J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEW ARTICLE

Fatty Acyl-AMP Ligases as Mechanistic Variants of ANL Superfamily and Molecular Determinants Dictating Substrate Specificities

*Kinatukara Priyadarshan and Rajan Sankaranarayanan**

Abstract | Fatty acyl-AMP ligases (FAAL) are enzymes that establish the crosstalk between fatty acid synthases and polyketide synthases or non-ribosomal peptide synthetases by activating newly synthesized fatty acids, shuttling them towards polyketide synthesis. These enzymes are unique as they have evolved a strict rejection mechanism for the coenzyme-A but can recognize and react with the phosphopantetheine of acyl-carrier proteins. A strategically placed insertion in the N-terminal domain and the rigidity of the hydrophobic anchor holding the insertion is at the heart of such a discrimination mechanism, the molecular details of which is yet to be clearly understood. The unique structural features of the insertion have been exploited to filter out FAALs from other members of the superfamily, in silico. Interestingly, several independent studies have characterized FAALs from different organisms such as *Legionella*, *Myxococcus*, *Ralstonia*, *Pseudoalteromonas*, *Sorangium*, etc. to name a few, laying emphasis on the usage of a FAAL-specific biochemistry in these organisms, particularly for the production of important bioactive molecules. These bioactive molecules help in improving the fitness of these systems to tide over the competition for nutritional resources in their local niche. Substrate specificity of these enzymes is another interesting aspect, which may hint at their potential roles in the cell. Various substrate-bound crystal structures have been used to identify the determinants of chain-length specificity and predict the preferences for different FAALs (both mycobacterial and non-mycobacterial). The fascinating details of the mechanistic variation of these enzymes and the molecular determinants that define the chain-length specificity have been discussed herewith.

1 Introduction

Adenosine triphosphate (ATP) hydrolysis is a very commonly used strategy in many biochemical reactions to overcome the energy-barrier challenges. Several enzymes acting on substrates with carboxylate groups often rely on the ATP hydrolysis event because of the poorly reactive nature of a carboxylate group. The coupling of ATP hydrolysis is said to "activate" these substrates, simply to indicate that the adduct of the carboxylate group with the adenosine monophosphate group through a phosphoester bond is extremely reactive. The resulting product is often referred to as an acyl-adenylate which is a very good "leaving group" owing to the labile phosphoester bond, which facilitates the subsequent nucleophilic attack by either an amine, alcohol, or a thiol moiety or even the molecular oxygen. Such an elaborate reaction is catalyzed by a group of enzymes called adenylate-forming enzymes (AFE), which

[1] *CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India.*
**sankar@ccmb.res.in*

Springer IISc Press

has been extensively reviewed elsewhere.[1] There are three known independently evolved biochemical solutions to help a feebly reactive carboxylate moiety to form a labile acyl-adenylate and eventually facilitating their participation in the enzymatic reaction, which has been classified into three classes. The classification describes Class-I AFEs [adenylation domains from non-ribosomal peptide synthetases (NRPS) and Acyl/Aryl-CoA ligases; ANL superfamily], Class-II AFEs (aminoacyl-tRNA synthetases; aaRS) and Class-III AFEs (NRPS-independent siderophores synthetases; NIS). Class-I AFEs act on a wide range of substrates such as simple alkyl and aryl acids (fatty acids, amino acids, benzoic acids, and derivatives) and multi-ringed acids and lactones (coumarins and luciferins), etc., to name a few, while Class-II is known to act exclusively on amino acids and Class-III acts on dicarboxylic acid substrates such as citrate. These enzymes contribute to some of the most basic cellular processes such as translation, lipid metabolism, cofactor synthesis, etc. The fundamental similarity lies in that all the three systems form an acyl-adenylate, but the differences lie in the way they stabilize the adenylates and where they transfer the acyl-chains. Thus, adenylation-based activation is very common in biology that catalyzes an unusually difficult, but essential reaction on a myriad of substrates for the sustenance of life.

Recently, Estrada et al. and Wang et al. have independently characterized a novel structural fold in pimeloyl-CoA synthetase (BioW) that produces the adenylate of a dicarboxylic acid called pimelic acid, the building block for biotin (vitamin B7) biosynthesis.[2, 3] It is also interesting to note that BioW brings together the ability of Class-I AFEs to form thioesters after adenylation and Class-II AFEs to exercise proofreading of their respective products to ensure fidelity. Thus, variations of the existing themes are not uncommon in these structural folds of enzymes. The current review focuses on one such variant called FAALs that belong to Class-I AFEs. The unique structural and mechanistic aspects that differentiate the FAALs modus operandi from other members of Class-I AFEs such as adenylation domains (A-domains) and Fatty acyl-CoA ligases (FACLs) are discussed herewith. The sequence–structure relationship of such variants is also discussed which has hinted at their substrate specificities, and can possibly shed light on their function in many of the orthologs identified across different forms of life.

2 Structural and Mechanistic Principles of ANL Superfamily of Enzymes

ANL superfamily of adenylating enzymes includes three main subfamilies, namely, the Acyl/Aryl-CoA synthetases, the A-domains of NRPS, and the luciferase enzymes. The Acyl/Aryl-CoA synthetases and the A-domains of NRPS utilize the adenylate intermediate in the formation of thioesters, while luciferases utilize it for oxidation with molecular oxygen. Luciferases are the most commonly used biochemical tools in the form of a reporter gene for a wide variety of cellular and molecular biology experiments. The success and widespread use of these enzymes is because luciferase uses biomolecules available abundantly within the cell such as ATP and molecular oxygen to first form adenylate of luciferins (literally translates in Latin to "light-bearers") which is then converted to oxidized intermediates that emit light (termed as bioluminescence).[4–7] Unlike luciferases, the other members of ANL superfamily transfer the formed adenylate to 4′-phosphopantetheine (4′-PPant) moiety through a thioester bond with CoA or CoA-derived post-translational modification of an acyl-carrier protein (ACP). The resultant thioester derivatives are essential for both anabolic processes such as fatty acid biosynthesis and catabolic processes such as β-oxidation of fatty acids along with other important biological processes such as membrane lipid synthesis, secondary messengers, polyketide (related bioactive small molecules) synthesis, etc. (Fig. 1a).

The earliest member of the superfamily to be studied was acetyl-CoA synthetase, about 7 decades ago by Fritz Lipmann,[9] and hence, the superfamily was earlier named as "Acetyl-CoA synthetase superfamily". These enzymes convert an acetate to acetyl-adenylate (acetyl-AMP) and then transfer it to coenzyme-A to form acetyl-CoA in a two-step reaction utilizing one molecule of ATP and releasing one pyrophosphate moiety (PPi). It was also proposed later that large conformational changes are the central hallmark of the catalysis through a series of biochemical studies on different members of this superfamily.[10–13] Ten conserved motifs (referred to as A1–A10-motifs) were identified, which play an important role in either maintaining the architecture (structural role) or the enzymatic function (biochemical role).[14–16] The conformational changes and the conserved motifs could repetitive use of only be rationalized only after the crystal structures of these proteins were solved in multiple conformations.[17–19] The crystal structure of

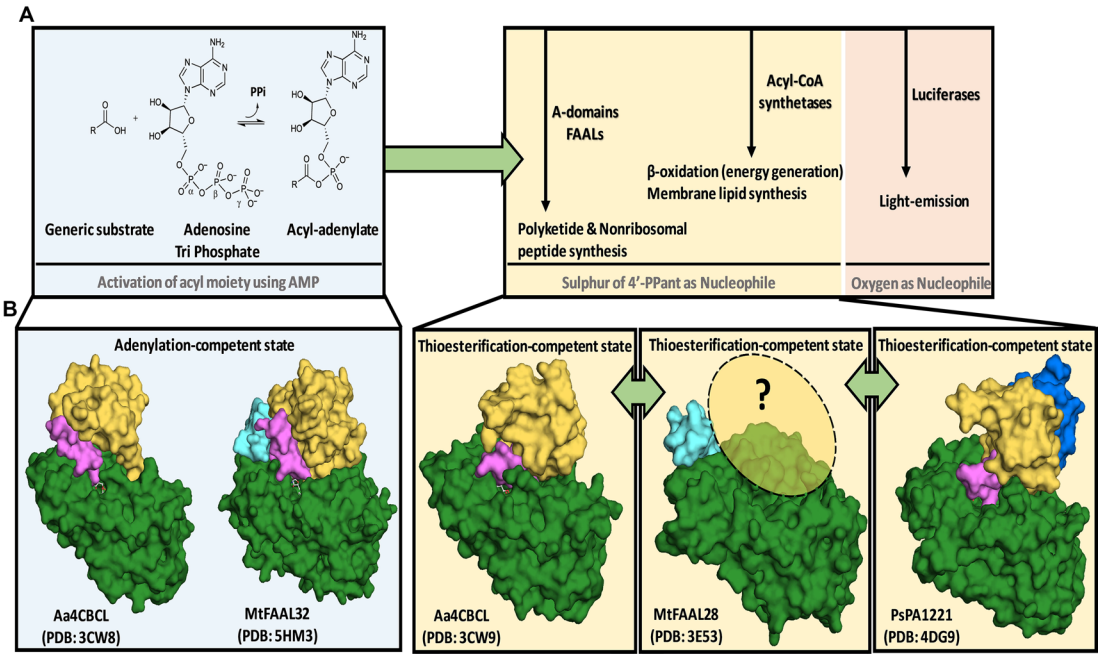


Figure 1: Modus operandi (structural and biochemical) of the ANL superfamily of enzymes is described. A canonical two-domain architecture supports the adenylation reaction (blue panel, top), where the specific conformation of the C-terminal domain (orange), A8-motif (violet) and the N-terminal domain (green) are defined as the Adenylation-competent conformation or A-state (blue panel) is shown. Large-scale domain rotation reorganizes the active site and allows the reaction with a nucleophile, a sulphur, or an oxygen (orange panel), in a conformation referred to as thioesterification-competent conformation or T-state (orange panel). The presence of FAAL-specific insertion (cyan) is the basis for rejection of CoASH, while the orientation of the C-terminal domain or A8-motif in case of FAALs is a mystery. 4-chlorobenzoyl-CoA ligase (4-CBCL; PDB: 3CW8) and MtFAAL32 (PDB: 5HM3) represent the A-state, while 4-chlorobenzoyl-CoA ligase (4-CBCL; PDB: 3CW9) and PA1221 (PDB: 4DG9; with its ACP shown in dark blue) represent the T-state. All the structural figures discussed were rendered using PyMOL[8].

Photonis Luciferase in 1995, the first crystal structure of the superfamily helped shed light on the fold and the unique two-domain architecture, which clearly outlined the potential to exhibit conformational changes during the chemical reaction.[20] Later, over the years, several crystal structures were solved in different conformations which clearly suggested an elegant ballet of conformational dynamics in tune with the biochemistry of the enzyme. For instance, the two crystal structures of 4-chlorobenzoyl-CoA ligases in adenylation-competent conformation or A-state (PDB: 3CW8) and thioesterification-competent conformation or T-state (PDB: 3CW9) led to the proposition of domain alternation mechanism[21, 22] (Fig. 1b). The proposition states that the C-terminal alters its orientation against the N-terminal domain during the two-step catalysis, facilitating reconfiguration of residues facing the common active site at the interface, which utilizes an identical-binding pocket. Currently, there are about 200 crystal structures of more than 50 different proteins in the PDB defining the major conformations sampled by the protein as a part of their catalysis. The two major conformational states defined by ANL members are referred to as "Adenylation-competent state" (A-state) and "Thioesterification-competent state" (T-state). The A-state has A8-motif, one of the ten conserved motifs, in a β-hairpin architecture that orients C-terminal domain, such that a conserved catalytic lysine of the A10-motif is poised for the formation of acyl-adenylate. The T-state, on the other hand, has A8-motif also in a β-hairpin architecture that orients C-terminal domain (140° away from A-state), such that an invariant glycine of A8-motif interacts with the –NH of the β-alanine moiety in the 4′-PPant, while forming the PPant-binding tunnel. Similar interactions are observed when PPant of ACP also binds in case of A-domains, suggesting a similar mode of operations in A-domains.[23] Structure solution of enzymes trapped by proximity-based chemical cross-linking using maleimide allowed capturing the enzymes in action during the biochemical reaction, and these crystal structures showed

A-state or T-state, suggesting that these states are not crystallization artefacts.[24] Thus, an interplay of the domain dynamics allows the enzyme system to change the residues at the active site, eventually performing an important biochemical reaction.[16]

3 Modulation of Domain Dynamics in FAALs Makes Them Different from FACLs

Fatty acyl-activating enzymes can easily be identified by simple sequence-based searches, where the similarity in the A1 to A10 canonical motifs can allow identification of these enzymes. Sequence analysis in *Mycobacterium tuberculosis* revealed homologous systems and subsequent biochemical analysis revealed the existence of the fascinating variants called FAALs. Canonical FACLs bind to a fatty acid of defined length and unsaturation along with ATP activating them to form acyl-AMP in A-state. Following a C-terminal domain rotation, a cofactor-binding pocket is formed where CoASH binds facilitating the formation of acyl-CoA, simultaneously liberating PPi in the T-state. Gokhale and co-workers initially showed that unique systems called FAALs form acyl-AMP but do not react with CoASH.[25] It was not clear why FAALs operated by bypassing the canonical role of reaction with CoASH until the first crystal structure of the N-terminal domain of MtFAAL28 was solved.[26]

The crystal structure showed an FAAL-specific insertion (FSI) in the C-subdomain of the N-terminal domain, which seemed to indicate that it may obstruct the canonical domain rotations preventing the enzyme from achieving a T-state (Fig. 2). It was eventually demonstrated that deletion of the FSI in FAALs allowed, although weak, acyl-AMP reaction with CoASH. Interestingly, it was found that the same FAALs that could not react with PPant of CoASH, could transfer the acyl-AMP to PPant of ACP domains of multimodular multi-domain PKS and NRPS. The ability of FAAL to discriminate between PPant from CoASH and ACP was attributed to the FSI, which allowed the enzyme to attain the functional T-state only in the presence of ACP, while the small diffusible CoASH was incapable of doing the same. Such a hypothesis was also supported by the fact that FAALs are often found in the same operon or proximity of ACP-carrying PKS and NRPS[25] or even stand-alone ACPs.[29] Some of the recently characterized FAALs such as FAAL domain in micC (biosynthesis of micacocidin)[30] and FAAL-ACP didomains in AmbG (biosynthesis of Ambruticin)[31] or TamA (biosynthesis of tambjamine)[32] are fused with ACP or into the PKS as a single polypeptide replacing the canonical adenylation domains. However, it still remains

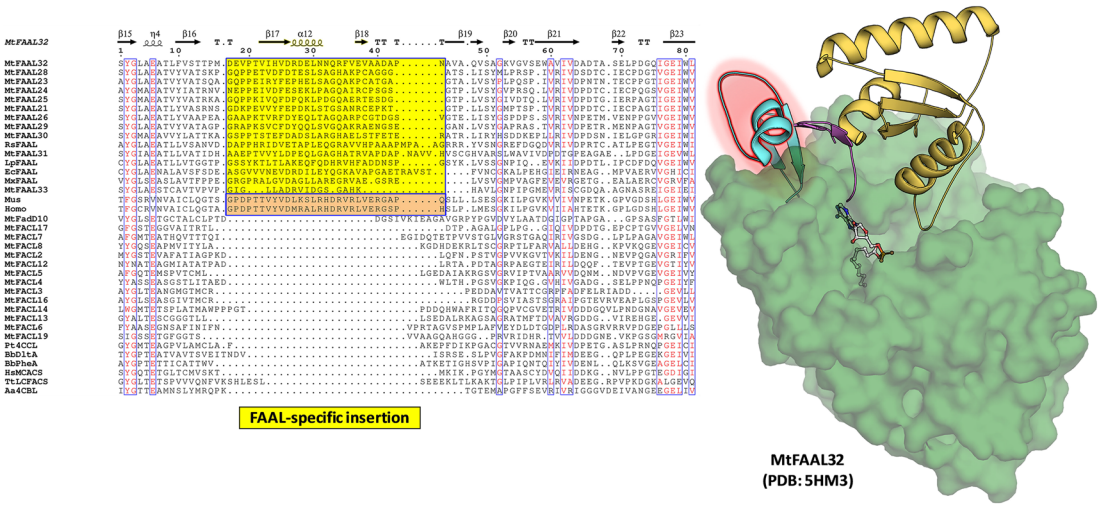


Figure 2: FAAL-specific insertion is found only in FAALs but missing in other members of the ANL superfamily as shown in the structure-based sequence alignment (highlighted in a box). A representative structure of an FAAL; FAAL32 from *Mycobacterium tuberculosis* is shown for reference (PDB: 5HM3). The surface view of the N-terminal domain (green) is presented, while the FAAL-specific insertion (cyan) and the C-terminal domain (orange) are shown in cartoon representation. The insertion presents a potential obstruction to domain rotation which is highlighted in red. The alignment was generated using MAFFT server[27] and rendered using ESPript server[28].

to be understood, how the T-states are finally achieved and how the FSI-based obstructions in domain rotations are bypassed in the presence of ACP.

It should also be mentioned that the rejection of CoASH and maintenance of fidelity for ACP in FAALs may happen through unexpected mechanisms. An alternate proposition comes in the form of the crystal structure of FadD10 from *M. tuberculosis*, which lacks FSI but still preferentially reacts with PPant of ACPs (Fig. 3a). The dimerization-based interlocking of domains with a buried surface area of 7% of the surface area (1588 out of 22,018 Å^2), as calculated by PISA,[33] has been proposed to keep these enzymes from moving to the T-state, resulting in their inability to react with CoASH but capable of reacting with PPant of ACPs. It is surprising that FadD10 forms dimers, considering that most of the ANL superfamily members known till date exist only as monomers except the very-long-chain fatty acid synthetases from *Thermus thermophilus* HB8 owing to a specific set of residues favouring intermolecular contacts, which is uncommon in ANL superfamily of enzymes.[17] The crystal structure of a FAAL from *Legionella pneumophila* (LpFAAL) also presents an unusual C-terminal domain orientation, which neither resembles the canonical A-state or the T-state.[34] The unusual domain orientation is possibly stabilized by a unique helical conformation of the critical A8-motif, which, otherwise, is always seen in a β-hairpin conformation. The FSI in LpFAAL, usually a two antiparallel β-strands followed by a helix, is seen as a loop with no helical region suggesting some localized melting of FSI architecture (Fig. 3b). The functional relevance of such a localized melting of the insertion architecture or the unusual C-terminal domain orientation still remains to be characterized in the context of the unique ability of FAALs. It still remains to be seen if these observations are merely a crystallization artefact or are they functionally relevant. The LpFAAL crystal structure along with crystal structure of FAAL from *Escherichia coli* O6:H1 (EcFAAL) has been solved with an acyl-adenylate which was captured during purification (lauroyl-adenylate and myristoyl-adenylate in EcFAAL; lauroyl-adenylate in LpFAAL) and also confirmed by mass spectrometry.[34] Similar observations were also

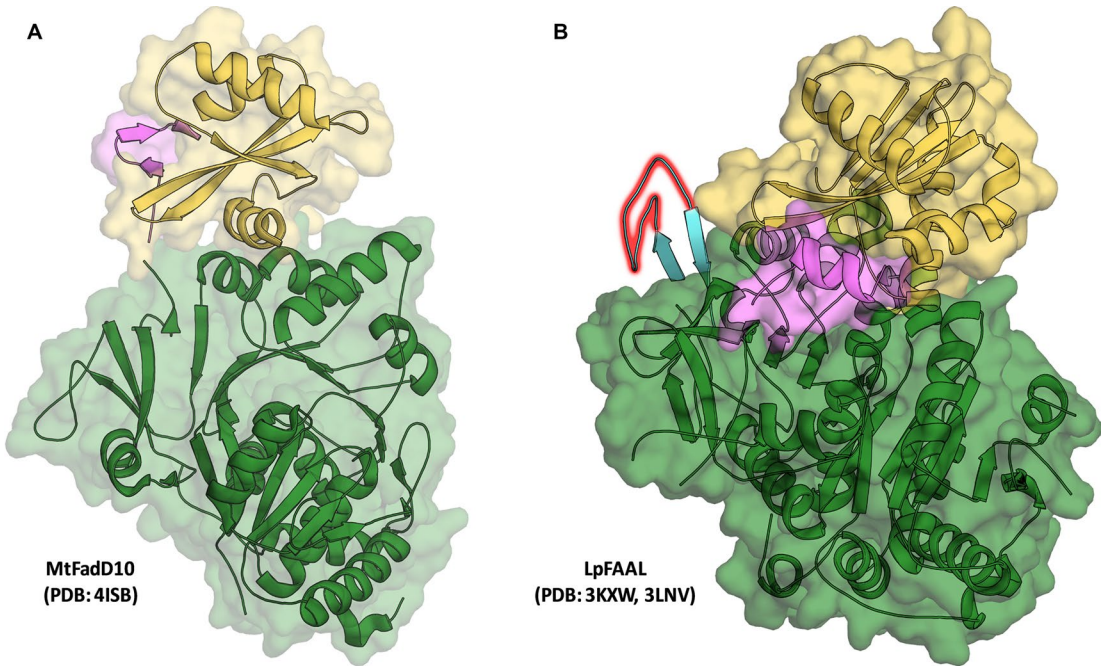


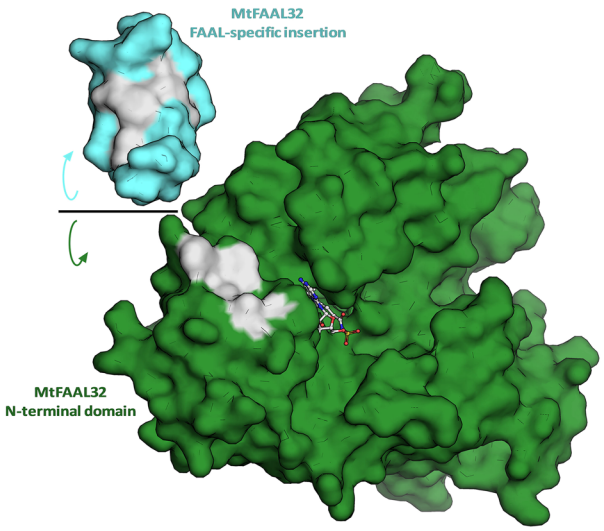
Figure 3: ANL members with unusual relative domain orientations and structural conformations are represented. A protomer from the domain-interlocked structure of FadD10, which is insertion-less but is unable to react with CoASH, from *Mycobacterium tuberculosis* is shown in both cartoon and surface representations (**a**). The second instance shows the unusual relative domain orientation of FAAL from *Legionella pneumophila* (**b**) shown in both cartoon and surface representation. The helical part of the FAAL-specific insertion (cyan) is observed to be disturbed (highlighted in red) which is shown in a cartoon representation.

made with FAAL domains from TamA involved in tambjamine biosynthesis, where it was shown to be co-purified with lauroyl-adenylate by mass spectrometry.[32] It remains to be seen if FAALs are also different in their ability to latch on to the acyl-adenylate, which may have implications on their ability to remain faithful to PKS and NRPS systems. The mechanism by which an FAAL operates, whether by domain-interlocking or destabilizing the FSI along with unusual C-terminal orientation or any alternate modes should form the basis for the fidelity allowing FAALs to distinguish the PPant of ACP from that of CoASH. The current understanding of the biochemical and structural aspects of FAALs suggests that it is very likely that FAALs have evolved to modulate the domain dynamics in interesting ways to achieve its unique ability to remain inert to CoASH. It also follows that FAALs are mechanistic variants of the existing mode of operation seen in an FSI-lacking A-domains, which show a preference for ACP over CoASH, and FACLs, which show no selective preferences.

4 Identification of FAALs Across Different Forms of Life

The FSI in FAALs is clearly playing a central role in dictating its fidelity to PPant in ACP and its architecture, perhaps, is crucial to such a unique mechanism. It should be mentioned that FSI is a part of the C-subdomain of the N-terminal domain, which is the most poorly conserved region of the protein. However, certain structural attributes of the FSI, including the highly non-polar interface between the insertion and the N-terminal domain, are unique to FAALs (Fig. 4).

The absence of FSI in A-domains or FACLs excludes the possibility of the same surface being hydrophobic and indeed, these surfaces are marked by polar residues in a non-FAAL ANL superfamily member. There are two interfaces that an FSI extends across-viz., (1) the helical region of FSI with C-terminal domain and (2) the FSI with the N-terminal domain, the former being a polar interface, while the latter is exclusively a non-polar interface. Structure-guided mutagenesis and biochemical analysis have emphasized that mutations destabilizing the non-polar interface between the FSI with the N-terminal domain can allow FAALs to start reacting with CoASH.[35] These experiments suggest that FSI is rigidly anchored into the N-terminal domain and the rigidity of FSI hinders the domain rotation eventually preventing CoASH binding (as shown in Fig. 2a). It immediately follows that ACP domains, in a hitherto uncharacterized mode, are capable of altering the rigidity of the insertion, thereby facilitating the reaction. It also should be noted that a deletion of the helical region from FSI can also allow the FAALs to start showing mild activity with CoASH,[26] but it



Protein	Residues from FAAL-specific Insertion			Residues from N-terminal domain			
MtFAAL28	F352	L357	A362	V400	F449	F456	I458
EcFAAL	V360	L365	A371	V411	Y448	Y455	T457
LpFAAL	L350	F355	V360	V399	F445	Y452	T454
Segniliparus	V365	L370	V375	V416	V472	Y479	T481
Bradyrhizobium	L342	L347	A352	V392	F438	F446	T448
Synechococcus	V345	I350	V355	I395	F438	Y445	T447
Tsukamurella	V365	L370	V376	V416	V473	Y480	T482
Nostoc	V354	L359	V364	V405	F450	F457	T459
Trichodesmium	V344	F349	L355	V395	F440	F447	T449
Acaryochloris	L350	L355	V360	V401	F446	Y453	T455
Streptomyces	V348	L353	A361	V397	A441	H448	T450
Streptomyces	F373	L378	A386	V424	F469	Y476	V478
Shigella	V357	L362	A368	V408	Y445	Y452	T454
Enterobacteriaceae	V346	L351	A357	I397	Y434	Y441	T443
Citrobacter	V346	L351	A357	V397	Y434	Y441	T443
Erwinia	I357	L362	A368	V408	Y445	Y452	T454
Nitrococcus	V334	L339	A347	V388	Y426	F433	T435
Pelobacter	I360	L365	I372	C413	Y451	V458	T460
deltaproteobacterium	V347	L352	P360	C403	Y441	F448	T450
Nitrosomonas	V383	L388	A394	V434	F472	F479	T481
Zymomonas	V376	L380	A386	V424	Y461	Y468	V470
Myxococcus	V358	L362	V369	V403	F441	F448	T450
Nostoc	V353	L358	V363	I404	F448	F455	T457
Homo	L678	L683	V690	I730	F780	F788	V790
Mus	V1320	L1325	V1330	L1371	F1420	Y1438	V1440

Figure 4: Surface representation of the N-terminal domain (green) and the insertion (cyan) of MtFAAL32 (PDB: 5HM3) are separated and rotated in opposite directions to highlight the unique FAAL-specific non-polar patch (grey region) that anchors the insertion to the N-terminal domain. The residues forming the non-polar patch from different identified organisms are tabulated alongside after structure-based alignment. The table is an adaptation from the previous work[35].

is likely due to the loss of hydrophobic interactions from the helical region that faces the N-terminal domain, rather than loss of polar interface with the C-terminal domain. It is not very clear if the polar interface between the helical regions of FSI with a C-terminal domain can influence the domain rotation mechanisms by latching on the C-terminal domain and can be a subject for further investigation.

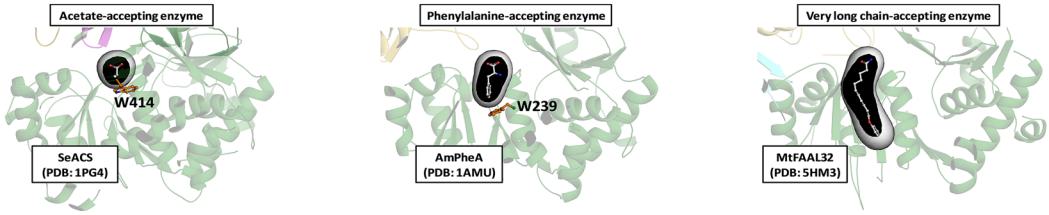
Structure-based sequence alignment of structurally and/or biochemically characterized FAAL and non-FAAL members of the ANL superfamily clearly shows the highly divergent insertion in FAALs along with the hydrophobic interfaces, which are very specific to FAALs. Such unique insertion can be used to filter out FAAL sequences from the rest of the family members, even though the sequence homology is on the lower side (15–25% identity). Previously, a simple sequence search allowed FAALs to be identified only in Mycobacteria or closely related genera such as *Nocardia*, *Rhodococcus,* and a few species of *Streptomyces*, mainly belonging to the Actinobacteria group. The structure-guided bioinformatic analyses, for the first time, revealed that FAAL-like domains are found in many prokaryotes along with eukaryotes. The presence of FSI and the non-polar patch across the identified FAAL-like domains (Fig. 4) is only suggestive, and hence, it still needs to be validated biochemically if the identified domains, indeed, show FAAL activity. The physiological relevance of the FAAL-specific biochemistry (selective preference for PPant of ACP over PPant of CoASH) has been explained in the context of the unique lipid metabolism in a prokaryotic *M. tuberculosis* cell. Fatty acid synthases (FAS) synthesize fatty acids, which can be "activated" for anabolic processes such as mycobacterial lipid biosynthesis by PKS/NRPS systems[36–39] or catabolic processes such as β-oxidation systems by FadAB complexes along with ECHs.[40] Thus, the fate of free fatty acids shows a general dichotomy which may be regulated by controlling relative cellular levels of FAALs and FACLs temporally.[35] Recently, it was shown that regulatory control of mycolic acid biosynthesis overlaps with enzymes of peptidoglycan biosynthesis, including enzymes for cross-linking to tide over environmental stresses.[41] In this context, it will be extremely interesting to understand how FAAL-specific biochemistry has been exploited in eukaryotic systems and how it may reshape the metabolic networks in eukaryotic cells, an investigation currently underway in our laboratory.

5 Substrate-Binding Pocket: The Ruler Defining Specificity in FAALs and FACLs

The current understanding of the structural reorganization process described in the earlier section was made possible through the crystallization of these enzymes with different substrates and/or substrate analogues. Moreover, the knowledge of the substrate specificity of ANL enzymes can be of immense use in defining their possible contribution to a specific function, particularly in the case of enzymes with unknown function. Identification of the determinants of substrate specificity can allow us to rationally alter the specificity, which, therefore, alters the starter molecules, and hence, new product biosynthesis can be engineered. Previously, different groups have worked towards understanding the "substrate-specificity code" of A-domains which dictate the final product biosynthesis by NRPS and PKS domains. Structural analysis of substrate-bound complexes of PheA,[42] the adenylation domain involved in Gramicidin biosynthesis, allowed identification of residues that dictated the specificity code for amino acids (Fig. 5).

The information on residues identification through this process was then extended to molecules of interest whose structures were not available using sequence analysis to identify the "putative binding-pocket signature sequence" that dictates the substrate-specificity code. Subsequently, rational mutagenesis and biochemical analyses also confirmed that these specificities can be swapped through simple point mutation(s).[43] Similar studies were extended to make in silico prediction of the substrate specificity by extracting "binding-pocket signature sequences" in ANL superfamily members on a genome-wide scale.[44]

The first crystal structure of the N-terminal domains of mycobacterial FAALs and FACLs (FAAL28; PDB: 3E53 and FACL13; PDB: 3T5B) served as a template for performing fatty acyl-chain-length specificity analysis to decode the chain-length specificity in FAALs and FACLs previously in our laboratory.[35] The substrate is accommodated within the crevices formed between helices (α7 and α8) and 3 strands (β10, β12, and β12) from the A- and B-subdomains (Fig. 5). It has been observed previously that bulkier residues are placed along the length of strands/helices and serve as a molecular ruler that dictates the specificity of chain length. Thus, identification of bulkier residues along these elements can allow prediction of the chain-length specificity of each of the mycobacterial FAALs and FACLs along with other systems. For instance, enzyme systems such



	221	222	300	330	299	336	223	328	269	337	227	267	297	195	340	198	199	192	213	196	265	237
MtFAAL28	H	D	S	G	G	V	M	S	P	Y	I	A	L	L	T	G	Y	F	L	M	F	Y
MtFAAL21	H	D	A	G	G	L	M	S	P	Y	L	A	T	L	S	N	Y	F	I	M	W	Y
MtFAAL23	H	D	S	G	G	V	M	A	P	Y	L	A	L	M	T	A	Y	F	I	I	F	C
MtFAAL24	H	D	S	G	G	V	M	A	P	Y	L	A	L	I	T	D	F	F	V	V	I	V
MtFAAL25	H	D	S	G	G	V	M	S	P	Y	L	A	L	I	T	S	Y	F	V	M	F	I
MtFAAL26	H	D	S	G	G	L	M	S	P	Y	L	A	V	S	A	G	Y	V	V	M	F	R
MtFAAL29	H	D	A	G	G	V	M	S	P	Y	L	S	V	L	T	D	Y	C	P	M	I	D
MtFAAL30	H	D	A	G	G	V	M	S	P	Y	V	A	I	I	T	N	Y	V	V	I	F	C
MtFAAL31	H	D	S	G	G	L	M	S	P	L	M	A	I	M	T	S	I	L	G	I	V	G
MtFAAL32	H	D	S	G	G	L	M	S	P	F	T	A	L	V	T	A	L	V	G	L	F	H
MtFAAL33	H	D	G	G	G	C	M	S	P	A	F	A	L	L	V	R	V	L	G	N	M	A
MtFAAL34	H	D	S	G	A	V	M	G	P	A	G	F	I	M	M	G	S	M	G	A	H	V
MxFAAL	H	D	A	G	G	L	M	V	P	A	G	P	L	L	F							
RsFAAL	H	D	A	G	A	L	M	A	P	L	M											
LpFAAL	H	D	A	G	G	L	M	C	P	L	G	G	F									
EcFAAL	H	D	A	G	G	L	M	C	P	A	G	V	G	S	F							
PheA	F	D	G	G	A	I	A	A	P	C	W											
ACS	W	V	G	W																		
	Top (Short-chain)							Middle (Medium-chain)						Bottom (Long-chain)								

Figure 5: Molecular ruler in the form of a bulky residue lining the binding pocket, shown as sticks in the figure, defines the chain-length specificity of the ANL superfamily of enzymes. The ruler W414 restricts the specificity to acetate (C2) in acyl-CoA synthetase from *Salmonella enterica* (PDB: 1PG4), while W239 restricts the specificity to phenylalanine in phenylalanine-accepting PheA of Gramicidin synthetase from *Aneurinibacillus migulanus*. A very-long-chain fatty acid accepting FAAL32 from *Mycobacterium tuberculosis* without a ruler residue along the binding pocket represents an open tunnel. The tunnel is drawn as a surface around the bound substrate and the ruler residues are shown as sticks. The residues proposed to line the pocket of various FAALs from *Mycobacterium tuberculosis* and other organisms are tabulated. The ruler residues in FAALs whose specificity has been characterized from *Ralstonia solanacearum* and *Myxococcus xanthus* restrict the specificity to 6–8 carbon and 18 carbon, respectively, which are also indicated. The residues acting as rulers are highlighted in yellow for characterized systems and the structurally analogous residues in FAAL28 (used as reference for numbering) and FAAL32 are marked in pale orange, indicating the long-chain specificity in FAAL28 and absence of bulky residues results in an open-ended tunnel in FAAL32. The specificity table is modified from the previous work[35].

as acetyl-CoA synthetase (ACS) which is known to accommodate 2-carbon acetate, possesses a bulkier tryptophan side chain (W414; PDB:1PG4), while a smaller glycine residue occupies the structurally analogous position in MtFAAL28 (G330; PDB: 3E53). The chain-length specificity can be swapped by a G330W mutation in MtFAAL28, where the biochemical analysis of the mutant clearly suggests that it does not react with C12 acyl-chain but reacts with C2 acyl-chains. Similar strategies have been used previously in our lab to understand the substrate specificity of mycobacterial type-III PKSs. We had earlier shown that a subtle change in torsional angles allowed the common thiolase-fold to accommodate long-chain fatty acids and the introduction of bulky residue in the pocket can alter the chain-length specificity of these enzymes.[45]

The side chains from the specificity-defining elements can also dictate the volume of the pocket, where aryl chain accepting enzyme systems such as PheA (PDB: 1AMU) and 4CBCL (PDB: 3CW8) have broad pockets than fatty acyl-chain-binding enzymes. The length and the breadth of the pocket in PheA and 4CBCL are sufficient to accommodate the side chain of phenylalanine and 4-chlorobenzoic acid, respectively. Mycobacterial FAALs such as FAAL21, FAAL23, FAAL25, and FAAL28 have large tunnels that can accommodate fatty acids of chain lengths C12–C16, while some FAALs such as FAAL31, FAAL32, and FAAL34 have open tunnels that can accommodate unusually long fatty acids (Fig. 5). Indeed, FAAL32 has been implicated in the activation of very-long-chain mero-mycolates during biosynthesis of mycolic acids, but it is still unclear how the open tunnel helps in accommodating very-long-chain fatty acids. An interesting proposition was made by Hogbom and co-workers that a positive patch at the base of the N-terminus, where the open-ended tunnel leads to, can interact with the negatively charged head groups of the membrane lipids and the non-polar very-long acyl-chains of fatty acid can be buffered within the membrane.[46] Such analyses can be extended to the newly characterized FAALs and some of these already have substrate-bound crystal structures such as LpFAAL (PDB: 3KXW, 3LNV) and

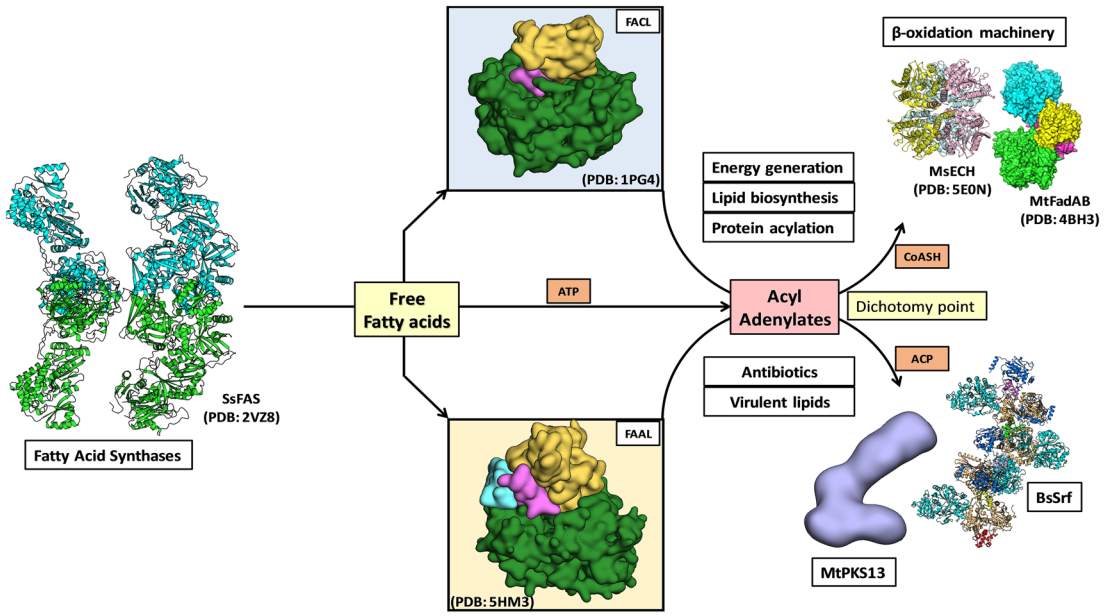


Figure 6: Multisystem crosstalk dictates the fate of free fatty acid by fatty acid synthases (PDB: 2VZ8). FACLs channel these to multiple pathways such as energy generation by β-oxidation (PDB: 4BH3; 5E0N), membrane lipid biosynthesis, etc., by FACLs (PDB: 1PG4), while A-domains (PDB: 1AMU) or FAALs (PDB: 5HM3) transfer them to PKS (MtPKS13; unpublished data) and NRPS, as proposed previously,[48] for the synthesis of metabolites such as antibiotics or the virulent lipids of *Mycobacterium*, etc. The temporal regulation of these systems regulates the crosstalk and hence the fate of free fatty acids, while the role of FAAL, a mechanistic variant in metabolic dichotomy, is still unknown.

EcFAAL (PDB: 3PBK)[34] where they have been shown to purify with bound acyl-adenylates with chain lengths to C12 and C14, respectively. Non-mycobacterial FAALs have also been characterized biochemically such as MxFAAL from *Myxococcus xanthus*[29] and RsFAAL from *Ralstonia solanacearum*,[30] where the substrate specificities have been fixed at C18 and C6, respectively, which concurs with the analysis presented here. Thus, it is possible to predict the substrate specificity of molecules of unknown function such as eukaryotic FAALs which may assist in delineating their function.

6 Discussion

The ANL superfamily is a group of fascinating enzyme systems that perform a complex reaction in two halves by large-scale domain rotations utilizing the same active site. FAALs are enigmatic members of this superfamily, because these systems are known to transfer the activated acyl-adenylate to PPant of ACP in PKS/NRPS through a hitherto uncharacterized mechanism. The presence of the rigidly anchored FSI somehow manages to facilitate the reaction specifically with PPant of ACP and rejecting PPant of CoASH. It also indicates that FAALs must have evolved certain additional features that have rendered themselves absolutely specific for PPant linked to acyl-carrier proteins (ACP), unlike the A-domain system which allows a mild reaction with CoASH, in vitro. Marahiel et al.[47] have suggested that the A-domains may have evolved from an ancestral form of adenylate-forming enzymes capable of utilizing CoASH as the acceptor molecule. It is possible that a diffusible small molecule such as CoASH was the "initial choice" and the catalytic strategy was built around it. The utilization of CoASH to generate post-translational modifications in proteins such as an ACP domain paved the way for bringing forth the importance of interfacial interactions that bring two proteins together for catalysis, eventually moving to a point of strict fidelity to its cognate partner.

All these flavours of catalysis are available in nature today (Fig. 6), with FACLs representing one end of the spectrum which is capable of accepting both CoASH and holo-ACP, while FAALs on the opposite side of the spectrum are almost incapable of accepting CoASH. The A-domains span the spectrum in the middle, which is capable of acylation of ACPs along with

some residual ability to acylate CoASH.[47] It has already been demonstrated that FAALs show "gain of function" and acylate CoASH after a deletion of the FSI.[35] Such an ability indicates its intrinsic ability to acylate CoASH, the evolutionary reminiscence of the past, upon a simple deletion of FSI. The gain of function in the deletion mutant is remarkable; however, the weak catalytic efficiency of the reaction suggests that the evolutionary divergence in FAALs has necessitated its crosstalk with ACPs with strict rejection mechanisms to prevent participation of a small diffusible CoASH in thioesterification reaction. It still remains to be understood how the FSI-based rejection of CoASH is bypassed only in the presence of a holo-ACP. Such a discrimination is fascinating, because the PPant participates in the reaction, whether in ACP or CoASH is identical as the PPant of ACP is derived from CoASH.

The FAAL-specific hydrophobic anchor of the N-terminal domain that holds the FSI rigidly can be a hallmark for segregating these systems from other members of the superfamily. It is now shown, in silico, that FAALs are ubiquitous in all prokaryotes and few eukaryotes, an identification made possible by exploiting the structural feature of FSI. These bioinformatic predictions need to be validated and extended both biochemically and structurally, which can serve as template systems to understand the unique mechanistic variations the FAALs have to offer. The substrate specificities in terms of the chain length of these systems can also possibly hint at their actual role, which has been now made possible through multiple substrate-bound complexes of the FAALs as well as other orthologs. The binding-pocket volume and depth is dictated by strategically placed bulky residues that act as molecular rulers defining the specificity. The substrate specificities can be swapped by simple mutagenesis that replace a bulky residue to the desired location, thereby altering the chain length. Such specificity-altered enzymes, particularly the A-domains and FAALs, can serve as tools for a novel polyketide generation, because the starter molecules from altered enzymes will be different. Similar strategies along with subdomain swapping have been previously used for the synthesis of novel molecules.[49] Thus, understanding substrate specificity can not only help in understanding the physiological role of the enzymes but also help in synthetic biology approaches.

Acknowledgements

R. S. acknowledges funding from JC Bose Fellowship of Science and Engineering Research Board (SERB), India, and Centre of Excellence Project of Department of Biotechnology, India.

Received: 5 May 2018 Accepted: 2 July 2018
Published online: 16 July 2018

References

1. Schmelz S, Naismith JH (2009) Adenylate-forming enzymes. Curr Opin Struct Biol 19:666–671
2. Estrada P, Manandhar M, Dong SH, Deveryshetty J, Agarwal V, Cronan JE, Nair SK (2017) The pimeloyl-CoA synthetase BioW defines a new fold for adenylate-forming enzymes. Nat Chem Biol 13:668–674
3. Wang M, Moynie L, Harrison PJ, Kelly V, Piper A, Naismith JH, Campopiano DJ (2017) Using the pimeloyl-CoA synthetase adenylation fold to synthesize fatty acid thioesters. Nat Chem Biol 13:660–667
4. Harvey EN (1957) A history of luminescence from the earliest times until 1900. American Philosophical Society, Philadelphia
5. Green AA, McElroy WD (1956) Crystalline firefly luciferase. Biochim Biophys Acta 20:170–176
6. Koo JA, Schmidt SP, Schuster GB (1978) Bioluminescence of the firefly: key steps in the formation of the electronically excited state for model systems. Proc Natl Acad Sci USA 75:30–33
7. Schuster GB, Dixon B, Koo JY, Schmidt SP, Smith JP (1979) Chemical mechanisms of chemi- and bioluminescence. Reactions of high energy content organic compounds. Photochem Photobiol 30:17–26
8. The PyMOL Molecular Graphics, Version 1.8, Schrödinger LLC
9. Lipmann F (1944) Enzymatic Synthesis Of Acetyl Phosphate. J Biol Chem 155:55–70
10. Berg P (1956) Acyl adenylates; an enzymatic mechanism of acetate activation. J Biol Chem 222:991–1013
11. Bar-Tana J, Shapiro B (1964) Studies on palmitoyl-coenzyme A synthetase. Biochem J 93:533–538
12. Bar-Tana J, Rose G (1968) Studies on medium-chain fatty acyl-coenzyme A synthetase. Enzyme fraction I: mechanism of reaction and allosteric properties. Biochem J 109:275–282
13. Bar-Tana J, Rose G, Shapiro B (1968) Studies on medium-chain fatty acyl-coenzyme A synthetase. Purification and properties. Biochem J 109:269–274
14. Chang KH, Xiang H, Dunaway-Mariano D (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. Biochemistry 36:15650–15659

15. Marahiel MA, Stachelhaus T, Mootz HD (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem Rev 97:2651–2674
16. 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
17. Hisanaga Y, Ago H, Nakagawa N, Hamada K, Ida K, Yamamoto M, Hori T, Arii Y, Sugahara M, Kuramitsu S, Yokoyama S, Miyano M (2004) Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. J Biol Chem 279:31717–31726
18. Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. J Mol Biol 388:997–1008
19. Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. J Biol Chem 283:32484–32491
20. Conti E, Franks NP, Brick P (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure 4:287–298
21. Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D (2009) The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. Biochemistry 48:4115–4125
22. Wu R, Cao J, Lu X, Reger AS, Gulick AM, Dunaway-Mariano D (2008) Mechanism of 4-chlorobenzoate:coenzyme A ligase catalysis. Biochemistry 47:8026–8039
23. Mitchell CA, Shi C, Aldrich CC, Gulick AM (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry 51:3252–3263
24. Sundlov JA, Fontaine DM, Southworth TL, Branchini BR, Gulick AM (2012) Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. Biochemistry 51:6493–6495
25. Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D, Gokhale RS (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature 428:441–445
26. Arora P, Goyal A, Natarajan VT, Rajakumara E, Verma P, Gupta R, Yousuf M, Trivedi OA, Mohanty D, Tyagi A, Sankaranarayanan R, Gokhale RS (2009) Mechanistic and functional insights into fatty acid activation in Mycobacterium tuberculosis. Nat Chem Biol 5:166–173
27. Yamada KD, Tomii K, Katoh K (2016) Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. Bioinformatics 32:3246–3251
28. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324
29. Hayashi T, Kitamura Y, Funa N, Ohnishi Y, Horinouchi S (2011) Fatty acyl-AMP ligase involvement in the production of alkylresorcylic acid by a Myxococcus xanthus type III polyketide synthase. ChemBioChem 12:2166–2176
30. Kage H, Kreutzer MF, Wackler B, Hoffmeister D, Nett M (2013) An iterative type I polyketide synthase initiates the biosynthesis of the antimycoplasma agent micacocidin. Chem Biol 20:764–771
31. Hemmerling F, Lebe K, Wunderlich J, Hahn F (2018) An unusual FAAL-ACP didomain in ambruticin biosynthesis. ChemBioChem 19:1006–1011
32. Marchetti PM, Kelly V, Simpson JP, Ward M, Campopiano DJ (2018) The carbon chain-selective adenylation enzyme TamA: the missing link between fatty acid and pyrrole natural product biosynthesis. Org Biomol Chem 16:2735–2740
33. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797
34. Zhang Z, Zhou R, Sauder JM, Tonge PJ, Burley SK, Swaminathan S (2011) Structural and functional studies of fatty acyl adenylate ligases from *E. coli* and *L. pneumophila*. J Mol Biol 406:313–324
35. Goyal A, Verma P, Anandhakrishnan M, Gokhale RS, Sankaranarayanan R (2012) Molecular basis of the functional divergence of fatty acyl-AMP ligase biosynthetic enzymes of Mycobacterium tuberculosis. J Mol Biol 416:221–238
36. Gavalda S, Leger M, van der Rest B, Stella A, Bardou F, Montrozier H, Chalut C, Burlet-Schiltz O, Marrakchi H, Daffe M, Quemard A (2009) The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in Mycobacterium tuberculosis. J Biol Chem 284:19255–19264
37. Trivedi OA, Arora P, Vats A, Ansari MZ, Tickoo R, Sridharan V, Mohanty D, Gokhale RS (2005) Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. Mol Cell 17:631–643
38. Siméone R, Léger M, Constant P, Malaga W, Marrakchi H, Daffe M, Guilhot C, Chalut C (2010) Delineation of the roles of FadD22, FadD26 and FadD29 in the biosynthesis of phthiocerol dimycocerosates and related compounds in Mycobacterium tuberculosis. FEBS J 277:2715–2725
39. Vergnolle O, Chavadi SS, Edupuganti UR, Mohandas P, Chan C, Zeng J, Kopylov M, Angelo NG, Warren JD, Soll CE, Quadri LE (2015) Biosynthesis of cell envelope-associated phenolic glycolipids in Mycobacterium marinum. J Bacteriol 197:1040–1050
40. Srivastava S, Chaudhary S, Thukral L, Shi C, Gupta RD, Gupta R, Priyadarshan K, Vats A, Haque AS, Sankaranarayanan R, Natarajan VT, Sharma R, Aldrich CC, Gokhale RS (2015) Unsaturated lipid assimilation by mycobacteria requires auxiliary cis-trans enoyl CoA isomerase. Chem Biol 22:1577–1587

41. Toyoda K, Inui M (2018) Extracytoplasmic function sigma factor sigma(D) confers resistance to environmental stress by enhancing mycolate synthesis and modifying peptidoglycan structures in Corynebacterium glutamicum. Mol Microbiol 107:312–329
42. Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J 16:4174–4183
43. Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol 6:493–505
44. Khurana P, Gokhale RS, Mohanty D (2010) Genome scale prediction of substrate specificity for acyl adenylate superfamily of enzymes based on active site residue profiles. BMC Bioinform 11:57
45. Sankaranarayanan R, Saxena P, Marathe UB, Gokhale RS, Shanmugam VM, Rukmini R (2004) A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. Nat Struct Mol Biol 11:894–900
46. Andersson CS, Lundgren CA, Magnúsdóttir A, Ge C, Wieslander A, Martinez Molina D, Hogbom M (2012) The Mycobacterium tuberculosis very-long-chain fatty acyl-CoA synthetase: structural basis for housing lipid substrates longer than the enzyme. Structure 20:1062–1070
47. Linne U, Schafer A, Stubbs MT, Marahiel MA (2007) Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains. FEBS Lett 581:905–910
48. Marahiel MA (2016) A structural model for multimodular NRPS assembly lines. Nat Prod Rep 33:136–140
49. Kries H, Niquille DL, Hilvert D (2015) A subdomain swap strategy for reengineering nonribosomal peptides. Chem Biol 22:640–648

Kinatukara Priyadarshan joined Dr. Sankaranarayanan's laboratory after finishing his post-graduation from Madurai Kamaraj University, India, and obtained his Ph.D. from CSIR-Centre for Cellular and Molecular Biology, India, in 2016. He is presently working as a Project-Research Associate at the CSIR-Centre for Cellular and Molecular Biology, India.

Dr. Rajan Sankaranarayanan obtained his Ph.D. from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore, in 1996, and was a postdoctoral research fellow at IGBMC, Strasbourg, France, from 1996 to 2002. He is currently at the CSIR-Centre for Cellular and Molecular Biology, India, and is carrying out research in the field of Structural Biology. The major areas of interest in his laboratory are translational quality control mechanisms and the enzymology of lipid metabolism pathways.