



**NPR**

**Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis**

Journal:	<i>Natural Product Reports</i>
Manuscript ID:	NP-REV-09-2014-000120.R1
Article Type:	Highlight
Date Submitted by the Author:	12-Jan-2015
Complete List of Authors:	Labby, Kristin; Beloit College, Chemistry Watsula, Stoyan; University of Michigan, Medicinal Chemistry Garneau-Tsodikova, Sylvie; University of Kentucky, Pharmaceutical Sciences

SCHOLARONE™  
Manuscripts

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

**HIGHLIGHT**

# Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis

Kristin J. Labby,<sup>a</sup> Stoyan G. Watsula,<sup>b</sup> and Sylvie Garneau-Tsodikova<sup>\*c</sup>

Received (in XXX, XXX) Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

Covering up to 2014

Nonribosomal peptides (NRPs) account for a large portion of drugs and drug leads currently available in the pharmaceutical industry. They are one of two main families of natural products biosynthesized on megaenzyme assembly-lines composed of multiple modules that are, in general, each comprised of three core domains and on occasion of accompanying auxiliary domains. The core adenylation (A) domains are known to delineate the identity of the specific chemical components to be incorporated into the growing NRPs. Previously believed to be inactive, A domains interrupted by auxiliary enzymes have recently been proven to be active and capable of performing two distinct chemical reactions. This highlight summarizes current knowledge on A domains and presents the various interrupted A domains found in a number of nonribosomal peptide synthetase (NRPS) assembly-lines, their predicted or proven dual functions, and their potential for manipulation and engineering for chemoenzymatic synthesis of new pharmaceutical agents with increased potency.

1	<b>Introduction</b>
2	<b>Adenylation domain core signature sequences and structures</b>
3	<b>Interrupted adenylation domains</b>
3.1	<b>Overview of interrupted adenylation domains</b>
3.2	<b>Adenylation domains with methyltransferase activity</b>
3.3	<b>Adenylation domains with ketoreductase activity</b>
3.4	<b>Adenylation domains with oxidase or monooxygenase activity</b>
4	<b>Perspectives</b>
5	<b>Acknowledgements</b>
6	<b>Notes and references</b>

## 1. Introduction

Nonribosomal peptides (NRPs) are natural products that exhibit a wide range of pharmaceutically relevant biological activities, including anti-cancer, anti-fungal, anti-malarial, anti-microbial, and anti-viral functions. NRPs are biosynthesized on nonribosomal peptide synthetase (NRPS) assembly-lines, which are divided into multiple modules that are each generally composed of at least three core domains: a condensation (C), an adenylation (A), and a thiolation (T) domain (in the literature also referred to as peptidyl carrier protein (abbreviated as PCP, PC, or P) or carrier protein (CP)).

The A domain plays a critical role in dictating the structure of

the NRP to be formed as it is responsible for the selection and activation of the proteinogenic or non-proteinogenic amino acid building block<sup>1</sup> to be incorporated into the growing peptide chain prior to covalent attachment onto the 4'-phosphopantetheinyl (Ppant) arm of the active (holo) downstream T domain partner. The C domains, normally located between consecutive pairs of A and T domains, serve to link the upstream and downstream building blocks tethered to the T domains that surround them. A thioesterase (TE) domain, located at the C-terminus of the module acting last during the natural product biosynthesis normally releases the peptidic chain from the NRPS assembly-line.

In addition to the three core domains, in order to add functional groups and structural diversity to the natural products to be biosynthesized, NRPS modules may contain strategically embedded auxiliary domains, such as cyclization (Cy), dehydratase (DH), epimerase (E), formylase (F), ketoreductase (KR), methyltransferase (M), monooxygenase (MOx), oxidase (Ox), and reductase (R) domains.

More recently, A domains interrupted by portions of auxiliary domains including M, KR, Ox, and MOx have been observed in the NRPSs responsible for the production of thiocoraline,<sup>2</sup> kutznerides,<sup>3</sup> pyochelin,<sup>4</sup> yersiniabactin,<sup>5</sup> micacocidin,<sup>6</sup> enniatins,<sup>7</sup> microcystin,<sup>8</sup> micropeptin,<sup>9</sup> tubulysin,<sup>10</sup> cereulide,<sup>11,12</sup> valinomycin,<sup>13</sup> myxothiazol,<sup>14</sup> melithiazole,<sup>15</sup> and indigoidine.<sup>16</sup> Originally thought to be inactive as the auxiliary domains are usually not found to interrupt A domains or other NRPS domains, some of these interrupted A domains have now been found to take an active role in the biosynthesis of a variety of natural products, including kutznerides,<sup>17</sup> thiocoraline, and cereulide.<sup>18</sup>

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

**HIGHLIGHT****Table 1.** Highly conserved core motifs of A, M, KR, Ox and MOx domains of NRPSs.

Domain	Core(s)	Consensus sequence	Role
<b>Adenylation</b>	a1	L(T/S)YxEL	Structural. At N-terminus of domain, caps an $\alpha$ -helix. <sup>19</sup>
	a2	LKAGxAYL(V/L)P(L/I)D	Structural. Properly aligns Gly78. <sup>19</sup>
	a3	LAYxxYTSG(S/T)TGxPKG	Substrate binding. Acts as a loop and positions the $\beta,\gamma$ -phosphates correctly. <sup>19</sup>
	a4	$\Phi$ DxS	Substrate binding. Aromatic residue terminates an $\alpha$ -helix that forms side of acyl-binding pocket. Aromatic residue also switches conformation between two enzyme states. <sup>19</sup>
	a5	NxYGPTE	Structural and substrate binding. Invariant glutamic acid coordinates $Mg^{2+}$ ion. Adenine ring of ATP is stacked against aromatic residue. <sup>19</sup>
	a6	GELxIGx(V/L)ARGYL	Structural. Stabilizes distorted $\beta$ -sheets in the N-terminal domain. <sup>19</sup>
	a7	Y(R/K)TGDL	Substrate binding and catalytic. Aspartic acid residue is 100% conserved and hydrogen bonds with ATP through the ribose hydroxyls. <sup>19</sup>
	a8	GRxDxxxKxxGxxxELxxxE	Structural and substrate binding. Arginine stabilizes the ribose through its hydroxyls. A hinge is formed at aspartic acid residue (or occasionally lysine). In the thioester-forming conformation, the glycine forms part of the pantetheine tunnel. <sup>19</sup>
	a9	(L/V)Px $\Phi$ M(L/V/I)P	Catalytic. Stabilizes thioester-forming conformation by properly positioning residues to interact with T domain. <sup>20</sup>
	a10	NGK(V/L)DR	Catalytic. In the adenylate-forming conformation, lysine is within the active site. <sup>19</sup>
<b>Methylation</b>	M1	VL(D/E)GxGxG	S-adenosylmethionine (SAM) binding. Methyltransferase: catalyze N- or C-methylation of amino acid residues.
	M2	NELsYRYxAV	
	M3	VExSxARQxGxLD	
<b>Oxidase</b>	Ox1	KYxYxSxGxxY(P/G)VQ	
	Ox2	GxxxG(L/V)xxGxYYY(H/D) P	
	Ox3	IxxxYG	
<b>Ketoreductase</b>	KR	GGxGxxGxxxA	21
		YxxxN	21
<b>Monoxygenase</b>	MOx1	GFxxxxxxExH	14
	MOx2	VxPxxxPxxxEExxxxDxxxx GRxxxxxxG	14

$\Phi$  designates an aromatic amino acid residue and x is any residue.

In this highlight, we present an overview of these interrupted A domains as well as what is known about the structures and core recognition sequences of A domains in general.

## 2. Adenylation domain core signature sequences and structures

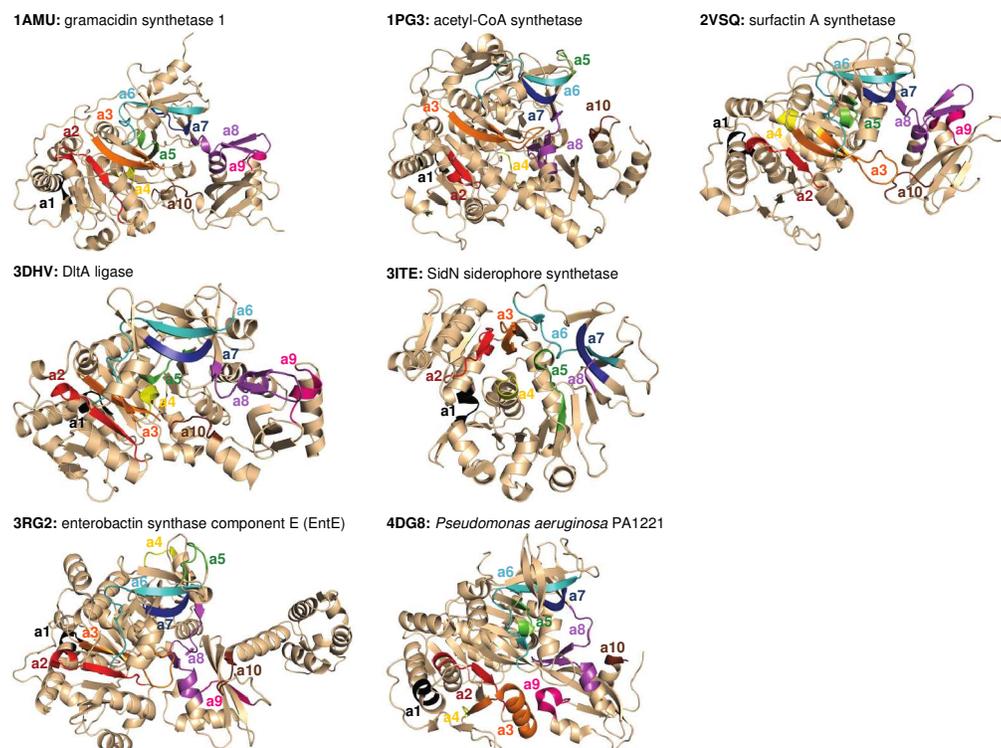
Adenylation domains determine the selection of the next amino acid residue to be incorporated along the NRPS assembly-line. A domains can be identified by ten conserved core signature

sequence motifs, a1-a10 (Table 1).<sup>22</sup> These highly conserved core motifs serve a variety of functions, including structural, substrate binding, and catalytic roles.

While the ten signature sequence motifs are highly conserved across A domains, their structures are diverse (Fig. 1). A domains contain approximately 550 amino acid residues. The crystal structures reveal that A domains typically consist of a large N-terminal domain containing conserved signature sequences a1-a7 and a smaller C-terminal domain consisting of a8-a10. The active site, where ATP binds, is located between these two subunits.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

**HIGHLIGHT**

**Fig. 1 A.** Structures of adenylation domains from gramicidin synthetase 1 (1AMU)<sup>23</sup> and SidN siderophore synthetase (3ITE).<sup>24</sup> Crystal structures are also shown of other adenylate-forming enzymes: acetyl-CoA synthetase (1PG3),<sup>25</sup> surfactin A synthetase C (2VSQ),<sup>26</sup> D-alanine-poly(phosphoribitol) (DltA) ligase (3DHV),<sup>27</sup> enterobactin synthase component E (EntE) (3RG2),<sup>28</sup> and *Pseudomonas aeruginosa* PA1221 (4DG8).<sup>29</sup> **Notes:** (i) The 10 conserved signature sequences of A domains with the a1-a10 core signature sequences highlighted in black (a1), red (a2), orange (a3), yellow (a4), green (a5), aqua (a6), dark blue (a7), purple (a8), magenta (a9), and brown (a10), respectively. (ii) The sequence of 1PG3 does not contain an a9 motif; the crystal structure of 3ITE does not include a9 or a10.

In recent years, a class of small (approximately 70 amino acid residues) proteins called Mbth-like proteins found to be encoded within many NRPS gene clusters have been studied in detail.<sup>30,31</sup> These small proteins are reported to have strong interactions with their corresponding A domains and are suspected to play a role in increasing binding affinity between the A domain and its amino acid substrate. Mbth-like proteins are not necessary for all A domains expression and/or activity.<sup>32</sup> They seem to have high specificity for their particular A domain partners, but some examples of weak complementarity are reported.<sup>33</sup> The roles of Mbth-like proteins have been explored in a variety of biochemical and structural studies,<sup>17,33-46</sup> but their exact functions and mechanisms of action still remain elusive.

### 3. Interrupted adenylation domains

#### 3.1 Overview of interrupted adenylation domains

NRPs obtain their unique structures in part from the diversity of the hundreds of viable substrates incorporated by A domains,<sup>47</sup> and also from the incorporation of auxiliary domains into specific modules of NRPS assembly-lines. As additional means to incorporate structural diversity, Nature has evolved interrupted A

domains, in which an additional domain, or part of a domain, is inserted between two core sequences, most commonly a8 and a9. In this section, we cover fourteen natural product scaffolds made by NRPSs containing interrupted A domains: thiocoraline, kutznerides, pyochelin, yersiniabactin, micacocidin A, enniatins, micropeptins, microcystins, tubulysins, cereulide, valinomycin, myxothiazols, melithiazoles, and indigoidine. These interrupted A domains are interrupted by methyltransferase (M), ketoreductase (KR), oxidase (Ox), and monooxygenase (MOx) domains.

#### 3.2 Adenylation domains with methyltransferase activity

The NRPs thiocoraline, kutznerides, pyochelin, yersiniabactin, micacocidin A, enniatins, micropeptins, microcystins, and tubulysins are biosynthesized by NRPSs that contain A domains interrupted with M domains (Fig. 2 and Fig. 3). These M domains are capable of *S*-, *N*-, or *O*-methylation (Fig. 4). Typically *O*-methyltransferases exist as stand-alone domains, as exemplified in type I polyketide biosynthesis. However, in NRPSs, M domains are normally embedded in modules of the assembly-lines as intact full-length domains.<sup>48</sup> We note that other A domains interrupted by methyltransferases exist; some are

functional, some are not functional.<sup>48</sup> In the unique cases presented in this section, portions of M domains are found embedded between different consensus sequences of A domains, either between a2 and a3 or between a8 and a9 (Table 1 and Figs. 2 and 3).

The NRP thiocoraline is a bisintercalator with a twofold symmetric bicyclic scaffold that contains two rare *S*-methylated L-Cys moieties (Fig. 2).<sup>49,50</sup> Originally, TioN, which consists of an A domain interrupted between motifs a2 and a3 by part of an M domain containing an *S*-adenosylmethionine (SAM)-binding sequence, was hypothesized to be non-functional due to this disruption.<sup>2</sup> Upon further examination, however, and because of its A site specificity for L-Cys (DLYDLSLV), TioN was proposed to be responsible for side-chain *S*-methylation of this residue just after its incorporation during biosynthesis (Fig. 4A).<sup>51</sup> Recently, it was reported that TioN, in the presence of its MbtH-like partner TioT, is capable of both adenylation of L-Cys followed by its *S*-methylation.<sup>32</sup> This means of *S*-methylation is unique in NRPS biosynthesis and moreover, the position of interruption, between a2 and a3 is unprecedented. Furthermore, the TioS thiocoraline biosynthesis module includes two A domains interrupted by *N*-methyltransferase domains that methylate backbone amine nitrogens.<sup>52</sup> This more common type of interrupted A domain will briefly be discussed later.

Kutznerides are cyclic depsipeptides isolated from soil actinomycetes and have antifungal and antimicrobial activity.<sup>53</sup> Kutznerides feature an *O*-methylated L-Ser residue unprecedentedly incorporated by the interrupted A<sub>4</sub> domain of KtzH (Figs. 3 and 4A).<sup>17</sup> KtzH contains four A domains, the second of which is interrupted between a8 and a9 by a SAM-binding containing portion of an M domain. The importance of the MbtH-like protein partner KtzJ for expression of KtzH(A<sub>4</sub>MA<sub>4</sub>T<sub>4</sub>) was recently reported and the order of adenylation of L-Ser by the A domain portion of KtzH(A<sub>4</sub>MA<sub>4</sub>T<sub>4</sub>) followed by loading onto the T<sub>4</sub> domain and further *O*-methylation by the M portion of KtzH(A<sub>4</sub>MA<sub>4</sub>T<sub>4</sub>) established.<sup>17</sup>

Similarly to thiocoraline and kutznerides, the biosynthesis of the NRP pyochelin involves an A domain interrupted by a SAM-binding portion of an M domain, but this M domain has *N*-methyltransferase activity.<sup>4</sup> Pyochelin is a siderophore produced by *Pseudomonas aeruginosa* during iron starvation and contributes to the virulence of *P. aeruginosa* in animals.<sup>54</sup> The PchF subunit of the four-protein, 11-domain pyochelin NRPS assembly-line contains a SAM-binding M domain located within A<sub>2</sub> between a8 and a9 (Fig. 3). L-Cys is incorporated by PchF, cyclized by the C<sub>2</sub> domain, and then, after AMP activation by the A<sub>2</sub> domain and reduction from thiazoline to thiazolidine by PchG, the M domain is responsible for its *N*-methylation. Mutation of a key residue within the M domain signature sequence, Gly1167, to arginine eliminated *N*-methyltransferase activity and prevented subsequent release by TE from PchF.<sup>4</sup> Interestingly, PchE was also found to contain a similar M domain insertion between a8 and a9 of A<sub>1</sub>, but the conserved signature motif for SAM-binding naturally contains an arginine in place of a key glycine residue (Gly1130), much like the aforementioned PchG G1167R mutant. Furthermore these genes are similar to those involved in the biosynthesis of structurally similar siderophores yersiniabactin and micacocidin A (Fig. 2); *Yersenia pestis* HMWP2, which is

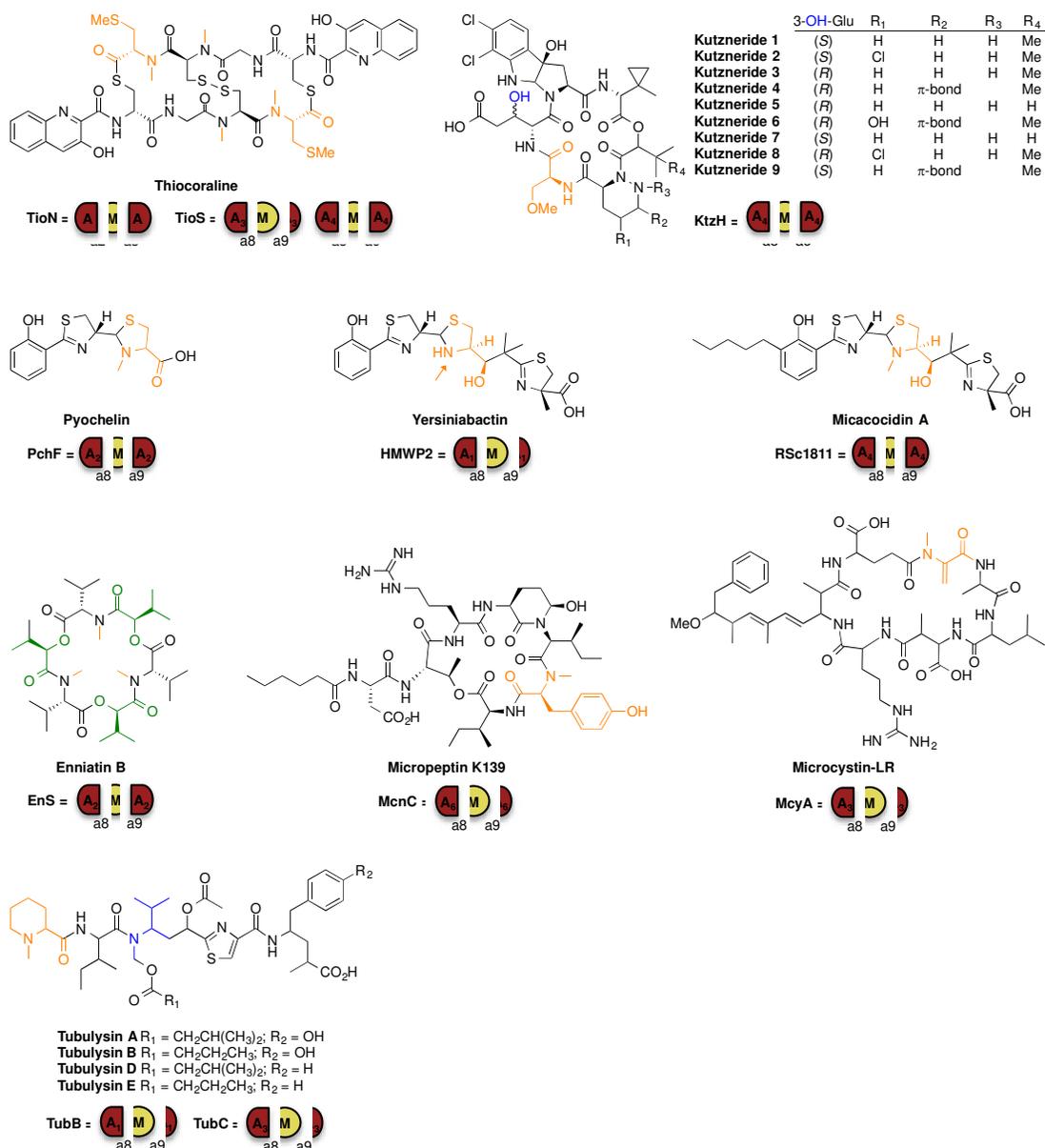
involved in the biosynthesis of yersiniabactin contains an M domain between a8 and a9 of its A<sub>1</sub> domain which also bears the Gly to Arg mutation within the SAM-binding motif. As expected, this mutation prevents methylation of the yersiniabactin thiazolidine nitrogen (indicated by an orange arrow in Fig. 2), which corresponds to the methylated nitrogen in pyochelin. RSc1806, the HMWP2 homologue found in *Ralstonia solanacearum* GM11000 and involved in micacocidin A biosynthesis, also contains a non-functional M domain inserted between a8 and a9 of its A<sub>2</sub> domain.<sup>6</sup> Interestingly, another M domain is inserted into a different protein in the micacocidin A NRPS gene cluster, RSc1811 (homologous to the latter half of HMWP1, Fig. 3). This M domain, located between a8 and a9 of A<sub>4</sub>, does not bear the Gly to Arg mutation, so is most likely responsible for the methylation of the L-Cys residue added by A<sub>4</sub>, though this NRPS has yet to be thoroughly characterized.

The most common interrupted A domains identified are those involved in *N*-methylation of peptide backbones.<sup>48</sup> While this is not an exhaustive list, examples other than those discussed herein include the cyclosporines,<sup>55</sup> actinomycin,<sup>56</sup> PF1022A,<sup>57</sup> anabaenopeptilide,<sup>58</sup> complestatin,<sup>59</sup> pristinamycin,<sup>60</sup> thaxtomin,<sup>61</sup> and barbamide.<sup>62</sup> These *N*-methylated NRPs all use SAM as a cosubstrate and are interrupted between a8 and a9. Examining the crystal structures of A domains, this position would be accommodating to an additional moiety (Fig. 1). The structure of enniatin B is shown (Fig. 2) as a representative NRP structure with *N*-methylation of its backbone. Enniatins are depsipeptides (where one or more amide bonds are replaced by ester bonds) produced by NRPS gene clusters from the fungal genus *Fusarium* (Fig. 3). Enniatin structures consist of six alternating residues of D-hydroxyisovaleric acid (D-HIV) and *N*-methylated amino acids linked by amide and ester bonds.<sup>7</sup>

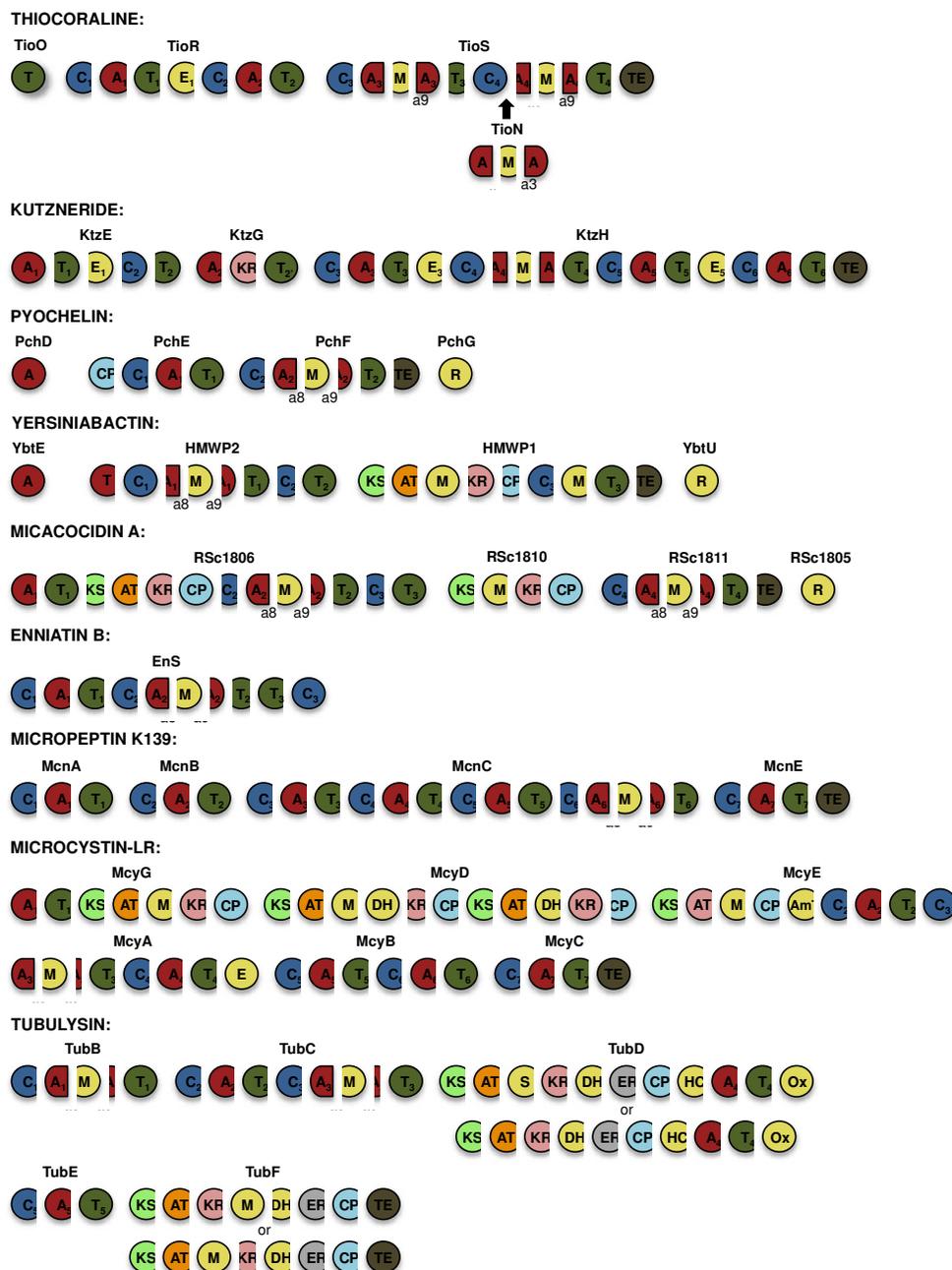
Another family of NRP compounds featuring backbone *N*-methylation are the cyclic heptapeptides microcystins from the *Microcystis* cyanobacterial genus (Fig. 4B).<sup>9</sup> Over 80 microcystins have been identified, including microcystin-LR (Fig. 2). Microcystins feature a mix of natural and unnatural amino acids; the most common scaffold is cyclic Adda-D-Glu-Mdha-D-Ala-L-X-D-MeAsp-L-Y (in which Adda is 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid, D-MeAsp is 3-methylaspartic acid, Mdha is *N*-methyldehydroalanine, and X and Y are varying amino acids). Installation of these *N*-methyl groups also occurs by M domains embedded between a8 and a9 of A domains. Similar in structure to microcystins are the cyanobacterial depsipeptides micropeptins. The micropeptin NRPS from *M. aeruginosa* K-139, which produces micropeptin K139, contains an M domain within the A<sub>6</sub> domain (Fig. 3). A<sub>6</sub> incorporates Tyr and the embedded M domain *N*-methylates it (Fig. 2). Micropeptin biosynthesis gene clusters from many other *M. aeruginosa* strains (e.g., NIES-843, NIVA-CYA 172/5, and PCC 7806) also contain M domains embedded between a8 and a9 of an A domain that are responsible for peptide backbone *N*-methylation.<sup>9</sup> Nodularin is another cyanobacterial peptide that is structurally similar to microcystins and micropeptins, but nodularin is composed of only five amino acids in the peptide ring. It too contains *N*-methyl moieties added by an M domain spliced into an A domain between a8 and a9.<sup>63</sup> Microcystins, micropeptins, and nodularin are potent inhibitors of

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

**HIGHLIGHT**

**Fig. 2** Structures of natural products with adenylation (A) domains interrupted by methyltransferase (M) domains in their biosynthetic NRPS machinery. Regions of the structure synthesized by the interrupted A domains are highlighted in color. See Fig. 3 for a definition of domain abbreviations.



**Fig. 3** Structural organization of thiocoraline, kutzneride, pyochelin, yersiniabactin, micacocidin, enniatin B, micropeptin K139, microcystin-LR, and tubulysin NRPSs. Each domain of the NRPSs is depicted in a different color. Abbreviations used for each domain: A, adenylation; C, condensation; T, thiolation (T domains have also been termed peptidyl carrier protein [PCP, PC, or P] and carrier protein [CP] in the literature); KR, ketoreductase; M, methyltransferase; MOx, monooxygenase; Ox, oxidase; TE, thioesterase; KS,  $\beta$ -ketoacyl synthase; AT, acyltransferase; AmT, aminotransferase; E, epimerization; DH,  $\beta$ -hydroxy-dehydratase; ER, enoyl reductase; HC, heterocyclization; CP, acyl carrier protein; Cy, cyclization; R, reductase; S, spacer.

eukaryotic serine/threonine protein phosphatases.<sup>8</sup> These cyanobacterial cyclic peptides are hepatotoxic and are a significant threat to human and animal health because they contaminate the water when they are produced during cyanobacteria (blue-green algae) blooms.

Tubulysins (Fig. 2) are produced by an NRPS that contains two interrupted A domains, both interrupted between a8 and a9 with M domains (Fig. 3).<sup>10</sup> The first, located within TubB, is responsible for the incorporation and subsequent *N*-methylation of piperolic acid. The second interrupted A domain, located within TubC, incorporates valine, which is also subsequently *N*-

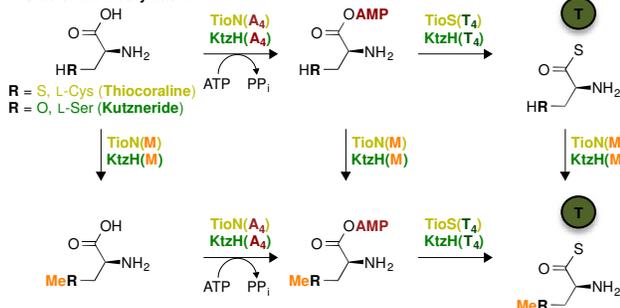
methyated. It remains unclear how this methyl group is further oxidized and acylated as observed in the final product.<sup>64</sup>

### 3.3 Adenylation domains with ketoreductase activity

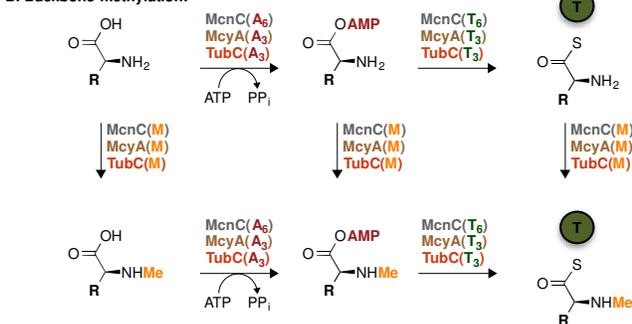
Nature incorporates functionalities into NRPSs using various mechanisms; for example aforementioned enniatin NRPS incorporates D-HIV directly *via* an A domain coding specifically for D-HIV, while cereulide incorporates D-HIV *via* an A domain that codes for  $\alpha$ -ketoisovaleric acid, which is then modified by the interrupting KR domain (Fig. 5A). Cereulide is an emetic toxin from *Bacillus cereus*, and is structurally similar to

valinomycin from *Streptomyces* spp. Both cereulide and valinomycin are macrocyclic potassium ionophores. They both are cyclododecapeptides consisting of alternating  $\alpha$ -keto acids (six) and amino acids (six). The NRPSs of cereulide and valinomycin are also very similar to one another (Fig. 6).<sup>18</sup> Both contain two A domains interrupted by KR domains between a8 and a9. These KR domains are responsible for the reduction of the  $\alpha$ -ketone to a hydroxyl group, which then forms an ester bond to the next residue incorporated (Fig. 7).

#### A. Side-chain methylation:



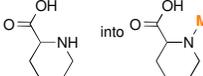
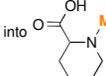
#### B. Backbone methylation:



R = CH<sub>2</sub>OH, L-Ser (Microcystin)

R = H<sub>2</sub>C-C<sub>6</sub>H<sub>4</sub>-OH, L-Tyr (Micropeptin)

R = CH(CH<sub>3</sub>)<sub>2</sub>, L-Val (Tubulysin)

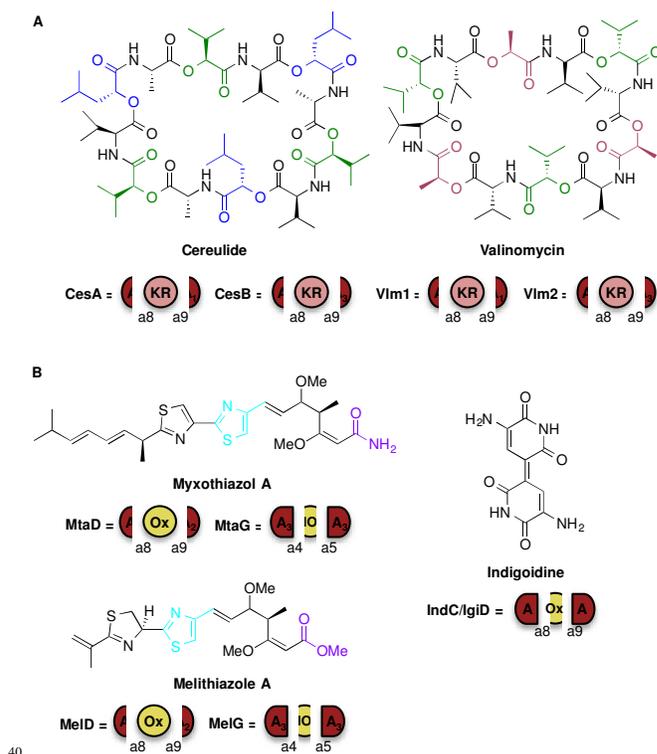
**Note:** In Tubulysin biosynthesis, TubB(M) similarly converts  into 

**Fig. 4** Representative examples of potential proposed pathways for interrupted A domain with methyltransferase activity during natural product biosynthesis. **A.** Side-chain methylation produces thiocoraline and kutzneride. **B.** Backbone *N*-methylation produces micropeptin, microcystin, and tubulysin.

### 3.4 Adenylation domains with oxidase or monooxygenase activity

The structurally similar myxobacterial secondary metabolites myxothiazols and melithiazoles are both biosynthesized from combined polyketide synthases (PKSs)/NRPSs containing two interrupted A domains (Fig. 5B).<sup>14,15</sup> The first interrupted A domains, the A<sub>2</sub> domains of MtaD or MelD are interrupted between a8 and a9 by an Ox domain (Fig. 6). They incorporate an L-Cys residue, then cyclize and oxidize it, forming the resultant thiazole ring (Fig. 8A).<sup>65</sup> The second interrupted A domains, A<sub>3</sub> located within MtaG and MelG, are interrupted between a4 and a5 by an MOx domain. The location of the MOx domain, inserted between a4 and a5, is very unusual and these are the only interrupted A domains spliced at this position known to date. These A domains add glycine, which is then hydroxylated at the

$\alpha$ -position by the MOx (Fig. 8B). The A<sub>3</sub> and C<sub>3</sub> domains of MtaG/MelG act before the MOx of the interrupted A domain (Fig. 8B).<sup>15</sup> Biosynthesis of melithiazole A differs from myxothiazol A in the identity of the starter unit (isobutyrate for Mel) added to MtaB/MelB (Fig. 5) as well as in the formation of the terminal amide (myxothiazol A) versus the terminal methyl ester (melithiazole A) (Fig. 8B).<sup>15</sup> Additional genes MelK and MelJ are responsible for melithiazole terminal methyl ester formation (Fig. 8B).<sup>66</sup>



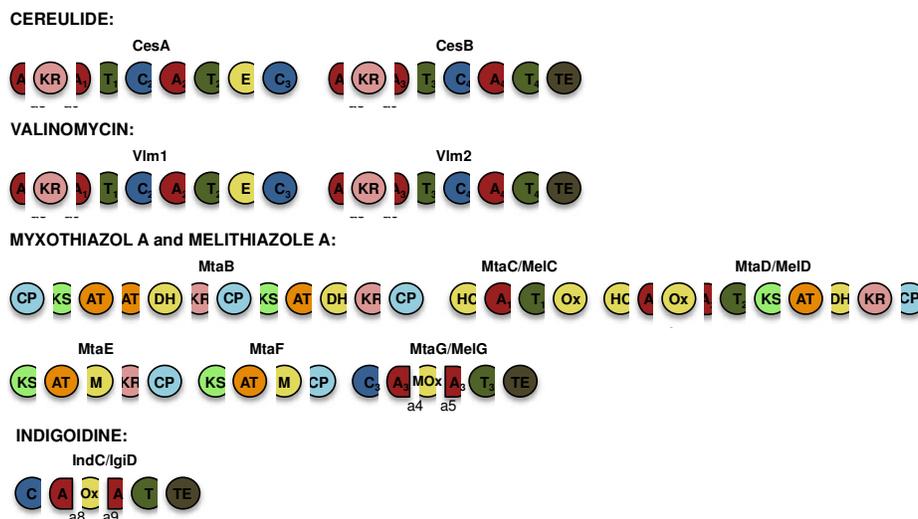
**Fig. 5** Structures of natural products with A domains interrupted by A, KR or B. Ox and MOx domains in their biosynthetic NRPS machinery. When applicable, regions of the structure synthesized by the interrupted A domains are highlighted in color. See Fig. 3 for a definition of domain abbreviations.

Indigoidine (Fig. 5B) is a blue pigment produced by the plant pathogenic bacteria *Erwinia chrysanthemi*<sup>67</sup> as well as by other bacteria including certain *Photorhabdus*,<sup>68</sup> *Phaeobacter*,<sup>69</sup> and *Streptomyces*<sup>70</sup> species. Indigoidine is composed of two molecules of 5-amino-3*H*-pyridine-2,6-one, which are produced by the small NRPS IndC (Fig. 6, Fig. 8C). (*Note:* What is termed IndC in *E. chrysanthemi* is homologous to “IgiD” in *Phaeobacter*). The IndC A domain codes for L-glutamine as a substrate. The IndC A domain is interrupted between a8 and a9 by a flavin-dependent Ox domain that installs a carbon-carbon double bond either immediately after loading the substrate onto the T domain (Fig. 8C), or perhaps after intramolecular cyclization and release.<sup>71</sup> The final dimerization step is proposed to be non-enzymatic and yields the conjugated product indigoidine, which is thought to be produced by these pathogenic bacterial species as a defence mechanism against reactive oxygen species.

Cite this: DOI: 10.1039/c0xx00000x

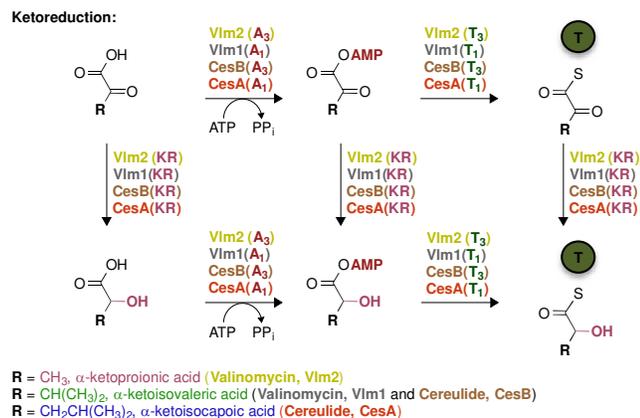
www.rsc.org/xxxxxx

HIGHLIGHT



**Fig. 6** Structural organization of cereulide, valinomycin, myxothiazol A, melithiazole A, and indigoidine NRPSs. Each module of the NRPSs is depicted in a different color. See Fig. 3 for a definition of domain abbreviations.

5



**Fig. 7** Proposed potential pathways for interrupted adenylation domain with ketoreductase activity during valinomycin and cereulide biosynthesis.

## 10 4. Perspectives

Because of their modular architectures, NRPSs provide a platform for engineering the biosynthesis of NRPs beyond those found to exist in Nature.<sup>72,73</sup> The most common and most successful method of engineering NRPSs is domain swapping,<sup>15</sup> often termed combinatorial biosynthesis.<sup>74,75</sup> Variants of the antibiotic daptomycin have been produced by NRPSs containing exchanged domains.<sup>76</sup> Additionally, modification of A domain active sites can be used to incorporate alternative amino acid substrates, including unnatural amino acids.<sup>77-80</sup> Alteration of A domain specificity can lead to “tailor made enzyme systems”.<sup>20</sup> The existence of A domains interrupted with auxiliary enzymes

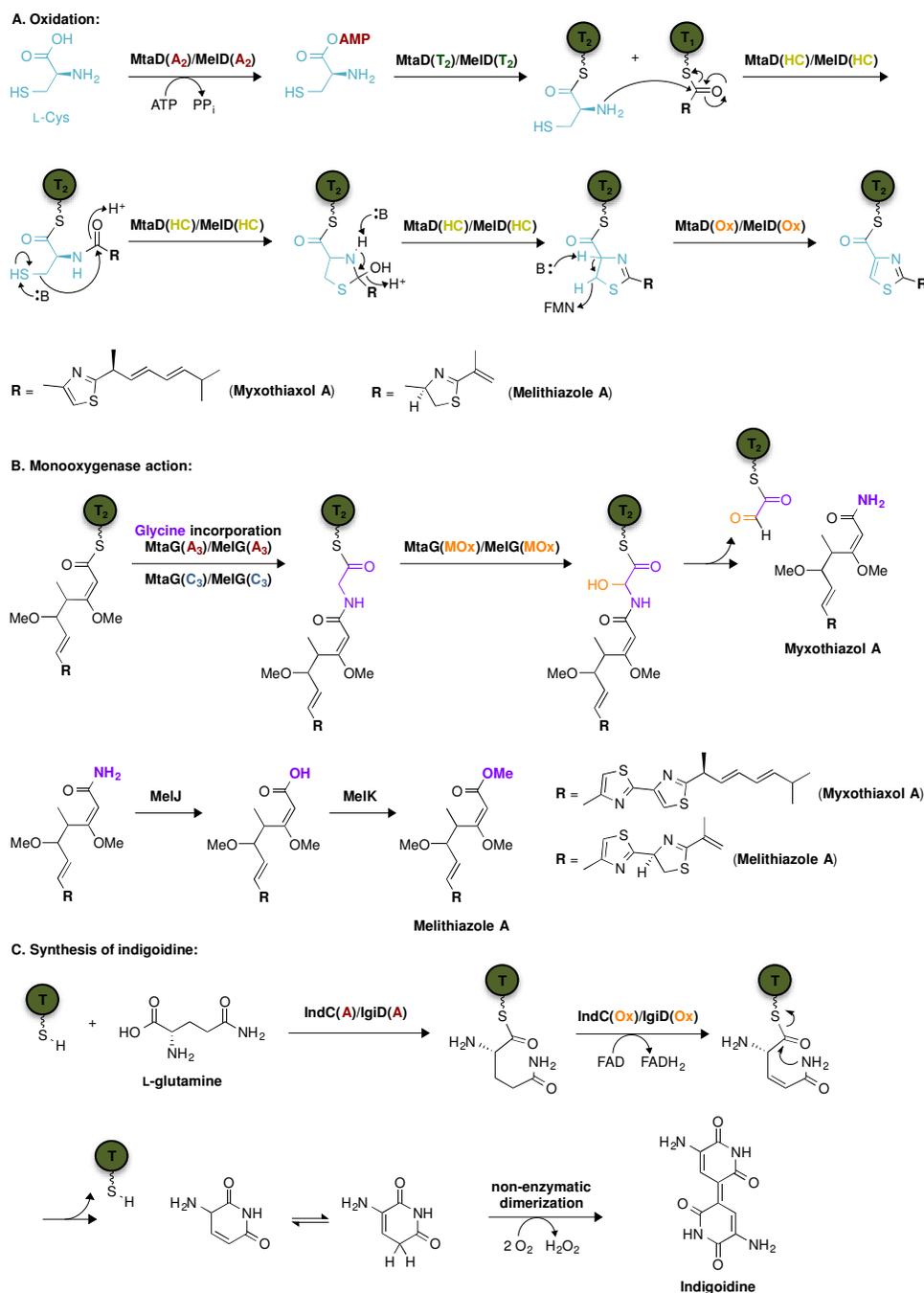
provides yet another viable means of controlling biosynthesis of novel NRPs, which has yet to be explored.

For many of these systems containing interrupted A domains, mechanistic and structural details remain to be resolved. Within the family of siderophores, for example, if the inactive M domains within PchE or HMWP2 were mutated active (R back to G), would methyl transfer occur? In the case of PchE, an additional reductase domain would possibly be required, but a second *N*-methylation could occur on that subsequent thiazolidine. For yersiniabactin, it is reasonable to expect that an M domain mutated to active would produce an *N*-methylated thiazolidine analogue.<sup>30</sup>

To better exploit A domains and other NRPS motifs, we need to understand where we can cut/insert portions of other auxiliary enzymes into A domains. The known interrupted A domains are commonly disrupted between motifs a8 and a9, but in some instances, between a2 and a3 or a4 and a5. Are there other positions in A domains where one could insert additional functionalities? What are the boundaries? These rules need to be understood and leave much to be explored, especially with regards to less common interruptions such as those between a2 and a3 as well as a4 and a5. The unique biosynthetic mechanisms featured in the interrupted A domains discussed here (*e.g.*, *S*-Me-L-Cys within thiocoraline and *O*-Me-L-Ser within kutznerides) are rare, but could be powerful for biosynthetic engineering. Furthermore, it remains to be answered whether or not domains other than M, KR, Ox, or MOx may be inserted to produce functional A domains capable of achieving two separate chemical reactions. We, and others, seek answers to these questions; our laboratory is currently working towards a better understanding

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

**HIGHLIGHT**

**Fig. 8** Proposed pathway for interrupted adenylation domains with **A.** oxidase (Ox) and **B.** monooxygenase (MOx) activity during myxothiazol and melithiazole biosynthesis. **C.** Proposed biosynthesis of indigoidine *Note:* oxidation may happen before (as depicted here) or after the cyclization step.

and engineering of a variety of interrupted A domains.

## 5. Acknowledgements

The authors would like to acknowledge the work on adenylation domains of those not cited in this highlight due to the scope of the

manuscript. Support for this work is provided by an NSF CAREER award MCB-1149427 (S.G.-T.) and by startup funds from the University of Kentucky. S.G.W. was supported by a University of Michigan NSF funded interdisciplinary REU program DBI-1156741.

## 6. Notes and references

- <sup>a</sup> Beloit College, Department of Chemistry, 700 College Street, Beloit, WI 53511, USA.
- <sup>b</sup> University of Michigan, Department of Medicinal Chemistry in the College of Pharmacy, 210 Washtenaw Ave, Ann Arbor, MI 48109, USA.
- <sup>c</sup> University of Kentucky, Department of Pharmaceutical Sciences in the College of Pharmacy, BioPharm Complex, 789 South Limestone Street, Lexington, KY 40536-0596, USA. Fax: +1 859-257-7585; Tel: +1 859-218-1686; E-mail: sylviegtsoodikova@uky.edu
- 1 C. T. Walsh, R. V. O'Brien, and C. Khosla, *Angew. Chem., Int. Ed.*, 2013, **52**, 7098-7124.
  - 2 F. Lombó, A. Velasco, A. Castro, F. de la Calle, A. F. Brana, J. M. Sanchez-Puelles, C. Mendez and J. A. Salas, *ChemBioChem*, 2006, **7**, 366-376.
  - 3 D. G. Fujimori, S. Hrvatin, C. S. Neumann, M. Strieker, M. A. Marahiel and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16498-16503.
  - 4 H. M. Patel and C. T. Walsh, *Biochemistry*, 2001, **40**, 9023-9031.
  - 5 D. A. Miller, L. Luo, N. Hillson, T. A. Keating and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 333-344.
  - 6 M. F. Kreutzer, H. Kage, P. Gebhardt, B. Wackler, H. P. Saluz, D. Hoffmeister and M. Nett, *Appl. Environ. Microbiol.*, 2011, **77**, 6117-6124.
  - 7 T. Hornbogen, S. P. Riechers, B. Prinz, J. Schultchen, C. Lang, S. Schmidt, C. Mugge, S. Turkanovic, R. D. Sussmuth, E. Tauberger and R. Zocher, *ChemBioChem*, 2007, **8**, 1048-1054.
  - 8 D. Tillett, E. Dittmann, M. Erhard, H. von Dohren, T. Borner and B. A. Neilan, *Chem. Biol.*, 2000, **7**, 753-764.
  - 9 T. Nishizawa, A. Ueda, T. Nakano, A. Nishizawa, T. Miura, M. Asayama, K. Fujii, K. Harada and M. Shirai, *J. Biochem.*, 2011, **149**, 475-485.
  - 10 A. Sandmann, F. Sasse and R. Muller, *Chem. Biol.*, 2004, **11**, 1071-1079.
  - 11 M. Ehling-Schulz, N. Vukov, A. Schulz, R. Shaheen, M. Andersson, E. Martlbauer and S. Scherer, *Appl. Environ. Microbiol.*, 2005, **71**, 105-113.
  - 12 M. Ehling-Schulz, M. Fricker, H. Grallert, P. Rieck, M. Wagner and S. Scherer, *BMC Microbiol.*, 2006, **6**, 20.
  - 13 Y. Q. Cheng, *ChemBioChem*, 2006, **7**, 471-477.
  - 14 B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hofle, S. Beyer and R. Muller, *J. Biol. Chem.*, 1999, **274**, 37391-37399.
  - 15 S. Weinig, H. J. Hecht, T. Mahmud and R. Muller, *Chem. Biol.*, 2003, **10**, 939-952.
  - 16 D. Yu, F. Xu, J. Valiente, S. Wang and J. Zhan, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 159-168.
  - 17 O. E. Zolova and S. Garneau-Tsodikova, *J. Antibiot.*, 2014, **67**, 59-64.
  - 18 N. A. Magarvey, M. Ehling-Schulz and C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 10698-10699.
  - 19 A. M. Gulick, *ACS Chem Biol.*, 2009, **4**, 811-827.
  - 20 V. Bucevic-Popovic, M. Sprung, B. Soldo and M. Pavela-Vrancic, *ChemBioChem*, 2012, **13**, 1913-1920.
  - 21 Y. He, Y. Sun, T. Liu, X. Zhou, L. Bai and Z. Deng, *Appl. Environ. Microbiol.*, 2010, **76**, 3283-3292.
  - 22 D. Schwarzer, R. Finking and M. A. Marahiel, *Nat. Prod. Rep.*, 2003, **20**, 275-287.
  - 23 E. Conti, T. Stachelhaus, M. A. Marahiel and P. Brick, *EMBO J.*, 1997, **16**, 4174-4183.
  - 24 T. V. Lee, L. J. Johnson, R. D. Johnson, A. Koulman, G. A. Lane, J. S. Lott and V. L. Arcus, *J. Biol. Chem.*, 2010, **285**, 2415-2427.
  - 25 A. M. Gulick, V. J. Starai, A. R. Horswill, K. M. Homick and J. C. Escalante-Semerena, *Biochemistry*, 2003, **42**, 2866-2873.
  - 26 N. E. Mikkelsen, B. Munch-Petersen and H. Eklund, *FEBS J.*, 2008, **275**, 2151-2160.
  - 27 L. Du, Y. He and Y. Luo, *Biochemistry* 2008, **47**, 11473-11480.
  - 28 J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, and A. M. Gulick, *Chem. Biol.*, 2012, **19**, 188-198.
  - 29 C. A. Mitchell, C. Shi, C. C. Aldrich and A. M. Gulick, *Biochemistry*, 2012, **51**, 3252-3263.
  - 30 R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2011, **38**, 1747-1760.
  - 31 R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 357-369.
  - 32 A. H. Al-Mestarihi, G. Villamizar, J. Fernandez, O. E. Zolova, F. Lombó and S. Garneau-Tsodikova, *J. Am. Chem. Soc.*, 2014, **136**, 17350-17354.
  - 33 W. Zhang, W.; J. R. Jr. Heemstra, C. T. Walsh and H. J. Imker, *Biochemistry*, 2010, **49**, 9946-9947.
  - 34 E. Stegmann, C. Rausch, S. Stockert, D. Burkert and W. Wohlleben, *FEMS Microbiol. Lett.*, 2006, **262**, 85-92.
  - 35 S. Lautru, D. Oves-Costales, J. L. Pernodet and G. L. Challis, *Microbiology*, 2007, **153**, 1405-1412.
  - 36 M. Wolpert, B. Gust, B. Kammerer and L. Heide, *Microbiology*, 2007, **153**, 1413-1423.
  - 37 E. J. Drake, J. Cao, J. Qu, M. B. Shah, R. M. Straubinger and A. M. Gulick, *J. Biol. Chem.*, 2007, **282**, 20425-20434.
  - 38 G. W. Buchko, C. Y. Kim, T. C. Terwilliger and P. J. Myler, *Tuberculosis*, 2010, **90**, 245-251.
  - 39 E. A. Felngale, J. J. Barkei, H. Park, A. M. Podelvels, M. D. McMahon, D. W. Drott and M. G. Thomas, *Biochemistry*, 2010, **49**, 8815-8817.
  - 40 H. J. Imker, D. Krahn, J. Clerc, M. Kaiser and C. T. Walsh, *Chem. Biol.*, 2010, **17**, 1077-1083.
  - 41 B. Boll, T. Taubitz and L. Heide, *J. Biol. Chem.*, 2011, **286**, 36281-36290.
  - 42 E. Tatham, S. Sundaram Chavadi, P. Mohandas, U. R. Edupuganti, S. K. Angala, D. Chatterjee, and L. E. Quadri, *BMC Microbiol.*, 2012, **12**, 118.
  - 43 D. A. Herbst, B. Boll, G. Zocher, T. Stehle and L. Heide, *J. Biol. Chem.*, 2013, **288**, 1991-2003.
  - 44 B. Boll and L. Heide, *ChemBioChem*, 2013, **14**, 43-44.
  - 45 J. M. Davidsen, D. M. Bartley and C. A. Townsend, *J. Am. Chem. Soc.*, 2013, **135**, 1749-1759.
  - 46 C. Zhang, L. Kong, Q. Liu, X. Lei, T. Zhu, J. Yin, B. Lin, Z. Deng and D. You, *PLoS one* 2013, **8**, e56772.
  - 47 S. A. Sieber and M. A. Marahiel, *Chem. Rev.*, 2005, **105**, 715-738.
  - 48 M. Z. Ansari, J. Sharma, R. S. Gokhale and D. Mohanty, *BMC Bioinf.*, 2008, **9**, 454.
  - 49 F. Romero, F.; Espliego, J. Perez Baz, T. Garcia de Quesada, D. Gravalos, F. de la Calle and J. L. Fernandez-Puentes, *J. Antibiot.*, 1997, **50**, 734-737.
  - 50 J. Perez Baz, L. M. Canedo, J. L. Fernandez Puentes and M. V. Silva Elipse, *J. Antibiot.*, 1997, **50**, 738-741.
  - 51 O. E. Zolova, A. S. Mady and S. Garneau-Tsodikova, *Biopolymers*, 2010, **93**, 777-790.
  - 52 L. Robbel, K. M. Hoyer and M. A. Marahiel, *FEBS J.*, 2009, **276**, 1641-1653.
  - 53 A. Broberg, A. Menkis and R. Vasiliauskas, *J. Nat. Prod.*, 2006, **69**, 97-102.
  - 54 C. D. Cox, *J. Bacteriol.*, 1980, **142**, 581-587.
  - 55 G. Weber, K. Schorgendorfer, E. Schneider-Scherzer and E. Leitner, *Curr. Genet.*, 1994, **26**, 120-125.
  - 56 X. Wang, J. Tabudravu, M. E. Rateb, K. J. Annand, Z. Qin, M. Jaspars, Z. Deng, Y. Yu and H. Deng, *Mol. Biosyst.*, 2013, **9**, 1286-1289.
  - 57 W. Weckwerth, K. Miyamoto, K. Inuma, M. Krause, M. Glinski, T. Storm, G. Bonse, H. Kleinkauf and R. Zocher, *J. Biol. Chem.*, 2000, **275**, 17909-17915.
  - 58 L. Rouhiainen, L. Paulin, S. Suomalainen, H. Hyytiainen, W. Buikema, R. Haselkorn and K. Sivonen, *Mol. Microbiol.*, 2000, **37**, 156-167.
  - 59 H. T. Chiu, B. K. Hubbard, A. N. Shah, J. Eide, R. A. Fredenburg, C. T. Walsh and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8548-8553.
  - 60 Y. Mast, T. Weber, M. Goltz, R. Ort-Winklbaauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol.*, 2011, **4**, 192-206.
  - 61 R. Loria, D. R. Bignell, S. Moll, J. C. Huguet-Tapia, M. V. Joshi, E. G. Johnson, R. F. Seipke and D. M. Gibson, *Antonie van Leeuwenhoek*, 2008, **94**, 3-10.
  - 62 Z. Chang, P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis and D. H. Sherman, *Gene*, 2002, **296**, 235-247.
  - 63 M. C. Moffitt and B. A. Neilan, *Appl. Environ. Microbiol.*, 2004, **70**, 6353-6362.
  - 64 Y. Chai, S. Shan, K. J. Weissman, S. Hu, Y. Zhang and R. Muller, *Chem. Biol.*, 2012, **19**, 361-371.

- 
- 65 H. Chen, S. O'Connor, D. E. Cane and C. T. Walsh, *Chem. Biol.*, 2001, **8**, 899-912.
- 66 I. Muller and R. Muller, *FEBS J.*, 2006, **273**, 3768-3778.
- 67 S. Reverchon, C. Rouanet, D. Expert and W. Nasser, *J. Bacteriol.*, 2002, **184**, 654-665.
- 5 68 A. O. Brachmann, F. Kirchner, C. Kegler, S. C. Kinski, I. Schmitt and H. B. Bode, *J. Biotechnol.*, 2012, **157**, 96-99.
- 69 W. N. Cude, J. Mooney, A. A. Tavanaei, M. K. Hadden, A. M. Frank, C. A. Gulvik, A. L. May and A. Buchan, *Appl. Environ. Microbiol.*, 2012, **78**, 4771-4780.
- 10 70 R. Novakova, Z. Odnogova, P. Kutas, L. Feckova and J. Kormanec, *Folia Microbiol.*, 2010, **55**, 119-125.
- 71 C. T. Walsh and T. A. Wencewicz, *Nat. Prod. Rep.*, 2013, **30**, 175-200.
- 15 72 G. J. Williams, *Curr. Opin. Struct. Biol.*, 2013, **23**, 603-612.
- 73 H. L. Conurso and S. D. Bruner, *Curr. Opin. Chem. Biol.*, 2012, **16**, 162-169.
- 74 C. T. Walsh, *ChemBioChem*, 2002, **3**, 125-134.
- 75 H. G. Menzella and C. D. Reeves, *Curr. Opin. Microbiol.*, 2007, **10**, 238-245.
- 20 76 K. T. Nguyen, D. Ritz, J. Q. Gu, D. Alexander, M. Chu, V. Miao, P. Brian and R. H. Baltz, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 17462-17467.
- 77 G. C. Uguru, C. Milne, M. Borg, F. Flett, C. P. Smith and J. Micklefield, *J. Am. Chem. Soc.*, 2004, **126**, 5032-5033.
- 25 78 K. Eppelmann, T. Stachelhaus and M. A. Marahiel, *Biochemistry*, 2002, **41**, 9718-9726.
- 79 B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao and N. L. Kelleher, *Chem. Biol.*, 2011, **18**, 601-607.
- 30 80 J. Thirlway, R. Lewis, L. Nunns, M. Al Nakeeb, M. Styles, A. W. Struck, C. P. Smith and J. Micklefield, *Angew. Chem., Int. Ed.*, 2012, **51**, 7181-7184.