Structural studies of natural product biosynthetic proteins

Craig A Townsend

The first high-resolution structures of key proteins involved in the biosynthesis of several natural product classes are now appearing. In some cases, they have resulted in a significantly improved mechanistic understanding of the often complex processes catalyzed by these enzymes, and they have also opened the way for more rational efforts to modify the products made.

Address: Department of Chemistry, The Johns Hopkins University, Baltimore, MD 21218, USA.

Correspondence: Craig A Townsend
E-mail: townsend@jhunix.hcf.jhu.edu

Chemistry & Biology October 1997, 4:721–730
http://biomednet.com/elecref/1074552100400721

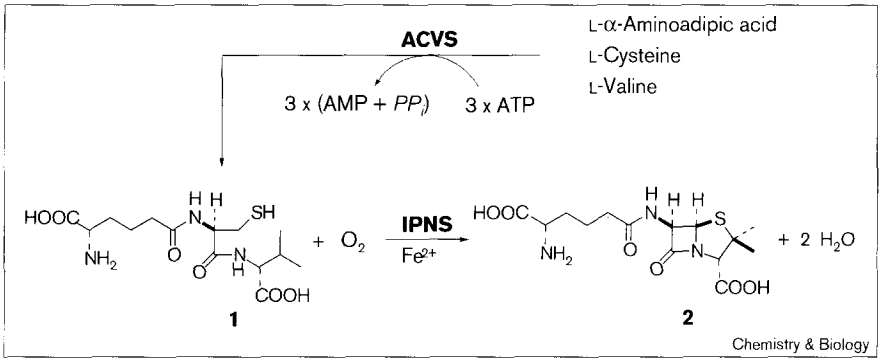
Introduction

X-ray crystallography has dramatically advanced our understanding of protein structure and the macromolecular association of proteins with DNA, RNA, drugs, cofactors and other proteins. Progress has, however, been considerably slower in obtaining detailed insights into the enzymes that catalyze natural product biosynthesis. The reasons for this are several; in particular the enzymes concerned have modest, even poor, kinetic properties, they are available in only minute amounts from wild-type organisms and they are often not very stable. As a consequence, they have proved difficult to detect, isolate and purify. Efforts towards this end have been successful in the last decade, however, and now a number of genes encoding synthetic enzymes for natural products of interest have been cloned and recombinant enzymes have been made available for thorough characterization. Given the commercial importance of natural products, the associated enzymes have become targets for manipulation by chemical modification, mutagenesis and combinatorial methods, not only to improve the production of these high-value materials but also, through deeper mechanistic understanding, to achieve the efficient synthesis of new chemical entities. It is axiomatic that both dynamic and structural information are required in order to understand the behavior of a chemical system. Bioorganic chemists (or chemical biologists!) have proved resourceful, even ingenious, in devising experiments to reveal in exquisite mechanistic detail the frequently complex catalytic cycles of biosynthetic enzymes. On the other hand, high-resolution structures of these proteins, which would be immensely useful guides to any program of protein modification for mechanistic or biotechnological ends, have not been available until very recently. Among the first of a growing wave of these structures is that of isopenicillin N synthase at 1.3 Å resolution [1], which underscores this broader development and prompts this short review.

Penicillin

The biosynthesis of penicillin from a tripeptide by a single enzyme, isopenicillin N synthase (IPNS), is among the most impressive transformations known in natural products chemistry. In the reaction, a monomeric protein containing a single ferrous ion, in the presence of one molecule of dioxygen, mediates a stepwise double oxidative cyclization of δ-(L–aminoadipyl)-L-cysteinyl-D-valine (ACV, **1**) to isopenicillin N (**2**) and two molecules of water (Figure 1) [2,3]. The intermediate L,L,D-tripeptide is in turn synthesized from its constituent L-amino acids and ATP by ACV synthetase (ACVS), a member of the wider class of nonribosomal peptide synthetases responsible for the analogous synthesis of modified polypeptides such as

Figure 1



The biosynthesis of isopenicillin **(2)** from its primary amino acid precursors. The reaction proceeds via an intermediate tripeptide, ACV (**1**). ACVS, ACV synthetase; IPNS, isopenicillin N synthase.

gramicidin and cyclosporin [4,5]. Thus, in only two enzymic steps, a compound of enormous value to human health is produced by chemical reactions that have an efficiency unmatched by contemporary synthetic organic methods and at a lower cost than that of the primary amino acid building blocks.

The X-ray crystal structure of IPNS has recently been obtained under anaerobic conditions, in the presence of Fe(II) and ACV [1]. An earlier crystal structure of the catalytically inactive Mn(II) enzyme showed the metal ion to be octahedrally coordinated by His214, Asp216, His270, Gln330 and two water molecules [6]. This ligand geometry was thought to represent the resting state of the enzyme. Notably, when ACV was bound to the native ferrous enzyme, the cysteine thiolate was found to displace the glutamine residue and the loss of a water molecule resulted in a coordination change to square pyramidal. Yet when nitric oxide, a surrogate for dioxygen, was bound to the native enzyme, the ligation geometry returned to octahedral, the conformation that is believed to represent the catalytically active enzyme complex, as was suggested by earlier electron paramagnetic resonance (EPR) and other spectroscopic studies [7,8]. Interestingly, in both of the latter X-ray structures, the valine β-hydrogen, which is abstracted late in the catalytic cycle, is directed away from the iron center, perhaps contributing to the selectivity of the early steps. This intimate view has led to new revisions of the proposed mechanism of IPNS [1]. The critical difficulty in any such proposal is to account for the formation of two rings, which necessitates the removal of four hydrogen atoms and the reduction of a molecule of dioxygen to two molecules of water. The essential features of the newly proposed catalytic cycle are outlined in Figure 2.

In the light of the crystal structure, a series of chemically interesting steps has been proposed for isopenicillin N formation. In the first ring-forming sequence, dioxygen binds in *trans* to Asp216 to generate octahedral ferric superoxide **3** (Figure 2). A radicaloid oxygen (peroxyl-like) from **3** abstracts the cysteine β-hydrogen to give the ferrous hydroperoxide **4** and the electronic equivalent of an iron-ligated thioaldehyde. Generation of the ferryl species appearing in **5** is invoked to release a hydroxide ion in the active site where deprotonation of the amide sets the stage for transannular closure to the monocyclic β-lactam **5**. Formation of the thiazolidine ring follows suggestions made previously [9] that the ferryl species **5** carries out homolytic C–H bond cleavage at the valyl β-carbon, presumably now rotated to correctly position this hydrogen so as to initiate formation of the second ring — a process indirectly supported by studies of other non-heme [iron] oxygenases such as clavaminate synthase (although they are dependent on α-ketoglutarate to activate oxygen) [10–12]. The *cis* relationship between the ACV sulfur atom and the reacting iron–oxygen species is critical to these oxidative cyclization processes [13–15].

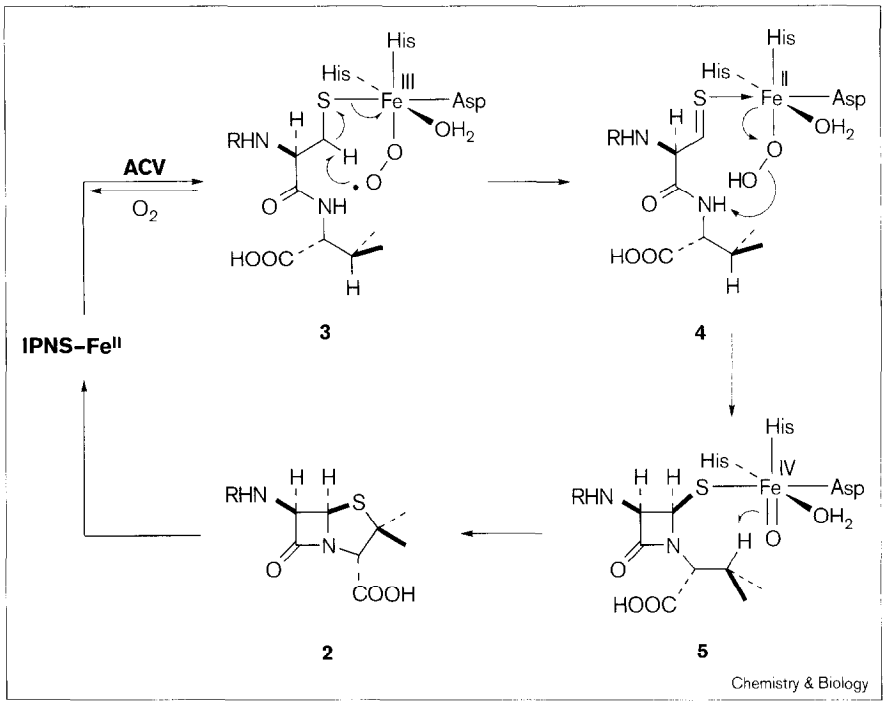
The synthesis of the tripeptide ACV and other non-ribosomal peptides

Although the sequential oxidative cyclizations carried out by IPNS are the virtuoso performance of a single active site, the assembly of its tripeptide substrate, ACV, involves the orchestration of many individual chemical steps. A single protein contains the three modular synthetase units that are required for the different steps; the units select, activate, modify (as necessary) and condense three amino acids in an order directed by the organization of the full-enzyme structure [4,5,16]. In contrast to IPNS (~35 kDa), ACVS (405–425 kDa) and other non-ribosomal peptide synthetases (NRPSs; Figure 3) are very large multifunctional enzymes. For example, gramicidin synthetase is composed of two separately encoded proteins of five modules in total (590 kDa altogether) that catalyze the assembly and head-to-tail dimerization of a pentapeptide to a cyclic decapeptide product. Cyclosporin synthetase is a single giant protein of molecular mass 1700 kDa and a total of 11 modules that combine and N-methylate natural and unnatural amino acids [4,5,17].

Contained within each of the modules of the NRPSs are the first two or more of the following catalytic domains,

Figure 2

The proposed mechanism of isopenicillin N (**2**) formation from the L,L,D-tripeptide ACV (**1**; see Figure 1). The reaction is catalyzed by IPNS.



the functions of which have been assigned on the basis of biochemical studies and high levels of sequence identity. Detailed sequence comparisons have led to the definition of domains responsible for: (1) amino acid recognition and adenylation by ATP, (2) thioesterification of the resulting mixed anhydrides by phosphopantetheine units bound at highly conserved serine residues, (3) occasional epimerization and/or N-methylation, (4) amide bond formation (condensation) and, finally, (5) intramolecular cyclization or hydrolysis by carboxy-terminal thioesterase domains, as is the case with ACVS, to release the fully elaborated peptide. The reactions of a single module are depicted in Figure 4. Type I NRPSs comprise about 1000 amino acids. N-Methylation domains, which distinguish type II from type I NRPS enzymes, are approximately 400 amino acids long and are located in the carboxy-terminal region of adenylation domains [5].

Over 300 precursor molecules are found in non-ribosomal polypeptides. The ability to assemble these units in a

Figure 3

The modular arrangement of some non-ribosomal peptide synthetases (NRPSs) and the overall reactions that they catalyze (• = SH terminus of phophopantetheine bound at a conserved serine). ACVS, ACV synthetase; GRS, gramicidin synthetase and SIM, cyclosporin synthetase.

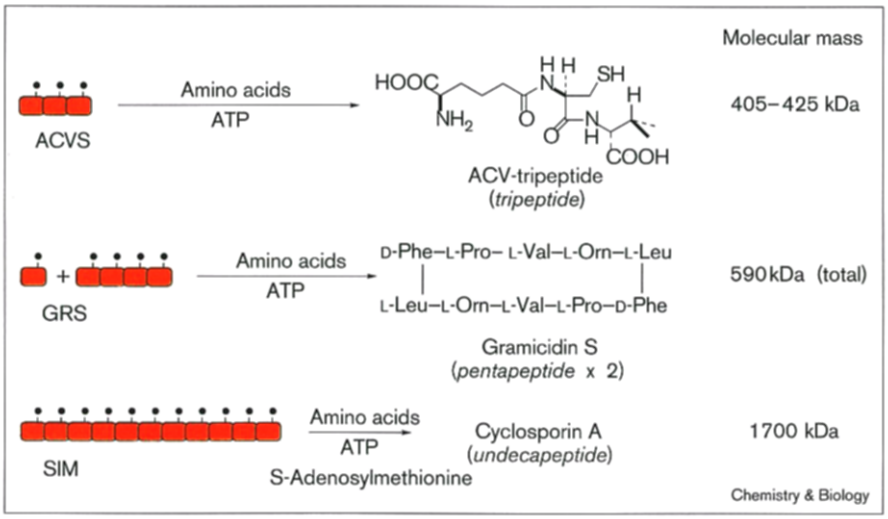
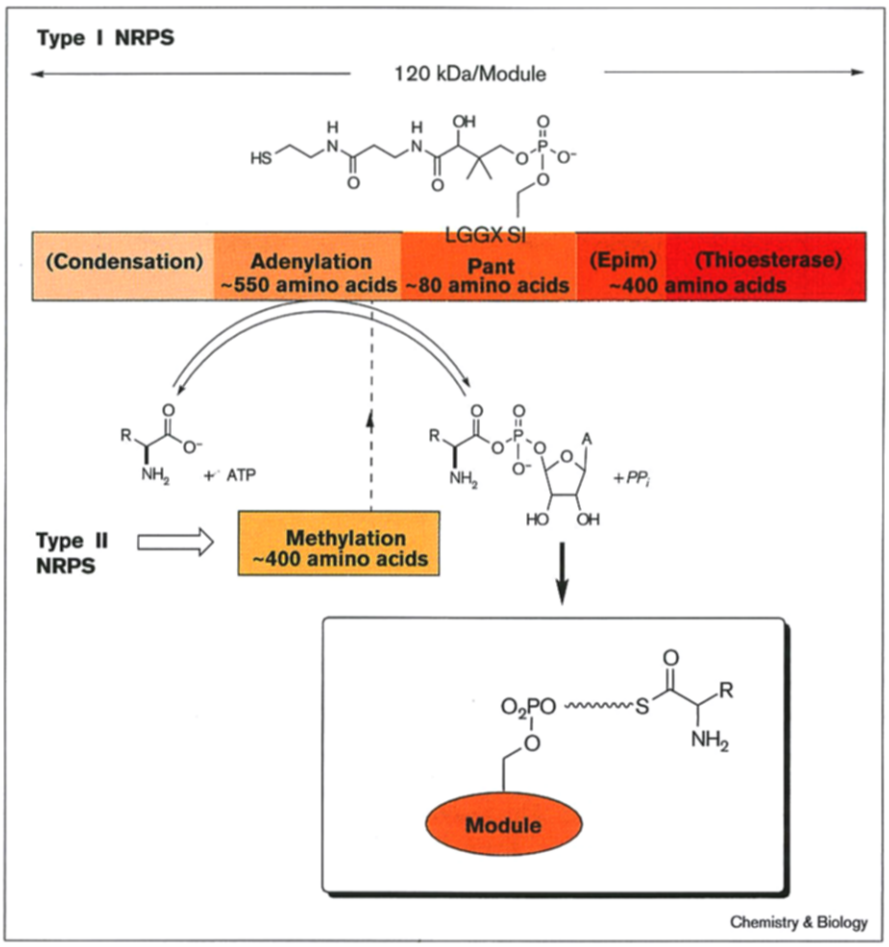


Figure 4



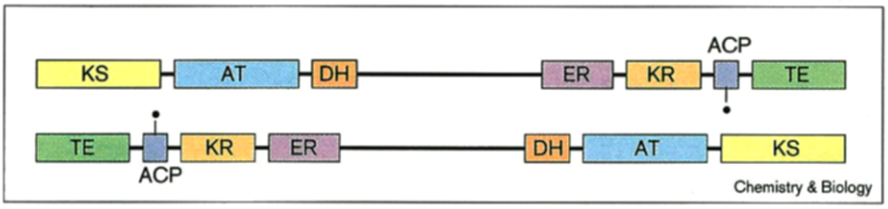
The domain organization of a generic type I NRPS module and an illustration of the thiotemplate mechanism of polypeptide synthesis. Activation of an amino acid occurs by recognition and adenylation followed by transfer to the module-bound phosphopantetheine. Epimerization ('Epim' domain) and N-methylation can take place at this stage followed, finally, by condensation to afford the polypeptide product. Type II NRPS enzymes are distinguished by an N-methylation domain (orange) in addition to those present in a type I system. The module with the product bound to phosphopantetheine is shown in the box.

predetermined order and number by combining the appropriate modules to encode a chimeric synthetase has clear mechanistic and practical attractions. In particular, it could be imagined that combinatorial methods could be applied to the synthesis of new chemical species. Indeed, experiments have been carried out to show the feasibility of such a goal [18]. The X-ray structure of an entire module would be of great value for determining the three-dimensional relationships between its several catalytic domains, and an important advance toward this goal has just been reported in the 1.9 Å structure of the phenylalanine-activating subunit of gramicidin synthetase (PheA) [17,19]. Adenylation domains play a central role in peptide formation by selecting and activating specific building blocks for synthesis. Nine well-conserved signature sequences have been identified within adenylation domains and they are believed to be involved in ATP and amino acid binding and catalysis [4]. These assignments have been supported by mutagenesis studies. The PheA model structure consists of 512 residues which is in good agreement with the size of an entire adenylation domain with both phenylalanine and AMP bound to it. The structure reveals several invariant or highly conserved residues within the signature motifs that are thought to be involved in AMP binding or the interaction of the enzyme with the α-amino and α-carboxyl of the substrate phenylalanine. Although proteins with this level of sequence identity are expected to have highly similar main chain conformations, the determinants of amino acid binding specificity are governed by the nature of the protein residues lining the substrate-binding pocket. As would be anticipated, differences in these regions are evident in individual adenylation domains.

The model polypeptide folds into two compact subdomains; a large amino-terminal portion that contributes most substrate contacts and a significantly smaller carboxy-terminal part. Although it shares only 16% amino acid identity, principally in short, highly conserved motifs, the X-ray crystal structure of unligated firefly luciferase shows remarkable topological similarity to the PheA model structure [20]. One would expect that all adenylation domains, which have 30–60% sequence identity, have very similar three-dimensional structures. Interestingly, in the model PheA structure, the small carboxy-terminal domain is

Figure 5

A linear domainal representation of an animal fatty acid synthase. KS, β-ketoacyl synthase; AT, acyl transferase (acetyl and malonyl); DH, dehydratase; ER, enoyl reductase; KR, β-ketoacyl reductase; ACP, acyl carrier protein; TE, thioesterase. The relative sizes of the domains and the DH–ER linker are approximately to scale.



rotated approximately 90° relative to its orientation in firefly luciferase. Although the model PheA structure includes substrate and AMP, the luciferase structure does not. These conformational differences may reflect distinct points in the catalytic cycles of these proteins — an important dynamic consideration to any attempt at protein engineering.

Modular polyketide synthases

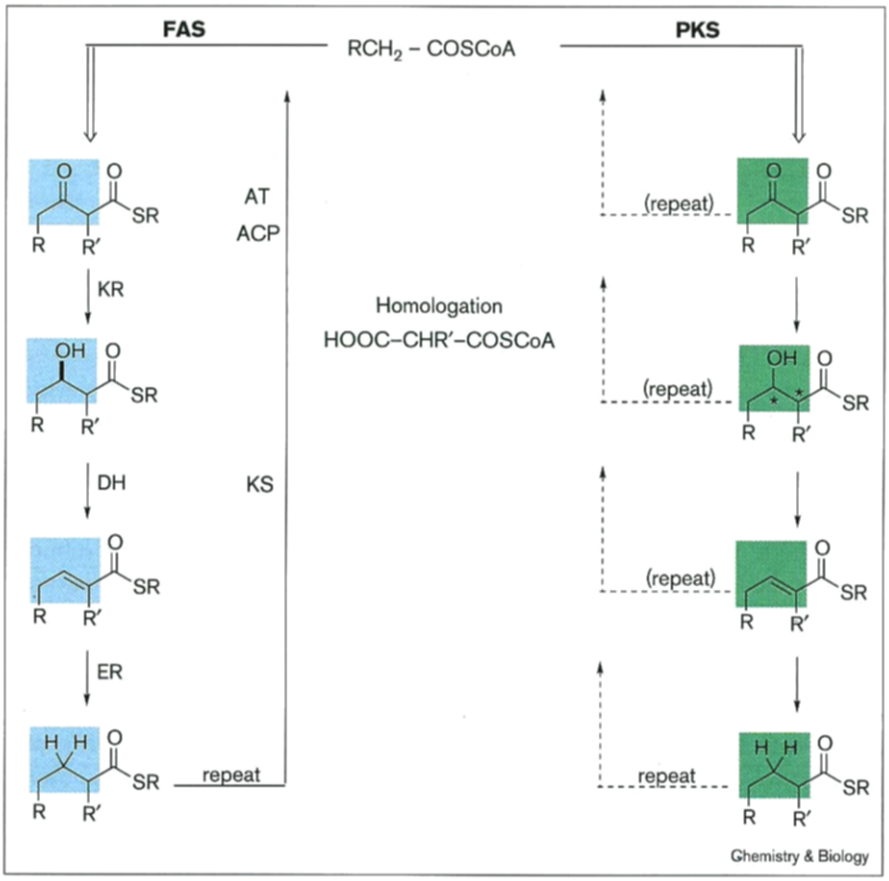
Another major group of multidomainal biosynthetic proteins that has achieved prominence recently is the modular bacterial polyketide synthases (PKSs), responsible for the biosynthesis of, for example, erythromycin [21–24] and rapamycin [25]. Interestingly, both the primary sequences and the organization of individual catalytic domains in PKSs closely mimic vertebrate fatty acid synthases (FASs). Animal FASs are highly evolved polyproteins (Type I) that contain seven catalytic sites: β-ketoacyl synthase (KS), acyl transferase (AT), dehydratase (DH), enoyl reductase (ER), β-ketoreductase (KR), acyl carrier protein (ACP) and thioesterase (TE) arranged in that order from the amino terminus to the carboxy terminus of the protein (Figure 5). The active form of the protein exists as a head-to-tail homodimer (α_2) in which the penultimate ACP domains, each with its bound phosphopantetheine arm (Figure 5; • = SH), 'carry' the intermediates of synthesis through successive rounds of two-carbon extension (homologation) by decarboxylative Claisen additions of malonyl-ACP, followed by reduction/dehydration/reduction steps until a chain length of C_{16} or C_{18} is reached. At this point the acylpantetheinate of ACP is hydrolyzed at the thioesterase and released as the free acid. The whole process is shown schematically in Figure 6.

The two central features of the process are that the active sites are used iteratively and that synthesis requires the cooperation of catalytic centers located on both of the antiparallel subunits [26,27]. Clever complementation experiments by Stuart Smith's group [28,29] have established that a single catalytic complex consists of the DH–ER–KR–ACP–TE of one strand and the AT–KS of the other, despite the occurrence of a ~600 residue linker separating the DH and ER domains (Figure 5). The subunits of FAS therefore cannot be positioned in a strictly extended, side-by-side manner as was suggested earlier by the analysis of electron microscope (EM) images [30,31], but must be coiled in some manner to accommodate the physical constraints to catalysis that are imposed by the length of the phosphopantetheine arm.

In contrast to vertebrate fatty acid biosynthesis in which acetyl-CoA (Figure 6, R = H) and malonyl-CoA (Figure 6, R′ = H) cycle through FAS in a repeated manner, microbial polyketide synthesis extends these basic chemical reactions dramatically, especially in bacteria. Starter units are variable (Figure 6, R = H, CH_3, ethyl, butyl, isobutyl, alkyl, arylalkyl, etc.), and homologations by substituted malonyl units are common (Figure 6, R′ = H, CH_3, ethyl, butyl, etc.). Similarly, reductive transformations that proceed in lock step at the β-carbon in fatty acid synthesis can stop at any step and homologation can continue. Moreover, the β-hydroxyl intermediate can, in principle, exist in any of four possible stereochemical configurations (Figure 6, *) at the α and β carbons.

It can be readily appreciated that compounds of great structural diversity can be rapidly assembled using the simple synthetic reactions available to PKSs. How Nature has done this is perhaps best understood through investigations of the modular PKS of 6-deoxyerythronolide B (6-dEB; **6**; Figure 7) biosynthesis in *Saccharopolyspora erythraea*. In this organism, three large open reading frames were discovered that encode three large proteins, DEBS 1–DEBS 3 [21-24]. Assignments of individual catalytic domains were readily made on the basis of their high sequence homology to equivalent domains in animal FAS. In keeping with a 'processive' model of polyketide biosynthesis, six modules could be assigned that corresponded to the initial loading of a propionate starter unit and five chain extensions of (2*S*)-methylmalonyl-CoA [32] and accompanying modifications (Figure 7). In contrast to animal FASs, modular PKSs are elaborately constructed such that each active site is used non-iteratively (if at all) [25]. A telling early experiment both supported the validity of the multifunctional catalytic roles proposed for the DEBS proteins and raised unmistakably the possibility of engineering these proteins in a rational way for the production of new 'natural' products. Katz and his coworkers at Abbott Laboratories [22] generated a mutant of DEBS 3 in which the ketoreductase in module 5 (*KR*; Figure 7) was disrupted. The mutant produced an analog (**7**) of 6-dEB bearing a keto group at C–5 rather than a hydroxyl group as had been hoped.

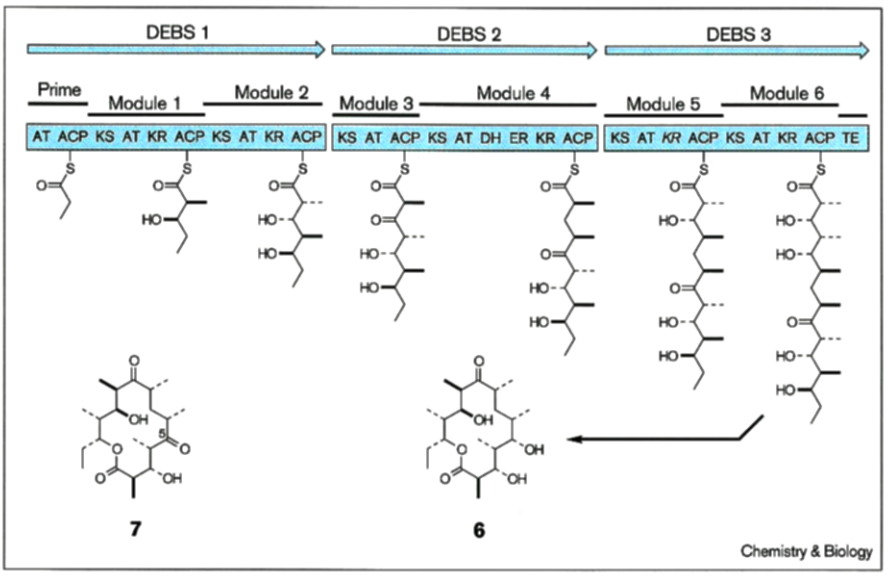
Figure 6



A comparison of FAS homologation of acetyl-CoA with malonyl-CoA (and their substituted derivatives with different groups at R and R′) with a generic PKS that shows a varying extent of reduction/dehydration/reduction at the thioacyl β-carbon prior to further homologation (highlighted in color). Both *R*- and *S*-configurations are possible at the positions marked (*). Abbreviations as for Figure 5.

Erythromycin is a front-line antibiotic and the commercial importance of other polyketide metabolites is also well appreciated [33–35]. Using the 6-dEB system, in particular, remarkable strides are currently being made to exploit the biotechnological possibilities of the PKSs [36–38]. Similarly, massive up-regulation of human FAS has been associated

Figure 7



The modular organization of 6-deoxyerythronolide B synthase (DEBS) and the proposed build up of 6-deoxyerythomycin B (**6**, 6-dEB) in a processive fashion from the initiation of synthesis at module 1 to final macrocyclization occurring at the thioesterase domain of module 6. Other abbreviations as for Figure 5.

with cancer *inter alia* of the breast, ovary and prostate [39]. Detailed tertiary and quarternary information about these proteins would be very useful to guide the design of drugs targeted to human FAS and to allow selective domainal modifications or combinatorial methods of *de novo* domain assembly in PKSs. The methods that have been applied to modular PKS structures have followed those that were used earlier for FAS. The Leadlay and Staunton groups [40,41] have examined the behavior of the DEBS proteins during gel filtration and analytical ultracentrifugation and found them, like vertebrate FAS, to be homodimers. Cross-linking and partial proteolysis studies clearly indicate a structural motif in which the inter-subunit arrangement is overall a parallel rather than an anti-parallel arrangement [40,41]. But, just as the extended, side-by-side cylinders view of animal FAS is now seen as oversimplified [29], so must the PKS subunits be intertwined to achieve synthesis. Experiments by the Khosla and Cane groups [42] suggest that, like Type I FASs, the ACP on one chain must cooperate with the KS on the partner chain. They go on to propose a dimer model in which the individual modules of one chain zig-zag back and forth in one plane and are arrayed with their partner in such a way that each module pair is oriented head-to-tail [42]. In contrast, Leadlay and Staunton [41] have advanced a 'double helical' model in which the chains are twisted together in a parallel head-to-head, tail-to-tail fashion; this model can also account for the new results of Smith for the Type I FAS [29].

Electron microscopy, physical measurements, cross-linking and proteolysis experiments have provided low-resolution views of FAS [30,31] and, to a lesser extent, of the PKS enzymes [40–42]. Closely related to the studies described above have been those carried out with the fungal Type I PKS 6-methylsalicylic acid synthase [43] and the first, seminal experiments in this area by Lynen [44] on yeast FAS two decades ago. Unfortunately, high resolution information about Type I systems from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy is lacking. In bacteria, however, most fatty acid biosynthesis is carried out by particulate enzymes that have single active sites (Type II FASs). Some bacterial PKSs are particulate as well, such as those involved in the synthesis of actinorhodin [33] and tetracenomycin [35]. These bacterial enzymes are smaller, more tractable targets for detailed structural analysis and have started to become a focus for experimental effort. They are also significantly homologous to individual domains of their Type I FAS and PKS brethren and, hence, their structures are likely to be broadly representative of the corresponding domains of these multifunctional proteins.

The first advance in the structural analysis of the bacterial enzymes was reported by Kim and Prestegard [45,46] with the solution structure of *holo*-ACP from the Type II FAS of *Escherichia coli*. In an important correlation with Type II PKSs, Simpson and coworkers [47] have recently described a high-resolution NMR structure of the *apo*-ACP of actinorhodin (*act*) synthesis from *Streptomyces coelicolor*; nuclear Overhauser enhancement spectroscopy (NOESY) experiments suggest that there is little interaction between phosphopantetheine and ACP. The tertiary folds of the two proteins show strong structural homology, despite the generally observed failure of FAS ACPs to substitute for PKS ACPs in genetically engineered systems. Both ACPs share a hydrophobic core that presumably accommodates the growing acyl chain as synthesis proceeds. It is quite flexible, and differing extents of mobility, and different charged and hydrophilic residues lie buried within this core. These dynamic and structural differences may account for the synthetic discrimination shown by the different ACPs [47].

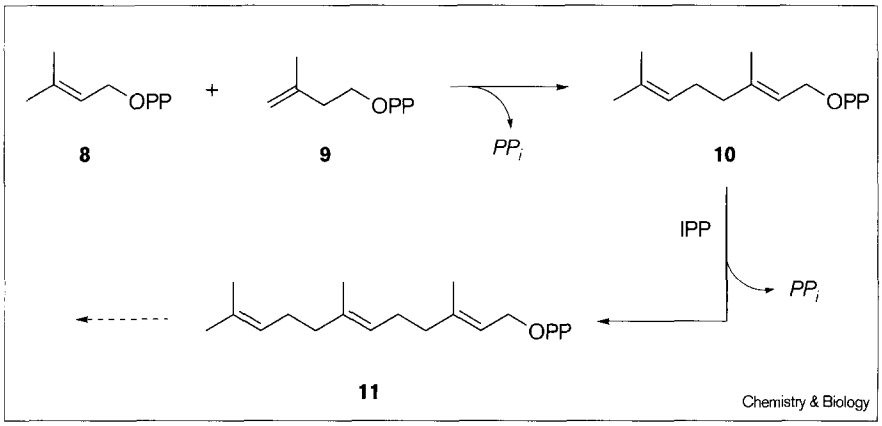
X-ray structures of two further proteins of the *E. coli* FAS apparatus have been determined recently. The acyl transferase that transfers malonate from CoA to ACP has been crystallized to afford a structure at 1.5 Å [48]. The catalytic Ser92 is revealed as hydrogen bonded to a histidine, His201, which is typical of hydrolases, but the catalytic triad appears to be completed by Gln250 rather than the usual carboxylic acid of aspartate or glutamate. The enzyme is known to be highly discriminating for malonyl-CoA in preference to acetyl-CoA, but the origin of this substrate specificity is not fully clear from the X-ray structure, although a nearby arginine may be involved. Enoyl reductase catalyzes the last reaction in the catalytic cycle of FAS and requires NADPH (Figure 6); the *E. coli* enzyme has been crystallized in the presence of NAD^+ and its structure solved at 2.1 Å [49]. Although the NAD^+ cofactor presumably occupies the expected binding site, the location of the α,β-unsaturated thioester substrate is not known.

Terpenes

The terpenes are the final class of natural products in which X-ray crystallography has begun to provide fundamental mechanistic insights. More than 23,000 terpenes are known of the widest structural variety of any known class of natural product; they range from being essential membrane components, steroid hormones and insecticides to carotenoids, flavors and fragrances. Central to the formation of all of these products are the prenyl transferases in which dimethylallyl diphosphate (**8**; DMAPP; Figure 8) and isopentyl diphosphate (**9**; IPP), the basic building blocks of isoprenoids, are condensed to give geranyl diphosphate (**10**; GPP) and inorganic pyrophosphate. In an analogous reaction, IPP can react again with this allylic diphosphate intermediate to afford farnesyl diphosphate (**11**; FPP) and so forth (Figure 8).

Avian FPP synthase has been crystallized as a dimer and its structure has been determined at 2.6 Å resolution [50]. The structure is highly α-helical; each subunit is composed

Figure 8



The assembly of linear isoprenoids by prenyl transferase enzymes.

of ten core helices surrounding a large central cavity containing the putative active site. Highly conserved residues line this site: notably two Asp–Asp–X–X–Asp motifs, which reside on opposing walls of the cavity, as well as specific arginine residues. The roles of these amino acids in substrate binding and catalysis have been demonstrated by site-specific mutagenesis experiments. Mg(II)–IPP has been co-crystallized with the prenyl transferase but, unfortunately, was insufficiently resolved to allow its precise placement within the active site. Although the mechanism of fundamental chain elongation reactions has been studied in detail, a further understanding of the process on the basis of the FPP synthase structure has been largely elusive, apart from the organization of conserved amino acids in the active site.

The central issue of the control of chain length during the prenyl transferase reaction has been addressed in an acutely rational manner [51]. An examination of the X-ray structure of FPP synthase and comparisons among 35 protein sequences of GPP and FPP synthases and higher chain-length synthases points to the potentially important roles of a pair of phenylalanine residues that form a 'floor' to the presumed allylic substrate-binding pocket. In contrast, higher chain-length synthases have smaller or more flexible sidechain residues, such as serine, and alanine or methionine, at these sites. Mutants of the avian FPP synthase bearing alanine and/or serine substitutions at these aromatic amino acid sites were constructed and their reactions were compared with those of the wild-type enzyme [51]. A single amino acid replacement, Phe112→Ala112, extended the chain length cleanly by one isoprene unit to C_{20}. Substitution of both aryl groups in the Phe112→Ala112/Phe113→Ser113 double mutant, on the other hand, dramatically affected the products formed. A distribution of C_{20}–C_{70} isoprenoids resulted, with the peak at C_{35}–C_{40}. The *apo*-enzyme and structures containing GPP and FPP could be obtained, plainly showing binding of the allylic diphosphate substrates and a pair of Mg(II) ions at one of the Asp–Asp–X–X–Asp sites with their hydrocarbon tails extending toward the mutated residues constituting the altered floor of the active site. Although longer chain allylic diphosphates could not be co-cystallized, molecular modeling exercises disclosed that the rotation of just two sidechains opened a passageway that was lined mainly with hydrophobic residues. This molecular picture provides the sharpest insight into how prenyl transferases catalyze the fundamental elongation of C_5 units and control their length. This view of terpene biosynthesis will be widened by the X-ray structures of three terpene cyclases that are soon to be published (these structures have been published whilst this review was in press [52–54]). The terpene cyclases take the reaction products of prenyl transferases, like FPP synthase, and catalyze the formation of the familiar cyclic structural types that are characteristic of this natural product class.

Conclusions

At the heart of mechanistic enzymology lies a central interest in catalysis — how do enzymes achieve their extraordinary rate accelerations? As a rule, the enzymes of natural product biosynthesis are not so impressive for their rates of synthesis but rather for the reactions that they catalyze. For natural product enzymes, investigations tend to be skewed towards determining the mechanism and its relationship to gross structure, as well as to the pragmatic considerations of how that knowledge can be applied. The point I wish to make is that a threshold has been crossed in the study of natural product biosynthetic enzymes that, while the inevitable consequence of hard-won advances in protein purification and molecular biology, has elevated the level of mechanistic discussion to include detailed three-dimensional information about protein structure. Caveats to interpreting site-specifically mutated enzyme behavior to correspond to that of wild-type enzymes remain as they were when pointed out a decade ago [55], and limitations can be seen for cases in which substrates and/or cofactors cannot be co-crystallized or successfully

diffused into presumed active sites. Less obvious is the fact that the catalyzed reactions are multistep and inescapably involve multiple transition states. The effects of protein structural changes, or indeed substrate structural changes, on the relative energies of these states (and on overall reaction rates and mechanism) will be difficult to assess or predict. Linked to this conundrum is the probable conformational plasticity of active sites that direct reactions through multiple intermediates to one or a number of products. IPNS, clavaminate synthase [56] or the terpene cyclases may be broadly cited as examples. An X-ray crystal structure affords a snapshot and does not in itself give a picture of whatever dynamic flexibility may be required to guide all phases of these catalytic events. Nonetheless, it is clear that altered substrate specificities will be achieved (greatly aided by the availability of high-resolution protein structures) and multidomain proteins will be created that synthesize new 'natural' products. These are exciting prospects. It is important to note that given the synthetic capability of natural product enzymes, only a small subset of possible structures exists (or has yet been discovered). Although the absence of some theoretically possible products may be a result of unfavorable thermodynamics, other selective pressures must be at work in Nature as well. It will be interesting to see if, for example, the antibiotic properties of the macrolide class of natural products can be improved beyond the outcome of these forces.

Acknowledgements

I am pleased to acknowledge and thank the National Institutes of Health for its support of the biosynthetic projects ongoing in my laboratory (AI14937 and ES01670).

References

1. Roach, P.L., *et al.*, & Baldwin, J.E. (1997). Structure of isopenicillin *N* synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827-830.
2. Baldwin, J.E., Adlington, R.M., Moroney, S.E., Field, L.D. & Ting, H.-H. (1984). Stepwise ring closure in penicillin biosynthesis. Initial β-lactam formation. *J. Chem. Soc., Chem. Commun.*, 984-986.
3. White, R.L., John, E.-M.M., Baldwin, J.E. & Abraham, E.P. (1982). Stoichiometry of oxygen consumption in the biosynthesis of isopenicillin from a tripeptide. *Biochem. J.* **203**, 791-793.
4. Kleinkauf, H. & Von Döhren, H. (1996). A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.* **236**, 335-351.
5. Stachelhaus, T. & Marahiel, M.A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* **125**, 3-14.
6. Roach, P.L., *et al.*, & Baldwin, J.E. (1995). Crystal structure of isopenicillin *N* synthase is the first from a new structural family of enzymes. *Nature* **375**, 700-704.
7. Orville, A.M., *et al.*, & Lipscomb, J.D. (1992). Thiolate ligation of the active site Fe^{2+} of isopenicillin N synthase derives from substrate rather than endogenous cysteine: spectroscopic studies of site-specific Cys→Ser mutated enzymes. *Biochemistry* **31**, 4602-4612.
8. Que, J., L. & Ho, R.Y.N. (1996). Dioxygen activation by enzymes with mononuclear non-heme iron active sites. *Chem. Rev.* **96**, 2607-2624.
9. Baldwin, J.E. & Abraham, E. (1988). The biosynthesis of penicillins and cephalosporins. *Nat. Prod. Reports* **5**, 129-145.
10. Krol, W.J., Basak, A., Salowe, S.P. & Townsend, C.A. (1989). Oxidative cyclization chemistry catalyzed by clavaminate synthase. *J. Am. Chem. Soc.* **111**, 7625-7627.
11. Basak, A., Salowe, S.P. & Townsend, C.A. (1990). Stereochemical course of the key ring-forming reactions in clavulanic acid biosynthesis. *J. Am. Chem. Soc.* **112**, 1654-1656.
12. Salowe, S.P., Krol, W.J., Iwata-Reuyl, D. & Townsend, C.A. (1991). Elucidation of the order of oxidations and identification of an intermediate in the multistep clavaminate synthase reaction. *Biochemistry* **30**, 2281-2292.
13. Townsend, C.A. & Basak, A. (1991). Experiments and speculations on the role of oxidative cyclization chemistry in natural products biosynthesis. *Tetrahedron* **47**, 2591-2602.
14. Townsend, C.A. (1992). Oxidative amino aid processing in β-lactam antibiotic biosynthesis. *Biochem. Soc. Trans.* **21**, 208-213.
15. Leising, R.A., Zang, Y. & Que, J., L. (1992). Oxidative ligand transfer to alkanes: a model for iron-mediated C–X bond formation in β-lactam antibiotic biosynthesis. *J. Am. Chem. Soc.* **113**, 8555-8557.
16. Cooper, R.D.G. (1993). The enzymes involved in biosynthesis of penicillin and cephalosporin; their structure and function. *Bioorg. & Med. Chem.* **1**, 1-17.
17. Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chem. Biol.* **4**, 561-567.
18. Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69-72.
19. Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174-4183.
20. Conti, E., Franks, N.P. & Brick, P. (1990). Crystal structure of firefly luciferase throws light on a super family of adenylate-forming enzymes. *Structure* **4**, 287-298.
21. Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. & Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176-178.
22. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675-679.
23. Donadio, S. & Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111**, 51-60.
24. Bevitt, D.J., Cortes, J., Haydock, S.F. & Leadlay, P.F. (1992). 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.* **204**, 39-49.
25. Schwecke, T., *et al.*, & Leadlay, P.F. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl Acad. Sci. USA* **92**, 7839-7843.
26. Wakil, S.J. (1989). Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* **28**, 4523-4530.
27. Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* **8**, 1248-1259.
28. Witkowski, A., Joshi, A. & Smith, S. (1996). Fatty acid synthase: *in vitro* complementation of inactive mutants. *Biochemistry* **35**, 10569-10575.
29. Joshi, A.K., Witkowski, A. & Smith, S. (1997). Mapping of functional interactions between domains of the animal fatty acid synthase by mutant complementation *in vitro*. *Biochemistry* **36**, 2316-2322.
30. Stoops, J.K., Wakil, S.J., Uberbacher, E.C. & Bunick, G.J. (1987). Small-angle neutron-scattering and electron microscope studies of the chicken liver fatty acid synthase. *J. Biol. Chem.* **262**, 10246-10251.
31. Kitamoto, T., Nishigai, M., Sasaki, T. & Ikai, A. (1988). Structure of fatty acid synthetase from the harderian gland of guinea pig. Proteolytic dissection and electron microscopic studies. *J. Mol. Biol.* **203**, 183-195.
32. Marsden, A.F.A., Caffrey, P., Aparicio, J.F., Loughran, M.S., Staunton, J. & Leadlay, P.F. (1994). Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263**, 378-380.
33. Hopwood, D.A. & Sherman, D.H. (1990). Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24**, 37-66.
34. Katz, L. & Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* **47**, 875-912.
35. Hutchinson, C.R. & Fujii, I. (1995). Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **49**, 201-238.
36. Pieper, R., Kao, C. & Khosla, C. (1996). Specificity and versatility in erythromycin biosynthesis. *Chem. Soc. Rev.*, 297-302.
37. Jacobsen, J.R., Hutchinson, C.R., Cane, D.E. & Khosla, C. (1997). Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277**, 367-369.
38. Staunton, J. & Wilkinson, B. (1997). The biosynthesis of erythomycin and rapamycin. *Chem. Rev.*, in press.
39. Kuhajda, F.P., *et al.*, & Pasternack, G.R. (1994). Fatty acid synthesis: a potential selective target for antineoplastic therapy. *Proc. Natl Acad. Sci. USA* **91**, 6379-6383.

40. Aparicio, J.F., Caffrey, P., Marsden, A.F.A., Staunton, J. & Leadlay, P.F. (1994). Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *J. Biol. Chem.* **269**, 8524-8528.
41. Staunton, J., Caffrey, P., Aparicio, J.F., Roberts, G.A., Bethel, S.S. & Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3**, 188-192.
42. Kao, C.M., Pieper, R., Cane, D.E. & Khosla, C. (1996). Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry* **35**, 12363-12368.
43. Child, C.J., Spencer, J.B., Bhogal, P. & Shoolingin-Jordan, P.M. (1996). Structural similarities between 6-methylsalicylic acid synthase from *Penicillium patulum* and vertebrate type I fatty acid synthase: evidence from thiol modification studies. *Biochemistry* **35**, 12267-12274.
44. Lynen, F. (1980). On the structure of fatty acid synthetase of yeast. *Eur. J. Biochem.* **112**, 431-442.
45. Kim, Y. & Prestegard, J.H. (1989). A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28**, 8792-8797.
46. Kim, Y. & Prestegard, J.H. (1990). Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins* **8**, 377-385.
47. Crump, M.P., *et al.*, & Simpson, T.J. (1997). Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **36**, 6000-6008.
48. Serre, L., Verbree, E.C., Dauter, Z., Stuitje, A.R. & Derewenda, Z.S. (1995). The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5 Å resolution. *J. Biol. Chem.* **270**, 12961-12964.
49. Baldock, C., *et al.*, & Rice, D.W. (1996). A mechanism of drug action revealed by structural studies of enoyl reductase. *Science* **274**, 2107-2110.
50. Tarshis, L.C., Yan, M., Poulter, C.D. & Sacchettini, J.C. (1994). Crystal structure of recombinant farnesyl diphosphate synthase at 2.6 Å resolution. *Biochemistry* **33**, 10871-10877.
51. Tarshis, L.C., Proteau, P.J., Kellog, B.A., Sacchettini, J.C. & Poulter, C.D. (1996). Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl Acad. Sci. USA* **93**, 15918-15023.
52. Wendt, K.U., Poralla, K. & Schultz, G.E. (1997). Structure and function of a squalene cyclase. *Science* **227**, 1811-1815.
53. Starks, C.M., Back, K. Chappell, J. & Noel, J.P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **227**, 1815-1820.
54. Lesburg, C.A., Zhai, G., Cane, D.E. & Christianson, D.W. (1997). Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science* **227**, 1820-1824.
55. Knowles, J. (1987). Tinkering with enzymes: what are we learning? *Science* **236**, 1252-1258.
56. Busby, R.W. & Townsend, C.A. (1996). A single monomeric iron center in clavaminate synthase catalyzes three nonsuccessive oxidative transformations. *Bioorg. & Med. Chem.* **4**, 1059-1064.