The Journal of Microbiology, March 1997, p.40-46
Vol. 35, No. 1
Copyright © 1997, The Microbiological Society of Korea

Nucleotide Sequence Analysis of a Second Set of the Polyketide Synthase β-Ketoacyl synthase and Chain Length Factor Genes from the Salinomycin-producing *Streptomyces albus*

Chang-Gu Hyun, Kwan-Hyung Park, C. Richard Hutchinson[1] and Joo-Won Suh*

Department of Biological Science, Biotechnology Research Institute, Myong Ji University, Yongin 449-728, Korea
[1]School of Pharmacy & Department of Bacteriology, University of Wisconsin, Madison, WI 53706

(Received November 22, 1996 / Accepted December 20, 1996)

The pWHM220 cosmid with a 24-kb insert cloned from *Streptomyces albus* ATCC 21838 induces the biosynthesis of a polyether antibiotic similar to salinomycin in *Streptomyces lividans*. We have analyzed this region by DNA sequencing as well as Southern blot hybridization with type I and type II polyketide synthase (PKS) probes. Surprisingly, we found another set of type II PKS genes only 10-kb from the original PKS genes, *salABCDE*. The DNA sequence revealed two complete open reading frames (ORFs) named *salB2* and *salC2*, and one partial ORF that does not resemble any known DNA or deduced protein sequence. The *salC2* should code for chain length determining factor while the deduced amino acid sequence encoded by *salB2* exhibits high similarity to β-ketoacyl synthase from different PKS gene clusters. The highest identity was found for β-ketoacyl synthases from *S. argillaceus* (MtmP, 59.1% identity), the mithramycin producer and from *S. venezuelae* ISP5230 (JadA, 52.3% identity), the jadomycin producer. The SalC2 protein clearly resembles its counterparts in other aromatic PKS gene clusters that are believed to influence the length of the polyketide chain. The highest identities observed were to that of *S. argillaceus* (MtmK, 62.3%) and *S. venezuelae* ISP5230 (JadB, 55.1%) proteins. Moreover, the deduced amino acid sequences of the *salB2* and *salC2* products were 29.0% identical.

Key words: *Streptomyces albus*, β-ketoacyl synthase, chain length factor

Polyketides are a large group of structurally diverse secondary metabolites, several of which have applications as antibiotics, immunosuppressants, anticancer agents, and veterinary products. In recent years genetic studies have yielded detailed information on the organization and function of genes involved in the biosynthesis of polyketides in microorganism (16, 17).

Cloning, sequencing, and functional analysis of these genes have improved our understanding of the genetic programming of polyketide synthases (16, 24, 25, 26). Two classes of PKSs have been studied from microorganisms (type I and type II (Fig. 1)). Various examples of the type II PKS have been identified (3, 5, 7, 8, 9, 12, 14, 15, 21, 22, 32, 35), which consists of a multienzyme complex formed by several polypeptides with individual enzymatic activities. In contrast, there is a paucity of information about type I PKS where the enzyamtic activities are present as separate domains in one multifunctional polypeptide. This is the case for the PKS involved in biosynthesis of the macrolide antibiotics, such as erythromycin (11) and rapamycin (30).

All of these metabolites share a common mechanism of biosynthesis that has largely resisted in vivo analysis. The carbon skeleton of a polyketide results from the sequential condensation of fatty acids like acetate, propionate, and butyrate. This process is catalyzed by polyketide syhthases in a manner that is conceptually similar to the biosynthesis of long-chain fatty acids catalyzed by the fatty acids synthases found in all organisms (17).

We reported recently the type II PKS gene cluster from the salinomycin producer (8). Five of the genes cloned in pWHM220 were characterized by sequence analysis to confirm that the 3.9-kb DNA fragment contained type II PKS homologs (Fig. 2). The five open reading frames were identified by sequence analysis and named *salA-salE*. Since *salA*, *salB*, *salC*, *salD*, and *salE* genes encode proteins

* To whom correspondence should be addressed

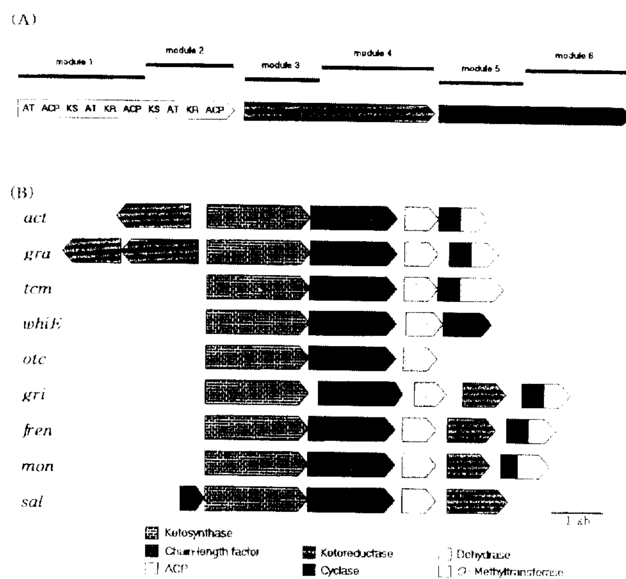

Fig. 1. Comparison of the organization between type I and type II PKSs. (a) Type I PKS for erythromycin biosynthesis. (b) The known gene clusters encoding type II PKSs. References: *act* (12), *fren* (5), *gra* (33), *gri* (35), *mon* (3), *otc* (21), *sal* (8), *tcm* (7), *whiE* (9).

highly similar to the following enzymes : cyclase, β-ketoacyl synthase, chain length determining factor, acyl carrier protein, and ketoreductase; it is very likely that these five *S. albus* genes govern the synthesis of some type of aromatic polyketide.

In this paper we report the additional PKS genes which was found by further sequencing of pWHM 220. These new genes, named *salB2* and *salC2* showed homology with *tcmKL*(7). The two genes seems to encode β-ketoacyl synthase and chain length factor homologs, compose of minimal PKS together with acyl carrier protein.

Materials and Methods

Bacterial strains, bacteriophage, and plasmids

S. albus ATCC21838 was obtained from the American Type Culture Collection (Rockville, Md). *Escharichia coli* DH5α was purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md) and was used for subcloning and squencing. *E. coli* vector used for subcloning was pUC19 (GIBCO-BRL, Gaithersburg Md) and for sequencing, M13mp18 and M13mp19 were used (33). The cosmid clone pWHM220 has been described elsewhere (8). The cosmid vector pKC505 (27) was obtained from Richard Baltz, Eli Lilly Co. Ltd. (Indianapolis, Ind.).

Media, enzymes, and chemicals

Bacterial strains were grown on LB medium where 50 μg/ml ampicillin was added, when necessary.

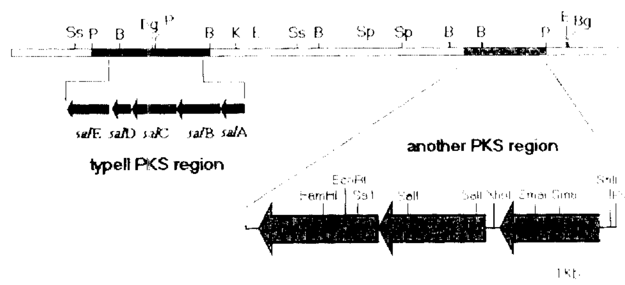

Fig. 2. Restriction map of 24-kb region of *S. albus* ATCC 21838 genomic DNA cloned in pWHM220.

Restriction enzymes, ligase, and klenow enzyme were purchased from Takara Co. All enzyme reactions were carried out according to the recommendations of manufacturers, unless indicated otherwise. All chemicals were purchased from Sigma Chemical Co. (U.S.A).

DNA manipulation

The general DNA manipulation was carried out following Sambrock *et al* (28). Plasmid DNA and bacteriophage RF DNA were isolated from *E. coli* by using alkaline SDS extraction of cell lysate. A large-scale DNA purificaton was performed by banding in CsCl-ethidium bromide density gradients. Single-stranded DNA was isolated by precipitating phage particles. Agarose gel electrophoresis was performed in Tris-acetate or Tris-borate buffer.

Nucleotide sequencing and analysis of sequence

Nested deletions were constructed with the Erase a Base system (Promega Biotech) according to the manufacturer's instruction. Derivatives of pUC19 and M13mp19 containing *S. albus* DNA were digested with *Xba*I-*Pst*I and *Bam*HI-*Kpn*I, respectively before exonuclease III treatment.

Double-stranded templates were subcloned into M13mp18 and single-stranded templetes were sequenced by the dideoxy chain-termination method using (α-^{32}P)dATP and Sequenase version 2.0 (United States Biochemicals) according to the manufacturer's instructions. To avoid compressions, 7-deaza-dGTP was used instead of dGTP. Labeled DNA was separated on 6% polyacrylamide wedge gels.

Primary DNA sequence data were analyzed and assembled by using software from DNASIS and BLAST programs (1,13). DNA and protein sequence were analyzed with the Genetics Computer Group software package (10).

Results and Discussion

Southern analysis of pWHM220 and isolation

of DNA homologous to *tcmKL*

The *actI* gene encoding a type II PKS involved in the biosynthesis of actinorhodin (12), an aromatic polyketide which has played a key role in the elucidation of polyketide synthesis mechanism and β-ketoacyl synthase domain of the *eryA* type I PKS gene(11) were used for Southern analysis of cosmid pWHM220 (Fig. 2). This established the presence of 1.4-kb, 2.4-kb, and 2.9-kb *Bam*HI-*Pst*I fragments that hybridized to *actI* probe, but very faintly to *ery*KS. Because previously described (8) type II polyketide synthase genes from *S. albus* were found in the 3.8-kb *Bam*HI fragment that contains one *Pst*I site and yields 1.4-kb and 2.4-kb *Bam*HI-*Pst*I fragments, we concluded that the 2.9-kb *Bam*HI-*Pst*I fragment should be located in different place from the 3.8-kb *Bam*HI fragment.

Sequence analysis of another PKS genes in pWHM220

The DNA sequence of additional type II PKS region located in the 2.9-kb *Bam*HI-*Pst*I fragment and adjacent 1.3-kb *Bam*HI fragment, encodes another set of the β-ketoacyl synthase and chain length factor (Fig. 3). We identified two complete open reading frames (Fig. 5) with the characteristics of *Streptomyces* genes (overall G+C content, 75.2%; high bias toward G and C in the third codon position, using the CODON PREFERENCE program (10)).

Two open reading frames were preceded by potential ribosome binding sites (GGGAGG and AAGGAG) that showed reasonable degree of complementarity to the 3' end of the *Streptomyces lividans* 16S rRNA (6), suggesting that coding region of upstream gene begins at GTG-71 and coding region of down stream gene begins at ATG-1300 (Fig. 3).

Translation of two open reading frames would result in a polypeptide with 411 amino acids and

```
   1    G AGC CTG GCC CAG CAC GTC GGC CGG GCC TGC CGC GGC CAC GTG GAC     46
  47  CGC GCT GCC CAC GGG AGG CGG CCG GTG AAC CGC GTC GTG ATC ACC GGG     94
   1                                   V   N   R   V   V   I   T   G      8
  95  ATC GGC GTC GTC GCC CCC GGC GCG GTG GGC ACC GCC CGC TTC TGG GAC    142
   9   I   G   V   V   A   P   G   A   V   G   T   A   R   F   W   D     24
 143  CTG CTC ACC GTC GGC CGC ACC GCC ACC CGC CGC GTC ACC CTC TTC CGC    190
  25   L   L   T   V   G   R   T   A   T   R   R   V   T   L   F   R     40
 191  ACC TCG CGG CTA CGC TCC CGC GTC GCC GCC GAG GTC GAC TTC ACC CCC    238
  41   T   S   R   L   R   S   R   V   A   A   E   V   D   F   T   P     56
 239  GCC GCC CAC GGA TCG GAC TCG CCG ACA CCG AAC GCC TGG ACC CGC GCG    286
  57   A   A   H   G   S   D   S   P   T   P   N   A   W   T   R   A     72
 287  GCA CAG TTC GCG CTG GTC GCC GCA CGC GAA GCC GTC GCC GAC AGC GGC    334
  73   A   Q   F   A   L   V   A   A   R   E   A   V   A   D   S   G     88
 335  GTC GCG GAC CGC ATC GGC CGC AAC CCC CTG CGC ACC GGC GTC AGC CTG    382
  89   V   A   D   R   I   G   R   N   P   L   R   T   G   V   S   L    104
 383  GGC AGC GCC ATC GGC TGC ACC ACG AGC CTG GCC ACC CAG TAC GCC ATC    430
 105   G   S   A   I   G   C   T   T   S   L   A   T   Q   Y   A   I    120
 431  CTC AGC GAC TGC GGC ACC ACC TGG ACC CTC GAC CAC ACC GAG GCC GCC    478
 121   L   S   D   C   G   T   T   W   T   L   D   H   T   E   A   A    136
 479  GAA TCC CTC TAC GAC TAC TTC GTG GCC CAG CTC CCT GCC GCC ACC GTC    526
 137   E   S   L   Y   D   Y   F   V   A   Q   L   P   A   A   T   V    152
 527  GCC CGC GAC CGG GGC GCA CAG GGC CCC GTC GCT CTC GTC TCC AGC GGC    574
 153   A   R   D   R   G   A   Q   G   P   V   A   L   V   S   S   G    168
 575  TGC ACC TCC GGC CTG GAC GCC GTC GGC CAC GGC GCC GAC CTG ATC CGG    622
 169   C   T   S   G   L   D   A   V   G   H   G   A   D   L   I   R    184
 623  GAA GGC AGC GCC GAC ATC GTC GTC GCC GGC GGA ACG GAA GCA CCC ATC    670
 185   E   G   S   A   D   I   V   V   A   G   G   T   E   A   P   I    200
 671  GTG CCC ATC GCC ATG GCC TGC TTC GAC CGC CTG CGC CTC ACC AGC TCC    718
 201   V   P   I   A   M   A   C   F   D   R   L   R   L   T   S   S    216
 719  CGC AAC GAC GAC CCC GCC ACC GCC AGC CGC CCC TTC GAC CGC ACC CGC    766
 217   R   N   D   D   P   A   T   A   S   R   P   F   D   R   T   R    232
 767  GAC GGA TCG TGC TCG GCG AGG GCG CCG CCG TCC TGG TCC TGG AAG AAC    814
 233   D   G   S   C   S   A   R   A   P   P   S   W   S   W   K   N    248
 815  TCG AAC ACG CCC GCC GCC GCG GCG GCC CGC CCC TAC GCG GAA CTG TCC    862
 249   S   N   T   P   A   A   A   A   A   R   P   Y   A   E   L   S    264
 863  GCC GTC ACC GCC CAC AGC AGC GCC CAC CAC ATG ACG GGA CTG CGC CCC    910
 265   A   V   T   A   H   S   S   A   H   H   M   T   G   L   R   P    280
 911  GGG GCA CTG GAG ATG GCC GAC GCC ATC CGC GCC GCC CTC GAC CAG GCA    958
 281   G   A   L   E   M   A   D   A   I   R   A   A   L   D   Q   A    296
 959  CGG CTG AAC CCC GCC GAC GTC GAC TAC ATC AGC GCC CAC GGC GCG GGA   1006
 297   R   L   N   P   A   D   V   D   Y   I   S   A   H   G   A   G    312
1007  ACC CGG CAC AAC GAC CGG CAC GAG ACA CAC GCC CTC AAG GAA AGC CTG   1054
 313   T   R   H   N   D   R   H   E   T   H   A   L   K   E   S   L    328
1055  GGC GGC AGC GCC CAC CGC GTG CCC GTC AGC TCC ATC AAG AGC ATG ATC   1102
 329   G   G   S   A   H   R   V   P   V   S   S   I   K   S   M   I    344
1103  GGG CAC AGC CTG GGC GCC GCC GGC GCC CTG GAA CTG GCC GCC AGC GCC   1150
 345   G   H   S   L   G   A   A   G   A   L   E   L   A   A   S   A    360
1151  CTG GCC ATC CGG CAC GAC ACC GTC CCG CCC ACC GCC AAC CTC CAC GAA   1198
 361   L   A   I   R   H   D   T   V   P   P   T   A   N   L   H   E    376
1199  CCC GAG CCC CAC TGC ACC AGC ACC GTC CTC ACC GTC GCC AGC GGC TTC   1246
 377   P   E   P   H   C   T   S   T   V   L   T   V   A   S   G   F    392
1247  GGC GGC TCC ACA CGG CCG CCG TAT CTC ACC CGG CCG GTG CTC AAG GAG   1294
 393   G   G   S   T   R   P   P   Y   L   T   R   P   V   L   K   E    408
1295  GCCGCATGACC AGG CGA CCC AGG CCC CCG AAC CGG CCC CGC CCC GCC GGC   1344
 409   A   A   *                                                         411
   1           M   T   R   R   P   R   P   P   N   R   P   R   P   A   G     15
1345  GCC GAA GGC CCC CGC CAG ACC CTG GTC ACC GGC ATC GGA GTC GCC GCA   1392
  16   A   E   G   P   R   Q   T   L   V   T   G   I   G   V   A   A     31
1393  CCC AAC GGC CTG GGC ACC AGA GCC TGG TGG GAC GCC GTG CTG TGC GGG   1440
  32   P   N   G   L   G   T   R   A   W   W   D   A   V   L   C   G     47
1441  CGC ACC GGA CTG GGC CCC ATC ACC CGC TTC GAC CGT CGG CTA CCC CCG   1488
  48   R   T   G   L   G   P   I   T   R   F   D   R   R   L   P   P     63
1489  GTA CGC ATC GCC CGC GAG ATC CCC GGC TTC GTC GAC GAG GAC CAC ATC   1536
  64   V   R   I   A   G   E   I   P   G   F   V   D   E   D   H   I     79
1537  CCC AGC AGA CTG CTG CCC TCC ACC GAC CGC GGC ACC CGC ATC GCC CTG   1584
  80   P   S   R   L   L   P   S   T   D   R   G   T   R   I   A   L     95
1585  GTC GCG GCC GAA CAA GCA CTG CGC GAC GCT AAC GTG AGC CCG GCC GAC   1632
  96   V   A   A   E   E   A   L   R   D   A   N   V   S   P   A   D    111
1633  CTG CCC GCA TAC GGC GCC GGC GTG ATC ACC GCC AGC TCC GCG GGC GGC   1680
 112   L   P   A   Y   G   A   G   V   I   T   A   S   S   A   G   G    127
1681  GCC GAA TTC GGC GAA CGG GGA CTG GCC GCA CTG TGG AGC AAA GGC GCC   1728
 128   A   E   F   G   E   R   G   L   A   A   L   W   S   K   G   A    143
1729  CAG CAC GTC AGC GCC TAC CAG TCC TTC GCG TCC TTC CAC GCG GCA GCC   1776
 144   Q   H   V   S   A   Y   Q   S   F   A   S   F   H   A   A   A    159
1777  CCC GCA CAG ATC TCC ATC CGG CAC CGG CTG CGC GGC CAC GGC TCG ACC   1824
 160   P   A   Q   I   S   I   R   H   R   L   R   G   H   G   S   T    175
1825  GTC GTC AGC GAA CAG GCC GGC GGC ATC GAC GCA CTC GCC CGC GCC CGG   1872
 176   V   V   S   E   Q   A   G   G   I   D   A   L   A   R   A   R    191
1873  CGG CGG ATC CGC GAC GGG GCA TGC CTC ATG GTC ACC GGC GGG ATC GAC   1920
 192   R   R   I   R   D   G   A   C   L   M   V   T   G   G   I   D    207
1921  TCC ACA CTG TGC GCA TGG GGC TGG GCC GCG CAC CTG GCG GAC GGC CGG   1968
 208   S   T   L   C   A   W   G   W   A   A   H   L   A   D   G   R    223
1969  TTC AGC CCC GCC ACC GAA CCC GCC CGG GCC TAC CGG CCC TTC GCG GCC   2016
 224   F   S   P   A   T   E   P   A   R   A   Y   R   P   F   A   A    239
2017  ACG GCC GAC GGC CAC GCG GTC GGC GAG GGC GGC GCC CTG CTC GTC CTG   2064
 240   T   A   D   G   H   A   V   G   E   G   G   A   L   L   V   L    255
2065  GAG GAC GCC CGG GCC GCC GCC CGC CGG GGC GCC ACC GGC TAC GGC GTC   2112
 256   E   D   A   R   A   A   A   R   R   G   A   T   G   Y   G   V    271
2113  ATC GCC GGC TGC GCC GCC ACC TTC GAC GGC CCC GAC CGC CCC ACA CTG   2160
 272   I   A   G   C   A   A   T   F   D   G   P   D   R   P   T   L    287
2161  CGC CAG GCC GCG GAA CTC GCC CTG GCC AAG GCC GGC CTG GCC CCC GAA   2208
 288   R   Q   A   A   E   L   A   L   A   K   A   G   L   A   P   E    303
2209  CAC GTG GAT GTG GTT TTC CCC CGA ACG GCC CCC GAG CGG CGC GCC GAC   2256
 304   H   V   D   V   V   F   P   R   T   A   P   E   R   R   A   D    319
2257  CTC GTC GAG AGC CAG GCG CTG TGC GCG CTG TTC GGA CCC TAC GGA GTA   2304
 320   L   V   E   S   Q   A   L   C   A   L   F   G   P   Y   G   V    335
2305  CCG GTC ACC GTG CCG AAG ACG ATG ACC GGG CGG CTG GGC GCG GGC GGC   2352
 336   P   V   T   V   P   K   T   M   T   G   R   L   G   A   G   G    351
2353  TCG GCC CTG GAC GTG GCA GCC GCG CTG CTC GCC CTG CGC GAG AAG GTC   2400
 352   S   A   L   D   V   A   A   A   L   L   A   L   R   E   K   V    367
2401  GTA CCC CCG ACC ACC GGA ACC GGA CGC GTC GCC GAC GAC TGC CCG CTG   2448
 368   V   P   P   T   T   G   T   G   R   V   A   D   D   C   P   L    383
2449  GAC CTG GTC ACC GGA GCC CCA CGG GAC TGC CCC GGC TGC GGG TGG CGC   2496
 384   D   L   V   T   G   A   P   R   D   C   P   G   C   G   W   R    399
2497  TGG GTA CTG GCC CGC GGA CGG GGC GGC TTC AAC TCC GCC GCG GTG CTC   2544
 400   W   V   L   A   R   G   R   G   G   F   N   S   A   A   V   L    415
2545  CGG GCC CCT CAG ACG CAG TGA CGG CCC GCC CCT AAC GAG ACG GAG CGG   2592
 416   R   A   P   Q   T   Q   *                                         422
2593  AAC ACA CGG GCG GTC AGG GAC GCC CGC CCC GCG GGC CGC GCA AAG GGA   2640
2641  AAC ACC CCC GGC GCC GAC TCC CCG GCC CTC CCG GCC CTC CTG GCG CCT   2688
2689  TCC GCC TTC CGG CTT TCG CCT TCC GGC CTC CTG CCT TCC GGC CT        2732
```

Fig. 3. Nucleotide sequence of a 2,732-bp region encoding *salB2* and *salC2*. The predicted peptide sequence is shown below the nucleotide sequence. Putative ribosome binding sites are underlined.

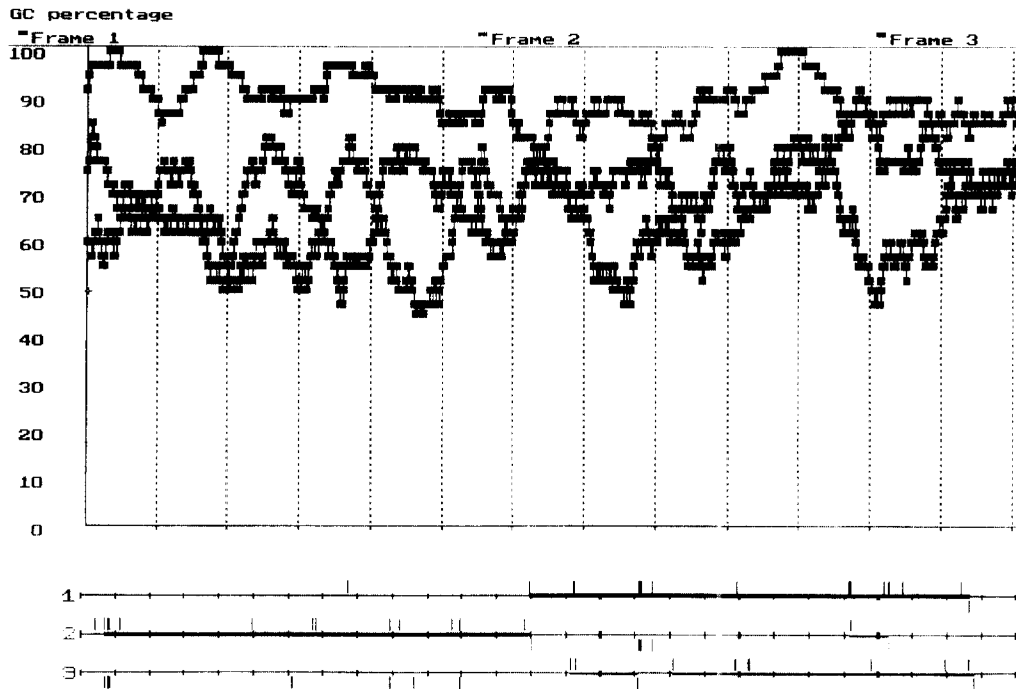

Fig. 4. Analysis of guanine and cytosine content at third positions, for 2,732-bp with the *salB2* and *salC2* genes of pWHM220. ATG and GTG codons are indicated by vertical bars, stop codons are designated by crosses.

422 amino acids, respectively, and a molecular weight of 43,621 and 43,738, and a calculated isoelectric point of 7.72 and 9.60, respectively. The gene organization for the two genes seemed to be translationally coupled.

Deduced function of the proteins

The two new genes cloned in pWHM220 were proved to have homology with *tcmKL* by sequence analysis. TcmK and TcmL compose minimal PKS for tetracenomycin biosynthesis (7). The separate components of the minimal PKS are ketoacyl synthase (KS) which catalyzes the condensation reaction and may also carry an acyltransferase (AT) domain for loading the starter unit to the KS, a second protein with a high degree of homology to the KS which is involved, at least in part, in determination of the polyketide chain length, and an acyl carrier protein (ACP).

The two open reading frames were identified by CODONPREFERENCE and named *salB2*, *salC2*. Sequence comparisons with GAP program (10) indicated that SalB2 polypeptide bears strong similarity to many other β-ketoacyl synthases. The similarity is especially strong in the region surrounding 169-Cys and 347-Ser, which corresponds to the

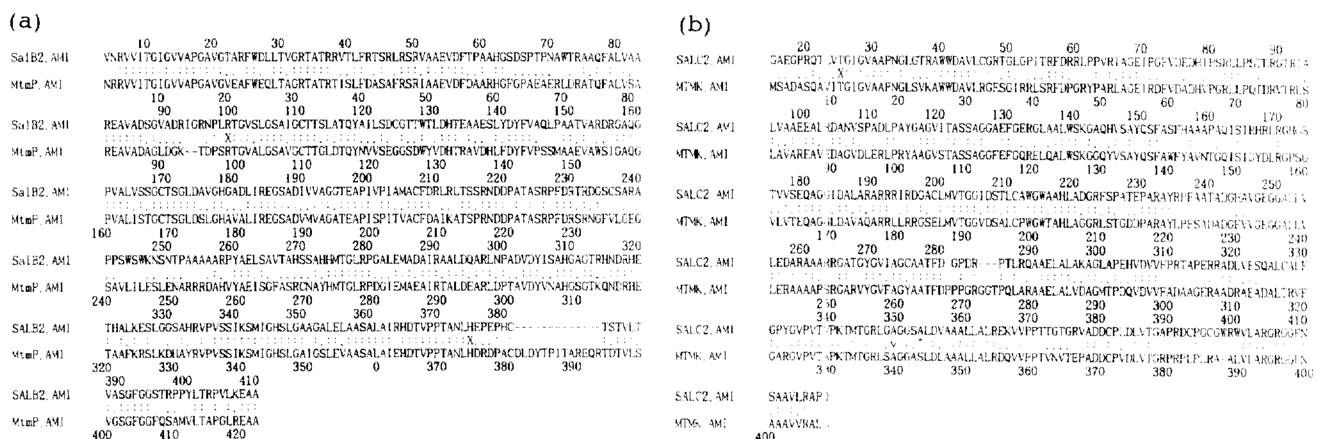

Fig. 5. Amino acid sequence comparisons using Genetic Computer Group program. (a) SalB2 and MtmP, β-ketoacyl synthase for mithramycin biosynthesis (59.1% identity); (b) SalC2 and MtmK, chain length factor from *S. argillaceus* (62.3% identity).

A. β ketoacyl synthase active site

		Ref
S. albus SalB2	RGAQGPVALVSSGCTSGLDAVGHGA	this study
S. argillaceus MtmP	IGAQGPVALISTGCTSGLDSLGHAV	22
S. venezuelae JadA	VGAEGPNTVVSTGCTSGLDSVGYAR	15
S. glaucescens TcmK	AGAEGPVTVVSTGCTSGLDAVGYGT	7
S. griseus Gri	VGAEGPATVVSTGCTAGIDAVGHAV	35
S. rimosus Otc	AEAEGPAGVVSAGCTSGIDVLTHAA	21
S. violaceoruber Gra	AGAEGPVTMVSDGCTSGLDSVGYAV	32
S. peucetius Dau	AGAEGPVNIVSAGCTSGIDSIGYAC	14
S. coelicolor Act	VGAEGPVTMVSTGCTSGLDSVGNAV	12
S. albus SalB	VGAEGPSTVVSTGCTSGLDSVGYAV	8
Active site system	*A*GP *S GCT*G ** *	

B. Acyltransferase active site

		Ref
S. albus SalB2	PVSSIKSMIGHSLGAAGALELAASAL	this study
S. argillaceus MtmP	PVSSIKSMIGHSLGAIGSLEVAASAL	22
S. venezuelae JadB	PVSSIKSMVGHSLGAIGSIEIAASAL	15
S. glaucescens TcmK	PVSSIKSMIGHSLGAIGSLELAACAL	7
S. griseus Gri	PVSAIKSMVGHSLGAIGSIEIAACAL	35
S. rimosus Otc	PISSIKSMIGHSLGAICALEVAASAL	21
S. violaceoruber Gra	PVSSIKSMGGHSLGAIGSIEIAASVL	32
S. peucetius Dau	PISSIKSMIGHSLGAVGSLEVAATAL	14
S. coelicolor Act	PVSSIKSMVGHSLGAIGSLEIAACVL	12
S. albus SalB	PVSSIKSMVGHSLGAIGSIEIAASAL	8
Active site system	P*S**KSM GHSLGA **E*AA* L	

Fig. 6. Sequence alignment of amino acids of β-ketoacyl synthase (KS) and acyltransferase (AT) with conserved active sites of other KS and AT, respectively. Conserved amino acids for all peptides are shown under each alignment. Amino acids at active sites are marked in box. References for each peptide are shown.

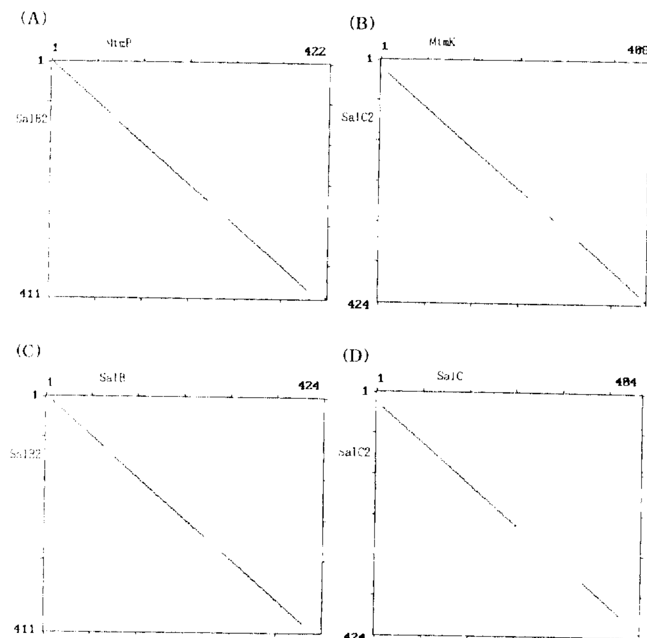

Fig. 7. COMPARE-DOTPLOT analysis of the deduced protein products of (A) SalB2 versus MtmP, 59.1% identity; (B) SalC2 versus MtmK, 62.3% identity ; (C) SalB2 versus SalB, 54.9% identity; (D) SalC2 versus SalC, 54.8% identity. A window size of 30 was used at a stringency of 15.

active site in β-ketoacyl synthase enzyme (Fig. 6).

The deduced *salC2* product resembles several

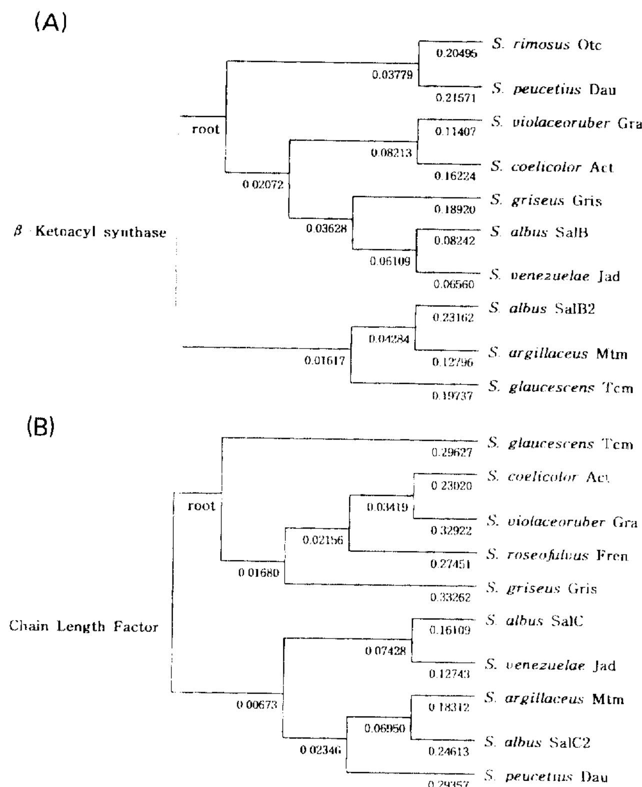

Fig. 8. Dendrogram showing relationahips of type II PKS components, β-ketoacyl synthase and chain length determining factor. PILEUP program of GCG (10) was used to analyze the data.

chain length factor with the following percentages of identical amino acids (Fig. 4, 7): MtmK for mithramycin biosynthesis, 62.3% (22), JadB for jadomycin biosynthesis, 55.1% (15), and SalC for polyketide biosynthesis in *S. albus*, 54.8% (8).

All of the genes are transcribed in the same direction. A phylogenetic tree, involving a large set of sequence including those of *salB2* and *salC2*, is shown in Fig 8. Recent studies have shown that combination of genes from different PKSs can be expressed to produce functional hybrid PKS. The PKS gene set identified in this study is an important addition to the range of available PKSs and will enable further analysis of the molecular basis for PKS programming to be made.

Relationship between the two polyketide gene clusters

We had previously shown that *S. albus* has a gene cluster (*salA-E*) encoding type II PKS. And this region could complement mutations in the *tcmKL* region (8). In this study, we showed that pWHM220 cosmid clone had another copy of genes for β-ketosynthase and chain length factor. The presence of multiple copies of the same or similar genes in

Streptomyces is not usual but not odd(11, 23).

Here we report a second set of PKS genes (*salB2, salC2*) in *S. albus* that are similar to another set, the *salBC* genes from another PKS gene cluster (8); *salB2* encodes the β-ketoacyl synthase and *salC2* encodes chain length factor in the same orientation. Since the original PKS gene cluster and the additional PKS gene cluster are close each other (only 10-kb apart) and have the same orientation, we speculate that these two gene clusters might have arisen by tandem genetic duplication (2).

However, we do not exclude that the two PKS gene clusters have different functions. Since *act*- and *whiE*-PKS genes, the two sets of related genes, are expressed at different stages in the life-cycle of the organism, *S. coelicolor*, to produce antibiotic and spore pigments, respectively. It was found that each of the three subunits of the *whiE* minimal PKS could complement lesions in the *act*-PKS to produce actinorhodin. Conversely, the corresponding *act*-PKS subunits would complement mutations in the *whiE* locus to restore spore pigmentation (34).

Acknowledgments

This work was supported by a grant of Genetic Engineering Research Program of the Ministry of Education, Korea, in 1995.

References

1. **Altchul S.F., W. Gish, W. Miller, E.W. Myers, and D. J. Lipman.** 1990. Basic local alianment search tool. *J. Mol. Biol* **215**, 403-410.
2. **Anderson, R.P. and J.R. Roth.** 1977. Tandem genetic duplication in phage and bacteria. *Annu. Rev. Microbiol.* **31**, 473-506.
3. **Arrowsmith, T.J., F. Malpartida, D.H. Sherman, A. Birch, D.A. Hopwood. and J. A. Robinson.** 1992. Characterization of actI-homologous DNA encoding polyketide synthase genes from the monensin producer *Streptomyces cinnamonensis. Mol Gen Genet* **234**, 254-64.
4. **Ben Shen and C. Richard Hutchinson.** 1993. Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme *A. Science* **262**, 1535-1540.
5. **Bibb, M.J., D.H. Sherman, S. Omura, and D.A. Hopwood.** 1994. Cloning, sequencing and deduced functions of a cluster of *Streptomyces* genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. *Gene* **142**, 31-39.
6. **Bibb, M.J., P.R. Findlay, and M.W. Johnson.** 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* **30**, 157-66.
7. **Bibb, M.J., S. Biro, H. Motamedi, J.F. Collins, and C.R. Hutchinson.** 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.
8. **Borell, Chris W., Y.G. He, H.J. Kwon, C.G. Hyun, K. Fumio, J.W. Suh, H.J. Jin, H.G. Floss and C.R. Hutchinson.** 1996. Production of antibiotics salinomycin by introduction of a gene cluster containing type II polyketide synthase genes from *Streptomyces albus* ATCC 21838 into *Streptomyces lividans. Microbiology* (UK) in submitted.
9. **Davis, N.K. and K.F. Chater.** 1990. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol.* **4**, 1679-1691.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.
11. **Donadio, S., M.J. Staver, J.B. McAlpine, S.J. Swanson and L. Katz.** 1991. Modular organizaion of genes required for complex polyketide biosynthesis. *Science* **252**, 675-679.
12. **Fernandez-Moreno, M.A., E. Martinez, L. Boto, D.A. Hopwood, and F. Malpartida.** 1992. Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin *J. Biol. Chem.* **267**, 19278-19290.
13. **Gish, W. and D.J. States.** 1993. Identification of protein coding regions by database similarity search. *Nature Genet* **3**, 266-272.
14. **Grimm, A., Madduri, K., Ali, A. and Hutchinson, C. R.** 1994. Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151**, 1-10.
15. **Han, L., Yang, K., Ramalingan, E., Mosher, R.H. & L.C. Vining.** 1994. Cloning and characterization of polyketide synthase genes for Jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. *Microbiology* **140**, 3379-3389.
16. **Hopwood, D.A., C. Khosla, D.H. Sherman, M.J. Bibb, S. Ebert-Khosla, E.-S. Kim, R. McDaniel, W.P. Revill, R. Torres, T.-W. Yu.** 1993. Toward an understanding of the programming of aromatic polyketide synthases a genetics-driven approach, p. 267-275. *In* G. D. Hegeman, R.H. Baltz, and P.L. Skatrud (ed.), Genetics and molecular biology of industrial microorganisms. American Society for Microbiology, Washington, D. C.
17. **Hopwood, D.A., and D.H. Sherman.** 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24**, 37-66.
18. **Hutchinson, C.R. and I. Ikeda, I.** 1995. Polyketide synthase gene manipulation; A structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **49**, 201-238.
19. **Khosla, C., R. McDaniel, S. Ebert-Khosla, R. Torres, D.H. Sherman, M.J. Bibb, and D.A. Hopwood.** 1993. Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins. *J. bacteriol.* **175**, 2197-2204.
20. **Kim, E.S., D.A. Hopwood, and D.H. Sherman.** 1994.

Analysis of type II β-ketoacyl synthase specificity in *Streptomyces coelicolor* A3(2) by *trans* complementation of actinorhodin synthase mutants. *J. Bacteriol.* **176**, 1801-1804.

21. **Kim, E.S., M.J. Bibb, M.J. Butler, D.A. Hopwood, and D.H. Sherman.** 1994. Sequences of the oxytetracycline polyketide synthase-encoding *otc* genes from *Streptomyces rimosus*. *Gene* **141**, 141-142.
22. **Lombo, F., G. Blanco, E. Fernandez, C. Mendez, and J.A. Salas.** 1996. Characterization of *Streptomyces argillaceus* genes encoding a polyketide synthase involved in the biosynthesis of the antitumor mithramycin. *Gene* **172**, 87-91.
23. **Magdalena, J., M. Forsman, M.V. Lenzini, A. Brans, and J. Dusart.** 1992. Two different β-lactamase genes are present in *Streptomyces cacaoi*. *FEMS Microbiol. Lett.* **99**, 101-106.
24. **McDaniel, R., S. Ebert-Kosla, D.A. Hopwood, and C. Khosla.** 1993. Engineered biosynthesis of novel polyketides. *Science* **262**, 1546-50.
25. **McDaniel, R., S. Ebert-Kosla, D.A. Hopwood, and C. Khosla.** 1993. Engineered biosynthesis of novel polyketides: manipulation and analysis of an aromatic polyketide synthase with unproven catalytic specificities. *J. Am. Chem. Soc.* **15**, 11671-11675.
26. **McDaniel, R., S. Ebert-Kosla, D.A. Hopwood, and C. Khosla.** 1995. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375**, 549-54.
27. **Richardson, M. A., S. Kuhstoss, P. Solenberg, N. Schaus, and R.N. Rao.** 1987. A new shuttle vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a *Streptomyces ambofaciens library*. *Gene* **61**, 231-241.
28. **Sambrock, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. 2nd ed., Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N.Y.
29. **Sanger, F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.
30. **Schweche, T., J.F. Aparicio, I. Molnar, A. Konig, L. E. Khaw, S.F. Haydock, M. Oliynyk, P. Caffrey, J. Cortes, J.J.B. Lester, G.A. Bohm, J. Staunton and P. F. Leadley.** 1995. The biosynthesis gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **93**, 7839-7843.
31. **Sherman, D.H., E.-S. Kim, M.J. Bibb, and D.A. Hopwood.** 1992. Functional replacement of genes for individual polyketide synthase components in *Streptomyces coelicolor* A3(2) by heterologous genes from a different polyketide pathway. *J. bacteriol.* **174**, 6184-6190.
32. **Sherman, D.H., F. Malpartida, M.J. Bibb, H.M. Kieser, and D.A. Hopwood.** 1989. Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. *EMBO J.* **8**, 2717-2725.
33. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors *Gene* **33**, 103-119.
34. **Yu, T.W. and D.A. Hopwood.** 1995. Ectopic expression of the *Streptomyces coelicolor whiE* genes for polyketide spore pigment synthesis and their interaction with the *act* genes for actimorhodin biosynthesis. *Microbiology* **141**, 2779-2791.
35. **Yu, T.W., M.J. Bibb, W.P. Revill, and D.A. Hopwood.** 1994. Cloning, sequencing, and analysis of the griseusin polyketide synthase gene cluster from Streptomyces griseus. *J Bacteriol.* **176**, 2627-34.