Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains

Gregory L Challis, Jacques Ravel and Craig A Townsend

Background: Nonribosomal peptide synthetases (NRPSs) are large modular proteins that selectively bind, activate and condense amino acids in an ordered manner. Substrate recognition and activation occurs by reaction with ATP within the adenylation (A) domain of each module. Recently, the crystal structure of the A domain from the gramicidin synthetase (GrsA) with L-phenylalanine and adenosine monophosphate bound has been determined.

Results: Critical residues in all known NRPS A domains have been identified that align with eight binding-pocket residues in the GrsA A domain and define sets of remarkably conserved recognition templates. Phylogenetic relationships among these sets and the likely specificity determinants for polar and nonpolar amino acids were determined in light of extensive published biochemical data for these enzymes. The binding specificity of greater than 80% of the known NRPS A domains has been correlated with more than 30 amino acid substrates.

Conclusions: The analysis presented allows the specificity of A domains of unknown function (e.g. from polymerase chain reaction amplification or genome sequencing) to be predicted. Furthermore, it provides a rational framework for altering of A domain specificity by site-directed mutagenesis, which has significant potential for engineering the biosynthesis of novel natural products.

Address: Department of Chemistry, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA.

Correspondence: Craig A Townsend
E-mail: townsend@jhunix.hcf.jhu.edu

Key words: adenylation domain, amino acid activation, nonribosomal peptide synthetase, substrate binding, substrate specificity

Received: **7 October 1999**
Revisions requested: **2 November 1999**
Revisions received: **6 December 1999**
Accepted: **14 December 1999**

Published: **23 February 2000**

Chemistry & Biology 2000, **7**:211–224

1074-5521/00/$ – see front matter
© 2000 Elsevier Science Ltd. All rights reserved.

Introduction

A wide range of biologically active peptides is synthesized by nonribosomal peptide synthetases (NRPSs) in bacteria and fungi. These metabolites show functional diversity and include antibiotics (e.g. penicillins, vancomycin, bacitracin and gramicidin), siderophores (e.g. mycobactin and enterobactin), toxins (e.g. syringomycin and HC-toxin) and immunosuppressive agents (e.g. rapamycin and cyclosporin). The biosyntheses of nonribosomal peptides share a common mechanism, which has been widely investigated (for reviews see [1–3]). Briefly, NRPSs are large multifunctional enzymes of 100–1700 kDa [2] that have a modular structural organization and are thought to link amino acid residues according to the multiple-carrier thiotemplate mechanism [4–8]. The minimal module is composed of an amino-acid-activating domain, a thiolation domain and a condensation domain. The activation domain recognizes a substrate amino (or hydroxy) acid, usually specifically, and activates it as its acyl adenylate by reaction with ATP. This active ester is then covalently linked as its thioester to the enzyme-bound 4′-phosphopantetheine located within the module. The reaction continues by the direct transfer to another acylamino acid intermediate on the adjacent downstream module mediated by the condensation domain to form a peptide bond. In some cases, modifications (epimerization, N- or C-methylation or cyclization) are catalyzed by additional domains or by modified domains within a module. A terminal thioesterase is frequently present to release the peptide from the enzyme by cyclization or hydrolysis.

In most NRPSs the organization and order of the modules maps in a 1:1 manner to the amino acid sequence of the peptide products (co-linearity rule). This paradigm has allowed researchers to assign a specific amino acid activation function to each module. Exceptions have appeared recently, however, that seem to violate this paradigm. Nocardicin A is a monocyclic β-lactam antibiotic known to be derived from nocardicin G [9]. A tripeptide composed of L-serine and two units of 4-hydroxy-L-phenylglycine (HPG) would seem the logical precursor of nocardicin G [10]. We have recently determined, however, that a pair of NRPSs is contained in the nocardicin A biosynthetic cluster encompassing five modules rather than the expected three (S.D. Breazeale, M. Gunsior and C.A.T., unpublished observations). In addition, *Pseudomonas syringae* syringomycin is a cyclic lipodepsinonapeptide whose synthetase genetic organization does not respect the co-linearity rule [11]. The module activating the carboxy-terminal threonine encoded by the gene *syrB1* is located upstream of *syrE*, which encodes the first eight amino-acid-activating modules. *Mycobacterium smegmatis* produces exochelin MS, a pentapeptide siderophore [12]. Genetic analysis has revealed the presence of six NRPS modules rather than the expected five. The role of the

sixth module could not be assigned [13]. The genetic organization of the bleomycin biosynthetic cluster has been elucidated recently and has indicated the presence of 11 NRPS modules for eight amino acids units in the peptide product [14].

In the past decade, enormous progress has been made in the discovery and sequencing of NRPS genes from bacteria and fungi (for a comprehensive list see [3] and Tables S1 and S2 in the Supplementary material section). Many of these putative peptide synthetases produce peptides of unknown structure, but all share the well-conserved characteristic signature sequences of NRPSs summarized by Marahiel *et al.* [2].

Specificity is principally, but not exclusively, controlled at the adenylation/ pantetheinylation step [1,15]. Conti *et al.* [16] have recently identified, by crystal structure analysis, the substrate-binding pocket of the phenylalanine adenylation (A) domain of the gramicidin S synthetase (GrsA) from *Bacillus brevis*. The amino acids lining the substrate-binding pocket mediate amino acid specificity. The authors concluded that, because of the high degree of sequence identity among NRPS activation domains, the GrsA structure represents a structural model that all NRPS activation domains could fit [16].

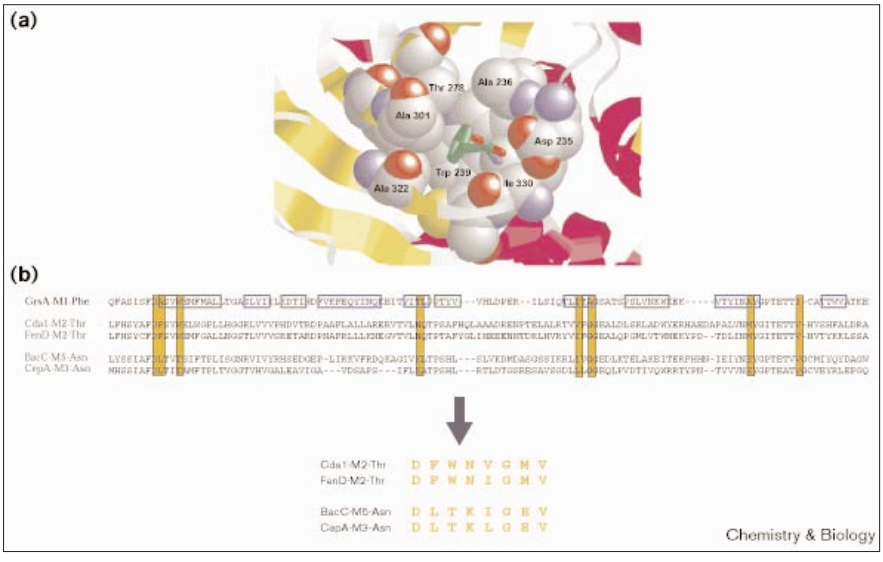
Motivated by the exceptions to the co-linearity rule seen in our laboratory and cases emerging in the literature, we have developed a method to reveal the amino acid building blocks activated by individual adenylation domains on the basis of their sequence. Here we describe the protein sequence analysis of over 150 NRPS activation domains to pinpoint the essential amino acid residues involved in substrate specificity and binding. As this manuscript was nearing completion, a similar approach was described by Stachelhaus *et al.* [17], in which phylogenetic and structure-based analyses identified these essential residues and allowed a correlation with substrate specificity to be established. Although the X-ray structure of GrsA formed the common basis for our two approaches, we have both reinforced, and extended the analysis to include 33 amino acid substrates (Table S1) for which the function of adenylation domains has been assigned by either direct enzymic assay, or convincing deduction, and for which multiple protein sequences are available of A domains recognizing specific amino acids. Care has been taken to assemble an accurate and comprehensive database to ensure the integrity of these correlations. In many cases it is possible to suggest the chemical basis for substrate selectivity, whereas in others, distinct binding motifs have been identified that pose interesting questions for further research to understand the origins of their specificity at a fundamental level. We present an algorithm at the first level for the prediction of function of uncharacterized A domains translated from newly isolated gene sequences, which provides at a second level a framework for rational mutagenesis to alter the selection of amino acid for activation at a given module. Eight amino acid residues have been identified whose order and alignment can be readily applied to refine and advance current phylogenetic analyses. Application of this structural mask has made it possible for A domains of diverse evolutionary origins to be correlated directly to function with far greater accuracy than anticipated at the outset of this study.

Results and discussion

Analysis of adenylation domains using phylogenetic tools

Phylogenetic analysis and sequence alignments have been applied to A domains and revealed limited correlation with function [18–23]. The small number of sequences, or

Figure 1



(a) The substrate-binding pocket of the gramicidin S synthase (GrsA) phenylalanine-activating domain, viewed from above. The eight residues thought to be important for substrate selection are indicated (Ile299 is obscured by Ala301). The figure was generated from 1amu.pdb using RasMol. **(b)** Identification of the eight amino acids lining the substrate binding pocket of four A domains from different genera: *Streptomyces* calcium-dependent antibiotic synthetase Cda1-M2 and *Bacillus* fengycin synthetase FenD-M2, which both activate L-threonine; *Amycolatopsis* chloroeremomycin synthetase CepA-M3 and *Bacillus* bacitracin synthetase BacC-M5, which both activate L-asparagine. The sequence of each A domain was aligned to the sequence of GrsA, and critical amino acids were identified. The secondary structural elements in the GrsA sequence are indicated: α helices are boxed in green and β sheets are boxed in blue.

only a single genus, used and the choice of a large number of residues make these analyses of limited value.

On the basis of the assumption that the eight amino acids lining the binding pocket (Figure 1a) determine substrate specificity, we used phylogenetic tools to establish a relationship between these important residues and substrate specificity in 154 A domains in the database. We believe that these residues are sufficient to both predict and, in many cases, rationalize the substrate specificity of NRPS A domains.

To validate the theory, the eight residues lining the binding pocket of the A domain were obtained from all known sequences in the database, after aligning each to that of the phenylalanine A domain of GrsA using the ClustalX multiple alignment software (Figure 1b) [24]. Two phylogenetic trees were constructed, one based on the 180–200 amino acids spanning motifs A3–A6 (Figure S1 in the Supplementary material section) as described previously by Konz *et al.* [20] and the other based on the eight critical residues (Figure S2). Tables S1 and S2 in the Supplementary material list the identity of these residues and their position according to the numbering of GrsA [16]. In contrast to the analysis of Stachelhaus *et al.* [17], we chose to exclude Cys331 because the sidechain of this residue points away from the specificity pocket in GrsA [16]. This observation does not, however, exclude the possibility that residues in this position could participate in substrate selection in other A domains. Lys517 was omitted because it is strictly conserved and is not involved in substrate discrimination.

The functional clustering observed when limiting the analysis to the eight amino acids lining the binding pocket (Figure S2) is far greater than that observed in the analysis of 180–200 amino acids encompassing motif A3–A6, which shows clustering predominantly among genus lines (Figure S1). This is certainly because the phylogenetic analysis of 180 amino acids is more likely to emphasize genus-specific mutational evolution than functional evolution of the protein studied. This result supports the importance of the residues lining the binding pocket in defining the function of each binding pocket, as observed similarly by Stachelhaus *et al.* [17]. A dichotomy between eukaryotic and prokaryotic domains is still observed, however, but to a lesser extent.

Prokaryotic activation domains appear to be the result of an evolutionary mechanism involving horizontal transfer. The analysis of whole modules by sequence alignments shows that a greater homology is observed in the regions flanking the A domain (data not shown) [16]. Moreover, the functional clustering (Figure S1) among prokaryotic A domains makes it unlikely that they arose by convergent evolution, and there is no evidence for an ancient activation domain originating by transfer from a eukaryotic ancestor (Figures S1, S2 in the Supplementary material). In some cases, such as the leucine A domains of GrsB and TycC, it is possible to imagine an independent evolutionary mechanism. These A domains do not cluster according to their function on either tree. A more detailed phylogenetic analysis would be needed to confirm such an evolutionary hypothesis for this important group of enzymes.

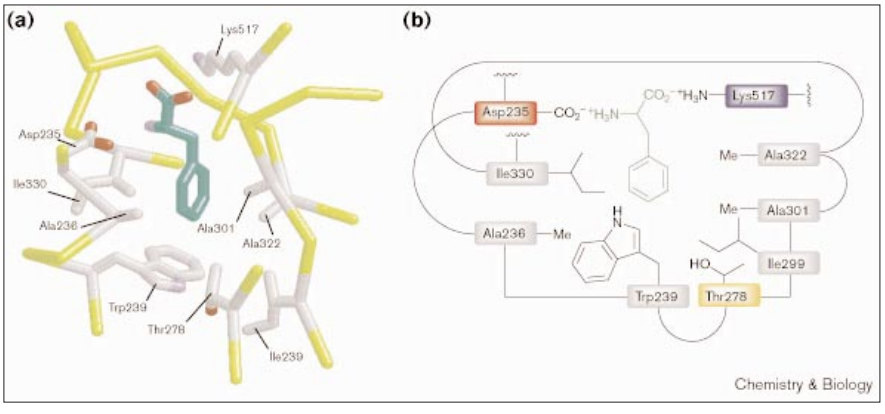
The phylogenetic tree in Figure S2 can serve as a tool for predicting the amino acid selectivity of activation domains and for discerning evolutionary relationships among synthetases of diverse origin. A more detailed analysis is required, however, to identify important residues for substrate recognition through, for example, hydrogen bonding, electrostatic or Van der Waals interactions.

Functional analysis of amino-acid-activating domains

We first generated a schematic of the specificity pocket based on the crystal structure of the phenylalanine-activating domain of the gramacidin S synthetase GrsA (Figure 2). Schematics of the predicted specificity pocket structures

Figure 2

(a) Three-dimensional structure generated from 1amu.pdb using RasMol (version 2.6) and **(b)** two-dimensional representation of the specificity pocket from the phenylalanine-activating domain of GrsA. In all the two-dimensional representations, residue coding is as follows: acidic, red; basic, blue; neutral polar, yellow; hydrophobic, gray. Hy represents a variable hydrophobic residue; N represents a nonconserved residue.



for each A domain in the database were then produced by mapping to this template the eight critical residues identified by the sequence alignments. In general, the predicted specificity pocket structures fall into one of two groups, depending on the chemical nature of the substrate's sidechain. For domains that activate amino acids with polar sidechains, one or two universally conserved polar residues can be identified that presumably interact with the substrate sidechain through hydrogen bonding and electrostatic interactions. The remaining residues in the specificity pocket are hydrophobic and vary among domains that activate the same amino acid. In contrast, for substrates with hydrophobic sidechains, the residues lining the specificity pockets are all hydrophobic, except for proline and pipecolate. Although some of these residues appear to be highly conserved, variation in other residues seems to be tolerated. This finding, coupled with the observation that many of these domains have a lower substrate selectivity than those that activate polar amino acids, makes it difficult to assign functional roles for individual residues. The following discussion is, therefore, split into two main sections on domains that activate substrates with polar sidechains and those that activate substrates with hydrophobic sidechains.

Substrates with polar sidechains

Serine-, threonine- and 4-hydroxyphenylglycine-activating domains

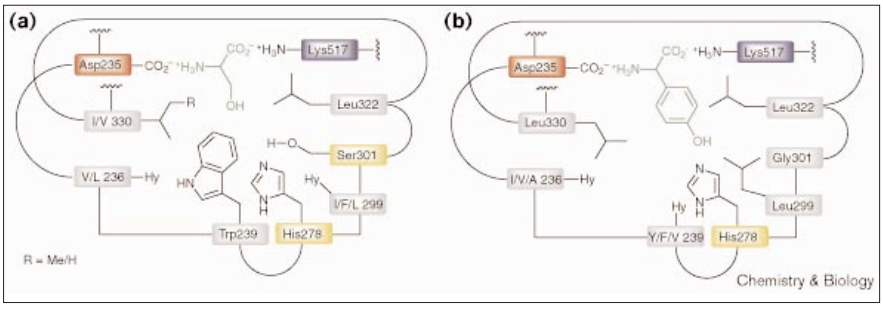
The enterobactin synthetase EntF and the first two modules of the syringomycin synthetase SyrE have been cloned, overexpressed, purified and shown to specifically activate L-serine [11,25]. Sequence alignment of these proteins and the putative L-serine-activating domain from the calcium-dependent antibiotic (CDA) synthetase Cda1 predicts highly conserved structures for the specificity pockets (Figure 3a). In particular, the His278/Ser301 dyad is universally conserved. Mapping of this dyad to the GrsA structure suggests a functional role in serine selection, which is hydrogen bonding of the His278 nitrogen to the Ser301 hydroxyl group, which in turn hydrogen bonds to the substrate hydroxyl group.

Although no domains that specifically activate 4-hydroxy-L-phenylglycine (HPG) have been biochemically characterized, the predicted specificity pocket structures for the putative HPG-activating domains from the chloroeremomycin synthetase CepB and the CDA synthetase are highly conserved (Figure 3b). Furthermore, comparing the predicted structures for the L-serine- and HPG-activating domains reveals that they have a remarkably similar architecture. The principal difference is the substitution of Ser301 in the former with Gly301 in the latter. From a functional perspective, this could be explained by direct hydrogen bonding of the HPG hydroxyl group to the His278 nitrogen. In contrast, the analysis of Stachelhaus *et al.* [17] places a leucine residue at this position. Our analysis indicates that serine- and HPG-activating domains could be functionally interconverted by the single point mutation Ser301↔Gly301.

Functional characterization of the cloned syringomycin synthetase SyrB has shown that it specifically activates L-threonine [11]. Wild-type actinomycin synthetase AcmB also activates L-threonine [26]. Sequence alignment of these domains with the putative L-threonine-activating domains from fengycin synthetases FenD and Pps2, pristinamycin I synthetase SnbC, exochelin synthetase FxbC, pyoverdin synthetase PvdD and CDA synthetase Cda1 reveals highly conserved predicted structures for the specificity pocket (Figure 4a). Interestingly, and unexpectedly, these structures are quite different from those predicted for the specificity pocket of the L-serine-activating domains (Figure 3a). This difference presumably reflects the need to specifically activate threonine in the presence of serine. A comparison of the predicted structures of the threonine-activating domains with that of the mycobactin synthetase MbtB module 1 (Figure 4b), which ostensibly contains a domain capable of activating both L-serine and L-threonine, suggests how this might be achieved. In our analysis, Asn278 is universally conserved (unlike in [17]) and probably interacts with the threonine hydroxyl. There are, however, a number of other amino acid substitutions (Phe236→Met, Trp239→Leu and Met322→Leu), that presumably diminish specificity in the MbtB A domain.

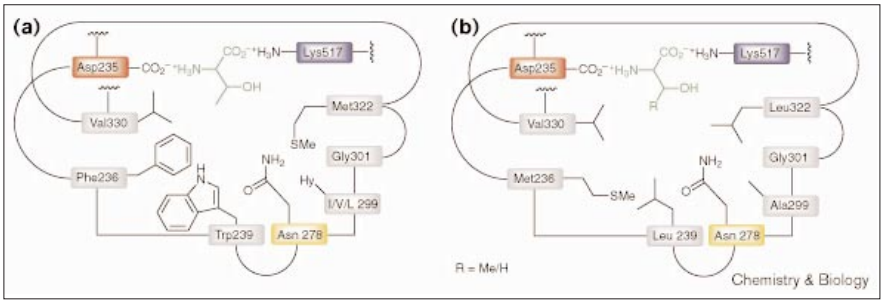
It should be noted that the excluded residue 331 is histidine in these cases. For this residue to be involved in hydrogen

Figure 3



(a) Predicted structure for the specificity pocket of L-serine-activating domains based on alignment of the Cda1-M1, EntF, SyrE-M1 and SyrE-M2 sequences. **(b)** Predicted specificity pocket structure for HPG-activating domains based on alignment of the Cda1-M6, CepB-M1 and CepB-M2 sequences. For description of the domain nomenclature see the legend for Table S1.

Figure 4



(a) Predicted structure for the specificity pocket of L-threonine-activating domains based on alignment of SyrB, AcmB-M1, Cda1-M2, FenD-M2, Pps2-M2, FxbC-M2 and SnbC-M1. **(b)** Structure predicted for the L-threonine/L-serine-activating domain from MbtB-M1.

bonding to the substrate hydroxyl, however, a conformational change in the specificity pocket would be required.

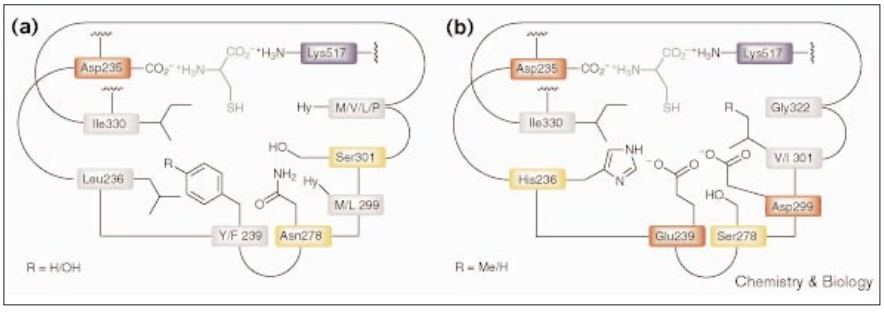
To test further our approach for assigning the specificity of amino-acid-activating domains, the predicted specificity pocket structures of the putative dehydrothreonine-activating domain of syringomycin synthetase SyrE [17] and L-threonine-activating domains were compared. This analysis indicated that the 'dehydrothreonine' and threonine specificity pockets are identical in structure. It is suggested, therefore, that the substrate for this domain is actually L-threonine; the dehydroamino acid probably being formed through dehydration of the peptide product of the NRPS. Furthermore, exclusion of a dehydroamino acid is consistent with chemical expectation that such an enamine species would be neither stable in solution, nor capable of propagating peptide elongation.

Cysteine-activating domains

Cloning and expression of residues 1–1491 of the yersiniabactin synthetase Irp2 (HMPW2) have shown that it activates L-cysteine [27]. Sequence alignment of this protein with the putative cysteine-activating domains from bacitracin synthetase BacA, anguibactin synthetase AngR and pyochelin synthetases PchE and PchF predicts a highly conserved structure for the specificity pocket (Figure 5a). The universally conserved Asn278 is presumably involved in substrate selection through interaction with the cysteine thiol. A comparison of these structures with those predicted for the threonine-activating domains (Figure 4a) reveals that they share this residue. Aside from relatively conservative changes among the hydrophobic residues in the specificity pocket structures for these two substrates, the only major difference is the Ser301→Gly substitution in the threonine-activating domains. Ser301 therefore probably plays an important role in selecting cysteine over threonine. As hydrogen bonding between the Ser301 sidechain hydroxyl group and the sidechain sulfur of cysteine is unlikely to occur, it is not easy to assign a functional role for this residue. A test of its role in substrate selectivity would be to examine the effect of the mutation Ser301↔Gly.

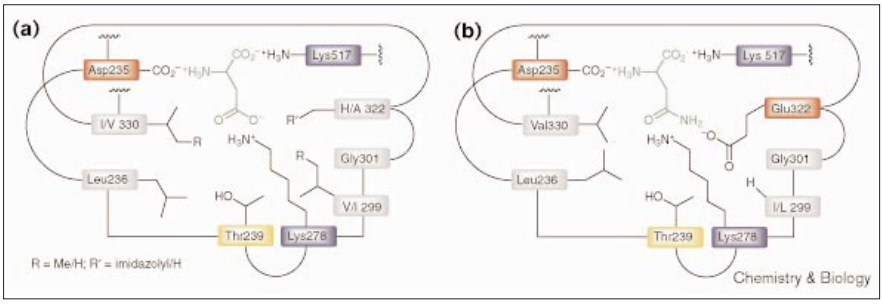
A comparison of the structures for the cysteine-activating domains discussed above with those predicted for the L-cysteine-activating domains from the several known ACV synthetases (e.g. PcbAB and AcvA from *Nocardia lactamdurans* and *Aspergillus nidulans*, respectively) shows that they are quite different (Figure 5b). In fact, the very highly conserved ACV synthetase cysteine-activating domain is closer in structure to that predicted for the L-valine-activating domain from the same synthetases, consistent with the isolation of L-(α-aminoadipoyl)-L-valinyl–D-valine from large scale fermentations of *Cephalosporium acremonium* [28]. These observations are in keeping with the evolution of ACV synthetases through gene duplication and subsequent functional divergence.

Figure 5



(a) Predicted structures for the specificity pocket for the L-cysteine-activating domain from Irp2 and the putative L-cysteine-activating domains from BacA-M2, AngR and PchE and PchF. **(b)** Predicted specificity pocket structures for AcvA and PchAB.

Figure 6



(a) Predicted structures for the specificity pockets of L-aspartate-activating domains based on alignment of LicB-M2, LchAB-M2, SrfAB-M2, BacC-M4, Cda1-M4, Cda1-M5 and Cda2-M1. **(b)** Structures predicted for the L-asparagine-activating domains from BacC-M5, TycC-M1, CepA-M3 and Cda2-M3. Where two amino acid residues are specified at a given position the first refers to a *Bacillus*-derived domain and the second refers to a *Streptomyces*-derived domain.

Aspartate-, asparagine-, glutamate- and glutamine-activating domains

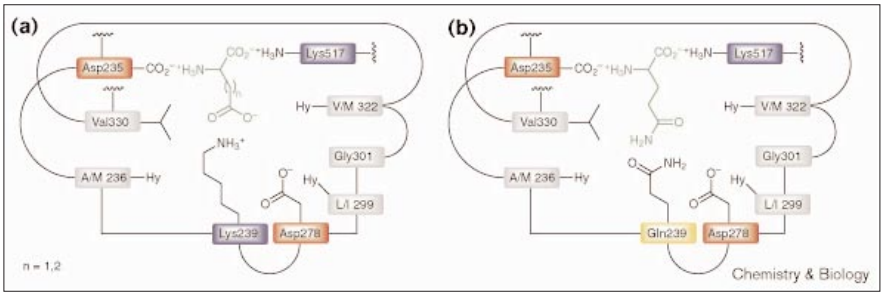
Recombinant A domain from the second module of lichenysin synthetase LicB was shown to specifically activate L-aspartate [20]. The purified trimodular surfactin synthetase SrfAB has also been shown to activate L-aspartate, consistent with the deduced role of the second module incorporating this amino acid into surfactin [29]. Alignment of these sequences with the putative aspartate-activating domains from bacitracin synthetase BacC, lichenysin synthetase LchAB and CDA synthetases Cda1 and Cda2 predicts a highly conserved structure for the specificity pocket. The three differences, which are genus specific, are the substitutions His322→Ala, Val299→Ile and Ile330→Val (*Bacillus*-derived→*Streptomyces*-derived; Figure 6a). It seems unlikely that these differences are functionally significant, although the specificity of any of the *Streptomyces*-derived domains has yet to be examined experimentally. The universally conserved Lys278 probably forms salt bridges with the carboxylate group of the substrate sidechain. In contrast, Stachelhaus *et al.* [17] have suggested a key interaction between the substrate and His322. This residue, however, is replaced by alanine in *Streptomyces coelicolor*, making the proposed importance of this interaction unlikely. Additionally, Thr239 is conserved and might contribute to substrate selection.

The A domain of the first module of tyrocidine synthetase TycC has been shown to activate specifically L-asparagine [30]. Alignment of this sequence with the putative L-asparagine A domains from bacitracin synthetase BacC, chloroeremomycin synthetase CepA and the CDA synthetase Cda2 predicts highly conserved structures for the asparagine specificity pockets, which are very similar to the aspartate-activating domains. The predicted structures for both domains contain Lys278. The His/Ala322 residue in the structure of the aspartate-activating domains is, however, replaced by a universally conserved Glu322 residue in the asparagine-activating domain structure. Again, there are genus-specific variations in the hydrophobic residues 239 and 299 (Figure 6b). A functional role for Lys278 and Glu322 through hydrogen bonding to the substrate amide can be envisioned. Salt-bridge formation between the sidechain amino and carboxylate groups of these residues may anchor them close to each other in the pocket. Furthermore, the presence of both Lys278 and Glu322 renders the overall charge of the pocket neutral in accord with the overall charge on the substrate sidechain.

The functional significance of the Glu322 residue in the asparagine-activating domains has been demonstrated recently by Stachelhaus *et al.* [17]. It was shown that the single point mutation His322→Glu in the aspartate-specific activating domain from surfactin synthetase SrfAB completely altered the specificity to asparagine.

Comparison of the predicted specificity pocket structure for the apparent L-aspartate-activating domain from the syringomycin synthetase SyrE with those for the aspartate-activating domains discussed above reveals extensive differences. In fact, the structure is much closer to those predicted for the putative L-glutamate-activating domains from bacitracin synthetase BacA and surfactin synthetase SrfAA (Figure 7a). Partially purified trimodular SrfAA has been shown to activate not only L-glutamate, but also L-aspartate at approximately one third the rate [29]. It seems likely that the L-glutamate-activating domain also accepts L-aspartate as a substrate, because the other two modules in SrfAA are thought to incorporate L-leucine (see below). The activation domains from SyrE-M8, BacA-M4 and SrfAA-M1 therefore appear to have relaxed specificity that allows incorporation of glutamate or aspartate into their peptide products, depending on either the availability of the substrates in the cell, or perhaps a downstream selectivity mechanism [15]. For these domains, a conserved lysine residue occurs at position 239, as opposed to position 278 in the aspartate-specific domains. Additionally, Asp278 is conserved in the aspartate/glutamate-activating domains and may help modulate the relaxed specificity of these domains. As L-3-hydroxy-L-aspartate occurs in syringomycin, it is also possible, therefore, that this amino acid, not L-aspartate, is the substrate for the activation domain in SyrE.

Figure 7



(a) Predicted specificity pocket structures for the putative L-glutamate activating domains from SrfAA-M1 and BacA-M4, and the putative L-aspartate-activating domain from SyrE-M8; **(b)** Predicted specificity-pocket structure for the L-glutamine-activating domain from LicA-M1.

Cloning and overexpression of the activation domain from the first module of the lichenysin synthetase LicA has shown that it specifically activates L-glutamine [20]. The predicted structure for the specificity pocket is very similar to those predicted for the glutamate/aspartate-activating domains discussed above. The only significant difference is the Lys239→Gln substitution (Figure 7b), which could reflect requirement for a neutral functional group to ensure selectivity for glutamine over glutamate. Of course, this observation may not be general because the sequences of only two closely related isoenzymes are currently available.

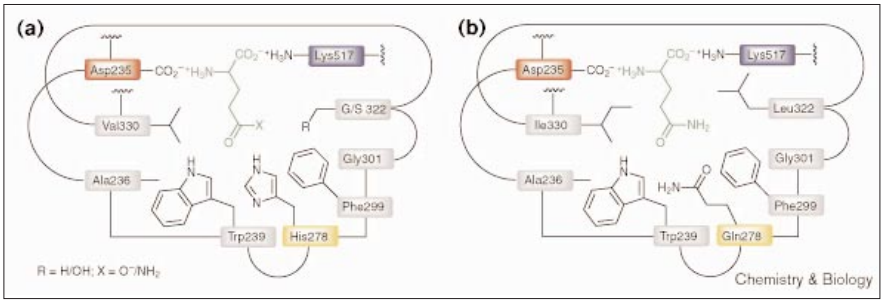
Sequence alignment of the apparent glutamate-activating domains from fengycin synthetases Pps1/FenC, Pps3/FenE and Pps4/FenA predicts a highly conserved group of specificity pocket structures. Indeed, the first two pairs of isoenzyme share identical template sequences and differ from the third pair only in the substitution Ser322→Gly (Table S1). Interestingly, an additional set of isoenzymes (Fen1, Fen3 and Fen4), of undetermined sequence, have been experimentally demonstrated to activate L-glutamate, L-glutamate and L-glutamine, respectively [31]. Further research will be required to clarify the inferred L-glutamine activation by Pps4/FenA. Unexpectedly, when these structures are compared with those predicted for L-serine- and L-HPG-activating domains a close relationship is revealed (see above and Figure 4a). It might be possible, therefore, to interconvert the substrate specificity of these domains by point mutations at Ser301↔Gly and Leu322↔Gly for serine↔glutamate, and Leu322↔Ser for HPG↔glutamine.

The A domain from the second module of the tyrocidine synthetase TycC has been cloned, overexpressed and shown to specifically activate L-glutamine [30]. The structure of the specificity pocket is quite similar to those predicted for the apparent L-glutamate-activating domains from the fengycin synthetases. Interestingly, the principal difference is not Gly322→Ser but His278→Gln, which presumably selects glutamine over glutamate through head-to-head hydrogen bonding between the amide groups of Gln278 and the substrate (Figure 8b).

Ornithine-, N-hydroxyornithine-, lysine- and N-hydroxylysine-activating domains

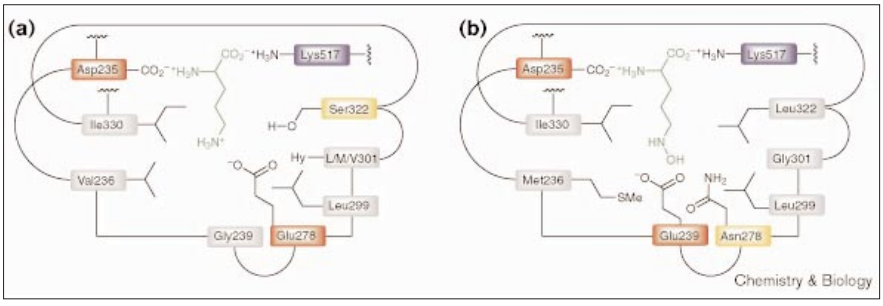
The recombinant activating domain from the third module of the gramicidin synthetase GrsB has been shown to be specific for L-ornithine [32]. Alignment of the sequence from GrsB with the putative L-ornithine-activating domains from fengycin synthetases Pps1 and FenC, tyrocidine synthetase TycC and bacitracin synthetase BacB predicts a highly conserved structure for the specificity pocket. Conservative substitutions occur in the hydrophobic residues 301 and 330 (Figure 9a). Among the remaining universally conserved residues, Glu278 and Ser322 are probably involved in salt-bridge formation and hydrogen bonding with the sidechain amino group of the substrate.

Figure 8



(a) Selectivity pocket structures predicted for the apparent L-glutamate-activating domains from Pps1/FenC-M1 and Pps3/FenE-M1 (R = H; X = O^-), and Pps4/FenA-M2 (R = OH; X = NH_2). **(b)** Selectivity-pocket structure predicted for the L-glutamine-activating domain from TycC-M2.

Figure 9



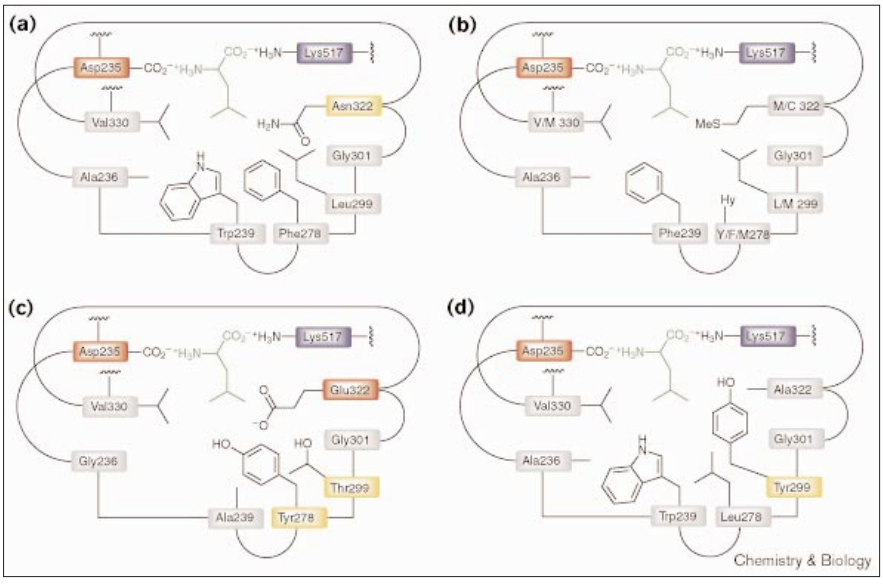
(a) Predicted specificity-pocket structures for the L-ornithine-activating domains from GrsB and the putative ornithine-activating domains from FenC, Pps1, TycC and BacB. **(b)** Predicted specificity-pocket structures for the putative 5-N-hydroxyornithine-activating domains from FxbC.

A comparison of these structures with those predicted for the putative ornithine-activating domains from the exochelin synthetase FxbC reveals substantial differences. In particular, the glutamate residue at position 278 in the former is shifted to position 239 in the latter (Figure 9b). As the identity of the corresponding peptide residues in exochelin itself is not ornithine, but 5-N-hydroxyornithine [13], it is likely that these domains actually activate 5-N-hydroxyornithine. This interpretation differs from that of Stachelhaus *et al.* [17], who assigned this domain as an ornithine-activating domain. Indeed, it has been shown this modified amino acid is incorporated into the related siderophore ferrichrome from *Ustilago sphaerogena* [33]. Furthermore, the loss of ferrichrome synthesis upon disruption of an ornithine 5-N-hydroxylase gene in *Ustilago maydis* indicates amino acid oxidation preceeds polypeptide formation [34].

The predicted structure for the specificity pocket of exochelin synthetase FxbB domain is dissimilar to both the ornithine and proposed 5-N-hydroxyornithine activating domains discussed above. As 5-N-acetyl-5-N-hydroxyornithine is also incorporated into the ferrichrome of *U. sphaerogena*, it seems likely that the domain from FxbB activates 5-N-formyl-5-N-hydroxyornithine — the corresponding amino acid in the exochelin structure from *M. smegmatis* [13].

For the analogous family of substrates, lysine, 6-N-hydroxylysine and 6-N-acyl-6-N-hydroxylysine, a similar pattern emerges for the predicted specificity-pocket structures. Thus, the structures for the putative lysine-activating domain from bacitracin synthetase BacB, the putative 6-N-hydroxy- and 6-N-acyl-6-N-hydroxylysine-activating domains from the mycobactin synthetases MbtF and MbtE, respectively, all show substantial differences (Table S1). It is, however, not possible to derive consensus specificity-pocket structures for these substrates as these are the only domains of known sequence.

Figure 10



(a) Structures predicted for the L-leucine specificity pockets from SrfAA-M3, SrfAB-M3, LicB-M3, LicA-M3, LchAB-M3, LchAA-M3 and BacA-M3. **(b)** Structures predicted for the specificity pockets from CepA-M1, LchAA-M2, LicA-M2, SrfAA-M2 and SrfAC-M1. **(c)** Structures predicted for the specificity pockets from TycC-M6 and GrsB-M4. **(d)** Structures predicted for the specificity pockets from CssA-M2, CssA-M3, CssA-M8 and CssA-M10.

Substrates with hydrophobic sidechains

Leucine-, isoleucine- and valine-activating domains

The surfactin synthetases SrfAA, SrfAB and SrfAC have been partially purified and shown to activate L-leucine. Competition experiments have indicated that the putative leucine-activating domains of these synthetases also activate L-valine and L-isoleucine to a lesser extent [29]. Wild-type gramicidin synthetase GrsB is also known to activate L-leucine [35]. Sequence alignment of these domains with other putative leucine-activating domains from eubacteria — bacitracin synthetase BacA, chloroeremomycin synthetase CepA, lichenysin synthetases LchAA, LchAB, LicA and LicB, and tyrocidine synthetase TycC — predicts three families of highly conserved specificity pocket structures (Figure 10a–c). In accordance with the lipophilicity of the substrate sidechain, the residues lining this pocket are principally hydrophobic. It is, therefore, difficult to assign functional roles for particular residues in these structures. A fourth family of specificity pocket structures for leucine-activating domains is formed by the four from the eukaryotic cyclosporin synthetase CssA (Figure 10d). These structures, which are again composed almost entirely of hydrophobic residues, are all identical.

Experimental characterization of the recombinant fengycin synthetase FenB and purification of the isoenzyme Fen5 have demonstrated that these enzymes both activate L-isoleucine [31,36]. FenB was also shown to activate L-valine with about one tenth of the efficiency. The activating domain from the third module of the lichenysin synthetase LicC has been cloned, overexpressed, purified and shown to activate L-isoleucine [20]. This domain similarly activates L-leucine and L-valine with about one quarter of the efficiency. Sequence alignment of the isoleucine-activating domains from these synthetases with the putative isoleucine-activating domains from bacitracin synthetases BacA, BacC and lichenysin synthetase LchAC predicts a well-conserved family of structures for the specificity pockets. As expected, the residues lining this pocket are largely hydrophobic (Figure 11a).

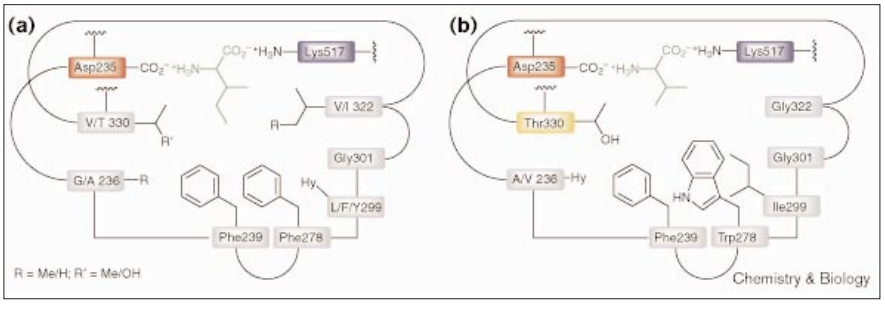
The wild-type surfactin, fengycin and gramicidin synthetases, SrfAB, Fen3 and AcmB, respectively, have been shown to activate L-valine [29,31,35]. In addition, the recombinant A domain from the first module of SrfAB activates valine and isoleucine to a lesser extent [37]. Sequence alignment of these domains with the putative valine-activating domains from fengycin synthetases FenE and Pps3 and lichenysin synthetase LicB and LchAB predicts well-conserved structures, broadly similar to those for the isoleucine-activating domains discussed above. One striking difference is the universally conserved Phe239/Phe278 dyad in the isoleucine-activating domains is replaced with a Phe239/Trp278 dyad in the valine-activating domains (Figure 11b). This is in accord with the smaller steric size of the valine sidechain than the isoleucine sidechain and might be important in biasing the substrate selectivity of these domains. All of these isoleucine- and valine-activating domains are from the *Bacillus* genus.

Like the *Bacillus* L-isoleucine- and L-valine-activating domains discussed above, the remaining valine-activating domains of known sequence cluster in a genus-dependent manner (Table S1). Their predicted specificity-pocket structures do, however, consist mostly of hydrophobic residues as expected. It appears, therefore, that a number of genus-specific solutions have evolved selecting for valine from the cellular amino acid pool, although more sequence data will be required to substantiate this conjecture. One interesting exception to this generalization, however, is the case of AVC synthetases, which cluster across genus lines, as noted above.

Phenylalanine-, tryptophan- and tyrosine-activating domains

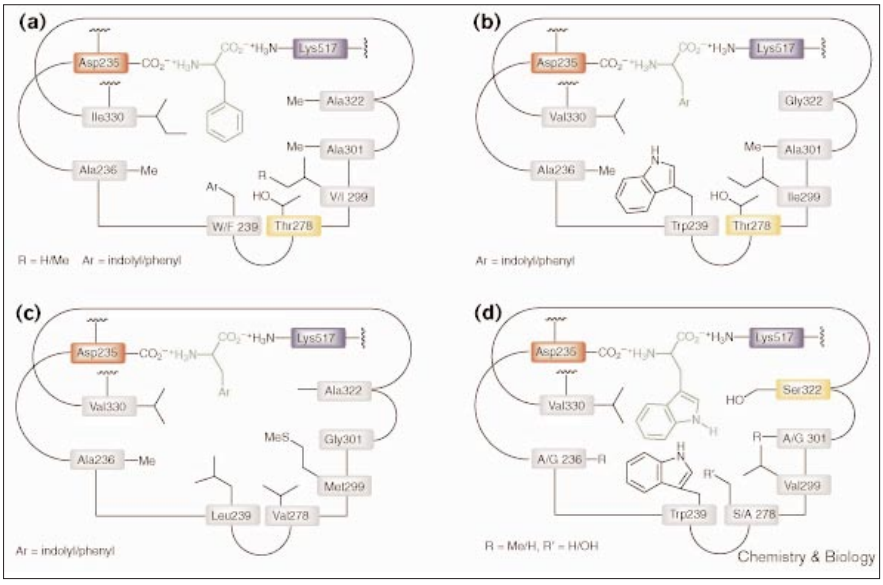
Wild-type GrsA, three truncated fragments of this monomodular synthetase and recombinant tyrocidine synthetase TycA have been shown to activate both L- and D-phenylalanine [38–40]. Sequence alignment of the activation domains from these synthetases with that from the second module of bacitracin synthetase BacC predicts a

Figure 11



(a) Specificity pocket structures predicted for the L-isoleucine-activating domains from fengycin synthetase FenB-M1 and lichenysin synthetase LicC-M3, and for the putative L-isoleucine-activating domains from fengycin synthetase Pps5-M1 and bacitracin synthetases BacA-M1, BacA-M5 and BacC-M1. **(b)** Specificity pocket structures predicted for the L-valine-activating domain from gramicidin synthetase GrsB-M2 and the putative L-valine-activating domains from fengycin synthetases FenE-M2 and Pps3-M2, lichenysin synthetases LicB-M1 and LchAB-M1 and surfactin synthetase SrfAB-M1.

Figure 12



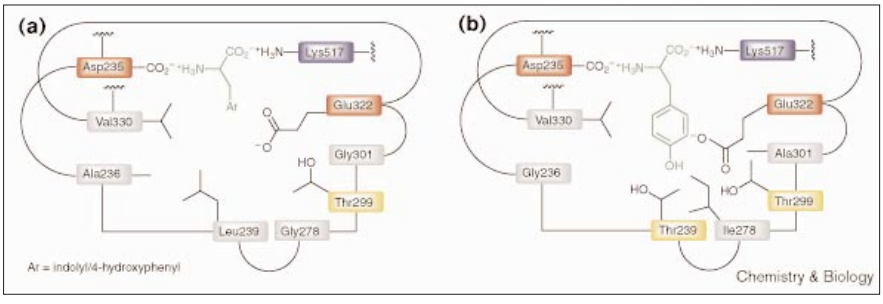
(a) Specificity-pocket structures predicted for the D-/L-phenylalanine-activating domains from gramacidin synthetase GrsA and tyrocidine synthetase TycA and the putative D-/L-phenylalanine-activating domain from bacitracin synthetase BacC-M2; **(b)** specificity pocket structure predicted for the D-/L-phenylalanine- and L-tryptophan-activating domain from tyrocidine synthetase TycB-M3; **(c)** specificity pocket structure predicted for the putative L-phenylalanine-/L-tryptophan-activating domain from tyrocidine synthetase TycB-M2; **(d)** specificity-pocket structures predicted for L-tryptophan-activating domains from the calcium-dependent antibiotic synthetases Cda1-M3 and Cda3-M2.

structure containing only two highly conservative substitutions (Figure 12a). In contrast, cloning, overexpression and purification of the activation domain from the third module of the tyrocidine synthetase TycB has shown it to activate L-tryptophan, in addition to L- and D-phenylalanine [30]. As both this module and TycA contain an epimerization domain, these results are in accord with the isolation of tyrocidines containing only D-phenylalanine at the first position of their peptide sequence, but either D-phenylalanine or D-tryptophan at the fourth position. A comparison of the predicted specificity pocket structures for the A domains of these modules reveals two substitutions: Ala322→Gly and Ile330→Val (Figure 12a,b). An inspection of the GrsA crystal structure (Figures 1,3a) indicates that the latter substitution should not alter the substrate specificity predicted for the TycB domain. In contrast, the former substitution would increase the volume of the pocket enough to allow the larger aromatic sidechain of tryptophan to bind preferentially. Recently, Stachelhaus *et al.* [17] have confirmed this hypothesis by a mutagenesis experiment (Ala322→Gly).

As tyrocidines containing either L-phenylalanine or L-tryptophan at the third position of their peptide sequence are also known [30], TycB-M2 might be expected to contain an activating domain of similar architecture to TycB-M3. For reasons that are unclear, however, a comparison of the structures predicted for the specificity pockets of these two domains shows that they are quite different (Figure 12b,c).

Comparing the structure for the third module of the tyrocidine synthetase TycB with those predicted for the putative tryptophan-activating domains from the second and third modules of the CDA synthetases Cda1 and Cda3, respectively, reveals that they are of similar structure. The principal nonconservative substitution is Gly322→Ser (Figure 12). As variants of the CDA containing L- or D-phenylalanine in place of the L-tryptophan residues are not known it seems that Ser322 plays a pivotal role in selecting tryptophan over phenylalanine. This could be achieved through hydrogen bonding between the tryptophan N-H and Ser322.

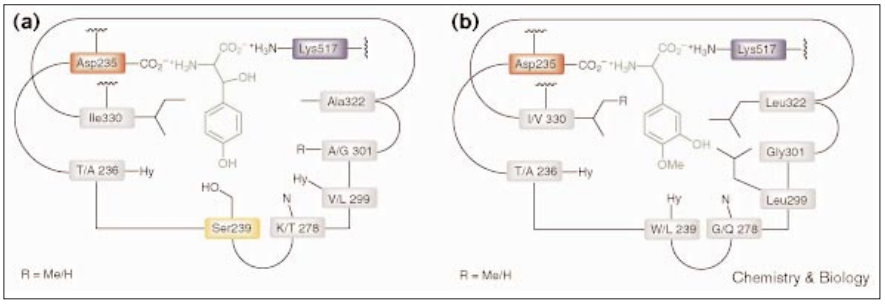
Figure 13



(a) Predicted structure for the specificity pocket of the L-tyrosine/L-tryptophan-activating domain from tyrocidine synthetase TycC-M3. **(b)** Predicted structures for the specificity pockets of the putative L-tyrosine-activating domains from fengycin synthetases FenA-M3/Pps4-M3 and FenD-M1/Pps2-M1.

Figure 14

(a) Predicted structures for the specificity pockets of the putative 3-hydroxy-L-tyrosine-activating domains from chloroeremomycin synthetases CepA-M3 and CepB-M3. **(b)** Predicted specificity pocket structures for the putative 3-hydroxy-4-methoxy-L-phenylalanine-activating domains from the saframycin MX1 synthetases SafA-M1 and SafB-M2.



Cloning and overproduction of the activating domain from the third module of tyrocidine synthetase TycC has shown that it activates either L-tyrosine or L-tryptophan with approximately the same efficiency [30]. This finding is in accord with the structure of tyrocidine D, which contains tryptophan in place of tyrosine at the seventh position. In contrast, the purified wild-type fengycin synthetases Fen2 and Fen4 are reported to activate L-tyrosine, but not L-tryptophan [31]. In line with these observations, no tryptophan-containing analogs of fengycin appear to be known and, indeed, tyrosine at position 3 is lactonized in the product. Comparison of the structures for the specificity pocket for the A domain from Fen2 and Fen4 isoenzymes FenD-M1/Pps2-M1 and FenA-M3/Pps4-M3, respectively, with that from TycC-M3 indicates a possible reason for these differences in substrate selectivity. Glu322 is therefore universally conserved in the five domains and could have a role in substrate selection by hydrogen bonding to the aromatic hydroxyl group of tyrosine. Residue Thr239, however, which is universally conserved in the fengycin structures, is replaced by leucine in the tyrocidine structure (Figure 13). This residue could also be involved in tyrosine selection through hydrogen bonding with the hydroxyl group of this substrate. The replacement of this residue with a hydrophobic one could therefore lead to relaxed substrate specificity and allow tryptophan to bind as well.

A comparison of the predicted specificity-pocket structures for the A domains from chloroeremomycin synthetases CepA and CepB with those for the tyrosine-activating domain, discussed above, reveals several differences. The differences may confer specificity for 3-hydroxy-tyrosine, which has been shown to be incorporated into the closely related glycopeptide vancomycin [41].

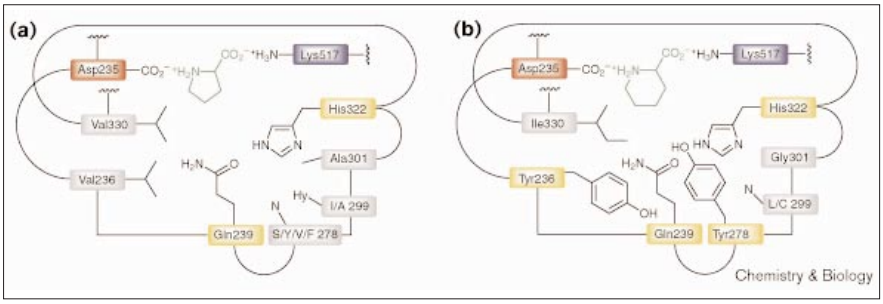
The two aromatic rings in the myxobacterial metabolite saframycin MX1 have been shown to derive from tyrosine by precursor incorporation experiments [42]. Other experiments with blocked mutants, however, have suggested that the substrate for the putative tyrosine-activating domains from saframycin synthetases SafA and SafB is actually 3-hydroxy-4-methoxy-L-phenylalanine [43]. In line with this suggestion, the predicted specificity-pocket structures for these domains are quite different from the other tyrosine-activating domains; they curiously lack polar residues capable of hydrogen bonding with the substrate (Figure 14b).

Proline- and pipecolate-activating domains

The recombinant activating domains from the first modules of the tyrocidine and gramicidin synthetases, TycB and GrsB [30,44], and the wild-type fengycin synthetase Fen1 have been shown to activate L-proline [31]. A comparison of the predicted structure of the specificity pocket from the former domains with those from the putative L-proline-activating domains from the Fen1 isoenzymes Pps4/FenA and the pristinamycin/virginiamycin synthetases SnbDE reveals that they are highly conserved. Given that proline is a hydrophobic amino acid, it is at first sight surprising that a number of polar residues is present in these predicted structures (Figure 15a). Proline, however, has much greater solubility in water than comparable aliphatic amino acids, which might account for the substrate specificity conferred by these binding pockets.

The synthetase responsible for pipecolate incorporation in the biosynthesis of immunomycin has been purified from *Streptomyces hygroscopicus* var. *ascomyceticus*. This enzyme activates L-proline with about one fifth the efficiency, and a number of proline and pipecolate derivatives as well [45]. Consistent with these observations, disrupting of the putative cyclodeaminase responsible for conversion of L-lysine to L-pipecolate in *S. hygroscopicus*, which produces the related immunosuppressant rapamycin, led to the synthesis of prolylrapamycin as the major product [46]. A comparison of the predicted specificity pocket structures for the pipecolate-incorporating enzymes from the rapamycin and FK-506 biosynthetic gene clusters reveals that they are similar to those for the proline-activating enzymes discussed above (Figure 15). The ability for these enzymes to activate and incorporate L-proline is therefore readily explained.

Figure 15



(a) Predicted structures for the specificity pockets of the L-proline-activating domains from tyrocidine synthetase TycB-M1 and gramicidin synthetase GrsB, and the putative L-proline-activating domains from fengycin synthetases FenA/Pps4 and pristinamycin/virginiamycin synthetases SnbDE; **(b)** the putative L-pipecolate-activating domains from rapamycin synthetase RapP and FK-506 synthetase FkbP.

Other amino-acid-activating domains

For the domains that activate amino acids other than those discussed above, either more than one domain has not been sequenced (histidine, 3-methylglutamate, β-alanine, arginine, phenylglycine, 3,5-dihydroxyphenylglycine and 2-butenyl-3-methyl-L-threonine), or the specificity pocket structures deduced from the available information show weak homology (alanine and glycine). For these cases, more sequence data will be required to derive consensus structures for the specificity pockets. As glycine has no sidechain and the sidechain of alanine is very small, it may transpire that there are many degenerate solutions to selectively activating these substrates.

One interesting case worthy of comment, however, is β-alanine seen in the siderophore exochelin [13]. In the predicted specificity pocket, the otherwise universally conserved aspartate residue at position 235 is displaced to position 236. This striking change presumably accommodates the greater separation between the amino and carboxyl termini of this substrate.

Limitations

A clear shortcoming of the available biochemical data is that it is limited mostly to A domains from *E. coli*, *Bacillus* sp. and *Pseudomonas* sp. The important actinomycetes, notably *Streptomyces*, are largely missing from this essential database. Additional potential deficiencies are that the predictive value of the model is obviously dependent on accurate sequence information available in the public domain, the accuracy of which we have come to question in certain instances. Moreover, in the existing literature the limited number of potential substrates (amino acids and stereochemistry) that have been tested for activity in some cases against a particular A domain gives the impression, perhaps falsely, of high substrate specificity. This is especially relevant for recognition of hydrophobic amino acids in which relaxed specificity is frequently observed. Looking ahead, the substrate recognition patterns identified by this model and that of Stachelhaus *et al.* [17] will provide guides for more thoughtful consideration of potential substrates both in overall structure, and absolute stereochemistry.

Although 19 of the 20 common amino acids are represented among the eight critical binding pocket residues, arginine is notably absent from any of the >150 A domains examined. The reasons for this are not understood. Similarly, although some binding pockets lend themselves readily to chemical interpretation of their substrate specificity, particularly those recognizing polar amino acids, others do not; for example, cysteine or the strikingly polar pockets selective for proline and pipecolate. More broadly, a number of structural motifs have evolved to bind nonpolar amino acids, and often less selectively than their polar counterparts. This initial, quite empirical analysis could lead to more rigorous consideration of the many questions of molecular recognition raised by this model of amino acid recognition and activation.

Significance

One application of this model is for the prediction of the order and identity of amino acids activated by a nonribosomal peptide synthetase (NRPS) based on its sequence. As many organisms appear to contain multiple NRPS gene clusters, the ability to do this may greatly aid identification of a cluster of interest using the polymerase chain reaction (PCR) with appropriately designed primers. A second application is to rationally alter subsets of the eight critical residues to change the amino acid recognized and activated by these domains. Although this experiment has not yet been carried out with a full NRPS, encouragement can be taken from existing, albeit limited, data that mutagenesis of recombinant A domains to modify substrate specificities has been achieved [17]. From the perspective of genetic engineering, a useful comparison to the modular polyketide synthases can be made. A large base of genetic experiments exists here, where, in general, the swapping of functional domains has given proteins of significantly reduced synthetic ability [47,48]. More recent experiments suggest that improved efficiency can be achieved by swapping entire modules [49]. Clearly, both of these strategies are available for the modification of NRPSs, and, indeed, have been attempted [37,50–52]. The unknown tertiary and quarternary structures of these

giant synthetases, however, make the effect of gross mutational changes on catalytic efficiency impossible to predict, but experience has shown them to be substantially detrimental. The micro changes that may now be possible to engineer into existing NRPSs on the basis of the new model will not greatly affect their macromolecular structure and may, as a consequence, enable the comparatively efficient synthesis of new peptide products. For the moment, however, the promise of this strategy for creating of structurally altered natural products will benefit from further protein sequence information to refine the model discussed in this article and, ideally, extend the analysis to unusual amino acid substrates of choice.

Materials and methods

Rasmol was used to view and analyze the crystal structure of the phenylalanine activating domain of gramicidin A synthetase, GrsA (1amu.pdb), to determine the binding pocket residues believed to be involved in substrate selection. Multiple protein sequence alignments of candidate adenylation domains were performed with ClustalX [24]. The eight critical amino acid residues lining the binding pockets of each A domain were obtained after aligning each sequence to the GrsA sequence using its secondary structure mask as illustrated in Figure 1. For phylogenetic analysis, peptide sequences were aligned using the ClustalX multiple alignment software described above. The distance matrices were calculated from these alignments using the Kimura protein distance algorithm [53], and trees were generated using the unweighted pair group method arithmetic averages (UPGMA) [54] and the GROWTREE program of the GCG Wisconsin package (version 9.0).

Identification of the predicted amino acid activated by a specific unknown NRPS A module can be performed online in automated fashion at the forthcoming website: http://raynam.chm.jhu.edu/~nrps/.

Supplementary material

Supplementary material including a list of amino acids lining the binding pockets of A domains of assigned (Table S1) and unassigned (Table S2) functions, and phylogenetic analysis of residues between the A3 and A6 core motif (Figure S1) and the eight putative binding-pocket residues (Figure S2) is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

Financial support of this work by the National Institute of Health (AI-14937) and the Wellcome Trust to G.L.C. (053086) is gratefully acknowledged.

References

1. von Döhren, H., Keller, U., Vater, J. & Zocher, R. (1997). Multifunctional peptide synthetases. *Chem. Rev.* **97**, 2675-2705.
2. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651-2673.
3. Konz, D. & Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? *Chem. Biol.* **6**, R39-R48.
4. Laland, S.G., Froyshov, O., Gilhuus-Moe, C. & Zimmer, T.L. (1972). Gramicidin S synthetase, an enzyme with an unusually large number of catalytic functions. *Nat. New. Biol.* **239**, 43-44.
5. Laland, S.G. & Zimmer, T.L. (1973). The protein thiotemplate mechanism of synthesis for the peptide antibiotics produced by *Bacillus brevis*. *Essays Biochem.* **9**, 31-57.
6. Lipmann, F. (1973). Nonribosomal polypeptide synthesis on polyenzyme templates. *Accounts Chem. Res.* **6**, 361-367.
7. Kurahashi, K. (1974). Biosynthesis of small peptides. *Annu. Rev. Biochem.* **43**, 445-459.
8. Stein, T.*, et al.,* & Morris, H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.* **271**, 15428-15435.
9. Townsend, C.A. & Wilson, B.A. (1988). The role of nocardicin G in nocardicin A biosynthesis. *J. Am. Chem. Soc.* **110**, 3320-3321.
10. Wilson, B.A., Bantia, S., Salituro, G.M., Reeve, A.M. & Townsend, C.A. (1988). Cell-free biosynthesis of nocardicin A from nocardicin E and *S*-adenosylmethionine. *J. Am. Chem. Soc.* **110**, 8238-8239.
11. Guenzi, E., Galli, G., Grgurina, I., Gross, D.C. & Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**, 32857-32863.
12. Sharman, G.J., Williams, D.H., Ewing, D.F. & Ratledge, C. (1995). Isolation, purification and structure of exochelin MS, the extracellular siderophore from *Mycobacterium smegmatis*. *Biochem. J.* **305**, 187-196.
13. Yu, S., Fiss, E. & Jacobs, W.R., Jr. (1998). Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. *J. Bacteriol.* **180**, 4676-4685.
14. Shen, B., Du, L., Sanchez, C., Chen, M. & Edwards, D.J. (1999). Bleomycin biosynthesis in *Streptomyces verticillus* ATCC15003: a model of hybrid peptide and polyketide biosynthesis. *Bioorg. Chem.* **27**, 155-171.
15. 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.
16. 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.
17. 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.
18. Turgay, K., Krause, M. & Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**, 529-546.
19. Zhang, J.H., Quigley, N.B. & Gross, D.C. (1995). Analysis of the *syrB* and *syrC* genes of *Pseudomonas syringae* pv. syringae indicates that syringomycin is synthesized by a thiotemplate mechanism. *J. Bacteriol.* **177**, 4009-4020.
20. Konz, D., Doekel, S. & Marahiel, M.A. (1999). Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J. Bacteriol.* **181**, 133-140.
21. de Crecy-Lagard, V.*, et al.,* & Blanc, V. (1997). Streptogramin B biosynthesis in *Streptomyces pristinaespiralis* and *Streptomyces virginiae*: molecular characterization of the last structural peptide synthetase gene. *Antimicrob. Agents Chemother.* **41**, 1904-1909.
22. Cosmina, P.*, et al.,* & van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **8**, 821-831.
23. Husi, H., Schorgendorfer, K., Stempfer, G., Taylor, P. & Walkinshaw, M.D. (1997). Prediction of substrate-specific pockets in cyclosporin synthetase. *FEBS Lett.* **414**, 532-536.
24. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.
25. Reichert, J., Sakaitani, M. & Walsh, C.T. (1992). Characterization of EntF as a serine-activating enzyme. *Protein Sci.* **1**, 549-556.
26. Keller, U. (1987). Actinomycin synthetases. Multifunctional enzymes responsible for the synthesis of the peptide chains of actinomycin. *J. Biol. Chem.* **262**, 5852-5856.
27. Gehring, A.M., Mori, I., Perry, R.D. & Walsh, C.T. (1998). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637-11650.
28. Byford, M.F., Baldwin, J.E., Shiau, C.-Y. & Schofield, C.J. (1997). The mechanism of ACV synthetase. *Chem. Rev.* **97**, 2631-2649.
29. Galli, G.*, et al.,* & Grandi, G. (1994). Characterization of the surfactin synthetase multi-enzyme complex. *Biochim. Biophys. Acta* **1205**, 19-28.
30. Mootz, H.D. & Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843-6850.
31. Steller, S.*, et al.,* & Vater, J. (1999). Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem. Biol.* **6**, 31-41.
32. Krause, M., Marahiel, M.A., von Dohren, H. & Kleinkauf, H. (1985). Molecular cloning of an ornithine-activating fragment of the

gramicidin S synthetase 2 gene from *Bacillus brevis* and its expression in *Escherichia coli*. *J. Bacteriol.* **162**, 1120-1125.
33. Emery, T.F. (1966). Initial steps in the biosynthesis of ferrichrome. Incorporation of δ-N-hydroxyornithine and δ-N-acetyl-δ-N-hydroxyornithine. *Biochemistry* **5**, 3694-3701.
34. Wang, J., Budde, A.D. & Leong, S.A. (1989). Analysis of ferrichrome biosynthesis in the phytopathogenic fungus *Ustilago maydis*: cloning of an ornithine-N5-oxygenase gene. *J. Bacteriol.* **171**, 2811-2818.
35. Vater, J. & Kleinkauf, H. (1975). Substrate specificity of the amino acyl adenylate activation sites of gramicidin S-synthetase (GSS). *Acta. Microbiol. Acad. Sci. Hung.* **22**, 419-425.
36. Lin, G.H.*, et al.,* & Liu, S. T. (1998). Molecular cloning and characterization of fengycin synthetase gene *fenB* from *Bacillus subtilis*. *J. Bacteriol.* **180**, 1338-1341.
37. Elsner, A.*, et al.,* & Bernhard, F. (1997). Substrate specificity of hybrid modules from peptide synthetases. *J. Biol. Chem.* **272**, 4814-4819.
38. Vater, J. & Kleinkauf, H. (1976). Gramicidin S-synthetase. A further characterization of phenylalanine racemase, the light enzyme of gramicidin S-synthetase. *Biochim. Biophys. Acta* **429**, 1062-1072.
39. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270**, 6163-6169.
40. Pfeifer, E., Pavela-Vrancic, M., von Dohren, H. & Kleinkauf, H. (1995). Characterization of tyrocidine synthetase 1 (TY1): requirement of posttranslational modification for peptide biosynthesis. *Biochemistry* **34**, 7450-7459.
41. Zmijewski, M.J. & Fayerman, J.T. (1995). Glycopeptides. In *Genetics and Biochemistry of Antibiotic Production*. (Vining, L.C. & Stuttard, C., eds.), pp 269-281, Butterworth-Heinemann, Boston, MA.
42. Mikami, Y.*, et al.,* & Okuda, S. (1985). Biosynthetic studies on saframycin A, a quinone antitumor antibiotic produced by *Streptomyces lavendulae*. *J. Biol. Chem.* **260**, 344-348.
43. Pospiech, A., Bietenhader, J. & Schupp, T. (1996). Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from *Myxococcus xanthus*. *Microbiology* **142**, 741-746.
44. Hori, K.*, et al.,* & Saito, Y. (1991). The nucleotide sequence for a proline-activating domain of gramicidin S synthetase 2 gene from *Bacillus brevis*. *J. Biochem.* **110**, 111-119.
45. Nielsen, J.B., Hsu, M.J., Byrne, K.M. & Kaplan, L. (1991). Biosynthesis of the immunosuppressant immunomycin: the enzymology of pipecolate incorporation. *Biochemistry* **30**, 5789-5796.
46. Khaw, L.E., Bohm, G.A., Metcalfe, S., Staunton, J. & Leadlay, P.F. (1998). Mutational biosynthesis of novel rapamycins by a strain of *Streptomyces hygroscopicus* NRRL 5491 disrupted in *rapL,* encoding a putative lysine cyclodeaminase. *J. Bacteriol.* **180**, 809-814.
47. McDaniel, R.*, et al.,* & Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel 'unnatural' natural products. *Proc. Natl Acad. Sci. USA* **96**, 1846-1851.
48. Katz, L. (1997). Manipulation of modular polyketide synthases. *Chem. Rev.* **97**, 2557-2575.
49. Gokhale, R.S., Tsuji, S.Y., Cane, D.E. & Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482-485.
50. Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69-72.
51. Schneider, A., Stachelhaus, T. & Marahiel, M.A. (1998). Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Mol. Gen. Genet.* **257**, 308-318.
52. de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C. & Grandi, G. (1997). Engineering of peptide synthetases. Key role of the thioesterase-like domain for efficient production of recombinant peptides. *J. Biol. Chem.* **272**, 25304-25309.
53. Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambrige University Press, Cambridge.
54. Sneath, P.H. & Sokal, R.R. (1973). *Numerical Classification.* Freeman, San Francisco, CA.