The EMBO Journal vol.8 no.9 pp.2717 - 2725, 1989

Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22

David H.Sherman, Francisco Malpartida[1], Maureen J.Bibb, Helen M.Kieser, Mervyn J.Bibb and David A.Hopwood

John Innes Institute and AFRC Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH, UK

[1]Present address: Instituto de Investigaciones Biomedicas, Fac. Medicina, Universidad Autonoma de Madrid, c/Arzobispo Morcillo 4, 28029 Madrid, Spain

Communicated by D.Hopwood

A 6.5 kb region of DNA from *Streptomyces violaceoruber*, which contains polyketide synthase (PKS) genes for production of the benzoisochromane quinone moiety of the antibiotic, granaticin, was cloned and sequenced. Of six open reading frames (ORFs) identified, four (ORFs 1-4) would be transcribed in one direction and two (ORFs 5 and 6) divergently from ORFs 1-4. ORF1 and ORF2, which show evidence for translation coupling, encode (deduced) gene products which strongly resemble each other and the *Escherichia coli* fatty acid ketoacyl synthase (condensing enzyme), FabB. We conclude that ORF1 (which contains a characteristic cysteine residue) functions as a condensing enzyme, possibly as part of a heterodimeric protein including the product of ORF2. The predicted ORF3 gene product strikingly resembles acyl carrier proteins (ACPs) of fatty acid synthase (FAS), particularly in the region of the active site motif, while the predicted ORF5 and ORF6 gene products resemble known oxidoreductases, suggesting that they function at reductive steps required during assembly of the granaticin carbon skeleton. Comparison of the deduced ORF4 gene product with available protein databases failed to elucidate its potential function. The overall conclusion is that the granaticin-producing PKS would consist of at least six separate enzymes involved in carbon chain assembly, thus resembling a Type II, rather than a Type I, FAS.

Key words: antibiotic biosynthesis/fatty acid biosynthesis/granaticin/polyketide synthase/*Streptomyces violaceoruber* Tü22

Introduction

Polyketides make up a large family of structurally varied and complex secondary metabolites produced by bacteria (mostly actinomycetes), fungi and plants. Many have proven useful in human and veterinary medicine and in animal husbandry as effective antibiotics, chemotherapeutic and growth promoting agents. Others are important as fungal aflatoxins or as plant pigments and flavour compounds.

The biosynthesis of polyketides appears to be (at least conceptually) analogous to that of long-chain fatty acids (Birch and Donovan, 1953; Lynen, 1980). The biosynthesis of such fatty acids proceeds by an iterated cycle of reactions in which two-carbon units (acetate residues), derived from malonyl CoA, are joined by a condensation, with loss of CO_2, to form an even-numbered chain (for example of 16 carbon atoms, requiring seven successive condensations, in the case of palmitic acid). Since the β-carbon of each malonate unit carries an oxygen atom, which is not found in the final product, the oxygen has to be removed. This occurs by a set of three reactions after each two-carbon addition: reduction to a hydroxyl, dehydration to remove oxygen with introduction of a double bond in the carbon chain, and further reduction to saturate the system (Figure 1). The four reactions—condensation, β-ketoacyl reduction, dehydration and enoyl reduction—represent only part of the repertoire of the fatty acid synthase (FAS). In addition, acyl transferase reactions are required to attach the acetate unit which constitutes the 'starter' of the carbon chain (acetyl transferase), and the malonate units which provide the successive 'extenders' for its elongation (malonyl transferase), to appropriate sites on the FAS, together with a palmityl transferase to detach the mature fatty acid from the FAS. The acetyl and malonyl transferases attach the acyl building units, not directly to the FAS, but rather to a phosphopantetheine 'arm' attached to an active site serine on an acyl carrier protein (ACP) that forms part of the FAS; successive malonate units are received by this 'arm' and then transferred to a cysteinyl residue in the condensing enzyme for addition to the growing fatty acid chain.

A typical FAS is therefore a multivalent system involving eight functional units—acetyl, malonyl and palmityl transferases, ACP, ketoacyl synthase (condensing enzyme), ketoacyl reductase, dehydratase and enoyl reductase. The organization of these functional units varies in different life forms (McCarthy and Hardie, 1984; Wakil, 1986). In most bacteria, and in plant plastids (which are themselves probably of prokaryotic origin), they occur as eight separate polypeptides (the products of only distantly linked genes in *Escherichia coli*) which come together to form a multi-enzyme complex (Type II FAS). In vertebrates, in contrast, in which fatty acid synthesis occurs in the cytosol rather than in organelles as in plants, the same eight functions are carried out by domains on a single, multifunctional, polypeptide (Type I FAS). The yeast *Saccharomyces cerevisiae* represents an intermediate case in which the FAS consists of two classes of polypeptide, components 1 and 2, carrying five and three functions, respectively (Schweizer *et al.*, 1986, 1987; Chirala *et al.*, 1987; Mohamed *et al.*, 1988). [When the FAS consists of one or two multifunctional polypeptides there may be fewer than three acyl transferases: thus in vertebrates acetyl and malonyl transferase reactions are catalysed by a single active site, and in yeast a single malonyl-palmityl transferase is found (McCarthy and Hardie, 1984).]

Despite the formal similarity in the chemistry of biosynthesis of fatty acids and of the primary carbon

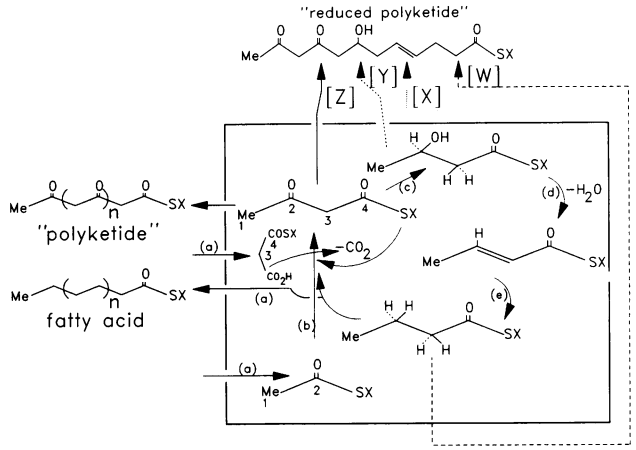


Fig. 1. General scheme for the assembly of the carbon chains of fatty acids, 'polyketides' and 'reduced polyketides'. The enzyme activities that make up the fatty acid synthase (represented as a box) are as follows: (**a**) acetyl, malonyl and palmityl transferases; (**b**) ketoacyl synthase (condensing enzyme); (**c**) keto reductase; (**d**) dehydratase; (**e**) enoyl reductase. (Note: the structures labelled 'polyketide' and 'reduced polyketide' would not exist in the forms illustrated but would be stabilized by further chemical changes—such as ring formation—after the completion of carbon chain assembly.) Possible polyketide functional group substitutions: [W], alkyl functionality; [X], enoyl functionality; [Y], hydroxyl functionality; [Z], keto functionality. After Sherman *et al.* (1988).

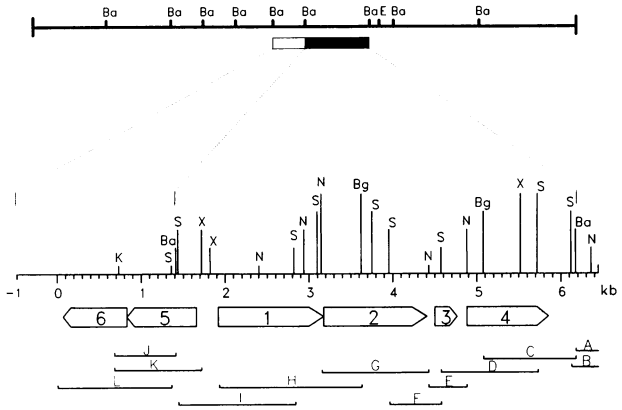


Fig. 2. Fine restriction map and gene structure of the *gral,III* regions of *S. violaceoruber* Tü22. *gral* (filled rectangle) represents ORFs 1–4, and *gralII* (open rectangle) represents ORFs 5–6, based on homology in Southern blots of these gene segments with *actI* and *actIII*, respectively (Malpartida *et al.*, 1987). Vertical lines represent restriction enzyme sites: Ba, Bg, E, K, N, S, X = *Bam*HI, *Bgl*II, *Eco*RI, *Kpn*I, *Nco*I, *Sal*I, *Xho*I. Orientation of arrows denotes direction of transcription for each ORF. Bars at bottom of the figure represent M13 clones used for sequencing, derived from the following subcloning experiments: (**A**) 0.8 kb pIJ2361 *Bam*HI–*Eco*RI fragment cloned into *Bam*HI+*Eco*RI-digested M13mp18; (**B**) 0.9 kb pIJ2361 *Sal*I–*Eco*RI fragment ligated to similarly digested M13mp18; (**C**) 1.1 kb pIJ5200 *Bgl*II–*Bam*HI fragment cloned in both orientations into *Bam*HI-digested M13mp18; (**D**) 1.2 kb pIJ5200 *Sal*I fragment cloned in both orientations into *Sal*I-digested M13mp19; (**E**) 450 bp pIJ5200 *Nco*I fragment which was blunt-ended and cloned in both orientations into the *Sma*I site of M13mp18; (**F**) 600 bp pIJ5200 *Sal*I fragment cloned in both orientations into *Sal*I-digested M13mp19; (**G**) 1.3 kb pIJ5200 *Nco*I fragment which was blunt-ended and cloned in both orientations into the *Sma*I site of M13mp18; (**H**) 1.6 kb pIJ5200 *Pvu*II–*Bgl*II fragment ligated to *Hinc*II+*Bgl*II-digested M13mp19; (**I**) 1.4 kb pIJ5200 *Sal*I fragment cloned in both orientations into *Sal*I-digested M13mp19; (**J**) 800 bp pIJ2361 *Bam*HI–*Eco*RI fragment ligated to similarly digested M13mp18 and M13mp19; (**K**) 1.0 kb pIJ2361 *Xho*I–*Eco*RI fragment cloned into *Sal*I+*Eco*RI-digested M13mp18; (**L**) 2.4 kb pIJ5201 *Bam*HI fragment cloned into *Bam*HI-digested M13mp18. Clone L was sequenced using a series of overlapping oligonucleotide primers.

chains of polyketides, polyketide biosynthesis is potentially much more complex in three important respects. (i) The cycle of reduction–dehydration–reduction that follows every two-carbon addition in fatty acid biosynthesis is omitted, or curtailed, at some or all points in the polyketide chain (Figure 1). Thus, for each step in chain assembly, the polyketide synthase (PKS) chooses between four possibilities; condensation–reduction–dehydration–reduction, condensation–reduction–dehydration, condensation–reduction and condensation alone, giving alkyl [W], enoyl [X], hydroxyl [Y] or keto [Z] functionalities at different carbon atoms (Figure 1). (ii) Whereas the FAS is constrained to use an acetate unit as starter and a malonate unit for every extender, many PKSs make a choice of starter unit and of each successive extender unit from a range of possibilities, including acetate, propionate, butyrate and occasionally more complex residues (Ōmura and Tanaka, 1984; Simpson, 1985). (iii) The product of the FAS is achiral, unlike many polyketides where chirality at multiple carbon atoms is stereospecifically induced, presumably by the PKS. It is these three degrees of flexibility in polyketide carbon chain assembly which account for the great diversity of structures produced. Yet, a particular PKS usually shows little variation in the product it makes. The molecular mechanisms which control the construction of a single type of carbon chain (both stereospecifically and enantiospecifically) by the PKS is the fundamental challenge of this area of research. However, before it will be possible to understand these processes of polyketide synthase function, the precise structure and organization of the PKS gene complexes for representative polyketides must be delineated. In this study, we describe the arrangement, nucleotide sequences and probable functions of genes in the PKS gene cluster of *Streptomyces violaceoruber* which produces the benzoisochromane quinone antibiotic, granaticin (Floss *et al.*, 1986). In a companion paper (Bibb *et al.*, 1989), a similar approach to analysis of PKS genes for the anthracycline tetracenomycin is described.

Results

Isolation and sequencing of DNA encoding the granaticin PKS

The construction of the carbon skeleton of granaticin requires the condensation of eight acetate residues. The basic subunit structure (of granaticin and its direct precursor, dihydrogranaticin) (Floss *et al.*, 1986) is identical to that of actinorhodin, which is produced by *Streptomyces coelicolor* A3(2), and has been the subject of considerable study, both genetically (Rudd and Hopwood, 1979; Hopwood *et al.*, 1985; Malpartida and Hopwood, 1984, 1986) and chemically (Brockmann, 1966; Floss *et al.*, 1986). Biochemical genetic analysis of the actinorhodin (*act*) biosynthetic pathway identified seven classes of mutants blocked at different stages which allowed an ordered biosynthetic sequence to be deduced: I,III,VII,IV,VI,VA,VB (Rudd and Hopwood, 1979; Cole *et al.*, 1987). *ActI,III* mutants failed to secrete any biosynthetic intermediate active in co-synthesis with any other classes of mutants, while being able to convert to actinorhodin intermediates secreted by the other mutant classes. This indicated that *actI,III* mutants were likely to

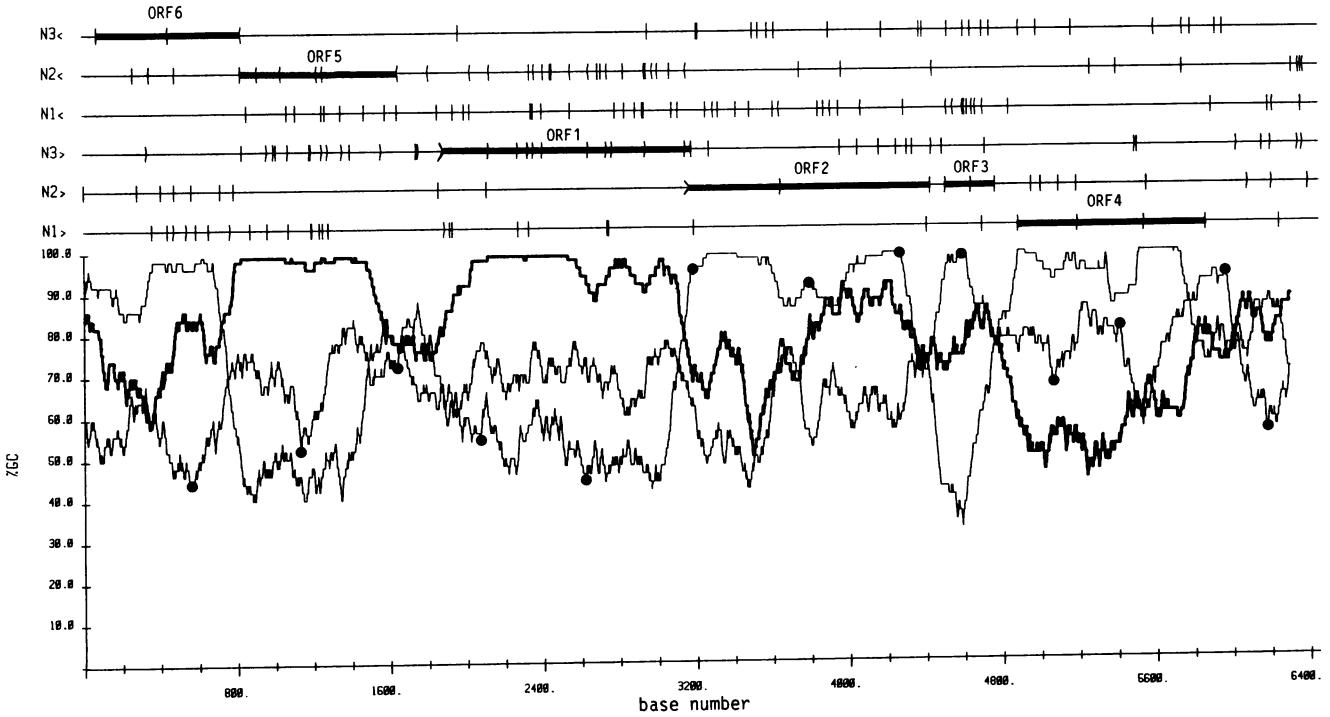


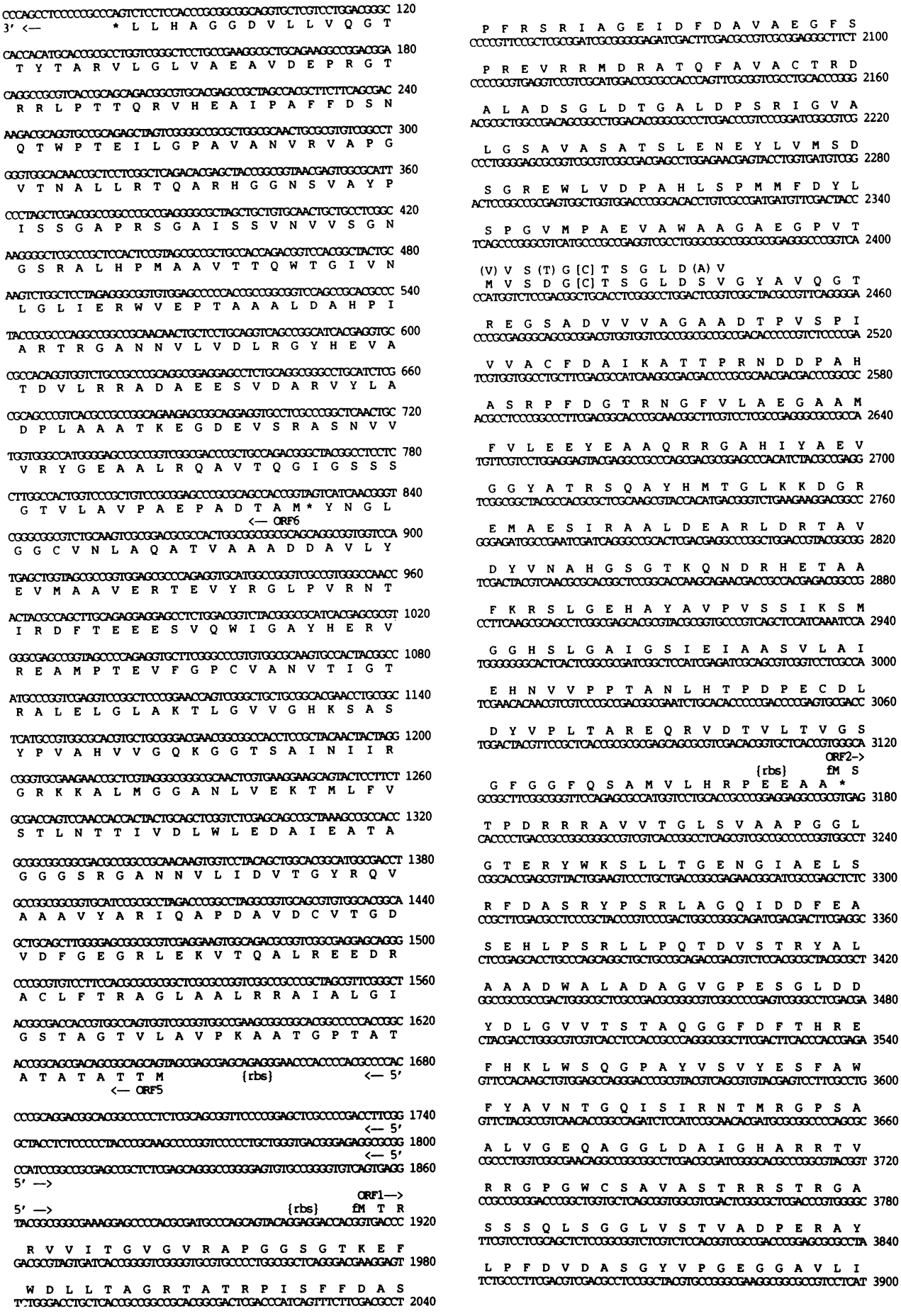
Fig. 3. FRAME plot (Bibb *et al.*, 1984) of the 6441 bp *gral,III* region using a bandwidth of 50 triplets. The plot indicates four protein coding regions for *gral* (ORFs 1−4) and two for *graIII* (ORFs 5−6): ORF1 = frame N3>, bold line from bp 1914 to 3178; ORF2 = frame N2>, bold line from bp 3176 to 4423; ORF3 = frame N2>, bold line from bp 4505 to 4762; ORF4 = frame N1>, bold line from bp 4882 to 5856; ORF5 = frame N2<, bold line from bp 1645 to 827; ORF6 = frame N3<, bold line from bp 827 to 68. Symbols > and < indicate ATG codons, and symbol | indicates stop codons. The symbol > indicates a GTG start codon.

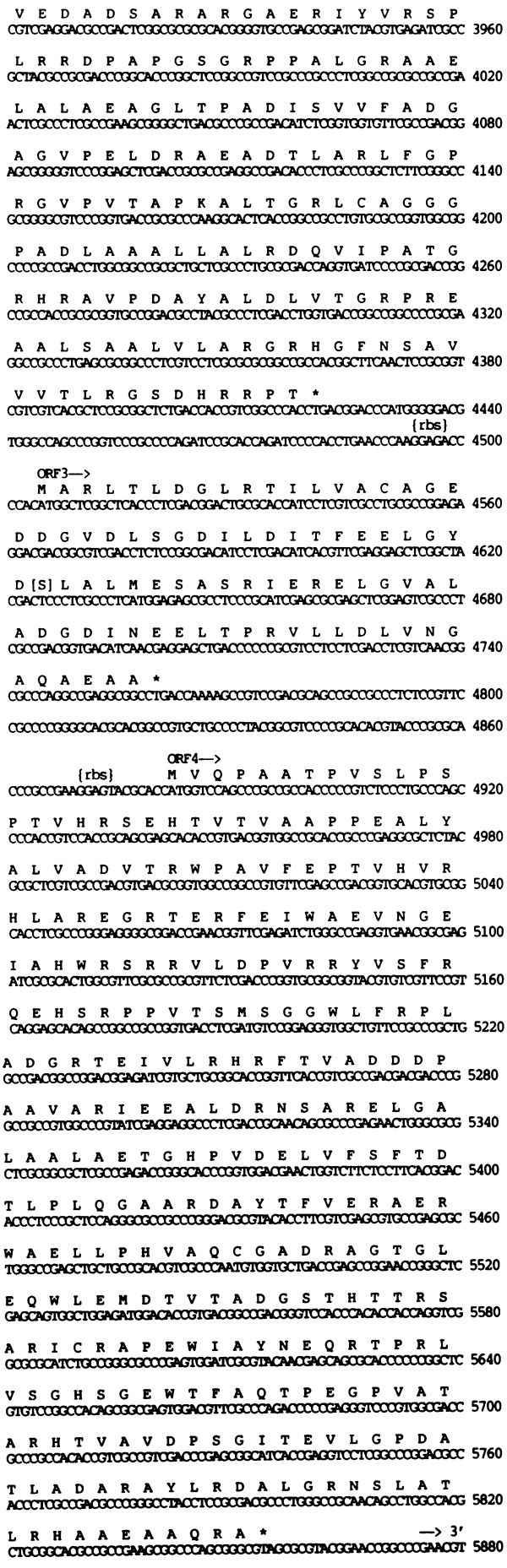
be defective in the genes which encode enzymes catalysing the construction of the polyketide carbon skeleton of actinorhodin. A 2.2 kb *Bam*HI fragment of *S.coelicolor* DNA was shown to contain the *actI* region, and an adjacent 1.1 kb *Bam*HI fragment corresponded to *actIII* (Malpartida and Hopwood, 1986). The derived amino acid sequence of a single open reading frame (ORF) revealed by sequencing the *actIII* fragment indicated that it encodes a ketoreductase, presumably involved in polyketide biosynthesis (Hallam *et al.*, 1988). The *actI* and *actIII* DNA fragments, encoding components of the actinorhodin PKS of *S.coelicolor*, were used as probes in Southern blots against the DNA of other polyketide-producing strains of *Streptomyces* (Malpartida *et al.*, 1987). *S.violaceoruber*, the producer of granaticin, contained DNA which hybridized particularly strongly to *actI* and *actIII*. This DNA was isolated [as a series of EMBL4 (Frischauf *et al.*, 1983) clones] using the combined *actI* and *actIII* fragments of *S.coelicolor* as a probe. Specific segments of the *S.violaceoruber* clones were shown to complement both *actI* and *actIII* blocked mutants of *S. coelicolor*. Furthermore, a small fragment of DNA derived from the *actI* homologue of *S.violaceoruber* was capable of generating granaticin non-producing mutants when inserted in a bacteriophage cloning vector, ϕC31 KC516 (Rodicio *et al.*, 1985), and used to form lysogens in a gene disruption experiment. These results, taken together, demonstrated the functional role of the cloned *S.violaceoruber* DNA in granaticin biosynthesis and formed the basis for the next phase of the study, described below.

The granaticin (*gra*) PKS cluster resides on a 4.6 kb *Bam*HI fragment which contains the *actI*-homologous region, and a 2.4 kb *Bam*HI fragment immediately adjacent which contains the *actIII*-homologous DNA. Figure 2 shows the organization of the *gra* region and an outline of the nucleotide sequencing strategy.

Analysis of the sequence for ORFs, using the program FRAME, is shown in Figure 3 and the complete sequence and its deduced gene products are given in Figure 4. There are six ORFs. The predicted direction of transcription for the ORF 1−4 cluster is from left to right, while ORF5 and ORF6 would be transcribed divergently from ORF 1−4. The predicted start site for ORF1 (1266 bp) is a GTG codon (fMet) at nucleotide position 1914 and a potential ribosome-binding site is centred ~9 bp upstream (Figure 4). The 3′ end of ORF1 has an overlap of 4 nucleotides with the predicted start of ORF2 (1260 bp), which begins at a GTG (nucleotide 3176), preceded by a ribosome-binding site centred ~8 nucleotides upstream. ORF2 and ORF3 are separated by an 81 bp segment of non-coding sequence. ORF3 (258 bp) encodes the smallest protein in the *gra* PKS cluster with 85 amino acid residues. The predicted ATG start (nucleotide 4505) has a potential ribosome-binding site centred ~10 bp upstream (Figure 4). ORF3 is separated from ORF4 (735 bp) by 120 bp of non-coding DNA. A potential ribosome-binding site for ORF4 is centred ~12 bp upstream of the predicted start site (ATG, nucleotide 4882). The FRAME analysis (Figure 3) suggests that an additional, incompletely sequenced, ORF just downstream of ORF4 is transcribed from right to left.

The diverging ORF 5−6 cluster which hybridized to *actIII* is separated from the ORF 1−4 cluster by a 268 bp segment of non-coding DNA. A potential ribosome-binding site is centred ~15 bp upstream of the predicted ATG start site (nucleotide 1645) of ORF5 (832 bp). As was found for ORF1 and ORF2, there is evidence that ORF5 and ORF6 are 'translationally coupled'. In this case there is an overlap


3′ <— * L L H A G G D V L L V Q G T
T Y T A R V L G L V A E A V D E P R G T
R R L P T T Q R V H E A I P A F F D S N
Q T W P T E I L G P A V A N V R V A P G
V T N A L L R T Q A R H G G N S V A Y P
I S S G A P R S G A I S S V N V V S G N
G S R A L H P M A A V T T Q W T G I V N
L G L I E R W V E P T A A A L D A H P I
A R T R G A N N V L V D L R G Y H E V A
T D V L R R A D A E E S V D A R V Y L A
D P L A A A T K E G D E V S R A S N V V
V R Y G E A A L R Q A V T Q G I G S S S
G T V L A V P A E P A D T A M* Y N G L
<— ORF6
G G C V N L A Q A T V A A A D D A V L Y
E V M A A V E R T E V Y R G L P V R N T
I R D F T E E E S V Q W I G A Y H E R V
R E A M P T E V F G P C V A N V T I G T
R A L E L G L A K T L G V V G H K S A S
Y P V A H V V G Q K G G T S A I N I I R
G R K K A L M G G A N L V E K T M L F V
S T L N T T I V D L W L E D A I E A T A
G G G S R G A N N V L I D V T G Y R Q V
A A A V Y A R I Q A P D A V D C V T G D
V D F G E G R L E K V T Q A L R E E D R
A C L F T R A G L A A L R R A I A L G I
G S T A G T V L A V P K A A T G P T A T
A T A T A T T M {rbs} <— 5′
<— ORF5
<— 5′
<— 5′
5′ —>
ORF1—>
5′ —> {rbs} fM T R
R V V I T G V G V R A P G G S G T K E F
W D L L T A G R T A T R P I S F F D A S
P F R S R I A G E I D F D A V A E G F S
P R E V R R M D R A T Q F A V A C T R D
A L A D S G L D T G A L D P S R I G V A
L G S A V A S A T S L E N E Y L V M S D
S G R E W L V D P A H L S P M M F D Y L
S P G V M P A E V A W A A G A E G P V T
(V) V S (T) G [C] T S G L D (A) V
M V S D G [C] T S G L D S V G Y A V Q G T
R E G S A D V V V A G A A D T P V S P I
V V A C F D A I K A T T P R N D D P A H
A S R P F D G T R N G F V L A E G A A M
F V L E E Y E A A Q R R G A H I Y A E V
G G Y A T R S Q A Y H M T G L K K D G R
E M A E S I R A A L D E A R L D R T A V
D Y V N A H G S G T K Q N D R H E T A A
F K R S L G E H A Y A V P V S S I K S M
G G H S L G A I G S I E I A A S V L A I
E H N V V P P T A N L H T P D P E C D L
D Y V P L T A R E Q R V D T V L T V G S
ORF2->
{rbs} fM S
G F G G F Q S A M V L H R P E E A A *
T P D R R R A V V T G L S V A A P G G L
G T E R Y W K S L L T G E N G I A E L S
R F D A S R Y P S R L A G Q I D D F E A
S E H L P S R L L P Q T D V S T R Y A L
A A A D W A L A D A G V G P E S G L D D
Y D L G V V T S T A Q G G F D F T H R E
F H K L W S Q G P A Y V S V Y E S F A W
F Y A V N T G Q I S I R N T M R G P S A
A L V G E Q A G G L D A I G H A R R T V
R R G P G W C S A V A S T R R S T R G A
S S S Q L S G G L V S T V A D P E R A Y
L P F D V D A S G Y V P G E G G A V L I



of one nucleotide encompassing the TGA codon of ORF5 and the predicted ATG (nucleotide 827) start of ORF6 (756 bp). ORF6 is not preceded by a strong ribosome-binding site.

Deduced functions of the gra PKS ORFs

Computer-assisted comparison of the deduced product of ORF1 with other proteins revealed end-to-end similarity with the *fabB*-specified ketoacyl synthase of *E.coli* (Figure 5) and the ketoacyl synthase region of the FAS of *S.cerevisiae* (data not shown). For FabB, this includes a co-linear alignment with the known active site cysteine residue. This suggests that *gra* ORF1 may carry out a ketoacyl synthase function in polyketide chain assembly. Further comparison of the FabB amino acid sequence with *gra* ORF2 revealed an overall similarity of the proteins (Figure 5b). Alignment of the deduced amino acid sequences of *gra* ORF1 and ORF2 shows very strong overall similarity (Figure 5c). The lack of a co-linear cysteine residue, which would provide the putative active site of ORF2, is a major difference between these two sequences.

A well characterized feature of the FAS from all systems studied (Hale and Leadley, 1985; Hale *et al.*, 1987) is an ACP of ~9 kd. Comparisons of the deduced amino acid sequence of the 9 kd predicted product of *gra* ORF3 with FAS ACPs revealed considerable amino acid similarity. This was especially striking in the region of the active site serine residue (Figure 6) which contains a conserved motif (Asp-Ser-Leu) for all the known FAS ACPs and now includes the putative *gra* PKS ACP. Interestingly, the amino acid residues on either side of this triplet motif are quite different: a Tyr replaces Met/Ala on the amino-terminal side, and Ala replaces Asp on the carboxy-terminal side.

No meaningful similarities were found with comparisons of the predicted product of *gra* ORF4 to sequences in major databases or to available FAS-related sequences. Thus the possible function of *gra* ORF4 remains unknown.

A number of molecular genetic experiments provide very strong evidence that the *actIII* gene product carries out a keto to hydroxyl reductive function in polyketide biosynthesis (Hallam *et al.*, 1988). As expected from Southern blots using an *actIII* probe (data not shown), comparisons of the deduced amino acid sequence of ORF5 with the *actIII* ORF revealed strong co-linear similarity throughout the length of the proteins (Figure 7a), excluding the extreme amino- and carboxy-terminals. A similar comparison for the *gra* ORF6 and *actIII* also showed similarity at the amino acid level (Figure 7b). Interestingly, *gra* ORF6 appears to be more similar to *actIII* than to *gra* ORF5 (Figure 7c). Since *actIII*

Fig. 4. Nucletide sequence of *graI* ORFs 1–4 and *graIII* ORFs 5–6 of *S.violaceoruber*. The deduced amino acid sequence is given [below the DNA sequence (3′ ← 5′) in ORF5 and ORF6, which are transcribed from right to left, and above the DNA sequence (5′ → 3′ strand switch at nucleotide 1801) for ORFs 1–4 which are transcribed from left to right] in the standard one letter amino acid code. The regions upstream of a presumed ATG (or fMet, GTG) start codon which show complementarity to the 16S rRNA of *S.lividans* are indicated by { } (above the sequence for ORFs 1–4 and below the sequence for ORFs 5–6) and the letters rbs (ribosome-binding site). Each deduced ORF start site is identified with an arrow showing the direction of transcription and each termination codon by an asterisk. The putative active site cysteine residue in ORF1 [along with an alignment of the active site region of *tcm* ORF1 (Bibb *et al.*, 1989)] and the active site serine in ORF3 are indicated by [ ].

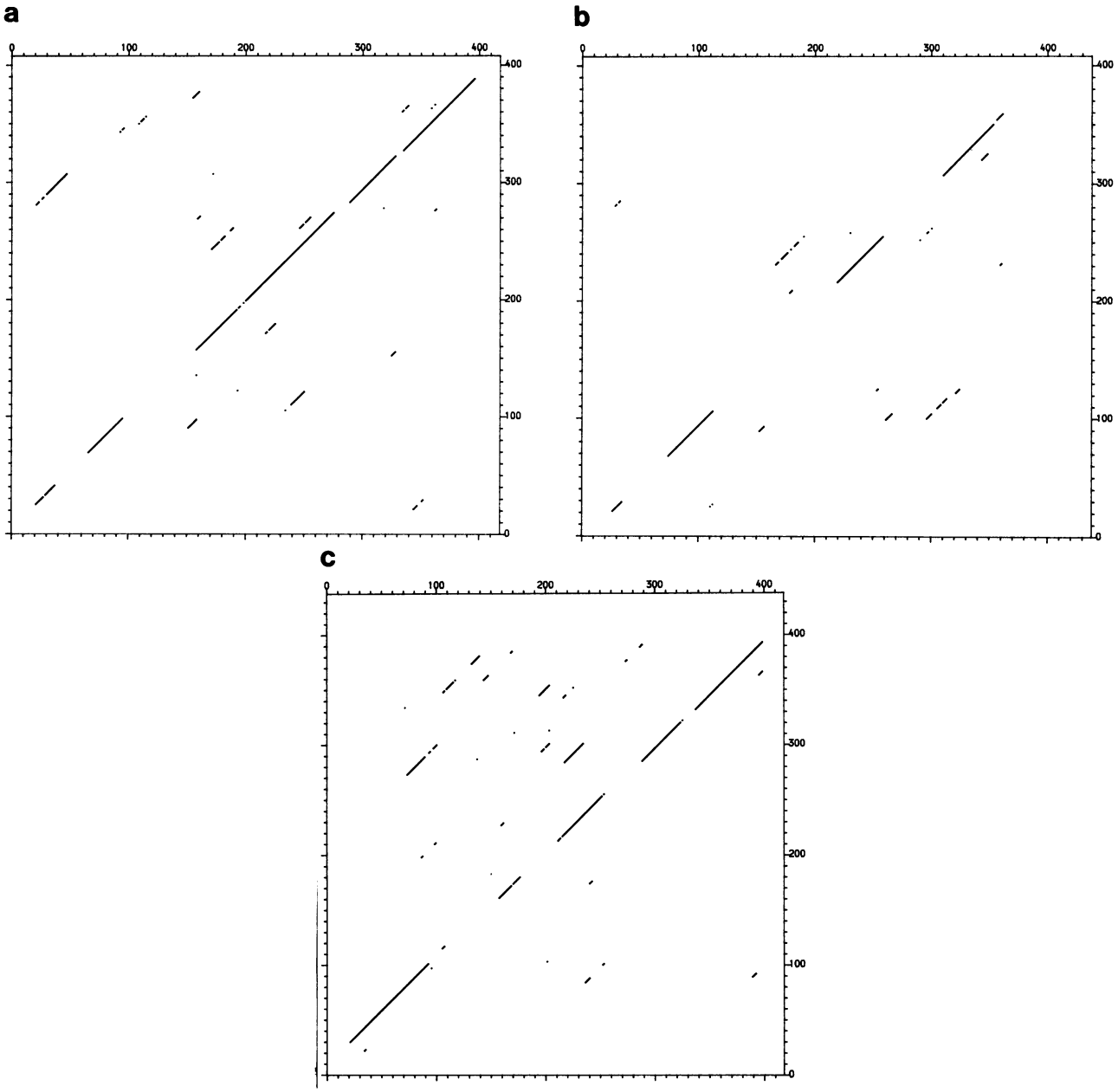


Fig. 5. DOTPLOT comparisons of *gra* ORF1 and *gra* ORF2 proteins, and FabB protein of *E.coli* (Kauppinen *et al.*, 1988). (**a**) *gra* ORF1 (vertical axis) and FabB; (**b**) *gra* ORF2 and FabB; (**c**) *gra* ORF1 and *gra* ORF2. For an explanation of stringency, see Computer analysis section of Materials and methods.

probably encodes an oxidoreductase, it seems likely from these results that *gra* ORF5 and ORF6 also both encode oxidoreductases.

Discussion

Analysis of the *graI* and *graIII* DNA segments has clarified a number of issues concerning the *S. violaceoruber* PKS gene organization. First, the PKS is specified by several separate ORFs rather than a single ORF that encodes a protein with multiple catalytic sites. Second, potential functions have been assigned to most of the *gra* PKS genes. Third, the direction of transcription of the PKS genes and, therefore, the probable location of a bidirectional promoter region that controls them have been revealed.

The apparent 'translational coupling' of ORF1 and ORF2,

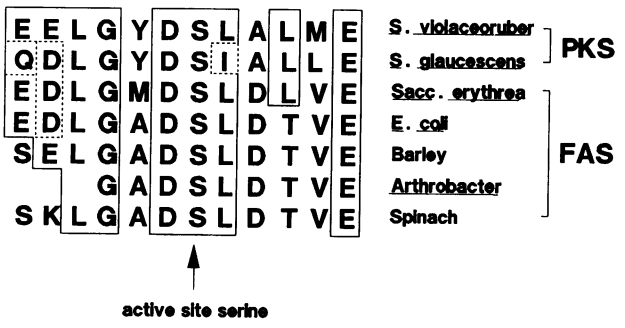


Fig. 6. Amino acid sequences encompassing the presumed active site serine residues of the *S. violaceoruber* and *S. glaucescens* (Bibb *et al.*, 1989), PKS acyl carrier protein and those of FAS acyl carrier proteins from *Saccharopolyspora erythrea*; *E.coli*, Barley, *Arthrobacter* and Spinach (Hale and Leadley, 1985; Hale *et al.*, 1987).

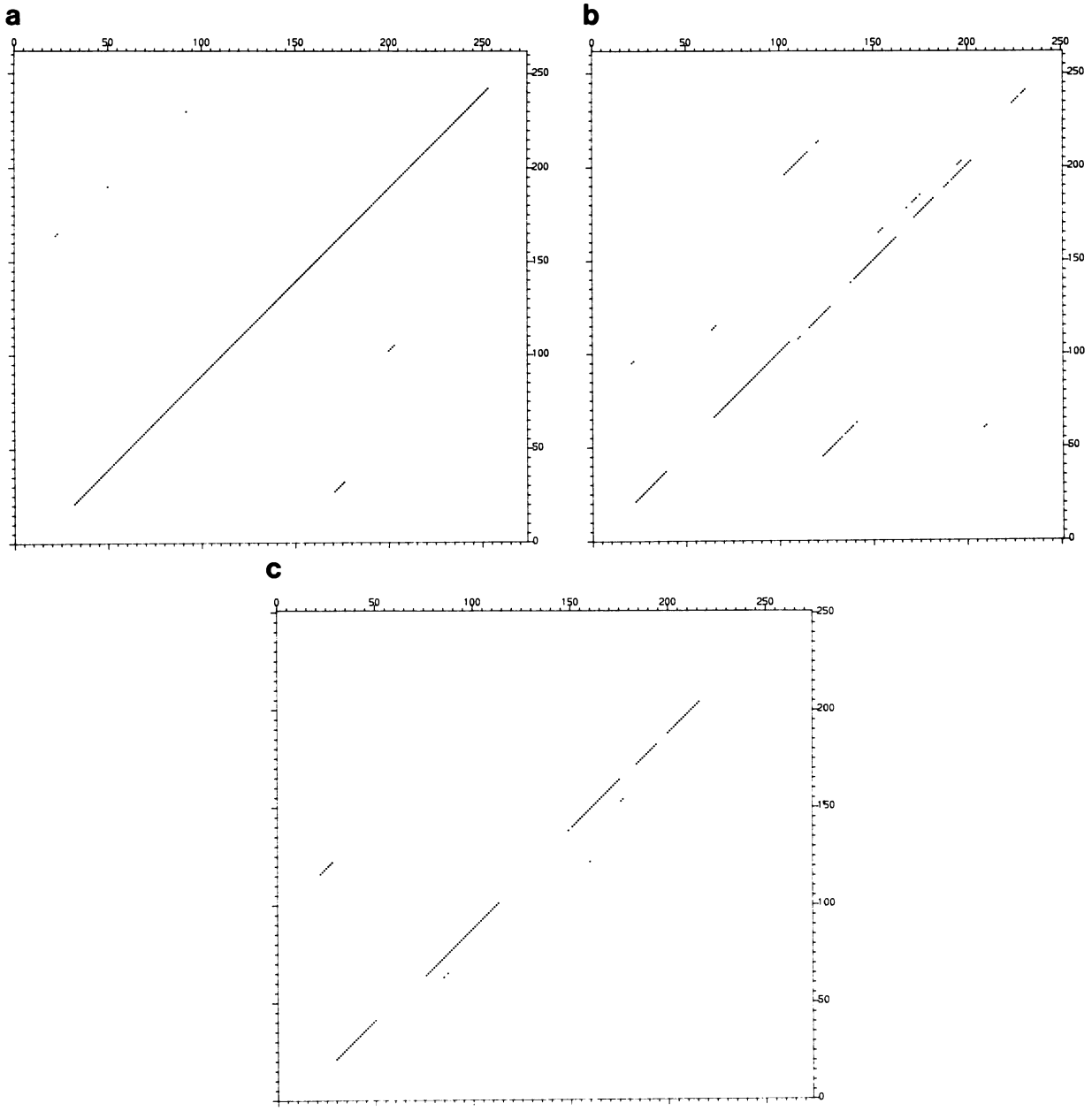


Fig. 7. DOTPLOT comparisons of *gra* ORF5, *gra* ORF6 and *actIII* proteins: (**a**) *gra* ORF5 and *actIII*; (**b**) *gra* ORF6 and *actIII*; (**c**) *gra* ORF5 and *gra* ORF6. An explanation of stringency conditions is in the Computer analysis section of Materials and methods.

which has been observed in many bacterial operons (Zalkin and Ebbole, 1988), implies the equimolar production of the two gene products. This could have an important functional role in the programmed construction of an oligomeric protein. The lack of a co-linear cysteine residue, which would provide the putative active site residue in ORF2, virtually eliminates the possibility that its gene product could be a second condensing enzyme, even though the existence of multiple ketoacyl synthase enzymes in the *E.coli* FAS complex is well established (Rock and Cronan, 1982; de Mendoza and Cronan, 1983; Jackowski and Rock, 1987; Siggaard-Andersen, 1988). The formation of a ketoacyl synthase heterodimer derived from the *gra* ORF1 and ORF2 gene products seems quite plausible and is consistent with the implication of 'translational coupling'.

The identification of *gra* ORF3 as a PKS ACP is significant in terms of the overall mechanism of polyketide antibiotic biosynthesis in *Streptomyces*. The ACP is a universal feature of all known FAS systems. However, chalcone synthase (Schüz *et al.*, 1983) from *Petroselinum hortense*, which catalyses the formation of naringenin from malonyl CoA and 4-coumaryl CoA, is a notable exception and has been found to contain neither an ACP nor a pantetheinyl residue. In this regard, the *gra* PKS more closely resembles the synthase for fatty acids than that of naringenin chalcone, which is presumably derived from polyketide intermediates.

Our inability to assign a potential function to *gra* ORF4 may reflect the lack of sequence data available from bacterial FAS genes and proteins. [Only two bacterial FAS genes (both from *E.coli*) have been sequenced, the *fabB* ketoacyl synthase (Kauppinen *et al.*, 1988), and the *fabA* β-hydroxy-

decanoyl thioester dehydrase (Cronan *et al.*, 1988).] We might expect the gene product of *gra* ORF4 to possess acyl transferase or dehydrase activity since these are significant functions which have not been assigned in the *gra* cluster. Alternatively, *gra* ORF4 may encode a protein involved in a function which distinguishes polyketide biosynthesis from fatty acid biosynthesis; specifically, the ability to alter the sequence of condensation–reduction–dehydration–reduction after each malonyl condensation during chain elongation. It is also possible that other genes, encoding enzymes involved in granaticin polyketide chain assembly, will be found outside of the regions analysed in this study.

The presence of *gra* ORF5 and ORF6 (which show evidence of translational coupling) represents an important difference between *graIII* and *actIII*, which comprises a single ORF, shown to be encoded by a monocistronic transcript (Hallam *et al.*, 1988).

A companion paper (Bibb *et al.*, 1989) reports similar results to those described here. Sequence analysis of an *actI*-homologous segment of *S.glaucescens* DNA which complements *tcmI* mutants of *S.glaucescens* (Malpartida *et al.*, 1987), and is involved in the production of the anthracycline antibiotic tetracenomycin, revealed three ORFs which correspond to *gra* ORFs 1–3. The DNA sequence and deduced protein sequences of *tcm* ORF 1–2 and *gra* ORF 1–2 show very strong similarity, particularly in the region of the presumed active site (see Figure 4, ORF1 for active site sequence comparison). Alignment of the deduced amino acid sequences of *gra* ORF3 and *tcm* ORF3 show very strong similarity and the active site regions are compared in Figure 6. These data will provide a framework to begin investigating the unique properties of different PKSs which produce a variety of secondary metabolites. It will now be possible to make direct comparisons of other systems which are involved in the production of acetate-derived polyketides with various numbers of carbon atoms, as well as those involved in the production of compounds derived from mixures of precursors. These comparisons may pin-point certain regions of different PKSs to be studied more closely in order to begin to investigate the molecular mechanisms that control the specificity of carbon chain assembly.

Materials and methods

Bacterial strains and plasmids

S.lividans TK64 (Hopwood *et al.*, 1985) was the host for plasmids derived from the low copy number vector pIJ941 (Lydiate *et al.*, 1985) which carry the *gra* cluster gene segments. The plasmids were obtained from subclones of three EMBL4 clones (1.1, 1.2, 7.1) which contain DNA homologous to *actI* and *actIII* (Malpartida *et al.*, 1987). pIJ2361 was obtained by subcloning a 6.4 kb *Eco*RI fragment from EMBL4::1.2 into the *Eco*RI site of pIJ941. pIJ2360 was constructed by subcloning a 9.0 kb *Eco*RI fragment from EMBL4::1.1 into pIJ941. DNA involved in this subcloning is derived from a region to the right of the *actI*-homologous *gra* genes. pIJ5200 contains a 4.6 kb *Bam*HI fragment derived from pIJ2361 which was subcloned into pUC18. pIJ5201 was derived by subcloning an *actIII*-homologous 2.4 kb *Bam*HI fragment from EMBL4::7.1 into M13mp18. Additional subclones for sequencing were constructed using specific fragments of pIJ5200 (Figure 2). *E.coli* strain *lacZ* ΔM15 *recA* (Rüther *et al.*, 1981) was used as the host for plasmids derived from pUC18 (Norrander *et al.*, 1983). *E.coli* JM101 was used as the host for phages M13mp18 and M13mp19 (Yanisch-Perron *et al.*, 1985).

Standard media and manipulations

Procedures for growth and manipulation of *Streptomyces* and for general recombinant DNA manipulations were as described (Hopwood *et al.*, 1985). *E.coli* was transformed by the procedure of Hanahan (1983) and transformants were selected on L agar (Miller, 1972) containing 100 μg/ml carbenicillin. Procedures for the growth and manipulation of M13 were as described by Messing (1983).

DNA isolation

Streptomyces plasmid DNA was prepared as described by Kieser (1984). The alkaline lysis method (Maniatis *et al.*, 1982) 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 (1981).

Sequence analysis

Fragments of *graI* and *graIII* DNA were cloned in both orientations (see Figure 2 for sequencing strategy) in either M13mp18 or M13mp19 (Yanisch-Perron *et al.*, 1985). Nested deletions were made by the method of Henikoff (1984). Both strands of every clone were sequenced by the dideoxy method (Sanger *et al.*, 1977). Regions of compression or sequence ambiguity were resolved using Sequenase (US Biochemicals).

Computer analysis

Sequencing gels were read directly into a VAX11/750 computer (AFRC Computing Centre, Harpenden) using the UWGCG program SEQED. The sequence was compiled using the UWGCG program ASSEMBLE. FRAME plot (Bibb *et al.*, 1984) was carried out to identify ORFs and direction of transcription. Comparison of amino acid sequences (Figures 5a–c, 7a–c) 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 a 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 (Devereux *et al.*, 1984). Amino acid sequence comparisons were made with the NBRF protein database (Protein Identification Resource, National Biomedical Research Foundation, Georgetown University, Washington; release 13.0, June 1987) and the PSEQIP database using previously described methods (Coulson *et al.*, 1987; Collins *et al.*, 1988).

Acknowledgements

We thank Penny von Wettstein-Knowles for providing the *fabB* DNA sequence and Sandor Biró and C.Richard Hutchinson for communicating results prior to publication, and John Collins for performing the PSEQIP database search. We are grateful to John Robinson and Tom Simpson for helpful discussions, and to Keith Chater and Tobias Kieser 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

Bibb,M.J., Findlay,P.R. and Johnson,M.W. (1984) *Gene*, **30**, 157–166.
Bibb,M.J., Biró,S., Motamedi,H., Collins,J.F. and Hutchinson,C.R. (1989) *EMBO J.*, **8**, 2727–2736.
Birch,A.J. and Donovan,F.W. (1953) *Austral. J. Chem.*, **112**, 431–442.
Brockmann,H., Zeeck,A., van der Merve,K. and Müller,W. (1966) *Justus Liebigs Annln. Chem.*, **698**, 3575–3579.
Chirala,S.S., Kuziora,M.A., Spector,D.M. and Wakil,S.J. (1987) *J. Biol. Chem.*, **262**, 4231–4240.
Cole,S.P., Rudd,B.A.M., Hopwood,D.A., Chang,C.-J. and Floss,H.G. (1987) *J. Antibiotics*, **40**, 340–347.
Collins,J.F., Coulson,A.F.W. and Lyall,A. (1988) *CABIOS*, **4**, 67–71.
Coulson,A.F.W., Collins,J.F. and Lyall,A. (1987) *Computer J.*, **30**, 420–424.
Cronan,J.E., Li,W.-B., Coleman,R., Narasimhan,M., de Mendoza,D. and Schwab,J.M. (1988) *J. Biol. Chem.*, **263**, 4641–4646.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Floss,H.G., Cole,S.P., He,X.G., Rudd,B.A.M., Duncan,J., Fujii,I., Chang,C.-J. and Keller,P.J. (1986) In Kleinkauf,H., von Döhren,D., Dornauer,H. and Nesemann,G. (eds), *Regulation of Secondary Metabolite Formation*. VCH, Weinheim, pp. 283–304.
Frischauf,A.-M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.

Hale,R.S. and Leadley,P.F. (1985) *Biochimie*, **67**, 835–839.
Hale,R.S., Jordan,K.N. and Leadley,P.F. (1987) *FEBS Lett.*, **224**, 133–136.
Hallam,S.E., Malpartida,F. and Hopwood,D.A. (1988) *Gene*, **74**, 305–320.
Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557–580.
Henikoff,S. (1984) *Gene*, **28**, 351–359.
Hopwood,D.A., Bibb,M.J., Chater,K.F., Kieser,T., Bruton,C.J., Kieser,H.M., Lydiate,D.J. Smith,C.P., Ward,J.M. and Schrempf,H. (1985) *Genetic Manipulation of Streptomyces. A Laboratory Manual.* John Innes Foundation, Norwich, UK.
Hopwood,D.A., Malpartida,F., Kieser,H.M., Ikeda,H., Duncan,J., Fujii,I., Rudd,B.A.M., Floss,H.G. and Omura,S. (1985) *Nature*, **314**, 642–644.
Ish-Horowitz,D. and Burke,J.F. (1981) *Nucleic Acids Res.*, **9**, 2989–2998.
Jackowski,S. and Rock,C.O. (1987) *J. Biol. Chem.*, **262**, 7927–7931.
Kauppinen,S., Siggaard-Anderson,M. and von Wettstein-Knowles,P. (1988) *Carlsberg Res. Commun.*, **53**, 357–370.
Kieser,T. (1984) *Plasmid*, **12**, 19–36.
Lydiate,D.J., Malpartida,F. and Hopwood,D.A. (1985) *Gene*, **35**, 223–235.
Lynen,F. (1980) *Eur. J. Biochem.*, **112**, 431–442.
McCarthy,A.D. and Hardie,D.G. (1984) *Trends Biochem. Sci.*, **9**, 60–63.
Malpartida,F. and Hopwood,D.A. (1984) *Nature*, **309**, 462–464.
Malpartida,F. and Hopwood,D.A. (1986) *Mol. Gen. Genet.*, **205**, 66–73.
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. and Hopwood,D.A. (1987) *Nature*, **325**, 818–821.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
de Mendoza,D. and Cronan,J.E. (1983) *Trends Biochem. Sci.*, **8**, 49–52.
Messing,J. (1983) In Wu,R., Grossman,L. and Moldave,K. (eds), *Methods in Enzymology.* Academic Press, New York, Vol. 101, pp. 20–28.
Miller,J. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mohamed,A.H., Chirala,S.S., Mody,N.H., Huang,W.-Y. and Wakil,S.J. (1988) *J. Biol. Chem.*, **263**, 12315–12325.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Omura,S. and Tanaka,Y. (1984) In Omura,S. (ed.), *Macrolide Antibiotics.* Academic Press, New York, pp. 199–225.
Rock,C.O. and Cronan,J.E. (1982) *Curr. Top. Membr. Transp.*, **17**, 207–233.
Rodicio,M.R., Bruton,C.J. and Chater,K.F. (1985) *Gene*, **34**, 283–292.
Rudd,B.A.M. and Hopwood,D.A. (1979) *J. Gen. Microbiol.*, **114**, 35–43.
Rüther,U., Koenen,M., Otto,K. and Müller-Hill,B. (1981) *Nucleic Acids Res.*, **9**, 4087–4098.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schüz,R., Heller,W. and Hahlbrock,K. (1983) *J. Biol. Chem.*, **258**, 6730–6734.
Schweizer,E., Roberts,L.M., Höltke,H.-J., Takabayashi,K., Höllerer,K., Hoffmann,E., Müller,G., Köttig,H. and Schweizer,E. (1986) *Mol. Gen. Genet.*, **203**, 479–486.
Schweizer,E., Müller,G., Roberts,L.M., Schweizer,M., Rösch,J., Wiesner,P., Beck,J., Stratmann,D. and Zauner,I. (1987) *Fat Sci. Technol.*, **89**, 570–577.
Sherman,D.H., Malpartida,F., Bibb,M.J., Kieser,H.M., Hallam,S.E., Robinson,J.A., Bergh,S., Uhlen,M., Simpson,T.J. and Hopwood,D.A. (1988) In Durand,G., Bobichon,L. and Florent,J. (eds), *Proceedings of the 8th International Biotechnology Symposium, Paris.* Société Française de Microbiologie, Vol. 1, pp. 123–137.
Siggaard-Andersen,M. (1988) *Carlsberg Res. Commun.*, **53**, 371–379.
Simpson,T.J. (1985) *Natural Products Rep.*, **2**, 321–347.
Wakil,S.J. (1986) *Ann. NY Acad. Sci.*, **478**, 203–219.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Zalkin,H. and Ebbole,D.J. (1988) *J. Biol. Chem.*, **263**, 1595–1598.

Received on March 30, 1989