



Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products

Fumitaka Kudo¹ · Akimasa Miyanaga¹ · Tadashi Eguchi¹

Received: 9 August 2018 / Accepted: 25 September 2018 / Published online: 5 October 2018
© Society for Industrial Microbiology and Biotechnology 2018

Abstract

Nonproteinogenic amino acids are the unique building blocks of nonribosomal peptides (NRPs) and hybrid nonribosomal peptide–polyketides (NRP–PKs) and contribute to their diversity of chemical structures and biological activities. In the biosynthesis of NRPs and NRP–PKs, adenylation enzymes select and activate an amino acid substrate as an aminoacyl adenylate, which reacts with the thiol of the holo form of the carrier protein to afford an aminoacyl thioester as the electrophile for the condensation reaction. Therefore, the substrate specificity of adenylation enzymes is a key determinant of the structure of NRPs and NRP–PKs. Here, we focus on nonproteinogenic amino acid selective adenylation enzymes, because understanding their unique selection mechanisms will lead to accurate functional predictions and protein engineering toward the rational biosynthesis of designed molecules containing amino acids. Based on recent progress in the structural analysis of adenylation enzymes, we discuss the nonribosomal codes of nonproteinogenic amino acid selective adenylation enzymes.

Keywords Nonproteinogenic amino acid · Nonribosomal code/specificity-conferring code · Structure of adenylation enzyme · Engineering of adenylation enzyme · Natural product biosynthesis

Introduction

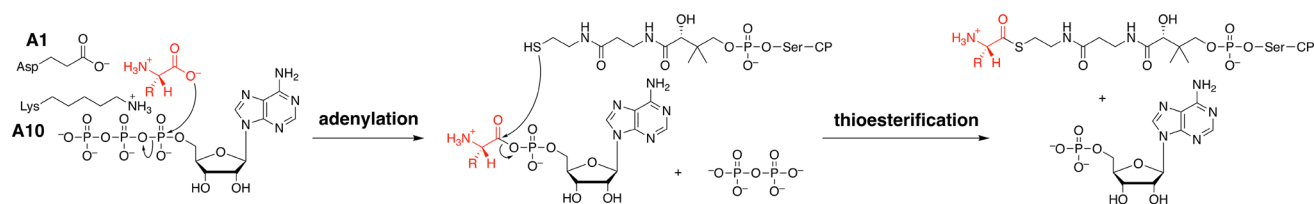
Amino acids are the central building blocks of proteins, peptides, nonribosomal peptides (NRPs), and hybrid nonribosomal peptide–polyketides (NRP–PKs). In the biosynthesis of these biomolecules, the carboxy group of an amino acid is generally activated as an aminoacyl adenylate or aminoacyl phosphate with adenosine triphosphate (ATP). The mixed acid anhydrides generated between the carboxylic acids and phosphoric acids are excellent electrophiles, and thus efficiently react with nucleophiles, such as thiols, to form covalent linkages. In the biosynthesis of NRPs and hybrid NRP–PKs, an adenylation enzyme, which is generally composed of 500–600 amino acid residues, activates the carboxy group of an amino acid as an aminoacyl adenylate and ligates the aminoacyl

group with the thiol group of the phosphopantetheine arm in the holo form of the carrier protein (CP), which is generally composed of 90–100 amino acid residues, to give an aminoacyl thioester (Scheme 1). Thus, the adenylation enzyme catalyzes the two nucleophilic substitutions for adenylate and thioester formation by releasing inorganic diphosphate (PPi) and adenosine monophosphate (AMP), respectively. The generated aminoacyl-CP thioester functions as an electrophile for amide formation in the nonribosomal peptide synthetase (NRPS) system (Scheme 2a) and for transesterification with the cysteine thiol group in the β -ketosynthase domain leading to subsequent C–C bond formation in the hybrid NRPS–polyketide synthase (PKS) system (Scheme 2b). In the NRPS system, an aminoacyl-CP thioester undergoes nucleophilic attack by the amino group of another aminoacyl intermediate in the condensation enzyme to form an amide bond to give a dipeptidyl-CP thioester in a stepwise manner. It should be noted that the reaction order is strictly regulated by these two types of enzymes to biosynthesize the programmed structures of natural products in the genome of producer strains. In general, the adenylation enzymes accept the holo form of a CP thiol, but do not accept the aminoacyl-CP thioester, to prevent undesired amide bond formation (Scheme 2c).

This article is part of the Special Issue “Natural Product Discovery and Development in the Genomic Era 2019”.

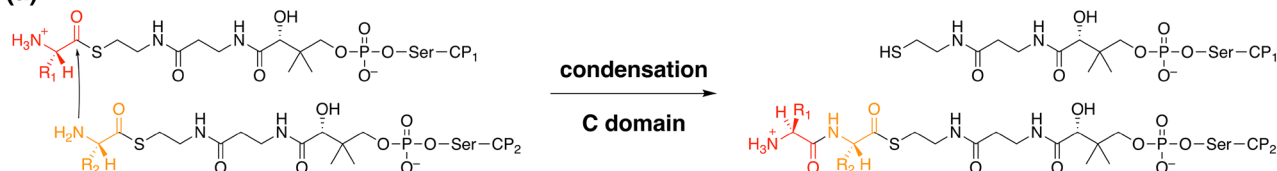
✉ Fumitaka Kudo
fkudo@chem.titech.ac.jp

¹ Department of Chemistry, Tokyo Institute of Technology, 2-12-1 Meguro-ku, O-okayama, Tokyo 152-8551, Japan

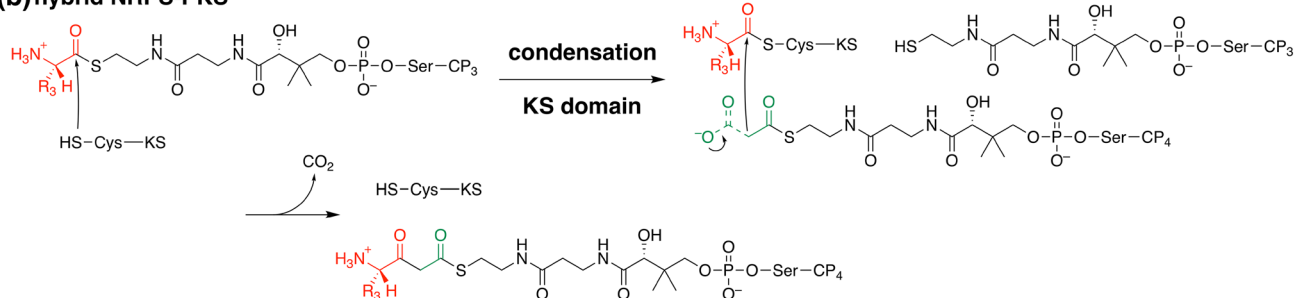


Scheme 1 Two-step nucleophilic substitution in the adenylation enzymatic reaction

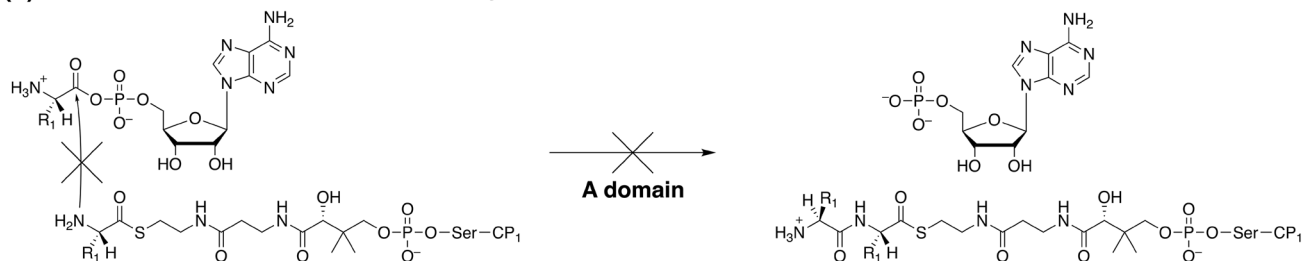
(a) NRPS



(b) hybrid NRPS-PKS



(c) Condensation does not occur in the adenylation domain.



Scheme 2 Condensation reaction in the condensation domain of NRPS and ketosynthase domain in hybrid NRPS-PKS

The amino acid substrate specificity of the adenylation enzymes makes a large contribution in determining the chemical structure of natural products. The information of the substrate specificity of the adenylation enzymes is thus important for rational genetic engineering of NRPS assembly lines to generate novel peptides for drug discovery [5, 6, 12]. Structural information of adenylation enzymes also helps rational engineering of adenylation domains of NRPSs to introduce unnatural amino acids into the peptide structures [80]. Therefore, the mechanism of substrate recognition and the conformational dynamics of adenylation enzymes have been extensively studied [36]. The recognition

of an amino acid substrate and the nucleophilic substitution with ATP to afford the aminoacyl adenylate and PPI occur at the active site of the adenylation enzyme. Thus, the positions of ATP and the carboxy group of the substrate amino acid must be precisely fixed to make the appropriate Bürgi–Dunitz angle for nucleophilic attack of the carboxylic acid to the α -phosphorous atom of ATP. To form the required in-line situation, in addition to the carboxy group, the α -amino group of an L- α -amino acid and the side-chain at the α position must be properly accommodated at the active site of the adenylation enzyme. The structural analysis of adenyating enzymes and detailed primary sequence

comparisons using the Hidden Markov Model have provided important information to allow the prediction of the substrate specificity of functionally uncharacterized adenylation enzymes [15, 50, 95]. A bioinformatic prediction tool for the amino acid substrate is available with high accuracy for the bacterial adenylation enzymes selective for the 20 proteinogenic L- α -amino acids [8, 50], although the tool does not truly apply to fungal adenylation enzymes because of significant sequence differences. The tool was established based on the structural analysis of the adenylation domain PheA in GrsA, which activates L- α -phenylalanine in gramicidin biosynthesis (Fig. 1) [19]. The crucial ten amino acid residues (A1-position—A10-position), the so-called nonribosomal code [108] or specificity-conferring code [95], are believed to be involved in the substrate recognition, based on information from the PheA structure, and thus are used to predict the amino acid substrates for adenylation enzymes. Many excellent reviews concerning the nonribosomal code for proteinogenic L- α -amino acids with structural information on the adenylation enzymes have been published [14, 62, 96, 98]. However, information regarding adenylation enzymes that recognize nonproteinogenic amino acids is limited, even though nonproteinogenic amino acids make a substantial contribution to the structural diversity of NRPs and hybrid NRP-PKs, resulting in the important biological activities of these compounds [109] (Fig. 2). For example, the well-known immunosuppressant FK506 (tacrolimus) contains pipercolic acid; the lipopeptide antibiotic daptomycin contains L-ornithine, (2*S*,3*R*)-3-methylglutamic acid, and kynurenine; the antibiotic rifamycin belonging to the ansamycin family contains 3-amino-5-hydroxybenzoic acid as the starter unit of the polyketide skeleton; and the anti-cancer drug bleomycin contains β -Ala in its unique hybrid NRP-PK structure. Although some of the proteinogenic

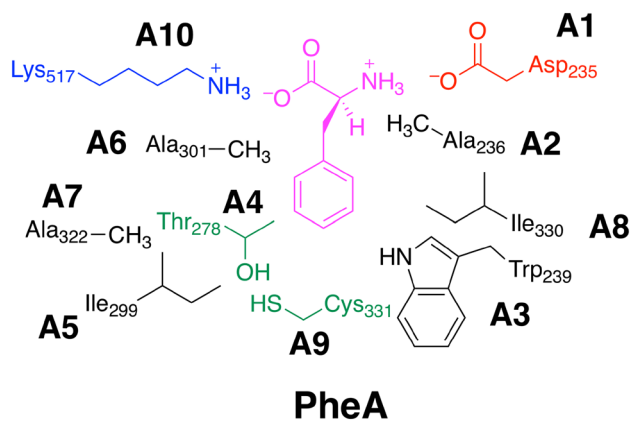


Fig. 1 Active site of L-Phe selective adenylation enzyme PheA. Magenta color represents substrate amino acid. Red color represents acidic residues. Blue color represents basic residues. Green color represents polar residues

amino acids can be modified by enzymes, including epimerases, *N*-methyltransferases, *O*-methyltransferases, and heterocyclization and oxidation domains, in the NRPS assembly line, many nonproteinogenic amino acids that are found in NRPs and hybrid NRP-PKs are independently biosynthesized and incorporated into the main NRPS and hybrid NRPS-PKS assembly lines (see the daptomycin structure in Fig. 2) [98, 109]. Thus, the selective activation and thioesterification of nonproteinogenic amino acids by adenylation enzymes is naturally occurring. Structural analysis of NRPSs and PKSs has provided valuable information to better understand the substrate recognition mechanism in the megasynthase [14, 62, 96, 98]. In this review, we describe the structural basis of nonproteinogenic amino acid selective adenylation enzymes in NRPSs and hybrid NRPS-PKs and discuss the nonribosomal code for nonproteinogenic amino acids. Although the associated proteins, such as condensation enzymes and CPs, affect the enzymatic activity by interacting with the C-terminal A_{sub} domain of adenylation enzymes [58, 68], this review will focus on the active site of the adenylation enzymes, the N-terminal A_{core} domain, which must accommodate the correct substrate amino acid.

Structural analysis of amino acid adenylation enzymes from 1997 to 2018

Accumulation of information on the genome sequence and the functional characterization of adenylation enzymes have accelerated accurate bioinformatics predictions. However, structural information regarding nonproteinogenic amino acid selective adenylation enzymes is still limited. The amino acid adenylation enzymes that have been structurally characterized to date are summarized in Table 1. In this review, we focus on amino acid adenylation enzymes and do not include enzymes that recognize hydroxybenzoic acid derivatives, such as salicylic acid and 2,3-dihydroxybenzoic acid. The active sites of the representative adenylation enzymes PheA [19] and VinN [73], the latter of which incorporates (2*S*,3*S*)-3-methyl-L-aspartic acid (MeAsp), a β -amino acid, in vicenistatin biosynthesis, are shown in Fig. 3.

Common features of amino acid adenylation enzymes

The Asp235 (A1-position) and Lys517 (A10-position) residues of PheA (according to the amino acid residue of PheA, the nonribosomal code numbering will be used throughout) are highly conserved in amino acid adenylation enzymes and recognize the substrate amino acids, including D- α -amino acids and β -amino acids (Table 2). The C-terminal Lys517

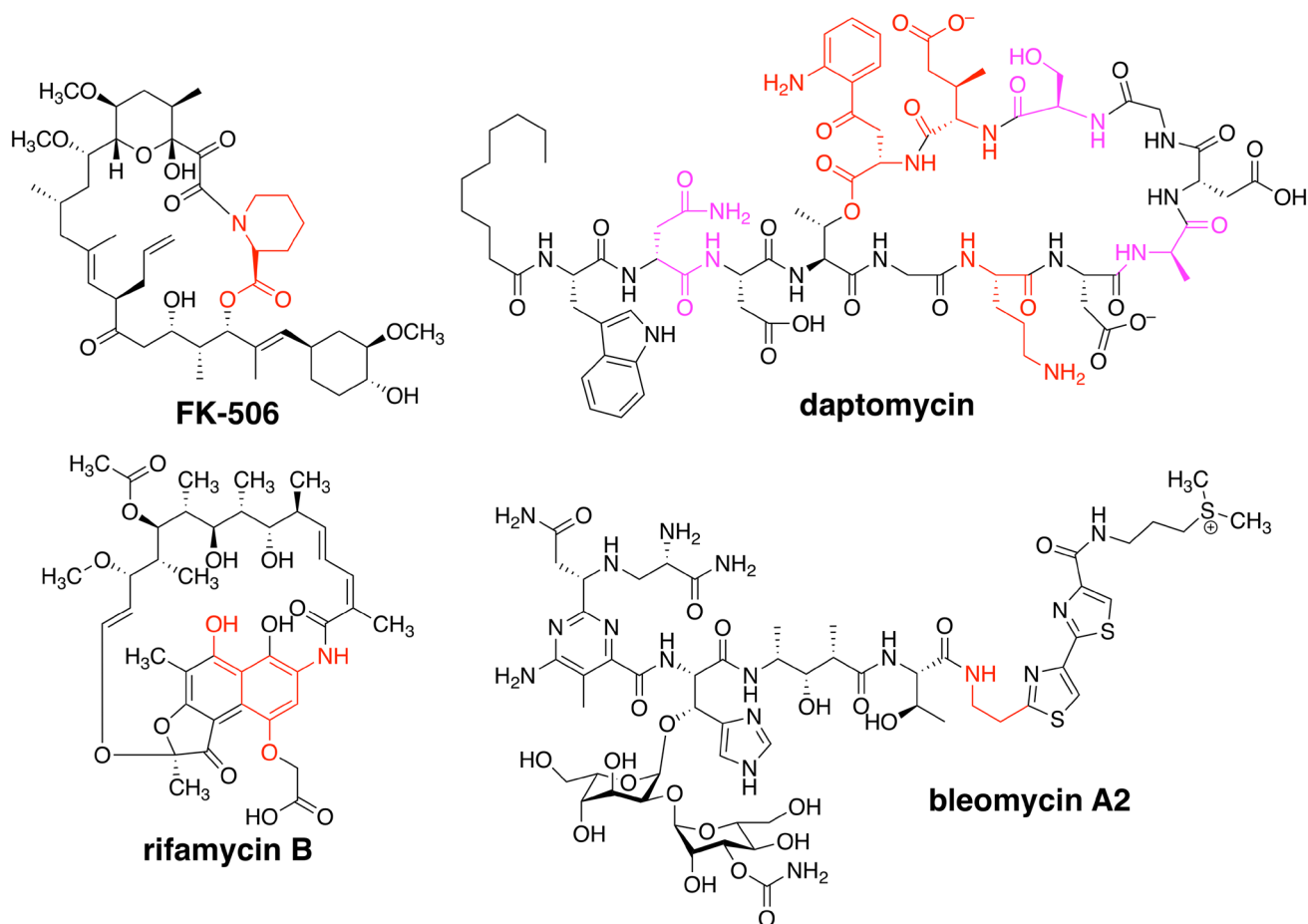


Fig. 2 Representatives of nonproteinogenic amino acid containing natural products. Color indicates nonproteinogenic amino acids. Red color nonproteinogenic amino acids are recognized by adenylation

enzymes. Magenta color nonproteinogenic amino acids are D-amino acids, which are constructed in the NRPS assembly line

(A10-position) in the Asub domain, which is involved in the recognition of the carboxy group of an amino acid and the α -phosphate group of ATP, which react to afford the acyl adenylate intermediate, is completely conserved in the ANL (Acyl-CoA synthetase, NRPS adenylation domain, and Luciferase) family of enzymes. The Lys residue at the A10-position activates the α -phosphate of ATP and prompts nucleophilic substitution with the carboxy group by displacing the inorganic diphosphate. Along with the release of the inorganic diphosphate, the Asub domain is rotated 140° and the Lys517 (A10-position) residue is thus oriented away from the active site where the acyl adenylate intermediate waits for the nucleophile CP [86]. The acceptor CP accesses the thioesterification form of the adenylation enzyme and then the adenylation enzyme catalyzes the second nucleophilic substitution of the aminoacyl adenylate with the thiol of the holo CP to afford the aminoacyl thioester by releasing AMP. In the conformational movement of the C-terminal Asub domain, an aromatic amino acid residue (Trp, Phe, or His in usual NRPS adenylation enzymes) that is placed

before the highly conserved Asp235 (A1-position) is critically involved.

The highly conserved Asp235 (A1-position) in amino acid adenylation enzymes is involved in the recognition of the amino group of the substrate amino acids (Table 2). The adenylation enzyme IdnL1 that is selective for the β -amino acid (*S*)-3-aminobutyric acid in incednine biosynthesis has an Asp residue at the A1-position that appears to interact with the β -amino group [17]. The D-glutamic acid selective adenylation enzyme McyE in microcystin biosynthesis has an Asp residue at the A1-position that appears to interact with the amino group of D-glutamic acid despite the fact that the γ -carboxylic acid is activated [69]. The L- α -amino adipic acid selective adenylation enzyme in α -amino adipate reductase also has an Asp residue at the A1-position that appears to interact with the amino group despite the δ -carboxylic acid being activated [47]. In contrast, aryl acid selective adenylation enzymes do not have an acidic residue at the A1-position and are able to accommodate the aromatic ring of aryl acids [24, 65, 83, 98]. The acidic amino acid residue

Table 1 Structural analysis of amino acid adenylation enzymes

Amino acid substrate	Protein	Final natural products	Ligands, note	PDB ID ^b	Year	References
Proteinogenic L-α-amino acids						
L- α -Phe	PheA in GrsA	Gramicidin	L- α -phenylalanine and AMP	1AMU	1997	[19]
L- α -Phe	McyG	Microcystin	L-Phe-AMP, A-CP didomain	4R0M	2015	[99]
L- α -Arg/Tyr	ApnA-A1	Anabenopectin	Arg/Tyr adenylyate	4D57/56	2015	[48]
L- α -Leu	SrfA-C	Surfactin	None, C-A-T-TE full module	2VSQ	2008	[101]
L- α -Val	PA1211	Unknown	Valine-adenosine vinyl-sulfonamide, A-CP	4DG8, 4DG9	2012	[72]
L- α -Val	LgrA	Linear gramicidin	Val-NH-phospho-pantetheine attached to the PCP domain, F-A-CP initiation module	5ES8	2016	[87]
Gly	AlmE	LPS modification	Gly-AMP	4OXI	2014	[38]
Gly	AB3404	Unknown	Gly and AMP, C-A-T-TE full module	4ZXI	2016	[23]
L-Ser	EntF	Enterobactin	Ser-AVS, C-A-T-TE full module and with MbtH	5T3D	2016	[23][71]
L- α -Thr	Thr1	Chloro Thr	Thr and ATP, Thr-AMP	5N9 W, 5N9X	2017	[90]
Nonproteinogenic amino acids						
D-α-amino acids						
D-Ala	DltA	D-alanylation of lipoteichoic acid	D-Ala-AMP	3DHV	2008	[25][26][114]
Modified α-amino acids:						
(2 <i>S</i> ,3 <i>S</i>)- β -methyl-L-aspartic acid	SlgN1	Streptolydigin	AMP-cpp, as α -amino acids	4GR5	2013	[39]
<i>N</i> -cis-anhydromevalonyl- <i>N</i> -hydroxy-L-ornithine	SidNA3	Fungal siderophore	Fungal adenylation enzyme	3ITE	2010	[56]
Norcoronamic acid (and L-Val) ^a	ThioS(A _{4a} M ₄ A _{4b})	Thiocoraline	MT domain insertion type	5WMM	2018	[77]
2-aminobutyric acid	CytC1	Cytotrienin	Only pdb data. L-valine is the best substrate	3VNR, 3VNQ, 3VNS	2007	[105]
β-amino acids:						
(2 <i>S</i> ,3 <i>S</i>)- β -methyl-L-aspartic acid	VinN	Vicenistatin	β -methyl-L-aspartic acid	3WV5	2014	[73]
(3 <i>S</i>)-3-aminobutyric acid	IdnL1	Incednine	3-ABA-adenylate	5JJQ	2017	[17]
3-aminononanoic acid	CmiS6	Cremimycin	None	5JJP	2017	[17]
(<i>S</i>)- β -phenylalanine	engineered TycA	–	β -Phe-AMS (sulfamoyl adenosine)	5N82	2018	[80]
Others						
Anthranilic acid	AuaEII	Aurachin	Anthraniloyl-AMP, CoA ligase	4WV3	2016	[43]

^aNatural substrate would be different from norcoronamic acid based on the structure of thiocoraline

^bProtein Data Bank (PDB) ID are listed for each deposited structure

that is the first amino acid residue of the nonribosomal code (A1-position) appears to interact with the amino group, even of β -amino, γ -amino, and δ -amino acids. The carboxy and

the amino groups of the substrate amino acids appear to form two salt bridge interactions with the enzyme Lys and Asp (occasionally Glu in the 2-amino adipic acid selective

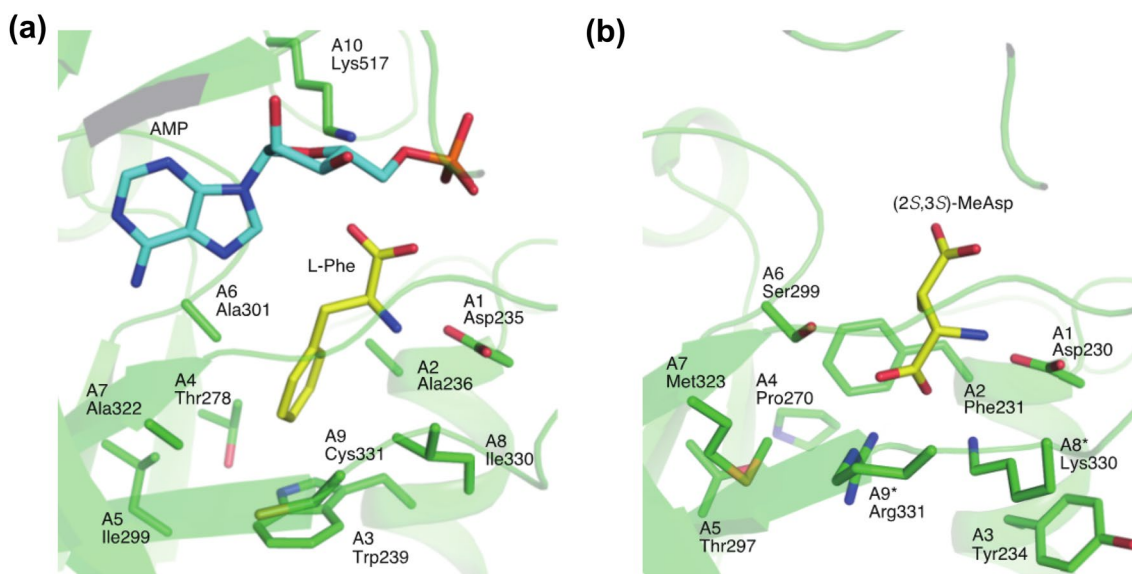


Fig. 3 The active sites of the crystal structures of PheA **a** [19] and VinN **b** [73]. **a** PheA contains L-Phe (yellow stick) and AMP (blue stick) at the active site. Nonribosomal-code amino acids are shown

as green sticks. **b** VinN contains (2*S*,3*S*)-MeAsp (yellow stick) at the active site. Nonribosomal code amino acids are shown as green sticks. The A10 Lys residue is not shown

enzyme) residues [50], respectively. The long main chains of nonproteinogenic amino acids (in the case of β -amino, γ -amino, and δ -amino acids) seem to be somehow folded, and the corresponding adenylation enzymes form an appropriate space to accommodate the long main chains and the unique side-chains. The other eight amino acid residues of the nonribosomal code (A2-position—A9-position) are located at the active site and seem to be involved in the recognition of the side-chains of the amino acids. In this review, we focus on the eight amino acid residues of the nonribosomal code (A2-position—A9-position), which are believed to be key determinants for particular amino acid substrates, including nonproteinogenic amino acids. Based on the structures of some nonproteinogenic amino acid selective adenylation enzymes, we discuss the current knowledge of the substrate recognition mechanisms of nonproteinogenic amino acids.

D- α -Amino acid selective adenylation enzyme, DltA (a structurally elucidated enzyme)

DltA selectively recognizes D-Ala over L-Ala in the D-allylation of lipoteichoic acid in the cell wall of Gram-positive bacteria (Fig. 4a) [25, 26, 114]. DltA has a Cys residue (Cys269) at the A6-position, which can only accommodate the small hydrogen atom and would clash with the methyl group of L-Ala. The C269A mutant showed a similar affinity for both D-Ala and L-Ala, suggesting that the role of this Cys

residue is as a gatekeeper to accommodate D-Ala only [25, 114]. In L- α -amino acid selective adenylation enzymes, the amino acid residue at the A6-position is often Gly or Ala, which are likely small enough to accommodate the side-chain of L- α -amino acids (Table 2). The small Ala residue at the A9-position of DltA seems to indirectly expand the active site of DltA, with the Val residue at the A8-position and the Leu residue at the A2-position, to accommodate only the methyl group of D-Ala. This nonribosomal code is conserved in LnmQ, which selectively recognizes D-Ala in leinamycin biosynthesis [100].

Modified α -amino acid selective adenylation enzymes (structurally elucidated enzymes)

SlgN1 is an adenylation enzyme that has a Mbth-like domain at the N-terminus and recognizes (2*S*,3*S*)-MeAsp, as an α -amino acid, in streptolydigin biosynthesis (Fig. 4b) [39]. SlgN1 has a Gly residue at the A6-position suggesting that it is able to accommodate the side-chain of an L- α -amino acid. The nonribosomal code of SlgN1 is different to that of L-Asp selective enzymes [50]. SlgN1 has a Gln residue at the A4-position, which seems to interact with the β -carboxy group. In contrast, L-Asp selective adenylation enzymes have a Lys residue at the A4-position that interacts with the β -carboxy group of L-Asp. The weak interaction between the Gln residue at the A4-position of SlgN1 and the β -carboxy group of the substrate amino acid might provide the flexibility to accommodate the β -methyl group of (2*S*,3*S*)-MeAsp.

Table 2 Nonribosomal codes of adenylation enzymes in figures in this manuscript

Adenylation enzyme	Substrate amino acid	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Figure	Natural product	Refs.
PheA	L-Phe	D	A	W	T	I	A	A	I	C	K	1	Gramicidin	[19]
DltA	D-Ala	D	L	M	T	L	C	T	V	A	K	4a	D-alanylation of lipoteichoic acid	[25, 26, 114]
SlgN1	(2 <i>S</i> ,3 <i>S</i>)-MeAsp	D	A	L	Q	I	G	G	G	F	K	4b	Streptolydigin	[39]
SidNA3	<i>cis</i> -AMHO ^a	D	V	G	G	G	G	V	I	G	K	4c	Fungal siderophore	[56]
CmaA	L- <i>allo</i> -isoleucine	D	A	F	L	Y	G	G	V	F	K	5a	Coronatine	[20]
Alb04A2	L-cyanoalanine	D	L	T	K	I	G	E	V	G	K	5b	Albicidin	[18]
SyID	3,4-dehydro-L-Lys	D	V	S	E	R	L	V	C	E	K	5c	Syringolin	[112]
SuII A1	D-Glu	D	A	M	H	V	G	G	T	A	K	5d	Sulfazecin	[58]
SuIM A3	L-2,3-diaminopropionic acid	D	V	W	E	M	G	N	D	D	K	5e	Sulfazecin	[58]
FkbP	L-pipecolic acid	D	Y	Q	Y	C	G	H	L	I	K	6a	FK520	[35]
GetE	L-pipecolic acid	D	V	Q	D	I	A	H	M	V	K	6b	GE81112	[9]
GriA2	(2 <i>S</i> ,4 <i>R</i>)-4-methyl-L-Pro	D	V	Q	F	A	G	H	A	V	K	6c	Griselimycins	[59]
LmbC	(2 <i>S</i> ,4 <i>R</i>)-4-propyl-L-Pro	D	V	A	L	V	A	I	G	C	K	6d	Lincomycin	[45]
CcbC	L-proline	D	V	F	Y	C	A	L	V	C	K	6e	Celesticetin	[45]
SnbDE-A2-domain	L-4-dimethylaminophenylalanine	D	A	W	A	I	A	A	V	C	K	7a	Pristinamycin I _A	[64, 103]
MbtF	N ⁶ -hydroxy-L-Lys	D	A	W	D	A	G	C	V	E	K	7b	Mycobactin	[66]
MbtE	N ⁶ -decanoyl-N ⁶ -hydroxy-L-Lys	D	V	A	H	P	G	F	I	N	K	7c	Mycobactin	[66]
VioG	capreomycidine	D	P	Q	D	V	G	I	G	N	K	7d	Viomycin	[30, 31]
HrmO1 _A	L-3-(<i>trans</i> -2-nitrocyclopropyl)Ala	D	P	I	V	V	G	G	V	A	K	7e	Hormaomycin	[21, 40]
HrmO3 _A	L-β-methyl-Phe	D	A	W	T	T	A	V	A	A	K	7f	Hormaomycin	[21, 40]
FmoA1	6-chloro-L-Trp	D	G	W	C	V	A	V	I	A	K	8a	JBIR-34	[78]
FmoA3	α-methyl-L-Ser	D	A	Y	N	V	S	S	I	C	K	8b	JBIR-34	[78]
FmoA4	D-Ala	D	G	I	T	L	G	L	V	Y	K	8c	JBIR-34	[78]
TioJ	3-hydroxyquinidic acid	N	S	C	T	Q	G	W	L	T	K	8d	Thiocoraline	[76]
EcmI	2-quinoxalinecarboxylic acid	Q	S	C	T	Q	G	W	L	T	K	8e	Echinomycin	[76]
SpsQ	L-methylthiomethylcysteine	D	V	D	Y	R	G	E	A	G	K	8f	Sparsomycin	[89]
VinN	(2 <i>S</i> ,3 <i>S</i>)-MeAsp	D	F	Y	P	T	S	M	K	R	K	9a	Vicenistatin	[73]
CmiS6	3-aminononanoic acid	D	F	G	H	L	S	L	M	A	K	9b	Cremimycin	[17]
IdnL1	(<i>S</i>)-3-aminobutyric acid	D	F	L	N	L	S	C	V	A	K	9c	Incednine	[17]
HitB	(<i>S</i>)-β-Phe	D	F	F	H	L	T	G	M	A	K	9d	Hitachimycin	[53, 75]
McyB-A2	(2 <i>R</i> ,3 <i>S</i>)-MeAs	D	A	R	H	V	G	I	F	V	K	10a	Microcystin	[69]
Orf5 (StnE)	β-Lys	D	T	E	V	V	G	T	A	L	K	10b	Streptothricin	[57, 63]
BlmIV A1	β-Ala	V	D	W	V	I	S	L	A	D	K	10c	Bleomycin	[34]
AdmJ	(<i>S</i>)-β-Phe	D	M	L	S	L	G	A	M	C	K	10d	Andrimid	[61]
SgcC1	(<i>S</i>)-β-Tyr	D	P	A	Q	L	M	L	I	A	K	10e	C-1027	[106]
CmdD-A7	(<i>R</i>)-β-Tyr	D	G	S	T	I	T	A	T	V	K	10f	Chondramides	[85]

Table 2 (continued)

Adenylation enzyme	Substrate amino acid	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Figure	Natural product	Refs.
McyE	D-Glu	D	P	R	H	S	G	V	V	G	K	11a	Microcystin	[69]
Nps3	L-2-aminoadipic acid	D	P	R	H	F	V	M	R	A	K	11b	Lysine	[46, 47]
RifA	3-amino-5-hydroxybenzoic acid	S	L	I	A	L	G	A	G	A	K	12a	Rifamycin	[1]
Alb01A3	<i>p</i> -aminobenzoic acid	A	V	K	Y	V	A	N	D	A	K	12b	Albicidin	[18]
AnaPS	anthranilic acid	G	A	L	F	F	A	A	G	V	K	12c	Acetylazonalenin	[3]
AtaEII	anthranilic acid	T	F	G	Y	V	A	G	H	I	K	12d	Aturachin	[43, 81]

^a*cis*-AMHO, *N*^δ-*cis*-anhydromevalonyl-*N*^δ-hydroxy-L-ornithine

Because there is no biochemical data available, it is difficult to speculate further on the recognition mechanism of SlgN1.

SidNA3 (the third adenylation domain of NRPS SidN) recognizes the large *N*^δ-*cis*-anhydromevalonyl-*N*^δ-hydroxy-L-ornithine (*cis*-AMHO) residue in fungal siderophore biosynthesis in *Neotyphodium lolii* Lp19 (Fig. 4c) [56]. Docking analysis of the SidNA3 structure with *cis*-AMHO suggests that the main-chain carbonyl oxygen of Ile (A8-position) is involved in the recognition of the *N*^δ-hydroxy group. SidNA3 has a Gly residue at the A6-position suggesting that it is able to accommodate the long side-chain of an L- α -amino acid in a similar manner to a typical L- α -amino acid adenylation enzyme. Because of the large size of the amino acid, the additional amino acid residues that are involved in the recognition of the side chain of *cis*-AMHO are proposed to be residues lining the binding pocket. Accumulation of structural information on fungal adenylation enzymes with ligand and biochemical characterizations will enhance our knowledge of the substrate recognition mechanisms of non-proteinogenic amino acid adenylation enzymes selective for large substrates.

Modified α -amino acid selective adenylation enzymes (functionally characterized, but with no protein structures)

Nowadays, the genome sequence analysis of microorganisms is easily accessible, and thus, the target biosynthetic gene clusters of natural products can be rapidly identified based on the chemical structures [10, 22, 67, 93, 94]. The target nonproteinogenic amino acid selective adenylation domains can be predicted based on the domain structures of NRPS and hybrid NRPS–PKS enzymes without functional analysis. Based on bioinformatics analysis, many adenylation enzymes have been characterized in vitro that are selective for nonproteinogenic amino acids in NRPs and hybrid NRP–PKs. In most reports of the functional analysis of adenylation domains, the nonribosomal codes for L- α -amino acid selective enzymes are used to discuss the substrate amino acid recognition mechanisms. The similarity in nonribosomal codes among structurally similar amino acid selective enzymes suggests there may be a plausible substrate amino acid recognition mechanism to allow prediction of the substrates. However, in most cases, the precise role of the nonribosomal codes is unclear, and thus, structural analysis of the enzymes is required to understand the recognition mechanism for accurate amino acid substrate prediction.

CmaA is a standalone adenylation enzyme that recognizes L-*allo*-isoleucine in coronatine biosynthesis [20]. Kinetic analysis of CmaA with several substrates showed that the k_{cat} and K_m values varied depending on the type of the fused FLAG-Tag or His-Tag. L-Ile selective adenylation

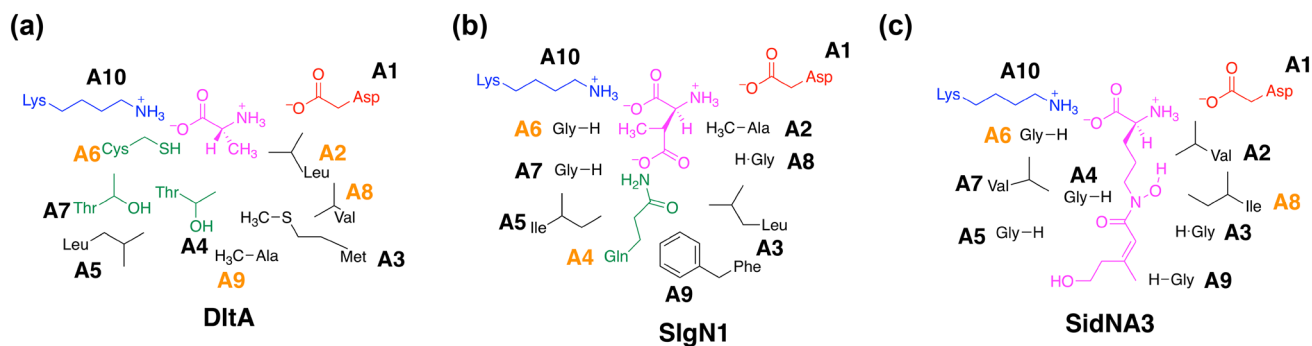


Fig. 4 Active sites of structurally elucidated enzymes. **a** Active site of *D*-Ala selective adenylation enzyme DltA. **b** Active site of MeAsp selective adenylation enzyme SlgN1. **c** Active site of *cis*-AMHO

selective adenylation enzyme SidNA3. Orange color codes are possibly involved in the recognition of the side chain of substrate amino acid

enzymes usually have a Phe residue at the A4-position and a Leu residue at the A5-position, while CmaA has a Leu residue at the A4-position and a Tyr residue at the A5-position (Fig. 5a). These residues may determine the substrate specificity. The adenylation domain Alb04A2* recognizes *L*-cyanoalanine in albicidin biosynthesis [18]. Alb04A2* has an *L*-Asn type nonribosomal code and also recognizes *L*-Asn (Fig. 5b). It is unclear how Alb04A2* distinguishes between *L*-cyanoalanine and *L*-Asn in the producer strain. The adenylation domain of the hybrid NRPS–PKS SyID recognizes

3,4-dehydro-*L*-lysine in syringolin biosynthesis [112]. *L*-Lys is also activated with approximately 50% efficiency. The nonribosomal code of the SyID adenylation domain is different from that of any proteinogenic *L*- α -amino acid adenylation enzyme. The two Glu residues at the A4-position and at the A9-position might be involved in the recognition of the ϵ -amino group of 3,4-dehydro-*L*-lysine (Fig. 5c). In sulfazecin biosynthesis, *D*-Glu and *L*-2,3-diaminopropionic acid are recognized by the adenylation domain of Sull (Sull A1-domain) and the second adenylation domain of SulM

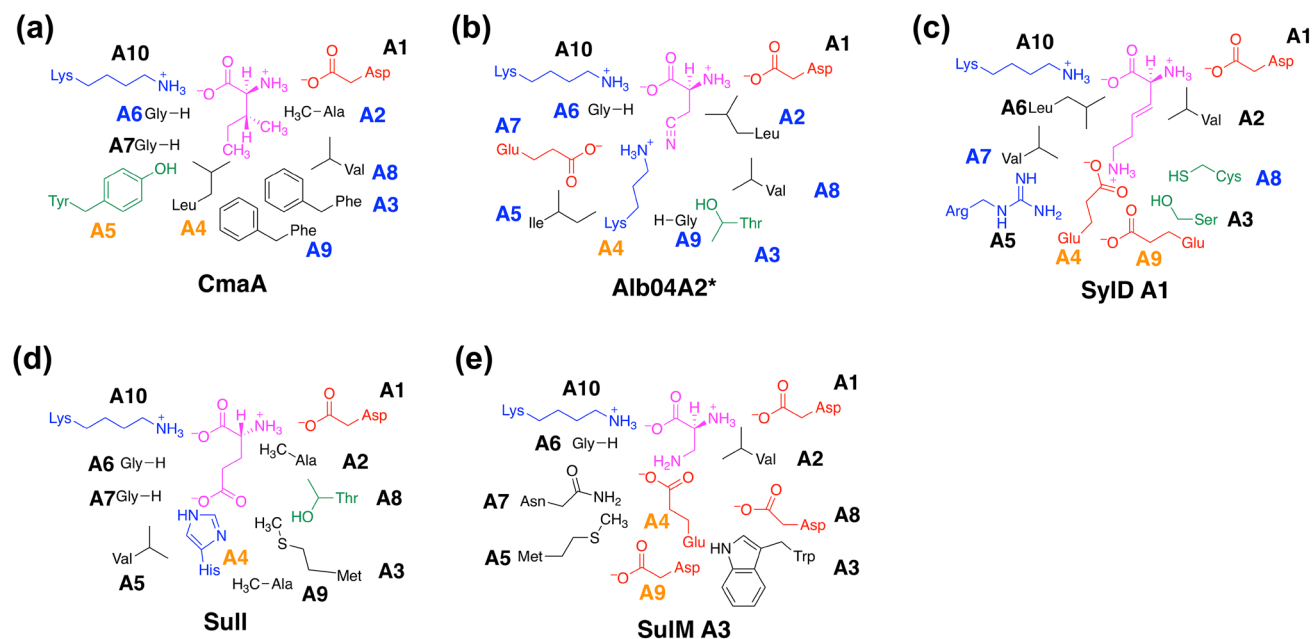


Fig. 5 Model structures of modified α -amino acid selective adenylation enzymes. **a** Model active site of *L*-allo-Ile selective adenylation enzyme CmaA. **b** Model active site of *L*-cyano-Ala selective adenylation enzyme Alb04A2*. **c** Model active site of 3,4-dehydro-*L*-Lys selective adenylation enzyme SyID A1. **d** Active site model of *D*-glutamic acid selective adenylation enzyme Sull. **(E)** Active site model

of *L*-2,3-diaminopropionic acid selective adenylation enzyme SulM A3. Orange color codes are possibly involved in the recognition of the side chain of substrate amino acid. Blue color codes are the conserved residues in the related proteinogenic α -amino acids selective adenylation enzymes

(SulM A3-domain), respectively [58]. The nonribosomal code of SulI suggests that the His residue at the A4-position may be involved in the recognition of the γ -carboxy group of D-Glu (Fig. 5d). The nonribosomal code of SulM A3-domain suggests that three acidic residues, Glu at the A4-position, Asp at the A8-position, and Asp at the A9-position, may be involved in the recognition of the β -amino group of L-2,3-diaminopropionic acid (Fig. 5e). The amino acid residue at the A4-position seems one of the main contributors to the recognition of the terminal side-chain functional groups of the amino acid substrates (Fig. 5).

Proline-related amino acids, including pipercolic acid and 4-alkylproline derivatives, are also found in NRPs and hybrid NRP-PKs. The adenylation domain of the NRPS module FkbP selectively recognizes L-pipercolic acid in FK520 biosynthesis [35]. Homologous adenylation domains in rapamycin and FK506 biosynthesis are responsible for the activation of L-pipercolic acid. The FkbP adenylation domain has a different nonribosomal code at the A2-, A4-, and A5-positions from that of L-Pro selective adenylation enzymes (A2, Val; A4, Phe; and A5, Ile) (Fig. 6a). GetE, which is a standalone adenylation enzyme in GE81112 biosynthesis, recognizes L-pipercolic acid and also recognizes L-Pro [9]. GetE has a similar nonribosomal code to L-Pro selective adenylation enzymes, except for the A4- and A8-positions (Fig. 6b). 4-Alkylproline derivatives are found, not only in the NRP griselimycin, but also in

licosamides and pyrrolbenzodiazepines [44]. Griselimycins contain two or three (2*S*,4*R*)-4-methyl-L-Pro units, which are recognized selectively by the corresponding adenylation domains in module two and module five of GriA, and module eight of GriB [59]. The nonribosomal code of the adenylation domain in module two of GriA (GriA2) shows a subtle difference from that of L-Pro selective adenylation enzymes (Fig. 6c). The small Ala residue at the A8-position may be important to accommodate the 4-methyl group of 4-methyl-L-Pro. LmbC is a standalone adenylation enzyme that selectively recognizes (2*S*,4*R*)-4-propyl-L-proline in lincomycin biosynthesis [45]. A homologous adenylation enzyme, CcbC, in celesticetin biosynthesis recognizes L-proline selectively [45]. Because LmbC and CcbC have quite different nonribosomal codes from that of L-Pro selective adenylation enzymes, the substrate recognition mechanism is likely to be different (Fig. 6d, e). Although extensive protein engineering of LmbC has been investigated, it has proved difficult to change the substrate specificity without structural information [107]. The size of the amino acid residues at the A4-position might be important to accommodate the side chains of 4-alkyl-L-proline derivatives.

Pristinamycin I_A contains several nonproteinogenic amino acids, including 3-hydroxypicolinic acid, L-2-aminobutyric acid, L-4-dimethylaminophenylalanine, 4-oxo-L-pipercolic acid, and L-phenylglycine [64, 103]. Intact native NRPSs have been purified from the producer *Streptomyces*

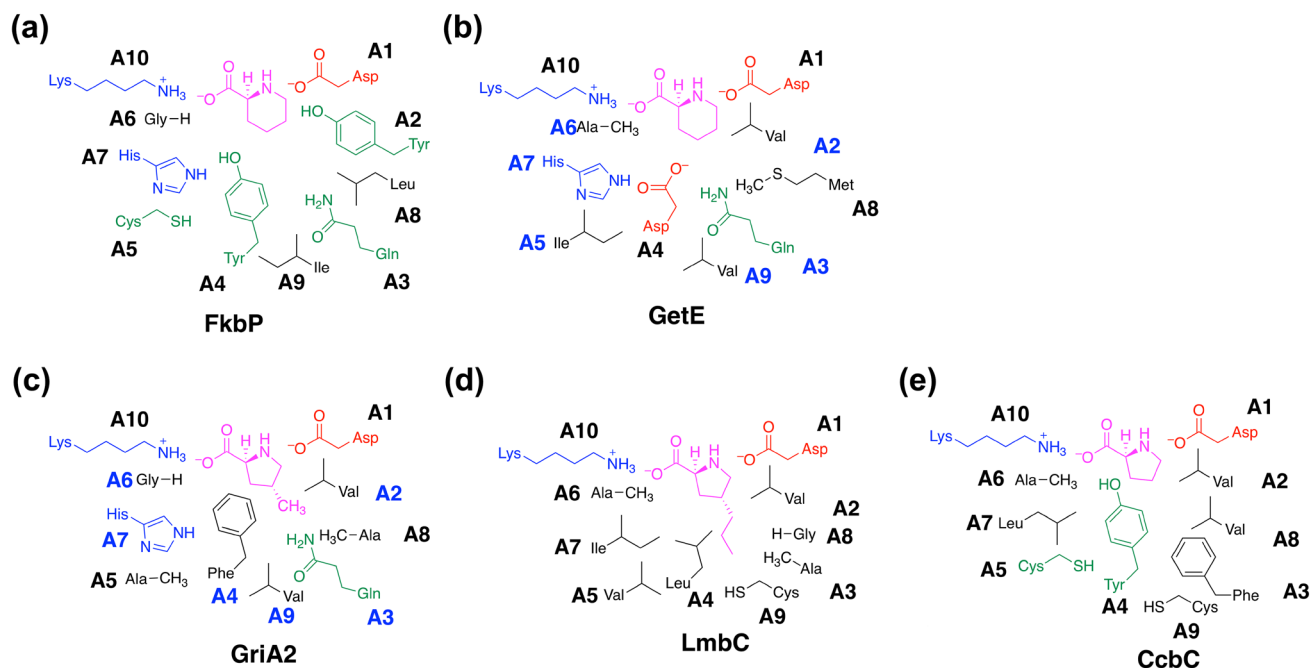


Fig. 6 Model structures of proline-related amino acids including pipercolic acid and 4-alkylproline derivatives. **a** Active site model of L-pipercolic acid selective adenylation domain of FkbP. **b** Active site model of L-pipercolic acid selective adenylation enzyme GetE. **c**

Active site model of 4-methyl-L-proline selective adenylation domain GriA2. **d** Active site model of 4-propyl-L-proline selective adenylation enzyme LmbC. **e** Active site model of L-proline selective adenylation enzyme CcbC

pristinaespiralis SP92 and the adenylation activities were examined to show the expected substrate specificity of SnbC and SnbDE. SnbC, consisting of two NRPS modules, recognizes L-Thr and L-2-aminobutyric acid. SnbDE, consisting of four NRPS modules, recognizes L-Pro, L-4-dimethylaminophenylalanine, L-pipecolic acid, and L-phenylglycine. The nonribosomal codes of SnbC-A1-domain and A2-domain are predicted to be L-Thr and L-Val selective codes, respectively. The nonribosomal codes of SnbDE-A1-, A2-, A3-, and A4-domains are predicted to be selective for L-Pro, L-Phe, L-pipecolic acid, and an unknown amino acid, respectively. The adenylation domain of SnbA is predicted to recognize 3-hydroxypicolinic acid. SnbDE-A2-domain has an Ala residue at the A4-position instead of Thr, as is the case for PheA, which appears to expand the room available to be able to accommodate the 4-dimethylamino moiety of L-4-dimethylaminophenylalanine (Fig. 7a). Pyridomycin contains 3-(3-pyridyl)-L-alanine and 3-hydroxypicolinic acid units in a hybrid NRP-PK structure [41]. The adenylation domain of PyrA that is a loading module of NRPS recognizes 3-hydroxypicolinic acid as a starter unit of NRPS. The adenylation domain of PyrA shows a 63% identity/71% similarity to the adenylation domain of SnbA. In congoicidine biosynthesis, 4-acetamidopyrrole-2-carboxylic acid and guanidinoacetic acid are recognized by the standalone adenylation enzyme Cgc3* and the adenylation domain of

Cgc18, respectively [2]. Cgc3* recognizes pyrrole-2-carboxylic acid derivatives, but does not activate L-proline or picolinic acid. The PKS/NRPS analysis (<http://nrps.igs.umaryland.edu>) [4] of the target adenylation enzymes that recognizes nonproteinogenic amino acids often afford “no hit” as the nonribosomal codes. The other web-based NRPS substrate prediction tools such as SEQL-NRPS (<https://services.birc.au.dk/seq-nrps/>) [51] and NRPSp (<http://www.nrpsp.com>) [82] predict wrong amino acid residues presumably due to less structural information. In fact, because of the insertion or truncation of sequences in the multiple sequence alignment with the PheA sequence, it is difficult to predict the nonribosomal codes for L-phenylglycine, 3-hydroxypicolinic acid, pyrrole-2-carboxylic acid derivatives, and guanidinoacetic acid selective adenylation enzymes.

In mycobactin biosynthesis, *N*⁶-hydroxy-L-Lys and *N*⁶-decanoyl-*N*⁶-hydroxy-L-Lys are recognized by MbtF and MbtE, respectively [66]. The *N*⁶-hydroxy group is critical for the recognition. In addition, it is noteworthy that the associated MbtH protein is necessary for the adenylation activity of MbtE and MbtF. From examining the nonribosomal codes for MbtF and MbtE, it is difficult to find common features that contribute to the recognition of the *N*⁶-hydroxy group of substrates (Fig. 7b, c).

The adenylation domain of the NRPS module VioG strictly recognizes capreomycin in viomycin biosynthesis

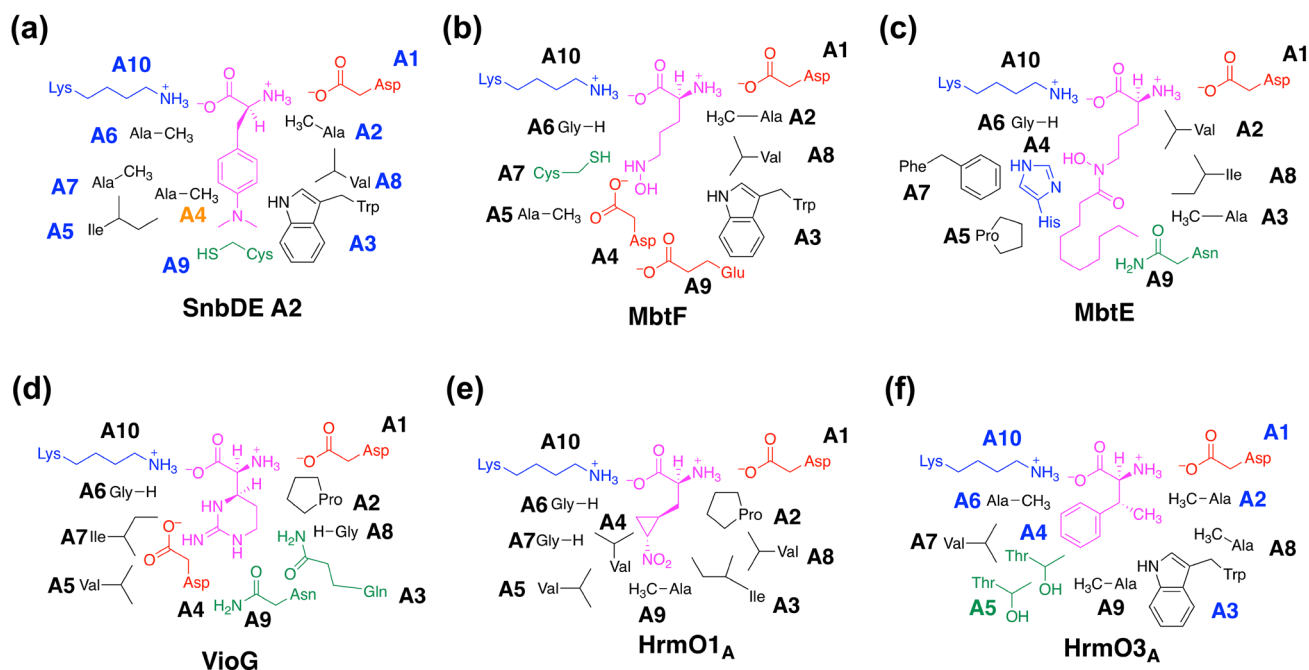


Fig. 7 Model structures of highly modified α -amino acid selective adenylation enzymes. **a** Active site of L-4-dimethylamino-Phe selective adenylation enzyme SnbDE A2. **b** Active site model of *N*⁶-hydroxy-L-Orn selective adenylation domain of MbtF. **c** Active site model of *N*⁶-decanoyl-*N*⁶-hydroxy-L-Orn selective adenylation

domain of MbtE. **d** Active site model of capreomycin selective adenylation domain of VioG. **e** Active site model of L-3-(*trans*-2-nitrocyclopropyl)Ala selective adenylation domain HrmO1_A. **f** Active site model of β -methyl-Phe selective adenylation domain HrmO3_A

[30, 31]. The nonribosomal code of the adenylation domain of VioG suggests that the Asp residue at the A4-position might be responsible for recognition of the guanidine moiety of capreomycin (Fig. 7d). The adenylation domains of VioF and CmnF are predicted to recognize 2,3-diaminopropanoic acid in viomycin and capreomycin biosynthesis, respectively, although the nonribosomal codes are different from that of the SullI adenylation domain in sulfazecin biosynthesis [31].

HrmO1_A (also HrmO4_A), HrmO3_A (also HrmP1_A), and HrmP3_A recognize L-3-(*trans*-2-nitrocyclopropyl)Ala, L-β-methyl-Phe, and L-4-(*Z*)-propenylproline, respectively, in hormaomycin biosynthesis [21, 40]. The MbtH homolog HrmR is necessary for the adenylation enzymatic activity [21]. HrmO1_A has a unique nonribosomal code and thus it is difficult to speculate on the recognition mechanism (Fig. 7e). HrmO3_A has a similar nonribosomal code to the L-Phe-selective PheA, except for the A7-, A8-, and A9-positions (Fig. 7f), which cause the different substrate recognition mechanism. HrmP3_A has a similar nonribosomal code to GriA2 that recognizes (2*S*,4*R*)-4-methyl-L-Pro. Hormaomycin also contains 5-chloro-1-hydroxypyrrole-2-carboxylic acid at the starter position.

JBIR-34 is a 4-methyloxazoline-containing NRP with 6-chloro-4-hydroxy-indole-3-carboxylic acid, D-Ala, and L-Ser amino acid units [78]. The A-CP didomain enzyme FmoA1 recognizes 6-chloro-L-Trp in the biosynthesis of

6-chloro-4-hydroxy-indole-3-carboxylic acid. The nonribosomal code of FmoA1 is different from that of L-Trp selective adenylation enzymes indicating that the 6-Cl group on 6-chloro-L-Trp greatly affects the substrate recognition mechanism (Fig. 8a). 4-Methyloxazoline is constructed from D-Ala via α-methyl-L-Ser, which is recognized by the adenylation domain of FmoA3. The nonribosomal code of FmoA3 contains hydrophilic amino acids (Fig. 8b), so it is difficult to speculate how the α-methyl group is accommodated at the active site. D-Ala and L-Ser are recognized by the adenylation domains of FmoA4 and FmoA5, respectively. The nonribosomal code of FmoA4 (Fig. 8c) is different from that of DltA, which also recognizes D-Ala (Fig. 4a). It was difficult to predict the nonribosomal code for the adenylation domain of FmoA2, which recognizes 6-chloro-4-hydroxy-indole-3-carboxylic acid, because the sequence of the adenylation domain of FmoA2 is largely different from that of PheA.

TioJ in thiocoraline biosynthesis recognizes 3-hydroxyquinaldic acid, while Ecm1 in echinomycin biosynthesis recognizes 2-quinoxalinecarboxylic acid [76]. The nonribosomal codes of TioJ and Ecm1 are the same, except for the A1-position, which is Asn in TioJ and Gln in Ecm1 (Fig. 8d, e). Thus, this subtle difference might affect the substrate recognition mechanism and/or other structural differences may contribute to distinguishing between these amino acid substrates. It should be noted that the ring nitrogen atom in

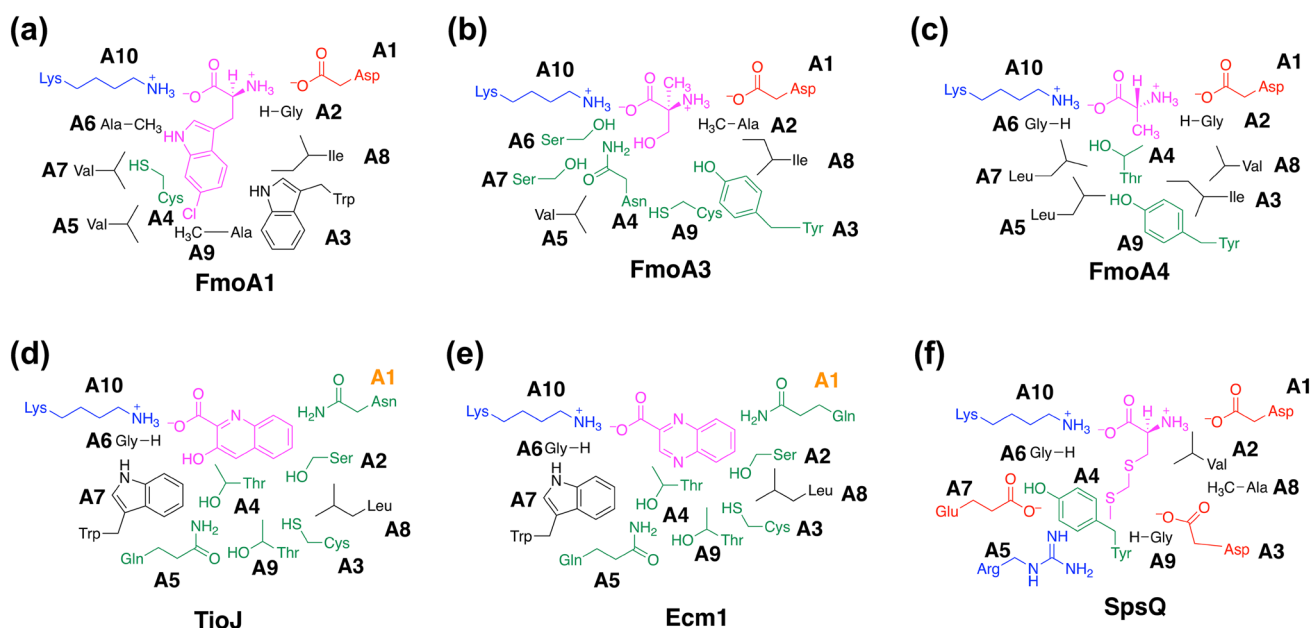


Fig. 8 Model structures of unique amino acid selective adenylation enzymes in JBIR-34, thiocoraline, echinomycin, and sparsomycin biosynthesis. **a** Active site model of 6-chloro-L-Trp selective adenylation domain FmoA1. **b** Active site model of α-methyl-L-Ser selective adenylation domain FmoA3. **c** Active site model of D-Ala selective

adenylation domain FmoA4. **d** Active site model of 3-hydroxyquinaldic acid selective adenylation domain of TioJ. **e** Active site model of quinoxaline-2-carboxylic acid selective adenylation domain of Ecm1. **f** Active site model of L-methylthiomethylcysteine selective adenylation domain of SpsQ

quinaldic acid and quinoxalinecarboxylic acid has different chemical properties from that of a typical amino acid amino group, and thus, the Asp residue at the A1-position may be replaced by polar amino acids, such as Asn and Gln.

In sparsomycin biosynthesis, 6-methyl-uracil acrylic acid and L-methylthiomethylcysteine (MTM-Cys) are recognized by the first adenylation domain of SpsR and the A-CP didomain enzyme SpsQ, respectively [89]. The nonribosomal code of SpsQ does not provide enough information to speculate on the substrate recognition mechanism of MTM-Cys (Fig. 8f).

In the biosynthesis of phosphinothricyl-alanyl-alanine and phosphinothricyl-alanyl-leucine, the adenylation enzyme PhsA recognizes *N*-acetyldemethylphosphinothricin or *N*-acetylphosphinothricin [11, 55, 92]. The amino acid residue at the A1-position of PhsA is Val instead of Asp, presumably because the amino group of the amino acid substrate is *N*-acetylated [11]. Therefore, the substrate recognition mechanism of PhsA is quite different from typical amino acid selective adenylation enzymes.

β -Amino acid selective adenylation enzymes (Figs. 9 and 10)

The protein structural analyses of three naturally occurring β -amino acid selective adenylation enzymes, VinN [73] in vicenistatin biosynthesis, CmiS6 [17] in cremimycin biosynthesis, and IdnL1 [17] in incednine biosynthesis, have been reported to date (Table 1) [75]. Vicenistatin, cremimycin, and incednine belong to a family of macrolactam antibiotics with a unique β -amino acid starter unit in the polyketide skeleton [54]. VinN recognizes (2*S*,3*S*)-MeAsp selectively and ligates it to a discrete CP, VinL. CmiS6 recognizes 3-aminononanoic acid (3-ANA, the stereochemistry is unknown) selectively and ligates it to a discrete CP, CmiS4. IdnL1 recognizes (*S*)-3-aminobutyric acid (3-ABA) selectively and ligates it to a discrete CP, IdnL6. These adenylation enzymes share a common β -amino acid selective cavity, which prevents the binding of proteinogenic L- α -amino acids (Fig. 9) [17, 73]. The VinN, CmiS6, and IdnL1 family of enzymes have a Phe or Tyr residue at the A2-position (Table 2). This aromatic residue at the A2-position seems

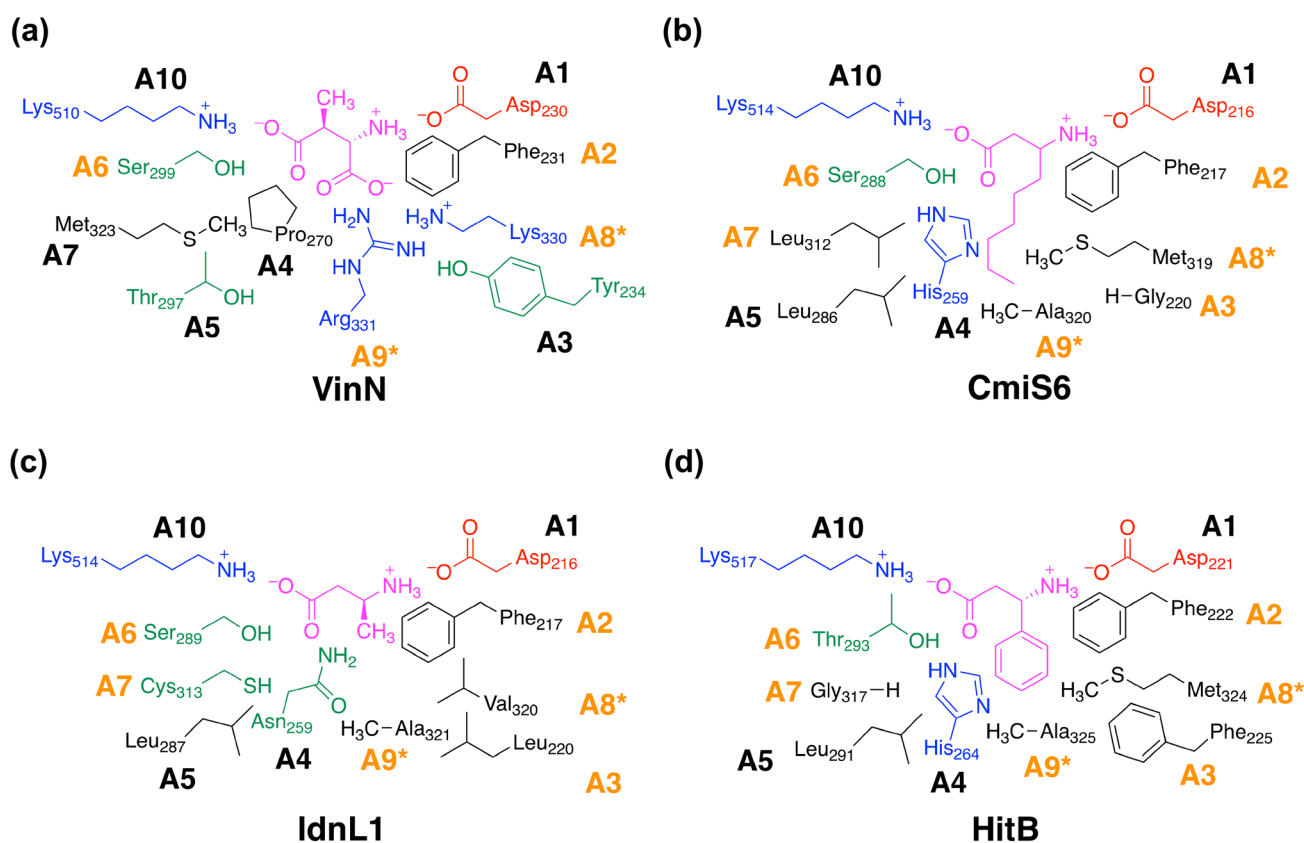


Fig. 9 Active sites of VinN/IdnL1/CmiS6 family of β -amino acid selective adenylation enzymes. **a** (2*S*,3*S*)-MeAsp selective adenylation enzyme VinN (structurally elucidated), **b** 3-aminononanoic acid selective adenylation enzyme CmiS6 (structurally elucidated), **c**

3-aminobutyric acid selective adenylation enzyme IdnL1 (structurally elucidated) **d** β -Phe selective adenylation enzyme HitB (model structure)

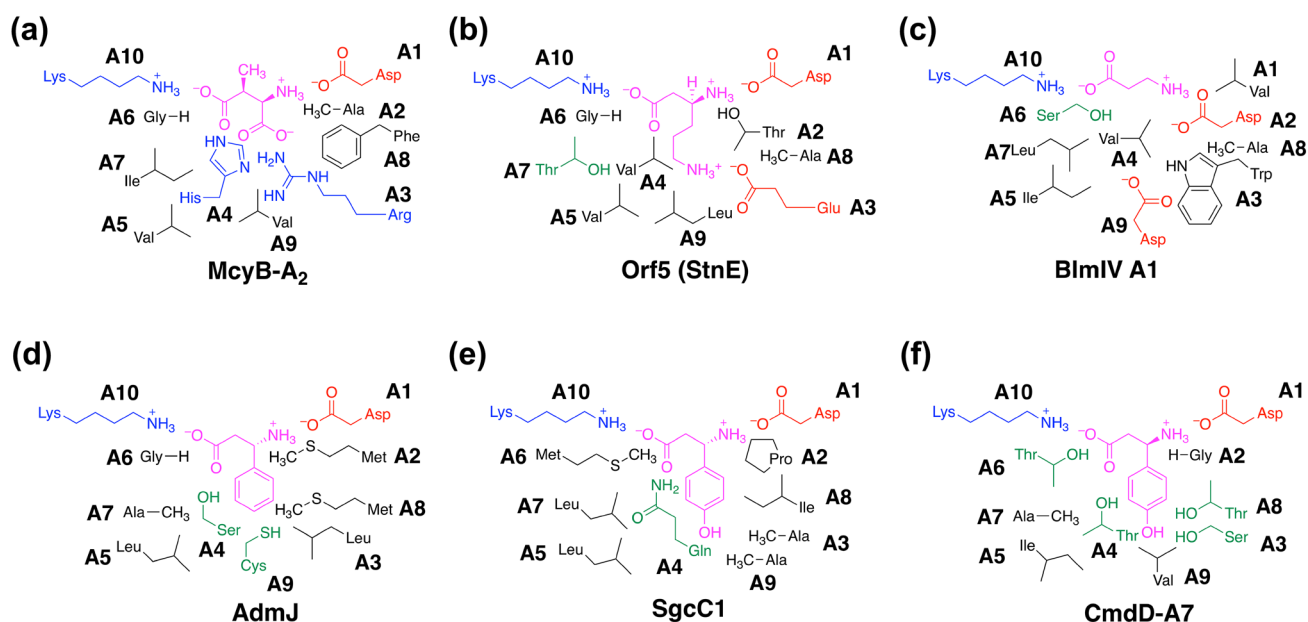


Fig. 10 Model structures of β -amino acid selective adenylation enzymes. **a** Active site model of (*2R,3S*)- β -methyl-D-aspartic acid selective adenylation domain McyB-A₂. **b** Active site model of (*S*)- β -Lys selective adenylation enzyme Orf5 (StnE). **c** Active site model

of β -Ala selective adenylation domain BlmIV A1. **d** Active site model of β -Phe selective adenylation enzyme AdmJ. **e** Active site model of β -Tyr selective adenylation enzyme SgcC1. **f** Active site model of β -Tyr selective adenylation domain CmdD-A7

to clash with the side-chain of *L*- α -amino acids. In contrast, the C3 methyl group of (*2S,3S*)-MeAsp seems to be well accommodated by VinN, because the K_m value of VinN for (*2S,3S*)-MeAsp is 35 times lower than that for *L*-Asp. The C3 methyl group of (*2S,3S*)-MeAsp seems to be recognized by the Phe residue at the A2-position through CH– π and van der Waals interactions. The second feature of this enzyme family is a Ser or Thr residue at the A6-position, which would clash with the side chain of *L*- α -amino acids. These two features at the A2- and A6-positions of the VinN/IdnL1/CmiS6 family of enzymes appear to strictly prohibit these enzymes from accommodating *L*- α -amino acids. The third feature is that “one amino acid is skipped” just before the A8/A9 position according to the protein structural analysis, although it is difficult to determine “the skipped sequence motif” only by sequence alignment [73]. The pentapeptide sequence motif (Ser or Thr/Glu/Cys or Ser or Thr/A8/A9) where the Glu residue is highly conserved contributes to making a wider active site to accommodate the side-chain at the β -position of β -amino acids. This structural motif was introduced into the *L*- α -Phe selective adenylation enzyme TycA to alter the substrate specificity resulting in a mutant enzyme that could accommodate (*S*)- β -Phe [80]. VinN has a Lys residue at the A8-position and an Arg residue at the A9-position that interact with the C3 carboxy group of (*2S,3S*)-MeAsp (Fig. 9a). IdnL1 and CmiS6 have a Val or Met residue at the A8-position and an Ala residue at the A9-position to accommodate the alkyl side chain at the β -position, the methyl and

hexyl groups, respectively (Fig. 9b, c). Leu220/Cys313 (A3/A7) in IdnL1 and Gly220/Leu312 (A3/A7) in CmiS6 seem to indirectly make the appropriate hydrophobic atmosphere to accommodate the appropriate length of alkyl side chain at the C3 position. The CmiS6 G220L variant can accommodate 3-ABA as a substrate, but has lost the affinity for 3-ANA, suggesting that the A3-position is involved in the selection of the alkyl side chain of 3-amino fatty acids [17]. The stereochemistry at C3 might affect the selective recognition by these enzymes, although only one enantiomer of β -amino acids is likely to be biosynthesized in the producer strains. In summary, the VinN/IdnL1/CmiS6 family of enzymes shares a common nonribosomal code (A1, A2, A6, A10, and pre A8/A9 positions). The combinations of A3/A7 and A8/A9 seem to be a determinant for selectivity for the substituents at the C3 position of β -amino acids [75]. FlvN in fluvirucin B2 biosynthesis is known to recognize *L*-Asp and has the same nonribosomal code as VinN, except for the A2- and A6-positions (Tyr at the A2-position and Thr at the A6-position) [7, 74]. These subtle differences may cause the different recognition mechanisms for *L*-Asp and MeAsp. HitB in hitachimycin biosynthesis is believed to recognize (*S*)- β -Phe in a similar manner to the VinN/IdnL1/CmiS6 family of enzymes [53, 75] (Fig. 9d).

Several other β -amino acid selective adenylation enzymes have been also characterized, although the protein structures have not been elucidated yet. The adenylation domain McyB-A2 in microcystin biosynthesis is known to recognize

the β -amino acid (2*R*,3*S*)-MeAsp [50, 69]. The nonribosomal code of McyB-A2 is different from that of the VinN/IdnL1/CmiS6 family of enzymes, because McyB-A2 does not have an aromatic residue at the A2-position or Ser/Thr at the A6-position (Fig. 10a and Table 2). It is also difficult to find the third feature of the VinN/IdnL1/CmiS6 family of enzymes in which one amino acid is skipped just before the A8/A9-position (please see Fig. 9 and the corresponding sentences to describe the structure of VinN) [50, 69]. McyB-A2 has an Arg at the A3-position and a His at the A4-position, both of which seem to be involved in the recognition of the carboxy group of MeAsp.

The standalone adenylation enzymes Orf5 (StnE) and Orf19 (StnS) in streptothricin biosynthesis are known to recognize β -Lys [50, 57, 63]. The nonribosomal codes of Orf5 (StnE) and Orf19 (StnS) are different from that of the VinN/IdnL1/CmiS6 family of enzymes, because Orf5 (StnE) and Orf19 (StnS) do not have an aromatic residue at the A2-position or Ser/Thr at the A6-position [73] (Fig. 10b). The Glu residue at the A3-position of Orf5 (StnE) and Orf19 (StnS) may be involved in the recognition of terminal amino group of β -Lys. The first adenylation domain of BlmIV (BlmIV A1-domain) in bleomycin biosynthesis is predicted to recognize β -Ala [34, 50]. The amino acid at the A1-position of BlmIV A1-domain is not Asp, but the amino acid at the A2-position is Asp, which is presumed to be involved in the recognition of the amino group of β -Ala (Fig. 10c). The nonribosomal code at the A6-position of BlmIV A1-domain has a Ser residue, which would clash with the side chain of L- α -amino acids, as seen with the VinN/IdnL1/CmiS6 family of enzymes. The Asp residue at the A9-position might be also involved in the recognition of the β -amino group of β -Ala.

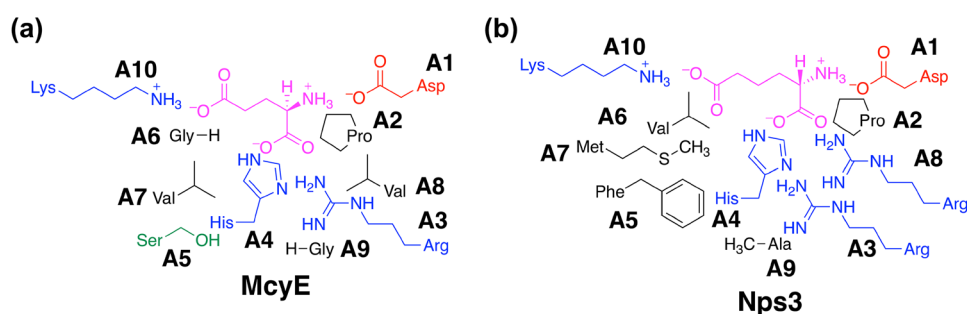
Many β -Phe derivatives are known to be involved in NRP and hybrid NRP–PK biosynthesis. AdmJ in andrimid biosynthesis recognizes (*S*)- β -Phe [61]. SgcC1 recognizes (*S*)- β -Tyr in the biosynthesis of the enediyne antibiotic C-1027 [106]. CmdD-A7-domain in the biosynthesis of chondramides recognizes (*R*)- β -Tyr [85]. It is difficult to find common features among these adenyating domains selective for β -Phe derivatives, except for the A1- and A10-positions, because the nonribosomal codes of these enzymes are different from those of the VinN/IdnL1/CmiS6 family of enzymes,

including HitB that recognizes (*S*)- β -Phe in hitachimycin biosynthesis (Figs. 9, 10d–f) [53, 73]. The A6-positions of SgcC1 and CmdD-A7 are Met and Thr, respectively, which may clash with the side chain of L- α -amino acids similar to the VinN/IdnL1/CmiS6 family of enzymes. In contrast, AdmJ has a Gly residue at the A6-position as is typical for L- α -amino acid selective adenylation enzymes. The Pro residue (P571) of SgcC1 at the A2-position seems to affect the catalytic reaction mechanism, because the k_{cat}/K_m value of the SgcC1 P571A variant is reduced 142-fold [106]. The successful protein engineering of TycA (engineered TycA, PDB ID 5N82), which was able to recognize (*S*)- β -Phe after the introduction of the skipped sequence motif at preA8/A9 [80] based on the VinN structure, indicate that a wide active site appears to be one of the key properties to accommodate β -amino acids. Several different recognition mechanisms for β -amino acids appear to be used depending on the different adenylation enzymes. Structural analysis of these unique β -amino acid selective adenyating enzymes will help to better understand the substrate recognition mechanisms.

γ -Amino acids, δ -amino acids, aminobenzoic acids, and others (no structural analysis to date)

The adenyating enzymes selective for γ -amino acids, such as D-glutamic acid in microcystin biosynthesis [69], δ -amino acids, such as L- α -aminoadipic acid in penicillin biosynthesis [49, 91], and other unusual amino acids have not been structurally elucidated. McyE in microcystin biosynthesis recognizes D-Glu as γ -amino acid and has a similar nonribosomal code to McyB-A2-domain, which recognizes the β -amino acid (2*R*,3*S*)-MeAsp (Fig. 11a) [69]. The conserved Arg at the A3-position and His residues at the A4-position might be involved in the recognition of the α -carboxy group [69]. The adenylation domain of α -aminoadipate reductase Nps3 recognizes L-2-aminoadipic acid as δ -amino acid in lysine biosynthesis of the basidiomycete *Ceriporiopsis subvermispota* [46, 47]. The nonribosomal code at the A6-position of Nps3 is Val, which would clash with the side chain of L- α -amino acids (Fig. 11b). Two Arg residues at the A3- and

Fig. 11 Model structures of γ -amino acids selective adenylation enzyme and δ -amino acids selective adenylation enzyme. **a** Active site model of D-Glu as γ -amino acids selective adenylation domain of McyE. **b** Active site model of L-2-aminoadipic acid as δ -amino acids selective adenylation domain of Nps3



A8-positions, and the His residue at the A4-position might be involved in the recognition of the α -carboxy group, as for McyB-A2 and McyE. The α -carboxy groups of D-glutamic acid and L-2-aminoadipic acid appear to be held distant from the adenylation site by interactions with several basic residues, resulting in only the terminal carboxy group of the substrates being able to react with ATP. Because only a comparison of nonribosomal codes has been discussed, structural analysis of unique adenylation enzymes will be required to understand the amino acid substrate recognition mechanisms.

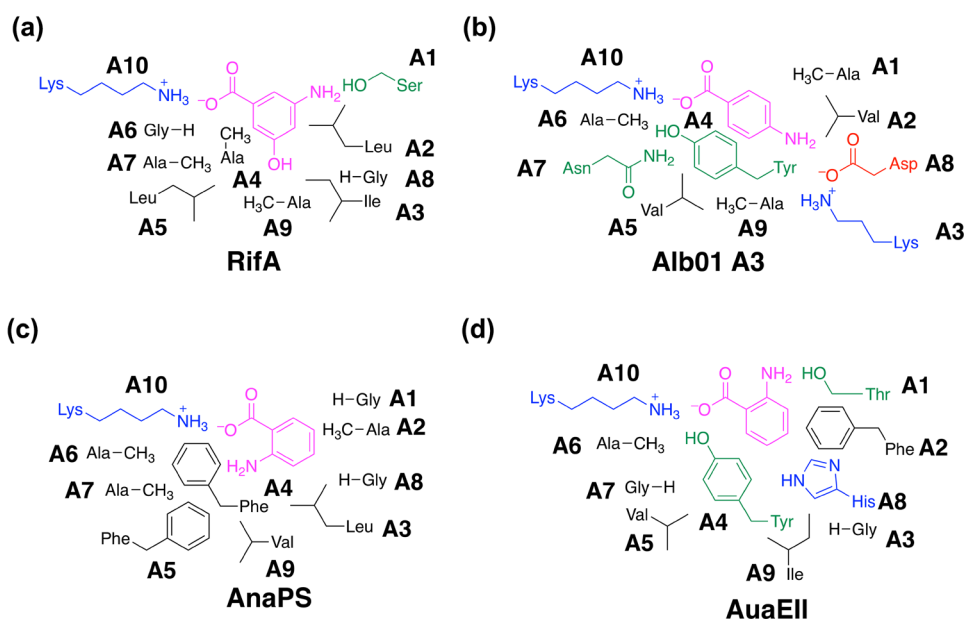
Aminobenzoic acid derivatives are also unique components of NRPs and hybrid NRP–PKs. The loading adenylation domain of RifA recognizes 3-amino-5-hydroxybenzoic acid (AHBA) in rifamycin biosynthesis [1]. The nonribosomal code for AHBA in the adenylation domain of RifA can be predicted, indicating a hydrophobic active site (Fig. 12a). In albicidin biosynthesis, four *p*-aminobenzoic acids (PABAs) are recognized by Alb01A1-domain, A3-domain, Alb09A4-domain, and A5-domain [18]. The nonribosomal code for PABA in Alb adenylation enzymes indicates that the Asp residue at the A8-position may interact with the amino group of the aniline moiety (Fig. 12b). The first adenylation domain of the fungal NRPS AnaPS recognizes anthranilic acid in acetylszonalenin biosynthesis [3]. The nonribosomal code for anthranilic acid in AnaPS indicates a hydrophobic active site (Fig. 12c). Interestingly, in aurachin biosynthesis, the standalone adenylation enzyme AuaEII recognizes anthranilic acid and ligates it with coenzyme A to afford anthraniloyl-CoA [43, 81]. Then, another ANL family enzyme AuaE, which lacks the Lys residue at the A10-position, transfers the anthranilate onto the CP AuaB.

The nonribosomal code for the anthranilic acid selective adenylation enzyme AuaEII is different from that of AnaPS (Figs. 12c, d). Based on the crystal structure of AuaEII with anthranilic acid adenylate, the Thr residue at the A1-position and the His residue at the A8-position appear to interact with the amino group of anthranilic acid [43]. Because the amino acid residue at the A1-position is not Asp in the adenylation enzymes of these aminobenzoic acid derivatives, the amino group of aminobenzoic acid derivatives seems to be recognized in a different manner to that of amino acids.

Mining of nonproteinogenic amino acid selective adenylation enzymes

Many adenyating enzymes for nonproteinogenic amino acids have been predicted based on the NRPS and hybrid NRPS–PKS assembly lines in the target biosynthetic gene cluster. Norcoronamic acid [(1*S*,2*S*)-1-amino-2-methylcyclopropane-1-carboxylic acid] is known to be involved in the biosynthesis of the quinomycin family of nonribosomal peptides, SW-163 s [110]. The results of feeding norcoronamic acid into the producer strain indicated that the adenylation domain in Swb17 is responsible for the recognition of norcoronamic acid. Polymyxins are cyclic nonribosomal lipopeptides that contain four to five 2,4-diaminobutyric acid residues [16, 33]. The highly conserved Glu residue at the A4-position of the NRPS adenylation domains is proposed to be involved in the recognition of the terminal amino group. Enduracidin contains enduracididine, citrulline, and six L-hydroxyphenylglycine derivatives [113]. The

Fig. 12 Active sites of adenylation enzymes of aminobenzoic acids. **a** Proposed active site model of 3-amino-5-hydroxybenzoic acid selective adenylation domain of RifA. **b** Proposed active site model of PABA selective adenylation domain Alb01 A3. **c** Proposed active site model of anthranilic acid selective adenylation domain of AnaPS. **d** Active site of anthranilic acid selective adenylation enzyme AuaEII



adenylation domains that recognize these nonproteinogenic amino acids can be predicted based on the domain organization. Daptomycin contains kynurenine as a nonproteinogenic amino acid unit. The second adenylation domain in the NRPS module DptD2 is responsible for the recognition of kynurenine in daptomycin biosynthesis [70, 88]. However, the adenylation activity of these enzymes has not been investigated yet.

The biosynthetic gene clusters of several NRPs and hybrid NRP–PKs containing piperazic acid units [28, 79], including sangliferrin A [84], himastatin [60], auranti-mycin [115], polyoxypeptin A [27], and kutznerides [32] have been identified. This family of compounds contains many nonproteinogenic amino acids. For example, kutznerides contain norcoronamic acid, *O*-methyl-L-Ser, 3-hydroxy-D-Glu, (2*S*,3*aR*,8*aS*)-6,7-dichloro-3*a*-hydroxy-hexahydropyrrolo[2,3-*b*]indole-2-carboxylic acid, and chlorinated piperazic acid. The adenylation domain responsible for recognizing piperazic acid can be predicted based on the domain organization. However, the adenylation activity of adenylation enzymes selective for piperazic acid has not been investigated yet. Because of the commercial unavailability of nonproteinogenic amino acids, functional analysis, including protein structural analysis of nonproteinogenic amino acid selective adenylating enzymes is limited. Efficient synthetic methodology for nonproteinogenic amino acids and mimics is required to overcome this situation. Further functional and structural analyses will pave a way to better understand the substrate recognition mechanisms.

Engineering of adenylation enzymes

Engineering of the substrate specificity of adenylation enzymes is an attractive method in synthetic biology for creating new molecules [5, 6, 29, 111]. The adenylation enzyme CdaPS3 selective for L-Glu (also 3-methyl-L-Glu) in the biosynthesis of calcium-dependent antibiotic (CDA) has been successfully engineered to recognize 3-methyl-L-Gln [104]. Double mutation of CdaPS3, Q236E at the A2-position and K278Q at the A4-position, resulted in the production of CDA derivatives containing L-Gln or 3-methyl-L-Gln. The L-Phe selective adenylation enzyme PheA (GrsA_A) in gramicidin S biosynthesis has been successfully engineered to recognize *O*-propargyl-L-Tyr by introduction of a single mutation W239S at the A3-position [52]. Furthermore, a successful attempt to alter the substrate specificity from an α -amino acid to a β -amino acid has been reported recently, based on structural analysis of the β -amino acid selective adenylation enzyme VinN [80]. Replacement of the tetrapeptide sequence Thr328–Ser329–Ile330(A8)–Cys331(A9) with randomized tripeptides and a random mutation at A236

(A2-position) in the W239S variant of TycA in tyrocidine biosynthesis resulted in an α/β switch. Therefore, structural elucidation of nonproteinogenic amino acid selective adenylation enzymes will have a significant impact on rational protein engineering to change the substrate specificity, leading to the engineered biosynthesis of designed NRPs and hybrid NRP–PKs.

Protein–protein interactions between adenylation enzymes and associated proteins

In addition to the nonribosomal code of adenylation enzymes that selectively recognizes amino acids; the adenylation enzymes must recognize the acceptor CPs. In fact, the associated proteins, such as condensation enzymes and CPs, affect the adenylation enzymatic activity by interacting with the A_{sub} domain of adenylation enzymes [68]. Successful combinatorial biosynthetic approach by manipulating NRPS assembly lines indicates that C–A di-domain and C–A–CP *tri*-domain may have coevolved by maintaining appropriate protein–protein interaction [5, 6]. However, it is still unknown exactly how the acceptor CP interacts with the partner adenylation enzyme and the condensation enzymes [71, 102]. It is anticipated that the structures of adenylation enzyme–CP complexes will be further elucidated in future to better understand this protein–protein interaction. The complex crystal structure of the EntE (dihydroxybenzoic acid adenylation enzyme)–EntB (CP) fusion protein with a vinyl sulfamoyl adenosine derivative as a molecular probe to form a covalent bond with the thiol of EntB, shows the interface between the adenylation enzyme and CP [97]. The recent crystal structures of several conformations of full NRPS modules show that the CP domain is delivered to each catalytic domain depending on the reaction stage [23, 37, 87]. Accumulation of structural data for the protein–protein interaction between adenylation enzymes and CPs will enhance our knowledge and promote being able to produce the CP targeting code of the adenylation enzyme, which is important for the rational engineering of NRPSs and hybrid NRPS–PKSs [13, 42].

Conclusion

In this review, we have summarized the current status of structural knowledge on nonproteinogenic amino acid selective adenylation enzymes. Currently, only limited structural information of nonproteinogenic amino acid selective adenylation enzymes is available. As we pointed out, protein structural information helps rational engineering of adenylation enzymes to introduce unique amino acids such as

nonproteinogenic amino acids into NRPS assembly lines. For future successful engineered biosynthesis of NRPs and hybrid NRP–PKs for drug discovery, we emphasize that more extensive efforts for protein structural analysis are necessary to overcome the current situation to better understand the substrate recognition mechanisms of nonproteinogenic amino acid selective adenylation enzymes.

Acknowledgements This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) through a Grant-in-Aid for Scientific Research in Innovative Areas (17H05434 to F. K. and 16H06451 to T. E.), The Uehara Memorial Foundation (to F. K.) and the Japan Society for the Promotion of Science A3 Foresight Program.

References

- Admiraal SJ, Walsh CT, Khosla C (2001) The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. *Biochemistry* 40:6116–6123
- Al-Mestarihi AH, Garzan A, Kim JM, Garneau-Tsodikova S (2015) Enzymatic evidence for a revised congocidine biosynthetic pathway. *ChemBioChem* 16:1307–1313. <https://doi.org/10.1002/cbic.201402711>
- Ames BD, Walsh CT (2010) Anthranilate-activating modules from fungal nonribosomal peptide assembly lines. *Biochemistry* 49:3351–3365. <https://doi.org/10.1021/bi100198y>
- Bachmann BO, Ravel J (2009) Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol* 458:181–217. [https://doi.org/10.1016/s0076-6879\(09\)04808-3](https://doi.org/10.1016/s0076-6879(09)04808-3)
- Baltz RH (2014) Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth Biol* 3:748–758. <https://doi.org/10.1021/sb3000673>
- Baltz RH (2018) Synthetic biology, genome mining, and combinatorial biosynthesis of NRPS-derived antibiotics: a perspective. *J Ind Microbiol Biotechnol* 45:635–649. <https://doi.org/10.1007/s10295-017-1999-8>
- Barajas JF, Zargar A, Pang B, Benites VT, Gin J, Baidoo EEK, Petzold CJ, Hillson NJ, Keasling JD (2018) Biochemical Characterization of β -Amino Acid Incorporation in fluvirucin B2 biosynthesis. *ChemBioChem* 19:1391–1395. <https://doi.org/10.1002/cbic.201800169>
- Baranasic D, Zucko J, Diminic J, Gacesa R, Long PF, Cullum J, Hranueli D, Starcevic A (2014) Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J Ind Microbiol Biotechnol* 41:461–467. <https://doi.org/10.1007/s10295-013-1322-2>
- Binz TM, Maffioli SI, Sosio M, Donadio S, Müller R (2010) Insights into an unusual nonribosomal peptide synthetase biosynthesis: identification and characterization of the GE81112 biosynthetic gene cluster. *J Biol Chem* 285:32710–32719. <https://doi.org/10.1074/jbc.M110.146803>
- Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dick-schat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>
- Blodgett JA, Zhang JK, Yu X, Metcalf WW (2016) Conserved biosynthetic pathways for phosalacine, bialaphos and newly discovered phosphonic acid natural products. *J Antibiot* 69:15–25. <https://doi.org/10.1038/ja.2015.77>
- Bozhuyuk KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert CP, Bode HB (2018) De novo design and engineering of non-ribosomal peptide synthetases. *Nat Chem* 10:275–281. <https://doi.org/10.1038/nchem.2890>
- Brown AS, Calcott MJ, Owen JG, Ackerley DF (2018) Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat Prod Rep*. <https://doi.org/10.1039/c8np00036k>
- Challis GL, Naismith JH (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* 14:748–756. <https://doi.org/10.1016/j.sbi.2004.10.005>
- Challis GL, Ravel J, Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* 7:211–224
- Choi SK, Park SY, Kim R, Kim SB, Lee CH, Kim JF, Park SH (2009) Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. *J Bacteriol* 191:3350–3358. <https://doi.org/10.1128/JB.01728-08>
- Cieślak J, Miyayama A, Takaku R, Takaishi M, Amagai K, Kudo F, Eguchi T (2017) Biochemical characterization and structural insight into aliphatic β -amino acid adenylation enzymes IdnL1 and CmiS6. *Proteins* 85:1238–1247. <https://doi.org/10.1002/prot.25284>
- Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noell J, Pieretti I, Hugelland M, Kemper S, Mainz A, Rott P, Royer M, Süssmuth RD (2015) The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat Chem Biol* 11:195–197. <https://doi.org/10.1038/nchembio.1734>
- Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* 16:4174–4183
- Couch R, O'Connor SE, Seidle H, Walsh CT, Parry R (2004) Characterization of CmaA, an adenylation-thiolation didomain enzyme involved in the biosynthesis of coronatine. *J Bacteriol* 186:35–42
- Cruesemann M, Kohlhaas C, Piel J (2013) Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem Sci* 4:1041–1045. <https://doi.org/10.1039/c2sc21722h>
- Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster AL, Magarvey NA (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat Chem Biol* 12:1007–1014. <https://doi.org/10.1038/nchembio.2188>
- Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC, Gulick AM (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* 529:235–238. <https://doi.org/10.1038/nature16163>
- Drake EJ, Nicolai DA, Gulick AM (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem Biol* 13:409–419. <https://doi.org/10.1016/j.chembiol.2006.02.005>
- Du L, He Y, Luo Y (2008) Crystal structure and enantiomer selection by *D*-alanyl carrier protein ligase DltA from *Bacillus cereus*. *Biochemistry* 47:11473–11480. <https://doi.org/10.1021/bi801363b>

26. Du L, Luo Y (2014) Thiolation-enhanced substrate recognition by D-alanyl carrier protein ligase DltA from *Bacillus cereus*. F1000Res 3:106. <https://doi.org/10.12688/f1000research.4097.1>
27. Du Y, Wang Y, Huang T, Tao M, Deng Z, Lin S (2014) Identification and characterization of the biosynthetic gene cluster of polyoxypeptin A, a potent apoptosis inducer. BMC Microbiol 14:30. <https://doi.org/10.1186/1471-2180-14-30>
28. Du YL, He HY, Higgins MA, Ryan KS (2017) A heme-dependent enzyme forms the nitrogen–nitrogen bond in piperazate. Nat Chem Biol 13:836–838. <https://doi.org/10.1038/nchembio.2411>
29. Eppelmann K, Stachelhaus T, Marahiel MA (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry 41:9718–9726
30. Fei X, Yin X, Zhang L, Zabriskie TM (2007) Roles of VioG and VioQ in the incorporation and modification of the capreomycin residue in the peptide antibiotic viomycin. J Nat Prod 70:618–622. <https://doi.org/10.1021/np060605u>
31. Felnagle EA, Rondon MR, Berti AD, Crosby HA, Thomas MG (2007) Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Appl Environ Microbiol 73:4162–4170. <https://doi.org/10.1128/AEM.00485-07>
32. Fujimori DG, Hrvatin S, Neumann CS, Strieker M, Marahiel MA, Walsh CT (2007) Cloning and characterization of the biosynthetic gene cluster for kutznerides. Proc Natl Acad Sci USA 104:16498–16503. <https://doi.org/10.1073/pnas.0708242104>
33. Galea CA, Han M, Zhu Y, Roberts K, Wang J, Thompson PE, Velkov T (2017) Characterization of the polymyxin D synthetase biosynthetic cluster and product profile of *Paenibacillus polymyxa* ATCC 10401. J Nat Prod 80:1264–1274. <https://doi.org/10.1021/acs.jnatprod.6b00807>
34. Galm U, Wendt-Pienkowski E, Wang L, Huang SX, Unsin C, Tao M, Coughlin JM, Shen B (2011) Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: bleomycin, tallysomycin, and zorbamycin. J Nat Prod 74:526–536. <https://doi.org/10.1021/np1008152>
35. Gatto GJ Jr, McLoughlin SM, Kelleher NL, Walsh CT (2005) Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme. Biochemistry 44:5993–6002. <https://doi.org/10.1021/bi050230w>
36. Gulick AM (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem Biol 4:811–827. <https://doi.org/10.1021/cb900156h>
37. Gulick AM (2016) Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. Curr Opin Chem Biol 35:89–96. <https://doi.org/10.1016/j.cbpa.2016.09.005>
38. Henderson JC, Fage CD, Cannon JR, Brodbelt JS, Keatinge-Clay AT, Trent MS (2014) Antimicrobial peptide resistance of *Vibrio cholerae* results from an LPS modification pathway related to nonribosomal peptide synthetases. ACS Chem Biol 9:2382–2392. <https://doi.org/10.1021/cb500438x>
39. Herbst DA, Boll B, Zocher G, Stehle T, Heide L (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylation enzymes. J Biol Chem 288:1991–2003. <https://doi.org/10.1074/jbc.M112.420182>
40. Hofer I, Crusemann M, Radzom M, Geers B, Flachshaar D, Cai X, Zeeck A, Piel J (2011) Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. Chem Biol 18:381–391. <https://doi.org/10.1016/j.chembiol.2010.12.018>
41. Huang T, Wang Y, Yin J, Du Y, Tao M, Xu J, Chen W, Lin S, Deng Z (2011) Identification and characterization of the pyridomycin biosynthetic gene cluster of *Streptomyces pyridomyces* NRRL B-2517. J Biol Chem 286:20648–20657. <https://doi.org/10.1074/jbc.M110.180000>
42. Izore T, Cryle MJ (2018) The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. Nat Prod Rep. <https://doi.org/10.1039/c8np00038g>
43. Jackson DR, Tu SS, Nguyen M, Barajas JF, Schaub AJ, Krug D, Pistorius D, Luo R, Müller R, Tsai SC (2016) Structural insights into anthranilate priming during type II polyketide biosynthesis. ACS Chem Biol 11:95–103. <https://doi.org/10.1021/acschembio.5b00500>
44. Janata J, Kamenik Z, Gazak R, Kadlcik S, Najmanova L (2018) Biosynthesis and incorporation of an alkylproline-derivative (APD) precursor into complex natural products. Nat Prod Rep 35:257–289. <https://doi.org/10.1039/c7np00047b>
45. Kadlcik S, Kucera T, Chalupska D, Gazak R, Koberska M, Ulanova D, Kopecky J, Kutejova E, Najmanova L, Janata J (2013) Adaptation of an L-proline adenylation domain to use 4-propyl-L-proline in the evolution of lincosamide biosynthesis. PLoS One 8:e84902. <https://doi.org/10.1371/journal.pone.0084902>
46. Kalb D, Lackner G, Hoffmeister D (2014) Functional and phylogenetic divergence of fungal adenylation-forming reductases. Appl Environ Microbiol 80:6175–6183. <https://doi.org/10.1128/AEM.01767-14>
47. Kalb D, Lackner G, Rappe M, Hoffmeister D (2015) Activity of α -aminoadipate reductase depends on the N-Terminally Extending domain. ChemBioChem 16:1426–1430. <https://doi.org/10.1002/cbic.201500190>
48. Kaljunen H, Schiefelbein SHH, Stummer D, Kozak S, Meijers R, Christiansen G, Rentmeister A (2015) Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. Angew Chem Int Ed Engl 54:8833–8836. <https://doi.org/10.1002/anie.201503275>
49. Kallow W, Neuhof T, Arezi B, Jungblut P, von Dohren H (1997) Penicillin biosynthesis: intermediates of biosynthesis of δ -L- α -aminoadipyl-L-cysteinyl-D-valine formed by ACV synthetase from *Acremonium chrysogenum*. FEBS Lett 414:74–78
50. Khayatt BI, Overmars L, Siezen RJ, Francke C (2013) Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. PLoS One 8:e62136. <https://doi.org/10.1371/journal.pone.0062136>
51. Knudsen M, Sondergaard D, Tofting-Olesen C, Hansen FT, Brodersen DE, Pedersen CN (2016) Computational discovery of specificity-conferring sites in non-ribosomal peptide synthetases. Bioinformatics 32:325–329. <https://doi.org/10.1093/bioinformatics/btv600>
52. Kries H, Wachtel R, Pabst A, Wanner B, Niquille D, Hilvert D (2014) Reprogramming nonribosomal peptide synthetases for “clickable” amino acids. Angew Chem Int Ed 53:10105–10108. <https://doi.org/10.1002/anie.201405281>
53. Kudo F, Kawamura K, Uchino A, Miyanaga A, Numakura M, Takayanagi R, Eguchi T (2015) Genome mining of the hitachimycin biosynthetic gene cluster: involvement of a phenylalanine-2,3-aminomutase in biosynthesis. ChemBioChem 16:909–914. <https://doi.org/10.1002/cbic.201500040>
54. Kudo F, Miyanaga A, Eguchi T (2014) Biosynthesis of natural products containing β -amino acids. Nat Prod Rep 31:1056–1073. <https://doi.org/10.1039/c4np00007b>
55. Lee JH, Evans BS, Li G, Kelleher NL, van der Donk WA (2009) In vitro characterization of a heterologously expressed nonribosomal Peptide synthetase involved in phosphinothricin

- tripeptide biosynthesis. *Biochemistry* 48:5054–5056. <https://doi.org/10.1021/bi900164d>
56. Lee TV, Johnson LJ, Johnson RD, Koulman A, Lane GA, Lott JS, Arcus VL (2010) Structure of a eukaryotic nonribosomal peptide synthetase adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J Biol Chem* 285:2415–2427. <https://doi.org/10.1074/jbc.M109.071324>
 57. Li J, Guo Z, Huang W, Meng X, Ai G, Tang G, Chen Y (2013) Mining of a streptothricin gene cluster from *Streptomyces* sp. TP-A0356 genome via heterologous expression. *Sci China Life Sci* 56:619–627. <https://doi.org/10.1007/s11427-013-4504-2>
 58. Li R, Oliver RA, Townsend CA (2017) Identification and characterization of the sulfazecin monobactam biosynthetic gene cluster. *Cell Chem Biol* 24:24–34. <https://doi.org/10.1016/j.chembiol.2016.11.010>
 59. Lukat P, Katsuyama Y, Wenzel S, Binz T, König C, Blankenfeldt W, Bronstrup M, Müller R (2017) Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem Sci* 8:7521–7527. <https://doi.org/10.1039/c7sc02622f>
 60. Ma J, Wang Z, Huang H, Luo M, Zuo D, Wang B, Sun A, Cheng YQ, Zhang C, Ju J (2011) Biosynthesis of himastatin: assembly line and characterization of three cytochrome P450 enzymes involved in the post-tailoring oxidative steps. *Angew Chem Int Ed Engl* 50:7797–7802. <https://doi.org/10.1002/anie.201102305>
 61. Magarvey NA, Fortin PD, Thomas PM, Kelleher NL, Walsh CT (2008) Gatekeeping versus promiscuity in the early stages of the andrimid biosynthetic assembly line. *ACS Chem Biol* 3:542–554. <https://doi.org/10.1021/cb800085g>
 62. Marahiel MA, Essen LO (2009) Chapter 13. Nonribosomal peptide synthetases mechanism and structural aspects of essential domains. *Methods Enzymol* 458:337–351. [https://doi.org/10.1016/s0076-6879\(09\)04813-7](https://doi.org/10.1016/s0076-6879(09)04813-7)
 63. Maruyama C, Toyoda J, Kato Y, Izumikawa M, Takagi M, Shinya K, Katano H, Utagawa T, Hamano Y (2012) A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat Chem Biol* 8:791–797. <https://doi.org/10.1038/nchembio.1040>
 64. Mast Y, Weber T, Gözl M, Ort-Winklbauer R, Gondran A, Wohlleben W, Schinko E (2011) Characterization of the ‘pristinamycin supercluster’ of streptomyces pristinaespiralis. *Microb Biotechnol* 4:192–206. <https://doi.org/10.1111/j.1751-7915.2010.00213.x>
 65. May JJ, Kessler N, Marahiel MA, Stubbs MT (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci USA* 99:12120–12125. <https://doi.org/10.1073/pnas.182156699>
 66. McMahon MD, Rush JS, Thomas MG (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol* 194:2809–2818. <https://doi.org/10.1128/JB.00088-12>
 67. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39:W339–346. <https://doi.org/10.1093/nar/gkr466>
 68. Meyer S, Kehr JC, Mainz A, Dehm D, Petras D, Süßmuth RD, Dittmann E (2016) Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem Biol* 23:462–471. <https://doi.org/10.1016/j.chembiol.2016.03.011>
 69. Meyer S, Mainz A, Kehr JC, Süßmuth RD, Dittmann E (2017) Prerequisites of isopeptide bond formation in microcystin biosynthesis. *ChemBioChem* 18:2376–2379. <https://doi.org/10.1002/cbic.201700389>
 70. Miao V, Coeffet-Legal MF, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* 151:1507–1523. <https://doi.org/10.1099/mic.0.27757-0>
 71. Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM (2016) Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J Biol Chem* 291:22559–22571. <https://doi.org/10.1074/jbc.M116.746297>
 72. Mitchell CA, Shi C, Aldrich CC, Gulick AM (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* 51:3252–3263. <https://doi.org/doi:10.1021/bi300112e>
 73. Miyana A, Cieślak J, Shinohara Y, Kudo F, Eguchi T (2014) The crystal structure of the adenylation enzyme VinN reveals a unique β -amino acid recognition mechanism. *J Biol Chem* 289:31448–31457. <https://doi.org/10.1074/jbc.M114.602326>
 74. Miyana A, Hayakawa Y, Numakura M, Hashimoto J, Teruya K, Hirano T, Shin-Ya K, Kudo F, Eguchi T (2016) Identification of the fluvirucin B2 (Sch 38518) Biosynthetic gene cluster from *Actinomadura fulva* subsp. indica ATCC 53714: substrate specificity of the β -amino acid selective adenylation enzyme FlvN. *Biosci Biotechnol Biochem* 80:935–941. <https://doi.org/10.1080/09168451.2015.1132155>
 75. Miyana A, Kudo F, Eguchi T (2016) Mechanisms of β -amino acid incorporation in polyketide macrolactam biosynthesis. *Curr Opin Chem Biol* 35:58–64. <https://doi.org/10.1016/j.cbpa.2016.08.030>
 76. Mori S, Shrestha SK, Fernandez J, Alvarez San Millan M, Garzan A, Al-Mestarihi AH, Lombo F, Garneau-Tsodikova S (2017) Activation and Loading of the starter unit during thio-coraline biosynthesis. *Biochemistry* 56:4457–4467. <https://doi.org/10.1021/acs.biochem.7b00661>
 77. Mori S, Pang AH, Lundy TA, Garzan A, Tsodikov OV, Garneau-Tsodikova S (2018) Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nature Chemical Biology* 14:428–430. <https://doi.org/10.1038/s41589-018-0014-7>
 78. Muliandi A, Katsuyama Y, Sone K, Izumikawa M, Moriya T, Hashimoto J, Kozono I, Takagi M, Shin-ya K, Ohnishi Y (2014) Biosynthesis of the 4-methylloxazoline-containing nonribosomal peptides, JBR-34 and -35, in *Streptomyces* sp. Sp080513GE-23. *Chem Biol* 21:923–934. <https://doi.org/10.1016/j.chembiol.2014.06.004>
 79. Neumann CS, Jiang W, Heemstra JR Jr, Gontang EA, Kolter R, Walsh CT (2012) Biosynthesis of piperazine acid via N^5 -hydroxyornithine in *Kutzneria* spp. 744. *ChemBioChem* 13:972–976. <https://doi.org/10.1002/cbic.201200054>
 80. Niquille DL, Hansen DA, Mori T, Fercher D, Kries H, Hilvert D (2018) Nonribosomal biosynthesis of backbone-modified peptides. *Nat Chem* 10:282–287. <https://doi.org/10.1038/nchem.2891>
 81. Pistorius D, Li Y, Mann S, Müller R (2011) Unprecedented anthranilate priming involving two enzymes of the acyl adenylation superfamily in aurachin biosynthesis. *J Am Chem Soc* 133:12362–12365. <https://doi.org/10.1021/ja203653w>
 82. Prieto C, Garcia-Estrada C, Lorenzana D, Martin JF (2012) NRPSp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics* 28:426–427. <https://doi.org/10.1093/bioinformatics/btr659>
 83. Qiao C, Wilson DJ, Bennett EM, Aldrich CC (2007) A mechanism-based aryl carrier protein/thiolation domain affinity probe.

- J Am Chem Soc 129:6350–6351. <https://doi.org/10.1021/ja069201e>
84. Qu X, Jiang N, Xu F, Shao L, Tang G, Wilkinson B, Liu W (2011) Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor. *Mol Biosyst* 7:852–861
 85. Rachid S, Krug D, Weissman KJ, Müller R (2007) Biosynthesis of (*R*)- β -tyrosine and its incorporation into the highly cytotoxic chondramides produced by *Chondromyces crocatus*. *J Biol Chem* 282:21810–21817. <https://doi.org/10.1074/jbc.M703439200>
 86. Reger AS, Wu R, Dunaway-Mariano D, Gulick AM (2008) Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* 47:8016–8025. <https://doi.org/10.1021/bi800696y>
 87. Reimer JM, Aloise MN, Harrison PM, Schmeing TM (2016) Synthetic cycle of the initiation module of a formylating non-ribosomal peptide synthetase. *Nature* 529:239–242. <https://doi.org/10.1038/nature16503>
 88. Robbel L, Marahiel MA (2010) Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J Biol Chem* 285:27501–27508. <https://doi.org/10.1074/jbc.R110.128181>
 89. Rui Z, Huang W, Xu F, Han M, Liu X, Lin S, Zhang W (2015) Sparsomycin biosynthesis highlights unusual module architecture and processing mechanism in non-ribosomal peptide synthetase. *ACS Chem Biol* 10:1765–1769. <https://doi.org/10.1021/acscchembio.5b00284>
 90. Scaglione A, Fullone MR, Montemiglio LC, Parisi G, Zamparelli C, Vallone B, Savino C, Grgurina I (2017) Structure of the adenylation domain Thr1 involved in the biosynthesis of 4-chloro-threonine in sp. OH-5093-protein flexibility and molecular bases of substrate specificity. *FEBS J* 284:2981–2999. <https://doi.org/10.1111/febs.14163>
 91. Schofield CJ, Baldwin JE, Byford MF, Clifton I, Hajdu J, Hensgens C, Roach P (1997) Proteins of the penicillin biosynthesis pathway. *Curr Opin Struct Biol* 7:857–864
 92. Schwartz D, Grammel N, Heinzelmann E, Keller U, Wohlleben W (2005) Phosphinothricin tripeptide synthetases in *Streptomyces viridochromogenes* Tu494. *Antimicrob Agents Chemother* 49:4598–4607. <https://doi.org/10.1128/AAC.49.11.4598-4607.2005>
 93. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster AL, Wyatt MA, Magarvey NA (2015) Genomes to natural products PRediction informatics for secondary metabolomes (PRISM). *Nucleic Acids Res* 43:9645–9662. <https://doi.org/10.1093/nar/gkv1012>
 94. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA (2017) PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* 45:W49–W54. <https://doi.org/10.1093/nar/gkx320>
 95. Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* 6:493–505. [https://doi.org/10.1016/S1074-5521\(99\)80082-9](https://doi.org/10.1016/S1074-5521(99)80082-9)
 96. Strieker M, Tanovic A, Marahiel MA (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* 20:234–240. <https://doi.org/10.1016/j.sbi.2010.01.009>
 97. Sundlov JA, Shi C, Wilson DJ, Aldrich CC, Gulick AM (2012) Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol* 19:188–198. <https://doi.org/10.1016/j.chembiol.2011.11.013>
 98. Süßmuth RD, Mainz A (2017) Nonribosomal peptide synthesis—principles and prospects. *Angew Chem Int Ed Engl* 56:3770–3821. <https://doi.org/10.1002/anie.201609079>
 99. Tan XF, Dai YN, Zhou K, Jiang YL, Ren YM, Chen Y, Zhou CZ (2015) Structure of the adenylation-peptidyl carrier protein didomain of the *Microcystis aeruginosa* microcystin synthetase McyG. *Acta Crystallogr D Biol Crystallogr* 71:873–881. <https://doi.org/10.1107/S1399004715001716>
 100. Tang GL, Cheng YQ, Shen B (2007) Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from *Streptomyces atroolivaceus* S-140. Discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly load *D*-alanine. *J Biol Chem* 282:20273–20282. <https://doi.org/10.1074/jbc.M702814200>
 101. Tanovic A, Samel SA, Essen LO, Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* 321:659–663. <https://doi.org/10.1126/science.1159850>
 102. Tarry MJ, Haque AS, Bui KH, Schmeing TM (2017) X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* 25:783–793. <https://doi.org/10.1016/j.str.2017.03.014>
 103. Thibaut D, Bisch D, Ratet N, Maton L, Couder M, Debussche L, Blanche F (1997) Purification of peptide synthetases involved in pristinamycin I biosynthesis. *J Bacteriol* 179:697–704
 104. Thirlway J, Lewis R, Nunns L, Al Nakeeb M, Styles M, Struck AW, Smith CP, Micklefield J (2012) Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew Chem Int Ed Engl* 51:7181–7184. <https://doi.org/10.1002/anie.201202043>
 105. Ueki M, Galonić DP, Vaillancourt FH, Garneau-Tsodikova S, Yeh E, Vosburg DA, Schroeder FC, Osada H, Walsh CT (2006) Enzymatic generation of the antimetabolite γ , γ -dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a *Streptomyces* sp. *Chem Biol* 13:1183–119. <https://doi.org/10.1016/j.chembiol.2006.09.012>
 106. Van Lanen SG, Lin S, Dorrestein PC, Kelleher NL, Shen B (2006) Substrate specificity of the adenylation enzyme SgcC1 involved in the biosynthesis of the enediyne antitumor antibiotic C-1027. *J Biol Chem* 281:29633–29640. <https://doi.org/10.1074/jbc.M605887200>
 107. Vobruba S, Kadlcik S, Gazak R, Janata J (2017) Evolution-guided adaptation of an adenylation domain substrate specificity to an unusual amino acid. *PLoS One* 12:e0189684. <https://doi.org/10.1371/journal.pone.0189684>
 108. von Dohren H, Dieckmann R, Pavela-Vrancic M (1999) The non-ribosomal code. *Chem Biol* 6:R273–279
 109. Walsh CT, O'Brien RV, Khosla C (2013) Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl* 52:7098–7124. <https://doi.org/10.1002/anie.201208344>
 110. Watanabe K, Hotta K, Nakaya M, Praseuth AP, Wang CC, Inada D, Takahashi K, Fukushi E, Oguri H, Oikawa H (2009) *Escherichia coli* allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product. *J Am Chem Soc* 131:9347–9353. <https://doi.org/10.1021/ja902261a>
 111. Winn M, Fyans JK, Zhuo Y, Micklefield J (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat Prod Rep* 33:317–347. <https://doi.org/10.1039/c5np00099h>
 112. Wuest WM, Krahn D, Kaiser M, Walsh CT (2011) Enzymatic timing and tailoring of macrolactamization in syringolin biosynthesis. *Org Lett* 13:4518–4521. <https://doi.org/10.1021/ol2016687>
 113. Yin X, Zabriskie TM (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* 152:2969–2983. <https://doi.org/10.1099/mic.0.29043-0>
 114. Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT (2008) Crystal structure of DltA. Implications for

- the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J Biol Chem* 283:32484–32491. <https://doi.org/10.1074/jbc.M800557200>
115. Zhao H, Wang L, Wan D, Qi J, Gong R, Deng Z, Chen W (2016) Characterization of the aurantimycin biosynthetic gene cluster and enhancing its production by manipulating two pathway-specific activators in *Streptomyces aurantiacus* JA 4570. *Microb Cell Fact* 15:160. <https://doi.org/10.1186/s12934-016-0559-7>