Tetrahedron Vol. 47, No. 31, pp. 6029-6043, 1991
Printed in Great Britain

0040-4020/91 $3.00+.00

Molecular Genetic Analysis Reveals a Putative Bifunctional Polyketide Cyclase/Dehydrase Gene from *Streptomyces coelicolor* and *Streptomyces violaceoruber*, and a Cyclase/O-methyltransferase from *Streptomyces glaucescens*.

David H. Sherman[1,2],* Maureen J. Bibb[1], Thomas J. Simpson[3], Darrin Johnson[4], Francisco Malpartida[5], Miguel Fernandez-Moreno[5], Eduardo Martinez[5], C. Richard Hutchinson[6], and David A. Hopwood[1]

[1]John Innes Institute, John Innes Centre for Plant Science Research, Norwich NR4 7UH, England
[2]Institute for Advanced Studies in Biological Process Technology, and Department of Microbiology, University of Minnesota, 1479 Gortner Avenue, St. Paul, Minnesota 55108, USA
[3]Department of Chemistry, University of Bristol, Bristol BS81TS, England
[4]Molecular Biology Computing Center, University of Minnesota, 1479 Gortner Avenue, St. Paul., Minnesota, 55108, USA
[5]Centro Nacional de Biotecnologia, Serrano 115, 28006 Madrid, Spain
[6]School of Pharmacy and Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706, USA

(*Received in USA* 5 *March* 1991)

Abstract—Molecular genetic analysis of the actinorhodin gene cluster from *Streptomyces coelicolor* has revealed a putative bifunctional cyclase/dehydrase gene. Open reading frame (ORF) 4 from both the actinorhodin (*act*) and granaticin (*gra*) polyketide synthase (PKS) gene clusters were able to relieve the block in a mutant strain of *S. coelicolor* which produces mutactin. This compound is a shunt product of the actinorhodin biosynthetic pathway which results from an aberrant intramolecular aldol cyclization, and a failure to dehydrate the hydroxyl group at C-9. These results provide compelling evidence for the existence of a novel polyketide aldolase which specifies the correct cyclization of a complex oligoketide chain.

Polyketide biosynthesis by bacteria, fungi, and plants provides an amazing range of diverse secondary metabolites. Many of these molecules from actinomycetes have proven to be important antibiotics or other chemotherapeutic agents, while polyketides from fungi and higher plants include toxins, pigments and flavor compounds. Although the diversity of polyketide structural classes is vast, the underlying biosynthetic mechanisms ultimately represent a single theme. This in turn is conceptually analogous to fatty acid biosynthesis, a relationship which was first proposed by Collie,[1] and later articulated by Birch.[2] Iterative condensation of acetate units (derived from the CoA esters of acetate and malonate), with a cycle of ketoreduction, dehydration, and enoylreduction after each condensation, leads to the characteristic saturated chemical composition of linear, long chain fatty acids. In contrast, polyketide biosynthesis can involve the use of residues of a single, simple carboxylic acid subunit (acetate or propionate) or a mixture of subunits (e.g. acetate, propionate, butyrate residues). Additionally, the cycle of condensation, reduction, dehydration, reduction can be curtailed in polyketide biosynthesis, resulting in the characteristic partially oxidized state of these molecules. Another important structural aspect of many polyketide-derived metabolites is the cyclic nature of the products, which often involves the formation of six-membered rings. Mechanistically, this process is the result of one or more intramolecular aldol or Dieckmann cyclization reactions involving the oligoketide intermediate. Chemical studies[3] have investigated the regiochemical course of a significant number of synthetic oligoketides, but there has been little insight into the role that enzymes might play in cyclization of biosynthetically derived polycarbonyl compounds *in vivo*.

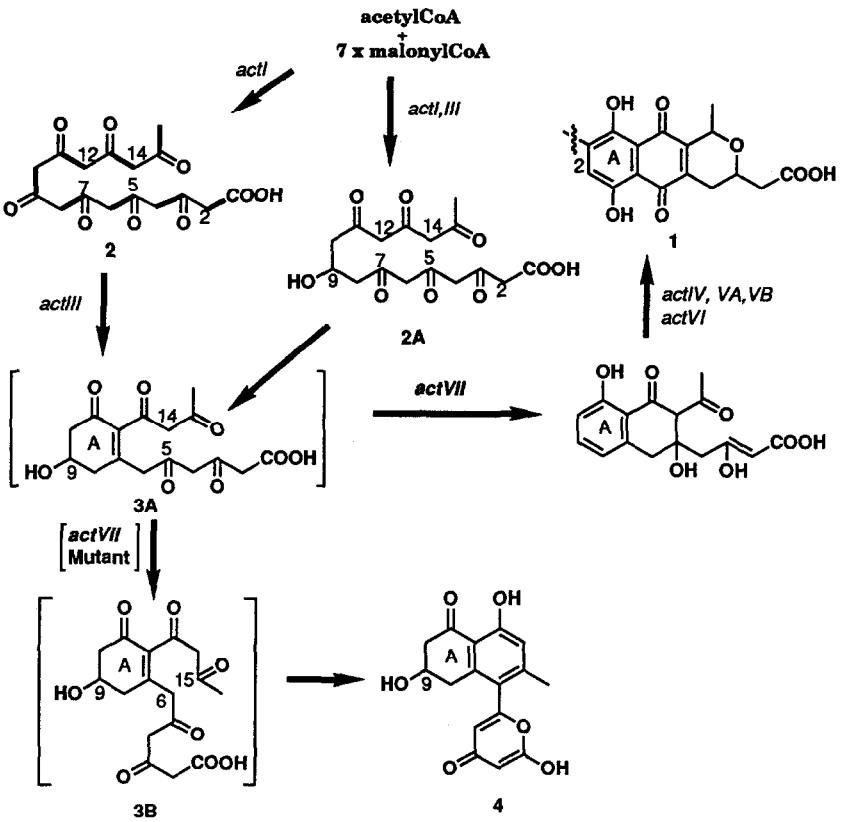

Figure 1. Proposed biosynthetic scheme for actinorhodin[7] and mutactin.[8] Symbols in italics denote biosynthetic genes or sets of genes in the actinorhodin cluster identified by biochemical genetics.[11] Molecules in brackets represent presumed intermediates leading to actinorhodin or mutactin.

This report describes the evidence for a novel polyketide aldolase which carries out a key step in the formation of the benzoisochromane quinone antibiotic, actinorhodin.[4] As shown in Figure 1, the monomeric unit of actinorhodin **1** is derived from an acetylCoA starter unit and seven malonylCoA extender units. For the sake of clarity, a distinct octaketide **2** is shown in the biosynthetic scheme. Whether this exists as shown, or whether reduction of the carbonyl group at C-9 to a hydroxyl would occur during chain building to yield **2A** is unclear: there is an argument for reduction at C-9 only after chain building is complete,[5] but this conflicts with the other currently held belief (reviewed in ref. 6) that modification of keto groups, when it occurs, does so during chain building. In any case, it is possible to surmise from previous studies on the biosynthesis[7] of actinorhodin, that an initial intramolecular aldol condensation must occur between C-12 and C-7. This regiospecific reaction is the first of two internal C-C bonds which are formed to construct the benzoisochromane quinone ring system. The second aldol condensation directly influences the final structure of the biosynthetic product. If internal C-C bond formation occurs between C-14 and C-5, the intermediate **3A** will proceed to give actinorhodin. However, if C-C bond formation occurs between C-6 and C-15 (intermediate **3B**), an alternative structure, mutactin **4** (ref. 8), results.

Work on the polyketides has been primarily limited to precursor incorporation and structural

elucidation studies. Molecular genetics has now provided a mechanism to begin analyzing genes and enzymes involved in the biosynthesis of polyketide-derived compounds. In this study we describe a method for identifying the genes and deduced enzymes which control the formation of the C-14/C-5 bond, and the dehydration step crucial to elaboration of benzoisochromane quinone antibiotics.

In order to put the problem of *in vivo* polyketide cyclization in a chemical perspective, it is useful to describe a number of *in vitro* cyclization studies with synthetic polycarbonyl compounds. The closest homologue to the presumed actinorhodin octaketide precursor **2** is a methyl-terminated heptaketide.[9] A highly significant finding is that *in vitro*, methyl-terminated polyketones undergo preferential cyclization by aldol processes involving the terminal carbonyl group. In order to determine how steric effects could influence ring formation, the ketal-protected heptaketone **5** (Figure 2) was synthesized.[9] Two initial cyclization reactions were obtained involving C-7/C-12 (Fig. 2, Scheme 1) and C-5/C-10 (Fig. 2., Scheme 2) aldol condensations. The former route was obtained in good yield with only traces of the latter detected. The C-5/C-10 cyclization of Scheme 2 could be further favored if larger R' groups were used in the synthesis of heptaketone **6.** Alternatively, if the central carbonyl group of heptaketide **9** is protected by formation of the corresponding ketal (Fig. 2, Scheme 3), the C-5/C-10 reaction occurred exclusively.[10] This is followed by *in situ* formation of the naphthalene derivative **10** which can be readily transformed to emodin **8.** In view of the actinorhodin biosynthetic pathway from *Streptomyces coelicolor*, and the mutactin pathway from the B40 mutant strain of *Streptomyces coelicolor* (Figure 1), these chemical studies provide important insight into the specificity of the *in vivo* cyclization reactions.

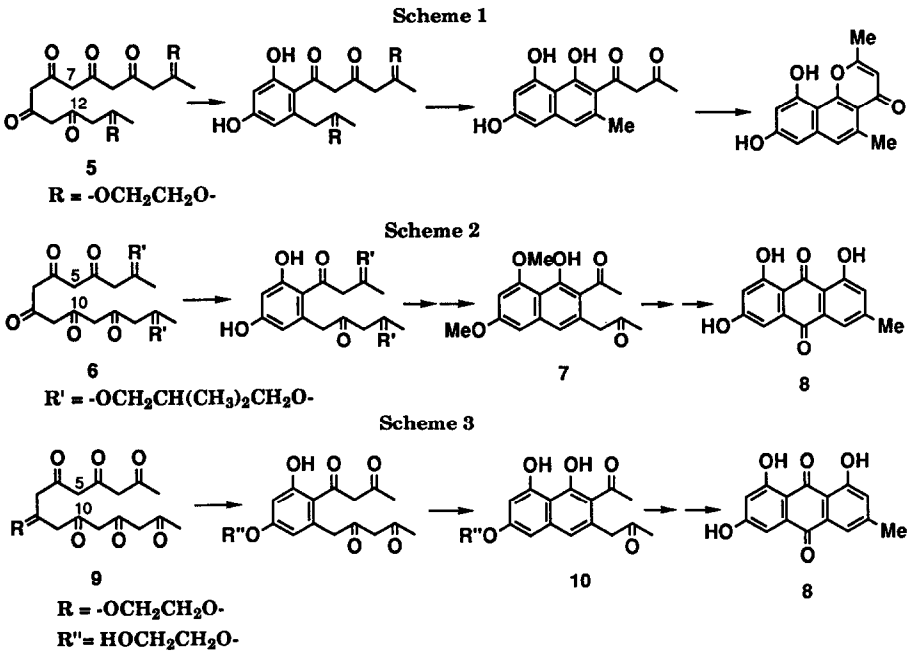

Figure 2. Cyclization pattern of synthetic oligoketides *in vitro*.[3] Scheme 1 shows cyclization of a heptaketide with terminal keto groups protected as the corresponding ketal. Scheme 2 shows cyclization of a heptaketide with terminal keto groups protected with the bulky ketal of 2,2-dimethyl-1,3-propanediol. Scheme 3 shows the cyclization of a heptaketide with the central keto group protected as the corresponding ketal.

First, given the propensity for *in vitro* internal aldol reactions to occur at the terminal ketones of oligoketides, what guides the initial (non-terminal ketone) octaketide cyclization in actinorhodin biosynthesis? Second, following formation of the A ring precursor (**3A, 3B**, Figure 1), what molecular mechanisms operate to influence the second internal aldol condensation, which will be determined in part by the stability of the two conformers (**3A, 3B**)? The former question can not be addressed directly in this study. However, a combination of genetics and chemistry has led to a more detailed understanding of the second cyclization in the actinorhodin pathway, which may prove to be a general process in the biosynthesis of benzoisochromane quinones and related classes of compounds.

Molecular genetic analysis of actinorhodin biosynthesis has provided a powerful complement to biosynthetic studies of the benzoisochromane quinones.[7] This in turn has introduced a general methodology for the study of polyketide systems. Initially, a series of mutants blocked at particular stages of the actinorhodin pathway were developed and characterized.[11] This led to the deduction of an ordered biosynthetic sequence: I,III,VII,IV,VI,V. *ActI,III* mutants failed to secrete any biosynthetic intermediate active in co-synthesis with any other classes of mutant, while being able to convert to actinorhodin intermediates secreted by the other mutant classes. The B40 *actVII* mutant was shown to produce mutactin **4**, a shunt product of the actinorhodin pathway which has been aberrantly cyclized and which retains the C-9 hydroxyl group (Figure 1). Additional shunt products and biosynthetic intermediates resulting from blocks in later stages of the actinorhodin pathway have also been structurally characterized.[12] Molecular genetics ultimately led to the cloning and analysis of a cluster of genes, spanning approximately 30 kilobases (kb) in the *Streptomyces coelicolor* chromosome,[13] which code for enzymes involved in construction of actinorhodin (Figure 3). The region of the DNA coding for the *act* polyketide synthase (PKS) was identified by its ability to complement mutants that were blocked in the earliest stages of actinorhodin biosynthesis, and other regions were implicated in later steps for their complementation of other classes of mutants.[14] Subsequently, the *act* PKS genes were used to probe the DNA of a series of other polyketide producing strains of *Streptomyces*.[15] Among these were *Streptomyces violaceoruber* and *Streptomyces glaucescens*, the producers of granaticin and tetracenomycin, respectively. Using this strategy, the genes (*gra*) which specify the PKS for granaticin biosynthesis were cloned, and the PKS genes for tetracenomycin were identified within the already cloned cluster of tcm genes.[23] The corresponding PKS regions were sequenced.[16,17,18] The number and organization of the *act*, *gra* and *tcm* PKS genes are shown in Figure 3.

A common feature of the organization of the *act*, *gra* and *tcm* PKS gene clusters is a group of three characteristic open reading frames (ORFs), labeled 1, 2, 3 (Figure 3). The deduced protein sequences of the first two ORFs show particularly high similarity with the *E. coli* FabB (fatty acid synthase) condensing enzyme.[20] A potential active site Cys occurs only in the first ORF in each case. This fact, along with the possibility of translational coupling between the first two ORFs (because of overlapping 3' stop/5' start codons), suggests that the gene products may be stoichiometrically produced to form a heterodimer, rather than being two distinct condensing enzymes. The deduced protein sequence of the third ORF in each cluster resembles type II FAS acyl carrier proteins (ACPs) from bacteria and plants and ACP domains of animal type I FASs, with a characteristic motif centered on a potential 4'-phosphopantetheine-binding Ser. This result is highly significant in revealing that PKSs involved in biosynthesis of complex metabolites in *Streptomyces* utilize an ACP-mediated pathway like that of fatty acids, in contrast to plant chalcone synthases which lack an ACP requirement.[21]

The divergently transcribed *actIII* region[22] (act ORF5) would encode a ketoreductase, based on similarity to known ketoreductases and chemical evidence.[5] *gra* ORF5 and ORF6 encode deduced protein sequences which are very similar to the *actIII* gene product.[16] Interestingly, *act* ORF5 and *gra* ORF5 (but not *gra* ORF6) deduced protein sequences contain the consensus nucleotide binding Gly-X-Gly-X-X-Ala motif.[23] The *tcm* PKS cluster reveals no sequence homologous with the *act* ketoreductase, consistent with a lack of a reductive step in polyketide chain

building for this compound.[15,243]

Comparison of the deduced protein sequences of *act*, *gra* and *tcm* ORF4 with the databases provided little insight into the potential function of these genes. However, based on the previously deduced location of the *actVII* gene in the cloned *act* DNA,[14] and determination of the chemical structure of mutactin,[8] the compound produced by an *actVII* B40 mutant, it appeared likely that the *act* and *gra* ORF4 gene products were involved in both cyclization and dehydration steps in benzoisochromane quinone biosynthesis. Experiments which support this hypothesis are described.

RESULTS

Our strategy to elucidate the function of the *act* and *gra* ORF4 genes involved complementation of the mutactin-producing B40 strain of *S. coelicolor*, which contains an *actVII* mutation. The precise nature of this mutation is not known, but recent Southern analysis has shown that there are no gross deletions or rearrangements of DNA in the *actVII* region of the B40 chromosome.[25] Two types of complementation studies were designed, which included homologous DNA from *S. coelicolor*, and heterologous DNA from *S. violaceoruber*, the granaticin producer. The first involved wild-type *act* gene constructs with overlapping DNA regions which served to pinpoint the complementing DNA to the *act* ORF4 gene. Three constructs are relevant to this study (Figure 4). The plasmid pIJ2311 (ref. 14) includes a segment of DNA from the actinorhodin gene cluster

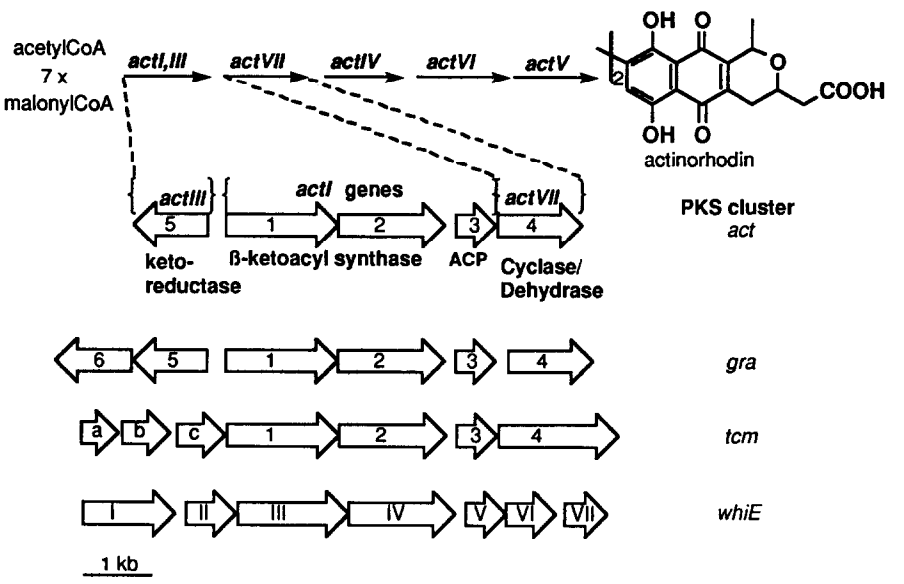

Figure 3. Number and organization of polyketide synthase (PKS) genes for *act*,[18] *gra*,[16] *tcm*,[17] and *whiE*[29] clusters. The closed arrows above denote the ordered biosynthetic sequence identified in a biochemical genetic analysis of the actinorhodin gene cluster.[11] The open arrows below represent genes which have been sequenced and identified as open reading frames (ORFs). The deduced protein sequences of ORFs 1,2 of the *act*, *gra*, *tcm* and *whiE* ORFs III,IV closely resemble the sequence of the *E. coli* FabB condensing enzyme.[6] The deduced protein sequence of ORF3 of *act*, *gra*, *tcm* and ORFV of *whiE* closely resemble fatty acid synthase acyl carrier proteins.[6] The deduced protein sequences of *act*, *gra*, *tcm* ORF4 and *whiE* ORFVI are compared and described in the text. The deduced protein sequences of *actIII* (ORF5) and *gra* ORF5,6 show high similarity to known oxidoreductases.[6] The functions of *tcm* ORFs a,b,c[28] and *whiE* ORFs I,II,VI,VII (ref. 29) have not been established.

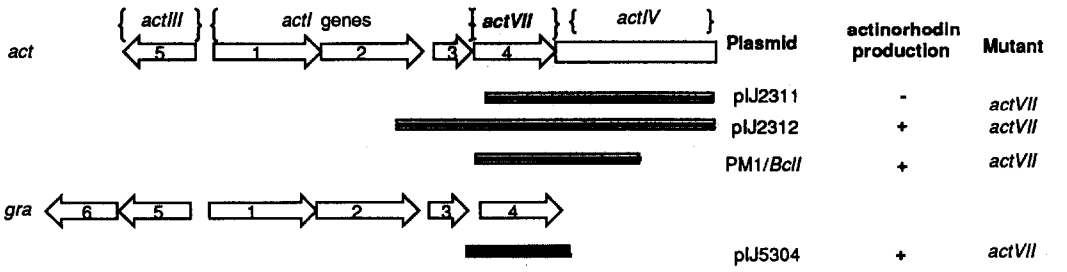

Figure 4. Gene constructs used for complementation of the *Streptomyces coelicolor* B40 *actVII* mutant. Open arrows denote sequenced genes that have been identified as open reading frames (ORFs). The plasmids pIJ2311 and pIJ2312, and the PM1/*BclI* phage construct contain segments of the *act* gene cluster (shown in the half-filled rectangles). The plasmid pIJ5304 contains *gra* ORF4 DNA as shown in the filled rectangle. Protoplasts of the *S. coelicolor* B40 mutant were transformed or transfected, and production of actinorhodin was determined (see Experimental section).

whose 5' end begins within the ORF4 gene; thus the plasmid lacks the amino-terminus of the *act* ORF4 gene product. This plasmid did not complement the B40 mutant strain of *S. coelicolor*.[14] A second plasmid, pIJ2312 (ref. 14), includes a longer segment of DNA whose 5' end falls in the *act* ORF2 gene, and which therefore includes the whole of ORF4. This plasmid, when introduced into the B40 mutant caused the production of actinorhodin.[14] The third construct included a *BclI* fragment from the *act* cluster, cloned into the PM1 phage vector,[14] having a 5' terminus about 200 bp upstream of the 5' terminus of the pIJ2311 *act* DNA insert. Some of the lysogens derived from infection of the B40 *actVII* mutant produced high levels of actinorhodin. This series of experiments showed that the beginning of the *BclI* fragment contains DNA which is required for complementation of the B40 mutant and that complementation is possible when the *actVII* gene is restored by a single crossover between the wild-type insert of the PM1 recombinant phage and the mutated *actVII* gene in the B40 chromosome (complementation would occur if crossover takes place before the position where the *actVII* mutation lies). This is the position occupied by *act* ORF4.

Based on the structure of granaticin,[7] we would predict that the early chemical steps leading to its biosynthesis are identical to those for actinorhodin, including the cyclization and dehydration reactions required to form the benzoisochromane quinone ring system. In order to test whether *gra* ORF4 was able to replace the homologous *act* ORF4 gene, a plasmid was constructed (pIJ5304) which contained *gra* ORF4 cloned into the single copy *Streptomyces* vector pIJ941 (ref 25, Figure 4). Transformation of *S. coelicolor* mutant B40 with pIJ5304 resulted in production of significant levels of actinorhodin, showing that *gra* ORF4 alone was sufficient to relieve the block in the B40 mutant strain. Transformation with pIJ941 alone gave no detectable production of actinorhodin. As a further control, transformation of an *actI* mutant strain (an early, PKS blocked mutant) with pIJ5304 gave no detectable production of actinorhodin. These results lead to the conclusion that *in vivo*, dehydration at C-9 and internal aldol cyclization between C-5/C-14 can be carried out by a single bifunctional enzyme encoded by ORF4 of either the *act* or the *gra* gene cluster.

Comparison of the *gra* and *act* ORF4 deduced proteins revealed a high degree of similarity over the entire length of the sequences (overall amino acid identity of 54.2%, similarity of 67.6%). This is most clearly shown in a COMPARE/DOTPLOT analysis (Figure 5a). Two additional homologues of the *act* ORF4 and *gra* ORF4 genes have been sequenced and they provide important comparisons. The first is the *tcmIa* ORF4 gene from the tetracenomycin biosynthetic gene cluster.[19] A total of four intramolecular aldol condensations would be required to construct the ring system of this molecule. Interestingly, COMPARE/DOTPLOT analysis of the deduced protein from the *tcm* ORF4 gene with *act* ORF4 reveals significant protein similarity over the amino-terminal region only (Figure 5b). Meanwhile a comparison of the *tcmIa* ORF4 deduced protein

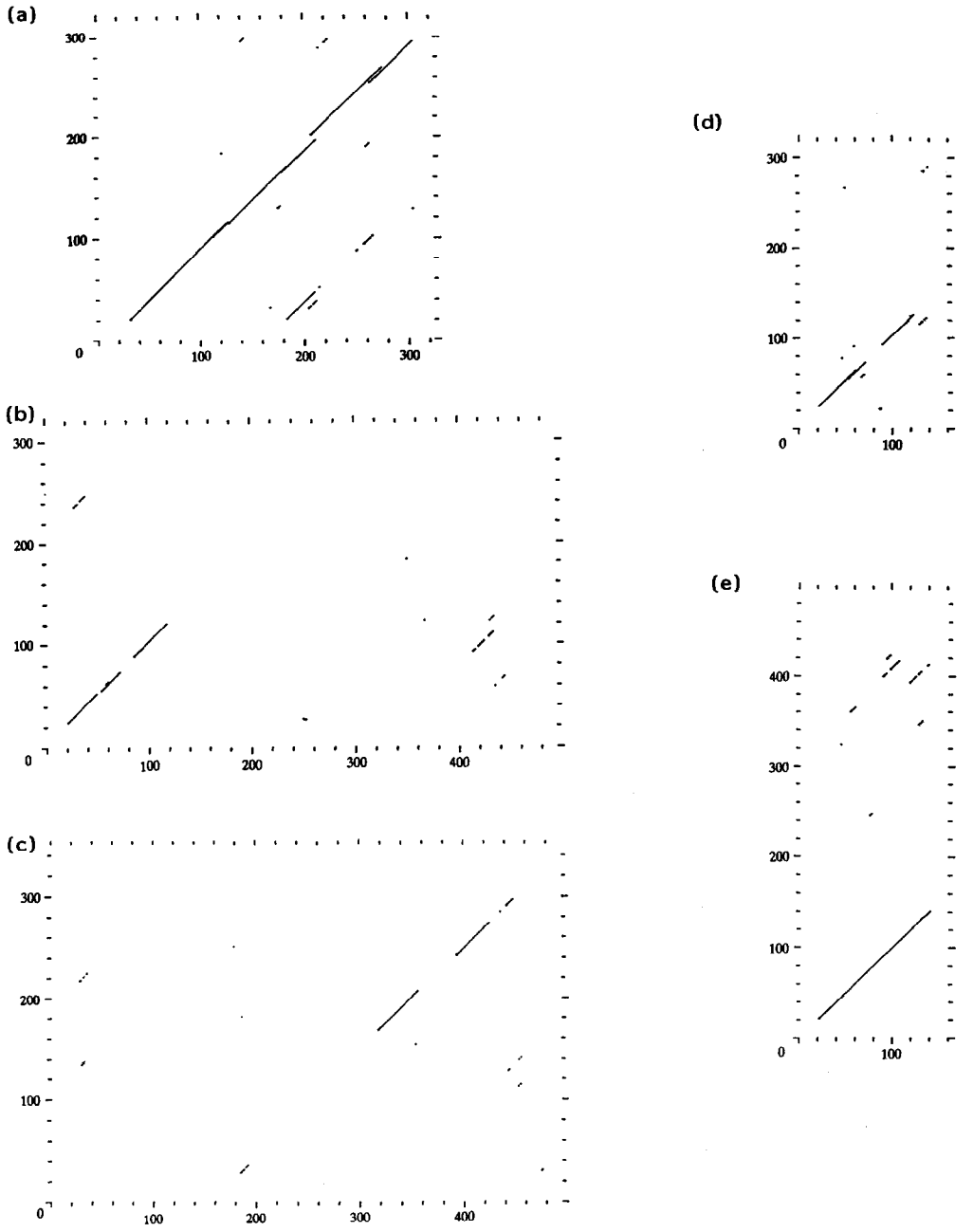

Figure 5. COMPARE/DOTPLOT comparisons of *act* ORF4, *gra* ORF4 (Ref. 16), *tcm* ORF4 and *whiE* ORFVI[29] proteins. (a) *act* ORF4 (vertical) and *gra* ORF4; (b) *act* ORF4 and *tcm* ORF4; (c) bovine hydroxyindole O-methyl transferase[27] (bovmet) and *tcm* ORF4; (d) *act* ORF4 and *whiE* ORFVI; (e) *tcm* ORF4 and *whiE* ORFVI. For an explanation of stringency, see the *Computer analysis* section of the Experimental section.

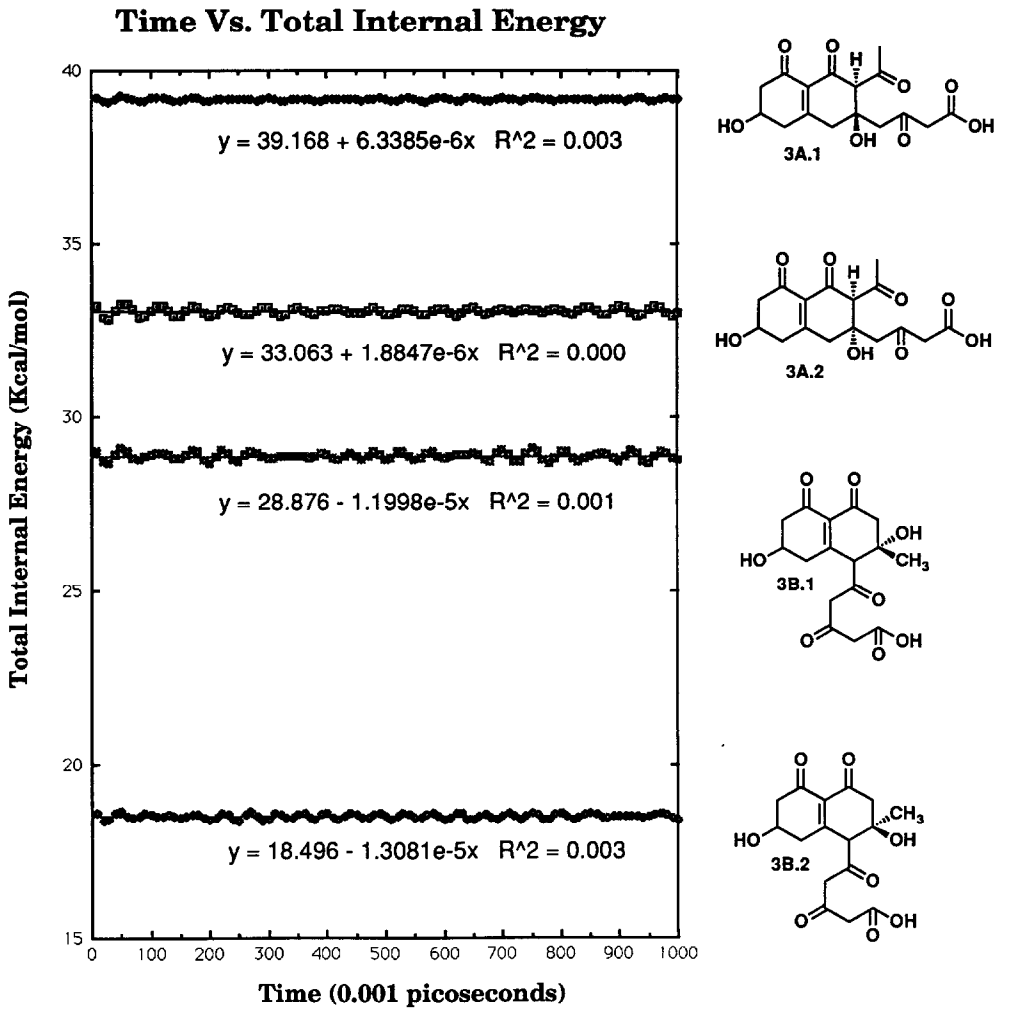

Figure 6. Plot of time vs total energy for the presumed primary cyclization products from intermediate 3A and 3B (Fig. 1) using energy minimization and detailed dynamics (see Experimental section for details).

sequence with the GenBank/EMBL database revealed a high degree of similarity to a bovine hydroxyindole O-methyl transferase,[27] but only over the carboxy terminus of the *tcmIa* ORF4 product (Figure 5c). The construction of the *tcm* polyketide (**11**, Fig. 7A), which is derived from one acetylCoA starter and nine malonylCoA extender units, does not require a ketoreduction or dehydration step. However, tetracenomycin C does contain two methyl ethers (one of them at the asterisked position equivalent to the hydroxyl that is removed by dehydration during the cyclization of the actinorhodin and granaticin polyketide precursors), which is consistent with the presence of a potential O-methyl transferase domain on the *tcmIa* ORF4-encoded protein. These data suggest that the amino terminal domain of the *act*, *gra*, and *tcmIa* ORF4 gene products contains intramolecular aldolase activity. The *whiE* locus of *Streptomyces coelicolor* has recently been found to consist of a series of ORFs, several of which clearly correspond to the *act*, *gra*, and

tcm PKS ORFs (Figure 3), suggesting that the brownish spore pigment lacking in *whiE* mutants is polyketide-derived.[29] There is no equivalent of the *act* and *gra* ORF5 (or *gra* ORF6) ketoreductase genes, but ORFs closely resembling in deduced protein sequence and arrangement, the *act*, *gra* and *tcmIa* ORFs 1, 2 and 3 are present (called *whiE* ORF III, IV, V). Interestingly, the next ORF to the right, *whiE* ORFVI, is a homologue of the *act*, *gra*, and *tcmIa* ORF4 genes. However, the deduced protein sequence is only half the length of the other three, and reveals significant similarity over its entire length with the amino-terminal domains of each of the other ORFs (Figure 5d, 5e).[27]

In vitro, substituent effects played a key role in directing the outcome of internal aldol cyclizations in oligoketide systems (Figure 2, see above). However, analysis of the enzymatic role played by an aldolase in formation of the primary cyclization products derived from **3A** and **3B** (Figure 1) will not be possible until the *act* ORF4 encoded protein has been overexpressed and tested on synthetic substrates. Until then, it is instructive to consider the implications of the molecular genetic results described in this study. First, the ability of the *gra* ORF4 gene to complement a corresponding mutant gene in the *act* pathway strongly suggests that the encoded protein provides either a direct C-14/C-5 bond-forming catalytic function, or provides scafolding which stabilizes the conformation of the nascent polyketide chain to insure formation of a C-14/C-5 bond. Second, complementation results obtained with the *act* PM1/*Bcl*1 phage construct suggests that the mutation in *act* ORF4 occurs at the 5'-end of the gene. Although the B40 *actVII* gene has not been sequenced, we expect to find a missense, frameshift, or nonsense mutation. Enzyme inactivation would result from the last two possibilities, whereas a missense mutation could result in formation of a gene product with altered catalytic properties. In order to address the possibility that internal aldol condensation between C-6/C-15 occurs spontaneously (which would suggest gene inactivation, or destruction of the normal catalytic potential of the *actVII* mutant enzyme), we calculated the internal total energies of the four initial intermediates which would result from cyclization of **3A** and **3B** (Figure 1) using energy minimization and detailed dynamics[30] (see Experimental section for details). Figure 6 represents the results of molecular modelling for the two stereoisomers (**3A.1** and **3A.2**) derived from C-14/C-5 cyclization of **3A**, and the two stereoisomers (**3B.1** and **3B.2**) derived from C-6/C-15 cyclization of **3B**. These data show that the intermediates of lowest internal energy (**3B.1** and **3B.2**) are precursors of mutactin. Specifically, **3B.2** is 14.5 Kcal lower in energy than **3A.2**, the likely cyclization product in the actinorhodin pathway. This finding is consistent with the idea that in the *S. coelicolor* B40 mutant strain, mutactin arises (after carbon chain construction) from a series of spontaneous chemical events beginning with formation of the C-6/C-15 bond.

DISCUSSION

The *in vivo* basis of regiochemical control of intramolecular aldol reactions is a primary goal in efforts to further our understanding of polyketide systems. Chemical studies of polyketide cyclizations have already provided an important model for catalytic reactions mediated by biosynthetic enzymes. Although our focus in this report has been on the second cyclization step in formation of the benzoisochromane quinone system (Fig. 1), the initial cyclization step is equally critical in determining the ultimate structure of the molecule. An interesting comparison involves regioselective aldol cyclization in heptaketide systems with bulky protecting groups at C-8 (Fig. 2, Scheme 2,3). In the biosynthetic formation of an octaketide precursor in benzoisochromane quinones, ketoreduction at the C-9 position (Fig. 1) may play a role in directing an intramolecular aldol reaction between C-12/C-7 (Fig. 1, **3A**). It is conceivable that steric effects (induced by the position of the oligoketide chain at the enzyme active-site) during this reduction step contribute to the regiospecific formation of the C-12/C-7 bond.

The intramolecular aldol and Dieckmann ring closures leading from **11** to tetracenomycin D3 (**13**, Fig. 7A) provide another example where a PKS directs the regiochemistry of cyclization, in this

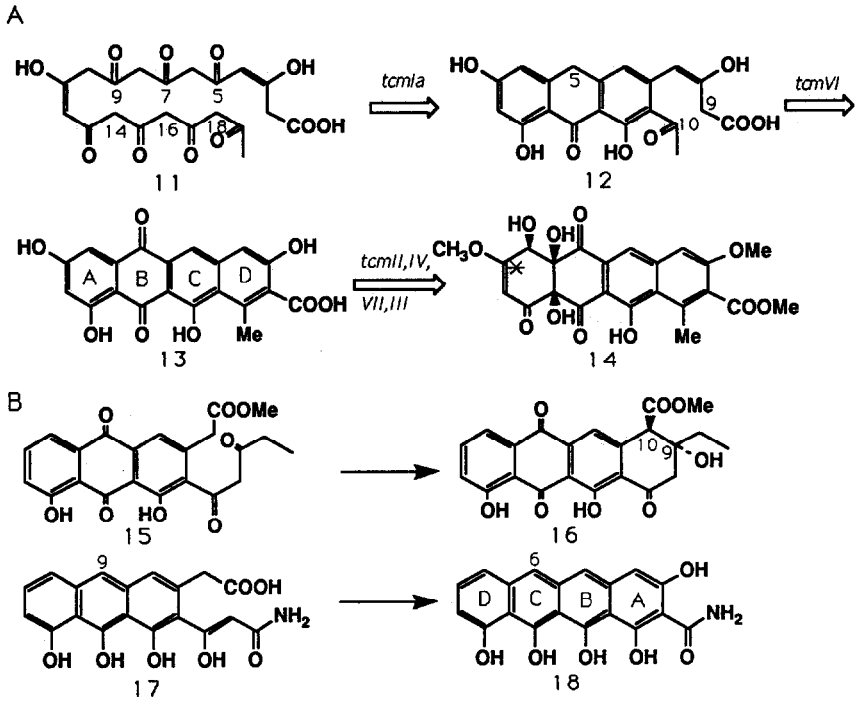

Figure 7. Cyclization pattern of decaketides. **(A)** shows biosynthetic steps in the *tcm* pathway. **(B)** shows cyclizations leading to aklaviketone **16** and pretetramide **18**.

case to form the naphthacene ring system rather than the benzo[a]anthraquinone alternative found in antibiotics like tetrangomycin,[31] urdamycin A,[32] and the precursor of kinamycin.[33] Chemical studies have explored the use of a central sp^3 carbon atom in polycarbonyl intermediates **9** (Fig. 2) and **19** (Fig. 8A) to provide conformational flexibility which allows the chain to fold back on itself. It is clear from this work[34] that intramolecular aldol cyclizations which involve the ß-ketomethylene groups of **9** or **19** compete favorably with those involving terminal methyl groups activated by only single keto groups, resulting in hydroxynaphthalenes **10** and **21** (and subsequently **8** and **22**). The possibility was suggested[34] that reaction between the C-5 ß-ketomethylene and the more electrophilic C-10 carbonyl creates the first six membered ring, (giving, e.g. **20**, Fig. 8A). This is followed by preferred closure of the C-3/C-12 bond because this aldolization involves the less crowded, more reactive C-12 carbonyl (it is flanked by two methylenes and only one ß-keto group) with subsequent aromatization of ring A by elimination of the C-8 substituents, creating the naphthalene instead of the isomeric tetralone ring system.

The preceeding rationale can be extended to the biosynthesis of actinorhodin, granaticin and tetracenomycin by invoking as intermediates, **3A** (Fig. 1) and the decaketide analog of **3A** formed from **11** by an aldol reaction between C-9 and C-14 (not shown in Fig 7A). It is unclear whether ring A or B would be aromatized first in the latter case, although the metabolite accumulated by an *S. coelicolor actIV* mutant[7] suggests that it should be ring A. Formation of mutactin (**4**, Fig. 1) via **3B** contradicts the above rationale, since cyclization involves the less reactive C-6 methylene and C-15 ß-keto groups. Our modelling studies (Fig. 6) suggest that a significant factor in determining the regiochemical course of intramolecular aldol cyclization could be the internal energy of the product. In this case, formation of the precursor leading to mutactin would be highly favored over the precursor leading to actinorhodin.

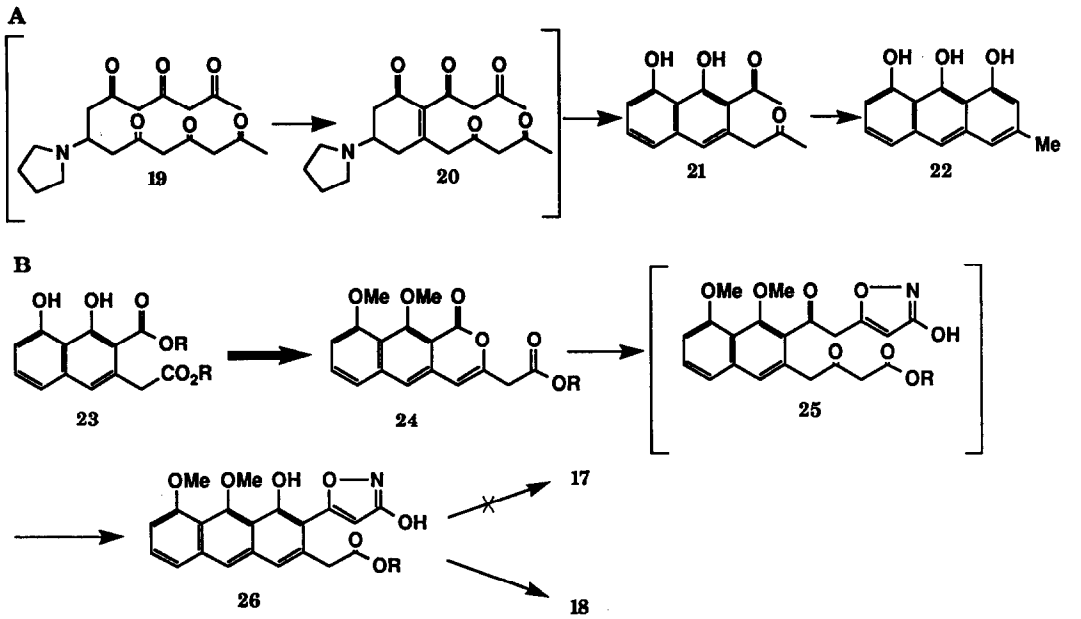

Figure 8. **(A)** shows the *in vitro* cyclization of a heptaketide containing a bulky substitutent at C-8. **(B)** shows the *in vitro* cyclization of a decaketide system.

Chemical studies by Harris and his colleagues involving biomimetic syntheses of naphthacene natural products have led to pretetramide[34,35] and 6-methylpretetramide,[36] intermediates in tetracycline biosynthesis.[37] This work provides an important precedent for the decaketide cyclization steps presumed to occur in tetracenomycin biosynthesis. Synthesis of hydroxynaphthalene **23** (Fig. 8B) was followed by formation of **24** which was subsequently reacted with the dilithium salt of 3-hydroxy-5-methylisoxazole. This transformation resulted in spontaneous cyclization to form ring B of **26** (ref. 34). Deprotection of **26** gave pretetramide **18** directly rather than 9-deoxyprotetrone (**17**, Fig. 7B), which had been isolated previously as a shunt product (in the form of the naphthacenequinone protetrone) from a non-tetracycline producing blocked mutant of *S. aureofaciens*.[38] The 52% overall yield of **18** from **24** suggests that the biological processes by which rings B and A are formed in tetracycline and rings C and D in tetracenomycin biosynthesis should be regiospecific as well as efficient. Ring formation also occurs stepwise *in vivo* as supported by the isolation of protetrone[38] and 9-methylprotetrone[39] from *S. aureofaciens* blocked mutants, and of tetracenomycin F2 (**12**, Fig. 7A) from *S. glaucescens tcmVI* mutants.[40] The conversion of **12** to **13** and thence to **14** observed in cosynthesis experiments between *tcmIa* and *tcmVI* mutants[40,41] confirms that tetracenomycin F2 **12** is a true intermediate. Significantly, protetrone and 9-methylprotetrone could not be converted to tetracyclines by *S. aureofaciens* since they are shunt products formed from the natural anthrone precursors by air oxidation.[38,39] Cyclization of methyl aklanonate (**15**, Fig. 7B) to aklaviketone **16**, a step in the biosynthesis of daunorubicin by *Streptomyces* species,[42] is a further example of stepwise ring formation among decaketide metabolites.

Why was spontaneous aromatization of ring A in the precursor of mutactin (Fig. 1) and in the corresponding intermediate lying between **9** and **10** (ref. 10, not shown in Fig. 2) not seen? The formation of compounds like these and the one accumulated by an *actIV* mutant[7] (Fig. 1), implies that the PKS cyclases could have distinct aldolase and dehydrase activities, perhaps as different domains of the enzyme (but not necessarily a different set of domains for each cyclization and dehydration event). For the conversion of **9** to **10**, we can assume that elimination of one of the

hydroxymethylenes of the C-8 ketal does not occur under neutral conditions. In the case of mutactin, the pH of the cellular milieu may not favor the facile acid catalyzed loss of the C-9 hydroxyl from **3A/3B** (Fig. 1), which could be further disfavored by a half-chair conformation of ring A that does not allow a trans, antiperiplanar orientation of one of the C-10 hydrogens and the C-9 hydroxyl. Ring A of **3A/3B** is similar to the A ring of aklaviketone **16**; both have nearly identical conformations. A stereochemical argument clearly can be used to explain the comparative stability of **16** to spontaneous aromatization of ring A as due to the cis arrangement of the C-9 hydroxyl and C-10 hydrogen. On the other hand, the ^{1}H NMR spectrum for **16** indicates that one of the C-8 hydrogens has a trans, diaxial relationship with the C-9 hydroxyl.[43] Whatever the reason for the apparent stability of mutactin precursors, support for distinct aldolase and dehydrase steps in the cyclization of octa- and decapolyketides can be found in studies of the formation of dehydroquinate and dehydroshikimate by bacteria, which occurs by the stepwise involvement of dehydroquinate synthase (equivalent to the aldolases discussed here) and 5-dehydroquinate dehydrase.[44]

EXPERIMENTAL

Bacterial strains, plasmids and phages

Streptomyces lividans TK64 (ref. 45) was the host for plasmids derived from the low copy number vector pIJ941 (ref. 26) which carry the *gra* cluster gene segment. pIJ5304 was constructed from an exoIII deletion derivative of pIJ5200 (ref. 16). The derivative, pIJ5315, contains the entire *gra* ORF4 gene including all of the non-coding DNA between *gra* ORF3 and *gra* ORF4 (ref. 16). The *gra* ORF4 gene segment was excised from pIJ5315 by digestion with *Hin*dIII/*Eco*RI and the ends blunted and cloned into the *Eco*RV site of pIJ941. Plasmids pIJ2311 and pIJ2312 were derived as described previously.[14] The lysogen carrying the *act* (early and intermediate region) *Bcl*I fragment in PM1 (ref. 14) was constructed by replacing the thiostrepton resistance gene from PM1 with the *Bcl*I fragment. *E. coli* strain *lacZ* DM15 *recA* [46] was used as the host for plasmids derived from pUC18 (ref. 47). *E. coli* JM101 was used as the host for phages M13mp18 and M13mp19 (ref. 48).

Standard media and manipulations

Procedures for growth and manipulation of *Streptomyces* and for general recombinant DNA manipulations were as described.[45] *E. coli* was transformed by the procedure of Hanahan[49] and transformants were selected on L agar[50] containing 100 mg/ml carbenicillin. Procedures for the growth and manipulation of M13 were as described by Messing.[51]

DNA isolation

Streptomyces plasmid DNA was prepared as described by Kieser.[52] The alkaline lysis method[53] was used to prepare plasmids from *E. coli*. Small-scale preparations of *E. coli* plasmids and replicative form M13 DNA were produced using a modification of the method of Ish-Horowitz and Burke.[54]

Computer analysis

Comparison of deduced amino acid sequences (Fig. 5a - e) was performed using the UWGCG programs COMPARE and DOTPLOT. COMPARE takes a segment of protein 1, the size of which is specified (window) and compares it to each window of protein 2, in every register. When the number of similarities is above a specified threshold (stringency) the region is noted. DOTPLOT takes the information from COMPARE and represents it in graphic form such that regions where the similarity was above the threshold are represented by a dot positioned at the central point of the window. For the comparisons described here, a window of 40 and a stringency of 19 was used with COMPARE in version 5.3 of the UWGCG package.[55] Amino acid sequence comparisons were

made with the GenBank/EMBL database using the program FASTA.[55] The *act*ORF4 DNA sequence has been submitted to the GenBank database.

Molecular modelling

The calculation of total internal energies of the two intermediate molecules, and their stereoisomers, in the proposed biosynthetic route to actinorhodin and mutactin was performed using the CHARMM algorithm,[30] as implemented in the program Quanta (Polygen Corp.). To obtain the total internal energies four steps were required: minimization, heating, equilibration, and detailed dynamics.[30]

Energy minimization of each structure was performed using the local harmonic approximations of the adopted-basis set Newton-Raphson (ABNR) minimization.[30] Approximately a 1000 steps were required to minimize each structure to produce values of 0.3755 kcal/mol, 4.9373 kcal/mol for the actinorhodin intermediates (**3A.1** and **3A.2**, respectively) and values of -0.3849 kcal/mol and -12.6570 kcal/mol for the mutactin intermediates (**3B.1** and **3B.2**, respectively). Values of 0.001 and 0.000 for energy value tolerance and step value tolerance, respectively, were used as criterion for successful minimization.

Minimization of each of the intermediates was followed by a 4020 step (0.0002 picosecond time step) heating period in which the minimized structures were heated to 300.0°K. In the equilibration period, random Gaussian velocity assignments were made with a mean temperature of 300° K. 4000 steps, with a 0.0002 picosecond time step equilibration was sufficient to reach a convergence allowing the various statistical properties to be independent of time.

Detailed dynamics was performed using a simple second-order predictor two-step method developed by Verlet.[56] A 2 x 10^4 step (0.0002 picoseconds per steps) detailed dynamics was performed on each of the stereoisomers. The total internal energies of the 40 picosecond simulation were determined by using the y-intercept of a linear fit of the detailed dynamics over the one picosecond simulation (Fig. 6). The total internal energy of the **3A.1** and **3A.2** isomers of the actinorhodin intermediates was 33.063 kcal/mol and 39.168 kcal/mol, respectively, and for the **3B.1** and **3B.2** isomers of the mutactin, 18.496 kcal/mol and 28.876 kca/mol respectively.

Isolation of Actinorhodin

Purification of actinorhodin and actinorhodin dimethyl ester tetraacetate was performed as described previously.[57] Characterization of the compounds was confirmed by comparison of their physical properties with published data.[57]

Acknowledgements

We thank Scott Rychnovsky for discussions on molecular modelling and Keith Chater and Chaitan Khosla for careful reading of the manuscript. This work was supported by the Agricultural and Food Research Council and the John Innes Foundation through grants-in-aid to the John Innes Institute, and NIH grant GM39784-02.

REFERENCES

1. Collie, J.N. *J. Chem Soc.* **1907**, *91*, 1806-1813.
2. Birch, A.J. *Science* **1967**, *156*, 202-206.
3. Harris, T.M., Harris, C.M. *Pure & Appl. Chem.* **1986**, *58*, 284-294.
4. Brockman, H., Zeeck, A., van der Merve, K., Muller, W. *Justus Liebigs Annln. Chem.* **1966**, *698*, 3575-3579.
5. Bartel, P.L., Zhu C.-B., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.M., Floss, H.G. *J. Bacteriol.* **1990**, *172*, 4816-4826.
6. Hopwood, D.A., Sherman, D.H. *Ann. Rev. Gen.* **1990**, *24*, 37-66.
7. Floss, H.G., Cole, S.P., He, X.G., Rudd, B.A.M., Duncan, J., Fujii, I., Chang, C.-J., Keller, P.J. In Kleinkauf, H., von Dohren, D., Dornauer, H. and Nesemann, G. (eds), *Regulation of Sec-*

ondary Metabolite Formation. **1986**, 283-304, VCH, Weinheim
8. Zang, H.-l., He, X.-g., Adefarati, A., Gallucci, J., Cole, S.P., Beale, J.M., Keller, P.J., Chang, C.-j., Floss, H.G. *J. Org. Chem.* **1990**, *55*, 1682-1684.
9. Harris, T.M., Wittek, P.J. *J. Am. Chem. Soc.* **1975**, *97*, 3270-3271.
10. Harris, T.M., Webb, A.D., Harris, C.M., Wittek, P.J., Murray, T.P. *J. Am. Chem. Soc.* **1976**, *98*, 6065-6067.
11. Rudd, B.A.M., Hopwood, D.A. *J. Gen. Microbiol.* **1979**, *114*, 35-43.
12. Cole, S.P., Rudd, B.A.M., Hopwood, D.A., Chang, C.-J., Floss, H.G. *J. Antiobiotics*, **1987**, *40*, 340-347.
13. Malpartida, F., Hopwood, D.A. *Nature* **1984**, *309*, 462-464.
14. Malpartida, F., Hopwood, D.A. *Mol. Gen. Genet.* **1986**, *205*, 66-73.
15. 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. *Nature* **1987**, *325*, 818-821.
16. Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.M., Bibb, M.J. Hopwood, D.A. *EMBO J.* **1989**, *8*, 2717-2725.
17. Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F., Hutchinson, C.R. *EMBO J.* **1989**, *8*, 2727-2736.
18. Miguel Fernandez-Moreno et al., unpublished.
19. E. Wendt-Pienkowski, R. Summers, C.R. Hutchinson, unpublished.
20. Kauppinen, S., Siggaard-Anderson, M., von Wettstein-Knowles, P. *Carlsberg Res. Commun.* **1988**, *53*, 357-370.
21 Schuz, R., Heller, W., Hahlbrock, K. *J. Biol. Chem.* **1983**, *258*, 6730-6734.
22. Hallam. S.E., Malpartida, F. Hopwood, D.A. *Gene* **1988**, *74*, 305-320.
23. Scrutton, N.S., Berry, A., Perham, R.N. *Nature* **1990**, *343*, 38-43.
24. Motamedi, H., Hutchinson, C.R. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4445-4449.
25. Kim, E.-S., Sherman, D.H., unpublished.
26. Lydiate, D.J., Malpartida, F., Hopwood, D.A. *Gene* **1985**, *35*, 223-235.
27. Ishida, I., Obinata, M. Deguchi, T. *J. Biol. Chem.* **1987**, *262*, 2895-2899.
28. H. Motamedi, R. Rubin, C. R. Hutchinson, unpublished
29. Davis, N.K., Chater, K.F. *Mol. Microbiol.* **1990**, *4*, 1679-1691.
30. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187-217.
31. Kunstmann, M.P., Mitscher, L.A. *J. Org. Chem.* **1966**, *31*, 2920-2925.
32. Drautz, H., Zahner, H., Rohr, J., Zeeck A. *J. Antibiotics* **1986**, *39*, 1657-1669.
33. Seaton, P. J., Gould, S. J. *J. Am. Chem. Soc.* **1987**, *109*, 5282-5284.
34. Gilbreath, S. G., Harris, C. M., Harris, T. M. *J. Am. Chem. Soc.* **1988**, *110*, 6172-6179.
35. Harris, T. M., Harris, C. M., Oster, T. A., Brown, Jr., L. E., Lee, J. Y.-C. *J. Am. Chem. Soc.* **1988**, *110*, 6180-6186.
36. Harris, T. M., Harris, C. M., Kuzma, P. C., Lee, J Y.-C., Mahalingam, S., Gilbreath, S. G. *J. Am. Chem. Soc.* **1988**, *110*, 6186-6192.
37. McCormick, J. R. D., Reichenthal, J., Johnson, S., Sjolander, N. O. *J. Am. Chem. Soc.* **1963**, *85*, 1692-1695.
38. McCormick, J. R. D., Jensen, E. R. *J. Am. Chem. Soc.* **1968**, *90*, 7126-7127.
39. McCormick, J. R. D., Jensen, E. R., Arnold, N. H., Corey, H. S., Joachim, U. H., Johnson, S., Miller, P. A., Sjolander, N. O. *J. Am. Chem. Soc.* **1968**, *90*, 7127-7129.
40. Nakayama, H., Hutchinson, C. R., unpublished results.
41. Motamedi, H, Wendt-Pienkowski, E., Hutchinson, C. R. *J. Bacteriol.* **1986**, *167*, 575-580.
42. Connors, N. C., Bartel, P. A., Strohl, W. R. *J. Gen. Microbiol.* **1990**, *136*, 1887-1894.
43. Eckardt, K., Schumann, G., Tresselt, D., Ihn, W. *J. Antibiotics* **1988**, *41*, 788-793.
44. Walsh, C. *Enzymatic Reaction Mechanisms*, **1979**, 923-924, W. H. Freeman, San Francisco.
45. Hopwood, D.A., Bibb, J.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J.,

Smith, C.P., Ward, J.M., Schrempf, H. *Genetic Manipulation of Streptomyces. A Laboratory Manual.* **1985**. John Innes Foundation, Norwich, U.K.
46. Ruther, U., Koenen, M., Oatto, K., Muller-Hill, B. *Nucleic Acids Res.*, **1981**, *9*, 4087-4098.
47. Norrander, J., Kempe, T., Messing, J. *Gene* **1983**, *26*,101-106.
48. Yanisch-Perron, C., Vieira, J., Messing, J. *Gene* **1985**, *33*, 103-119.
49. Hanahan, D. *J. Mol. Biol.* **1983**, *166*, 557-580.
50. Miller, J. *Experiments in Molecular Genetics.* **1972**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
51. Messing, J. *Methods in Enzymology* **1983**, *101*, 20-28, Wu, R., Grossman, L, Moldave, K. (eds), Academic Press, NY.
52. Kieser, T. *Plasmid* **1984**, *12*, 19-36.
53. Maniatis, T., Fritsch, E.F., Sambrook, J. *Molecular Cloning. A Laboratory Manual.* **1982**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
54. Ish-Horowitz, D., Burke, J.F. *Nucleic Acids Res.* **1981**, *9*, 2989-2998.
55. Devereaux, J., Haeberli, P., Smithies, O. *Nucleic Acids Res.* **1984**, *12*, 387-395.
56. Verlet, L., *Phys. Rev.*, **1967**, *159*, 98-103.
57. Gorst-Allman, C.P., Rudd. B.A.M., Chang, C.-j., Floss, H.G. *J. Org. Chem.* **1981**, *46*, 455-456.