PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Polyketide Synthase Complexes: Their Structure and Function in Antibiotic Biosynthesis

John A. Robinson

Phil. Trans. R. Soc. Lond. B 1991 **332**, 107-114
doi: 10.1098/rstb.1991.0038

Polyketide synthase complexes: their structure and function in antibiotic biosynthesis

JOHN A. ROBINSON

Organisch-Chemisches Institut, Universität Zürich, CH 8057 Zürich, Switzerland

SUMMARY

This paper gives an overview of existing knowledge concerning the structure and deduced functions of polyketide synthases active in antibiotic-producing streptomycetes. Using monensin A as an example of a structurally complex polyketide metabolite, the problem of understanding how individual strains of microorganism are 'programmed' to produce a given polyketide metabolite is first outlined. The question then arises, how is the programming of polyketide assembly related to the structural organization of individual polyketide synthase complexes at the biochemical and genetic levels? Experimental results that help to illuminate these relations are described, in particular, those giving information about the structures and deduced functions of polyketide synthases involved in aromatic polyketide biosynthesis (actinorhodin, granaticin, tetracenomycin, *whiE* spore pigment and an *act* homologous region from the monensin-producing organism), as well as the macrolide polyketide synthase active in the biosynthesis of 6-deoxyerythronolide A.

1. INTRODUCTION

The polyketide hypothesis enunciated by Birch stands out as an incisive contribution to the field of polyketide biogenesis (Birch & Donovan 1953; Birch 1957). It provided a valuable correlation between the structures of a large number of (largely) aromatic natural products and their probable modes of biosynthesis through the head-to-tail linkage of acetate units. The hypothesis was firmly grounded in mechanistic chemistry, and showed how a carbon chain of β-ketone groups, retained from the successive condensation of acetate units (instead of being serially removed as in fatty acid biosynthesis), might undergo well-precedented laboratory reactions, including aldol condensation, C-acylation, reduction, dehydration, methylation and oxidation, and thereby give rise to the large family of known polyketide metabolites. The biogenetic correlation was most apparent for certain plant phenolic compounds that have oxygen substituents attached β- to each other or to positions of ring closure, as a direct result of the β-positioning of ketone and methylene groups in the original chain. Birch also suggested that coenzyme-A esters other than acetyl-CoA might initiate chain assembly, while 'propionate' units (methylmalonyl-CoA) could be incorporated in place of malonyl-CoA during chain extension (Birch 1967).

Only a few years before, the first macrolide antibiotic had been reported by Brockmann & Henkel (1951), and the relevance of Birch's polyketide hypothesis to macrolide biosynthesis was quickly recognized. By 1957 at least six different macrolides had been discovered, and Woodward suggested that the macrolide rings could also arise by the stepwise condensation of acetate (or related) building blocks, with the oxygen atom of individual β-ketone groups retained as hydroxyl groups in the macrolide backbone (Woodward 1957). Since that time close to two hundred different macrolides have been isolated, largely from microorganisms of the genus *Streptomyces*, and many other important classes of antibiotics are now also known to be of polyketide origin. It is a remarkable aspect of natural products chemistry today that new, interesting, and therapeutically important polyketide metabolites continue to be discovered in Streptomycete screening programmes that are conducted largely within the pharmaceutical and agrochemical industries (e.g. avermectins, FK506, etc.).

The main focus of this paper, however, is not the chemistry of polyketides per se, but rather the enzymology of polyketide biogenesis as revealed over the past few years through the application of molecular genetic techniques. Although a good understanding of the central underlying chemical concepts in polyketide assembly had been laid by Birch's polyketide hypothesis, very little was known until recently about the structure of the polyketide synthases (PKSs) that catalyse carbon chain assembly, nor how their structure might be related to the 'programming' that directs individual strains of microorganism to produce a unique polyketide antibiotic. This problem of understanding PKS programming is discussed below with reference to the polyether ionophore antibiotic monensin A (an antibiotic of interest in the authors laboratory).

2. THE BIOSYNTHESIS OF MONENSIN; AND POLYKETIDE SYNTHASE PROGRAMMING

Monensin A is essentially a long highly functionalized fatty acid incorporating ether rings to enforce a

Phil Trans. R. Soc. Lond. B (1991) **332**, 107–114
Printed in Great Britain

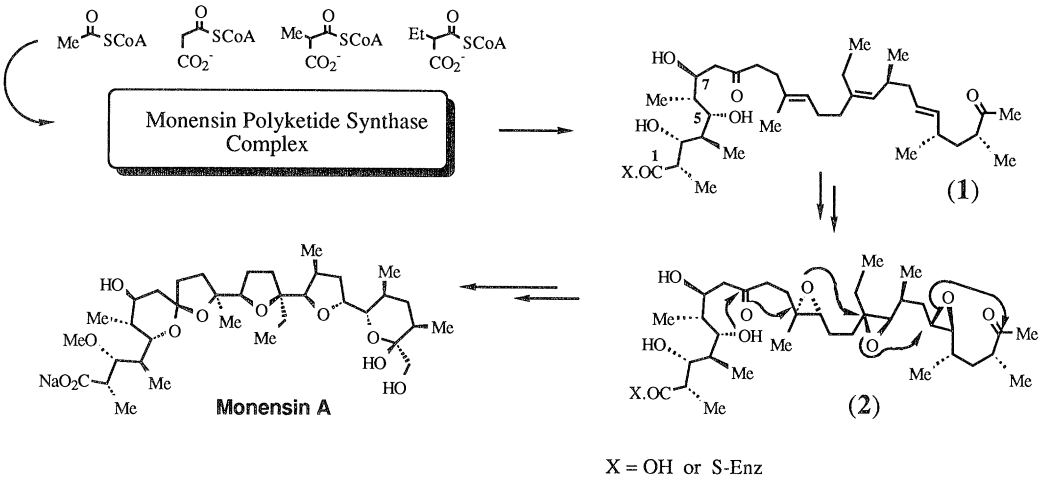

Figure 1. A hypothetical pathway to monensin A.

folded backbone conformation for binding Na^+ or K^+ cations. Carbon-13 labelling experiments revealed how five acetate units, seven propionates and one butyrate are linked in a head-to-tail (i.e. C(1) to C(2)) fashion during formation of the carbon backbone. The 3MeO methyl group is derived in addition from methionine, although the timing of methylation is uncertain. Similar experiments using ^{13}C- and ^{18}O-doubly labelled acetate and propionate, as well as fermentations under $^{18}O_2$, later defined the origins of all the oxygen atoms in monensin A (Cane *et al.* 1982; Ajaz *et al.* 1987). This information provided support for an attractive mechanism to account for ether ring formation, shown in figure 1, involving the triene (**1**) and triepoxide (**2**) as intermediates (Cane *et al.* 1983). The fact that each of the β-placed oxygen atoms from C1 through to C9 is incorporated intact shows that the absolute configuration of the C(3), C(5), and C(7) centres is established directly through stereodivergent reductions of β-ketone intermediates, rather than by additional dehydration and rehydration steps. Because C(3) is *R*, and C(5) and C(7) are *S*, this would seem to require at least two different ketoreductases having opposite stereospecificities. Similar results have been obtained in studies of macrolide biosynthesis, strengthening the idea of a common mode of stereocontrol during the assembly of these two classes of polyketide antibiotics (for a review see Robinson (1988)). In the biosynthesis of macrolides, and presumbly also polyethers, it is after completion of carbon chain assembly that the backbone is oxygenated, or glycosylated, steps that amplify the scope for structural diversity within each class of antibiotic.

The triene (**1**) is the presumed end product of a series of condensation, reduction and dehydration events, utilizing substrates derived from acetate, propionate and butyrate, catalysed by the monensin PKS multienzyme complex. An important question concerns the order of these various events during chain assembly. In macrolide biosynthesis, evidence has accumulated recently in support of the so-called processive stategy of backbone assembly (Yue *et al.* 1987; Cane & Yang 1987). In this strategy, the β-ketoacyl thioester formed by decarboxylative-condensation of an acyl chain with a malonyl, methylmalonyl, or ethylmalonyl extender unit, is modified (where necessary) before the next building block is incorporated. The modification can involve: reduction only to a β-hydroxythioester; reduction and dehydration to give an olefin; or reduction, dehydration and further reduction to give a fully saturated chain (see figure 2). However, the same modification need not occur during each cycle of chain extension. The evidence supporting this model of macrolide assembly includes the incorporation of putative chain elongation intermediates into tylactone (Yue *et al.* 1987), erythromycin B (Cane & Yang 1987) and nargenicin (Cane & Ott 1988), as well as the isolation of compounds considered to be chain elongation intermediates in the biosynthesis of protomycinolide IV (Takano *et al.* 1989). It is very likely that a processive strategy of chain assembly also operates during polyether biosynthesis, although successful incorporation experiments in support of this have not yet been described.

Chemical, biochemical and more recently molecular genetic studies have served to emphasize a fundamental similarity between fatty acid synthases (FASs) and PKSs, both in terms of their mechanisms of action and molecular organization. However, the synthesis of a polyketide must be more highly programmed than that of a fatty acid. For example, during assembly of the putative monensin triene intermediate (**1**), the PKS must: select an acetate starter unit (presumably acetyl-CoA) and then 12 further building blocks in a specific order from activated acetyl, propionyl and butyryl residues; incorporate the correct chemical functionality during each chain extension step, leaving a ketone, hydroxy, enoyl or saturated alkyl group; and establish the correct relative and absolute configuration at each new chiral centre and double bond. Leaving aside the question of stereochemistry, this programming must be translated into a series of choices regarding the type of extender unit, and the type of chemistry after each subunit addition, as shown in figure 2. Related schemes could be compiled in a similar way to arrive at other polyether intermediates, or macrolide rings. An important objective of current research is to understand how the programming, which allows the monensin PKS to assemble just one out of an enormous number of theoretically possible reduced polyketide chains, is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

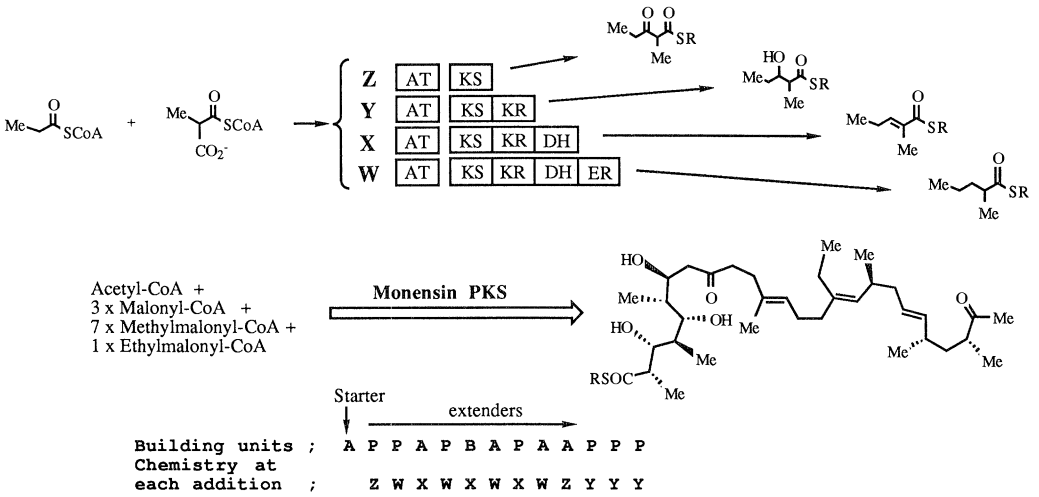

Figure 2. The monensin PKS must select the desired building unit (A = acetate, P = propionate, B = butyrate) and the correct chain extension chemistry (AT = acyl transferase, KS = β-ketoacyl synthase, KR = β-ketoacylreductase, DH = dehydrase, ER = enoyl reductase) at each round of chain elongation.

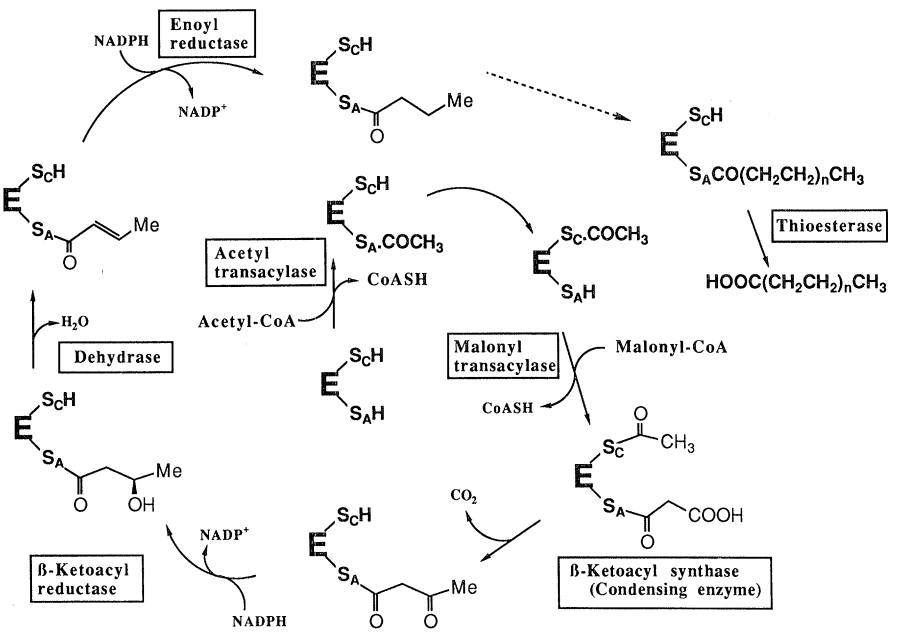

Figure 3. Steps in fatty acid biosynthesis (S_A = thiol on cofactor in the ACP; S_C = active site thiol in the condensing enzyme).

related to the structural organization of the polyketide synthase complex, at the biochemical and genetic levels.

3. THE STRUCTURE OF TYPE-I AND TYPE-II FATTY ACID SYNTHASES

In discussing the enzymology of PKS function it is helpful to draw upon comparisons to fatty acid synthases, where the relations between structure and function are becoming more clearly understood. The component activities of the type-II fatty acid synthases from plants and bacteria can be readily separated into discrete globular proteins, which together catalyse the consecutive steps of the fatty acid synthase cycle (see figure 3) (Fulco 1983). There is no evidence so far that these proteins form an aggregate within cells, and this is certainly not necessary to observe full fatty acid synthase activity *in vitro*. The following proteins compose the *E. coli* FAS responsible for palmitic acid biosynthesis (Volpe & Vagelos 1973): (i) the acyl carrier protein (ACP), a relatively small protein of around 80 amino acids, containing the phosphopantetheine cofactor. It plays a special role as it must bind the substrate at most steps of the cycle, and it must interact with all the component enzymes in the fatty acid synthase complex; (ii) and (iii) acetyl and malonyl transferases that load the substrates acetyl-CoA and malonyl-CoA onto ACPs, each via a mechanism involving discrete O-acyl(Ser)-enzyme intermediates: (iv) condensing enzymes (β-ketoacyl synthases), of which three have been isolated from *E.*

coli, that differ in their substrate specificities. The substrates for the synthase-I and synthase-II, both homodimeric proteins, are an acyl-ACP and malonyl-ACP, whereas the synthase-III condenses malonyl-ACP with acetyl-CoA (rather than with acetyl-ACP). The protein sequence of the β-ketoacyl synthase-I encoded by the *fabB* gene has been deduced from the gene sequence; (v) β-ketoacyl-ACP reductase, which catalyses the NADPH-dependent reduction of the β-ketoacyl-ACP to (3*R*)-β-hydroxyacyl-ACP; (vi) β-hydroxyacyl-ACP dehydrase, which catalyses a *syn*-elimination with formation of *E*-2-enoyl-ACP; (vii) enoyl-ACP reductase. Two distinct enoyl reductases are known in *E. coli* having slightly different acyl chain length specificities, one NADH dependent and the other NADPH dependent; and finally (viii) palmityl-ACP thioesterase, which catalyses hydrolysis of the acyl thioester when the acyl chain reaches C_{16} in length.

The type-I FASs from yeast and mammals have a quite different architecture. In animal cells the component activities of FAS are found on a single polypeptide chain around 2500 amino acids in length. The native enzyme from chicken and rat is then a homodimer (M_r 500000). Upon dissociating the polypeptide chains all the component activities are retained except that of the β-ketoacyl synthase. An elegant combination of protein biochemical and molecular genetic experiments have led to a detailed model for the architecture of the chicken FAS, in which the catalytic sites are arranged on a series of connected globular domains (Wakil 1989). Proteolytic cleavage of the chicken liver FAS leads initially to the release of three peptide fragments of M_r 127000, 107000 and 33000 corresponding to domains I, II and III in the intact protein. The smallest, domain III, contains the COOH terminus of the protein and the thioesterase. Domain I contains the NH_2 terminus, the β-ketoacyl synthase, and a single active Ser-OH used by both acetyl and malonyl transacylases. Domain II (the reduction domain) contains the dehydrase and enoyl and β-ketoacyl reductases, as well as the acyl carrier site that connects the β-ketoacyl reductase to the thioesterase. This physical map based on proteolysis experiments is in accord with the predicted locations for the component activities in the polypeptide, based on an analysis of the entire sequence of chicken and rat cDNAs (see figure 4) (for a review see Hopwood & Sherman (1990)).

Clearly, for the decarboxylative condensation to occur the malonyl extender unit on the ACP must be positioned next to the acyl chain attached to the active site cysteine on the β-ketoacyl synthase. In the sequence of both chicken and rat FASs, the ACP and β-ketoacyl synthase are located almost at opposite ends of the polypeptide. Further important biochemical information has come from studies with alkylating agents. 1,3-Dibromopropane reacts rapidly and specifically with the FAS, inactivating only the condensing enzyme, by alkylating *both* the cysteine-SH on one polypeptide *and* the pantetheinyl-SH in the ACP *of the other chain*. These and related observations with DTNB-inhibited FAS led to the proposal that the two chains are arranged head to tail so that the reactive thiols of

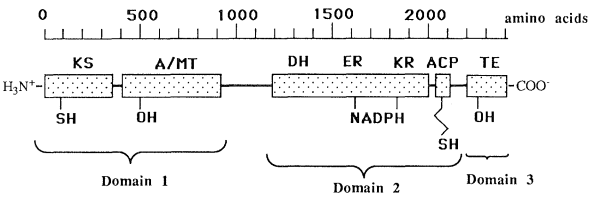

Figure 4. The location of activities along the type-I mammalian fatty acid synthase polypeptide deduced from the cDNA sequence (KS = β-ketoacyl synthase, A/MT = acetyl/malonyl transferase, DH = dehydrase, ER = enoyl reductase, ACP = acyl carrier protein domain, TE = thioesterase). The organization of activities into domains, as deduced by limited proteolysis, is also shown (Wakil, S. (1989) *Biochemistry* **28**, 4523).

the relevant ACP and condensing enzyme are juxtaposed (Wakil 1989). There are two full sites for fatty acid assembly. In this model each centre of palmitate synthesis derives the required component activities from both chains; one subunit contributes domain I whereas the second subunit contributes domains II and III.

The order of catalytic domains in the yeast FAS, deduced from cDNA sequence data, show some interesting differences to that found in the animal FAS (for a review see Hopwood & Sherman (1990)). The component activities are in this case distributed along two non-identical polypeptides. The α-subunit contains the acyl carrier site, the β-ketoacyl synthase and β-ketoacyl reductase, whereas the β-subunit contains domains for the remaining activities, including an FAD binding site in the enoyl reductase domain. In addition, there is no recognizable thioesterase domain, as expected because the end product is palmitoyl-CoA and not free palmitate. Thus the different end product of the FAS reflects a different combination of catalytic domains within the multifunctional protein, an observation that is of special interest in the context of the programming of polyketide synthase function.

4. THE STRUCTURE OF POLYKETIDE SYNTHASE COMPLEXES

Biochemical studies have proven a great success in the characterization of FASs from various organisms, however, the same is not true for PKSs from Streptomycetes. Several reasons for this can be found, including low levels of expression (typical for enzymes of secondary metabolism) under normal cellular conditions, and the very great problem of devising sensitive and specific assays applicable in cell free systems. Although the 6-methylsalicylic acid synthase has been purified from *Penicillium patulum*, and both chalcone synthase and resveratrole synthase have been purified from various plant species, no PKS active in polyketide antibiotic-producing Streptomycetes has yet been detected in an assay based on cell-free activity (for a review see Robinson (1989)). On the other hand, as a result of advances in molecular genetics, the genes encoding the biosynthetic pathways to several polyketide natural products have now been cloned from various *Streptomyces* species, including those for actino-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

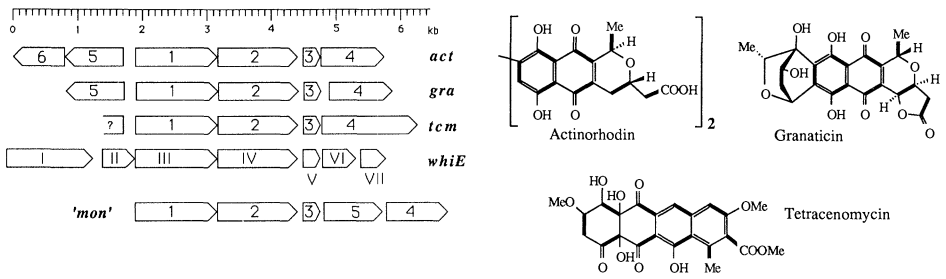

Figure 5. The organization of ORFs in the granaticin (*gra*), actinorhodin (*act*), tetracenomycin (*tcm*), *whiE* and '*mon*' polyketide synthase gene clusters. The deduced functions for these ORFs are: ORF1 + ORF2 (ORFIII + IV) = β-ketoacyl synthase; ORF3 (ORFV) = acyl carrier protein; ORF4 = cyclase/dehydrase (*gra*, *act*, and '*mon*'), cyclase/O-methyltransferase (*tcm*); ORF5 = ketoreductase; ORF6 = ketoreductase. See text for references.

rhodin, granaticin, oxytetracycline, tetracenomycin, daunorubicin, erythromycin, tylosin, spiramycin, carbomycin, avermectin, candicidin, as well as a spore pigment of unknown structure encoded by the *whiE* locus in *S. coelicolor* (for a review see Hopwood & Sherman (1990)). Two circumstances considerably facilitate the cloning of biosynthetic genes from Streptomycetes. One is that the genes of individual biosynthetic pathways are clustered on the host chromosome, and the second is that at least one resistance gene is always closely linked to the biosynthetic gene cluster. It is often relatively straightforward to clone an antibiotic resistance gene, and then to look in the flanking DNA for biosynthetic genes. An alternative, and more direct approach is by complementation of blocked mutants, and this was the method used to isolate biosynthetic genes for actinorhodin and tetracenomycin. From a genetic analysis of the DNA cluster, the region encoding the PKS genes can then usually be identified, and this upon sequencing leads directly to the primary structure(s) of the PKS protein(s).

The DNA for actinorhodin biosynthesis was the first antibiotic gene cluster to be cloned from a Streptomycete (Malpartida & Hopwood 1984). The DNA encoding the actinorhodin PKS was identified because it complemented mutants of the actinorhodin producing organism that were blocked in carbon chain assembly (*act*I and *act*III mutants). The sequence of this *act*I and *act*III DNA has been determined (Hopwood & Sherman 1990), but only that in the *act*III region has so far been reported in full (Hallam *et al.* 1988). Meanwhile, DNA sequences encoding PKSs required for granaticin and tetracenomycin biosynthesis have been described (Sherman *et al.* 1989; Bibb *et al.* 1989) and a computer assisted analysis revealed regions comprising a series of open reading frames (ORFs) (see figure 5) encoding discrete polypeptide chains. A third PKS cluster was cloned by complementing spore-pigment-deficient *S. coelicolor* mutants (*whiE*) (Davis & Chater 1990). When this DNA was sequenced a set of ORFs was again revealed, and several of the deduced protein sequences were closely related to those found in the actinorhodin and tetracenomycin PKS genes. The structure of the *whiE* spore pigment is presumably an aromatic polyketide, but its structure is presently unknown. Another putative polyketide synthase gene cluster has been isolated from the monensin producer *S. cinnamonensis* by using the *act*I and *act*III DNA as probes to screen a genomic library (Arrowsmith 1990). Earlier it had been shown that DNA from *act*I and *act*III hybridizes to homologous sequences in genomic DNA from a large number of polyketide-producing Streptomycetes (Malpartida *et al.* 1987). Indeed, the homologous regions of DNA in granaticin- and oxytetracycline-producing Streptomycetes were later identified as the PKS genes involved in the production of these two antibiotics. The sequence of the *act*-homologous DNA from *S. cinnamonensis* has also been determined, and although its function is not yet known (it may be part of the monensin PKS cluster), a similar arrangement of encoded protein ORFs was identified (see figure 5).

A feature common to the sequenced *gra*, *act*, *tcm*, *whiE* and '*mon*' PKS gene clusters is a group of three ORFs, labelled ORF1-2-3, or ORFIII-IV-V, in figure 5. When corresponding ORFs from each cluster are compared, a very high end-to-end protein sequence similarity is observed. Moreover, the deduced protein sequences of ORFs 1 and 2 (ORFIII and IV) show a particularly high similarity with the *fabB* gene encoding the β-ketoacyl synthase-I from *E. coli*. In each case a potential active site Cys occurs only in ORF1 (ORFIII) (see figure 6). This fact, along with the possibility of translational coupling between these first two ORFs (because of overlapping 3′ stop/5′ start codons), suggests that the gene products are produced stoichiometrically to form a heterodimer, rather than being two distinct condensing enzymes. The deduced protein sequence of ORF3 (ORFV) in each cluster shows high end-to-end similarity to ACPs from bacteria and plants, as well as to ACP domains of animal type-I FAS. This includes a conserved Ser to which the cofactor is covalently attached (see figure 6). It is notable that plant chalcone and resveratrole synthases do not contain a covalently bound 4′-phosphopantetheine cofactor.

There is genetic evidence that ORF5 from the *act*III region encodes a ketoreductase, and the *gra* ORF5 and ORF6 as well as the '*mon*' ORF5 deduced protein sequences are very similar to this deduced *act*III gene product. These *act* ORF5, *gra* ORF5 (but not *gra* ORF6) and '*mon*' ORF5 sequences contain a typical consensus nucleotide binding motif Gly-Xaa-Gly-Xaa-Xaa-Ala (see figure 6). The *tcm* PKS cluster reveals no sequence homologous with these putative keto-

(*a*) **ß-ketoacyl synthase**

gra ORF1	G P V T M V S D G **C** T S G L D S V	
tcm ORF1	G P V T V V S T G **C** T S G L D A V	
whiE ORFIII	G P V Q T V S T G **C** T S G L D A V	
'mon' ORF1	G P S T V V S T G **C** T S G I D A V	
6-MSAS	G P S T A V D A A **C** A S S L V A I	(*P. patulum*)
FAS *fabB*	G V N T S I S S A **C** A T S A H C I	(*E. coli*)
FAS rat	G P S I A L D T A **C** S S S L L A L	
FAS chicken	G P S L T I D T A **C** S S S L M A L	
ery ORFA (KS1)	G P A M T V D T A **C** S S G L T A L	
ery ORFA (KS2)	G P A V T V D T A **C** S S S L V A L	

(*b*) **ACP**

gra ORF3	E E L G Y D **S** L A L M E S A
tcm ORF3	Q D L G Y D **S** I A L L E I S
whiE ORFV	D T F G L D **S** L G L L G I V
'mon' ORF3	A L L G Y E **S** L A L L E T G
6-MSAS	A D L G V D **S** V M T V T L R
rat FAS	A D L G L D **S** L M G V E V R
chicken FAS	A D L G L D **S** L M G V E V R
S. erythraea	E D L G M D **S** L D L V E X V
E. coli FAS	E D L G A D **S** L D T V E L V
ery ORFA (ACP1)	R D L G F D **S** M T A V D L R
ery ORFA (ACP2)	T E L G F D **S** L T A V G L R

(*c*) **NAD(P) binding domains**

actIII	P V D V L V N N A G R P **G** G **G** A T **A** E L A **D**
gra ORF5	T V D I L V N N A G R S **G** G **G** A T **A** E I A **D**
'mon' ORF5	R I D V L V N N A G R S **G** G **G** V T **A** D L T **D**
6-MSAS	L P R P E G T Y L I T G **G** L **G** V L **G** L E V **A**
chick FAS^{1811} KR	S C P P T K S Y I I T G **G** L **G** G F **G** L E L **A**
rat FAS^{1800} KR	F C P E H K S Y I I T G **G** L **G** G F **G** L E L **A**
ery ORFA^{1121} (KR1)	L E P L A G T V L V T G **G** I **G** A H **L** A R W **L**
ery ORFA^{2558} (KR2)	S W E P A G T A L V T G **G** T **G** A L **G** G H V **A**
DHFR mouse	P L N C I V A V S Q N M **G** I **G** K N **G** D L P **W**
chick FAS^{1597} ER	K G E S V L I H S G S G **G** V **G** Q A **A** I A I **A**
rat FAS^{1587} ER	H G E T V L I H S G S G **G** V **G** Q A **A** I S I **A**

Figure 6. Alignments of segments of deduced protein sequence: (*a*) around the presumptive active site cysteine in β-ketoacyl synthases; (*b*) around the active site serine of putative ACPs; (*c*) NAD(P)H binding domains in β-ketoacyl reductases and enoyl reductases.

reductases, consistent with the fact that no reductive step is required during polyketide assembly; all the β-placed oxygens are retained! There is also no ORF in the *whiE* cluster whose deduced protein sequence shows high similarity to those of *gra* and *act* ORF5.

The deduced protein sequences in ORF4 from *act*, *gra*, *tcm* and '*mon*' show relatedness to each other, although the sequence similarities sometimes extend only over part of the deduced protein sequence. The *act*, *gra* and '*mon*' ORF4 sequences are similar throughout the length of the protein, but each resembles *tcm* ORF4 only in the N-terminal half of the protein. Since *act* and *gra* ORF4 complement *act* VII mutants, deduced to be defective in cyclization and dehydration events, it has been suggested that *act* and *gra* ORF4s encode a bifunctional cyclase/dehydrase. In contrast, the *tcm* ORF4 protein seems to be a cyclase/O-methyltransferase because its C-terminal half resembles bovine hydroxyindole-O-methyltransferase and *tcm* ORF4 complements *tcm* mutants lacking a specific O-methylation step (see Hopwood & Sherman 1990).

It is interesting that the deduced *gra* and *tcm* PKS clusters appear very similar at the protein level, despite the fact that polyketide chains are synthesized of different lengths and different patterns of folding. It seems likely that the folding of the polyketide chain and subsequent cyclization steps are catalysed by ORF4 and probably also by other proteins encoded outside the region sequenced so far. In any event, an important conclusion from this work is that these PKS genes for aromatic polyketides encode proteins most clearly resembling the type-II FAS from bacteria and plants, rather than being large multifunctional polypeptides as seen in the type-I FAS and the 6-methylsalicylic acid synthase from *Penicillium patulum* (Beck *et al.* 1990).

Other studies on macrolide biosynthesis have recently furnished a remarkable view of macrolide PKS structure, which begin to provide an insight into how the synthesis of a macrolide is programmed at the molecular level. 6-Deoxyerythronolide B, the first isolable intermediate on the pathway to erythromycin A, is assembled from a propionyl-CoA starter unit and six methylmalonyl-CoA extender units (see figure 7). The genes encoding erythromycin biosynthesis are clustered in the genome of *Saccharopolyspora erythraea* and in the middle of the cluster lies the resistance gene *erm*E. Located about 12 kb downstream from this resistance genes is a DNA segment capable of complementing EryA mutants blocked in the biosynthesis of the erythronolide ring. This *eryAI* locus encodes the macrolide PKS. A further DNA segment homologous to *eryAI*, designated *eryAII*, was localized to a region about 35 kb downstream from *ermE*, and was also shown to encode genes for the macrolide PKS (Tuan *et al.* 1990).

The *eryAI* and *eryAII* DNA has now been sequenced and this has provided crucial information, not only about the size, but also about the probable functions of ORFs encoded in the *ery* PKS cluster. The sequence of one such ORF (ORFA) extending over 9.5 kb of DNA has been reported by Leadlay's group (Cortes *et al.* 1990). The deduced gene product is predicted to

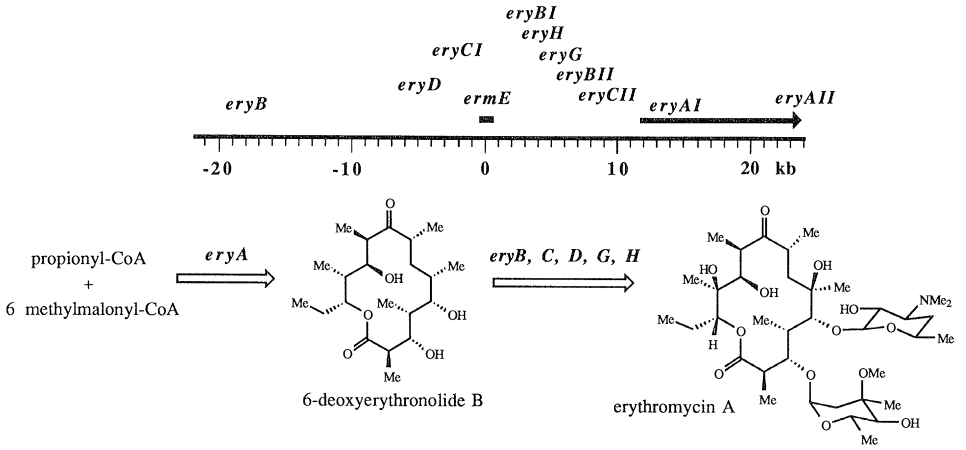

Figure 7. The erythromycin biosynthetic gene cluster, showing relative locations of the resistance gene (*ermE*) and biosynthetic genes (*eryA*-*eryH*) (Weber *et al.* 1990).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

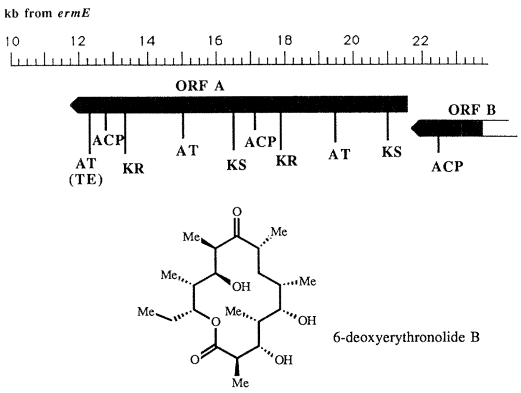

Figure 8. An open reading frame in the *ery* cluster (*eryAI*) encoding a multifunctional polypeptide, where the deduced activities are: AT(TE) = acyl transferase (thioesterase), ACP = acyl carrier protein domain, KR = ketoreductase, KS = β-ketoacyl synthase (Cortes *et al.* 1990).

contain 3178 amino acids. Upon comparison with available protein-sequence databases, nine separate portions of the deduced protein are found to be very similar to active-site sequences found in the constituent catalytic activities of known FASs and PKSs. The deduced activities for various regions of this predicted protein are shown in figure 8, and include in sequence: AT(or 'TE')-ACP-KR-AT-KS-ACP-KR-AT-KS (see figure 8 for definitions of symbols). A comparison of putative active site residues in this ORF with those found in corresponding components of other FAS and PKS complexes is shown in figure 6. These results reinforce the earlier conclusion that this portion of the *ery* cluster encodes a PKS, and show also that the bacterial PKSs are not necessarily complexes of monofunctional proteins, a trend evident with the aromatic PKSs.

In a parallel investigation Katz's group have also sequenced further downstream and found two additional large ORFs of comparable size to that described by Leadlay. Each of these ORFs again appears to encode a single large multifunctional protein containing multiple copies of units comprising AT, ACP, KS, KR activities, and one ORF contains also one set of dehydrase and enoyl reductase activities (L. Katz; personal communication). It seems, therefore, that the erythronolide PKS may comprise three large multienzyme complexes, possessing all the activities required to catalyse six rounds of chain extension and modification. With this information in hand it becomes possible to speculate as to how the individual steps in 6-deoxyerythronolide biosynthesis might be catalysed by the deduced activities encoded in the three *ery* ORFs. Katz has proposed a 'module-hypothesis' in which each polypeptide carries the functions for two rounds of chain elongation and modification. The DNA sequence for each round is called a 'module'. The deduced activities encoded in ORF A shown in figure 8 would comprise the last two modules required to add the final two propionate units and complete the synthesis of 6-deoxyerythronolide A. There are apparently two condensing enzymes and two ketoreductases, and at the C-terminus of the deduced protein is a sequence showing similarity to known thioesterases, which might therefore catalyse the final act of macrolide ring formation. This hypothesis implies that each of the deduced activities KS, KR, DH and ER is used once during the synthesis of each molecule of 6-deoxyerythronolide, and consistent with this is the total number of such activities; 6 × KS, 5 × KR, 1 × DH, 1 × ER. There are also multiple copies of deduced ACP domains and AT activities. This leaves sufficient scope for differences in stereospecificity to arise amongst the ketoreductases, for example, or among the acyl transferases ((*R*)- versus (*S*)-methylmalonyl-CoA (Hutchinson 1983)). There are no real clues yet as to how the growing acyl chain is transferred between active sites, and between the three different polypeptide chains. Leadlay has suggested that the two halves of the ORFA gene product might fold back on each other so as to bring the predicted *β*-ketoacyl synthase active sites into close proximity to the ACP domains (Cortes *et al.* 1990).

The remarkable picture emerging from these studies is one of a colinearity between the biochemical steps in macrolide assembly and the genetic order of ORFs encoding the requisite multifunctional proteins in the genome of the producing organism. Whether this picture has been interpreted correctly will no doubt be tested by future biochemical experiments. Although one swallow does not make a summer, should this correlation prove to be a general one among macrolide and polyether PKS complexes, then an important step forward will have been made in understanding the programming of these polyketide pathways at the biochemical and genetic levels. The prospect also arises of using rDNA methods to engineer new biosynthetic pathways. Protein engineering experiments might, for example, involve swapping modules or ORFs between polyketide pathways in order to make homologues of known antibiotics, or bringing modules or ORFs into novel combinations so that the host microorganism is programmed to biosynthesize entirely new classes of natural products. Alternatively, modules or simply portions of modules might be deleted or disrupted to generate new end products, or release intermediates in the assembly process. Although the prospects seem good, determining where the limits lie for the rational manipulation of polyketide biosynthesis will probably require a far more extensive knowledge of these processes, at all levels.

The author thanks Professor David Hopwood for valuable discussions, and Dr Leonard Katz and co-workers for providing information on the *ery* PKS prior to publication.

REFERENCES

Ajaz, A. A., Robinson, J. A. & Turner, D. L. 1987 Biosynthesis of the polyether ionophore antibiotic monensin A: assignment of the carbon-13 and proton NMR spectra of monensin A by 2D spectroscopy. Incorporation of oxygen-18 labelled molecular oxygen. *J. chem. Soc. Perkin Trans. I* 27–36.

Arrowsmith, T. J. 1990 Characterisation of putative polyketide synthase genes from *Streptomyces cinnamonensis*. Ph.D. thesis, University of Southampton.

Beck, J., Ripka, S., Siegner, A., Schiltz, E. & Schweizer, E. 1990 The multifunctional 6-methylsalicyclic acid syn-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

thase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. *Eur. J. Biochem.* **192**, 487–498.

Bibb, M. J., Birö, S., Motamedi, H., Collins, J. F. & Hutchinson, C. R. 1989 Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J* **8**, 2727–2736.

Birch, A. J. 1957 Biosynthetic relations of some natural phenolic and enolic compounds. *Prog. Chem. Org. Nat. Prod.* **14**, 186–216.

Birch, A. J. 1967 Biosynthesis of polyketides and related compounds. *Science, Wash.* **156**, 202–206.

Birch, A. J. & Donovan, F. W. 1953 Studies in relation to biosynthesis I. Some possible routes to derivatives of orcinol and phloroglucinol. *Aust. J. Chem.* **6**, 360–368.

Brockmann, H. & Henkel, W. 1951 Pikromycin ein bitter schmeckender Antibioticum aus Actinomyceten *Chem. Ber.* **84**, 284–288.

Cane, D. E., Liang, T.-C. & Hasler, H. 1982 Polyether Biosynthesis 2. Origin of the oxygen atoms of monensin A. *J. Am. chem. Soc.* **104**, 7274–7281.

Cane, D. E., Celmer, W. D. & Westley, J. W. 1983 Unified stereochemical model of polyether antibiotic structure and biogenesis. *J. Am. chem. Soc.* **105**, 3594–3600.

Cane, D. E. & Ott, W. R. 1988 Macrolide Biosynthesis 5. Intact incorporation of a chain-elongation intermediate into nargenicin. *J. Am. chem. Soc.* **110**, 4840–4841.

Cane, D. E. & Yang, C.-C. 1987 Macrolide biosynthesis 4. Intact incorporation of a chain-elongation intermediate into erythromycin. *J. Am. chem. Soc.* **109**, 1255–1257.

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, Lond.* **348**, 176–178.

Davis, N. K. & Chater, K. F. 1990 Spore colour in *Streptomyces coelicolor* A(3)2 involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Molec. Microb.* **4**, 1679–1691.

Fulco, A. J. 1983 Fatty acid metabolism in bacteria. *Prog. Lipid Res.* **22**, 133–160.

Hallam, S. E., Malpartida, F. & Hopwood, D. A. 1988 Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. *Gene* **74**, 305–320.

Hopwood, D. A. & Sherman, D. H. 1990 Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *A. Rev. Genet.* **24**, 37–66.

Hutchinson, C. R. 1983 Biosynthetic studies of macrolide and polyether antibiotics. *Acct. chem. Res.* **16**, 7–14.

Malpartida, F. & Hopwood, D. A. 1984 Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature, Lond.* **309**, 462–464.

Malpartida, F., Hallam, S. E., Kieser, H. M., Motamedi, H., Hutchinson, C. R., Butler, M. J., Sugden, D. A., Warren, M., McKillop, C., Bailey, C. R., Humphreys, G. O. & Hopwood, D. A. 1987 Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthesis genes. *Nature, Lond.* **325**, 818–820.

Robinson, J. A. 1988 Enzymes of secondary metabolism in microorganisms. *Chem. Soc. Rev.* **17**, 383–452.

Sherman, D. H., Malpartida, F., Bibb, M. J., Kieser, H. M., Bibb, M. J. & Hopwood, D. A. 1989 Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Sterptomyces violaceoruber* Tü22. *EMBO J.* **8**, 2717–2725.

Takano, S., Sekiguchi, Y., Shimazaki, Y. & Ogasawara, K. 1989 Stereochemistry of the proposed intermediates in the biosynthesis of mycinamicins. *Tetrahedron Lett.* **30**, 4001–4002.

Tuan, J. S., Weber, J. M., Staver, M. J., Leung, J. O., Donadio, S. & Katz, L. 1990 Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*E. coli* cosmid. *Gene* **90**, 21–29.

Volpe, J. J. & Vagelos, P. R. 1973 Saturated fatty acid biosynthesis and its regulation. *A. Rev. Biochem.* **42**, 21–60.

Wakil, S. 1989 Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* **28**, 4523–4530.

Weber, J. M., Leung, J. O., Maine, G. T., Potenz, R. H. B., Paulus, T. J. & DeWitt, J. P. 1900 Organisation of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J. Bact.* **172**, 2372–2383.

Woodward, R. B. 1957 Struktur und Biogenese der Makrolide; Eine neue Klasse von Naturstoffen. *Angew. Chem.* **69**, 50–58.

Yue, S., Duncan, J. S., Yamamoto, Y. & Hutchinson, C. R. 1987 Macrolide biosynthesis. Tylactone formation involves the processive addition of three carbon fragments. *J. Am. chem. Soc.* **109**, 1253–1255.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES