Biosynthesis of Erythromycin and Rapamycin

James Staunton*

Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK

Barrie Wilkinson

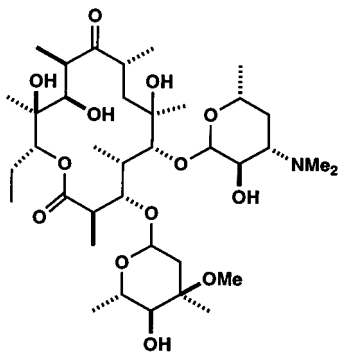
Bioprocessing Research Unit, GlaxoWellcome Research and Development, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

Received April 9, 1997 (Revised Manuscript Received June 26, 1997)

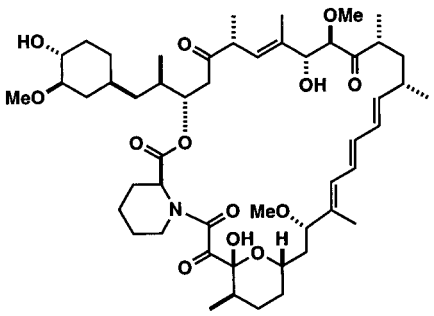
Contents

I. Introduction

Erythromycin and rapamycin are both macrolide antibiotics.[1,2] Erythromycin A (**1**) belongs to a large family of macrolides which are characterized by rings containing 12, 14, or 16 atoms. Members of the family show an intriguing common structural and stereochemical relationship which was first noted in Celmer's Rules.[3] Rapamycin (**2**) belongs to a different less populous family of metabolites in which a polyketide chain is linked to an amino acid in the macrocyclic ring. This type of metabolite is conceptually a hybrid of a polyketide and a polypeptide structure.[4] The reason for considering both compounds here is that they are the most thoroughly investigated aliphatic polyketides at the genetic level and so make an interesting comparative study.



(1)

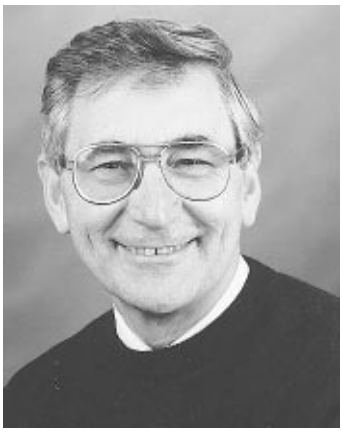


(2)

Erythromycin A was isolated in 1952 from *Saccharopolyspora erythraea*.[5] It is used in clinical medicine against infections caused by Gram-positive bacteria. It is also used for many pulmonary infections such as Legionnaire's disease and as an alternative for patients allergic to penicillins. The structure was elucidated in 1957,[6] and the absolute stereochemistry was assigned by X-ray crystallography in 1965.[7] The complete synthesis of erythromycin A has been accomplished only by Woodward and co-workers,[8] although other groups have made simpler analogues.[9]

Rapamycin was isolated in 1975 from a species of *Streptomyces hygroscopicus* native to Easter Island.[10,11] It showed antifungal, antitumor, and immunosuppressant activity.[12] Since the recognition that its immunosuppressant properties could be therapeutically useful, there has been much interest in the chemistry and biology of rapamycin, and of the structurally similar immunosuppressants FK506 (**3**) and FK520 (**4**). A model for their mode of action has been developed.[13] Biosynthetic studies have pro-

S0009-2665(96)00031-3 CCC: $28.00

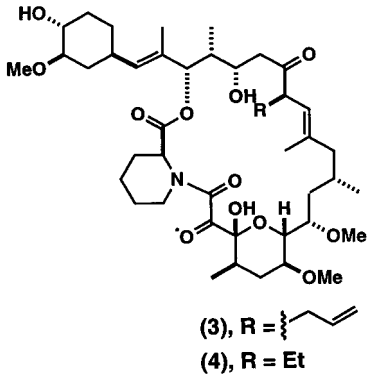


Jim Staunton graduated from Liverpool University with a B.Sc. in 1956, and was awarded a Ph.D. in 1959 for work on natural product structures. After further training at Stanford in the Djerassi group, he returned to the UK to take up a post at Liverpool University. For the next ten years he collaborated with Professor A. R. Battersby in studies of alkaloid biosynthesis. Following a move to Cambridge in 1969 he became fascinated by polyketides. Although primarily interested in biosynthesis, he has made contributions to methods of biomimetic synthesis and enantiospecific synthesis in this field. The need for improved methods of analysis in biosynthetic studies has also led him to develop novel NMR and mass spectrometry techniques for the study of isotopically labeled compounds and latterly of proteins. For the past ten years he has run jointly with Peter Leadlay a multidisciplinary team of chemists and biologists dedicated to studying polyketide biosynthesis by the full range of modern techniques.

Barrie Wilkinson was born in Sheffield, UK, in 1969. He studied chemistry at the Universtity of Leeds, UK, obtaining a B.Sc. in 1990 and a Ph.D. in 1993. His Ph.D. research was carried out under the guidance of Dr. Richard B. Herbert on studies concerning the isolation, characterization, and synthetic applications of a *Streptomyces* aldolase. This was followed by a first postdoctoral stint with Professor Heinz G. Floss at the University of Washington (Seattle, WA), where he applied chiral methyl group methodology to the study of the C–H bond functionalization mechanism of both soluble and particulate methane monooxygenases. Barrie obtained a second postdoctoral position with Drs. James Staunton and Peter F. Leadlay at the University of Cambridge, UK, where he investigated various aspects of aliphatic polyketide biosynthesis. In January 1997 he joined the Bioprocessing Research Unit of GlaxoWellcome (Stevenage, UK), and is presently enjoying a secondment back to Cambridge in order to learn molecular biology in the Leadlay–Staunton group.

ceeded in parallel on all three metabolites so cross reference will be made where appropriate later.

The structure of rapamycin was determined by X-ray crystallography in 1978.[14] Complementary NMR and degradation studies were reported in 1980[15] and revised NMR assignments were reported in 1991.[16] The solution conformation was recently



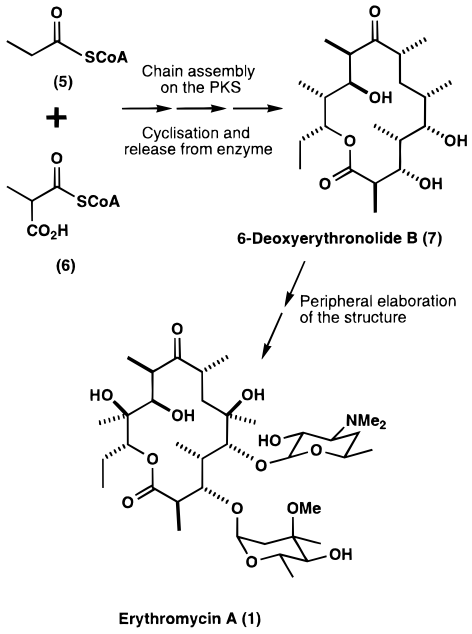
determined by NMR and molecular dynamics studies.[17] Four total syntheses have been described.[18−21]

II. Erythromycin Biosynthesis

A. Overview

The biosynthesis of erythromycin can be divided into two phases (Scheme 1). In the first phase the polyketide synthase (PKS) catalyzes sequential condensation of one unit of propionyl CoA (**5**) and six units of methylmalonyl CoA (**6**) to give 6-deoxyerythronolide B (**7**), the first enzyme-free intermediate.[22,23] In the second phase 6-deoxyerythronolide B is elaborated by a series of "tailoring" enzymes which include regiospecific hydroxylases, glycosyl transferases, and methyl transferases. From the biosynthetic point of view most interest is focused on the operations of the PKS in phase 1, but the late steps are essential to produce active antibiotics.

Scheme 1

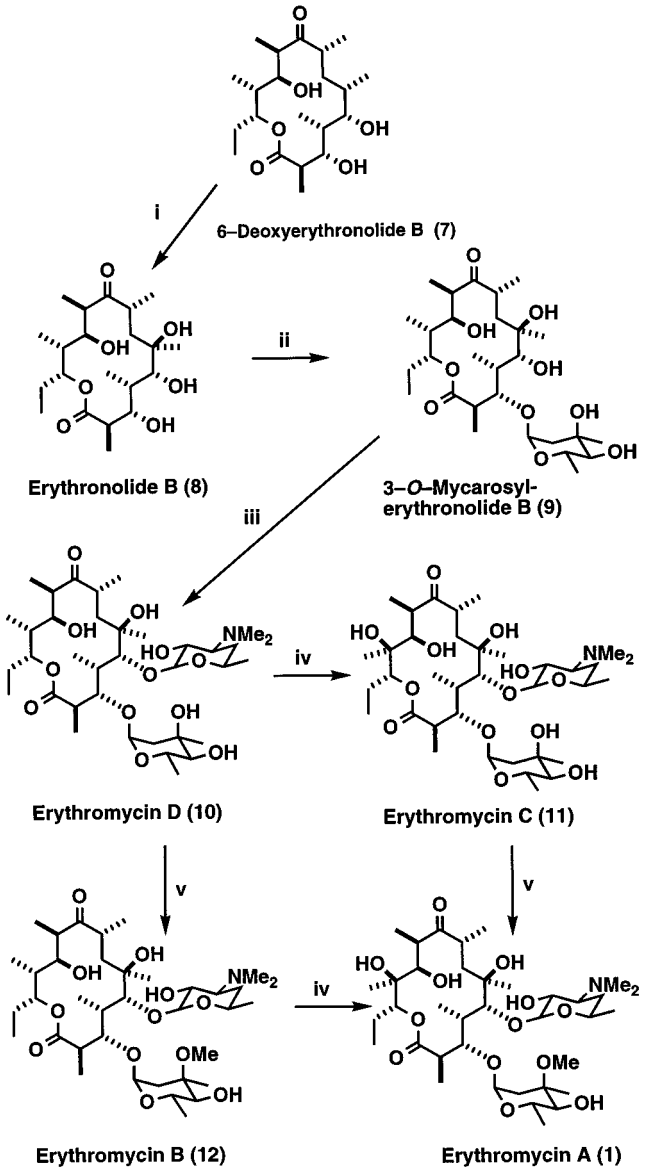


B. Post PKS Erythromycin Biosynthesis

Most investigations into the steps after 6-deoxyerythronolide B have been carried out using blocked mutants which are unable to produce one of the enzymes catalyzing the conversion of a late intermediate in the biosynthetic pathway.[24] This allows

accumulation of the intermediate in sufficient quantities for it to be isolated and identified. Mutants blocked in each step in the pathway from 6-deoxyerythronolide B (**7**) to erythromycin A (**1**) have been produced allowing the late intermediates to be identified (Scheme 2).

Scheme 2



i, C–6 erythronolide hydroxylase; ii, TDP–mycarose glycosyltransferase; iii, TDP–desosamine glycosyltransferase; iv, C–12 hydroxylase; v, (*O*)–methyltransferase

Erythronolide B (**8**) was the first intermediate to be isolated.[25] Its intermediacy was established by incorporation of radioactive **7**[26] and isotope dilution experiments.[27]

The C-6 hydroxylation of 6-deoxyerythronolide B (**7**) takes place with retention of configuration.[28] The enzyme responsible, eryF, is a cytochrome P450 hydroxylase which has been overproduced in *Escherichia coli* and for which an X-ray structure has been determined.[29] Several interesting features concerning this enzyme were investigated through specific mutants, and an unusual mechanism for substrate-assisted acid-catalyzed dioxygen bond cleavage during the catalytic cycle has been suggested.[30]

In the next step L-mycarose is attached to the C-3 hydroxyl group of **8** by a TDP-mycarose glycosyltransferase.[31,32] The amino sugar D-desosamine is then added to the C-5 hydroxyl of **9** by the action of the enzyme TDP-desosamine glycosltransferase.[32,33] The resulting intermediate, erythromycin D (**10**) is the first to show antibacterial activity and occurs at a branch in the biosynthetic pathway.

Either C-12 hydroxylation takes place with retention of configuration,[34] to produce erythromycin C (**11**), or *O*-methylation of the C-3″ hydroxyl of the mycarose moiety with SAM, catalyzed by an *O*-methyltransferase, produces erythromycin B (**12**). Finally erythromycin A (**1**) can be generated either by *O*-methylation of **11** catalyzed by an *O*-methyltransferase and SAM, or by C-12 hydroxylation of **12**.

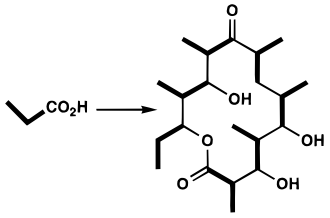
A single *O*-methyltransferase operates on both branches of the pathway.[35] Similarly a single cytochrome P450 hydroxylase has been implicated.[36] This enzyme, eryK, has been overproduced as inclusion bodies in *E. coli* and reconstituted to the active holo form. Since this has a 1200−1900-fold preference for erythromycin C over B, it is considered that the pathway via erythromycin B may be the minor one.

The biosynthetic pathways leading to the sugar residues are not well understood, although reasonable outline proposals have been put forward largely on the basis of sequence information.[37,38]

C. The Biosynthesis of 6-Deoxyerythronolide B

Studies using blocked mutants showed that no intermediates preceding 6-deoxyerythronolide B could be isolated, implying that these intermediates are enzyme bound.[24] Corcoran and co-workers fed labeled precursors to *S. erythraea* and detected their incorporation into 6-deoxyerythronolide B (**7**, Scheme 3).[39] These results gave the first evidence that the macrolide core of the erythromycins is built from seven C_3 units, now known to be derived from one unit of propionyl CoA and six of methylmalonyl CoA.

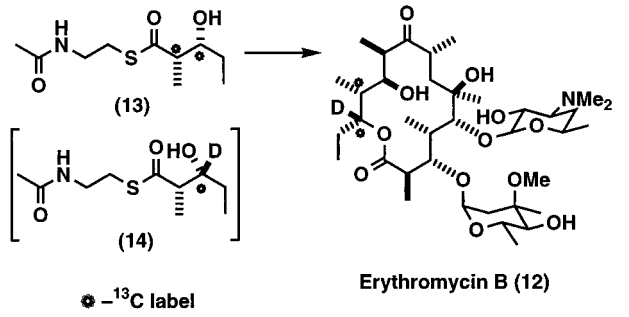
Scheme 3



In a further study Cane and co-workers[40] fed [^{18}O]propionate to *S. erythraea*. This revealed that all the oxygens attached to carboxyl-derived carbons in the macrolide aglycon core were retained from the propionate precursor and were not derived from molecular oxygen or water. These results are consistent with a processive mechanism in which each methylmalonyl condensation is followed by reduction to give the required oxidation state and stereochemistry before the addition of the next unit. Subsequently in an important pioneering experiment, Cane and Yang[41] investigated the proposed processive diketide intermediate. The free acid showed random incorporation into erythromycin B (**12**), indicating

that the diketide had been degraded to propionate before being used in the biosynthesis. However, the *N*-acetylcysteamine (NAC) thioester analogue **13** was incorporated intact with retention of the coupling between the two simultaneously labeled sites of ^{13}C enrichment (Scheme 4). The analogous deuterium

Scheme 4



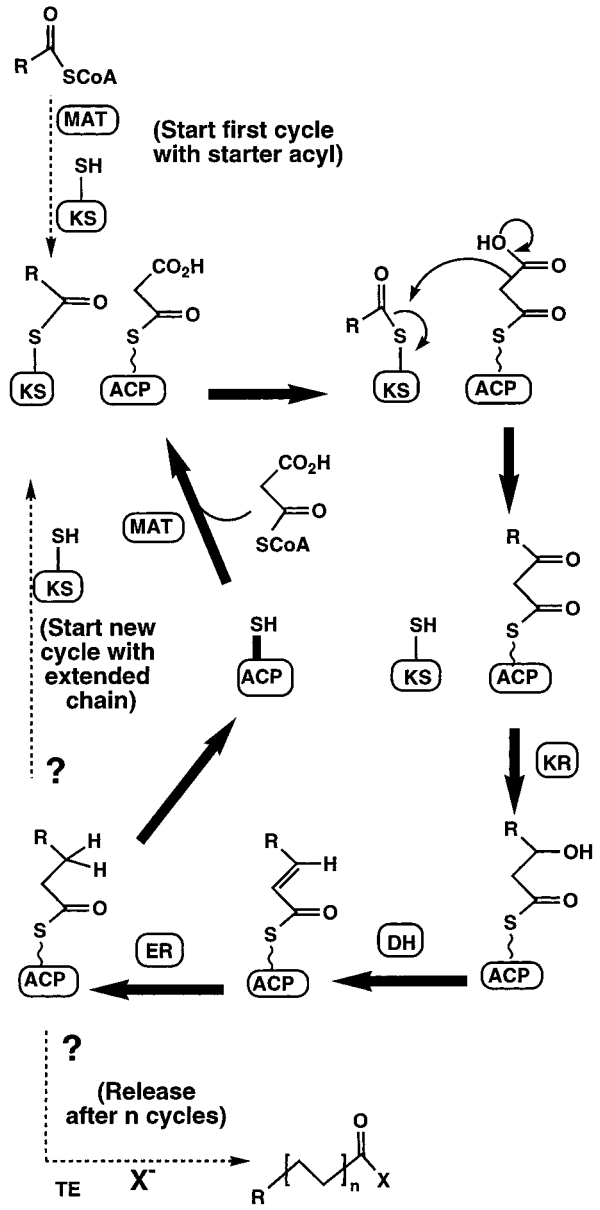
labeled diketide **14** was also incorporated intact demonstrating that **13** was not oxidized to the β-keto thioester before incorporation.[42] The *N*-acetylcysteamine (NAC) thioester was employed in these studies because it exhibits a high degree of structural homology with the thiol terminus of coenzyme A and of the 4′-phosphopantetheine group of the active acyl carrier protein. This experiment, with a similar contemporary pioneering contribution from Hutchinson and co-workers,[43] opened up a highly productive phase of research into the biosynthesis of other polyketide metabolites.

D. Structure and Function of the Erythromycin PKS

Up to 1990 many attempts were made to isolate active cell-free extracts of the PKS from the erythromycin-producing organism without success. It was generally recognized, however, that there is a strong parallel with the fatty acid biosynthetic pathway which has long been investigated at the enzyme level.[44,45] A fatty acid synthase (FAS) forms its fatty acid chain through a succession of reactions which is initiated by condensation of a starter unit (commonly acetate) with an extender unit (malonate) (Scheme 5). The resulting β-keto ester is then fully processed (reduced, dehydrated, and reduced again) to give an elongated saturated fatty chain, and the cycle resumes with the condensation of a new extender unit. To carry out these operations the FAS requires a set of catalytic activities, with each activity being responsible for one step of the cycle of reactions, *i.e.*, malonylacetyltransferase (MAT) to transfer both the acetyl starter unit and malonyl chain-extender units from the respective coenzyme A derivatives to the appropriate thiols of the synthase; β-ketoacyl synthase (KS) and acyl carrier protein (ACP) for chain elongation; β-ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) for reductive processing of the β-keto group; and a thioesterase (TE) for release of the full-length chain.

The constituent proteins of a FAS can be organized in markedly different ways depending on the organism. At one extreme of complexity, found in bacteria, the individual activities (domains) are freely dissociable and can be isolated separately.[46] At the other extreme, in animals or yeast, the domains are bound

Scheme 5



together by peptide links to form large nondissociable multidomain proteins.[44,45] The dissociable synthases are designated "type II" and the multidomain proteins are designated "type I". The same nomenclature has been adopted in the PKS field.

The sequence of domains from the amino to the carboxy terminus of the animal FAS (type I) is KS, MAT, DH, ER, KR, ACP, and TE (Figure 1). There is also a long stretch of amino acids sited between

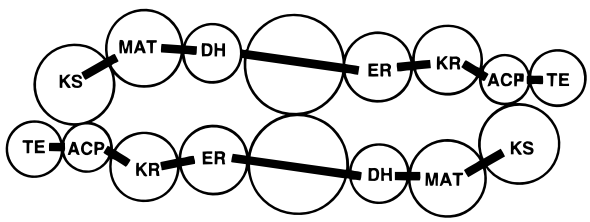


Figure 1. Linear "head to tail" model for the animal fatty acid synthase (FAS). Predicted active domains are represented by circles with activities coded as follows: ketoacyl synthase (KS); malonyl acetyl transferase (MAT); dehydratase (DH); enoyl reductase (ER); keto reductase (KR); acyl carrier protein (ACP); thioesterase (TE).

the DH and ER domains which appears to have no specific catalytic function and so is assigned a structural role.[44,45] In its spatial organization and structure this well-studied synthase shows strong similarities to the more recently discovered erythromycin PKS. It therefore merits special consideration as a possible model for that synthase and for PKS's more generally.

The animal FAS consists of a homodimer made up of two multidomain proteins. A structure has been proposed (Figure 1) on the basis of investigations by electron microscopy, chemical cross-linking, and proteolysis.[44,45] According to this model the core of the homodimeric structure is established by contacts between the long stretch of structural protein in one partner with the equivalent stretch of protein in the other. These regions are designated by unmarked spheres in the representation shown; the other domains are represented by spheres whose relative diameters are in proportion to their expected relative masses to give a sense of proportion to the model. The two chains run antiparallel in a linear "head to tail" arrangement forming two functionally independent catalytic clusters. This model leads to the prediction that each cluster relies on cooperation between sets of domains residing in *complementary* chains: KS+MAT+DH of one chain is predicted to cooperate with ER+KR+ACP+TE of the other. The general shape of the model resembles the theta shape seen in electron micrographs.[47,48]

This simple model is not easily reconciled, however, with the results of recent elegant complementation studies of FAS proteins containing defective domains (produced by genetic engineering).[49,50] These show that in each catalytic cluster the DH domain and the TE domain reside in the *same* chain not in complementary chains as implied by the model above. A fundamentally different topology for the animal FAS which satisfies this condition is proposed later.

E. Sequencing of the Genes Coding for the Erythromycin PKS

The first clue to the structure of the erythromycin PKS came with the sequencing of the corresponding genes. This immensely powerful strategy for probing biosynthetic pathways was pioneered in the polyketide field by Hopwood and coworkers who located and sequenced the genes responsible for the actinorhodin PKS.[51] This technology has since been adopted by most leading groups in the polyketide field. The genes coding for the erythromycin pathway were located in the vicinity of the gene coding for erythromycin resistance, *ermE*. On either side of *ermE* there were regions of DNA containing open reading frames encoding for non-PKS proteins responsible for the late stages of the biosynthesis, from 6-deoxyerythronolide B (**7**) to erythromycin A (**1**). This has been confirmed by targeted disruption of these genes to give mutants from which late intermediates have been isolated (see above). A map showing these regions of the genome is presented in Figure 2.

Sequencing further away from *ermE* revealed three large open reading frames (*eryAI*, *eryAII*, and *eryAIII*) each coding for a giant (>3000 amino acids) multifunctional protein.[52,53] These are named DEBS 1, DEBS 2, and DEBS 3, respectively.[52−56] In the linear representations of the primary sequence of the proteins shown in Figure 3, sections thought to be associated with specific catalytic activities are indicated by blocks with various markings. The assignments of these activities rested on sequence homology with the equivalent domains in fatty acid synthases. These catalytic centers are termed *domains*. Each domain is thought to be folded to form a local globular structure with a specific active site. The various regions between the domains which are unmarked in this diagram are thought to play a passive but vital structural role in maintaining the various domains in the correct topology for cooperation in the overall catalytic process. Many of these structural regions can be identified by short sections of amino acids rich in alanine, proline, and charged residues, up to 30 residues long, which are thought to serve as linkers between the folded domains or residues. There are also intriguing longer sections of primary sequence in all three proteins which cannot be associated with any specific catalytic activity.

In analyzing the synthetic implications of this organization of the domains it is helpful to consider a different representation of the PKS in which domains are represented by circles and the structural residues are ignored (Scheme 6). Taking all three proteins together the number of ketosynthase (KS) domains is six which allows for a separate KS for each chain extension cycle. Following each KS there is a set of additional domains appropriate for one chain-extension cycle: the sets include two essential domains, AT and ACP, plus an appropriate set of optional domains, KR, DH, ER, for modification of the new keto group. To aid the analysis, the optional domains are raised as loops above the line of the essential domains in this cartoon version of the structure. It can be seen that the activities are organized into six "*modules*" (two per protein), each of which is able to carry out a condensation step using the three essential domains followed by appropriate modification of the keto group in the newly formed keto ester. In module 3 there is a block of protein

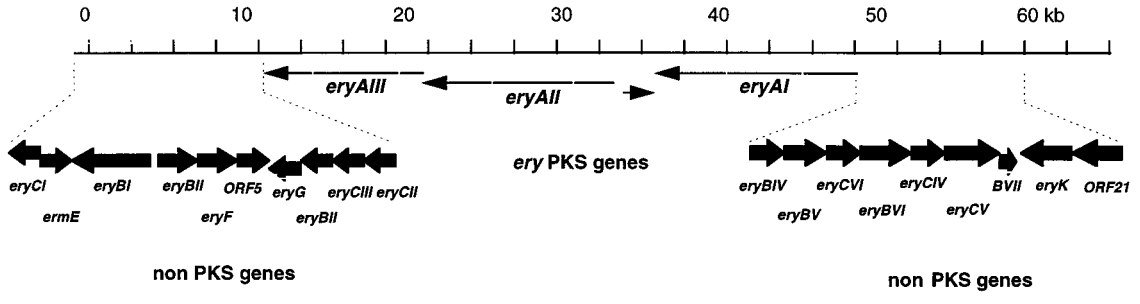


Figure 2. Map of regions of the *S. erythraea* genome containing genes associated with the late stages of the erythromycin biosynthetic pathway.

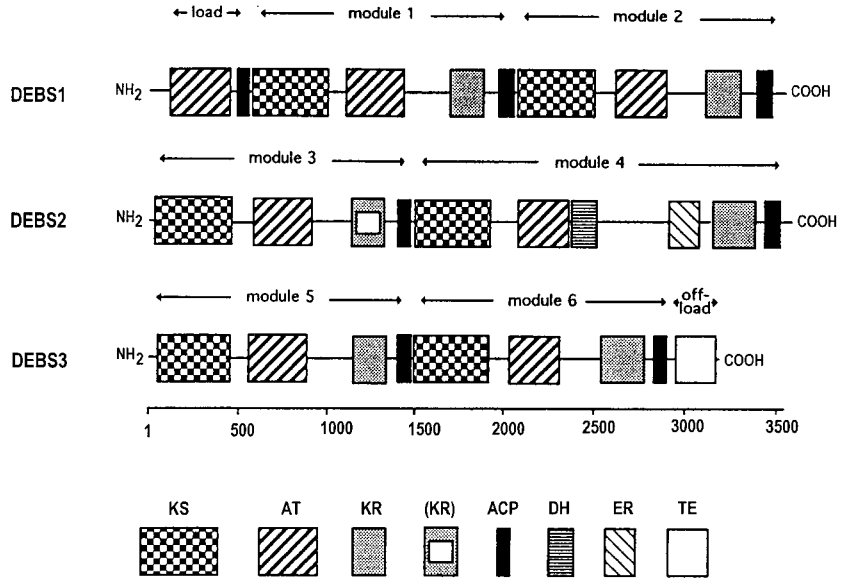
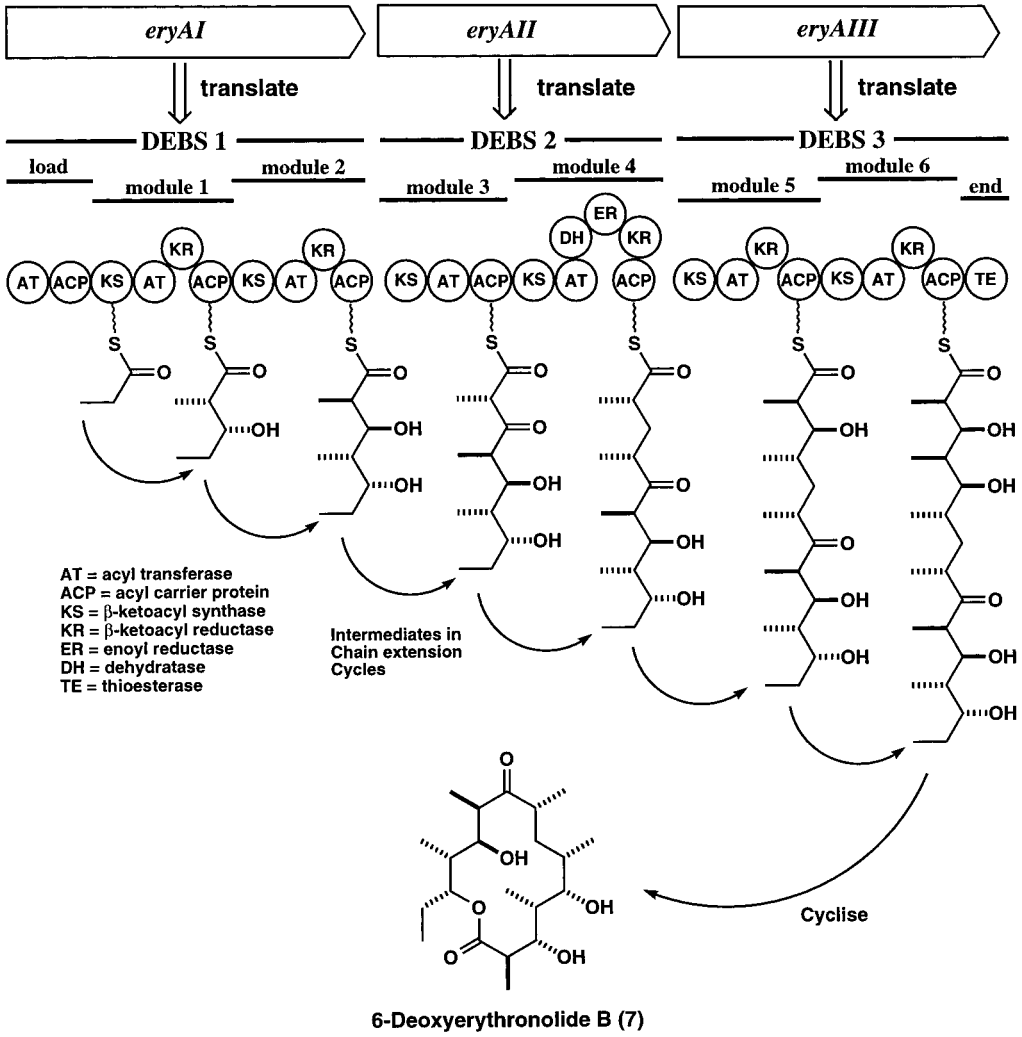


Figure 3. Predicted domain organization of the DEBS proteins. Each ketoacyl synthase (KS), malonyl acetyl transferase (MAT), dehydratase (DH), enoyl reductase (ER), keto reductase (KR), acyl carrier protein (ACP), and thioesterase (TE) domain is represented by a box with coded shading whose length is proportional to the size of the domain; (KR) indicates an inactive KR domain. The ruler indicates the residue number within the primary structure of the constituent proteins. Linker regions are shown in proportion.

Scheme 6



with sequence similarity to a KR at the expected position for a KR domain, but it appears to be nonfunctional and so is excluded from the analysis. The last module of DEBS 3 is followed by a thioesterase activity which is thought to catalyze lactonization of the polyketide chain and so release the macrolide from the enzyme. At the start of the first protein, DEBS 1, there is a loading module for the

starter acid (propionate) which consists of an AT linked to an ACP.

It is in module 4, which has the complete set of reductive domains required to form a methylene group, that the close similarity of domain sequence to the animal FAS is most apparent. The similarity is reinforced by comparing Figures 1 and 3, where it can be seen that in this module there is a large block of presumed structural protein at the same position (between the DH and ER domains) as a similar block of protein found in the animal FAS. In other modules of the erythromycin PKS, which lack the DH and ER domains, a similar block of protein structure lies between the AT and KR domains.

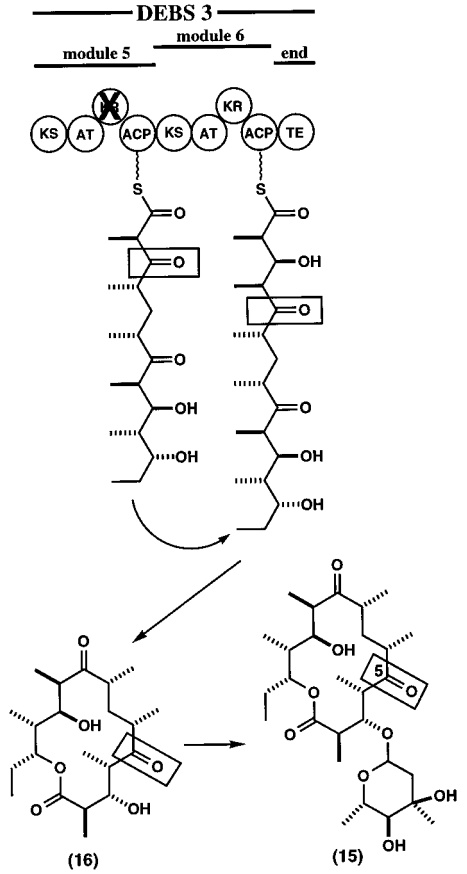
According to this modular analysis, each protein catalyzes two cycles of chain extension. The term "*cassette*" has been proposed for the giant proteins.[56] All three cassettes in the erythromycin cluster are bimodular, but in other systems, such as the rapamycin[57] and tylosin[58] PKS's, the size of cassette can vary from one to six chain extension modules. The three cassettes cooperate in some way to form an extraordinarily complex molecular assembly line. The biosynthetic intermediates remain PKS-bound throughout the whole synthetic sequence via *thioester* links. A challenging feature of this organization is the mechanism which controls the ordering of the cassettes in the assembly line so that transfer of the growing chain from one cassette to the next is correctly controlled.

The arrangement of the PKS genes and the ordering of the catalytic sites in relation to the postulated growing chain can also be seen in Scheme 6. There is a persuasive correlation between the structure of the newly added residue in the growing chain and the nature of the optional domains present in the module associated with its assembly. However, the protein could fold in many different ways to bring the complementary activities together to form functional modules, so it was a high priority to establish that individual domains are indeed used in the order shown.

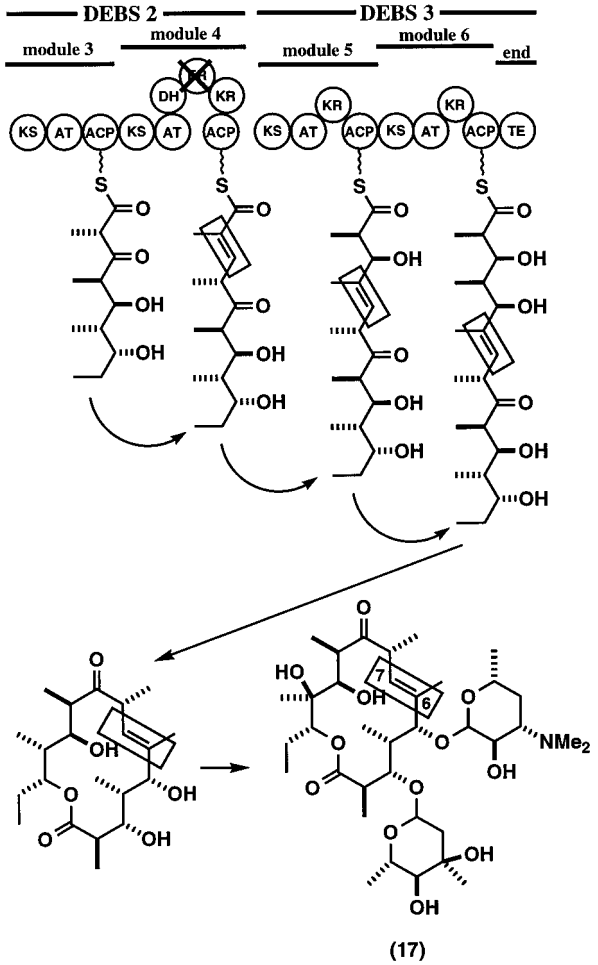
The first strong supporting evidence came from Katz and coworkers who carried out pioneering gene disruption experiments. They disrupted the β-ketoacyl reductase (KR) in module five and from the resulting mutant they isolated a partly processed erythromycin analogue **15**, in which a keto group has survived at the predicted position of the macrolide in place of the normal hydroxyl group.[59] The presumed product of the PKS is therefore the 5-keto analogue **16** of 6-deoxyerythronolide B (Scheme 7). More recently the enoyl reductase (ER) in module four was disrupted and an analogue **17** of erythromycin was isolated containing a double bond at C_6,C_7 (Scheme 8).[60] It can be concluded that the disrupted keto reductase operates in the fifth cycle of chain extension and that the disrupted enoyl reductase operates in the fourth cycle.

The formation of these two analogues and the fact that they are correctly lactonized shows that the structure of the nascent chain may not play a critical role in polyketide biosynthesis, and at least some altered polyketides can be substrates for further cycles of chain extension. This implies that the correct transfer of the growing chain from one synthase unit to the next may reside more in the specific juxtaposition of the various domains rather than in

Scheme 7



Scheme 8



the conventional substrate specificity of a particular synthase domain for the structure of the approaching chain. It is significant and encouraging for novel metabolite production that the product released from the PKS in both experiments was at least partly processed toward an analogue of erythromycin A, showing that the elaboration enzymes also exhibit a useful degree of relaxation in their substrate specificity.

F. Isolation and Structural Studies of the Erythromycin PKS

All three DEBS proteins have been overexpressed and successfully purified to homogeneity.[61] Gel filtration indicated that all were dimeric under native conditions,[61] and a DEBS multienzyme has since been confirmed to be exclusively homodimeric under the conditions of analytical ultracentrifugation.[62]

The first information concerning the structure of these homodimers came from limited proteolysis studies.[62−64] For each proteolytic enzyme and each DEBS protein the cleavage pattern was highly specific and reproducible to give a set of fragments containing one or more intact domains. The cleavages occurred between the domain boundaries predicted on the basis of sequence alignments, but many predicted inter domain linker regions were not cleaved, even under harsh conditions. The results are summarized for DEBS 1 in Figure 4.

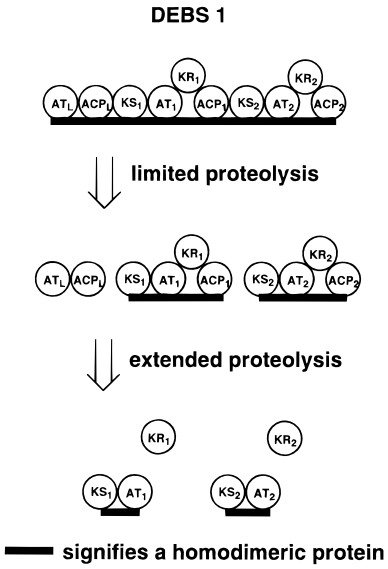


Figure 4. Pattern of fragments generated by controlled proteolysis of DEBS1.

The most rapid cleavages under conditions of limited proteolysis were always found to occur between the two chain extension modules housed in each multienzyme, irrespective of which cleavage enzyme was used to degrade the proteins.[62−64] Other initial cleavages split the N-terminal loading bidomain from DEBS 1 and a bidomain consisting of the C-terminal thioesterase from DEBS 3, together with its attached ACP domain.[62,63] All of the modular fragments have subunit molecular masses in agreement with those predicted from their amino acid sequences. With the exception of the loading bidomain, they all behave as *homodimers* on gel filtration. Each of the modular fragments gives a single N-terminal amino acid sequence, which confirms the identity of the fragment and also demonstrates that in a dimeric multienzyme there must be an element of symmetry that leads to identical cuts in both subunits.[62,63]

More extensive proteolysis of DEBS 1, DEBS 2, and DEBS 3 gave smaller fragments.[62−64] Most significantly, the six KS domains were released as bidomains with their attached AT domains and all of these were *homodimeric*. In contrast, fragments consisting of single domains derived from the side-loops associated with keto group modification were *monomeric*. The KR domains of modules 1, 3, 5, and 6 and the single ER domain from module 4 were all characterized as monomers.

Proteolysis of DEBS dimers in which the thiols at the active sites of cooperating ACP and KS residues had been cross-linked by prior treatment with 1,3-dibromopropanone gave a different pattern of fragments consistent with the formation of cross-links *between* two complementary protein chains.[62] It was concluded that the KS of one chain cooperates with the ACP residing in the complementary chain rather than in its own, as had been demonstrated earlier for the animal FAS.

A dimeric "double helical" structure for the DEBS proteins has been proposed which is consistent with all these results.[62] It is illustrated for DEBS 3 in Figure 5. To arrive at the double helix the two polypeptide chains of the DEBS are first folded to give domains and then lined up side by side in a parallel "head to head", "tail to tail" fashion as shown. The two chains are then twisted together to form a double helix with the KS, ACP, and AT domains of each module forming the core of the structure. The helix is shown in ribbon form to aid clarity in the positioning the domains. The sense of the helical twist is arbitrary at this stage. The chosen angle of twist in each module (180°) brings the ACP of one chain below the KS domain of the identical module in the complementary chain allowing the required interchain cooperation in the condensation step. To simplify the presentation of the helix, the ribbon diagram shows an arbitrary extra twist (60°) of the linker region between the ACP of the upper module and the KS of the next; the orientation of one module relative to the next could be varied depending on the constraints of the linker region between the ACP and KS domains. This is readily cleaved so it is likely to be relatively flexible as well as being open to attack by proteases. The optional domains in each module

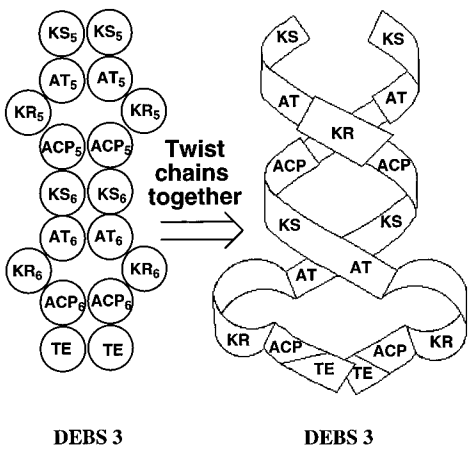


Figure 5. Proposed folding of DEBS 3 to give the "double helical" model.

needed for ketone group modification (DH, ER, and KR) form loops that protrude out sideways from the central core, while maintaining the active sites of all reductive domains within range of the phosphopantetheine group on the adjacent ACP in the *same* chain. There is therefore no contact between reductive domains of complementary chains which is consistent with the monomeric nature of these domains in the isolated state (see Figure 4). In contrast the KS (and TE domain of module 6) are in intimate contact in the core of the helical structure and so proteolytic fragments containing them would be likely to retain the homodimeric character of the intact protein. Three such double helices stacked on top of the other give rise to the complete erythromycin PKS.

Subsequently a complementation study of the erythromycin PKS was reported[65] which was modeled on the techniques developed by Smith in his studies of the animal FAS system.[49] In this work DEBS1+TE dimers were made containing first a dead version of the KS in module 1 of one chain and a dead version of the KS in module 2 of the complementary chain, and secondly another heterodimeric system containing a dead version of the KS of module 2 in one chain and a dead version of the ACP of module 2 in the complementary chain. The resulting proteins retained the ability to make the product triketide lactone, albeit at a reduced rate. These very elegant and highly significant experiments established a number of key facts about the structure and operation of the PKS. First, they confirmed the earlier results of cross-linking which showed that in a given module the KS of the chain operates with the ACP of the complementary chain.[62] Secondly, they showed that in going from one module to the next the polyketide residue is passed from the ACP to the KS residing in the *same* chain. Therefore, the growing polyketide is transferred both within the peptide chains and between the chains at different stages of its passage down the synthase.

The authors went on to suggest an alternative model for the erythromycin PKS based on the model for the linear head to tail model of the animal FAS. The domains in each module are arranged as a homodimer with the complementary chains running *anti*parallel side by side. The resulting flat (theta-shaped) modules are then stacked one above the next as in a stack of pancakes with linker regions running vertically between linked modules at the outside of the stack Figure 6. This linear "head to tail" model satisfactorily accommodates the data of the complementation study, as does the double helical model.

The crucial difference between this new model and the helical model can best be seen by considering module 4 of the erythromycin PKS. As noted earlier this most resembles the animal FAS with its full complement of reductive domains (but the analysis applies to all modules of the PKS). Both models are theta shaped when viewed from above. In the linear head to tail model the two KS domains are placed far apart and therefore could not possibly be in contact. In the double helical model the KS groups are placed in contact at the core with the sideways loops accommodating the reductive domains. As explained earlier, the helical model also brings the two key thioesterases in contact at the terminus of module 6, whereas the FAS-inspired model places

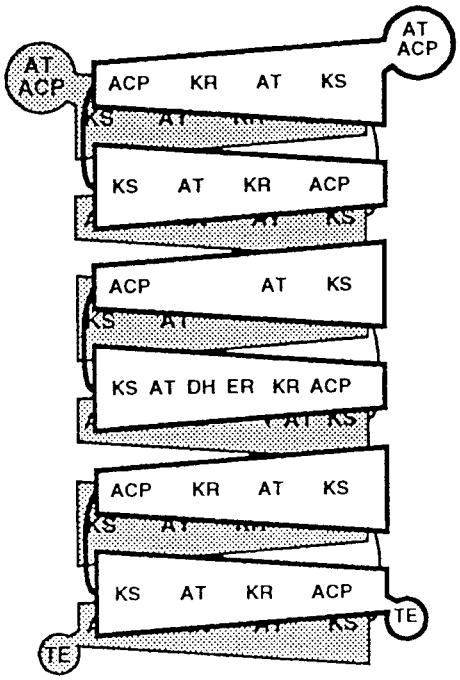


Figure 6. Proposed model for DEBS based on the proposed linear "head to tail" model for the animal FAS.

them far apart. The homodimeric behavior of key proteolysis fragments is consistent with the helical model, but is not easily reconciled with the alternative linear "head to tail" model.[62,63]

On the basis of the results of their complementation studies with the animal FAS Smith and coworkers have suggested that the homodimer might have an unspecified folded arrangement.[50] Interestingly, a suitable folded topology for the FAS dimer can be advanced on the basis of the double helical structure proposed for module 4 of the erythromycin PKS (Figure 7). The top diagram is intended to show that even in a two-dimensional model it is not necessary to invoke a "head-to-tail" arrangement to bring the KS and ACP domains of complementary chains close together. As before in Figure 1 each domain in the model is represented by a sphere with the diameter proportional to the relative expected molecular mass of the particular domain. In this model the two KS domains are placed in contact ("head-to-head") and the required contact each with the opposite ACP is achieved by folding the chains to form a spiral. The key difference between this arrangement and the simpler linear arrangement is that the core of the structure is formed by association of the KS domains, not the region of structural protein (blank sphere). In this two-dimensional model the TE is sited alongside the ACP in the same plane.

This simple idea is developed in the two lower diagrams which give two views of a three-dimensional model based on the "double helical" structure for the erythromycin PKS. The ACP and TE domains are slid below the KS domain of the partner chain, giving a helical twist to both chains (as before the sense of the helix is arbitrarily chosen). This new "double helical" model fits the *theta* shape observed for the FAS by electron microscopy.[47,48] It is also consistent with all the reliable data for the FAS structure including the unexpected result (for the linear "head to tail" model) of Smith's recent complementation studies in which the DH domain is pre-

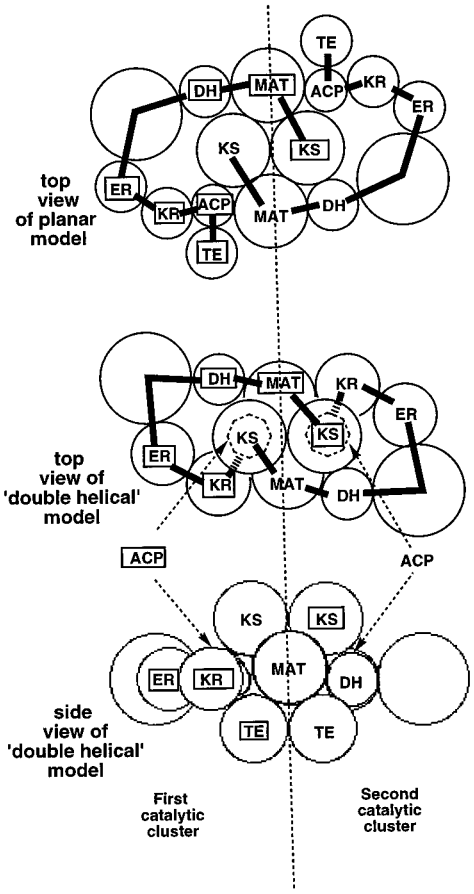


Figure 7. Proposed planar and "double helical" models for the animal FAS involving "head to head" association of the monomeric units.

dicted to cooperate with the catalytic cluster containing the ACP and TE of its own chain.[50] Finally, with possible future complementation experiments in mind, it should be noted that this model is compatible with any form of association for the ACP with the AT domains: in the planar variant shown the ACP of one chain partners the AT domain of the other, but in a helical variant with the ACP sited below the KS it could in principle partner either or both AT domains.

There is no direct evidence in proteolysis studies for association of the KS domains in the animal FAS, unlike the erythromycin PKS. However, there are so many close parallels between the two types of synthase, it is likely that they are folded in very similar ways. It is as reasonable to look to the PKS structure for guidance to the FAS structure as it is to argue in the reverse direction. At present the "double helical" model is the only one which is compatible with all the published data relating to both the FAS and PKS systems. It will be interesting to see if it stands the test of further complementation studies. The predicted composition of the two catalytic clusters is indicated in Figure 7.

G. Genetic Engineering of the Erythromycin PKS

The erythromycin PKS has been genetically engineered in many different ways which convincingly demonstrate the elegance and power of the genetic engineers' art to produce targeted mutations. Progress in the field was greatly facilitated by the development by Hopwood and Khosla of the expression plasmid pRM5, which has allowed engineered constructs to be expressed in *Streptomyces coelicolor* which normally does not produce erythromycin.[66] Most work to date on genetic engineering has relied upon this highly effective tool. Strikingly, it is possible to transfer the whole of the PKS to *S. coelicolor*.[67] The unelaborated PKS product, 6-deoxyerythronolide B (**7**) was produced in this foreign context. This mutant has also proven to be a more efficient vehicle for incorporation experiments with added precursors than the natural host, *S. erythraea*.[68]

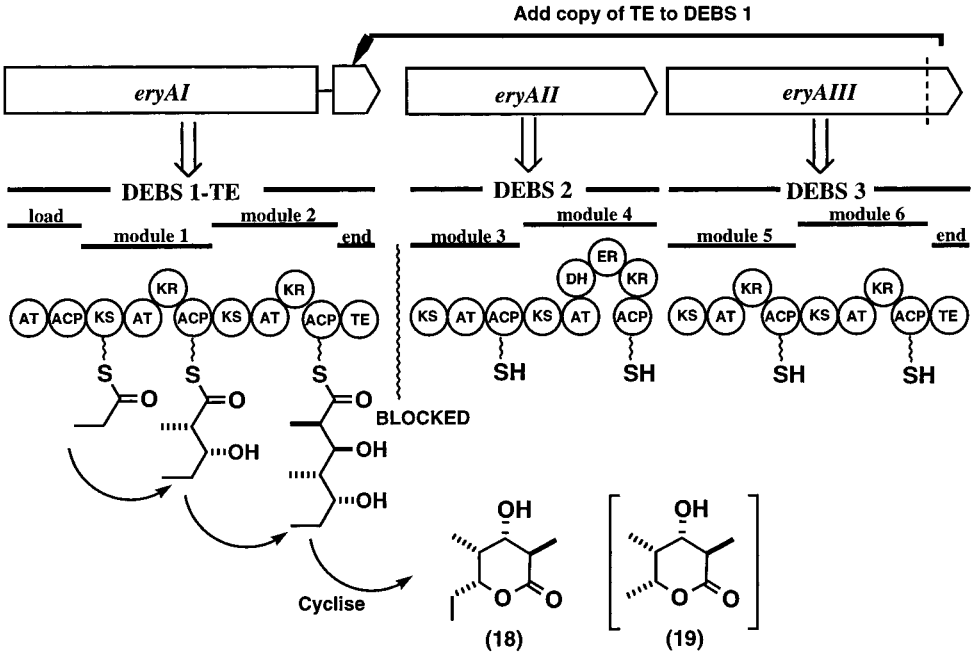
The important pioneering studies by the Katz group have already been discussed in support of the proposed modular structure of the PKS. A common characteristic of these experiments was the essentially destructive character of the mutations. The activity of individual domains could be destroyed without destroying the overall capacity of the PKS to make macrolide structures. This *destructive* strategy could therefore be used to generate novel products, but clearly the range of possible variations is relatively limited.

A potentially more versatile strategy is to *reposition* domains within the PKS, but this would only work if the repositioned domain could carry out its normal type of reaction in its new context on a foreign substrate. The first demonstration of the feasibility of this approach came with the relocation of the thioesterase domain from to the terminus of DEBS 1.[69] The aim was first to prevent further chain extension by blocking the docking of the C-terminus of DEBS 1 onto the N-terminal end of DEBS 2 (this was the predicted consequence of this change according to the "double helical" model for the PKS structure). Second, it was hoped that the TE domain would play an *active* role in the engineered protein by catalysing the release of the triketide chain as a δ-lactone.

A mutant of *S. erythraea* containing the engineered protein (called DEBS 1-TE) (Scheme 9) did give the anticipated δ-lactone **18**. As expected erythromycin production was shut down completely. A vital control experiment was then carried out to establish the second point—that the relocated domain was acting as an *active* agent of chain release rather than just a passive block to further chain extension. A further mutant was made in which a genetically engineered *dead* copy of the TE was placed at the end of DEBS 1. The inactivated TE differed from the normal enzyme only in the replacement of the key serine residue at the active site by an alanine; this conservative change should not alter the folding of the protein significantly. This control mutant did not produce erythromycin. It did produce the δ-lactone **18** but in much reduced yield, demonstrating that the TE was playing an active role in chain release in the first experiment.

A contemporary experiment was reported which mirrored the design of DEBS 1−TE in that it involved placing the thioesterase at the end of DEBS 1. The construct was engineered by cutting and pasting at a different place in the primary sequence of the proteins, however, to give a mutant protein with a significantly different primary structure, so it is appropriate that it was given a slightly different name: DEBS 1+TE.[70] The triketide lactone was formed but no control experiment with a *dead* version of the TE was carried out. Instead an experiment in which DEBS1 alone was expressed in *S. coelicolor*

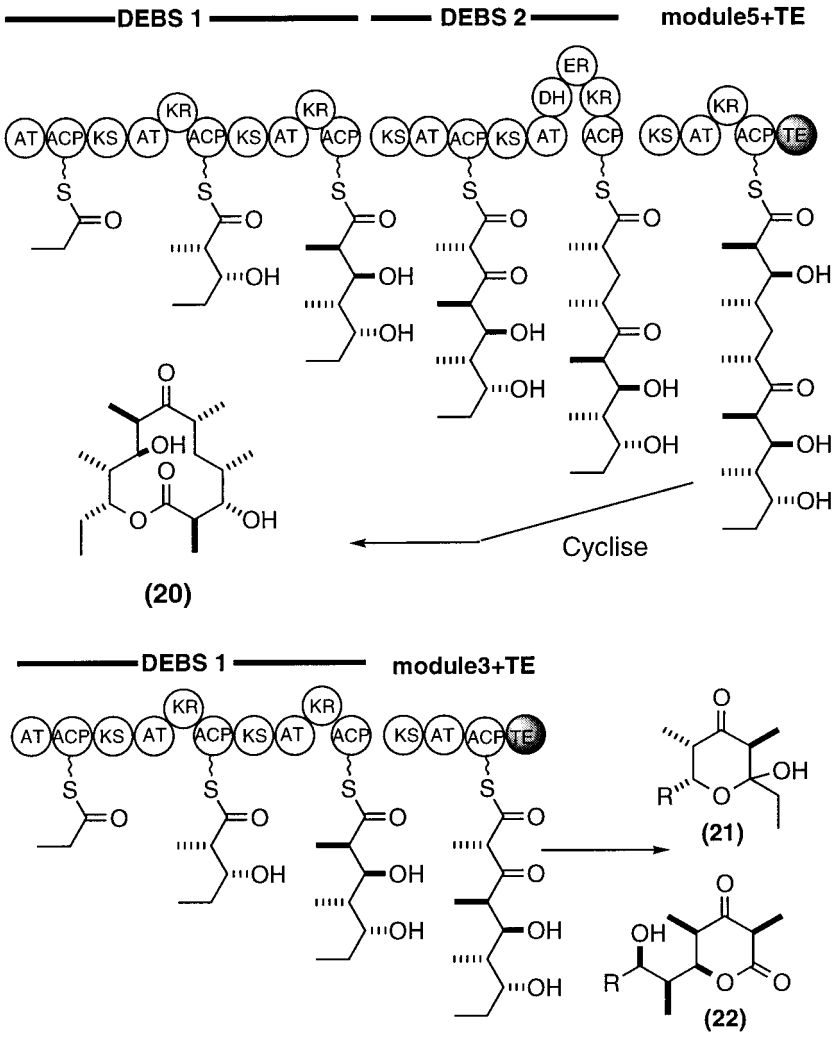
Scheme 9



was used as the basis for comparison.[71] Interestingly, the two closely similar constructs gave significantly different results when expressed in *S. coelicolor*: DEBS 1−TE, but not DEBS 1+TE, gave the δ-lactone analogue **19** with an acetate starter group, as well as **18**.[72]

In further applications of this relocation strategy (Scheme 10) the TE has been relocated to the termini of both module 5[70] and module 3.[73] Products consistent with truncation of chain extension at the expected stages were obtained in both cases. The truncation at module 5 was especially significant

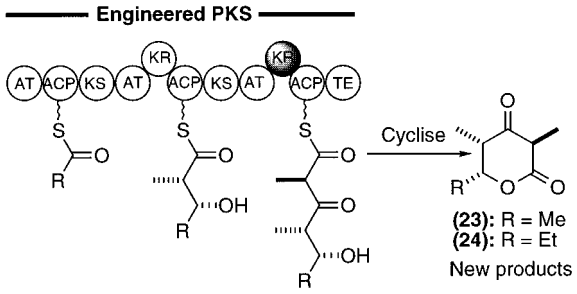
Scheme 10



because it lead to release of the hexaketide intermediate as a 12-membered ring macrolide **20**. The truncation after module 3 caused the formation of two tetraketide products, **21** and **22**; the first is possibly formed from a keto acid by decarboxylation after its release from the enzyme.

Recently, a different domain, the putative ketoreductase from module 3 (which is nonfunctional in its normal context) was relocated in place of KR_2 in module 2 of DEBS 1+TE (Scheme 11).[74] Novel

Scheme 11

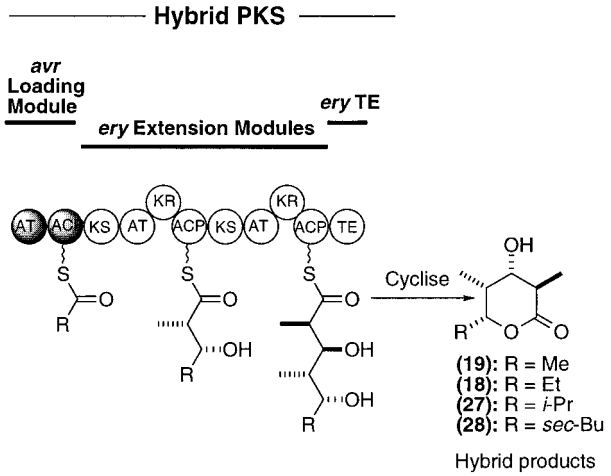


δ-lactone products **23** and **24** were observed in which the keto group generated by the condensation step in module 2 survived. In a sense this experiment offers an alternative strategy for destroying the activity of ketoreductase sites, although a question remains—would an intrinsically active KR also have been nonfunctional in the new construct?

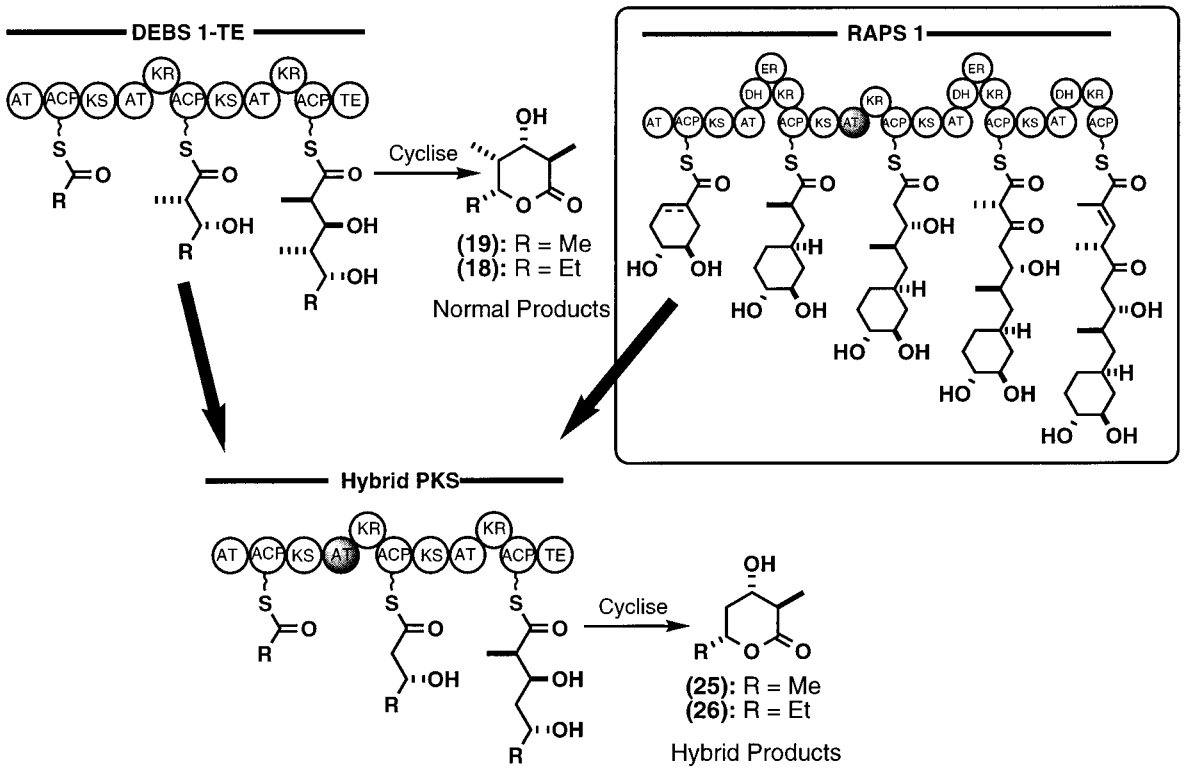
The potential range of novel products would be enormously expanded if domains could be transferred between different PKS clusters. This would amount in effect to "spare part surgery" or "mix and match" swapping of structural residues between different natural products. In the first experiment to realize this goal sections of two different *type I* PKS systems (rapamycin and erythromycin) were hybridized to produce a *hybrid type I* PKS (Scheme 12).[75] This produced the expected hybrid product containing elements of structure derived from the two parent natural products. The experimental model was DEBS1-TE and the experiment involved replacing the AT of module 1 by an AT derived from the rapamycin PKS (this will be described later). In its normal context in rapamycin biosynthesis the transplanted AT specifies a malonate unit as chain extender rather than methylmalonate. δ-Lactones **25** and **26** were produced lacking a methyl group at C-4, but otherwise the normal erythromycin pattern of structure was maintained at C-2, C-3, and C-5.

A more ambitious heterologous hybrid has also been described which involved transplanting a complete module rather than just a domain. The loading module (AT plus ACP) of the avermectin PKS was transplanted into DEBS1-TE in place of the normal erythromycin loading module (Scheme 13).[76] The

Scheme 13



Scheme 12



products of the fermentation of this mutant strain included two novel lactones **27** and **28** as well as the normal products **18** and **19**. The novel products have starter acyl groups characteristic of the avermectins and therefore can reasonably be viewed as hybrid molecules incorporating elements of the avermectin and erythromycin structures in positions dictated by the genetic engineering. A related example of heterologous hybridization was reported recently in which the loading module of the platenolide PKS was replaced by that of the closely related tylactone PKS to give a predicted novel natural product.[58]

The successful generation of these productive, truly hybrid, PKS's opens up exciting new avenues for genetic engineering of type I PKS systems. They show that domains or modules drawn from different synthases can cooperate to give a working synthase and therefore that different synthases are probably closely compatible in structure. Furthermore they point the way to more ambitious and versatile genetic engineering experiments leading to the production of a wide variety of novel structures including novel antibiotics, immunosuppressants, and other types of active compound.

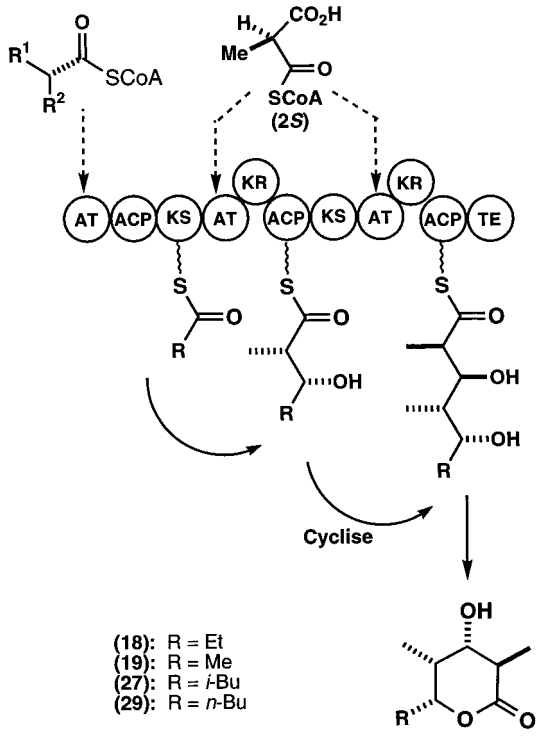
H. Catalytic Activity of the Isolated PKS as a Whole and in Part

In some of the early structural studies of the isolated DEBS proteins it was demonstrated that specific domains were catalytically active. For example cleavage of DEBS 1 with trypsin gave the N-terminal bidomain together with two other fragments, comprising modules 1 and 2 respectively. It was found that the N-terminal bidomain was specifically radiolabeled with [^{14}C]propionyl-CoA, after incubation, providing the first evidence for its proposed role as the "loading module" for the propionate starter unit, whereas the other two fragments were specifically acylated with [^{14}C]methylmalonyl-CoA, indicating that the other two acyltransferases were enzymatically active after proteolysis.[63,64] In other studies with the intact DEBS proteins it was demonstrated that all the AT domains of the chain extension modules specifically catalyzed the hydrolysis of (*S*)-methylmalonate but were inert to the *R* isomer. This was the first clue to the remarkable fact that the *S* isomer of methylmalonate serves as building block for both chain extension modules of DEBS 1 (see later).[64]

In 1995 the construct DEBS 1-TE was obtained in a high (95%) state of purity.[77] When incubated with the appropriate building blocks, propionyl CoA and methyl malonyl CoA, and a reducing agent, *specifically* NADPH, the protein catalyzed the formation of the δ-lactone triketide product **18** (Scheme 14). As the protein was substantially pure, it was established that the PKS is self sufficient and does not require assistance from any external catalytic activities. A further important discovery was that the loading domain had a more relaxed degree of substrate specificity than was apparent *in vivo*. Butyryl- and isobutyryl-CoA esters (**29** and **27**, respectively) were acceptable as starter acyl groups *in vitro* as well as the established acetyl and propionyl analogues (Scheme 14).

With completely purified protein it was also possible to make the rigorous determination of the

Scheme 14



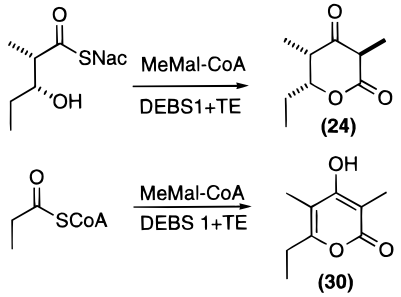
stereochemistry of the methylmalonate units used for chain extension in modules 1 and 2. The two modules of DEBS 1 are especially interesting because the two branching methyl groups in the lactone product have opposite stereochemistries. In principle it was possible that either the *R* isomer of the methylmalonate was incorporated at one site and the *S* isomer at the other by a common incorporation mechanism, or that a single enantiomer was incorporated at both sites with a different sequence of subsequent operations at one of the two sites to alter the methyl stereochemistry. It was unambiguously shown that the *S* isomer served for both chain extensions. It is unlikely, therefore that altered products with *different stereochemistry* will be generated by transferring AT domains and some other means will have to be devised to achieve this desirable end.

In contemporary studies DEBS 1+TE was isolated in a fully pure state.[78,79] A partially purified protein fraction taken from an early step of the purification procedure was also used for *in vitro* studies of triketide lactone production. The loading domain showed the same relaxed substrate specificity reported for DEBS 1−TE. In parallel studies a cell-free preparation containing all three DEBS proteins, an enzyme complex consisting of 28 distinct catalytic sites, was shown to support the synthesis of the macrolide lactone 6-deoxyerythronolide B (**7**).[78,79] These extracts were used to extend some of the earlier incorporation experiments with synthetic analogues of PKS-bound intermediates derivatized as *N*-acetylcysteamine thioesters.[78−80] The specificity of incorporation of the diketide could therefore be established with much higher confidence than was possible with the equivalent *in vivo* experiments.

New insights came from incubations of DEBS 1+TE in the absence of the reducing agent NADPH: of particular interest is the isolation in different

experiments of keto and pyrone analogues of the triketide lactone (**24** and **30**, Scheme 15).[78,80] This

Scheme 15



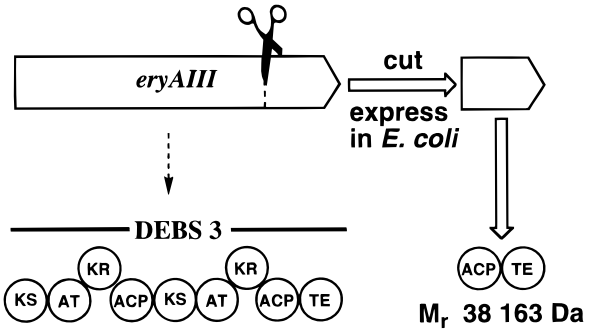
later result resembles the formation of a pyrone product by the eukaryotic FAS and the 6-methylsalicylic acid PKS under similar conditions.[81] Clearly, the β-keto ester intermediates can be transferred from one module to the next in the DEBS without further processing, even in modules which are geared to the production of a hydroxy ester intermediate. Much remains to be discovered before we understand how the PKS controls the extent of keto group modification in its synthetic operations.

Two kinetic studies with DEBS 1+TE have been reported, the first on partially purified protein[82] and the second on fully purified material.[83] The measured k_{cat} value was higher in the later experiments. It should be recognized that with these extraordinarily large enzymes, the available tests for homogeneity do not detect protein molecules which have suffered minor, but fatal, damage to one of the constituent domains in the extraction process. Such damage would render the individual molecular assembly with a consequent lowering of measured kinetic parameters.

A potentially very important advance for *in vitro* studies came with the isolation of a protein consisting of module 3+TE.[83] In collaboration with an extract of DEBS 1 this produced the tetraketide products **21** and **22** *in vitro* (see Scheme 10). It may be possible to use this system to catch the two cassettes in contact as the developing chain was passed from one to the other by treatment with appropriate cross-linking reagents. Unfortunately, attempts so far have not met with success, so the nature of cassette docking in multicassette PKS systems remains a mystery.

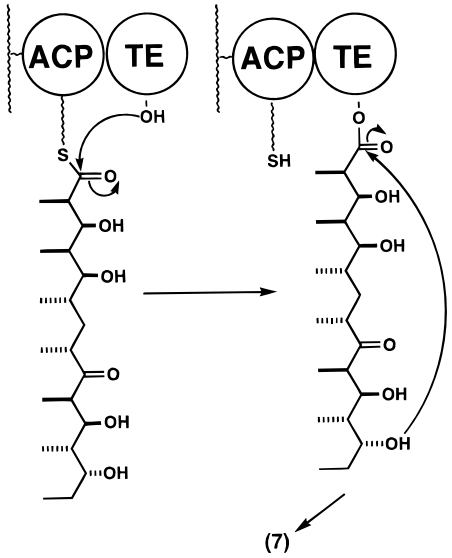
The substrate specificity of the thioesterase is of crucial concern in the quest for a rational basis for genetic engineering of PKS's to produce novel products: it would be futile to engineer a PKS system to produce a particular novel product if there is no effective mechanism for its release from the final ACP. In order to investigate this potential limitation, the bidomain from the end of DEBS 3 has been overexpressed in *E. coli* to give milligram amounts of protein (Scheme 16).[84] The resulting protein was shown by electrospray mass spectrometry to have the correct molecular weight for the bidomain with an *apo*-ACP in which the phosphopantetheine group had not been added. This deficiency did not matter as far as the planned studies of the TE specificity were concerned. In the preliminary work it was shown that the protein was able to bind *p*-methylphenylsulfonyl fluoride (PMSF), a standard inhibitor of

Scheme 16



chymotrypsin-like enzymes, to the hydroxyl of a specific serine residue located in the putative active site of the enzyme. This evidence strongly supported a proposed cleavage mechanism involving initial transfer of the developed acyl chain from the thiol of the neighbouring ACP to the hydroxyl group of an active serine residue on the TE (Scheme 17). The

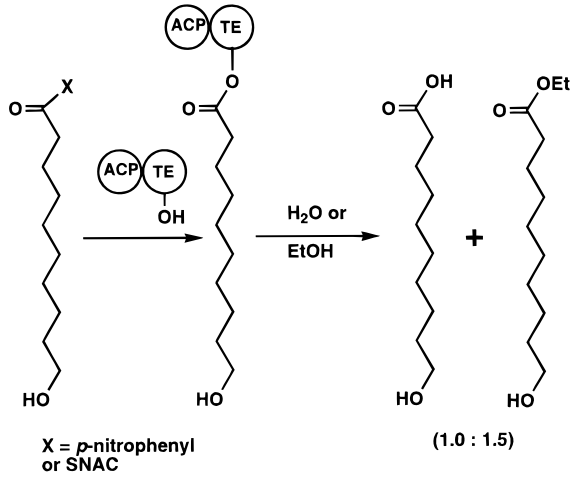
Scheme 17



resulting oxygen ester is then cleaved by attack of a suitable nucleophile to release the product. In forming a macrolide ring (the presumed normal role of the thioesterase), the nucleophile would be the appropriate hydroxyl group at the distant terminus of the polyketide chain.

In subsequent work the thioesterase was challenged with synthetic acyl esters derivatized as *p*-nitrophenyl and *N*-acetylcysteamine esters (exemplified in Scheme 18).[85] A wide range of structurally varied substrates with both leaving groups was cleaved. By electrospray mass spectrometry it was shown that acyl enzyme intermediates were formed as predicted. The resulting acyl enzyme intermediates could be cleaved by either water to give the corresponding acid, or by an added alcohol such as ethanol to give an ester. None of the substrates tested gave any macrocyclic lactone even when there was a hydroxyl group elsewhere in the chain at a suitable distance from the acyl terminus. The work established that the TE has alternative mechanisms for product release (hydrolysis or ester formation) when lactonization is not possible. Information from studies such as these can guide genetic engineering

Scheme 18



studies toward productive goals and away from experiments doomed to failure because of incompatibility of the TE domain with the target novel polyketide product.

I. Summary

In the seven years following the announcement of the sequence of the erythromycin PKS there have been extraordinarily rapid advances which have firmly established this system as the archetypal modular polyketide synthase. It provides the structural and mechanistic case against which other synthases may be compared both *in vitro* and *in vivo*. No less important are the genetic engineering advances which have been developed for modification of this synthase. It has to be said, however, that there are still enormous gaps in our understanding of the structure and function of this system, so the day of completely "rational" genetic engineering remains a distant prospect.

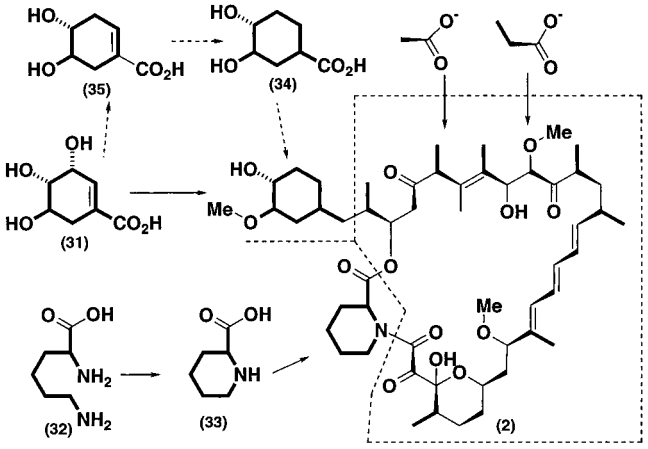
III. Rapamycin Biosynthesis

A. Origin of the Carbon Skeleton

Examination of the structure of rapamycin **2** leads to the conclusion that it is biosynthesized by a mainly polyketide pathway from seven acetate and seven propionate units, with *O*-methyl groups from *S*-adenosyl methionine (Scheme 19). The substituted cyclohexane ring could be derived by a reductive pathway from shikimic acid (**31**). The pipecolate unit would probably be derived from lysine (**32**) via pipecolic acid (**33**).

Initial precursor incorporation studies using ^{13}C-labeled acetate and propionate confirmed that the predicted part of the macrolide ring is of polyketide origin.[86] The three *O*-methyl groups originate from methionine. Later competitive incorporation studies with radiolabeled precursors proved that the pipecolate ring is derived from lysine via free pipecolic acid: radioactive lysine incorporation was reduced by the addition of unlabeled pipecolate to a greater extent than the reduction in radioactive pipecolate incorporation by unlabelled lysine.[87] Feeding of ^{13}C-labeled shikimic acid confirmed that the cyclohexane ring is derived intact from shikimic acid.[88] Recently it was shown that the saturated dihydroxy cyclohexane carboxylic acid **35** could serve as precursor of the cyclohexyl unit.[89]

Scheme 19



Related studies have filled in some details of the pathway from shikimic acid to the cyclohexane ring of FK520.[90] These have shown that shikimic acid undergoes a 1,4-conjugate elimination (dehydration) followed by reduction, isomerization, and reduction again to the saturated cyclohexane ring **34**.

Several rapamycin analogues have been isolated from alternative organisms[91] or different strains of *S. hygroscopicus*.[92,93] Further rapamycin analogues have been isolated using a high producing strain *Actinoplanes* sp. N902-109 to which alternative precursors and P-450 inhibitors were fed.[94]

B. The Rapamycin Biosynthetic Gene Cluster

The entire biosynthetic gene cluster for rapamycin has been sequenced.[57] The PKS genes were identified by hybridization with DNA from the PKS genes for erythromycin biosynthesis. Sequencing beyond the PKS region identified other genes that are predicted to be involved in rapamycin biosynthesis.[95] The organization of the gene cluster is shown in Figure 8.[96] Gene disruption experiments have confirmed that these genes are necessary for rapamycin biosynthesis.[97,98]

As might have been expected, the PKS for rapamycin shows a type I organization strongly reminiscent of the erythromycin PKS, with catalytic activities arranged in modules (Scheme 20) and with sets of modules housed in turn in three multimodular cassettes designated RAPS 1, RAPS 2, and RAPS 3. RAPS 1 contains modules 1−4, RAPS 2 modules 5−10, and RAPS 3 modules 11−14.[57] Previously, the largest PKS cassette identified contained only three modules so all the cassettes in the rapamycin PKS break new ground in terms of size.

Surprisingly, the putative domain structure of the rapamycin PKS may not correspond in every detail to the pattern expected for the proposed structure for the PKS product. In modules 3 and 6, there appear to be potentially active KR and DH domains which are not required; module 3 also contains a potentially active but functionally redundant ER domain. It is possible that the active sites of these extra domains have been inactivated in a way that is not apparent from the primary sequence, and that the now redundant protein residues have still to be edited out by the random processes of evolution. There is also a

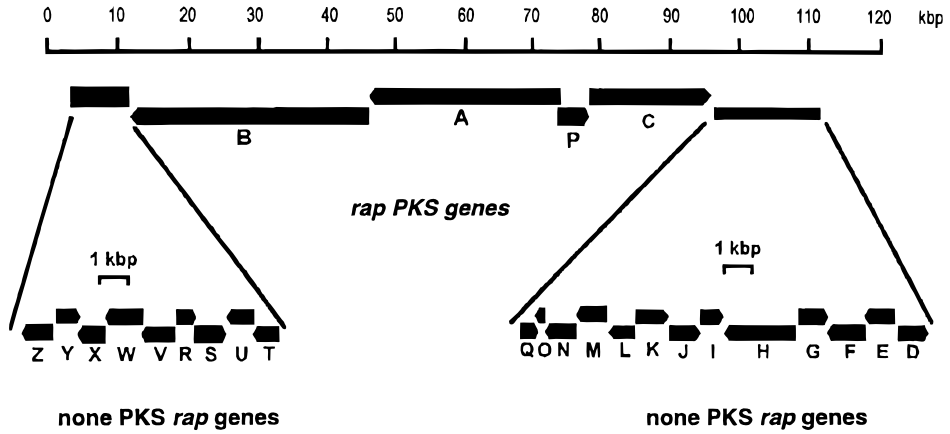
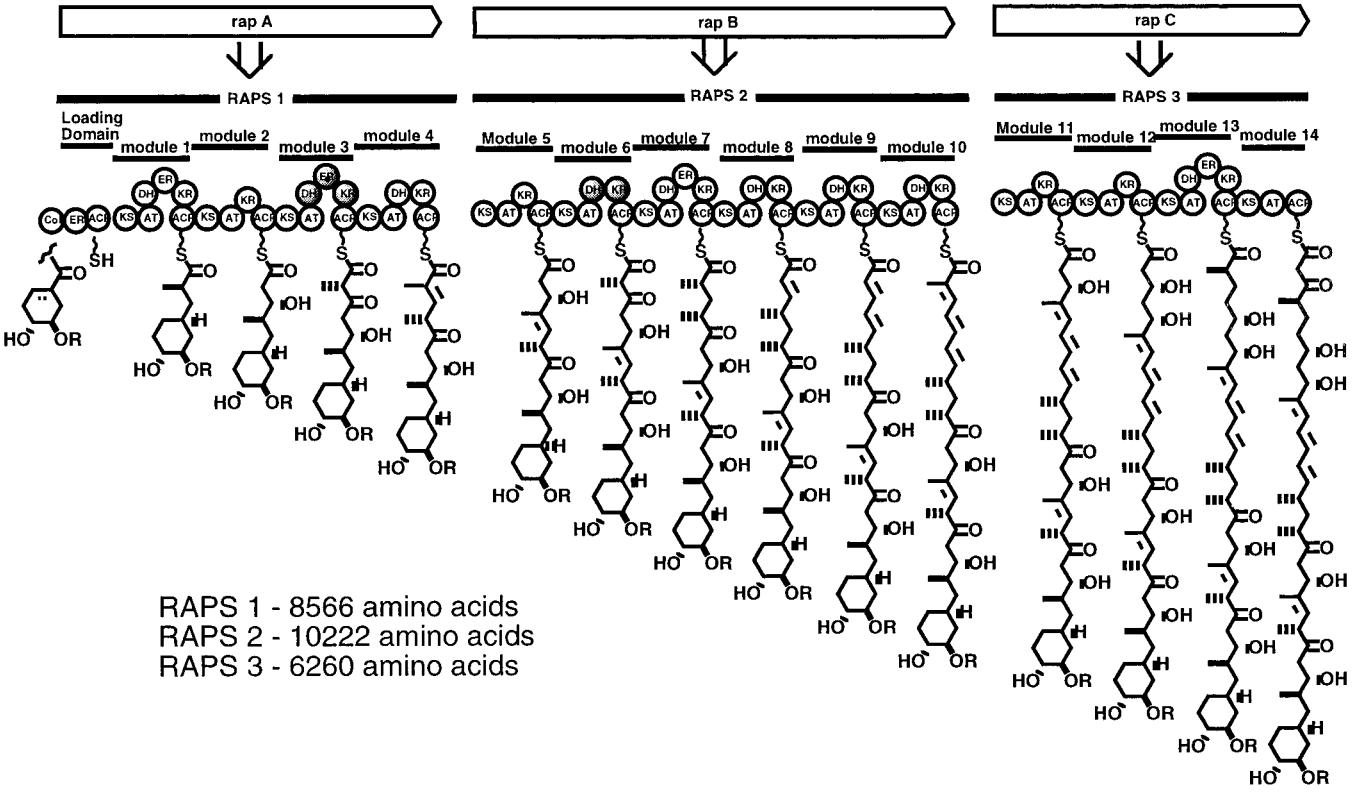


Figure 8. Map of the genes of the rapamycin biosynthetic cluster.

Scheme 20



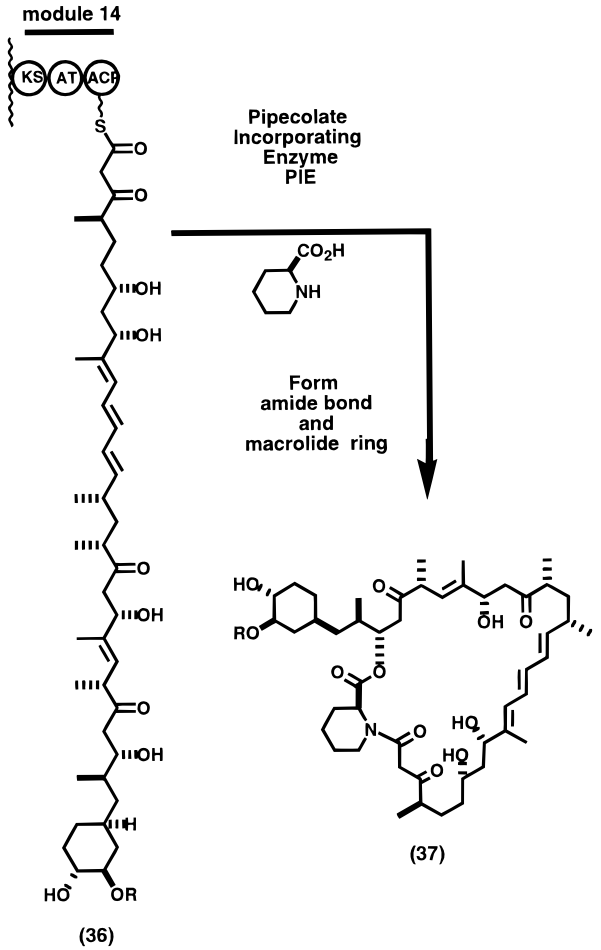
chance, possibly remote, that all these domains are indeed active and that the true rapamycin PKS product is more fully reduced than that shown. Extra post PKS reoxidations would then be required to reintroduce the oxygen functionality at the relevant sites in the final structure.

The loading module seems to require three domains. The first (CL) shows homology to ATP-dependent carboxylic acid-CoA ligases, the second is a putative enoyl reductase (ER), and the third, an ACP. The probable sequence of operations starts with the enoic acid **35** derived from shikimic acid (see Scheme 19). The ER domain will saturate the double bond of the cyclohexene ring. The first domain will activate a carboxylic acid to an active acyl derivative ready for transfer to the thiol residue of the ACP. The final saturated product will end up attached to the ACP as thioester derivative ready for transfer to the KS domain of the first chain extension module. The timing of the reduction in this sequence of operations cannot be predicted.

Chain termination is not carried out by a thioesterase as in the erythromycin PKS, but is thought to be effected by a specialized multidomain protein coded by the gene *rapP*.[95] This gene has strong similarity to genes involved in nonribosomal peptide biosynthesis and the corresponding protein is believed to catalyze the formation of the ester and amide bonds to pipecolic acid (Scheme 21).[99] The final product **36** of the PKS is coupled to the nitrogen of the pipecolate unit, and the carboxyl group of this unit is esterified with the first free hydroxyl at the remote end of the polyketide chain to give **37**. The order of these steps is not known. It is this key multidomain protein, designated the pipecolate-incorporating enzyme (PIE), that provides the pivotal link between polyketide and polypeptide biosynthesis. Isolation and characterization studies have recently demonstrated that this enzyme is monomeric.[98]

Most of the other non-PKS genes can be assigned to the late stages of rapamycin biosynthesis.[95] At

Scheme 21



least two oxygenation reactions are required at C-9 and C-27 of rapamycin. The *rapJ* and *rapN* genes have similarity to cytochrome P450 monooxygenases, and *rapO* has similarity to ferredoxin genes (required for the catalytic activity of P450s). There are three *O*-methyl groups in rapamycin and three genes, *rapI*, *rapM*, and *rapQ*, which are similar to SAM-dependent methyltransferase genes. Other genes can be assigned potential functions such as transcriptional regulation and export.

Detailed sequence analysis has revealed that the sequences for AT domains in the rapamycin PKS can be categorized into two groups depending on whether the predicted substrate is malonyl-CoA or methylmalonyl-CoA.[100] This analysis can be extended to include the AT domains of the erythromycin, oleandomycin, and avermectin PKSs. The AT domains of the recently published sequence for the FK 506 PKS also follow a similar pattern.[101]

C. Enzymology of Rapamycin, FK506, and FK520 Biosynthesis

Even without the benefit of genetic data, there has been some success in characterizing the enzymes of FK506 and FK520 biosynthesis. Enzymes designated 31-*O*-demethyl methyltransferases have been isolated and characterized from the producing organisms of both FK506 and FK520.[102] The FK520 enzyme has been used in the enzymatic synthesis of partially methylated FK520 analogues.[103]

An enzyme has been isolated from the FK520 producer which is believed to the key one responsible for inserting pipecolic acid into the macrocycle.[104] It is reported to be dimeric and activates pipecolic acid and several structural analogues in an ATP-dependent reaction to give an enzyme-bound amino-acyl adenylate. There is evidence that this then reacts to form a thioester linkage to the enzyme. This mechanism of activation is the same as that found in the nonribosomal biosynthesis of peptide natural products such as gramicidin.[99]

D. Summary

Although the study of rapamycin is in its infancy compared with erythromycin, there is abundant evidence to support the contention that the PKS stages of the biosyntheses are similar. This is apparent from the homology in the gene sequences and also in the pattern of domains and modules in the PKS. The evidence for a close similarity is reinforced by the demonstration that an AT domain can be transferred from one cluster to the other with retention of its synthetic capacity. This experiment points the way to future success in "mix and match" experiments with these remarkable synthetic machines to produce hybrid products with polyketide residues of mixed origin. The exciting new dimension opened up by the report of the rapamycin genes is the possibility of hybridizing the *polypeptide* portion of rapamycin with structures culled from the polypeptide natural products. The pipecolate incorporating enzymes show strong homology to those of the peptide synthases. As things stand this extension of structural "mix and match" may be considered no more than a remote possibility but it is one that deserves serious attention. The extra potential for structural diversity is enormous.

IV. Acknowledgments

We are indebted to many members of the Leadlay Staunton group in Cambridge, and especially to Dr. Peter Leadlay, for many helpful comments and suggestions concerning the contents of this review, and also for contributing to many of the ideas advanced in the discussion.

V. References

(1) *Macrolide Antibiotics: Chemistry, Biology and Practice*; Omura, S., Ed.; Academic: New York, 1984.
(2) O'Hagan, D. *The Polyketide Metabolites*; Ellis Horwood: Chichester, 1991.
(3) Celmer, W. D. *J. Am. Chem. Soc.* **1965**, *87*, 1801.
(4) Kleinkauf, H.; Von Döhren, H. *J. Antibiot.* **1995**, *48*, 563.
(5) McGuire, J. M.; Bunch, R. L.; Anderson, R. C.; Boaz, H. E.; Flynn, E. H.; Powell, M.; Smith, J. W. *Antibiot. Chemother.* **1952**, *2*, 281.
(6) Wiley, P. F.; Gerzon, K.; Flynn, E. H.; Sigal, M. V., Jr.; Weaver, O.; Quarck, U. C.; Chauvette, R. R.; Monahan, R. *J. Am. Chem. Soc.* **1957**, *79*, 6062.
(7) Harris, D. R.; McGeachin, S. G.; Mills, H. H. *Tetrahedron Lett.* **1965**, *6*, 679.
(8) Woodward, R. B.; Logusch, E.; Nambiar, K. P.; Sakan, K.; Ward, D. E.; Au-Yeung, B.-W.; Balaram, P.; Browne, L. J.; Gard, P. J.; Chen, C. H.; Chenevert, R. B.; Fliri, A.; Frobel, K.; Gais, H.-J.; Garrat, D. G.; Hayakawa, K.; Heggie, W.; Hesson, D. P.; Hoppe, D.; Hoppe, I.; Hyatt, J. A.; Ikeda, D.; Jacobi, P. A.; Kim, K. S.; Kobuke, Y.; Kojima, K.; Krowicki, K.; Lee, V. J.; Leutert, T.; Malchenko, S.; Martins, J.; Matthews, R. S.; Ong, B. S.; Press, J. B.; Rajan Babu, T. V.; Rosseau, G.; Sauter, H. M.; Suzuki, M.; Tatsuta, K.; Tolbert, L. M.; Truesdale, E. A.; Uchida, I.; Udea, Y.; Uyehara, T.; Vasella, A. T.; Vladuchick, W. C.; Wade, P. A.; Williams, R. M.; Wong, H. N.-C. *J. Am. Chem. Soc.* **1981**, *103*, 3210, 3213; 3215.

(9) Sample syntheses of erythromycin aglycon analogues areas follows: (a) Evans, D. A.; Kim, A. S. *Tetrahedron Lett.* **1997**, *38*, 53. (b) Sviridov, A. F.; Borodkin, V. S.; Ermolenko, M. S.; Yashunsky, D. V.; Kochetkov, N. K. *Tetrahedron*, **1991**, *47*, 2291; 2317. (c) Kinoshita, M.; Arai, M.; Tomooka, K.; Nakata, M. *Tetrahedron Lett.* **1986**, *27*, 1811; 1815. (d) Sviridov, A. F.; Ermolenko. M. S.; Yashunsky, D. V.; Borodkin, V. S.; Kochetkov, N. K. *Tetrahedron Lett.* **1987**, *28*, 3835.
(10) Vézina, C.; Kudelski, A.; Sehgal, S. N. *J. Antibiot.* **1975**, *28*, 721.
(11) Sehgal, S. N.; Baker, H.; Vézina, C. *J. Antibiot.* **1975**, *28*, 727.
(12) Martel, R. R.; Klicius, J.; Galet, S. *Can. J. Physiol. Pharmacol.* **1977**, *55*, 48.
(13) (a) Rosen, M. K.; Schreiber, S. L. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 384. (b) Schreiber, S. L.; Crabtree, G. R. *Immunol. Today* **1992**, *13*, 136. (c) Kunz, J.; Hall, M. N. *Trends Biochem. Sci.* **1993**, *18*, 334. (d) Dumont, F. J.; Su, Q. *Life Sci.* **1995**, *58*, 373.
(14) Swindells, D. C. N.; White, P. S.; Findlay, J. A. *Can. J. Chem.* **1978**, *56*, 2491.
(15) Findlay, J. A.; Radics, L. *Can. J. Chem.* **1980**, *58*, 579.
(16) McAlpine, J. B.; Swanson, S. J.; Jackson, M.; Whittern, D. N. *J. Antibiot.* **1991**, *44*, 688.
(17) Kessler, H.; Haessner, R.; Schüler, W. *Helv. Chim. Acta* **1993**, *76*, 117.
(18) (a) Nicolaou, K. C.; Chakraborty, T. K.; Piscopio, A. D.; Minowa, N.; Bertinato, P. *J. Am. Chem. Soc.* **1993**, *115*, 4419. (b) Nicolaou, K. C.; Piscopio, A. D.; Bertinato, P.; Chakraborty, T. K.; Minowa, N.; Koide, K. *Chem. Eur. J.* **1995**, *1*, 318.
(19) Romo, D.; Meyer, S. D.; Johnson, D. D.; Schreiber, S. L. *J. Am. Chem. Soc.* **1993**, *115*, 7906.
(20) Hayward, C. M.; Yohannes, D. Danishefsky, S. J. *J. Am. Chem. Soc.* **1993**, *115*, 9345.
(21) Smith, A. B., III; Condon, S. M.; McCauley, J. A.; Leazer, J. L.; Leahly, J. W.; Maleczka, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 947 & 962.
(22) Martin, J. R.; Rosenbrook, W. *Biochemistry* **1967**, *6*, 435.
(23) Corcoran, J. W.; Vygantas, A. M. *Biochemistry* **1982**, *21*, 263.
(24) Weber, J. M.; Wierman, C. K.; Hutchinson, C. R. *J. Bacteriol.* **1985**, *164*, 425.
(25) Tandrew, P. L.; Nyman, M. A. U.S. Patent 3,127,315, 1964.
(26) Vygantas, A. M.; Corcoran, J. W. *Fed. Proc.* **1974**, *33*, 1233.
(27) Hung, P. P.; Marks, C. L.; Tardrew, P. L. *J. Biol. Chem.* **1965**, *240*, 1322.
(28) Corcoran, J. W. In *Antibiotics, Volume IV: Biosynthesis*; Corcoran, J. W., Ed.; Springer-Verlag: New York, 1981; p132.
(29) Cupp-Vickery, J. R.; Poulos, T. L. *Nature Struct. Biol.* **1995**, *2*, 144.
(30) Cupp-Vickery, J. R.; Han, O.; Hutchinson, C. R.; Poulos, T. L. *Nature Struct. Biol.* **1996**, *3*, 632.
(31) Martin, J. R.; Perun, T. J.; Girolami, R. L. *Biochemistry* **1966**, *5*, 2852.
(32) Weber, J. M.; Leung, J. O.; Maine, G. T.; Potenz, R. H. B.; Paulus, T. J.; DeWitt, J. P. *J. Bacteriol.* **1990**, *172*, 2372.
(33) Majer, J.; Martin, J. R.; Egan, R. S.; Corcoran, J. W. *J. Am. Chem. Soc.* **1977**, *99*, 1620.
(34) Corcoran, J. W.; Vygantas, A. M. *Fed. Proc.* **1977**, *36*, 663.
(35) Corcoran, J. W. In *Methods in Enzymology XLIII*; Hash, J. H., Ed.; Academic Press: London, 1975; Chapter 33, p 487.
(36) Lambalot, R. H.; Cane, D. E.; Aparicio, J. J.; Katz, L. *Biochemistry* **1995**, *34*, 1858.
(37) Katz, L.; Donadio, S. In *Genetics and Biochemistry of Antibiotic Production*; Vining, L. C., Stoddard, C., Eds.; Butterworth-Heinmann: Boston, 1995; Chapter 14, p 223.
(38) Liu, H.-W.; Thorson, J. S. *Annu. Rev. Microbiol.* **1994**, *48*, 223.
(39) Kaneda, T.; Butte, J. C.; Taubman, S. B.; Corcoran, J. W. *J. Biol. Chem.* **1962**, *237*, 322.
(40) Cane, D. E.; Hasler, H.; Liang, T. *J. Am. Chem. Soc.* **1981**, *103*, 5960.
(41) Cane, D. E.; Yang, C.-C. J. *J. Am. Chem. Soc.* **1987**, *109*, 1255.
(42) Cane, D. E.; Prabhakaran, P. C.; Tan, W.; Ott, W. R. *Tetrahedron Lett.* **1991**, *32*, 5457.
(43) Yue, S.; Duncan, J. S.; Yamamoto, Y.; Hutchinson, C. R. *J. Am. Chem. Soc.* **1987**, *109*, 1253.
(44) Wakil, S. J. *Biochemistry* **1989**, *28*, 4523.
(45) Smith, S. *FASEB* **1994**, *8*, 1248.
(46) Fulco, A. J. *Prog. Lipid. Res.* **1983**, *22*, 133.
(47) Stoops, J. K.; Wakil, S. J.; Uberbacher, E. C; Bunick, G. J. *J. Biol. Chem.* **1987**, *262*, 10246.
(48) Kitamoto, T.; Nishigai, M.; Sasaki, T.; Ikai, A. *J. Mol. Biol.* **1988**, *203*, 183.
(49) Witkowski, A.; Joshi, A.; Smith, S. *Biochemistry* **1996**, *35*, 10569.
(50) Joshi, A. K.; Witkowski, A.; Smith, S. *Biochemistry* **1997**, *36*, 2316.
(51) Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet.* **1990**, *24*, 37.
(52) Cortés, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. *Nature* **1990**, *348*, 176.
(53) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675.
(54) Donadio, S.; Katz, L. *Gene* **1992**, *111*, 51.
(55) Bevitt, D. J.; Cortes, J.; Haydock, S. F.; Leadlay, P. F. *Eur. J. Biochem.* **1992**, *204*, 39.
(56) Staunton, J. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1302.
(57) Schwecke, T.; Aparicio, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B.; Böhm, G. A.; Staunton, J.; Leadlay. P. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839.
(58) Kuhstoss, S.; Huber, M.; Turner, J. R.; Pashal, J. W.; Rao, R. N. *Gene* **1996**, *183*, 231.
(59) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675.
(60) Donadio, S.; McAlpine, J. B.; Sheldon, P. L.; Jackson, M.; Katz, L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7119.
(61) Caffrey, P.; Bevitt, D. J.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1992**, *304*, 225.
(62) Staunton, J.; Caffrey, P.; Aparicio, J. F.; Roberts, G. A.; Bethell, S. S.; Leadlay, P. F. *Nature Struct. Biol.* **1996**, *3*, 188.
(63) Aparicio, J. F.; Caffrey, P.; Marsden, A. F. A.; Staunton, J.; Leadlay, P. F. *J. Biol. Chem.* **1994**, *269*, 8524.
(64) Marsden, A. F. A.; Caffrey, P.; Aparicio, J. F.; Loughran, M. S.; Staunton, J.; Leadlay, P. F. *Science* **1994**, *263*, 378.
(65) Kao, C. M.; Pieper, R.; Cane, D. E.; Khosla, C. *Biochemistry* **1996**, *35*, 12363.
(66) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546.
(67) Kao, C. M.; Katz, L.; Khosla, C. *Science* **1994**, *265*, 509.
(68) Cane, D. E.; Luo, G.; Khosla, C.; Kao, C. M.; Katz, L. *J. Antibiot.* **1995**, *48*, 647.
(69) Cortés, J.; Wiesmann, K. E. H.; Roberts, G. A.; Brown, M. J. B.; Staunton, J.; Leadlay, P. F. *Science* **1995**, *268*, 1487.
(70) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 9105.
(71) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 11612.
(72) Brown, M. J. B.; Cortés, J.; Cutter, A. L.; Leadlay, P. F.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995**, 1517.
(73) Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1996**, *118*, 9184.
(74) Bedford, D.; Jacobsen, J. R.; Luo, G.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1996**, *3*, 827.
(75) Oliynyk, M.; Brown, M. J. B.; Cortés, J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1996**, *3*, 833.
(76) Leadlay, P. F.; Staunton, J.; Marsden, A. F. A.; Wilkinson, B.; Dunster, N. J.; Cortés, J.; Oliynyk, M.; Hanefeld, U.; Brown, M. J. B. In *Industrial Micro-Organisms : Basic and Applied Molecular Genetics*; Batltz, R. H., Hegeman, G. D., Skatrud, P. L., Eds; American Society for Microbiology: Washington DC, 1997, in press.
(77) Wiesmann, K. E. H.; Cortés, J.; Brown, M. J. B.; Cutter, A. L.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1995**, *2*, 583.
(78) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 11373.
(79) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *Nature* **1995**, *378*, 263.
(80) Luo, G.; Pieper, R.; Rosa, A.; Khosla, C.; Cane, D. E. *Bioorg. Med. Chem. Lett.* **1996**, *4*, 995.
(81) Reference 2, Chapter 3, p 70.
(82) Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C. *Biochemistry* **1996**, *35*, 2054.
(83) Pieper, R.; Gokhale, R. S.; Luo, G.; Cane, D. E.; Khosla, C. *Biochemistry* **1997**, *36*, 1846.
(84) Caffrey, P.; Green, B.; Packman, L. C.; Rawlings, B. J.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **1991**, *195*, 823.
(85) Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995**, 1519.
(86) Paiva, N. L.; Demain, A. L.; Roberts, M. F. *J. Nat. Prod.* **1991**, *54*, 167.
(87) Paiva, N. L.; Demain, A. L.; Roberts, M. F. *Enzyme Microb. Technol.* **1993**, *15*, 581.
(88) Paiva, N. L.; Roberts, M. F.; Demain, A. L. *J. Ind. Microbiol.* **1993**, *12*, 423.
(89) Lowden, P. A. S.; Böhm, G. A.; Staunton, J.; Leadlay, P. F. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2249.
(90) Wallace, K. W.; Reynolds, K. A.; Koch, K.; McArthur, H. A. I.; Brown, M. S.; Wax, R. G.; Moore, B. S. *J. Am. Chem. Soc.* **1994**, *116*, 11600.
(91) Box, S. J.; Shelley, P. R.; Tyler, J. W.; Verall, M. S.; Warr, S. R. C.; Badger, A. M.; Levy, M. A.; Banks, R. M. *J. Antibiot.* **1995**, *48*, 1347.
(92) Findlay, J. A.; Liu, J.-S.; Burnell, D. J.; Nakashima, T. T. *Can. J. Chem.* **1982**, *60*, 2046.
(93) Salituro, G. M.; Zink, D. L.; Dahl, A.; Nielsen, J.; Wu, E.; Huang, L.; Kastner, C.; Dumont, F. J. *Tetrahedron Lett.* **1995**, *36*, 997.
(94) Nishida, H.; Sakakibara, T.; Aoki, F.; Saito, T.; Ichikawa, K.; Inagaki, T.; Kojima, Y.; Yamaguchi, Y.; Huang, L. H.; Guadliana, M. A.; Kaneko, T.; Kojima, N. *J. Antibiot.* **1995**, *48*, 657.
(95) Molnár, I.; Aparicio, J. F.; Haydock, S. F.; Khaw, L. E.; Schwecke, T.; König, A.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 1.
(96) Aparicio, J. F.; Molnár, I.; Schwecke, T.; König, A.; Haydock, S. F.; Khaw, L. E.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 9.

(97) Lomovskaya, N.; Fonstein, L.; Ruan, X.; Stassi, D.; Katz, L.: Hutchinson, C. R. *Microbiology* **1997**, *143*, 875.
(98) König, A.; Schwecke, T.; Molnár, I.; Böhm, G. A.; Lowden, P. A. S.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **1997**, *247*, 526.
(99) KleinKauf, H.; Von Döhren, H. *Eur. J. Biochem.* **1996**, *236*, 335.
(100) Haydock, S. F.; Aparicio, J. F.; Molnár, I.; Schwecke, T.; Khaw, L. E.; König, A.; Marsden, A. F. A.; Galloway, I. S.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1995**, *374*, 246.
(101) Motamedi, H.; Cai, S.-J.; Shafiee, A.; Elliston, K. O. *Eur. J. Biochem.* **1997**, *244*, 74.
(102) Byrne, K. M.; Shafiee, A,; Nielson, J. B.; Arison, B. H.; Monaghan, R. L.; Kaplan, L. In *Developments in Industrial Microbiology*; Nash, C., Hunter-Cevera, J., Cooper, R., Eveleih, D. E., Hamill, R., Eds.; Wm. C. Brown: Dubuque, Iowa, 1993; Vol. 32.
(103) Shafiee, A.; Chen, T. S.; Arison, B. S.; Dumont, F. J.; Colwell, L.; Kaplan, L. *J. Antibiot.* **1993**, *46*, 1397.
(104) Nielsen, J. B.; Hsu, M.-J.; Byrne, K. M.; Kaplan, L. *Biochemistry* **1991**, *30*, 5789.

CR9600316