Review

Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product

MANGAL SINGH, SANDEEP CHAUDHARY and DIPTI SAREEN*
Department of Biochemistry, Panjab University, Chandigarh, India

**Corresponding author (Email, diptsare@pu.ac.in)*

Non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) present in bacteria and fungi are the major multi-modular enzyme complexes which synthesize secondary metabolites like the pharmacologically important antibiotics and siderophores. Each of the multiple modules of an NRPS activates a different amino or aryl acid, followed by their condensation to synthesize a linear or cyclic natural product. The studies on NRPS domains, the knowledge of their gene cluster architecture and tailoring enzymes have helped in the *in silico* genetic screening of the ever-expanding sequenced microbial genomic data for the identification of novel NRPS/PKS clusters and thus deciphering novel non-ribosomal peptides (NRPs). Adenylation domain is an integral part of the NRPSs and is the substrate selecting unit for the final assembled NRP. In some cases, it also requires a small protein, the MbtH homolog, for its optimum activity. The presence of putative adenylation domain and MbtH homologs in a sequenced genome can help identify the novel secondary metabolite producers. The role of the adenylation domain in the NRPS gene clusters and its characterization as a tool for the discovery of novel cryptic NRPS gene clusters are discussed.

[Singh M, Chaudhary S and Sareen D 2017 Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product. *J. Biosci.* **42** 175–187]

1. Introduction

The post-genomic era has seen an upsurge in the knowledge of natural products' and secondary metabolites' biosynthetic machinery. The abundance of the microbial genome data being made available publicly and the advancements in sequencing and computational tools has helped to achieve higher success rate in predicting novel natural products from the previously unknown bacterial or plant sources. The thus-predicted natural compounds have an enormous potential to be important leads for the development of pharmaceutical drugs. The natural products are polyketides (PKs), non-ribosomal peptides (NRPs), terpenoids, alkaloids and saccharides (Medema and Fischbach 2015). The PKs and NRPs are complex and quite diverse in their structures owing to the multi-modular nature of their biosynthetic machinery, i.e. PKS and NRPS. Each of the multiple modules of NRPS/PKS activates a different carboxyl or amino acid, followed by their sequential condensation and/or modification. The studies on NRPS/PKS domain organization has made possible the *in silico* genetic screening of novel and diverse NRPSs/PKSs in the uncharacterized microbial strains or metagenomes. Among bacteria, most of the known secondary metabolites have been found to be produced by actinomycetes and myxobacteria. The sequence-guided identification of cryptic NRPS/PKS clusters has led to the successful prediction of many more novel metabolites in these microbial cultures (Reger *et al.* 2008; Boll *et al.* 2011; Herbst *et al.* 2013).

Various computational approaches have been adapted for the prediction of novel metabolites encoded by the putative cryptic gene clusters (Medema and Fischbach 2015). Evolving from the earlier BLAST-based approaches, many sophisticated bioinformatic tools are now available (table 1) for the identification of novel biosynthetic gene clusters (BGCs) of a known biosynthetic class, in a newly sequenced genome. These computational tools are based on the presence of specific signature domains which form the basis for the development of gene identification algorithms. Further on, the biggest challenge remains the prediction of the encoded natural product from

Keywords. Adenylation assay; genome mining; MbtH; natural product; NRPS; structure prediction

DOI: 10.1007/s12038-017-9663-z

Published online: 19 January 2017

Table 1. Currently available tools for scanning and analysis of NRPS and PKS clusters/domains

Bioinformatic tool	Prediction	Impact*	References
antiSMASH 3.0	NRPS, PKS, RiPPs and other secondary metabolites	1060	Medema *et al.* 2011; Blin *et al.* 2013; Weber *et al.* 2015
NRPSpredictor2	A-domain specificity	563	Röttig *et al.* 2011
SBS-PKS	NRPS, PKS and domain specificity	343	Ansari *et al.* 2004; Anand *et al.* 2010
PKS/NRPS	NRPS, PKS and domain specificity	194	Bachmann and Ravel 2009
NORINE	Database of non-ribosomal peptides	174	Caboche *et al.* 2008
NP.searcher	NRPS, PKS and others	128	Li *et al.* 2009
ClustScan	NRPS, PKS	106	Starcevic *et al.* 2008
NaPDoS	KS and C domain analysis	115	Ziemert *et al.* 2012
NRPSsp	A-domain specificity	43	Prieto *et al.* 2012
NRPS/PKS substrate predictor	A and AT domains specificity	29	Khayatt *et al.* 2013
PRISM	NRPS and PKS	26	Skinnider *et al.* 2015
Dynamite	NRPS and PKS	7	Ogasawara *et al.* 2015

*Total citations from all the published versions as on August 2016.

the identified putative biosynthetic cluster, especially those of PKSs and NRPSs, as the sequence and selection of the carboxyl/amino acid dictates the final product chemistry. For that, comprehensive training data sets are being generated by several groups (Baranašić *et al.* 2014). Minowa *et al.* have proposed an alternate method for prediction of NRP/PK compounds, encoded from the microbial genomes (Minowa *et al.* 2007). Their method is based not only on the specific domain sequences/motifs corresponding to specific substrates but also on the physical protein–protein interactions between the different modules of PKSs/NRPSs (Weissman and Müller. 2008). The advancements in analytical techniques, particularly mass spectrometry, has also contributed enormously to the success of natural product identification (Bouslimani *et al.* 2014; Krug and Müller 2014).

Of all the domains of PKS/NRPS, the substrate prediction of the adenylation domain (A-domain) computationally has been found to be the trickiest. This is due to the unavailability of computational methods trained on well-curated data sets for such a substrate diversity of A-domain or due to the presence of novel and uncharacterized substrate-recognizing motifs in A-domains. Adenylating protein is responsible for the chemical activation of the acyl/aryl substrates by ATP, forming an AMP adduct and releasing pyrophosphate. This process of adenylate formation is involved in a variety of metabolic pathways including ribosomal protein synthesis, fatty acid oxidation and enzyme regulation. A-domain is an integral part of NRPS and is the 'substrate selecting unit' for the final assembled NRP. However, the determination of substrate specificity of the adenylating protein became possible only due to the generation of the data sets based on A-domain's sequence specificity code, i.e. substrate-recognizing amino acid residues. Hence, there is a need for the generation of experimental data sets, based on the results of high-throughput adenylation activity assays, like ATP-pyrophosphate exchange assay (Mcquade *et al.* 2009). In other words, the characterization of A-domains can help in the structure prediction of the NRP, encoded by a cryptic NRPS BGC. A-domain specificity can also be altered for the synthesis of unnatural products with novel physicochemical properties over the original natural product (Hur *et al.* 2012). In this review, the role of A-domain in NRPS gene clusters and its characterization, as a tool for the discovery of novel cryptic clusters, is discussed. Using some of the computational analysis tools mentioned in this review, we have predicted the approximate chemistry of a secondary metabolite, the putative product of a cryptic NRPS cluster identified in the genome of an actinomycete, *Stackebrandtia nassauensis* DSM 44728.

1.1 *NRPS domain architecture*

NRPSs are organized into multiple modules encompassing long ORFs and these modules comprise of multiple catalytic

domains which work in coordination. A minimal module comprises of an adenylation (A) and thiolation (T) domain, which activates and contributes a single unit to the final assembled product. For example, the NRPS for actinomycin (a dimer of 4-MHA pentapeptide lactone) comprises 6 modules (1–6) spread over four genes, i.e. *acmA-acmD* (Keller *et al.* 2010) (figure 1A and B). The minimal first (1) and second (2) module activates and contribute 4-MHA (4-methyl-3-hydroxyanthranilic acid) and L-threonine to the final actinomycin structure (figure 1C and D). The number of NRPS modules and their sequence corresponds to the number of aminoacyl or aryl residues in the NRP, representing a co-linearity rule (Stachelhaus and Marahiel 1995; Kleinkauf and Von Döhren 1996). The A-domain is responsible for the substrate selection and activation, which is then transferred to the thiolation domain. The thiolation domain contains a phosphopantetheine arm which swings between the adenylation and condensation domain (C) and acts as a carrier of the aminoacyl/aryl thioester (Gocht and Marahiel 1994). The C-domain catalyses peptide bond formation between the two activated and thiolated substrate units, leading to chain elongation. The loading module (figure 1; module 1) from where the assembly begins differs from other modules by the absence of a condensation domain, which as apparent is not required in the first module. The process of activation and condensation is repeated until the last module, which contains an additional thioesterase (TE) domain at the C-terminus, to release the chain by hydrolysis or cyclization (figure 1B). Besides the thioesterase, several other accessory domains are present, as in the actinomycin NRPS, the D-valine introduced by the third module is generated from L-valine by the epimerization domain (E). The methyltransferase (MT) in the fifth module N-methylates the activated L-glycine to form *N*-methylglycine (sarcosine). Table 2 lists all the domains present in an NRPS and their equivalent domains present in a PKS, another class of analogous megasynthases involved in polyketide biosynthesis (Fischbach and Walsh 2006; Helfrich and Piel 2016).

1.1.1 Adenylation domain: A-domains are the gatekeeper enzymes of the NRPS (Cane and Walsh 1999) and each constituting unit in the NRP is activated by the respective A-domain of a module as discussed above in the example of actinomycin. A-domains catalyse a two-step adenylation-acylation reaction (Gulick 2009), i.e. binding of A-domain to a cognate residue and ATP in the first step leading to their condensation to form a tightly bound aminoacyl/aryl adenylate (figure 2A). In the second half reaction, the aminoacyl/aryl group of acyl/aryl adenylate is transferred to the phosphopantetheine arm by nucleophilic attack of the sulphur atom forming an acyl/aryl thioester (figure 2B). A-domains of NRPS belong to the ANL superfamily of adenylating enzymes, which also includes luciferase and acyl-CoA synthetases, all of which catalyse the partial adenylation half reaction (Gulick 2009). Despite of high structure conservation and same reaction mechanism, each of them selects and activates a different substrate from a pool of more than 20 proteinogenic amino acids as well as non-proteinogenic amino and aryl acids. The substrate selection by A-domain, out of a diverse cellular pool of molecules, is determined by the amino acid residues lining its binding pocket, which also hinders any catalytically favourable interaction with a non-cognate substrate. These residues were initially identified by Stachelhaus and co-workers by *in silico* and structure-function mutagenesis studies and were thus called Stachelhaus motif (Stachelhaus *et al.* 1999). A-domain contains 10 conserved regions, which are signature motifs of the A-domain named as A1 through A10. Among these, the binding pocket lies in a region between the A3-A6 motifs. In biochemically uncharacterized A-domains, Stachelhaus motif is identified by a comparison of the region between the A3-A6 motifs with that of GrsA (Gramicidin S synthetase, phenylalanine activating A-domain). A specific Stachelhaus code exists for a particular substrate and if an identical code is present in an uncharacterized A-domain, then the probability of their substrates being the same is high (Stachelhaus *et al.* 1999). This comparative analysis is used by many NRPS prediction tools like the SBS-PKS, PKS/NRPS and NP.searcher (table 1) to predict the approximate product chemistry of novel NRPS clusters by predicting substrates of all the A-domains. The prediction has been further refined in NRPSpredictor2 by including all the active site residues close to the substrate (Rausch *et al.* 2005). Still the prediction of the final NRP gets complicated due to the modification by the *cis* and *trans* acting tailoring enzymes.

MbtH, another small protein of 70 amino acids, was found to be a part of the mycobactin biosynthetic NRPS-PKS hybrid gene cluster discovered in *M. tuberculosis* (Quadri *et al.* 1998). Presence of such a small protein encoding gene was puzzling in NRPS-PKS cluster, and hence its study remained dormant for quite a long time. Later, consistent identification of MbtH homologs in other secondary metabolite gene clusters as well drew attention of the researchers and thus 'MbtH-like protein' became a common term to name a similar protein of ~8–9 kDa, found within the NRPS gene cluster. Interestingly, the first report of MbtH-like protein (MbtH-LP) in bahlimycin gene cluster of *Amycolatopsis bahlimycinia* ruled out the possibility of functional requirement of MbtH-LP in bahlimycin production (Stegmann *et al.* 2006; Wolpert *et al.* 2007; Lautru *et al.* 2007). At the same time, authors also hypothesized the possibility of complementation by other MbtH homologs present elsewhere in this bacterial genome. Later, two reports were published back to back, based on two MbtH-LP (cchK and cdaX) of *Streptomyces coelicolor* M145, encoded within the gene clusters of coelichelin and calcium-dependent antibiotic (CDA) respectively, which for the first time provided an

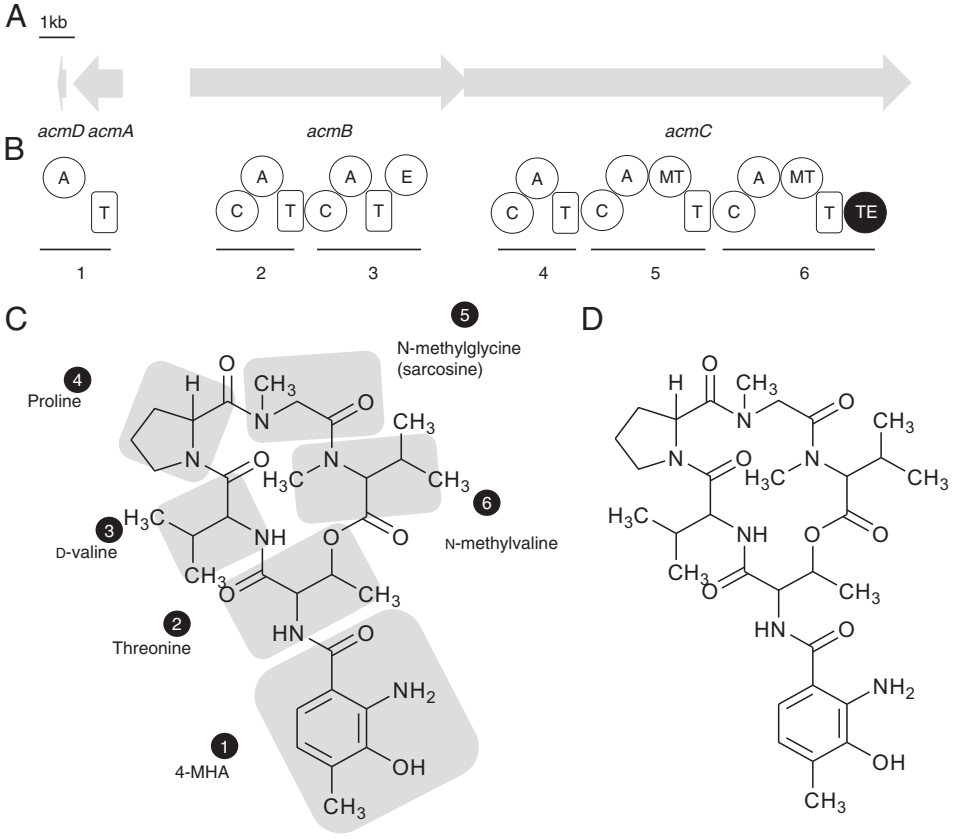

Figure 1. Modular organization of actinomycin biosynthetic cluster in *Streptomyces chrysomallus.* (**A**) Four NRPS genes *acm*A-*acmD.* (**B**) Modular organization (1-6) of the assembly line that catalyses the formation of a 4-MHA pentapeptide lactone. (**C**) Each contributed unit corresponding to the module is highlighted. (**D**) 4-MHA pentapeptide lactone is one half the molecule of the final actinomycin structure (not shown here). A, C and T are adenylation, condensation and thiolation domains, E – epimerase, MT – methyltransferase and TE – thioesterase.

evidence for functional role of MbtH-LP in synthesis of secondary metabolite (Wolpert *et al.* 2007; Lautru *et al.* 2007). Even the functional complementation of the two MbtH homologs (cchK and cdaX) was proven experimentally (Lautru *et al.* 2007), thus proposing that MbtH mediated cross-talk between two NRPS-dependent pathways might be a general phenomenon in those bacteria which have multiple NRPS gene clusters (Wolpert *et al.* 2007). Considering a functional link between MbtH-LP and secondary metabolism, the MbtH-LP PA2412 of pyoverdine gene cluster was crystallized in 2007 (Drake *et al.* 2007). The crystal structure showed the presence of a novel motif βββαα, along with manifestation of three conserved tryptophan residues on one face of the MbtH-LP. Thus came the hypothesis of protein–protein interaction between MbtH-LP and NRPS synthetase through this conserved surface supported by the co-purification studies of MbtH-LP (CmnN and VioN) with the corresponding NRPSs of capreomycin and viomycin, respectively (Felnagle *et al.* 2010). Structural study

Table 2. NRPS and PKS equivalent domains

NRPS domains	PKS equivalents	Role
Adenylation (A)	Acyl transferase (AT)	Selection unit
Thiolation (T)/Peptide carrier protein (PCP)	Thiolation (T)/Acyl carrier protein (ACP)	Carrier unit
Condensation (C)	Ketosynthase (KS)	Condensing unit
Thioesterase (TE)	Thioesterase (TE)	Releasing unit

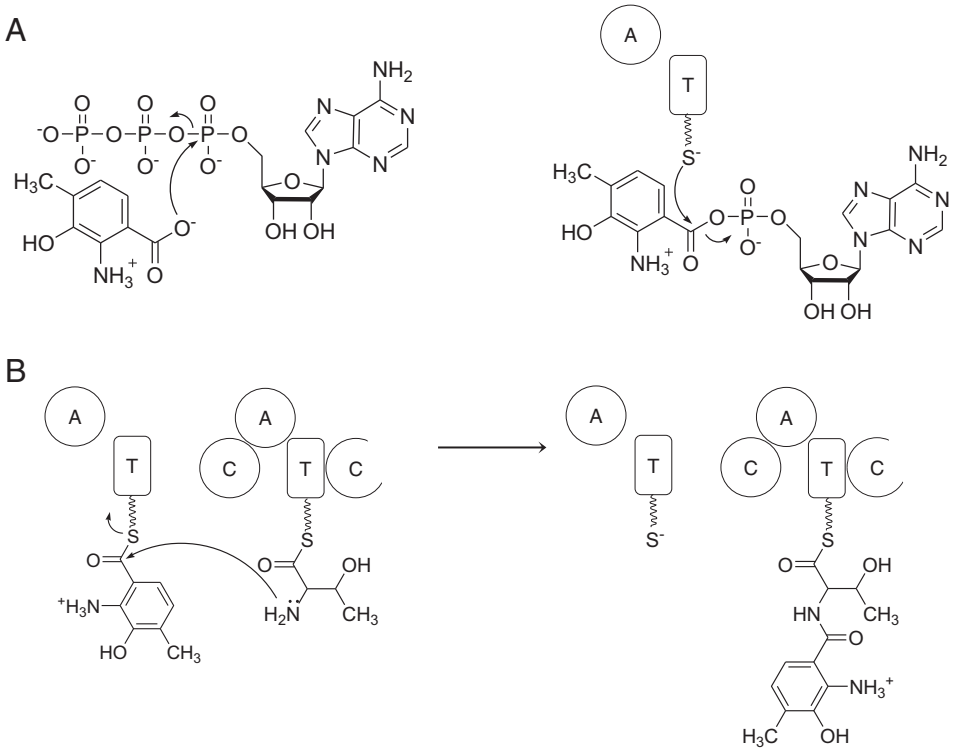

Figure 2. Activation and chain elongation during actinomycin biosynthesis. (**A**) Adenylation half reaction leading to 4-MHA-AMP formation (acyl adenylate) and transfer of 4-MHA (acyl group) onto the phosphopantetheine arm by nucleophilic attack of sulphur leading to formation of a thioester with release of AMP. (**B**) Peptide bond formation between two acyl thioesters catalysed by the condensation domain which leads to chain elongation.

of MbtH itself from *M. tuberculosis* using NMR, produced similar conclusion (Buchko *et al.* 2010).

Initially, a role of chaperone was attributed to MbtH-LP in mycobactin, glidobactin and thiocoraline biosynthesis (Imker *et al.* 2010; McMahon *et al.* 2012; Zolova and Garneau-Tsodikova 2012) as the soluble expression of adenylating enzyme could be achieved only after co-expression of the cognate MbtH-LP. Mutational analysis over the A-domain and its MbtH-LP pair, viz. PacL/PacJ, indicated strong structural interaction between them (Zhang *et al.* 2010). However, MbtH-LPs are required by some and not all adenylation reactions in the bacterial secondary metabolite synthesis. Among bacteria, Firmicutes and Bacteroidetes have been found to be devoid of MbtH homologs. The biosynthetic gene cluster of novobiocin had no MbtH-like gene in the near vicinity and NovH, the tyrosine adenylating enzyme of novobiocin was active without any requirement of MbtH-LP, despite of having 83% sequence identity with MbtH-LP dependent adenylating enzyme CloH (Boll *et al.* 2011). Even after the experimental proofs and crystal structure data of MbtH-LP, it was not certain whether the interaction between A-domain and MbtH-LP is at or near the active catalytic centre of the adenylating enzyme. Recently, in one of the path breaking study, the MbtH-like domain was found to bind at a site that is distant from the active centre of the adenylating enzyme (Herbst *et al.* 2013). Also, for the first time an MbtH-LP was crystallized with its cognate adenylating protein of NRPS, and was investigated to understand the mechanism of interaction between the two. The gene for the adenylating enzyme SlgN1 with its N-terminally fused MbtH-like domain from the streptolydigin biosynthetic NRPS gene cluster was synthesized, expressed in *E. coli*, purified and crystallized for its structure-function characterization (Herbst *et al.* 2013). It was known by the earlier studies that A-domain consists of two domains (Reger *et al.* 2008), N-terminal A_{core} and C-terminal A_{sub} domain, where A_{core} domain is responsible for adenylation and A_{sub} domain for the thioester bond formation. Here, it was confirmed that the MbtH-LP interacts only with A_{core} domain. The interface found in SlgN1 and MbtH-LP included highly conserved tryptophan residues of MbtH, which critically contributed in the adenylation activity.

Revelation of biochemistry of these small proteins has changed the course of NRPS-PKS study. Today, more than 600 MbtH-LP homologs have been identified, and this protein has become a marker of capability of secondary

metabolite synthesis in the gifted microbes due to its association directly with adenylating protein (Baltz 2011b). The small-sized MbtH-homolog genes, when used as a query to survey large-sized and large numbers of genomes, the search for cryptic clusters becomes straightforward (Baltz 2014). Thus, a 24-mer multiprobe was generated, using the multiple MbtH orthologous and paralogous sequences known to date, to identify the potential novel secondary metabolite producers (Baltz 2014). These potential gifted microbes identified by genome mining can further be explored using available bioinformatics tools for pathway assembly and annotation, followed by expression studies in heterologous hosts.

1.1.2 Chain elongation domains: The thiolation domain present in NRPS is also known as peptidyl carrier protein (PCP) as it carries amino acids and intermediate peptides. This domain has a highly conserved serine residue to which 4′-phosphopantetheine arm, derived from coenzyme A (CoA), is transferred by a phosphopantetheinyl transferase (PPTase) which is often encoded in the genes vicinal to the NRPS (Lambalot *et al.* 1996). Non-ribosomal peptide contains both L and D amino acids and these D-amino acids are especially generated by an epimerase encoded in the cluster, which acts on the activated aminoacyl thioester. The condensation domain just downstream of a D-amino acid contributing module is specific for a D-amino acid, to preclude condensation of non-epimerized substrate and is known as $^{D}C_{L}$ type condensation domain. For peptide bond formation between two L-amino acids, there is a $^{L}C_{L}$ domain. Condensation domain for two D-amino acids ($^{D}C_{D}$) does not exist but there exists a separate mechanism for such peptides. In total, five subtypes of condensation domains have been recognized (Rausch *et al.* 2007). Besides the two mentioned above, there can be a starter C-domain which acylates the first activated amino acid with β-hydroxy fatty acid, a dual E/C domain responsible for epimerization and peptide bond formation or a heterocyclization domain (Cy) responsible for peptide bond formation or cyclization of serine, threonine and cysteine residues. Inferring the subtype of the modular C-domains of uncharacterized NRPS clusters helps in the prediction of the NRPS product. Usually a condensation domain is absent in the starter or loading module except for the starter C-domain. Dual E/C domain containing module incorporates a D-amino acid and as both epimerization and condensation is carried out by this single domain, a separate epimerization domain is absent. The five membered ring formed from the serine, threonine or cysteine residues by cyclization domain (Cy) is responsible for chelation of metals or binding to DNA, RNA or protein (Rausch *et al.* 2007). The $^{L}C_{L}$ and $^{D}C_{L}$ condensation domains, both contain the HHxxxDG motif, dual E/C domain contains HH[I/L]xxxxGD and in the cyclization domains, the active site motif containing histidine is replaced by DxxxxD motif. The natural product domain seeker (NaPDoS) is an online tool (table 1) dedicated to identify the condensation domain and its subtype in the NRPS modules as well as the subtype of KS domains of the PKS (table 2). Chain elongation halts at the last module of the NRPS (and PKS) assembly line, which often encodes a thioesterase (TE) domain responsible for the release or cyclization by lactone formation of the assembled NRP. TE domain belongs to the family of alpha-beta hydrolase superfamily. If absent in the modules, a thioesterase often encoded by the genes vicinal to NRPS genes, acts in *trans* and is known as type II TE domain (Yeh *et al.* 2004).

1.1.3 Tailoring enzymes: One of the many tailoring reactions involves methylation of the activated substrate at N, C or O by the respective methyltransferases. Deciphering the function of methyltransferases (MT) becomes important for the correct structure prediction and it involves multiple sequence alignment with known MTs for the respective motifs identified by Ansari *et al.* (2008). N-MT consists of motif I, II/Y, IV and V while C-MT consists of motif I, I-Post, II and III, and O-MT consists of only motif I, II and III (Ansari *et al.* 2008). Most of the diversity generated in the NRP is as a result of the action of the accessory domains present in the modules, besides the specificity of the A-domains. These tailoring enzymes involve the epimerases, methyltransferases and thioesterase domain, both *cis*- and *trans*-acting reductases and glycosyltransferases.

2. Computational methods/tools for microbial genome mining

The predictive power of the bioinformatics sequence analysis tools for the discovery of novel PKS/NRPS products lies in the co-linearity logistics of these modular megasynthases (Mootz and Marahiel 1997; Staunton and Wilkinson 1997). However, there are several examples of the nonlinear modular synthesis of natural product as well (Haynes and Challis 2007). Despite of some constraints and exceptions, the ever-expanding genomic data is being actively analysed for the prediction and discovery of novel cryptic natural products in microbes. Computational tools such as antiSMASH, ClustScan, NP.searcher, etc., enable the genome wide scanning for the identification of BGCs and prediction of genetically encoded NRPs and PKs (Medema and Fischbach 2015). However, the automated structure prediction of the natural product encoded from the thus-identified clusters still remains a challenge, that has been addressed to some extent in the newly developed tool, PRISM (Skinnider *et al.* 2015). A special feature of PRISM is inclusion of the biosynthetic logic to include the exceptional BGCs, which do not obey the principle of collinearity including *trans*-acting adenylation and acyltransferase domains. The accuracy of the

predicted structure by these tools should still be analysed manually. Since the above-mentioned computational methods have the limitation for prediction of alkaloids, terpenoids, RiPPs, etc., from their BGCs, they are identified manually using basic *in silico* analysis tools. The simplest online tool available on NCBI to locate the conserved domains in a protein sequence, is the Conserved Domain Database (CDD). It can always identify an A-domain of an NRPS, however the prediction of substrate specificity of the A-domain and closest domain boundary in the NRPS requires additional bioinformatics tools, like those enlisted in the table 1. Following a genome mining strategy for RiPPs using BLAST and other basic utilities, we have identified novel lantibiotic encoding gene clusters in 24 bacterial genomes (Singh and Sareen 2014). In genome mining studies, a homology search is required with the domains exclusive to NRPS and PKS clusters for the identification of bacteria with potential for production of novel secondary metabolites (figure 3). The sequence conservation in the A-domain has been used as a successful tool for the discovery of novel NRP by homology search for A-domains and comparison of active site residues with that of the characterized adenylating proteins. The siderophore coelichelin from *S. coelicolor* is the first example of the metabolite predicted, based upon A-domain's active site residues, followed by a few more which have been extensively reviewed in earlier work (Challis 2008; Medema and Fischbach 2015). A-domains therefore can be exploited for the mining of the cryptic NRPS, NRPS-PKS hybrid clusters, standalone and unique adenylating proteins from the genomic or metagenomics database (Li *et al.* 2009). For the *in silico* analysis, many tools for both manual and automatic scanning of genomes (table 1) for the NRPS clusters are available. The *in silico* efforts are to be then substantiated by *in-vivo* techniques such as comparative genomic profiling, gene knockout, heterologous gene expression and isotopically labelled guided metabolite purification (Challis 2008).

2.1 *Microbial genome mining using adenylation domain*

As a proof of concept, using A-domain as a query against the recently sequenced genome of an actinobacteria, *Stackebrandtia nassauensis* DSM 44728 (Munk *et al.* 2009) which is a soil isolate from Nassau, Bahamas, we could identify at least four NRPS-PKS hybrid clusters. At present, the genus *Stackebrandtia* contains only one additional species: *S. albiflava*, another soil isolate from a tropical rainforest in China (Wang *et al.* 2009). The strategy mentioned in figure 3 was followed and taking the sequence of A-domain of GrsA as a query, led

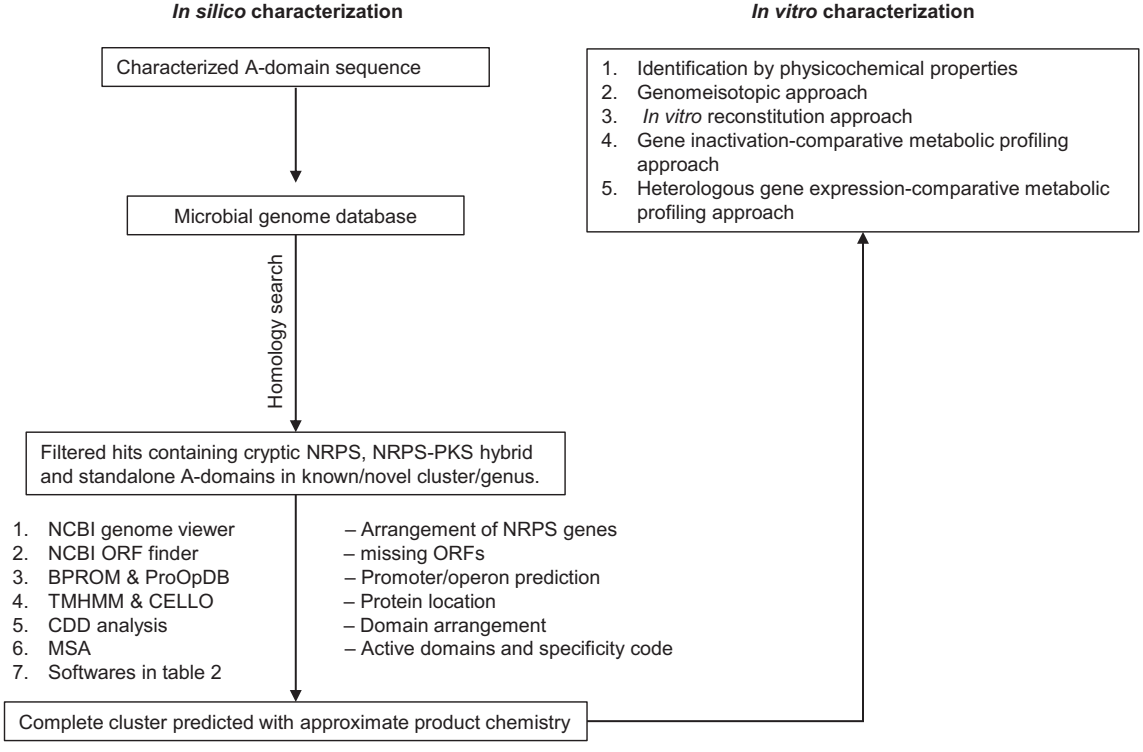

Figure 3. Schematic representation for the identification of novel NRPS clusters. A combined understanding from the manual and automatic screening tools for the NRPS clusters can lead to the closest possible chemistry of the associated product.

to the identification of novel cryptic NRPS clusters present in *S. nassauensis* genome (figure 4). The top hit, Snas_0188 which was 37% identical to the query, was found to be associated with a cryptic NRPS-PKS hybrid cluster present in a single operon (Taboada *et al.* 2012) (Snas_0188-0193) with four NRPS (1-4) modules and one PKS (5) module, besides an intermediating thioesterase present as a separate enzyme (Snas_0190) (figure 4A). In total, five elongation modules corroborated to a five membered product that might be involved in metal chelation or interaction with proteins, DNA or RNA. The respective substrates for the four A-domains and one AT domain, were identified by comparing their specificity code with that present in different databases (table 1). These were predicted as Ile for module 1, substrate for module 2 remained unpredictable, β-ala (derived from Asp) for module 3, Ser for module 4 and malonyl-CoA for module 5. The second (Snas_0191) and the fourth hit (Snas_0192) were 31% and 34% percent identical to the query, respectively and were also found to be the part of the same cluster discussed above. The order in which these 4 NRPSs and 1 PKS act in the NRP assembly was deduced by the identification of the loading module 1 in Snas_0188 (condensation domain being absent); termination module in Snas_0193 because of the presence of $^{L}C_{L}$ type condensation domains in all the other NRPS modules and a hybrid KS domain in this module and by the consideration that the order of remaining modules follow the order of their respective genes in the operon (figure 5). The loading module activates an isoleucine which is transferred to the phosphopantetheine arm of the thiolation domain, followed by the peptide bond formation with a second residue (unpredicted) by the $^{L}C_{L}$ domain of the module 2. The module 3 activates a non-natural amino acid, β-alanine which is probably generated from L-aspartate by the decarboxylation reaction catalysed by aspartate-1-decarboxylase (Snas_0195) present in the cluster (Miyanaga *et al.* 2016). The module 4 activates a serine residue which is further cyclized to a five membered oxazoline ring by the cyclization domain of the module. The cyclization domain was identified on the basis of DxxxxD motif present, instead of the HHxxxDC motif present in $^{L}C_{L}$ and $^{D}C_{L}$ domain (Rausch *et al.* 2007). We identified these motifs manually, though NaPDoS, antiSMASH and PRISM predicted it incorrectly as $^{D}C_{L}$ leading to non-consideration of serine cyclization to oxazoline ring in the structure predicted by these tools. The oxazoline ring further gets reduced to oxazolidine by a *trans*-reductase, as no cis reductase was present in any of the modules. This reduction step makes the oxazolidine ring nitrogen basic enough to act as a nucleophile, which is then N-methylated by the cis N-methyltransferase present in the module (Ansari *et al.* 2008). The genes responsible for carrying out the redox reaction in NRPS and PKS clusters are cytochrome P450/dehydrogenases, which were found present upstream of the NRPS-PKS operon and can thus work in trans for the NRP assembly. Finally, the assembled product is released as a linear or cyclized product by the *trans*-acting thioesterase (Snas_0190) domain. A major facilitator superfamily (MFS) transporter (Snas_0189) was found within this operon that might have a role in the transport of the putative siderophore (Chatfield *et al.* 2012).

Following in the sequence of these hits, the third hit (Snas_4054), was 34% identical to the query and was associated with a second cryptic NRPS-PKS hybrid cluster (Snas_4054-57) with two NRPS, one PKS and a thioesterase (figure 4B). Besides the modular genes, the operon included genes for an acyl-CoA dehydrogenase containing protein (Snas_4058) which might introduce a trans double-bond between C2 (α) and C3 (β) of the malonyl-CoA thioester substrate of the PKS module and an unusual lantibiotic dehydratase domain (Snas_4059), with two additional ABC transporter determinants (Snas_4060 and Snas_4061) (Taboada *et al.* 2012), that might be involved in the extracellular efflux of the assembled NRP. It will be intriguing to know what role the lantibiotic dehydratase domain have to play in the NRP synthesis. The domain is generally involved in dehydration reaction in lantibiotics biosynthesis. Lantibiotics are ribosomally synthesized antimicrobial peptides produced by a diverse array of bacteria and includes nisin, subtilin, etc. Lantibiotic peptides contain thioether bridges termed lanthionines that are generated by dehydration of serine and threonine residues followed by addition of cysteine residues (Yu *et al.* 2013). An MbtH homolog determinant of 68 residues (Snas_4056) was found in the cluster with 49% identity to dptG, the MbtH homolog important for the production of the antibiotic daptomycin (Cubicin®) in *Streptomyces roseosporus*, which is used as a probe for the identification of novel cryptic NRPS clusters (Baltz 2011a, 2014). The next hit Snas_1151 was 27% identical to the query and was associated with a non-conventional NRPS-PKS domain hybrid cluster (figure 4C; Snas_1143-52). Besides the D-alanine activating A-domain, the operon included a FAD-dependent oxidoreductase and PKS annotated gene with KS and PCP domains. The two genes (Snas_1143 and 44) located five genes upstream were two consecutive AT domains that seemed related to this operon as well. Using the same strategy with a KS domain as a query, our *in silico* analysis revealed a cryptic type I PKS cluster (figure 4D; Snas_4132 – 4135) with 5 modules present in a single operon. The cluster could be involved in the production of a novel macrolide like erythromycin, from three malonyl and one methylmalonyl precursors. The first module similar to NRPS comprised of only the AT and T domain for activation and tethering of the substrate. In some cases a KS domain is often found in the first module which is usually mutated to a non-functional KS domain. The last module always comprises a thioesterase, which in this case was also present as a separate enzyme among other modules, as was the case with the NRPS-PKS hybrid cluster discussed above.

S. nassauensis DSM 44728 genome contains multiple NRPS-PKS clusters (O'Brien *et al.* 2014) and till date no associated product has been predicted or identified from these

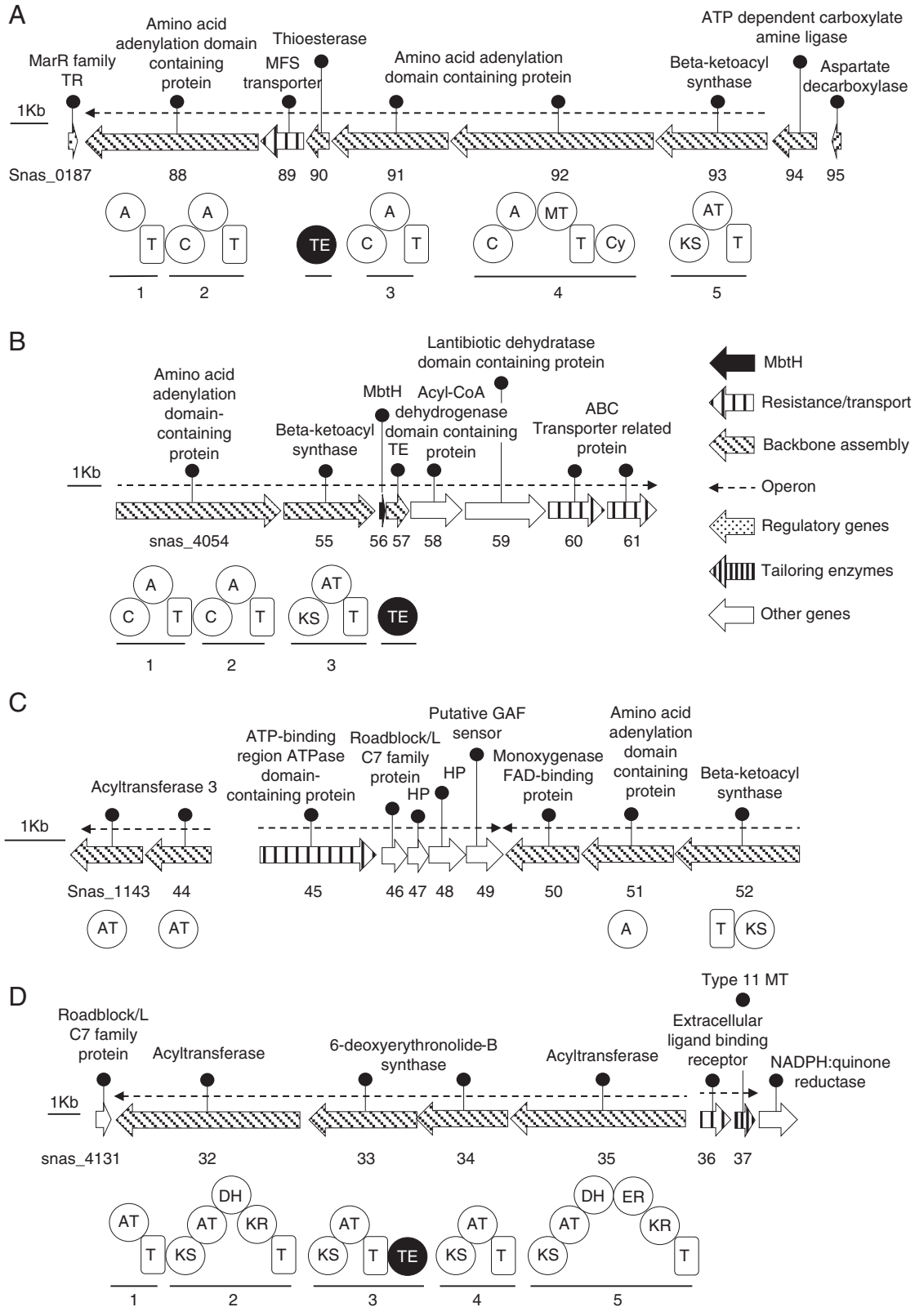

Figure 4. NRPS and PKS clusters identified in the genome of *Stackebrandtia nassauensis* DSM 44728 by the mining strategy outlined in figure 3. HP – hypothetical protein, TR – transcriptional regulator.

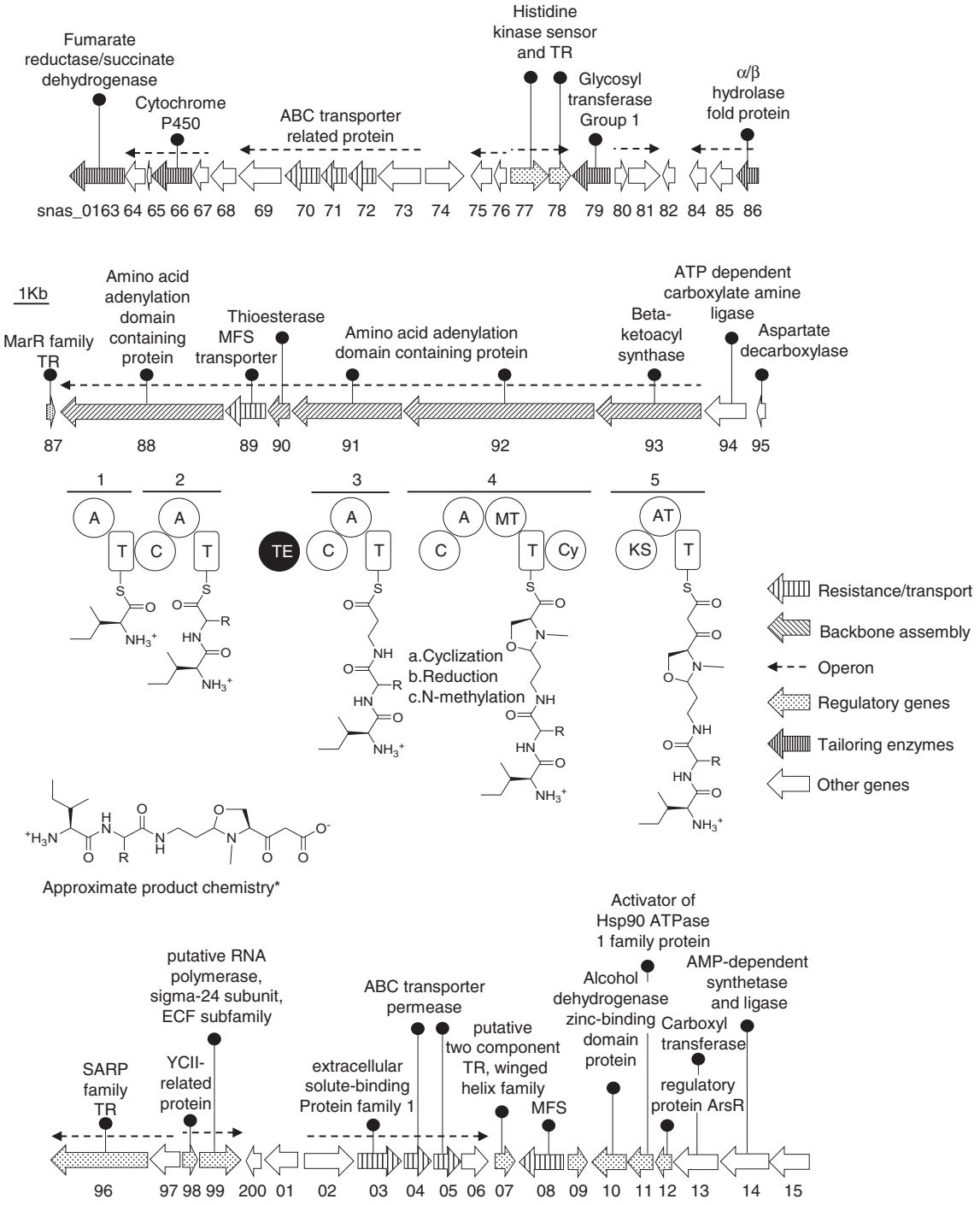

Figure 5. NRP assembly line and prediction of the associated product in one of the NRPS clusters identified in *S. nassauensis* DSM 44728. *with predicted modifications.

clusters. The identified cluster three (figure 4C), which has not been reported in the literature, has a putative stand-alone adenylating protein *snas_1151*. This adenylating protein is being taken up for activity analysis in our lab, as its presence in an incomplete cluster was intriguing for us. The gene has therefore been cloned, overexpressed and purified to characterize and assign its potential role in this or some other cluster. Prediction of the substrate for this adenylating protein could

not be achieved with confidence using the available *in silico* analysis tools, due to its novel specificity code (Lombó *et al.* 2006; Xia *et al.* 2012; Wang *et al.* 2014). This problem is at times remedied with the manual identification of the specificity code by multiple sequence alignment with the characterized proteins, followed by *in vitro* studies (Wang *et al.* 2014) based on the adenylation activity analysis. Adenylation activity is being assayed in our lab by the malachite green-pyrophosphatase assay to estimate pyrophosphate, which is formed as a by-product of adenylation reaction (unpublished data) (Geladopoulos *et al.* 1991; Mcquade *et al.* 2009).

3. Growing importance of adenylating enzymes

Adenylating proteins of ANL superfamily play an important role in biochemical processes as they possess a catalytic mechanism of its own kind. Adenylation protein of NRPSs has continued to be a topic of great interest as it determines the aminoacyl/aryl acid specificity, to be incorporated into the final natural product. Efforts are going on to modify this protein to understand the role of A-domain in diversity of NRPs. Recently, taking a giant leap towards the goal, a single mutation Try239Ser in phenylalanine-specific A-domain (PheA) of the gramicidin S synthetase (GrsA) altered its specificity from phenylalanine to accommodate the bulkier para substituted phenylalanine derivatives (Kries *et al.* 2014). The adenylation rates were not only found to increase with the size of the para-substituent in the order H<OH<Cl<OMe<OEt, but the now enlarged binding pocket of PheA could also accommodate the Clickable amino acids like p-azido-L-Phe and *O*-propargyl-L-Tyr, without any change in the catalytic efficiency (Kries *et al.* 2014). Clickable amino acids are non-natural amino acids having azide and alkyne functionality which can undergo bioorthogonal click reactions to enable isolation, labelling and modification of the molecule where introduced (in this case NRPs). Thus, a single point mutation not only successfully reprogrammed GrsA and TycA A-domain (from Tyrocidin A synthetase), but also demonstrated the permissivity of the downstream thiolation, condensation and epimerization domains for the modified substrate. These advancements indicate that A-domains with altered specificity have a vast potential as tools for creating unnatural assembly lines to produce completely novel NRPs. Many groups of researchers are moving out of the conventional model of NRPS and are working over what they call 'NRPS-like protein' (Forseth *et al.* 2013). Recently, a group has reported the identification of such a cluster, which consists of a stand-alone A-domain, acyl carrier protein and a reductase under one module with an NR-PKS (Non-reducing-PKS) gene in the near vicinity in the fungal genome of *Aspergillus terreus* (Wang *et al.* 2014). They have reported a distinct mechanism of aryl-aldehyde formation by cloning and heterologous expression of both the cryptic NRPS-like and NR-PKS genes, in which an NRPS-like protein activates and then reduces an aryl-acid to aryl-aldehyde, produced by the neighbouring NR-PKS to possibly produce a compound similar to cichorine in *Aspergillus terreus*. Hence, the identification of cryptic biosynthetic gene clusters using bioinformatics tools paves the way for identifying novel producers and structure prediction of novel natural products.

Acknowledgements

MS acknowledges the independent Senior Research Fellowship (SRF) No. 3/1/3/JRF-2011/HRD-99(11005), awarded by the Indian Council of Medical Research, New Delhi. The financial assistance received from DST-PURSE grant, Department of Science and Technology, and UGC-SAP, New Delhi, are also acknowledged.

References

Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, Ansari MZ and Mohanty D 2010 SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.* **38** W487–W496

Ansari MZ, Yadav G, Gokhale RS and Mohanty D 2004 NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32** W405–W413

Ansari MZ, Sharma J, Gokhale RS and Mohanty D 2008 *In silico* analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinf.* **9** 454

Bachmann BO and Ravel J 2009 Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. In Methods in Enzymology. Elsevier Inc., pp. 181–217.

Baltz RH 2011a Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38** 1747–1760

Baltz RH 2011b Strain improvement in actinomycetes in the post-genomic era. *J. Ind. Microbiol. Biotechnol.* **38** 657–666

Baltz RH 2014 MbtH homology codes to identify gifted microbes for genome mining. *J. Ind. Microbiol. Biotechnol.* **41** 357–369

Baranašić D, Zucko J, Diminic J, Gacesa R, Long PF, Cullum J, *et al*. 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

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E and Weber T 2013 antiSMASH 2.0 - a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41** 204–212

Boll B, Taubitz T and Heide L 2011 Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem.* **286** 36281–36290

Bouslimani A, Sanchez LM, Garg N and Dorrestein PC 2014 Mass spectrometry of natural products: current, emerging and future technologies. *Nat. Prod. Rep.* **31** 718–729

Buchko GW, Kim C-Y, Terwilliger TC and Myler PJ 2010 Solution structure of Rv2377c-founding member of the MbtH-like protein family. *Tuberculosis.* **90** 245–251

Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P and Kucherov G 2008 NORINE: A database of nonribosomal peptides. *Nucleic Acids Res.* **36** 326–331

Cane DE and Walsh CT 1999 The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* **6** R319–R325

Challis GL 2008 Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology.* **154** 1555–1569

Chatfield CH, Mulhern BJ, Viswanathan VK and Cianciotto NP 2012 The major facilitator superfamily-type protein LbtC promotes the utilization of the legiobactin siderophore by *Legionella pneumophila*. *Microbiology* **158** 721–735

Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM and Gulick AM 2007 The 1.8 A° crystal structure of PA2412, an MbtH-like protein from the Pyoverdine cluster of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **282** 20425–20434

Felnagle EA, Barkei JJ, Park H, Podevels AM, Matthew D, Drott DW and Thomas MG 2010 MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49** 8815–8817

Fischbach MA and Walsh CT 2006 Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106** 3468–3496

Forseth RR, Amaike S, Schwenk D, Affeldt KJ, Hoffmeister D, Schroeder FC and Keller NP 2013 Homologous NRPS-like gene clusters mediate redundant small-molecule biosynthesis in *Aspergillus flavus*. *Angew. Chem. Int. Ed. Engl.* **18** 1199–1216

Geladopoulos TP, Sotiroudis TG and Evangelopoulos A 1991 A malachite green colorimetric assay for protein phosphatase activity. *Anal. Biochem.* **192** 112–116

Gocht M and Marahiel MA 1994 Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. *J. Bacteriol.* **176** 2654–2662

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

Haynes SW and Challis GL 2007 Non-linear enzymatic logic in natural product modular mega-synthases and -synthetases. *Curr. Opin. Drug Discov. Dev.* **10** 203–218

Helfrich EJN and Piel J 2016 Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33** 231–316

Herbst DA, Boll B, Zocher G, Stehle T and Heide L 2013 Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288** 1991–2003

Hur GH, Vickery CR and Burkart MD 2012 Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29** 1074–1098

Imker HJ, Krahn D, Clerc J, Kaiser M and Walsh CT 2010 N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem. Biol.* **17** 1077–1083

Keller U, Lang M, Crnovcic I, Pfennig F and Schauwecker F 2010 The actinomycin biosynthetic gene cluster of *Streptomyces chrysomallus*: a genetic hall of mirrors for synthesis of a molecule with mirror symmetry. *J. Bacteriol.* **192** 2583–2595

Khayatt BI, Overmars L, Siezen RJ and 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

Kleinkauf H and Von Döhren H 1996 A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.* **236** 335–351

Kries H, Wachtel R, Pabst A, Wanner B, Niquille D and Hilvert D 2014 Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angew. Chem. Int. Ed. Engl.* **53** 1–5

Krug D and Müller R 2014 Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **31** 768–783

Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, *et al*. 1996 A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3** 923–936

Lautru S, Oves-Costales D, Pernodet J-L and Challis GL 2007 MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology.* **153** 1405–1412

Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S and Sherman DH 2009 Automated genome mining for natural products. *BMC Bioinf.* **10** 185

Lombó F, Velasco A, Castro A, de la Calle F, Braña AF, Sánchez-Puelles JM, *et al*. 2006 Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. *ChemBioChem.* **7** 366–376

McMahon MD, Rush JS and Thomas MG 2012 Analysis of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194** 2809–2818

Mcquade TJ, Shallop AD, Sheoran A, Delproposto JE and Tsodikov OV 2009 A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal. Biochem.* **386** 244–250

Medema MH and Fischbach MA 2015 Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11** 639–648

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, *et al*. 2011 antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39** W339–W346

Minowa Y, Araki M and Kanehisa M 2007 Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* **368** 1500–1517

Miyanaga A, Hayakawa Y, Numakura M, Hashimoto J, Teruya K, Hirano T, *et al*. 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 Adenylating Enzyme FlvN. *Biosci. Biotechnol. Biochem.* **8451** 1–7

Mootz HD and Marahiel MA 1997 The tyrocidine biosynthesis operon of *Bacillus brevis*: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179** 6843–6850

Munk C, Lapidus A, Copeland A, Jando M, Mayilraj S, Glavina Del Rio T, *et al*. 2009 Complete genome sequence of *Stackebrandtia nassauensis* type strain (LLR-40K-21). *Stand. Genomic Sci.* **1** 234–241

O'Brien RV, Davis RW, Khosla C and Hillenmeyer ME 2014 Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiot.* **67** 89–97

Ogasawara Y, Yackley BJ, Greenberg JA, Rogelj S and Melançon CE 2015 Expanding our understanding of sequence-function relationships of Type II polyketide biosynthetic gene clusters: Bioinformatics-guided identification of Frankiamicin A from *Frankia* sp. EAN1pec. *PLoS ONE* **10** 1–25

Prieto C, García-estrada C, Lorenzana D and Martín JF 2012 NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics* **28** 426–427

Quadri LE, Sello J, Keating TA, Weinreb PH and Walsh CT 1998 Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5** 631–645

Rausch C, Weber T, Kohlbacher O, Wohlleben W and Huson DH 2005 Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33** 5799–5808

Rausch C, Hoof I, Weber T, Wohlleben W and Huson DH 2007 Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7** 78

Reger AS, Wu R, Dunaway-mariano D and 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

Röttig M, Medema MH, Blin K, Weber T, Rausch C and Kohlbacher O 2011 NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39** 362–367

Singh M and Sareen D 2014 Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS ONE* **9** e91352

Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, *et al*. 2015 Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* **43** 9645–9662

Stachelhaus T and Marahiel MA 1995 Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125** 3–14

Stachelhaus T, Mootz H and Marahiel M 1999 The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6** 493–505

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J and Hranueli D 2008 ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. *Nucleic Acids Res.* **36** 6882–6892

Staunton J and Wilkinson B 1997 Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev.* **97** 2611–2630

Stegmann E, Rausch C, Stockert S, Burkert D and Wohlleben W 2006 The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. *FEMS Microbiol. Lett.* **262** 85–92

Taboada B, Ciria R, Martinez-Guerrero CE and Merino E 2012 ProOpDB: Prokaryotic operon database. *Nucleic Acids Res.* **40** 627–631

Wang Y, Zhi X, Zhang Y, Cui X, Xu L and Li W 2009 *Stackebrandtia albiflava* sp. nov. and emended description of the genus *Stackebrandtia*. *Int. J. Syst. Evol. Microbiol.* **59** 574–577

Wang M, Beissner M and Zhao H 2014 Aryl-aldehyde formation in fungal polyketides: Discovery and characterization of a distinct biosynthetic mechanism. *Chem. Biol.* **21** 257–263

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, *et al*. 2015 antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **39** W339–W346

Weissman KJ and Müller R 2008 Protein-protein interacions in multienzyme megasynthetases. *Chem. Bio. Chem.* **9** 826–848

Wolpert M, Gust B, Kammerer B and Heide L 2007 Effects of deletions of MbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiology* **153** 1413–1423

Xia S, Ma Y, Zhang W, Yang Y, Wu S, Zhu M, *et al*. 2012 Identification of Sare0718 as an alanine-activating adenylation domain in marine actinomycete *Salinispora arenicola* CNS-205. *PLoS ONE.* **7** e37487

Yeh E, Kohli RM, Bruner SD and Walsh CT 2004 Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem.* **5** 1290–1293

Yu Y, Zhang Q and van der donk WA 2013 Insights into the evolution of lanthipeptide biosynthesis. *Protein Sci.* **22** 1478–1489

Zhang W, Heemstra JR Jr, Walsh CT and Imker HJ 2010 Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry* **49** 9946–9947

Ziemert N, Podell S, Penn K, Badger JH, Allen E and Jensen PR 2012 The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **7** e34064

Zolova OE and Garneau-Tsodikova S 2012 Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *Med. Chem. Commun.* **3** 950

MS received 25 August 2016; accepted 30 December 2016

Corresponding editor: Debasisa Mohanty