Te domains of tyrocidine synthetase (TycC-Te) and surfactin synthetase (Srf-Te), catalysing, respectively, a macrolactamization and a macrolactonization (Kohli *et al.*, 2001, 2002a, b; Trauger *et al.*, 2000; Tseng *et al.*, 2002). From these studies, it appears that Te domains are selective for the side chain of residues adjacent to the site of formation of the peptidyl-O-Te intermediate (D-Leu6 and L-Leu7 for surfactin and L-Orn9 for tyrocidine). These data are in good agreement with information obtained from the crystal structure of Srf-Te, which has shown the presence of binding pockets for D-Leu6 and L-Leu7. Side chain selectivity and enantioselectivity is also observed for residues next to the intramolecular nucleophile (Glu1 for surfactin, D-Phe1 for tyrocidine), and results obtained by Kohli *et al.* (2002a) are strongly supportive of the existence of a binding pocket for

D-aromatic residues in the N-terminal position of the TycC-Te substrate. Nonetheless, in the case of Srf-Te, no binding pocket for the side chain of Glu1 was identified in the crystal structure.

Studies of TycC-Te have also underlined the importance of substrate preorganization in a product-like conformation, favoured by intramolecular hydrogen bonds between peptide bonds. Backbone amide bonds next to the cyclization junction seem particularly important (Trauger *et al.*, 2001). In the same manner, the twisted  $\beta$ -sheet structure of surfactin suggests some form of substrate preorganization prior to cyclization. This hypothesis is supported by the data of Tseng *et al.* (2002).

The influence of the leaving group (replacement of the thioester bond by an ester bond), of the length of the peptidyl chain, and of the nature of the cyclizing nucleophile have also been studied for both TycC-Te and Srf-Te. In these cases, however, no clear picture emerges from the results obtained, which are quite different for the two Te domains examined. Overall, Srf-Te appears to be relatively specific for its natural substrate, surfactin (Tseng *et al.*, 2002), whereas Tyc-Te has a rather broad substrate specificity (Kohli *et al.*, 2002b), which has allowed its use, in combination with solid phase peptide synthesis, for the

creation of libraries of cyclic molecules, which can then be screened for biological activities (Kohli *et al.*, 2002a).

Most of the studies on NRPS Te domains have focused on the cyclase activity of these domains. Schwarzer *et al.* (2001), however, have shown that Te domains are able, when fused to NRPSs, to hydrolyse the peptidyl chain from the enzyme, although with variable catalytic efficiency. The only Te domain with natural hydrolase activity was one of the less efficient Te domains tested, and it would be interesting to see if this reflects a more stringent substrate specificity of Te domains with hydrolase activity.

It seems difficult, from the studies carried out up to now, to draw general conclusions regarding the substrate specificity of Te domains, with situations as different as the quite specific Srf-Te and the largely permissive TycC-Te. Yet it is interesting that some Te domains appear to have a sufficiently relaxed substrate specificity to allow their use as cyclization catalysts for a wide range of linear peptidyl thioester substrates. It can be concluded that in at least some cases the high substrate specificity of Te domains will severely limit the range of analogues that can be generated by engineered biosynthesis strategies.

### Impact of interdomain recognition

Two cases must be considered regarding the role of interdomain recognition in NRPS substrate specificity: when domains are next to each other on the same polypeptide chain (domains acting *in cis*) or when they belong to two distinct proteins (domains acting *in trans*), as NRPSs can be composed of several polypeptide chains. Limited data are available regarding the importance of interdomain recognition for domains acting *in cis* and no systematic study on the subject has yet been undertaken. However, it appears that the specificity of a domain (A domains; Doekel & Marahiel, 2000) or its activity (heterocyclization domains; Marshall *et al.*, 2001) can be affected by the environment of this domain, i.e. by the presence or absence and by the nature of the domains surrounding it.

Several studies have brought to light the importance of protein–protein interactions for substrate recognition in the case of domains acting *in trans*. Thus in the tyrocidine synthetase, composed of three polypeptide chains TycA, TycB and TycC, Linne *et al.* (2003) demonstrated the major role played by the epimerization domains constituting the C-terminal domain of TycA and TycB in protein–protein recognition between domains interacting *in trans*. In another study, Marshall *et al.* (2002) established the importance of interdomain recognition between A and PCP domains interacting *in trans* in siderophore systems for A domain specificity. Their results also suggest that residues involved in interdomain recognition for domains interacting *in trans* are specific to each NRPS. Although this could be a problem for the design of hybrid NRPSs, Linne *et al.* (2003) have shown that domains interacting

*in trans* can be fused on a single polypeptide chain with only limited loss or even an increase in activity.

### Conclusion and future directions

Understanding the mechanisms of NRPS substrate recognition is a difficult task since these multi-enzymes are composed of numerous domains catalysing various reactions. Deconvoluting NRPS substrate specificity into the substrate specificity of each of its constitutive domains has been the principal strategy adopted up to now. It has shown that A domains are the principal domains exerting control over substrate selection. Yet, A domains are not the exclusive determinants of substrate selection and studies have demonstrated that C or Te domains also exhibit a certain level of substrate selectivity.

It is clear that our understanding of the mechanisms of substrate selection by the various catalytic domains of NRPSs – even A domains, which have arguably been studied the most extensively – is still relatively rudimentary. Further studies will be required to elucidate the molecular basis of substrate selectivity by NRPS domains, especially the optional domains for which very little data are available at the moment. Such studies will be invaluable aids in the design of engineered biosynthesis strategies for the production of new metabolites, in terms of both identifying appropriate modifications for producing novel structures and defining the limits of the structural diversity that will be accessible via such strategies. Information regarding solely the selectivity of isolated domains may, however, not provide a complete picture because interdomain interactions also seem to play a significant role in substrate recognition. Defining substrate selectivity at the level of a module, as done by Luo *et al.* (2001) for the initiation module of gramicidin S synthetase, may therefore prove a worthwhile general strategy.

In conclusion, there is no doubt that studies of substrate selection by NRPSs will continue to be an important area of research, which promises not only to enrich our understanding of the molecular mechanisms of selectivity in NRPSs, but also to serve as a model for elucidating selectivity mechanisms in other protein families that interact with a wide range of chemically diverse substrates. In an era when, thanks to genomics, new members are added to protein families far more rapidly than they can be characterized, methods for analysing, understanding and predicting the substrate selectivity of individual members of protein families are likely to become increasingly important.

### References

- Ackerley, D. F., Caradoc-Davies, T. T. & Lamont, I. L. (2003). Substrate specificity of the non-ribosomal peptide synthetase PvdD from *Pseudomonas aeruginosa*. *J Bacteriol* **185**, 2848–2855.
- Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489.

- Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M. & 40 other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.
- Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T. & Stubbs, M. T. (2002). Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure (Camb)* **10**, 301–310.
- Cerdeno, A. M., Bibb, M. J. & Challis, G. L. (2001). Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chem Biol* **8**, 817–829.
- Challis, G. L. & Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol Lett* **187**, 111–114.
- Challis, G. L., Ravel, J. & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211–224.
- Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174–4183.
- Doekel, S. & Marahiel, M. A. (2000). Dipeptide formation on engineered hybrid peptide synthetases. *Chem Biol* **7**, 373–384.
- Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol* **7**, 765–772.
- Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. (2002). Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726.
- Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* **9**, 522–526.
- Kohli, R. M., Trauger, J. W., Schwarzer, D., Marahiel, M. A. & Walsh, C. T. (2001). Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* **40**, 7099–7108.
- Kohli, R. M., Walsh, C. T. & Burkart, M. D. (2002a). Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658–661.
- Kohli, R. M., Takagi, J. & Walsh, C. T. (2002b). The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides. *Proc Natl Acad Sci U S A* **99**, 1247–1252.
- Linne, U. & Marahiel, M. A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**, 10439–10447.
- Linne, U., Stein, D. B., Mootz, H. D. & Marahiel, M. A. (2003). Systematic and quantitative analysis of protein-protein recognition between nonribosomal peptide synthetases investigated in the tyrocidine biosynthetic template. *Biochemistry* **42**, 5114–5124.
- Luo, L., Burkart, M. D., Stachelhaus, T. & Walsh, C. T. (2001). Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. *J Am Chem Soc* **123**, 11208–11218.
- Luo, L., Kohli, R. M., Onishi, M., Linne, U., Marahiel, M. A. & Walsh, C. T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**, 9184–9196.
- Marshall, C. G., Burkart, M. D., Keating, T. A. & Walsh, C. T. (2001). Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate. *Biochemistry* **40**, 10655–10663.
- Marshall, C. G., Burkart, M. D., Meray, R. K. & Walsh, C. T. (2002). Carrier protein recognition in siderophore-producing nonribosomal peptide synthetases. *Biochemistry* **41**, 8429–8437.
- May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. (2002). Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99**, 12120–12125.
- Peypoux, F. & Michel, G. (1992). Controlled biosynthesis of Val<sup>7</sup>- and Leu<sup>7</sup>-surfactins. *Appl Microbiol Biotechnol* **36**, 515–517.
- Ruttenberg, M. A. & Mach, B. (1966). Studies on amino acid substitution in the biosynthesis of the antibiotic polypeptide tyrocidine. *Biochemistry* **5**, 2864–2869.
- Schwarzer, D., Mootz, H. D. & Marahiel, M. A. (2001). Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem Biol* **8**, 997–1010.
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493–505.
- Thomas, M. G., Burkart, M. D. & Walsh, C. T. (2002). Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodiginosin and pyoluteorin biosynthesis. *Chem Biol* **9**, 171–184.
- Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215–218.
- Trauger, J. W., Kohli, R. M. & Walsh, C. T. (2001). Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* **40**, 7092–7098.
- Tseng, C. C., Bruner, S. D., Kohli, R. M., Marahiel, M. A., Walsh, C. T. & Sieber, S. A. (2002). Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**, 13350–13359.