Genetic, biosynthetic and functional versatility of polyketide synthases

Divya R. Nair[1], Swadha Anand[1], Priyanka Verma[1], Debasisa Mohanty[1] and Rajesh S. Gokhale[1–3,*]

[1]National Institute of Immunology, New Delhi 110 067, India
[2]Institute of Genomics and Integrative Biology, Council of Scientific and Industrial Research, Delhi 110 007, India
[3]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

Polyketide synthases (PKSs) are multi-functional enzymes that exhibit a great deal of diversity in their biosynthetic mechanism. This versatility allows them to influence an organism's biology in varied ways. Although many polyketides have been identified till now, and their biosynthetic pathways characterized, many more are waiting to be uncovered. With genome sequences of many organisms available now, the repository of putative *pks* genes is growing and so is the complexity involved in determining their possible products and functional roles in nature. Needless to say, efforts at the identification of novel products and pathways require a systematic, multi-pronged approach. In this review we discuss the growth of the PKS field and how it has undergone a pronounced transformation. We examine the various ways by which nature introduces variations in the products formed by these enzymes in different organisms to suit their biology. We further examine the different tools employed to crack the 'PKS puzzle' and how the future looks for research in this field.

Keywords: *Dictyostelium*, fatty-acyl AMP ligases, mycobacteria, polyketide synthases.

NATURE has an astounding repository of metabolites with diverse structures and varied functions. A significant number of these natural products are biosynthesized by enzymes known as polyketide synthases (PKSs), which are known to occur in various organisms ranging from bacteria to insects, and plants. PKSs and their metabolites (polyketides) have emerged as an area of intense research owing to their great commercial value to human beings. Their importance is amply exhibited by compounds such as erythromycin, streptomycin (antibiotics); lovastatin (anti-cholesterolemic agent); rapamycin (immunosuppressant drug); resveratrol (antioxidant); aflatoxin (fungal toxin) and many more. Thus for several years research in this field was focused on mining of novel metabolites from different organisms to explore new commercial products.

*For correspondence. (e-mail: rsg@igib.res.in)

Polyketides were first discovered in 1883 by James Collie while working on the elucidation of the structure of dehydroacetic acid[1]. He observed that boiling dehydroacetic acid with barium hydroxide yielded an aromatic compound called orcinol as one of the products. Further study revealed a polyketone intermediate to be involved in this conversion. He coined the term polyketide (i.e. polyacetate) for the compounds containing the structure CH_3-CO-$(CH_2$-CO$)_n$-X and suggested a complex series of reactions for the conversion of polyketones to their final form[2]. However, his ideas were met with indifference at that time and it was finally in the 1950s that the polyketide field was revived by Arthur Birch[3]. His work on the *Penicillium patulum* aromatic polyketide-6-methylsalicylic acid using ^{14}C acetate established that many polyphenolic aromatic compounds are biosynthesized from acetate units, according to what came to be known as the 'Collie–Birch polyketide hypothesis'. This pioneering study, combined with mass spectrometric and NMR tools paved the way for structural elucidation of many more complex polyketides like aflatoxin[4] and erythromycin[5], and provided insights into their biosynthetic pathways.

Initially, the plethora of these natural compounds far exceeded the information about genes and pathways involved in their biosynthesis. But over the years, there has been an interesting reversal in this trend owing to genome sequencing programmes initiated in different organisms. With increasing number of PKS homologues being discovered, today's researchers face the challenge of deciphering the metabolites synthesized by these genes. The recent past has witnessed another thematic shift, with due attention being given to comprehend how these 'secondary' metabolites could play a crucial role in the survival of an organism, to escape unfavourable conditions; in virulence and pathogenecity of pathogens, and as inducers of differentiation. It is evident that polyketides are more intimately involved with survival and viability of an organism than previously thought. It is particularly interesting to understand how nature employs these versatile enzymes to build unique metabolites while using apparently similar chemistry. Indeed it is true that work in past one decade has unravelled hitherto unknown

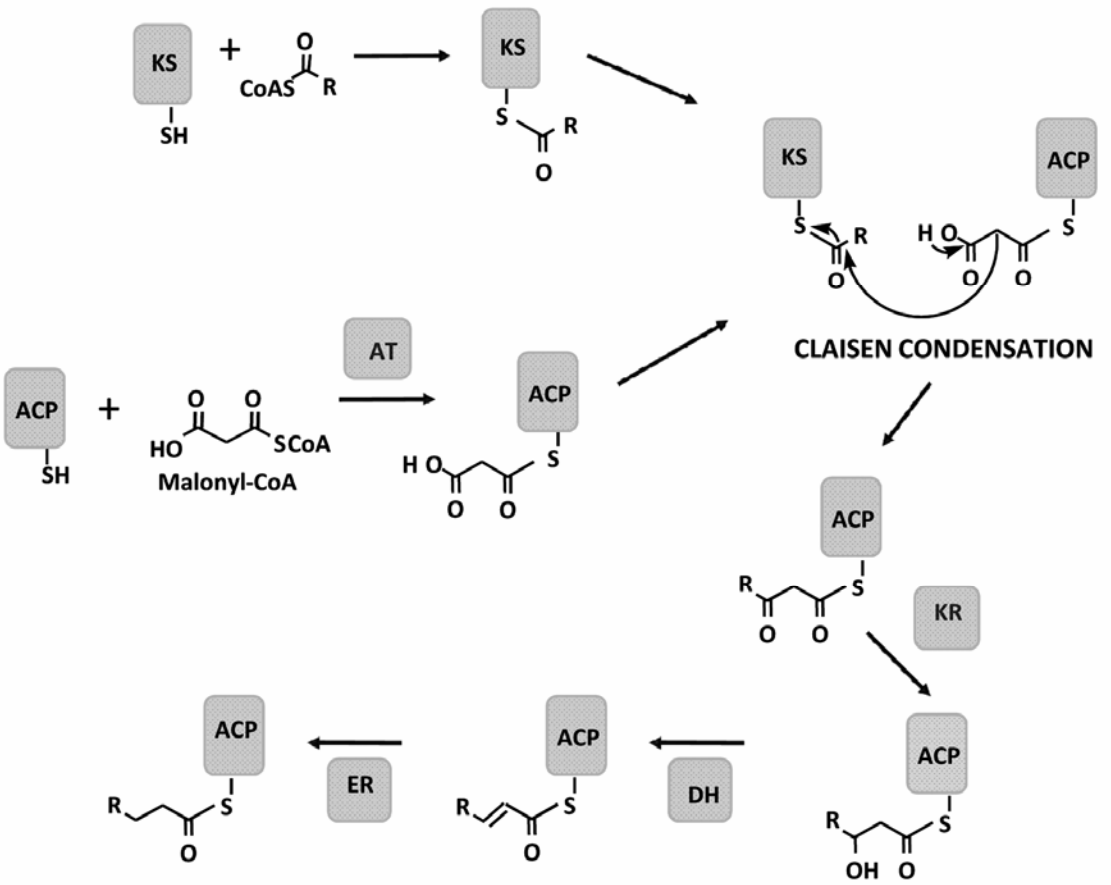

Figure 1. Polyketide biosynthesis mechanism.

aspects of PKSs, which deviate significantly from conventional wisdom on PKS chemistry in terms of substrate specificity, final product formation and so on. Subsequent sections in this review will highlight how the 'gene → metabolite' path can be traced down systematically; the emerging concepts in polyketide biosynthesis and their functional role in an organism's biology.

PKSs function by performing decarboxylative Claisen-type condensation of ketide units. A typical PKS is characterized by the presence of three core domains: ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP). Apart from these domains, auxiliary domains may also be present which are responsible for β-carbon modification. These include ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER). Figure 1 illustrates a general scheme for the biosynthesis of complex metabolites by the concerted action of these domains. Based on protein architecture, PKSs have been classified into three categories. The first category is that of type I PKSs in which various domains are covalently fused to each other and are further classified into two kinds – iterative and modular. Iterative PKSs use their domains in an iterative fashion, where a single set of domains is used repetitively for performing all the necessary chain elongation steps. Modular PKSs on the other hand, are organized into groups of active sites known as modules, in which each module is responsible for one cycle of chain extension and functional group modification. Type II PKSs form the second category. They also utilize their protein components in an iterative fashion; however, the different domains are present as discrete, independent proteins. The various components function together at the time of biosynthesis to yield the final product. Type III PKSs are homodimeric proteins that show homology to plant chalcone synthases (CHSs). They use CoA esters as substrates and synthesize polyketides using a single catalytic centre, without the involvement of the phosphopantetheine arm. Alternative patterns of intramolecular cyclization catalysed by type III PKSs result in remarkable biosynthetic diversity, as depicted in Figure 2.

Complexity of the PKS biosynthetic machinery makes it a daunting task to correlate a putative *pks* gene to its probable product. It is analogous to solving a jigsaw puzzle, where fitting together the first two pieces correctly lays the foundation for the bigger picture. Computational analyses of the newly discovered genes, with efficient prediction tools hold the key to get cracking the 'gene → metabolite' puzzle.

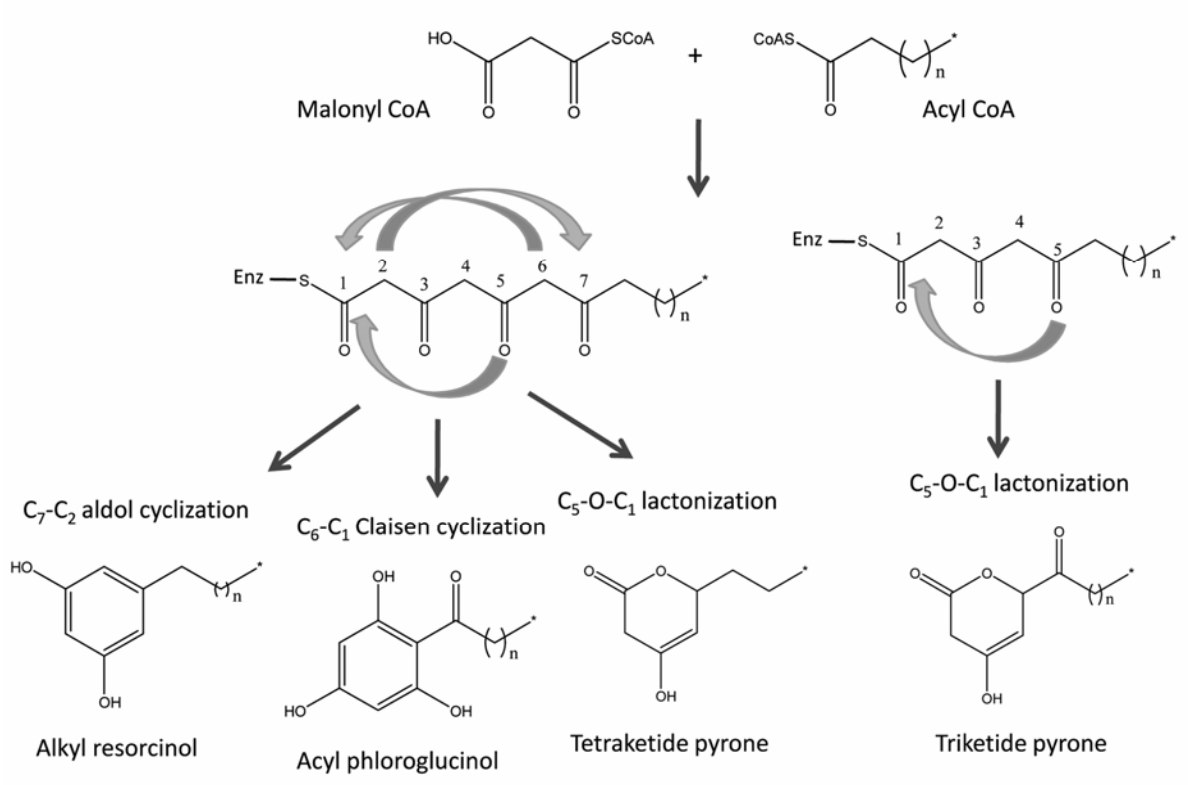

Figure 2. Type III polyketide synthase (PKS) metabolite versatility due to different patterns of intramolecular cyclization.

Bioinformatic analysis of PKS gene clusters

The functional group added by each module depends on the substrate specificity of the AT domain and types of chain-modification domains present in the corresponding module. Thus, the chemical structure of linear polyketide biosynthesized by a modular PKS can be predicted by knowledge of the organization of the catalytic domains, their substrate specificities and the order of substrate channelling, which is governed by protein–protein interactions between various ORFs in the cluster. The final structure is modified by cyclization and post-PKS modifications, and therefore, information on specificities of thioesterase (a class of chain-releasing domains) as well as tailoring enzymes is also required to predict the complete chemical structure.

Sequence-based analysis

Deciphering domain organization of PKS gene clusters: The standard databases used for domain identification like Conserved Domain Database (CDD)[6], InterPro[7], etc. fail to detect certain domains or predict their correct boundaries in case of PKS domains. Yadav *et al.*[8] and Ansari *et al.*[9] developed a knowledge-based method for automated identification of PKS catalytic domains using comprehensive analysis of 20 PKS gene clusters. The approach used BLAST alignment of the query sequence with representative homologous stand-alone structures of various catalytic domains to predict the domain boundaries. Later, various other web-servers like ASMPKS[10], ClustScan[11] and NP.searcher[12] were also developed with additional features like HMM-based domain search and scanning of complete genomes for PKS gene clusters. Recently, another webserver – SBSPKS[13] was developed, which predicts the structural and catalytic domain boundaries based on alignment with the recently available crystal structures.

Phylogenetic analysis of PKS domains: Evolutionary analysis of PKS genes has provided novel insights into their phylogeny and to find predictive rules for deciphering their function as well as substrate specificity, etc. Analysis of 203 KS domains from 23 modular PKS gene clusters shows a monophylectic relationship indicating evolution by gene duplication[14]. KS domains show distinct phylogeny-based clustering on the basis of modular or iterative condensation they catalyse, but KS domains in modular PKS do not show substrate-dependent clustering[15,16]. On the other hand, analysis of 187 AT domain sequences from different species showed clustering according to their substrate specificity rather than their species of origin[17–19]. Hence profiles/motifs based on conserved patterns across the group of sequences which cluster together phylogenetically help determine the

reaction type or substrate specificity. Incorporation of binding pocket information from the available crystal structures along with evolutionary analysis can help improve predictions given by various PKS prediction tools.

Structure-based analysis

Prediction of AT domain substrate specificity: The specificity-determining residues from 187 PKS AT domains, when analysed phylogenetically, showed distinct clusters for malonate and methylmalonate-specific AT domains and revealed conserved binding pocket motifs **QQGHS**[QMI]**GRSH**T[NS]**V** for methylmalonate-specific AT domains, and **QQGHS**[LVIFAM]**GR**[FP]**H**[ANTGEDS][NHQ]**V** for malonate-specific AT domains. Starter AT domains specific for monocarboxyl acid substrates also formed a separate cluster and lacked the conserved arginine present in the GRSH or GRFH motifs. The involvement of arginine in the GRSH or GRFH motif has been proposed in the recognition of dicarboxylic acid substrates[20]. This approach using 13 specificity-determining residues was implemented in the NRPS–PKS software for automated prediction of AT specificity. Other software like ASMPKS[10], ClustScan[11] and NP.searcher[12] have also used a similar protocol for the prediction of AT specificity. Minowa *et al.*[21] were able to find the distinction between methylmalonyl-CoA, ethylmalonyl-CoA and methoxymalonyl-CoA-specific ATs, which were not effectively distinguished by previous studies.

Prediction of KR domain stereo-specificity: The stereospecificity of the KR domain determines the orientation of the hydroxyl group added by it. The stereo-specificity-determining residues for KR domains in PKSs have been predicted using sequence- and structure-based studies[22,23], and these predictions have been verified using site-directed mutagenesis studies[24] and biochemical experiments[25]. A recent study divided the KR domain into six categories on the basis of α and β substituent chirality and elucidated the mechanistic aspects governing the stereo-control[23]. Rules for automated prediction of KR stereo-specificity are available in software like ClustScan[11] and SBSPKS[13].

Prediction of number of iterations catalysed by type I iterative PKSs: It might also be important to distinguish between modular and iterative type I PKSs and also predict the number of condensations catalysed by them. It has been shown that modular and iterative PKS proteins show distinct phylogeny-based clustering for the KS domains. Yadav *et al.*[16] could successfully classify type I iterative, modular, hybrid PKS/NRPS (NRPSs refer to non-ribosomal peptide synthetases – another family of secondary metabolite synthesizing enzymes) and enedyne-type KS domains using HMM profiles. The iterative KS domains showed distinct clustering on the basis of the number of iterations that are catalysed and the type of reductive chain modifications associated with each elongation step. Several iterative KS domains on analysis showed that the number of iterations as well as the reduction of the polyketide product was clearly dependent on the size and hydrophobicity of the KS active-site pocket. This structural analysis indicated that it was possible to predict the number of iterations using information on the residues lining the KS binding-site pocket as well as identified certain residues that could be mutated to change the number of iterations catalysed by an iterative type I PKS[16].

Prediction of the order of substrate channelling modular PKS clusters based on inter subunit interaction involving docking domain: Since modular PKS clusters contain multiple ORFs, the order of substrate channelling depends on the order in which ORFs are utilized, and often deviates from the order of their occurrence in the genome. The loading module can help identify the first ORF while last one can be identified using the TE domain, but the number of combinatorial possibilities for the other ORFs still remain very large. Therefore, several studies have attempted to predict the order of substrate channelling computationally. Minowa *et al.*[21] have utilized interdomain interactions between KS and ACP in PKS to distinguish cognate pairs, from non-cognate pairs, and predict the order of substrate channelling. The elucidation of the NMR structure of the so called 'docking domain' revealed the docking code or the structural basis of recognition between terminal linker regions of two interacting ORFs during substrate channelling[26–28]. Based on the analysis of inter-helical contacts in the docking domain structures, Yadav *et al.*[19] have developed an automated approach to distinguish between cognate and non-cognate ORF pairs in a modular PKS cluster. Benchmarking studies on a dataset of 17 modular PKS clusters indicated that, in 14 out of 17 cases the true cognate combination can be ranked within the top 20% in terms of total score. A web interface for automated extraction of linker sequences and prediction of the order of substrate channelling in case of modular PKS clusters is provided by the recent version of SBSPKS software[13].

It is to be noted that computational predictive tools need to be constantly upgraded and rewrapped for identification of novel domains and to explore their roles. Unusual domains like C-methyltransferase (C-Met), thioesterase/Claisen cyclase (TE/CLC), reductase, starter unit – ACP transacylase (SAT) and product template (PT) in fungal PKSs exemplify this requirement. Udwary *et al.*[29] developed an algorithm, Udwary–Merski Algorithm (UMA) that combines primary sequence similarity, predicted secondary structure, and local hydrophobicity to predict interdomainal linker regions. UMA made possible the identification of previously unidentified domains, SAT and PT, from norsolorinic acid synthase.

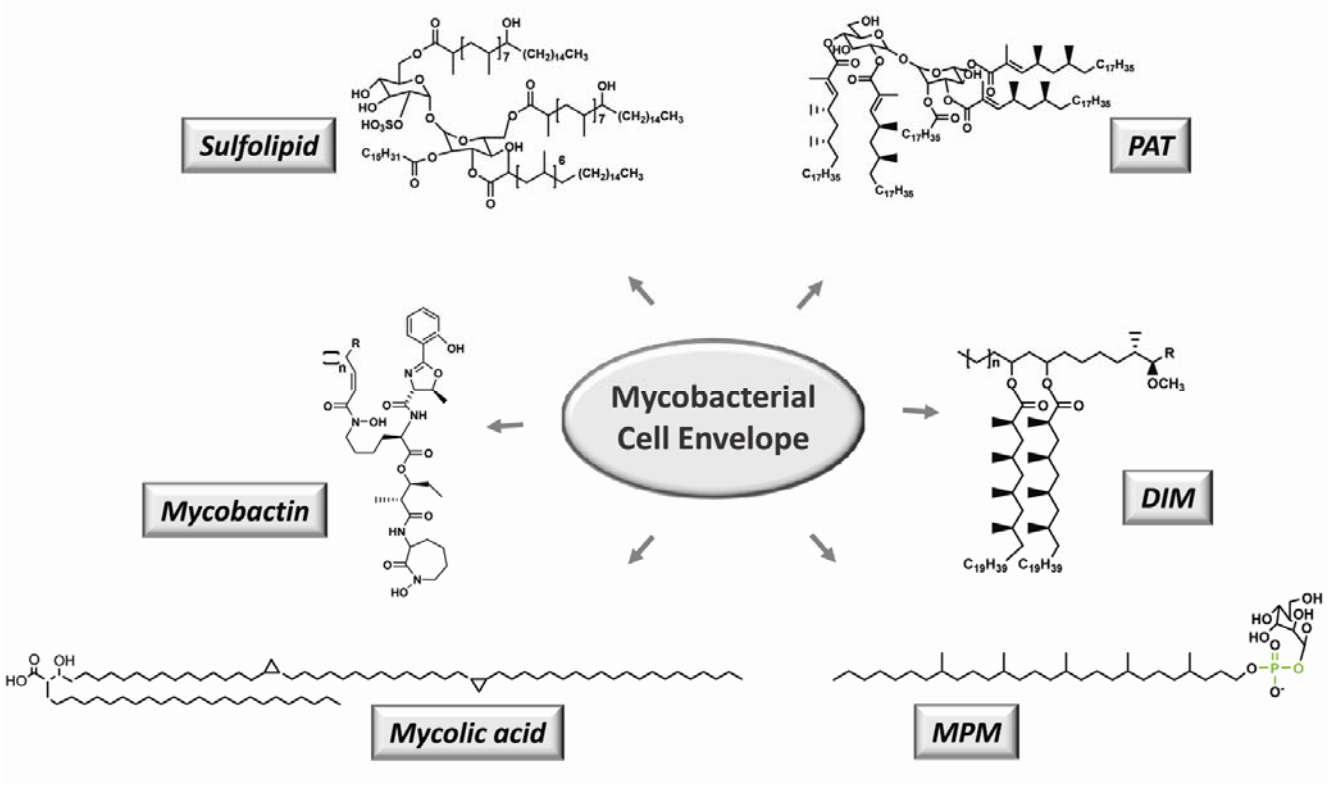

Figure 3. Mycobacterial cell wall-associated complex lipids synthesized by PKSs. PAT, Polyacyl trehalose; DIM, Dimycocersoates; MPM, Mannosyl phosphomycoketides.

In the examples cited below, we take an overview of instances where type I and type III PKSs exhibit interesting variations across mycobacteria, *Dictyostelium*, fungi and plants.

Mycobacterial PKSs in pathogenecity and virulence

Mycobacterium tuberculosis (*Mtb*) genome sequencing in 1998 was a tremendous boost to the pursuit of a suitable drug target against tuberculosis[30]. The presence of large number of PKS homologues in the genome came as a huge surprise, since *Mtb* is not known to produce secondary metabolites. Detailed investigation of these PKSs over the years has revealed their involvement in the biosynthesis and assembly of cell wall-associated complex lipids (Figure 3). It is remarkable to note the ability of *Mtb* PKSs to take up long-chain substrates and how they engage in a crosstalk with fatty acid synthases (FAS) via fatty acyl AMP ligases (FAALs). *Mtb* PKSs have indeed set the tone for exploring new paradigms in PKS functioning.

Mtb and *M. avium* cell wall is characterized by the presence of unusual phospholipids called mannosyl-β-1-phosphomycoketides (MPMs)[31]. These lipids contain a long phytanic acid-like alkyl chain with the carboxyl end capped with a phospho-mannose group. The alkane chain moiety was shown to be necessary for activation of CD1c-restricted cells, but its structure was intriguing because it did not conform to the expected products of any known fatty acyl, polyprenol or polyketide biosynthetic pathway[32]. Labelling studies coupled with mass spectrometry suggested a hypothetical mechanism for production of the isoprene-like alkyl backbone via malonyl and methylmalonyl units and PKSs[33]. Genome inspection revealed *pks12* to be the candidate gene encoding for a protein with 12 catalytic domains that corresponded to key steps in the proposed pathway. Gene knockout and complementation studies proved that this gene was involved in antigen production[33]. It was only recently that the biosynthetic mechanism of MPM became clear, when it was reported that the two ATs of bimodular PKS12 exhibit strict specificity for methylmalonate and malonate extender units respectively. Further biochemical analysis revealed a novel paradigm of 'modularly iterative' mechanism of catalysis in which the two modules function alternately to form the branching pattern associated with the mycoketide alkyl chain[34].

PDIMs form another class of complex lipids which are found abundantly on the cell wall of pathogenic species of mycobacteria and are known to be involved in the modulation of host immune responses[35,36]. Biosynthesis of these lipids requires both modular and iterative PKS machinery. The mycocerosic acid component is produced

by an iterative mycocerosic acid synthase (MAS) protein, and the β-diols are synthesized by modular PpsA–E proteins[37]. The starter fatty acyl chain is provided by a FAAL homologue, FadD26 (ref. 38). Notably, PpsD protein does not contain an ER domain and instead utilizes a stand-alone ER protein during the condensation cycle[37,38].

Mycolic acids are essential for the survival of *Mtb*[39] and their production also follows an unusual path, involving formation of long-chain products via the combined action of type I FAS, type II FAS and a PKS. Type I FAS forms the α-alkyl chain and the meromycolate precursor[40]. The latter undergoes type II FAS-mediated extension and modification, and is then activated by FAAL32 (refs 39 and 41). Eventually, PKS13 brings about a decarboxylative Claisen-like condensation between FAAL32-activated meromycolate chain and the α-alkyl chain, resulting in the final assembly of mycolic acid[42,43].

Dictyostelium PKSs in differentiation

Dictyostelium (*Dicty*) is a social amoeba that has held considerable interest among researchers as a model organism because of its ability to undergo a multicellular developmental pathway (www.dictybase.org)[44]. Starvation induces millions of such unicellular amoebae to aggregate towards each other through chemotaxis and differentiate into different cell types. The differentiated cell types then sort out to different regions of the developing structure to eventually form a fruiting body that contains a stalk (made of dead, vacuolated cells) and a spore sac encapsulating viable spores. This strategy helps the organism escape unfavourable conditions, with the spores germinating to release amoebae upon encountering conducive environment. Interestingly, the genome of this organism possesses an incredibly large number of type I PKSs (henceforth, referred to as DiPKSs)[45]. Computational analysis of these proteins has revealed their domain architecture to be closer to fungal PKSs, and they show the presence of unique domains. Of the 40 PKSs, 30 have a C-Met domain and 18 have NAD-dependent reductase domains in their C-terminal regions. It is noteworthy that two of the DiPKSs–DiPKS1 and DiPKS37 possess unprecedented type III PKS domain at their C-termini[46]. The similarity between the fungal and *Dicty* PKSs leads one to believe that complex iterative programming represented by the former (as exemplified in squalestatin and lovastatin biosynthesis, Figure 4) may also be emulated by the latter. Biosynthesis of squalestatin involves three rounds of chain extension by the enzyme SQTKS. Methyl group transfer by its C-Met domain occurs after the first and second rounds of iteration, but the domain is inactive in the final round of extension along with the ER domain[47]. Lovastatin, which is an HMG-CoA synthase inhibitor, is produced by the combined action of two PKSs – LDKS (non-iterative) and LNKS (iterative). It is intriguing that although LNKS possesses an ER domain, the domain is inactive and instead the enzyme utilizes a stand-alone ER (LovC) for correct enoyl reduction and chain extension[48].

It has been speculated that the tightly synchronized developmental process in *Dicty* is regulated by metabolites synthesized by PKSs. Although this slime mold has turned out to be a storehouse of many unique molecules like dictyopyrone[49,50], dictyomedin[51], brefelamide[52], etc., nothing is known about their biological relevance or the biosynthetic pathway (Figure 5). Much before the presence of PKSs was discovered in *Dicty*, it was known that a small metabolite called DIF-1 (Differentiation Inducing Factor) acts as an inducer of one of the cell types (pre-stalk cells) during the differentiation process[53–56]. Although many other metabolites (cAMP, ammonia, etc.) have been implicated in the development process[57–60], this molecule attracted attention because of its polyketide-like chemical structure. It was only recently that it was finally deciphered that DiPKS37 is responsible for the biosynthesis of this developmental regulatory factor. DiPKS37-mutant strain exhibited notable phenotypic defects during development. The mutant fruiting bodies showed lack of basal disc at the stalk base, causing the structure to topple over on the solid substratum. Transcript profile of DiPKS37 mRNA revealed stage-specific expression of the gene at later developmental stages. This study provided the first indication of the involvement of DiPKS in the *Dicty* differentiation[61]. Further evidence came from Zucko *et al.*[62], who reported that many of the DiPKSs showed temporally regulated expression through development.

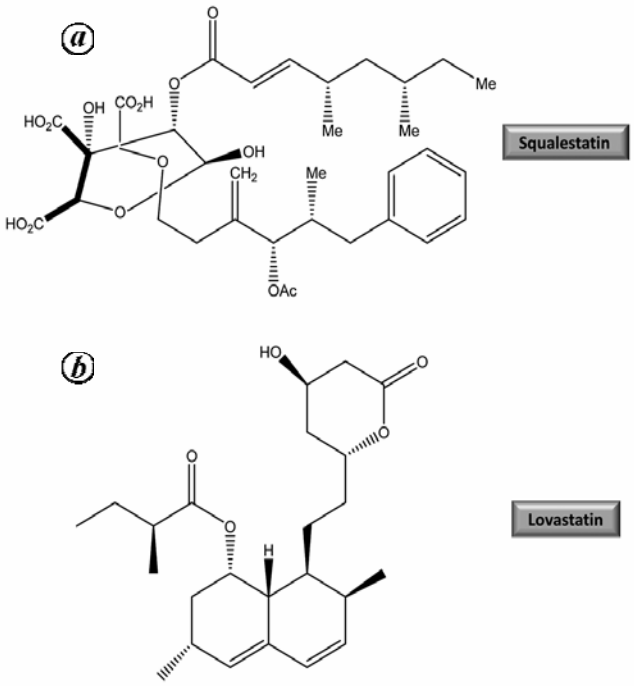

Figure 4. (***a***) Squalestatin and (***b***) lovastatin – fungal metabolites synthesized by complex programming of iterative cycles.

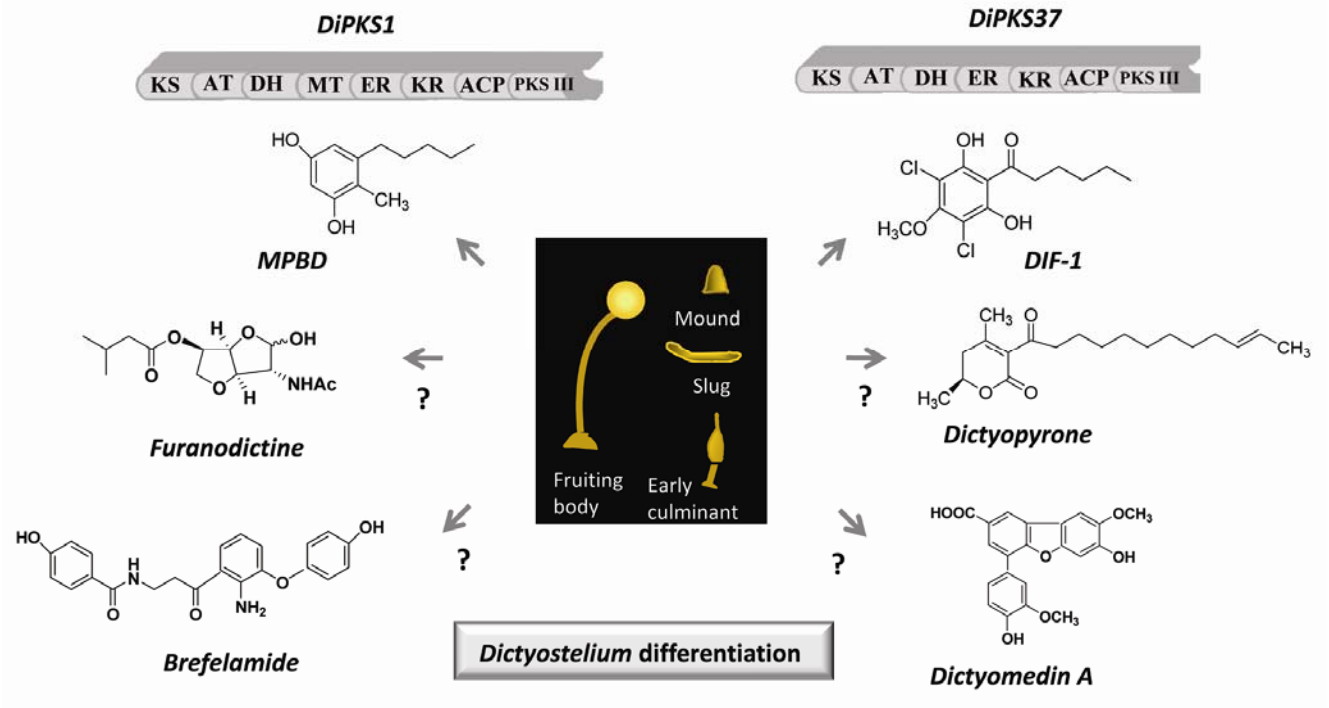

Figure 5. Metabolite repertoire of *Dictyostelium*. MPBD and DIF-1 are known to be involved in differentiation and are synthesized by DiPKS1 and DiPKS37 respectively. Biosynthetic pathway and functions of other metabolites are still unknown.

Close on the heels of these studies, DiPKS1 was also suggested to be involved in the developmental process[46]. Through elaborately performed biochemical studies, it was demonstrated that DiPKS1 type III PKS domain takes up hexanoyl CoA to catalyse three polyketide chain extensions followed by aldol cyclization to synthesize des-methyl MPBD. MPBD is a resorcinolic compound believed to be a general inducer of pre-stalk and pre-spore differentiation[63]. It is noteworthy that whereas the type III PKS domain by itself primarily produces acyl pyrone, the presence of interacting ACP domain modulates the catalytic activity to produce the alkyl resorcinol scaffold of MPBD. This work also suggested that post-PKS tailoring enzymes like O-methyltransferases could bring in subtle variations in the biological function of a PKS metabolite by introducing additional groups. It remains to be seen how other DiPKSs may modulate the life cycle of this amoeba.

Fungal PKSs

Melanin production

Melanin pigments are produced by a host of pathogenic organisms like bacteria, fungi and helminthes. PKS biosynthetic pathway is one of the major pathways by which microbes produce melanin. Melanization in fungi appears to contribute to their virulence by reducing susceptibility to host defence mechanisms and antifungal drugs. It is generally believed that acetyl CoA and malonyl CoA are the starter and extender units of PKSs involved in the fungal 1,8-dihydroxynaphthalene (DHN)-melanin pathway[64]. The typical biosynthetic pathway involves formation of a pentaketide precursor – 1,3,6,8-tetrahydroxynaphthalene (1,3,6,8-THN), which is reduced to scytalone by 1,3,6,8-THN reductase and subsequently converted to DHN via dehydration and reduction steps (Figure 6 *a*). Recognized human fungal pathogens that form melanin precursors by the polyketide pathway include *Wangiella dermatitidis*, *Sporothrix schenckii*, *Alternaria alternate*, *Cladosporium carionii*, *Fonsecaea pedrosoi*, *Aspergillus niger*, *A. nidulans* and *A. fumigatus*[65].

A. fumigatus is a ubiquitous fungus that causes life-threatening invasive pulmonary aspergillosis. Melanin production in this fungus follows a slightly different mechanistic scheme, as its PKS (Alb1) synthesizes a heptaketide rather than a pentaketide. The heptaketide formed, YWA1 (Figure 6 *b*), undergoes a unique polyketide-shortening mediated by Ayg1 to yield the pentaketide precursor 1,3,6,8-THN[66]. In *W. dermatitidis*, which is another fungal pathogen, disruption of melanin-producing PKS (WdPKS1) results in a decrease in virulence in mice, and cells deficient in the gene are more susceptible to killing by neutrophils. It also leads to reduction in the thickness of the cell wall, and the absence of a polymerized melanin layer results in decreased survival on exposure to the antifungal agents – amphotericin B, voriconazole and defensin NP-1 (ref. 67). *Penicillium*

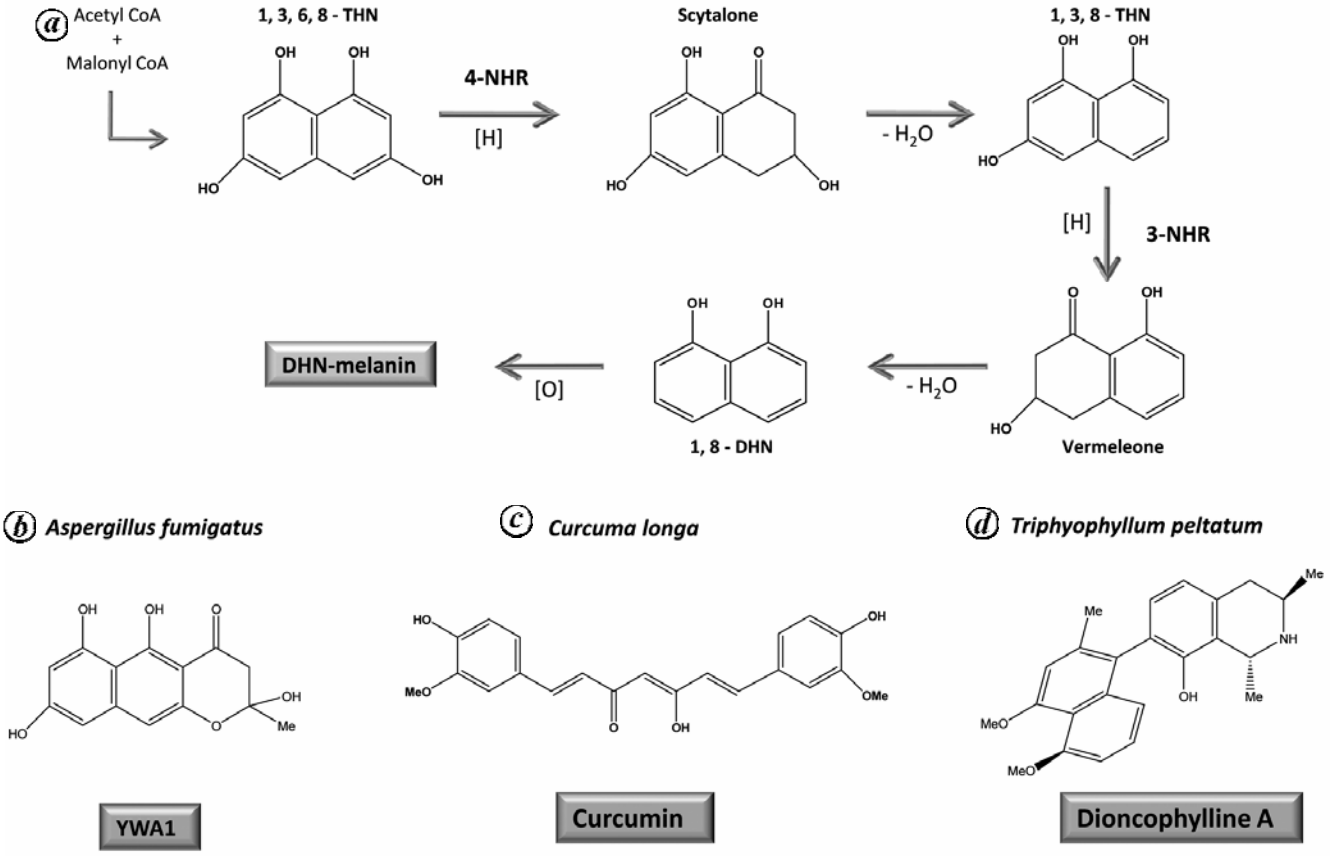

Figure 6. ***a***, DHN-melanin biosynthetic pathway; ***b***, YWA1 – precursor of 1,3,6,8-THN-synthesized by *Aspergillus fumigatus*; ***c***, curcumin produced by type III PKS of *Curcuma longa*; ***d***, alkaloid dioncophylline A synthesized by *Triphyophyllum peltatum*.

marneffei is a dimorphic fungus causing respiratory, skin and systemic mycosis. Its genome encodes for 23 PKSs, of which one of the PKSs lies in the melanin biosynthetic gene cluster. It has been demonstrated that knocking-down of this gene – *alb*1, results in the loss of melanin pigment production and reduced ornamentation in the conidia. Significantly, it was observed that survival of mice challenged with the *alb1* knock-down mutant was better than those challenged with wild type *P. marneffei*. It was also seen that survival time of the fungal conidia on exposure to hydrogen peroxide was reduced in the mutant compared to the wild type[68].

Type III PKS

A novel type III PKS protein was recently reported in *Neurospora crassa*, which had a notable feature – ability to take up long-chain starter molecules for synthesizing tetraketide resorcinols as the major products[69]. The only other known type III PKS with affinity for long-chain starters was shown in *Mtb*[70] (such complexity has now also been demonstrated in plant type III PKSs – which will be discussed in the following section). The observation that this gene is expressed during all three stages of the fungi – mycelia, conidia and ascospores, suggests the possible functional importance of resorcinolic lipids in fungal growth and reproduction[69].

Plant type III PKSs

Rhizome of turmeric (*Curcuma longa*) is a well-known source of a class of compounds known as curcuminoids that possess various biological activities. Curcumin (Figure 6 *c*) is of immense significance owing to its antioxidant, anti-inflammatory, anti-tumour and other pharmaceutical properties[71–73]. A lot of interest was generated in decoding its biosynthetic pathway due to its structurally complex nature. Radio-tracer studies performed decades ago[74] led to the proposal that curcuminoids are derived from the phenylpropanoid pathway and that type III PKSs could be involved[75]. ^{13}C feeding experiments carried out much later indeed showed that the curcuminoid scaffold originates from two phenylalanines[76]. Remarkably, it was eventually reported that curcumin synthesis actually requires participation of two type III PKSs – diketide CoA synthase (DCS) and curcumin synthase (CURS). DCS catalyses the formation of a diketide intermediate via condensation of feruloyl CoA with

malonyl CoA. CURS then acts upon this diketide intermediate and feruloyl CoA to yield curcumin. DCS is a unique type III PKS that releases the product as CoA-bound form[77].

Tetrahydroisoquinoline alkaloids (e.g. morphine) derived from plants are known to possess diverse biological properties. These alkaloids are known to originate from α-keto acids by Pictet–Spengler condensation with dopamine, thus ultimately arising from aromatic amino acids[78,79]. *Triphyophyllum peltatum* is a source of a unique class of naphthyloisoquinoline alkaloids like dioncophylline A with strong anti-malarial properties[76]. Dioncophylline A has an unusual structure that does not fit into the generally accepted biosynthetic scheme for alkaloids (Figure 6 *d*). Feeding experiments with ^{13}C-labelled precursors administered to callus cultures showed that biosynthesis of this alkaloid follows the acetogenic route, unlike other known alkaloids. It has been suggested that both molecular halves of the alkaloid, the naphthalene part and the isoquinoline portion, originate from six acetate units each, via joint polyketide intermediates. Interestingly, upon biotic, physical or chemical stress, the biosynthetic pathway is blocked and is instead diverted towards formation of naphthalene-related compounds – isoshinanolone, plumbagin and droserone. These acetogenic products are believed to play a chemo-ecological role by repelling plant parasites and helping in wound-healing[80].

Future perspective

It is clear from the above discussion that the PKS field has undergone incredible revamping and added new dimensions to our understanding of its versatility. Advent of the genome sequencing era promises that the task of unearthing new biosynthetic pathways only becomes bigger. To keep pace it is necessary that researchers evolve better computational software for predicting putative metabolites of a *pks* gene. Understanding the intricate programming that drives inter-domain and inter-modular interactions would help in the bioengineering of proteins that synthesize novel products which could be of immense use commercially. Recent experiments have revealed the cryptic nature of fungal iterative PKSs. For example, it has been indicated that PKS gene clusters could be silenced by histone methylation-mediated chromatin regulation. This silencing can be therefore reversed by removing genes responsible for repressive chromatin configuration[81]. It was shown in another instance that a cryptic gene cluster in *A. nidulans* could be activated by replacing the promoter of its transcriptional activator with an inducible *alcA* promoter. This manipulation led to the production of aspefuranone – an unknown fungal metabolite[82]. Decoding the mechanisms by which this complex networking is controlled is one of the major challenges in novel metabolite search. Possibilities are huge, ensuring that this area keeps mystifying scientists for long.

1. Collie, J. N. and Myers, W. S., The formation of orcinol and other condensation products from dehydroacetic acid. *J. Chem. Soc.*, 1893, **63**, 122.
2. Collie, J. N., Derivatives of the multiple keten group. *J. Chem. Soc*., 1907, **91**, 1806.
3. Birch, A. J., Massy-Westropp, R. A. and Moye, C. J., Studies in relation to biosynthesis VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Aus. J. Chem.*, 1955, **8**, 539.
4. Minto, R. E. and Townsend, C. A., Enzymology and molecular biology of aflatoxin biosynthesis. *Chem. Rev*., 1997, **97**(7), 2537–2556.
5. Cane, D. E., Prabhkaran, P. C., Tan, W. and Ott, W. R., Macrolide biosynthesis. 6 Mechanism of polyketide chain elongation. *Tetrahedron Lett.*, 1991, **32**, 5457–5460.
6. Marchler-Bauer, A. *et al*., CDD: a conserved domain database for protein classification. *Nucleic Acids Res*., 2005, **33** (Database Issue), D192–D196.
7. Mulder, N. J. *et al*., The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res*., 2003, **31**(1), 315–318.
8. Yadav, G., Gokhale, R. S. and Mohanty, D., SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res*., 2003, **31**(13), 3654–3658.
9. Ansari, M. Z., Yadav, G., Gokhale, R. S. and Mohanty, D., NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res*., 2004, **32** (Web Server Issue), W405–W413.
10. Tae, H., Kong, E. B. and Park, K., ASMPKS: an analysis system for modular polyketide synthases. *BMC Bioinformatics*, 2007, **8**, 327.
11. Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J. and Hranueli, D., 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*., 2008, **36**(21), 6882–6892.
12. Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S. and Sherman, D. H., Automated genome mining for natural products. *BMC Bioinformatics*, 2009, **10**, 185.
13. Anand, S., Prasad, M. V., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z. and Mohanty, D., SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res*., 2010, **38** (Web Server Issue), W487–W496.
14. Ginolhac, A., Jarrin, C., Robe, P., Perriere, G., Vogel, T. M., Simonet, P. and Nalin, R., Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol*., 2005, **60**, 716–725.
15. Moffitt, M. C. and Neilan, B. A., Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J. Mol. Evol*., 2003, **56**(4), 446–457.
16. Yadav, G., Gokhale, R. S. and Mohanty, D., Towards prediction of metabolic products of polyketide synthases: an *in silico* analysis. *PLoS Comput. Biol*., 2009, **5**(4), e1000351.
17. Challis, G. L., Ravel, J. and Townsend, C. A., Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol*., 2000, **7**(3), 211–224.
18. Stachelhaus, T., Mootz, H. D. and Marahiel, M. A., The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 1999, **6**(8), 493–505.
19. Yadav, G., Gokhale, R. S. and Mohanty, D., Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol*., 2003, **328**(2), 335–363.

20. Rangan, V. S. and Smith, S., Alteration of the substrate specificity of the malonyl-CoA/acetyl-CoA:acyl carrier protein S-acyltransferase domain of the multifunctional fatty acid synthase by mutation of a single arginine residue. *J. Biol. Chem.*, 1997, **272**(18), 11975–11978.
21. Minowa, Y., Araki, M. and Kanehisa, M., Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.*, 2007, **368**(5), 1500–1517.
22. Caffrey, P., Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem*, 2003, **4**(7), 654–657.
23. Keatinge-Clay, A. T., A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chemistry Biol.*, 2007, **14**(8), 898–908.
24. Baerga-Ortiz, A. *et al.*, Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. *Chemistry Biol.*, 2006, **13**(3), 277–285.
25. Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C. and Cane, D. E., The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.*, 2009, **131**(51), 18501–18511.
26. Weissman, K. J., The structural basis for docking in modular polyketide biosynthesis. *ChemBioChem*, 2006, **7**(3), 485–494.
27. Weissman, K. J., Single amino acid substitutions alter the efficiency of docking in modular polyketide biosynthesis. *ChemBioChem*, 2006, **7**(9), 1334–1342.
28. Buchholz, T. J., Geders, T. W., Bartley 3rd, F. E., Reynolds, K. A., Smith, J. L. and Sherman, D. H., Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.*, 2009, **4**(1), 41–52.
29. Udwary, D. W., Merski, M. and Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J. Mol. Biol.*, 2002, **323**(3), 585–598.
30. Cole, S. T. *et al.*, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, **393**(6685), 537–544.
31. Moody, D. B. *et al.*, CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. *Nature*, 2000, **404**(6780), 884–888.
32. Moody, D. B., Polyisoprenyl glycolipids as targets of CD1-mediated T cell responses. *Cell Mol. Life Sci.*, 2001, **58**, 1461–1474.
33. Matsunaga, I. *et al.*, *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J. Exp. Med.*, 2004, **200**(12), 1559–1569.
34. Chopra, T. *et al.*, Novel intermolecular iterative mechanism for biosynthesis of mycoketide catalyzed by a bimodular polyketide synthase. *PLoS Biol.*, 2008, **6**(7), e163.
35. Goren, M. B., Brokl, O. and Schaefer, W. B., Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: phthiocerol dimycocerosate and the attenuation indicator lipid. *Infect. Immun.*, 1974, **9**(1), 150–158.
36. Kirksey, M. A., Tischler, A. D., Simeone, R., Hisert, K. B., Uplekar, S., Guilhot, C. and McKinney, J. D., Spontaneous phthiocerol dimycocerosate-deficient variants of *Mycobacterium tuberculosis* are susceptible to gamma interferon-mediated immunity. *Infect. Immun.*, 2011, **79**(7), 2829–2838.
37. Trivedi, O. A. *et al.*, Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. *Mol. Cell*, 2005, **17**(5), 631–643.
38. Simeone, R., Constant, P., Guilhot, C., Daffe, M. and Chalut, C., Identification of the missing transacting enoyl reductase required for phthiocerol dimycocerosate and phenolglycolipid biosynthesis in *Mycobacterium tuberculosis*. *J. Bacteriol.*, 2007, **189**(13), 4597–4602.
39. Portevin, D. *et al.*, The acyl-AMP ligase FadD32 and AccD4-containing acyl-CoA carboxylase are required for the synthesis of mycolic acids and essential for mycobacterial growth: identification of the carboxylation product and determination of the acyl-CoA carboxylase components. *J. Biol. Chem.*, 2005, **280**(10), 8862–8874.
40. Bloch, K., Control mechanisms for fatty acid synthesis in *Mycobacterium smegmatis*. *Adv. Enzymol. Related Areas Mol. Biol.*, 1977, **45**, 1–84.
41. Trivedi, O. A., Arora, P., Sridharan, V., Tickoo, R., Mohanty, D. and Gokhale, R. S., Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature*, 2004, **428**(6981), 441–445.
42. Gokhale, R. S., Sankaranarayanan, R. and Mohanty, D., Versatility of polyketide synthases in generating metabolic diversity. *Curr. Opin. Struct. Biol.*, 2007, **17**(6), 736–743.
43. Portevin, D., De Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffe, M. and Guilhot, C., A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc. Natl. Acad. Sci. USA*, 2004, **101**(1), 314–319.
44. Weijer, C. J., *Dictyostelium* morphogenesis. *Curr. Opin. Genet. Develop.*, 2004, **14**, 392–398.
45. Eichinger, L. *et al.*, The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, 2005, **435**(7038), 43–57.
46. Ghosh, R. *et al.*, Dissecting the functional role of polyketide synthases in *Dictyostelium discoideum*: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1,3-diol. *J. Biol. Chem.*, 2008, **283**(17), 11348–11354.
47. Cox, R. J. *et al.*, Rapid cloning and expression of a polyketide synthase gene involved in squalestatin biosynthesis. *ChemComm*, 2004, 2260–2261.
48. Kennedy, J., Auclair, K., Kendrew, S. G., Park, C., Vederas, J. C. and Hutchinson, C. R., Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science*, 1999, **284**, 1368–1372.
49. Arai, A. *et al.*, Dictyopyrones, novel alpha-pyronoids isolated from *Dictyostelium* spp., promote stalk cell differentiation in *Dictyostelium discoideum*. *Differ. Res. Biol. Divers.*, 2005, **73**(7), 377–384.
50. Takaya, Y. *et al.*, Novel acyl alpha-pyronoids, dictyopyrone A, B, and C, from *Dictyostelium* cellular slime molds. *J. Org. Chem.*, 2000, **65**(4), 985–989.
51. Takaya, Y. K. H., Terui, Y., Komiya, J., Maeda, Y., Ito, A. and Oshima, Y., Novel aromatic substances, dictyomedin A and B, from *Dictyostelium* cellular slime molds and their inhibitory effects on *Dictyostelium* development. *Tetrahedron Lett.*, 2001, **3**, 61–63.
52. Kikuchi, H. *et al.*, Isolation and synthesis of a new aromatic compound, brefelamide, from dictyostelium cellular slime molds and its inhibitory effect on the proliferation of astrocytoma cells. *J. Org. Chem.*, 2005, **70**(22), 8854–8858.
53. Kay, R. R., The biosynthesis of differentiation-inducing factor, a chlorinated signal molecule regulating *Dictyostelium* development. *J. Biol. Chem.*, 1998, **273**(5), 2669–2675.
54. Saito, T., Kato, A. and Kay, R. R., DIF-1 induces the basal disc of the *Dictyostelium* fruiting body. *Dev. Biol.*, 2008, **317**(2), 444–453.
55. Thompson, C. R. and Kay, R. R., The role of DIF-1 signaling in *Dictyostelium* development. *Mol. Cell*, 2000, **6**(6), 1509–1514.
56. Brookman, J. J., Town, C. D., Jermyn, K. A. and Kay, R. R., Developmental regulation of a stalk cell differentiation-inducing factor in *Dictyostelium discoideum*. *Dev. Biol.*, 1982, **91**(1), 191–196.
57. Saran, S., Calcium levels during cell cycle correlate with cell fate of *Dictyostelium discoideum*. *Cell Biol. Int.*, 1999, **23**(6), 399–405.

58. Sternfeld, J. and David, C. N., Ammonia plus another factor are necessary for differentiation in submerged clumps of *Dictyostelium*. *J. Cell Sci*., 1979, **38**, 181–191.
59. Anjard, C., Chang, W. T., Gross, J. and Nellen, W., Production and activity of spore differentiation factors (SDFs) in *Dictyostelium*. *Development*, 1998, **125**(20), 4067–4075.
60. Serafimidis, I. and Kay, R. R., New prestalk and prespore inducing signals in *Dictyostelium*. *Dev. Biol*., 2005, **282**(2), 432–441.
61. Austin, M. B. *et al.*, Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nature Chem. Biol*., 2006, **2**(9), 494–502.
62. Zucko, J. S., N., Curk, T., Zupan, B., Long, P. F., Cullum, J., Kessin, R. H. and Hranueli, D., Polyketide synthase genes and the natural products potential of *Dictyostelium discoideum*. *Bioinformatics*, 2007, **23**, 2543–2549.
63. Saito, T., Taylor, G. W., Yang, J. C., Neuhaus, D., Stetsenko, D., Kato, A. and Kay, R. R., Identification of new differentiation inducing factors from *Dictyostelium discoideum*. *Biochim. Biophys. Acta*, 2006, **1760**(5), 754–761.
64. Fujii, I., Mori, Y., Watanabe, A., Kubo, Y., Tsuji, G. and Ebizuka, Y., Enzymatic synthesis of 1,3,6,8-tetrahydroxynaphthalene solely from malonyl coenzyme A by a fungal iterative type I polyketide synthase PKS1. *Biochemistry*, 2000, **39**(30), 8853–8858.
65. Taborda, C. P., da Silva, M. B., Nosanchuk, J. D. and Travassos, L. R., Melanin as a virulence factor of *Paracoccidioides brasiliensis* and other dimorphic pathogenic fungi: a minireview. *Mycopathologia*, 2008, **165**(4–5), 331–339.
66. Tsai, H. F. *et al.*, Pentaketide melanin biosynthesis in *Aspergillus fumigatus* requires chain-length shortening of a heptaketide precursor. *J. Biol. Chem.*, 2001, **276**(31), 29292–29298.
67. Paolo Jr, W. F., Dadachova, E., Mandal, P., Casadevall, A., Szaniszlo, P. J. and Nosanchuk, J. D., Effects of disrupting the polyketide synthase gene *WdPKS1* in *Wangiella* [Exophiala] *dermatitidis* on melanin production and resistance to killing by antifungal compounds, enzymatic degradation, and extremes in temperature. *BMC Microbiol*., 2006, **6**, 55.
68. Woo, P. C. *et al.*, High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. *FEBS J*., 2010, **277**(18), 3750–3758.
69. Goyal, A., Saxena, P., Rahman, A., Singh, P. K., Kasbekar, D. P., Gokhale, R. S. and Sankaranarayanan, R., Structural insights into the biosynthesis of resorcinolic lipids by a type III polyketide synthase in *Neurospora crassa*. *J. Struct. Biol*., 2008, **162**, 411–421.
70. Saxena, P., Yadav, G., Mohanty, D. and Gokhale, R. S., A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. *J. Biol. Chem*., 2003, **278**(45), 44780–44790.
71. Weber, W. M., Hunsaker, L. A., Abcouwer, S. F., Deck, L. M. and Vander Jagt, D. L., Anti-oxidant activities of curcumin and related enones. *Bioorg. Med. Chem.*, 2005, **13**(11), 3811–3820.
72. Ringman, J. M., Frautschy, S. A., Cole, G. M., Masterman, D. L. and Cummings, J. L., A potential role of the curry spice curcumin in Alzheimer's disease. *Curr. Alzheimer Res*., 2005, **2**(2), 131–136.
73. Maheshwari, R. K., Singh, A. K., Gaddipati, J. and Srimal, R. C., Multiple biological activities of curcumin: a short review. *Life Sci*., 2006, **78**(18), 2081–2087.
74. Roughley, P. J. W., D.A., Experiments in the biosynthesis of curcumin. *J. Chem. Soc. Perkins Trans 1*, 1973, **20**, 2379–2388.
75. Schroder, J., A family of plant-specific polyketide synthases: facts and predictions. *Trends Plant Sci.*, 1997, **2**, 373–378.
76. Kita, T., Imai, S., Sawada, H., Kumagai, H. and Seto, H., The biosynthetic pathway of curcuminoid in turmeric (*Curcuma longa*) as revealed by ^{13}C-labeled precursors. *Biosci. Biotechnol. Biochem*., 2008, **72**(7), 1789–1798.
77. Katsuyama, Y., Kita, T., Funa, N. and Horinouchi, S., Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. *J. Biol. Chem.*, 2009, **284**(17), 11160–11170.
78. Dalton, D. R., The alkaloids: the fundamental chemistry – a biogenetic approach. In *Studies in Organic Chemistry* (ed. Gassman, P. G.), Marcel Dekker Inc, New York, 1979, pp. 177–411.
79. Bentley, K. W., *The Isoquinoline Alkaloids*, Harwood Academic Publishers, United Kingdom, 1998.
80. Bringmann, G. and Feineis, D., Stress-related polyketide metabolism of Dioncophyllaceae and Ancistrocladaceae. *J. Exp. Bot*., 2001, **52**(363), 2015–2022.
81. Bok, J. W. *et al.*, Chromatin level regulation of biosynthetic gene clusters. *Nature Chem. Biol.*, 2009, **5**(7), 462–464.
82. Chiang, Y. M., Szewczyk, E., Davidson, A. D., Keller, N. P., Oakley, B. R. and Wang, C. C. C., A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, aspefuranone, in *Aspergillus nidulans*. *J. Am. Chem. Soc*., 2009, **131**, 2965–2970.

ACKNOWLEDGEMENTS. D.R.N., S.A. and P.V. are Senior Research Fellows of the Council of Scientific and Industrial Research, India. R.S.G. is supported by Swarnajayanti Fellowship from DST, India. This work is also supported by grants from the Council of Scientific and Industrial Research to Institute of Genomics and Integrative Biology and from Department of Biotechnology to National Institute of Immunology.