Chapter - 6

Nucelotide Sequencing
&
In silico Analysis

INTRODUCTION

Sequencing of the type II polyketide synthase cluster from *S. flavisclerotícus* was initiated in 2003, with only 2.8kb of DNA containing the most conserved Ketosynthase and Chain Length Factor deciphered, along with a partial sequence of O-methyl transferase gene. Parallely Mendez-Salas group in Oviedo, Spain in collaboration with Rohr group in Kentucky, USA, sequenced and published the whole cluster of chromomycin synthase gene cluster from *S. griseus* sub *griseus* (Mendez et al, 2004a). The minimal PKS sequence determined by our lab, in cluster from *S. flavisclerotícus,* was submitted to Pubmed prior to the release of sequence from *S. griseus.*

The paper had received commentary with marked attention, due to the heavy similarity between the chemical structures of chromomycin and another aureolic acid molecule-mithramycin. It was expected that the two clusters would be nearly identical. But it was found that, however there was high homology between individual mithramycin and chromomycin genes, the arrangement of the genes in the chromomycin cluster was substantially different. Owing to this phenomenon, a significant alteration in the sequences of two chromomycin producers, namely *S. griseus* and *S. flavisclerotícus* could not have been ruled out.

Also, it was proposed that if either mithramycin or chromomycin was acquired by horizontal gene transfer, the differences in the two clusters suggest that substantial gene reorganization must have taken place and that the driving force for such reorganization remains unclear. Interestingly, sequencing of cluster from *S. flavisclerotícus* provided, what can be considered as, a connecting link between the two clusters. A transposon appears in the cluster of *S. flaviscerloticus* which is absent in cluster from *S. griseus.* Though placed within cluster, the transposon is not involved in either chromomycin synthesis or resistance. Thus, due to the significance of origin and role of this transposon, it has been discussed separately in detail at the end of the chapter.

Putative proteins from *S. griseus* chromomycin cluster, involved in biosyntheis of the molecule, were assigned an alphabet with prefix *'cmm'* while those involved in

resistance had a prefix '*cmr*'. We have used prefix '*sfl*' (for *S. flavisclerotícus*) and '*sfr*' for genes of biosynthesis and resistance respectively, maintaining the alphabets assigned in *S. griseus* cluster.

Owing to the specificity of the cluster, a direct comparision of Sfl -cluster has been made with Cmm -cluster, along with the comparision with individual genes from different clusters. Thus in this chapter attempt has been made to compare the chromomycin cluster of *S. flavisclerotícus* in terms of structural organization (cluster level, operon level and gene level), functional organization (regulation, biosynthesis of chromophore & sugar moiety, time and order of activation of different genes), and evolutionary point of view.

STRATEGY

For DNA sequence determination, purified plasmid clones were sent in the form of single colony streaked culture, for which DNA isolation and sequencing was carried out by Sanger Dideoxy Method, on payment basis at Microsynth, Switzerland.
The strategies attempted for cloning and subsequent nucleotide sequencing of various fragments of cluster are as follows:

6.2.1 Primer Walking

Insert DNA was cloned in the homologous sites of pBluescriptKS. T3 –T7 phage promoter sites present in the plasmid flanking the MCS, were used as primers to initiate sequencing reaction. Primers required at both ends for subsequent reactions were generated from the new end of sequencing result of previous reactions. Sequencing was discontinued once the overlap was detected in the sequences.

6.2.2 Transposon mediated sequencing

Tn*1000,* also known as γδ, is replicative type of transposon of 5.7 kb. It is naturally found on the *E.coli* chromosome and on the conjugative plasmid F (Guyer, 1978); the experimental use of γδ described here used F as the transposon donor. γδ mediates transfer of non-conjugative plasmid via forming a co-integrate through the replicative transposition, which is converted by site-specific recombination in the recipient to the original F^+ factor plus the plasmid carrying one copy of γδ as described in fig 6.2.

Since the event is presumed to be reasonably random, clones with insertions at varied positions can be easily obtained that can be used for further sequencing. pSET (5.5kb) digestion with *Pst*I and subsequent intramolecular ligation could eliminate ~3kb DNA comprising of att-site and integrase gene not required for this experiment. This exercise would reduce the probablity of transposon insertion in backbone. p11EBSK

fragment of the cluster was cloned in reduced pSETΔ*Pst* to form p11EBT. The DNA was transformed into JP3301, which carries transposon Tn*1000* on a resident F^+. Conjugational cross was set up with DH5α as shown below (protocol in chapter 2).

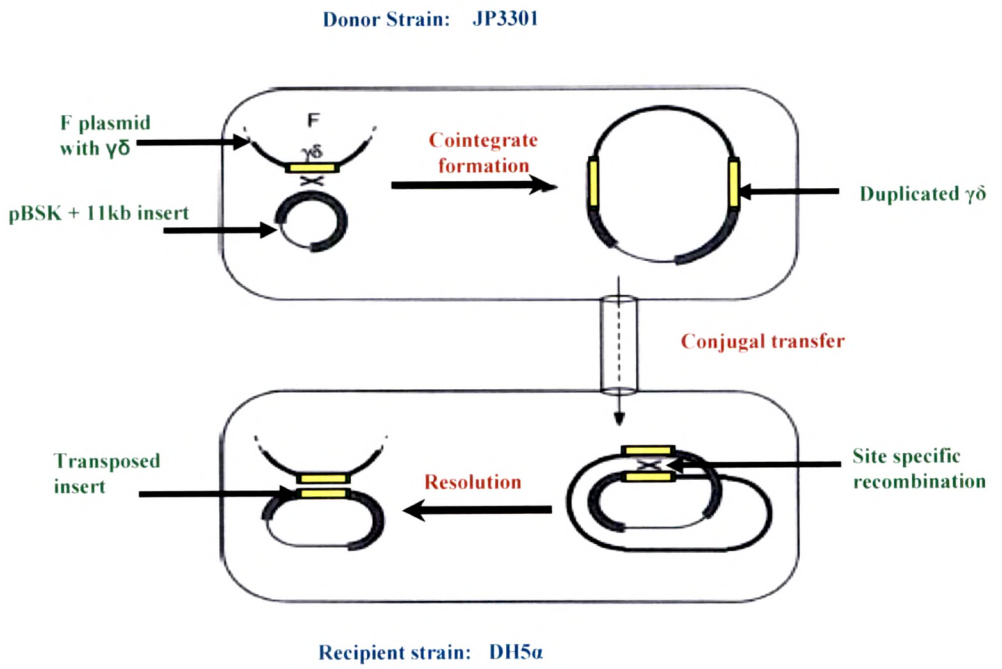


Fig 6.1: Diagrammatic representation of the conjugation mediated transposition process furnished by the presence of F plasmid containing Tn*1000*

---

p11EBT JP3301 (F^+ *recA*$^-$) X **DH5α** (F^- Nal^R *recA*$^-$)

DONOR RECEIPIENT

F is expected to form co-integrate with p11EBT and will move into recipient by conjugal transfer. Selection on media supplemented with nalidixic acid and apramycin would select against donor and recipient cells respectively.

Each colony obtained thereby has to be with a transposed plasmid, since it cannot undergo conjugal transfer without getting integrated with F in JP3301 and later get

A.

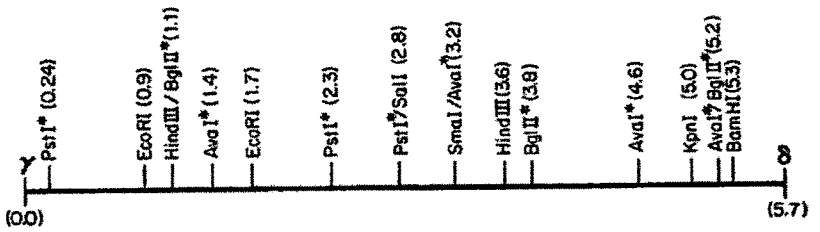


Fig 6.2: A. Physical map of Tn*1000* generated by Schwacha et al, 1990.

---

resolved by homologous recombination. Both these events are transposon mediated steps.

Inverted primers were designed from the end sequences of Tn*1000* that read in outward directions. This would enable direct sequencing of the insert. Ten purified colonies at random were sent for sequencing.

RESULT AND DISCUSSION

6.3.1 Nucleotide sequencing:

Eight PKS^+ clones from cosmid library with ~20kb fragment of genomic DNA of *S. flaviscleroticus* was initial source of DNA. A physical map with respect to four restriction enzymes generated in earlier work in our lab was source of DNA for sequencing purpose. Sequencing was done using *Eco*RI fragments which were in the order of 5kb, 11kb, 4kb, 8kb and 17kb in size from one end (Fig- 6.3).

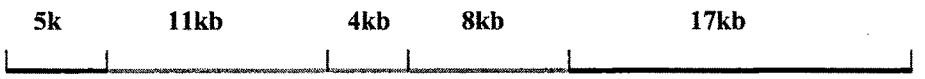


Fig- 6.3: Diagramatic representation of *Eco*R1 restriction sites in physical map of PKS DNA

For sequencing of 8kb fragment, sub-clones generated earlier were used. Subclones were generated with a view that it will involve handling of more than one clone at a time and so parallel sequencing of these clones would ensure faster processing of the objective against primer walking where only one reaction can be carried out at a time making it relatively slower process. The other advantage was in more frequent use of non-charged universal primers (T3-T7) against custom oligosynthesis for primer walking. But later it was realized that subclone generation not only consumes resource and time, but is not full-proof as well. There is always a chance that a very small fragment (<150 bases), not easily detectable on agarose gel, might get left out. To take care of this, single reaction of the 8kb clone at every subclone connection, using specific primers, had to be done. Thus, directly primer walking was used for sequencing 4kb and 17kb. Shot gun sequencing could have been attempted for 17kb fragment but for the cost intensive nature of the strategy, it was not attempted.

11kb *Eco*RI fragment was unique with flanking 4kb*Eco*R1 DNA downstream to it and a 5kb*Eco*R1 fragment upstream. Nucleotide sequencing of 5kb*Eco*R1 DNA with its T3 and T7 ends did not reveal any significant homology to type II PKS cluster genes. ORF of cyclase is split between 4kbE and 11kbE. 5' end of cyclase ORF is in 4kbE and 3' end in 11kbE. The rest of 11kbE DNA also does not contain Type II PKS genes but contain imperfectly conserved NRPS genes. Owing to the large size of the DNA and the aforementioned reason, transposon mediated sequencing technique was employed for it.

11kb *Eco*R1 was cloned in pSETΔpst. Routinely used pBSK was deliberately avoided in this experiment since it has been found in our lab, not to be suitable for such conjugation process as the satellite colonies very rapidly overgrow on ampicillin to develop into a lawn, making exconjugants indistinguishable. Sequencing results revealed the incomplete putative cyclase of 4kb fragment that continued in 11kb was infact the last gene of the cluster. Rest of the DNA appeared to belong to Non-Ribosomal Peptide Synthase (NRPS) cluster.

Tn*1000* showed a very high bias towards backbone DNA. Even though the insert (11kb) was much larger than backbone, out of ten clones sent for sequencing at random, seven had transposition in backbone. Further, this becomes more conspicuous since transposition event in ori and resistance gene regions are selected against.

Tn*1000* is of non-GC rich E. coli origin (Schwacha et al, 1990), while the insert DNA was from GC rich *Streptomyces* sp. Results suggested that Tn*1000* could be acting on DNA with nucleotide bias. Literature survey later revealed that Tn*1000* has high preference for AT rich five base pair DNA (Liu et al, 1987). Also, this transposon has been successfully used in past for other GC rich genomes, without any modification with *Drosophila melanogaster* (Strausbaugh et al, 1990) and its derivative in *S. avermitilis* (Berg et al, 1992). Atleast in Drosophila it was found that whenever transpositioning occurred in GC rich regions, they were invariably in the AT valleys of the GC rich DNA.

6.3.2 Cluster Organization:

The organization of genes in this polyketide cluster is complex. There are genes that function solitary as well as there are genes that are organized in operons. These genes/operons are convergently as well as divergently transcribed. Also, functionally related genes are placed in separate operons while those not having close functional relevance have been grouped in one operon. For better understanding of the cluster the analysis has been done, taking into consideration the structural organization as well as functional attributes of the components.

6.3.3 Structural organization of the cluster:

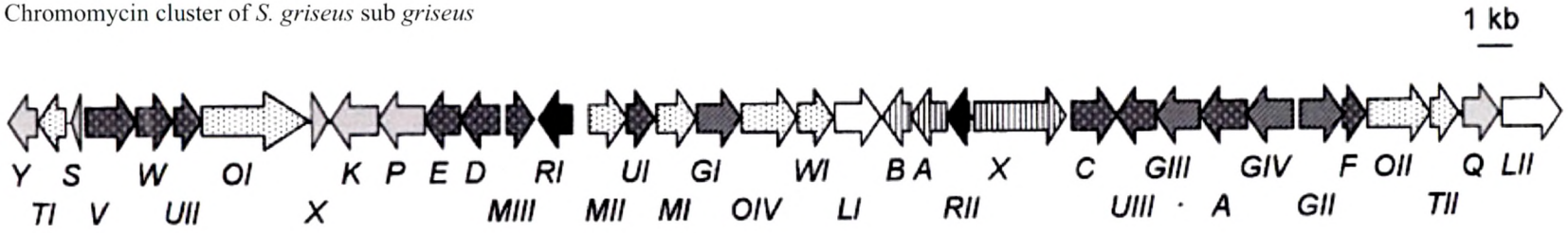
Eight overlapping cosmid clones, each with ~25kb DNA, represented over 45kb genomic DNA. Sequencing of 17kbE fragment shows an incomplete gene with high homology to *cmm*UIII, this would mean that the cluster would continue further downstream. Sequencing of 11kb*Eco*R1 fragment reveals putative cyclase to be the last gene of the cluster at one end, since ORFs upstream of cyclase did not show homology to known gene of aromatic polyketide biosynthetic pathway. Immediately upstream of cyclase is a small patch of DNA that shows homology with a putative transposase. Metallopeptidase and genes showing high homology to non-ribosomal Peptide Synthase (NRPS) cluster were present further upstream.

6.3.4 Operons:

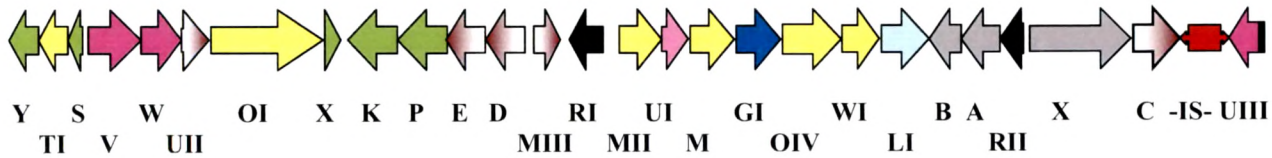
There are several proteins that play important roles in either synthesis or maturation of a polyketide molecule. The genes of these proteins are placed in close proximity to form clusters. The cluster contain genes independently as well as they are grouped together to form operon and are transcribed by a common promoter.

There are 26 complete and one incomplete open reading frames that show significant homology with type II polyketide synthase cluster of other producers. These genes are present in more than one operon. Surprisingly, only two genes appear to have

Chromomycin cluster of *S. griseus* sub *griseus*



Truncated Chromomycin Cluster from *S. flaviscleroticus*



Mithramycin cluster of S. *argillaceus*

Fig-6.4 : The Identical Cluster organization of two chromomycin producers compared to a much different cluster of structurally related Mithramycin

their independent promoters. Sequencing of ~30kb DNA reveals presence of alteast seven operons in this cluster. Six operons are complete and the seventh has been hypothesized based on the orientation of last gene.

Although there is significant similarity between structures of chromomycin and mithramycin which is also reflected in significant similarity between genes, the arrangement of the genes in the chromomyin cluster is organized in a substantially different way. A dramatic rearrangement in the order of the genes still results in successful expression of the aureolic acid antibiotic (O'connor, 2004). Mithramycin cluster appears to possess a relatively simple organization, wherein the aglycone polyketide chormophore synthesis genes are present in separate operons and the sugar synthesis genes in other. Thus, both the components are probably synthesized independently. As against this, chromomycin cluster appears to be disorganized and without any pattern, at first sight, with unrelated genes going together in same operon while the related ones present in separate operons. But a closer look into the cluster provides an insight about the co-ordination in the synthesis process exhibited by the cluster. Thus, contrary to the strategy of mithramycin cluster of grouping functionally related genes that perform similar activities and carry out synthesis of components independently, chromomycin cluster seem to have adopted policy of temporal regulation (grouping genes required for a particular stage of), thereby forming operons for each stage of synthesis irrespective of their relevance with each other. For example, in mithramycin synthesis genes of aglycon (*mtm*X, *mtm*P, *mtm*K, *mtm*S, *mtm*OI, *mtm*TI and *mtm*Y) and sugar moiety (*mtm*Z, *mtm*A, *mtm*H, *mtm*D and *mtm*E) form separate operons and remain functionally independent of each other, whereas in chromomycin it is ensured that synthesis of both these components initiate simultaneously and so the starter genes for sugar synthesis (*sfl*E, *sfl*D) and chromophore synthesis (*sfl*P, *sfl*K) have been probably placed in the same operon. Though SflS is an essential part of minimal PKS, it has not been included in this operon. It has been coupled with the cyclase (*sfl*Y) in a separate operon since cyclase is immediately required once carbon skeleton has been formed. Further, *cmm*V, *cmm*W and *cmm*UII have been coupled with *cmm*OI and *cmm*X as they are involved in the next step of sugar synthesis and chromophore synthesis respectively. Genes involved in subsequent maturation and modification of both these components is evenly distributed in operons.

Table-6.1: Comparative account of relevance of ORFs with the analogous protein in chromomycin cluster of *S. griseus* sub. griseus and other proteins based on homology search.

This study		Chromomycin Equivalent			Closest match from database			
Orf	a.a.	Name	a.a.	Homology	Organism	protein	a.a.	Homology
*sfl*UIII*	-	NDP-4-ketoreductase (*cmm*UIII)	356	-	-	-	-	-
*sfl*C	410	NDP-C methyl transferase (*cmm*C)	408	Identities = 317/408 (77%), Positives = 346/408 (84%), Gaps = 0/408 (0%)	*Streptomyces achromogenes subsp. rubradiris*	Hypothetical gene	411	Identities = 261/406(64%), Positives = 308/406 (75%) Gaps = 0/406 (0%)
*sfl*X	826	UV repair protein (SfrX)	826	Identities = 538/701 (76%) Positives = 602/701 (85%) Gaps = 0/701 (0%)	*Streptomyces argillaceus*	MtrX	828	Identities = 346/679 (50%) Positives = 430/679 (63%) Gaps = 7/679 (1%)
*sfl*RII	187	Transcriptional regulator (*cmm*RII)	198	Identities = 142/185 (76%) Positives = 154/185 (83%) Gaps = 0/185 (0%)	*Rhodococcus sp. RHA1*	possible transcriptional regulator, PadR family	203	Identities = 59/166 (35%), Positives = 90/166 (54%), Gaps = 6/166 (3%)
SfrA	321	ATP binding domain	325	Identities = 243/325 (74%) Positives = 267/325 (82%)	*S. argillaceus*	MtrA	320	Identities = 167/310 (53%) Positives = 217/310 (70%)

*Only small 3' region of *sfl*UIII was available and so limited analysis could be done

				Gaps = 5/325 (1%)				Gaps = 7/310 (2%)
*sfl*B	278	Membrane protein (SfrB)	250	Identities = 206/242 (85%) Positives = 225/242 (92%) Gaps = 1/242 (0%)	*Streptomyces avermitilis MA 4680*	ABC transporter transmembrane protein	249	Identities = 113/238 (47%) Positives = 163/238 (68%) Gaps = 1/238 (0%)
*sfl*LI	425	acyl coA ligase (*cmm*LI)	413	Identities = 275/406 (67%) Positives = 319/406 (78%) Gaps = 0/406 (0%)	*Streptomyces rishiriensis*	Putative ligase	529	Identities = 134/417(32%), Positives = 202/417 (48%), Gaps = 22/417 (5%)
*sfl*WI	272	side chain keto-reductase (*cmm*WI)	327	Identities = 202/270 (74%) Positives = 225/270 (83%) Gaps = 0/270 (0%)	*Streptomyces argillaceus*	Putative side chain reductase	326	Identities = 165/268(61%), Positives = 200/268 (74%), Gaps = 4/268 (1%)
*sfl*OIV	512	Oxygenase (*cmm*OIV)	506	Identities = 379/514 (73%) Positives = 408/514 (79%) Gaps = 11/514 (2%)	*Streptomyces argillaceus*	oxygenase	533	Identities = 283/506(55%), Positives = 326/506 (64%), Gaps = 14/506 (2%)
*sfl*G	395	glycosyl transferase (*cmm*G)	393	Identities = 299/391 (76%) Positives = 332/391 (84%) Gaps = 5/391 (1%)	*Streptomyces antibioticus*	ChlC7	411	Identities = 169/413(40%), Positives = 215/413 (52%), Gaps = 27/413 (6%)
*sfl*MI	345	O-methyltransferase (*cmm*MI)	357	Identities = 256/357 (71%) Positives = 285/357 (79%) Gaps = 13/357 (3%)	*Streptomyces argillaceus*	O-methyl transferase	345	Identities = 164/332(49%), Positives = 206/332 (62%), Gaps = 6/332 (1%)
*sfl*UI		NDP-4-ketoreductase (*cmm*UI)	249	Identities = 148/221 (66%) Positives = 172/221 (77%) Gaps = 0/221 (0%)	*Streptomyces fradiae*	NDP hexose 4-ketoreductase	247	Identities = 120/222(54%), Positives = 147/222 (66%), Gaps = 1/222 (0%)

*sfl*MII	376	C- Methyl transferase (*cmm*MII)	342	Identities = 254/343 (74%) Positives = 287/343 (83%) Gaps = 1/343 (0%)	*Streptomyces argillaceus*	C- methyl transferase	326	Identities = 185/328(56%), Positives = 238/328 (72%), Gaps = 3/328 (0%)
*sfl*RI	236	Transcriptional activator (*cmm*R1)	300	Identities = 166/228 (72%) Positives = 189/228 (82%) Gaps = 1/228 (0%)	*Streptomyces steffisburgensis*	SARP family regulator	290	Identities = 112/213(52%), Positives = 149/213 (69%), Gaps = 0/213 (0%)
*sfl*MIII	253	O - methyl transferase (*cmm*MIII)	250	Identities = 193/248 (77%), Positives = 217/248 (87%), Gaps = 0/248 (0%)	*Streptomyces olivaceous*	O-methyl transferase III / Rhamnose C4' methyltransferase	261	Identities = 158/247(63%), Positives = 185/247 (74%), Gaps = 0/247 (0%)
*sfl*D	347	NDP - Glucose synthase (*cmm*D)	355	Identities = 243/346 (70%), Positives = 286/346 (82%), Gaps = 0/346 (0%)	*Actinomadura hibisca*	PdmX	355	Identities = 197/341(57%), Positives = 243/341 (71%), Gaps = 0/341 (0%)
*sfl*E	327	NDP 4-6 Dehydratase (*cmm*E)	326	Identities = 243/326 (74%), Positives = 273/326 (83%), Gaps = 0/326 (0%)	*Streptomyces sp AM-7161*	NDP glucose 4-6 dehydratase	333	Identities = 210/319(65%), Positives = 256/319 (80%), Gaps = 1/319 (0%)
*sfl*P	423	Ketosynthase (*cmm*P)	422	Identities = 316/429 (73%), Positives = 338/429 (78%), Gaps = 14/429 (3%)	*Streptomyces antibioticus*	putative ketosynthase SimA1	422	Identities = 284/433(65%), Positives = 326/433 (75%), Gaps = 22/433 (5%
*sfl*K	437	Ketosynthase (*cmm*K)	422	Identities = 313/418 (74%), Positives = 347/418 (83%), Gaps = 1/418 (0%)	*Streptomyces antibioticus*	putative ketosynthase SimA2 / SimA3	404	Identities = 270/399 (67%), Positives = 314/399 (78%), Gaps = 2/399 (0%)

*sfl*UII	253	NDP- 4 Ketoreductase (*cmm*UII)	251	Identities = 163/247 (65%), Positives = 187/247 (75%), Gaps = 7/247 (2%)	*Streptomyces fradiae*	NDP-hexose 4 ketoreductase UrdR	247	Identities = 125/244 (51%), Positives = 168/244 (68%), Gaps = 3/244 (1%)
*sfl*W	366	NDP-3 Ketoreductase (*cmm*W)	336	Identities = 207/333 (62%), Positives = 244/333 (73%), Gaps = 9/333 (2%		3 Ketoreductase	349	Identities = 173/324 (53%), Positives = 205/324 (63%), Gaps = 5/324 (1%)
*sfl*V	345	NDP 2-3-Dehydratase (*cmm*V)	460	Identities = 276/456 (60%), Positives = 322/456 (70%), Gaps = 13/456 (2%)	*Streptomyces argillaceus*	D-olivose, D-oliose and D-mycarose 2,3-dehydratase	486	Identities = 217/424(51%), Positives = 270/424 (63%), Gaps = 34/424 (8%)
*sfl*S	86	Acyl Carrier Protein (*cmm*S)	83	Identities = 51/79 (64%), Positives = 62/79 (78%), Gaps = 0/79 (0%)	*Frankia CcI3*	Phospho-pentatheine binding	98	Identities = 51/79 (64%), Positives = 60/79 (75%), Gaps = 0/79 (0%)
*sfl*WI	259	Ketoreductase (*cmm*WI)	252	Identities = 162/253 (64%), Positives = 181/253 (71%), Gaps = 8/253 (3%)	*Strepomyces argillaceus*	Ketoreductase	254	Identities = 133/256 (51%), Positives = 155/256 (60%), Gaps = 7/256 (2%)
*sfl*Y	259	Cyclase (*cmm*Y)	258	Identities = 196/258 (75%), Positives = 221/258 (85%), Gaps = 0/258 (0%)		OxyN	257	Identities = 188/257(73%) Positives = 207/257 (80%) Gaps = 2/257 (0%)

Table-6. 2: Homology to Conserved Domain based on which putative proteins were predicted.

Putative Protein	Conserved Domain	Description
NDP-C methyl transferase	pfam08484: Methyltransf_14	This domain is found in bacterial C-methyltransferase proteins, often together with other methyltransferase domains such as pfam08241 or pfam08242.
UV repair protein	cd03271: ABC_UvrA_II	The excision repair protein UvrA domain II; Nucleotide excision repair in eubacteria is a process that repairs DNA damage by the removal of a 12-13-mer oligonucleotide containing the lesion. Recognition and cleavage of the damaged DNA is a multistep ATP-dependent reaction that requires the UvrA, UvrB, and UvrC proteins. Both UvrA and UvrB are ATPases, with UvrA having two ATP binding sites, which have the characteristic signature of the family of ABC proteins and UvrB having one ATP binding site that is structurally related to that of helicases.
Transcriptional regulator	pfam03551: PadR	Transcriptional regulator PadR-like family. Members of this family are transcriptional regulators that appear to be related to the pfam01047 family. This family includes PadR, a protein that is involved in negative regulation of phenolic acid metabolism.
ATP binding domain	cd03265: ABC_DrrA	DrrA is the ATP-binding protein component of a bacterial exporter complex that confers resistance to the antibiotics daunorubicin and doxorubicin. In addition to DrrA, the complex includes an integral membrane protein called DrrB. DrrA belongs to the ABC family of transporters and shares sequence and functional similarities with a protein found in cancer cells called P-glycoprotein. ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides,

		and more complex organic molecules. The nucleotide binding domain shows the highest similarity between all members of the family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region in addition to the Walker A motif/P-loop and Walker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins.
membrane protein	Pfam01061: ABC2_membrane	-
acyl coA ligase	PRK06188: PRK06188 acyl coA synthase	-
side chain keto-synthase	COG0667: Tas	Predicted oxidoreductases, related to aryl-alcohol dehydrogenases, involved in energy production and conversion
Oxygenase	PRK08244: PRK08244	hypothetical protein
glycosyl transferase	pfam06722: DUF1205	This family represents a conserved region of unknown function within bacterial glycosyl transferases. Many family members contain pfam03033.
O- methyltransferase	pfam08242: Methyltransf_12	Members of this family are SAM dependent methyltransferases.
NDP-4- ketoreductase	COG0451: WcaG	Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
C- Methyl transferase	pfam08242: Methyltransf_12	Methyltransferase domain. Members of this family are SAM dependent methyltransferases.
Transcriptional activator	COG3629: DnrI	DNA-binding transcriptional activator of the SARP family involved in signal

		transduction mechanisms.
O - methyl transferase	pfam05711: TylF	This family consists of bacterial macrocin O-methyltransferase (TylF) proteins. TylF is responsible for the methylation of macrocin to produce tylosin. Tylosin is a macrolide antibiotic used in veterinary medicine to treat infections caused by Gram-positive bacteria and as an animal growth promoter in the swine industry. It is produced by several *Streptomyces sp.* As with other macrolides, the antibiotic activity of tylosin is due to the inhibition of protein biosynthesis by a mechanism that involves the binding of tylosin to the ribosome, preventing the formation of the mRNA-aminoacyl-tRNA-ribosome complex.
NDP - Glucose synthase	pfam00483: NTP_transferase	This family includes a wide range of enzymes which transfer nucleotides onto phosphosugars.
NDP 4-6 Dehydratase	PRK10084: PRK10084	dTDP-glucose 4,6 dehydratase, has a putative role in NAD(P)-binding
Ketosynthase	cd00834: KAS_I_II	KASs are responsible for the elongation steps in fatty acid biosynthesis. KASIII catalyses the initial condensation and KAS I and II catalyze further elongation steps by Claisen condensation of malonyl-acyl carrier protein (ACP) with acyl-ACP.
Ketosynthase	cd00832: CLF	Chain-length factor (CLF) is a factor required for polyketide chain initiation of aromatic antibiotic-producing polyketide synthases (PKSs) of filamentous bacteria. CLFs have been shown to have decarboxylase activity towards malonyl-acyl carrier protein (ACP). CLFs are similar to other elongation ketosynthase domains, but their active site cysteine is replaced by a conserved glutamine.
NDP- 4 Ketoreductase	pfam01370: Epimerase	This family of proteins utilise NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions.

NDP-3 Ketoreductase	pfam01408: GFO_IDH_MocA	Oxidoreductase family, NAD-binding Rossmann fold. This family of enzymes utilise NADP or NAD. This family is called the GFO/IDH/MOCA family in swiss-prot.
NDP 2-3-Dehydratase	pfam03559: Hexose_dehydrat	This family includes a range of proteins from antibiotic production pathways. The family includes gra-ORF27 product that probably functions at an early step, most likely as a dTDP-4-keto-6- deoxyglucose-2,3-dehydratase. Its homologues include dnmT from the daunorubicin biosynthetic gene cluster in *S. peucetius*, a similar gene from the daunomycin biosynthetic cluster in *Streptomyces sp.* strain C5, eryBVI from the erythromycin cluster in *S. erythraea* and snoH from the nogalamycin cluster in *S. nogalater*. The proteins in this family are composed of two copies of a 200 amino acid long unit that may be a structural domain.
Acyl Carrier Protein	pfam00550: PP-binding	A 4'-phosphopantetheine prosthetic group is attached through a serine. This prosthetic group acts as a a swinging arm for the attachment of activated fatty acid and amino-acid groups. This domain forms a four helix bundle. The related domain of angR has the attachment serine replaced by an alanine.
Ketoreductase	PRK08063: PRK08063	
Cyclase	pfam04199: Cyclase	Proteins in this family are thought to be cyclase enzymes. They are found in proteins involved in antibiotic synthesis. However they are also found in organisms that do not make antibiotics pointing to a wider role for these proteins. The proteins contain a conserved motif HXGTHXDXPXH that is likely to form part of the active site.

6.3.5 Genes:

Of the 26 ORF's showing significant homology with different polyketide synthase genes (Table- 6.1), *sfl*X was the largest in the cluster and *cmm*S was found to be the smallest. Some genes that performed similar functions like *sfl*MI, *sfl*MII and *sfl*MIII showed a striking similarity in their DNA sequences. *sfl*MII and *sfl*RI, only genes that were transcribed individually, were contiguous in position but on opposite strands. Thus, they represent divergent transcriptional units. Genes in more than one place showed overlap of stop codon and start codon which is hallmark of ketosynthase and CLF genes. There were a few genes that even ran down, at their C-terminal end, in the ORF of next gene placed either on opposite strand or on the same strand. This could have probably happened due to addition/deletion mutation in the reading frame. This was not seen in chromomycin cluster of *S. griseus*. Two important genes of the cluster namely, *sfl*RI and *sfl*X involved in transcriptional regulation and self-resistance respectively, uses GTG as the start codon instead of usual ATG.

6.3.6 Organization of genes reflecting stages of their action:

Based on the putative functional attributes to the genes in the cluster and in the background of the putative pathway for chromomycin biosynthesis, we have attempted to study the step wise integrated progress in molecule formation that appears to be hallmark of this cluster. This attempt is not exhaustive since there are more genes/operons to be added in the cluster.

We hypothesize that operon -5 must be the first operon to activate since it contains a negative regulator (*sfl*RII) of chromomycin biosynthesis. Thus, activation of this operon is required before activation of biosynthesis genes. Also, it has ABC transporter genes (*sfr*A, *sfr*B), activation of which in advance ensures the cell from suicidal action of chromomycin biosynthesis. Under appropriate conditions *sfl*RI (transcriptional activator), which is transcribed individually, would initiate chromomycin biosynthesis by helping in transcription of operon – 3 which contains genes that initiate synthesis of chromophore backbone (*sfl*P and *sfl*K) and sugar moiety (*sfl*D and *sfl*E). Usually minimal PKS genes are found in same operon (Rawlings et al, 1999) but in this case it is not so. And for this reason activation of operon -1 containing essential *sfl*S

Table 6.3: Structural organization of various operons in the cluster and their presumptive role at different stages in chromomycin A_3 biosynthesis

Group	Strand	Order of genes in an operon	Putative stage of activation
Operon – 1	-	*sfl*S, *sfl*TI , *sfl*Y	3
Operon -2	+	*sfl*V, *sfl*W, *sfl*UII, *sfl*OI, *sfl*X	3
Operon -3	-	*sfl*K, *sfl*P, *sfl*E, *sfl*D	2
Operon -4	+	*sfl*MII, *sfl*UI, *sfl*MI, *sfl*GI, *sfl*OIV, *sfl*WI, *sfl*LI	4
Operon -5	-	*sfr*B, *sfr*A, *sfl*RII	1
Operon -6	+	*sfr*X, *sfl*C	4

gene along with *sfl*Y and *sfl*TI has to follow immediately. These are the same set of genes from mithramycin cluster that have been shown necessary for formation of SEK15 compound (Künzel et al, 1997). The two cyclases owing to a very high homology seem to be interchangeable. The role played by *mtm*X in formation of SEK15 appears to be assigned to other cyclase (*sfl*Y) in chromomycin cluster. Also, *mtm*OI which was present in the operon of mithramycin cluster, but was not required for formation of SEK15 and so, probably it has been removed from the operon in chromomycin cluster.

Ideally, operon -1 should have activated simultaneously with operon – 3. But the activator mediated, simultaneous activation of both operons was unlikely since there was no consensus DNA sequence in 5' upstream of the operons. Parallel activation of operon -2 is required for further synthesis of sugar moiety. This operon contains all genes involved in formation of basic sugar molecule (Lombó et al. 1997; González et al. 2001; Menéndez et al. 2004a) which is then variedly acted upon by different proteins to form five different sugars (Menéndez et al. 2006). Also, this operon contains genes required in subsequent maturation of chromophore. Based on the proposed pathways for mithramycin, next to follow would be operon – 4 that contain genes for final component

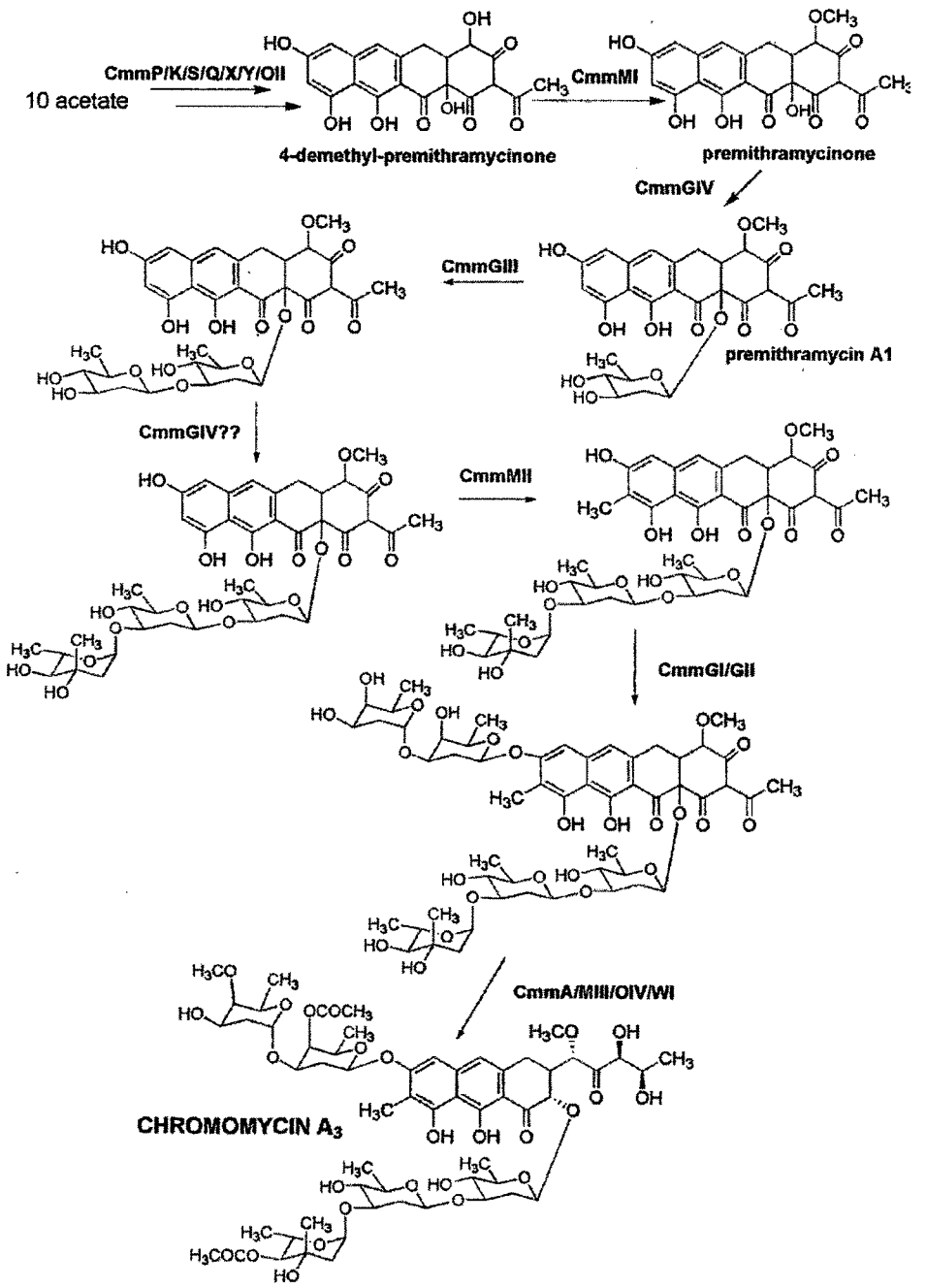


Fig 6.5: Putative chromomycin biosynthesis pathway showing role of various proteins in formation of molecule (Menendez et al, 2004a).

formation i.e. premithramycinone (Rohr et al, 1998) and sugars (Lombó et al, 1997). Either along with operon – 4 or just after it, activation of operon-6 would be required, as the later contains *sfr*X and *sfl*C genes. SflC converts dTDP-4-keto-D-olivose into dTDP-4-keto-D-mycarose, which is formed by the gene products of operon-4. Formation of dTDP-4-keto-D-mycarose is an important step, since it leads to formation of exclusive sugar-dTDP-L-chromose B after action of *cmm*F and *cmm*UIII (complete *sfl*F and 5' region of *sfl*UIII are yet to be explored in *S. flavisclerotícus* cluster) seen in case of *S. griseus* (Menéndez et al. 2004a). Also, *sfr*X with putative DNA damage repair activity will be required at this stage, since the active intermediates (9-demethyl-prechromomycin A3, prechromomycin B) will start forming as a result of appropriate glycosylation steps (Menéndez et al. 2006).

.3.7 Functional Groups in Chromomycin Cluster:

3.7.1 The Minimal PKS:

Several genes of the cluster probably participate in polyketide biosynthesis. Three genes namely Ketosynthase, Chain Length Factor and Acyl Carrier Protein forms a functional group called 'minimal PKS', which is responsible for building 20-carbon skeleton of the aglycon moiety of polyketide molecule. Due to their functional significance these three genes are usually found as a part of the same operon. Mithramycin cluster, another member of aureolic acid group and structurally very close relative of chromomycin has minimal PKS genes placed in tandem in a single operon. Strikingly in this cluster, while *sfl*P and *sfl*K are located together in the central region of the cluster, the acyl carrier protein gene (*sfl*S) is located at more than 8kb distance downstream, and in a separate operon which is unusually rare in the polyketide clusters known so far as seen in fig- 6.6. Such organizational anamoly has been reported only in case of daunomycin cluster (Grimm et al, 1994).

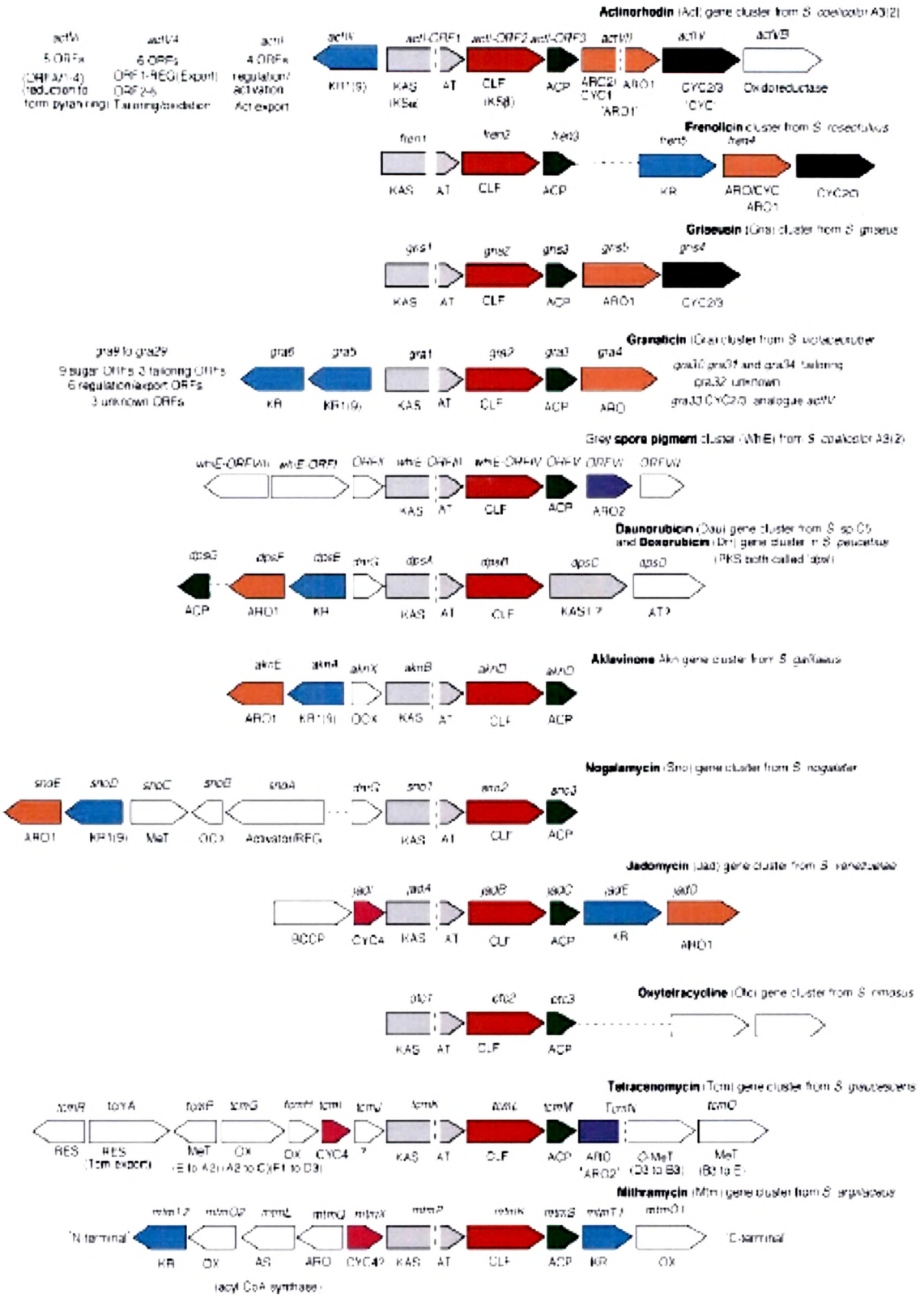


Fig 6.6: Minimal PKS region of different type II polyketide clusters (Rawlings et al, 1999). Note the common structural pattern of minimal PKS genes except for Daunomycin.

Distant location of these genes would require a simultaneous activation of both the operons, in order to ensure initiation of molecule formation. Interestingly, sequence analysis of the regions upstream of both the operons in *S. flavisclerotícus,* did not show any consensus region of homology, which could act as putative site for activator binding. Further, comparison of these sequences with their counterparts from *S. griseus* shows high conservation for regions upstream of SflP-SflK operon but there is total discrepancy for same regions of SflS. Also, no putative promoter region could be located even at a low stringency in *Dragon GC+ promoter finder* software, for ACP operon in both chromomycin clusters. Thus, it becomes important to find the mechanism by which the cluster is able to maintain its rhythm of genes/operons activation at right time.

*sfl*P and *sfl*K are inseparable due to a characteristic overlap (ATGA) between the stop codon (**ATGA**) of *sfl*P (**ATGA**) and the start codon of *sfl*K. They are highly conserved genes that show very high homolgy to KSα and KSβ homologous proteins, respectively. Owing to such conservation, the PKS library for this cluster was generated using KSα of *S. coelicolor* as probe, as has been done in past. The conservation of this gene has also been used in determining the evolution of these polyketide clusters (Metsa-Ketela et al, 2002). As explained in Chapter-4, we have also exploited this feature in studying phenomenon of horizontal gene transfer.

6.3.7.2 Genes involved in sugar biosynthesis :

From the structure it is clear that chromomycin A_3 (Fig- 6.7) contains five deoxysugars attached to the aglycon moiety, of which two are D – olivoses, one is 4-O-acetyl-L-chromose, one is 4-O-D-olivose, and one 4-O-methyl-D-olivose. Homology search shows presence of several putative genes that could be playing role in sugar biosynthesis.

Two genes, *sfl*D and *sfl*E are believed to carry out early steps during initiation of sugar biosynthesis. Positionally as well, they are located upstream in the operon that contains minimal pks genes, *sfl*P and *sfl*K. As mentioned earlier, this could be to initiate a simultaneous biosynthesis of sugar moiety along with the chromophore

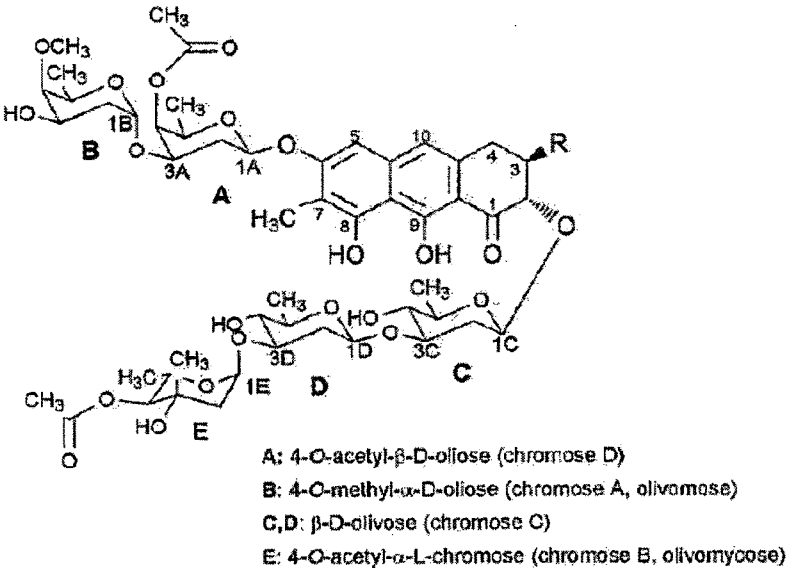


Fig 6.7: Structure and position of different sugars in structure of Chromomycin A_3

---

biosynthesis. Homolog of *sfl*V, a putative dehydratase has been shown to perform C-2 deoxygenation in other antibiotics (Draeger et al, 1999; Chen et al, 1999). *sfl*W with a putative reductase activity appears to bring about further maturation of these deoxysugars. The pathways for synthesis of all the deoxysugars involved in formation of sugar moiety seem to remain common till this step.

*sfl*UI would code for a 4-ketoreductase. There are two other genes (*sfl*UII and *sfl*UIII) whose products are similar to 4-ketoreductases and that are scattered in the cluster. SflUI and SflUII show highest similarities to UrdR, involved in the biosynthesis of D-olivose in the urdamycin pathway (Hoffmeister et al, 2000). On the other hand, SflUIII is more similar to 4-ketoreductases involved in L-sugar biosynthesis, such as AveBIV from the L-oleandrose biosynthetic pathway (Ikeda et al, 1999). SflUIII could be the 4-ketoreductase involved in the biosynthesis of 4-O-acetyl-L-chromose, and the two other ketoreductases (SflUI and SflUII) in the biosynthesis of the D-deoxysugars in chromomycin A_3. Like the chromophore, the sugar synthesis in chromomycin pathways follows the same route as for mithramycin except for one sugar intermediate (dTDP- 4-keto mycarose) which leads to formation of dTDP-D-mycarose carried out by *mtm*TIII.

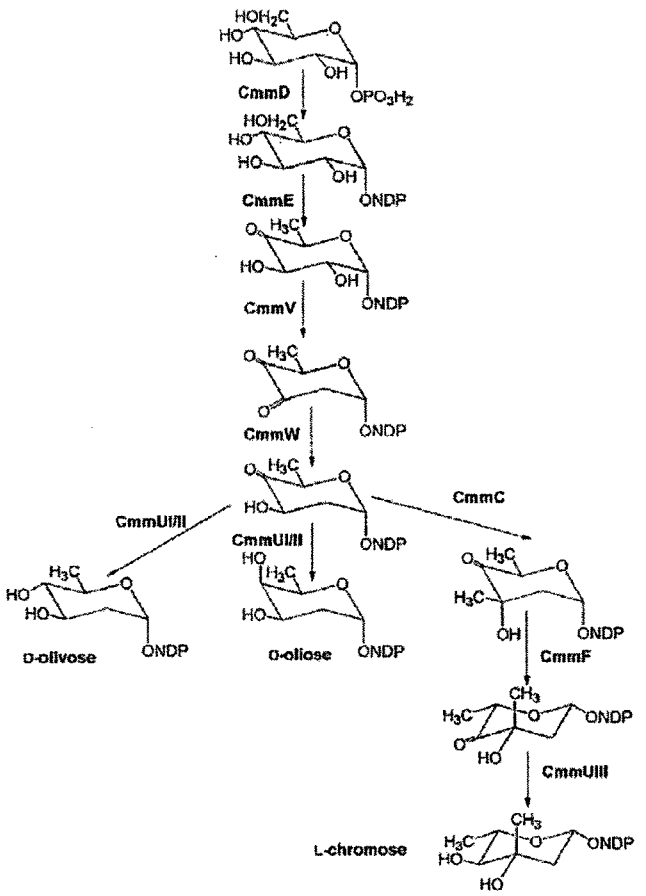


Fig 6.8: Putative biosynthesis pathway of different sugars of chromomycin A_3 as proposed for *S. griseus* cluster. (Mendez et al, 2004b)

In chromomycin, the same intermediate is acted upon by *cmm*F, which is an epimerase to generate dTDP- 4-keto chromose, followed by a reductase (*cmm*UIII) which finally brings reduction of keto sugar to generate dTDP-L-chromose (Menendez et al, 2004a). *sfl*F has not appeared in the cluster so far, and is expected to be present towards other end of the cluster as evidenced by *cmm*F position in chromomycin cluster of *S. griseus* (Menendez et al, 2004a). A mention of the gene has been made especially because it carries out a key step that differentiates the sugar moiety of mithramycin (Remsing et al, 2002) from chromomycin and thus, two molecules from each other, rest of the structure for both the molecules being the same.

6.3.7.3 Genes for regulation

The chromomycin biosynthetic cluster contains pathway specific regulatory genes that control the expression of the structural genes. Two genes have been identified namely *cmm*RI and *cmm*RII coding for proteins involved in regulation. *sfl*RI has a conserved domain for bacterial transcriptional activators marking its candidature in SARP (Streptomyces Antibiotic Regulatory Protein) family. Also, its importance in the cluster is reflected owing to its exclusive nature. This is one of the only two genes in the entire cluster, that are transcribed singly and do not form part of any operon. *sfl*RI showed highest homology to stfRI gene of a recently worked out steffimycin biosynthesis cluster from *S. steffiburgensis* (Gullon et al, 2006).

In mithramycin producer S. *argillaceus*, a mutation in *mtm*R gene could completely abolish the production of mithryamycin or any other intermediate. Further, expression of *mtm*R using a high-copy number vector in *S. argillaceus* caused a 16-fold increase in mithramycin production. In adition, *mtm*R was able to complement a mutation in the actinorhodin-specific activator actII-orfIV gene, and it also activated actionrhodin biosynthesis in *S. lividans* (Lombo et al. 1999).

*sfl*RII the other gene involved in regulation of chromomycin synthesis is positioned with resistance genes on its either side. This gene is present in an operon, upstream of a pair of ABC transporter genes. It contains a conserved domain of padR family regulatory proteins, which appears to be uncommon to Streptomycete. Interestingly, out of top twenty homologous proteins obtained by BlastX search, not even one belonged to *Streptomyces* sp (except for the homolog from *S. griseus*). Also, the maximum homology was with Rhodococcus sp., which is just 35% making it a unique protein. Highest homology with a Streptomycete protein (*S. avermitilis*) is just 24%. This makes *sfl*RII least conserved gene so far, in the whole cluster.

CmmRII has been shown to be involved in negative regulation of chromomycin biosynthesis in *S. griseus* (details in next chapter). The mithramycin analog (MtrY) of this protein shows very little homology and in contrast, inactivation of MtrY leads to decreased mithramycin production (Garcia-Bernardo et al, 2000). *cmm*RII has also been suggested to be controlling resistance genes, since in an experiment of *cmm*RII

inactivation by gene replacement using apramycin marker gene, it was observed that the mutant could withstand very high levels of mithramycin along with chromomycin, without any induction (Menendez et al, 2007). These results, and the fact that *cmm*RII itself is part of resistance gene operon, leads to hypothesize self regulation of *cmm*RII could be occuring through feedback inhibition mechanism.

6.3.7.4 Genes involved in self resistance:

Microbes produce antibiotics for self defense against other micro-organisms, in its vicinity. These organisms develop mechanisms to protect itself from deleterious effects of its own molecules (Cundliffe, 1989 ; Martin *et al.*, 2005 ; Méndez and Salas, 2001). In polyketide producers, such mechanisms of self resistance are invariably present within the polyketide biosynthesis cluster. Most of them possess ABC transporter genes to export the antibiotic product out of the cell and thereby impart self-resistance by maintaining intracellular environment free from the toxins (Fernández *et al.*, 1996 ; Guilfoile & Hutchinson, 1991 ; Olano *et al.*, 1995).

SfrA and SfrB form two component system of the efflux machinery. SfrA shows high homology to ATP binding protein component and SfrB with the membrane proteins. Interestingly, both proteins show higher homologies to proteins from different clusters/organisms. The CmrA-CmrB pump in *S. griseus* PKS cluster has been found to be more sensitive to the intermediate DDACA3 as compared to final product i.e. chromomycin A_3. Conversion of DDACA3 into chromomycin A_3 is carried out in one step reaction, by the action of the acetyltransferase (CmmA), which is supposed to be a transmembrane protein (Menéndez *et al.*, 2004b). Thus, the pro-drug which is not very active, gets matured (acetylated) to form an active drug, during its transport, outside the cell. The details about genes involved in resistance have been discussed in Chapter- 7A.

6.3.7.5 Genes for other modifications

There are several other genes that are involved in the formation of final molecule. These genes carry out a variety of changes either in the chromophore or in the sugar moiety.

A small protein SflX of just 142 amino acids (larger than only SflS) and SflY, loacated at one end of the cluster, are two putative cyclases that are expected to bring about cyclization of minimal PKS product at two different steps to generate chromophore rings.

There are three putative methyl transferase genes found in the cluster. Of the three genes, *sfl*MII and *sfl*MI are present in a single operon and based on the proposed role of analog in mithramycin (Lozano et al, 2000), are believed to act on the aglycon to carry out O-methylation and C-methylation of its intermediate. *sfl*MIII is transcribed individually with its own promoter. It is the only gene (other than *sfl*RI) in the whole cluster which is not part of any operon. It is responsible for methylation of the sugar moiety once it is attached to the aglycon (Patallo et al, 2001).

There are two accessory genes (*sfl*LI and *sfl*LII) the product of which share high homology with acyl co-ligase proteins. They could be involved in channelizing acetyl Co-A subunits in polyketide biosynthesis. From the open reading frames of the cluster, obtained so far, two putative monooxygenase genes have been obtained that show significant homology with FAD-dependent monooxygenases involved in hydroxylations of various aromatic polyketides. Especially *sfl*OIV which shares a very homology with *mtm*OIV must be playing an important role in aglycon cyclization. It has been shown to bring about oxidative cleavage of the fourth ring thereby converting the tetracyclic aglycon intermediate into a tricyclic one (Prado et al. 1999b) essential for making a molecule biologically active (Rodriguez et al. 2003; Gibson et al. 2005). One more oxygenase (CmmOIII) has been shown to be present in the chromomycin cluster in *S. griseus* (Menendez et al, 2004). Since region downstream of 17kbE is not available in PKS library, sequencing of that region of cluster is not determined.

Interestingly, four monooxygenase genes have been reported for mithramycin cluster, whereas only three appear in chromomycin cluster of *S. griseus*. BLAST analysis of the end sequences (126 bases) of chromomycin cluster in *S. griseus,* available in NCBI databank, showed homology with oxygenases from S. *rimosus*, *S. steffiburgensis*. This could be the missing analog of *mtm*OIII that has not been found so far in the chromomycin cluster of *S. griseus*. This is suggestive of that chromomycin cluster being incomplete and there being scope for exploring for more genes in *S. griseus* genome as

well. This could be a reason that heterologous expression of chromomycin cluster has not been reported, though clusters identified later, like steffimycin cluster by the same group (Gullon, 2006), have been heterologously expressed.

6.3.8 TRANSPOSON

6.3.8.1 Identification

Insertion sequence elements are small, genetically compact sequences, which generally encode no functions other than those involved in their mobility, although individual members of several families which include additional genes. IS-encoded functions include factors required in cis, in particular recombinationally active DNA sequences which define the ends of the element together with an enzyme, the transposase, which recognises and processes these ends. The Tpase is generally encoded by a single or perhaps two, open reading frames and consumes nearly the entire length of the element.

A majority of ISs exhibit short terminal inverted repeat sequences (IR) of between 10 and 40 bp. In those cases examined experimentally, the IRs can be divided into two functional domains. Domain A includes the two or three terminal base pairs, and is involved in the cleavages and strand transfer reactions leading to transposition of the element. Domain B is positioned within the IR and is involved in transposase binding.

Indigenous IS promoters are often located partially within the IR sequence upstream of the transposase gene, by convention IRL. A general pattern for the functional organisation of transposases suggests a sequence-specific DNA binding domain to be located in the N-terminal region while the catalytic domain is often localised towards the C-terminal end. Another general feature of IS elements is that, on insertion, most generate short directly repeated sequences (DR) of the target DNA flanking the IS. The length of the DR, between 2 and 14 bp, is characteristic for a given element.

Unexpectedly, sequencing of 17kbE fragment showed presence of a stretch of DNA which did not share any homology to PKS cluster. BLAST analysis suggested it to be an IS element (ISSfl1). It is positioned between the ORFs for *sfl*C and *sfl*UIII. Both these genes, involved in synthesis of sugar moiety, are parts of different operons and are diverging from each other. There is no such IS element detected in the chromomycin cluster of *S. griseus* sub *griseus*. Further, *cmm*C and *cmm*UIII overlap each other and have a common 3' region with shared 27 nucleotides. Surprisingly, the transpositioning event has occurred exactly at the end of *sfl*C gene and only *sfl*UIII is short of these 27 shared nucleotides. Also, insertion of transposon would alter the C-terminus of SflUIII.

ORF search for transposon DNA shows presence of two overlapping orfs of 190 and 150 amino acids that has homology to the IS701 and IS4 family of transposase respectively, from *S. avermitilis*. A rigorous analysis of this DNA suggests it to be a mutated version of ISSav4 trasposon present in *S. avermitilis*. BLAST2SEQ analysis of the two sequences showed significant homology throughout the region under study. But, a fragment of exactly 318 bases was missing from query DNA, and the homology is then completely restored till the C-terminus.

Homology observed at DNA level, does not get translated to aminoacid match. This was ascribed to point mutations in the query sequence. One to one sequence analysis with ISSav4 DNA revealed that at position 269 and 677 gaunosine residues were missing (deletion mutation) whereas at positions 385 and 553 an extra cytosine and adenosine residues were present (addition mutation) respectively. On incorporation (rather, resolving) of four mutations at positions 269, 385, 553 and 677 in existing sequence, a significant increase in aminoacid sequence was achieved. Aminoacid identities in BLASTX search increased from 41% (107/253) to 61% (153/249) and corresponding positive scores also increased from 50% (127/253) to 70% (176/249).

When searched for IRs, it was found that the left arm of ISSfl1 was considerably conserved with that of ISSav4 transposon. While the ends show intact sequences, there is variation in the central region of left arm which can be seen in the following comparision of sequences:

ISSfl1 :5'**TAGGGCTCCGT**C**AGGTCT**TCCCGTTTGCGCTGGTCA**GGA**G**CATGT**C**GA**G**G** 3'
ISSav4:5'**TAGGGCTCCGTTAGGTCT**GTCTCGCGGTCCTGTTTG**GGGGCATGTTGAAG** 3'

On the other hand, the right arm of transposon in ISSfl1 was not traceable. The corresponding sequences were completely missing in ISSfl1.

6.3.8.2 Evolutionary Significance:

Recent studies have shown that similar aromatic polyketide (Metsa-ketala et al, 2002), streptomycin (Egan et al, 2001) and penicillin (Buades and Moya, 1996) gene clusters or portions of clusters have been found to be present in otherwise distantly related organisms. These observations suggest that horizontal gene transfer of secondary metabolic pathways occurs among distantly related species [Lawrence and Henrickson, 2003; Walton, 2000). Mithramycin (the structurally close relative of chromomycin) also possesses a long stretch of inverted repeats flanking to its biosynthetic cluster.

Our results of chapter-4 clearly indicated the chromomycin cluster to have been acquired by *S. flavisclerotícus* via horizontal gene transfer. Though at that time the mechanism was not clear, one of the mechanisms proposed for the event was transposon mediated transfer. Presence of a transposon within the cluster is strong evidence to support that claim.

At first sight, presence of a mobile element in any genome is not a surprise. But when analyzed in totality, the presence of ISSfl1 in chromomycin cluster, its position and intactness leads to some interesting observations. This small, probably nonfunctional stretch of DNA has lot more to offer than what appears at first sight. *In silico* analysis, and interpretation in totality, surfaces a better picture about ISSfl1.
Before accounting ISSfl1 responsible for horizontal transfer of chromomycin cluster, it is very important to first determine the timing of transposition event in relation to the event of horizontal transfer. There are two possibilities:

1. PKS cluster acquisition by *S. flavisclerotícus*, followed by ISSfl1 transposition
2. ISSfl1 transposition, followed by PKS cluster acquisition by *S. flavisclerotícus*.

If case one is true, then ISSfl1 has to be an indigenous IS element of S. flavisclerotícus and should have been hopping randomly throughout the genome and its attack on PKS cluster is merely a co-incidence and has to relation with horizontal transfer. Data from the two completely sequenced Streptomyces genomes namely S. coelicor (Bentley et al, 2002) and S. avermitilis (Omura et al, 2001) shows presence of several copies of intact, truncated and frame shifted forms of transposons in their genomes. Whereas, there is only a single transposition event in ~30kb DNA of S. flavisclerotícus sequenced so far. Also, if transposition event occurred after integration of chromomycin cluster in S. flavisclerotícus genome, then the mechanism of transfer needs an explanation.

On the other hand, there are strong indicative evidences that favour transposition to have occurred before acquisition of the cluster.

a. If the transposon was a part of host genome, the cluster should have been attacked more than once, but there are no such signs throughout the cluster.
b. Positioning of the transposon is such that
    i. functionally, neither any of the seven operons in the cluster nor any of 26 genes in these operons, is disrupted.
    ii. Its insertion is between two divergent genes, so even the promoter activity of any genes is unhindered.

Probablity of such event to occur at random, appears to be very low. While transfer of such DNA from one genome to another, after selection, appears to be more convincing.

c. Presence of small stretch of transposase like sequence flanking one end of the cluster, absence of right arm (inverted repeat) in ISSfl1 and elimination of 318 bases from ISSfl1 leads us to hypothesize that due to lack of one arm IS element after insertion could not hop in isolation and required other IS elements for it. After transfer to new host, transposase might have lost 318 bases to get stabilized in that position.

Thus, circumstantial but strong evidences advocates that the latter case, of transposition attack on the PKS cluster followed by its transfer, to be true. A further work on the same would make the picture crystal clear.

CONCLUSION

Though the product of PKS cluster from S. flavisclerotícus was identical to that from S. griseus, there were chances that the two clusters could have evolved differently. This assumption was based on the similarity of chromomycin molecule with that of mithramycin from *S. argillaceus*. Though both the products have a very little difference in their structures, the clusters are entirely different. It has been suggested that the two similar products from differently organized clusters is an example of convergent evolution.

Comparision of the two gene clusters with same product, present in different genomes, provides an opportunity to study the plasticity of the genes in their structure and function. For 30kb DNA sequenced from *S. flavisclerotícus*, the two chromomycin clusters showed 75% homology at DNA level with just 2% gaps.

REFERENCES

Anthony Schwacha, Jill A. Cohen T, Kalle B, Gehring T, And Robert A B (1990) Tn*1000*-Mediated Insertion mutagenesis of the histidine utilization (hut) gene cluster from *Klebsiella aerogenes*: Genetic analysis of hut and unusual target specificity of Tn*1000.* J Bacteriol, 172: 5991-5998

Bentley S D, Chater K F, Cerdeño-Tárraga A M, Challis G L, Thomson N R, James K D, Harris D E, Quail M A, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen C W, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream M A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell B G, Parkhill J and Hopwood D A (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141-147

Berg C M, Vartak N B, Wang G, Xu X, Liu L, MacNeil D J, Gewain K M, Wiater L M and Berg D E (1992). The mγδ-1 element, a small γδ (Tn*1000*) derivative useful for plasmid mutagenesis, allele replacement and DNA sequencing. Gene 113:9-16.

Bisang C, Long P F, Cortes J, Westcott J, Crosby J, Matharu A L, Cox R J, Simpson T J, Staunton J and Leadlay P F (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. Nature 401: 502-505

Buades C and Moya A (1996) Phylogenetic analysis of the isopenicillin-N-synthetase horizontal gene transfer. J Mol Evol 42: 537-542.

Chakrabarti S, Bhattacharyya D, Dasgupta D (2000) Structural basis of DNA recognition by anticancer antibiotics, chromomycin A(3), and mithramycin: roles of minor groove width and ligand flexibility. Biopolymers 56:85–95

Chen H, G Agnihotri, Z Guo, N L S Que, H X Chen and H W Liu (1999). Biosynthesis of mycarose: isolation and characterization of enzymes involved in the C-2 deoxygenation. J Am Chem Soc. 121:8124-8125

Cundliffe, E. (1989). How antibiotic-producing organisms avoid suicide. Annu Rev Microbiol 43: 207–233.

Draeger G, Park S H and Floss H G (1999). Mechanism of the 2-deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin. J Am Chem Soc. 121:2611-2612

Egan S, Wiener P, Kallifidas D and Wellington E M H (2001). Phylogeny of Streptomyces species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. Antonie van Leeuwenhoek-Int. J Gen Mol Microb 79: 127-133.

Fernandez E, Lombo F, Mendez C and Salas J A (1996). An ABC transporter is essential for resistance to the antitumor agent mithramycin in the producer *Streptomyces argillaceus*. Mol Gen Genet 251: 692–698.

Furuya K and Hutchinson C R (1998). The DrrC protein of *Streptomyces peucetius*, a UvrA-like protein, is a DNA-binding protein whose gene is induced by daunorubicin. FEMS Microbiol Lett 168: 243–249.

Garcia-Bernardo J, Brana A F, Mendez C and Salas J A (2000). Insertional inactivation of mtrX and mtrY genes from the mithramycin gene cluster affects production and growth of the producer organism *Streptomyces argillaceus*. FEMS Microbiol Lett 186: 61–65.

Gibson M, Nur-e-alam M, Lipata F, Oliveira MA, Rohr J (2005) Characterization of kinetics and products of the Baeyer–Villiger oxygenase MtmOIV, the key enzyme of the biosynthetic pathway toward the natural product anticancer drug mithramycin from *Streptomyces argillaceus*. J Am Chem Soc 127:17594–17595

González A, Remsing L L, Lombo F, Fernández M J, Prado L, Brana A F, Künzel E, Rohr J, Méndez C, Salas J A (2001). The *mtmVUC* genes of the mithramycin gene cluster in *Streptomyces argillaceus* are involved in the biosynthesis of the sugar moieties. Mol Gen Genet 264:827–835

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

Guilfoile P G and Hutchinson C R (1991). A bacterial analog of the mdr gene of mammalian tumor cells is present in *Streptomyces peucetius*, the producer of daunorubicin and doxorubicin. Proc Natl Acad Sci 88: 8553–8557

Gullón S, Olano C, Abdelfattah M S, Braña A F, Rohr J, Méndez C and Salas J A (2006). Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin. Appl Env Microbiol. 72: 4172-4183

Hoffmeister D, Ichinose K, Domann S, Faust B, Trefzer A, Drager G, Kirschning A, Fischer C, Kunzel E and Bearden D (2000). The NDP-sugar co-substrate concentration and the enzyme expression level influence the substrate specificity of glycosyltransferases: cloning and characterization of deoxy- sugar biosynthetic genes of the urdamycin biosynthetic gene cluster. Chem. Biol. 7: 821–831.

Ikeda H, Nonomiya T, Usami M, Ohta T and Amura S (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. 96: 9509–9514.

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M and Omura S (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc Natl Acad Sci 98: 12215–12220.

Künzel E, Wohlert S E, Beninga C, Haag S, Decker H, Hutchinson C R, Blanco G, Mendez C, Salas J A and Rohr J (1997). Tetracenomycin M, a novel genetically engineered tetracenomycin resulting from a combination of mithramycin and tetracenomycin biosynthetic genes. Chem Eur *J* 3:1675–1678

Lawrence J G (2003). Lateral gene transfer: When will adolescence end? Mol Microbio 50:.739-749.

Liu L, Whalen W, Das A and Berg C M (1987). Rapid sequencing of cloned DNA using a transposon for bi-directional priming: sequence of the *Escherichia coli* K-12 *avt*A gene. Nucleic Acids Res. 15:9461-9469

Lombó F, Braña A F, Méndez C, Salas J A (1999) The mithramycin gene cluster of *Streptomyces argillaceus* contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster. J Bacteriol 181:642–647

Lombó F, Siems K, Braña A F, Méndez C, Bindseil K, Salas J A (1997). Cloning and insertional inactivation of *Streptomyces argillaceus* genes involved in the earliest steps of biosynthesis of the sugar moieties of the antitumor polyketide mithramycin. J Bacteriol 179:3354–3357

Lomovskaya N, Hong S K, Kim S U, Fonstein L, Furuya K & Hutchinson C R (1996). The *Streptomyces peucetius* drrC gene encodes a UvrA-like protein involved in daunorubicin resistance and production. J Bacteriol 178: 3238–3245

Lozano M J, Remsing L L, Quirós L M, Braña A F, Fernández E, Sánchez C, Méndez C, Rohr J, Salas J A (2000). Characterization of two polyketide methyltransferases involved in the biosynthesis of the antitumor drug mithramycin by *Streptomyces argillaceus*. J Biol Chem 275:3065–3074

Martin J F, Casqueiro J and Liras P (2005). Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. Curr Opin Microbiol 8: 282–293

Mendez C and Salas J A (2001). The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanism. Res Microbiol 152, 341–350

Menéndez N, Nur-e-Alam M, Braña A F, Rohr J, Salas J A, Méndez C (2004a) Biosynthesis of the antitumor chromomycin A_3 in *Streptomyces griseus*: analysis of the gene cluster and rational design of novel chromomycin analogs. Chem Biol 11:21–32

Menéndez N, Nur-e-Alam M, Braña A F, Rohr J, Salas J A, Méndez C (2004b) Tailoring modification of deoxysugars during biosynthesis of the antitumour drug chromomycin A3 by *Streptomyces* griseus ssp. griseus. Mol Microbiol 53:903–915

Menéndez N, Nur-e-Alam M, Fischer C, Braña A F, Salas J A, Rohr J, Méndez C (2006). Deoxysugar transfer during chromomycin A_3 biosynthesis in *Strepomyces griseus* subsp. *griseus*: new derivatives with antitumor activity. Appl Environ Microbiol 72:167–177

Menéndez N, Nur-e-Alam M, Fischer C, Braña A F, Salas J A, Rohr J, Méndez C (2006) Deoxysugar transfer during chromomycin A_3 biosynthesis in *Strepomyces griseus* subsp. *griseus*: new derivatives with antitumor activity. Appl Environ Microbiol 72:167–177

Metsa-Ketela M, Halo L, Munukka E, Hakala J, Mantsala P and Ylihonko K (2002). Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. Appl Env Microbiol 68:4472-4479.

Menendez N, Brana A F, Salas J A and Mendez C (2007). Involvement of a chromomycin ABC transporter system in secretion of a deacetylated precursor during chromomycin biosynthesis. Microbiol 153: 3061–3070

O'connor S. (2004). Aureolic acids: Similar antibiotics with different biosynthetic gene clusters. Chem Biol, 11: 8-10.

Olano C, Rodriguez A M, Mendez C and Salas J A (1995). A second ABC transporter is involved in oleandomycin resistance and its secretion by *Streptomyces antibioticus*. Mol Microbiol 16: 333–343.

Patallo E, Blanco G, Fischer C, Brana A F, Rohr J, Mendez C and Salas J A (2001). Deoxysugar methylation during biosynthesis of the antitumor polyketide elloramycin by *Streptomyces* olivaceus. J Biol Chem 276:18765-18774

Prado L, Fernández E, Weissbach U, Blanco G, Quirós L M, Braña A F, Méndez C, Rohr J and Salas J A (1999) Oxidative cleavage of premithramycin B is one of the last steps in the biosynthesis of the antitumor drug mithramycin. Chem Biol 6:19–30

Rawlings B J (1999) Biosynthesis of polyketides (other than actinomycete macrolides) Nat Prod Rep16: 425–484

Remsing L L, Garcia-Bernardo J, González A, Kunzel E, Rix U, Braña A F, Bearden D W, Méndez C, Salas J A and Rohr J (2002). Ketopremithramycins and ketomithramycins, four new aureolic acid-type compounds obtained upon inactivation of two genes involved in the biosynthesis of the deoxysugar moieties of the antitumor drug mithramycin by *Streptomyces argillaceus*, reveal novel insights into post-PKS tailoring steps of the mithramycin biosynthetic pathway. J Am Chem Soc 124:1606–1614

Rodriguez D, Quirós L M, Braña A F and Salas J A (2003). Purification and characterization of a monooxygenase involved in the biosynthetic pathway of the antitumor drug mithramycin. J Bacteriol 185:3962–3965

Rohr J, Weissbach U, Beninga C, Künzel E, Siems K, Bindseil K U, Lombó F, Prado L, Braña A F, Méndez C and Salas J A (1998). The structures of premithramycinone and demethylpremithramycinone, plausible early intermediates of the aureolic acid group antibiotic mithramycin. Chem Commun 437–438.

Strausbaugh L D, Bourke M T, Sommer M T, Coon M T and Berg C M (1990). Probe mapping to facilitate transposon-based DNA sequencing. Proc Natl Acad Sci 87: 6213-6217.

Walton J D (2000). Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: A hypothesis. Fungal Gen Biol 30:167-171.