ELSEVIER

Chemistry & Biology 8 (2001) 487–499

Chemistry & Biology

www.elsevier.com/locate/chembiol

Research Paper

Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*[1]

Clive Waldron [a,*], Patti Matsushima [b], Paul R. Rosteck Jr. [b], Mary C. Broughton [b], Jan Turner [b], Krishnamurthy Madduri [a], Kathryn P. Crawford [a], Donald J. Merlo [a], Richard H. Baltz [a,b]

[a] *Dow AgroSciences, Indianapolis, IN 46268, USA*
[b] *Eli Lilly and Company, Indianapolis, IN 46285, USA*

Received 16 October 2000; revisions requested 19 December 2000; revisions received 8 March 2001; accepted 16 March 2001
First published online 30 March 2001

Abstract

Background: Spinosad is a mixture of novel macrolide secondary metabolites produced by *Saccharopolyspora spinosa*. It is used in agriculture as a potent insect control agent with exceptional safety to non-target organisms. The cloning of the spinosyn biosynthetic gene cluster provides the starting materials for the molecular genetic manipulation of spinosad yields, and for the production of novel derivatives containing alterations in the polyketide core or in the attached sugars.

Results: We cloned the spinosad biosynthetic genes by molecular probing, complementation of blocked mutants, and cosmid walking, and sequenced an 80 kb region. We carried out gene disruptions of some of the genes and analyzed the mutants for product formation and for the bioconversion of intermediates in the spinosyn pathway. The spinosyn gene cluster contains five large open reading frames that encode a multifunctional, multisubunit type I polyketide synthase (PKS). The PKS cluster is flanked on one side by genes involved in the biosynthesis of the amino sugar forosamine, in *O*-methylations of rhamnose, in sugar attachment to the polyketide, and in polyketide cross-bridging. Genes involved in the early common steps in the biosynthesis of forosamine and rhamnose, and genes dedicated to rhamnose biosynthesis, were not located in the 80 kb cluster.

Conclusions: Most of the *S. spinosa* genes involved in spinosyn biosynthesis are found in one 74 kb cluster, though it does not contain all of the genes required for the essential deoxysugars. Characterization of the clustered genes suggests that the spinosyns are synthesized largely by mechanisms similar to those used to assemble complex macrolides in other actinomycetes. However, there are several unusual genes in the spinosyn cluster that could encode enzymes that generate the most striking structural feature of these compounds, a tetracyclic polyketide aglycone nucleus.

Keywords: Glycosylation; Macrolide; Polyketide biosynthesis; Spinosad; Spinosyn; *Saccharopolyspora spinosa*

1. Introduction

The spinosyns (previously referred to as A83543 factors; [1,2]) are novel macrolides produced by *Saccharopolyspora spinosa* [3]. Spinosyns are comprised of a tetracyclic macrolide containing forosamine and tri-*O*-methyl rhamnose, with different degrees of methylation on the polyketide or deoxysugars [1,2,4]. The two major factors in the *S. spinosa* fermentation, spinosyn A and spinosyn D, differ from each other by a single methyl substituent at position 6 of the polyketide (Fig. 1). Spinosad, a combination of spinosyn A and spinosyn D, is marketed by Dow AgroSciences for control of agricultural insect pests. Spinosad is highly effective against target insects and has an excellent environmental and mammalian toxicological profile [5–7]. Incorporation studies with ^{13}C-labeled acetate, propionate, and methionine established that spinosyns are assembled by a polyketide pathway, and that the two *N*-methyl groups of forosamine, and the three *O*-methyl groups of tri-*O*-methylrhamnose, are derived from S-adenosyl-methionine [2]. The polyketide portion of spinosyns differs

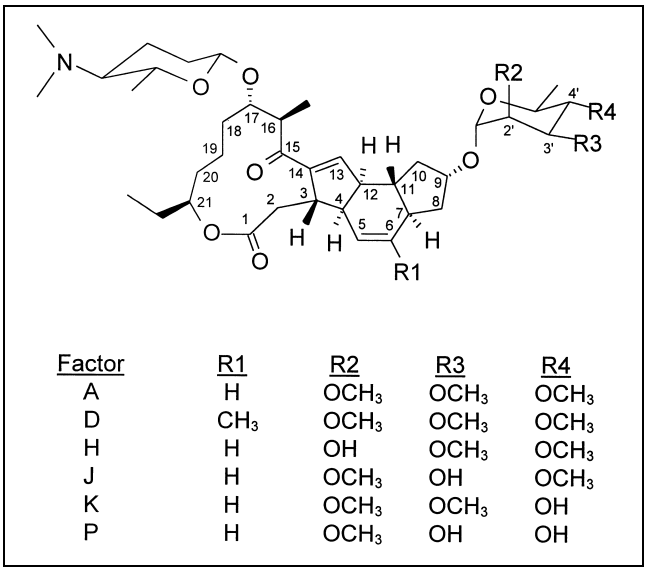
[1] The DNA sequence reported here was deposited in GenBank under the accession number AY007564.

* Correspondence: Clive Waldron; E-mail: cwaldron@dowagro.com

1074-5521/01/$ – see front matter
PII: S1074-5521(01)00029-1

from more common type I polyketides (e.g. erythromycin, rapamycin, or tylosin) in that it contains three intramolecular carbon–carbon bonds (Fig. 1). This added dimension poses interesting biosynthetic questions, and suggests that *S. spinosa* may encode novel enzymes to carry out these reactions. Candidate genes for these enzymes may be present as open reading frames (ORFs) that are closely linked to the other spinosyn synthetic genes, since most macrolide biosynthetic genes are clustered [8–10]. However, they are expected to have little similarity to known genes because few organisms produce such cross-linked macrolides [2,11,12], and their genetic or enzymatic basis has not yet been described. Spinosyn genes of this type may facilitate the isolation of the other genes, and (if the substrates can be synthesized) they could provide recombinant proteins for uncovering the enzyme mechanisms involved.

To aid in the cloning and analysis of the spinosyn biosynthetic genes, Matsushima et al. [13] developed a gene transfer system that employed conjugation from *Escherichia coli*. *S. spinosa* was shown to express potent restriction barriers to protoplast transformation and bacteriophage plaque formation [14], but conjugation bypasses these barriers [13,15]. Using conjugation, cosmids containing large inserts of *S. spinosa* DNA were introduced into *S. spinosa*, and were efficiently inserted into the chromosome by homologous recombination [13]. Chemically induced *S. spinosa* mutants blocked in spinosyn biosynthesis were also isolated and used as hosts for cloning the spinosyn biosynthetic genes by complementation. We report here the cloning, DNA sequence, and partial functional analysis of the spinosyn gene cluster.



Factor	R1	R2	R3	R4
A	H	OCH_3	OCH_3	OCH_3
D	CH_3	OCH_3	OCH_3	OCH_3
H	H	OH	OCH_3	OCH_3
J	H	OCH_3	OH	OCH_3
K	H	OCH_3	OCH_3	OH
P	H	OCH_3	OH	OH

Fig. 1. Structure of spinosyns. Forosamine and tri-*O*-methylrhamnose are attached at positions 17 and 9, respectively.

2. Results

2.1. Cloning of spinosyn biosynthetic genes

Mutants defective in spinosyn biosynthesis were identified during studies on conjugation in *S. spinosa* [13]. Some mutants were unable to bioconvert the spinosyn aglycone (lacking both sugars) or factor K (lacking a single *O*-methyl group; Fig. 1) to spinosyn A. These mutants lacked a 400 kb *Spe*I fragment that was easily separated from other *Spe*I fragments by PFGE [13]. Since spinosyn genes may have been deleted in these mutants, we used the 400 kb *Spe*I fragment, and a ketosynthase (KS) fragment [16], as hybridization probes to identify cosmids that contained spinosyn genes. Cosmids that hybridized with both probes were introduced by conjugation into *S. spinosa* A83543.3, a spinosad producer, and into strains S274, S229, S230 and S240, mutants blocked in the addition of forosamine to the pseudoaglycone of spinosyn (PSA) or blocked in the individual *O*-methylations of rhamnose (Fig. 1 and Table 1). Cosmid pRHB9A6 gave strong hybridization to the *Spe*I fragment and weak hybridization to the KS probe. When pRHB9A6 was introduced into each of the mutant strains, the recombinants produced spinosad. This indicated that pRHB9A6 contained genes involved in the biosynthesis or attachment of forosamine, and in the *O*-methylations of rhamnose. Cosmid pRHB2C10 gave strong

Table 1
Strains and plasmids

Strains and plasmids	Description[a]	Source/reference
E. coli S17-1	C600::RP4 2-Tc::Mu-Km::Tn7 *hsdR hsdM*$^+$ *recA*	[59]
S. spinosa A83543.3	produces spinosyn A	NRRL 18538
S. spinosa S229	produces spinosyn H	NRRL 18823
S. spinosa S230	produces spinosyn J	this report
S. spinosa S240	produces spinosyn K	NRRL 18743
S. spinosa S274	produces PSA	this report
pOJ436	cosmid vector; AmR, *oriT*, *attP*$^{\phi C31}$	[17]
pRHB2C10	pOJ436 containing spinosyn genes	this report
pRHB9A6	pOJ436 containing spinosyn genes	this report
pRHB3E11	pOJ436 containing spinosyn genes	this report
pOJ260	AmR, *rep*^{pUC}, *oriT*	[17]

[a]AmR, apramycin resistance gene; *attP*$^{\phi C31}$, attachment site for phage ϕC31; *oriT*, plasmid RK2 origin for conjugal transfer; *rep*^{pUC}, pUC origin of replication.

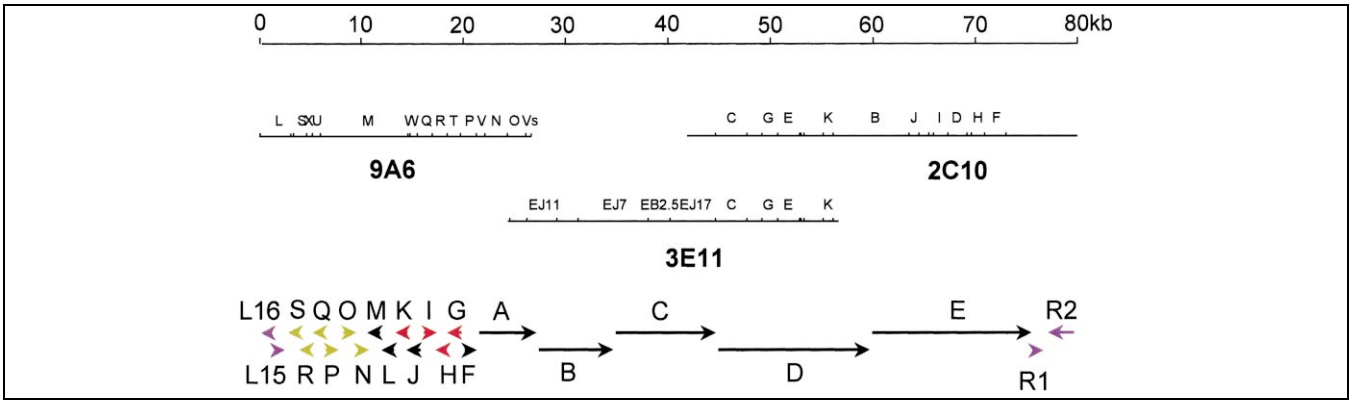


Fig. 2. Organization of the spinosyn biosynthetic gene cluster. The letters correspond to the *Bam*HI fragments that have been subcloned from the three overlapping inserts of the cosmid clones. The arrows represent ORFs, pointing in the direction of transcription. The arrows are colored to indicate which part of the spinosyn molecule has been tentatively associated with each gene product: black, aglycone; red, rhamnose; blue, forosamine; green, non-spinosyn.

hybridization to the 400 kb *Spe*I and the KS probe. Introduction of pRHB2C10 into strain A83543.3 caused inactivation of spinosad production. This suggested that pRHB2C10 contained genes involved in polyketide assembly. We used an end fragment of cosmid pRHB9A6 as a probe to identify an overlapping cosmid, pRHB3E11, and confirmed that it overlapped pRHB9A6 and pRHB2C10 by *Bam*HI restriction site mapping (Fig. 2).

2.2. Gene disruption and bioconversion analysis

To confirm that we had cloned polyketide synthase (PKS) genes involved in spinosyn biosynthesis, we subcloned several fragments from the three cosmids into pOJ260, and introduced them into *S. spinosa* by conjugation from *E. coli* S17-1. Stable recombinants were formed by recombination into the chromosome, because pOJ260 lacks replication functions for *Saccharopolyspora* species [15,17]. Insertion of plasmids containing segments from three different genes encoding putative PKS proteins (subsequently identified as *spnA*, *spnD*, and *spnE*) blocked spinosyn biosynthesis (Table 2). However, all three classes of insertion mutants converted the spinosyn aglycone (AGL) to spinosyn A, indicating that all of the post-PKS functions were functional in the recombinants. Having confirmed by complementation and gene disruption analyses that we had cloned genes involved in spinosyn polyketide assembly and tailoring, we proceeded to sequence the cloned DNA.

Table 2
Gene disruption and bioconversion by *S. spinosa* mutants

Gene	Internal segment (kb)	Mutant product	Bioconversion of spinosyns[a]			
			P	K	PSA	AGL
spnA	2.2	none				+
spnD	1.9	none				+
spnE	1.9	none				+
spnF	0.7	none		+	+	+
spnG	0.6	none	K	−	+	−
spnH	1.0	P		−	+	
spnJ	0.6	none		+		+
spnK	0.8	none				
spnL	0.6	none	+		+	+
spnM	0.7	A (trace)		+	+	
spnN	0.7	PSA				
spnO	1.0	PSA				
spnP	0.8	PSA				PSA
ORF-R2	0.5	A				
ORF-L15	0.6	A				
ORF-L16	0.6	A				

[a]Bioconversion of spinosyn factors to spinosyn A (+) or to other factors (J, K). No conversion detected (−).

2.3. DNA sequence analysis

We obtained DNA sequence of cosmids pRHB9A6, pRHB3E11 and pRHB2C10 and identified 23 ORFs in a cluster spanning ~80 kb (Fig. 2). We analyzed the deduced gene products of the individual ORFs by BLAST and FASTA programs [18,19] available through the National Center for Biotechnology Information, and in some cases by gene disruption and bioconversion of intermediates in spinosad biosynthesis.

2.3.1. Polyketide biosynthetic genes

DNA sequences encoding PKS functions (Table 3) were found in a 56 kb region of the overlapping cosmids. We identified 50 enzyme domains organized into one loading module and 10 extender modules (Fig. 3), as expected from the structure of the macrolide. The modules were arranged head-to-tail, like the erythromycin (*ery*) PKS [8], and not like the convergent avermectin (*avr*) or rapamycin (*rap*) PKSs [9,10]. There were no additional non-PKS sequences in this region, such as the insertion element that occurs between the *ery*AI and *ery*AII PKS genes [8]. Theoretically, the *spn* PKS genes could be transcribed as a single mRNA, but there is evidence from gene disruption

Table 3
Proposed functions of genes in the spinosyn biosynthetic cluster

Gene	Size (aa)	Proposed function	Closest homologue[a]	ID/SM[b]	Reference or accession number
spnA	2596	PKS	*S. venezuelae pikAI*	45/56	[37]
spnB	2153	PKS	*S. venezuelae pikAII*	46/56	[37]
spnC	3171	PKS	*S. erythraea* ORF3	43/53	[8]
spnD	4929	PKS	*S. natalensis* PKS	51/62	[60]
spnE	5589	PKS	*S. natalensis* PKS	48/58	[60]
spnF	276	polyketide bridging?	*S. avermitilis aveD*	39/54	[9]
spnG	391	rhamnosyltransferase	*S. cyanogenus lanGT1*	36/49	[31]
spnH	251	rhamnose; *O*-methyl-transferase	*S. fradiae tylF*	59/73	[29]
spnI	396	rhamnose; *O*-methyl-transferase	*S. fradiae tylE*	40/53	[30]
spnJ	540	polyketide bridging?	*S. peucetius dnrW*	35/47	U80222
spnK	396	rhamnose; *O*-methyl-transferase	*S. fradiae tylE*	43/58	[30]
spnL	284	polyketide bridging?	*S. avermitilis aveD*	30/50	[9]
spnM	321	polyketide bridging?	*S. coelicolor* SC10G8	32/48	AL158057
spnN	333	forosamine; 3-ketoreductase	*S. cyanogenus lanT*	53/62	[30]
spnO	487	forosamine; 2,3-dehydratase	*S. cyanogenus lanS*	50/62	[31]
spnP	456	forosamyltransferase	*S. venezuelae desVII*	41/56	[37]
spnQ	463	forosamine; 3,4-dehydratase	*S. violaceoruber gra*-ORF23	59/69	[27]
spnR	386	forosamine; transaminase	*S. lincolnensis lmbS*	31/46	[35]
spnS	250	forosamine; dimethyltransferase	*S. venezuelae desVI*	49/67	[37]
ORF-R1	199	unknown	*S. coelicolor* SCF81.27	33/43	AL133171
ORF-R2	282	exodeoxyribonuclease V	*C. pneumoniae recD*	38/57	AE001598.1
ORF-L15	256	putative oxidoreductase	*S. coelicolor* SCM1.23c	51/60	AL133422
ORF-L16	282	putative *lysR* regulator	*S. coelicolor* SC5G9.07c	60/69	AL17385

[a]Closest homologue, except *spnR* which showed higher sequence similarity to non-actinomycetes.
[b]% deduced amino acid sequence identities (ID) or similarities (SM).

experiments for at least three transcripts emanating from this region. When *Bam*HI fragments B or C from pRHB2C10 (Fig. 2) were cloned into pOJ260 and introduced into *S. spinosa* by conjugation and homologous recombination into the chromosome, spinosyn production was not affected. Had these *Bam*HI fragments been internal to a transcription unit, the termination signals in the vector would have disrupted expression of downstream information, and spinosyn biosynthesis would have been blocked. These particular fragments must therefore span transcriptional start points for *spnD* and *spnE*, respectively. As noted above, insertion of internal fragments of *spnA*, *spnD* and *spnE* disrupted spinosad biosynthesis.

The first PKS ORF (*spnA*) encodes a loading module and the first extender module. The loading module begins with an unusual KS domain that lacks the active site cysteine at position 174. Instead, it contains a glutamine residue that is characteristic of initiator decarboxylase domains [20]. The first extender module encodes the following domains: β-ketoacyl synthase (KS), acyltransferase (AT), β-ketoacyl reductase (KR) and acyl carrier protein (ACP). The second PKS ORF (*spnB*) encodes the second extender module, which has a full complement of domains: KS, AT, β-hydroxyacyl dehydratase (DH), enoyl reductase (ER), KR and ACP. The third ORF (*spnC*) encodes a bimodular polypeptide comprising extender modules 3 and 4, each containing KS, AT, KR and ACP domains. The fourth ORF (*spnD*) encodes extender module 5 (KS, AT, DH, KR and ACP), extender module 6 (KS, AT, KR and ACP) and extender module 7 (KS, AT, KR and ACP). The last PKS ORF (*spnE*) encodes extender modules 8, 9 and 10, each of which consists of KS, AT, DH, KR and ACP domains. The last module also encodes the carboxy-terminal thioesterase (TE) domain.

Most of the AT domains are predicted to incorporate

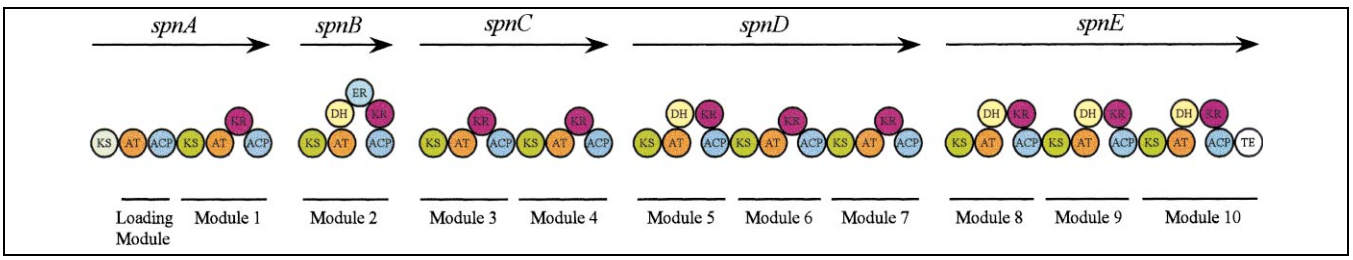


Fig. 3. Modular organization of the spinosyn PKS gene cluster. The direction of transcription of the *spn* genes is indicated by arrows. The multifunctional proteins and their functions in polyketide assembly are illustrated below the genes. KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; TE, thioesterase.

Table 4
Amino acid sequences around the active site of *spn* AT domains[a]

	51	100
AT1	GSDAQLLD**RTLWA**QSGLFAL**QVGLLKLL**DSWGVRPSVVL**GHSVG**ELAAAF	
AT2	GSDAQLLD**RTLWA**QSGLFAL**QVGLLKLL**DSWGVRPSVVL**GHSVG**ELAAAF	
AT4	GSDAQLLD**RTLWA**QSGLFAL**QAGLLGLL**GSWGVRPDVVM**GHSVG**ELAAAF	
AT5	GSQAWLLD**RTVWA**QSGLFAL**QIGLLRLL**GSWGVRPDVVL**GHSVG**ELAAVH	
AT6	GSDTQLLD**QTLWA**QPGLFAL**QVGLWELL**GSWGVRPAVVL**GHSVG**ELAAAF	
AT7	GADEQMLD**RTLWA**QSGIFAV**QVGLLGLL**RSWGVRPAAVL**GHSVG**ELAAAH	
AT9	GSDGSLLD**RTLWA**QSGLFAL**QVGLLSLL**GSWGVRPGVVL**GHSVG**EFAAAV	
AT10	GSDESLLD**RTLWA**QSGLFAL**QVGLWELL**GTWGVRPSVVL**GHSVG**ELAAAF	
Acet Motif	**-T-**YA QVAL**--LL** ***GHSVG***	
ATi	AAGAAHLD**HDDVV**QPALFAI**MVSLAELW**RSWGVRPVAVV**GHSQG**EIAAAC	
AT3	AENSPPLD**RVDVL**QPASFAV**MVSLAEVW**RSYGVRPAAVV**GHSQG**EIAAAC	
AT8	DPDAPALD**RDDVI**QLALFAM**MVSLAELW**RSCGVEPAAVV**GHSQG**EIAAAH	
Prop Motif	***RVDVV*** ***MVSLA*****A-*****W*** ***GHSQG***	

[a]The amino acids in bold in the AT sequences of *S. spinosa* represent a composite of sequences highlighted in earlier comparisons of AT domains for the incorporation of acetate or propionate [21,22]. The consensus acetate and propionate motifs were derived from sequences presented previously [21,22] and from the *S. spinosa* sequences above. Bold italics indicates >95% conservation; bold indicates 76–95% conservation; unbolded letters indicate 50–75% conservation.

only acetate and, as expected, they contained sequences resembling the acetate-specific motif ([21,22]; Table 4). The AT sequences in the initiator module and in extender module 3, as expected, included a motif resembling the propionate-incorporating domains of other PKSs. However, the AT domain of module 8 was also similar to a propionate-incorporating domain, despite the fact that it incorporates propionate only about 20% of the time (leading to the synthesis of spinosyn D). When making the major product, spinosyn A, module 8 incorporates acetate. Clearly, other factors influence the substrate specificity of the AT domain in module 8. An AT domain specifying two different extenders has also been observed in the epothilone PKS [23,24].

2.3.2. Tri-O-methylrhamnose biosynthetic genes

Biosynthetic studies indicated that the attachment of rhamnose to the aglycone is the obligatory first step in the conversion of the aglycone to spinosyns. The three methylations can occur after rhamnose is attached, and occur sequentially on the 2′-, 3′- and 4′-OH groups (C. Broughton, M.L.B. Huber, L.C. Creemer, H.A. Kirst and J.R. Turner, Abstracts of the Annual Meeting of the American Society for Microbiology, Dallas, TX, USA, May 5–9, 1991; H.A. Kirst, L.C. Creemer, M.C. Broughton, M.L.B. Huber and J. R. Turner, Abstracts of the Annual Meeting of the American Chemical Society, Anaheim, CA, USA, March 21–25, 1999). We did not find NDP-glucose synthase or NDP-glucose dehydratase genes within the cluster. These two genes are needed for the conversion of glucose-1-phosphate to NDP-4-keto-6-deoxy-D-glucose, the common intermediate in the biosynthesis of 6-deoxysugars such as rhamnose and forosamine [25]. Genes encoding these enzymes are not linked to the erythromycin biosynthetic cluster in *Saccharopolyspora erythraea* [26], but are present in other polyketide biosynthetic clusters such as avermectin [9], granaticin [27], and tylosin [28]. In addition, the genes required for the conversion of NDP-4-keto-6-deoxy-D-glucose to NDP-rhamnose (an epimerase and a ketoreductase) were not found in the spinosyn cluster. All four of these genes have been cloned and sequenced, and will be described elsewhere (K. Madduri, C. Waldron, and D. Merlo, manuscript submitted).

The cluster does contain three genes that could encode enzymes catalyzing the *O*-methylations of the rhamnosyl residue. They are closely related to genes known to encode deoxysugar *O*-methyltransferases in *Streptomyces fradiae*. The *spnH* gene product showed high sequence similarity to the product of the *tylF* gene that encodes macrocin-*O*-methyltransferase ([29,30]; Table 3). The *spnI* and *spnK* gene products showed high sequence similarity to the product of the *tylE* gene that encodes demethylmacrocin-*O*-methyltransferase ([29,30]; Table 3). In an attempt to determine the methylations specified by *spnH*, *spnI* and *spnK*, each gene was disrupted by using a cloned internal fragment to insert vector DNA into the target by a single homologous recombination event. This procedure generates two truncated, presumably inactive, copies of the gene on either side of the integrated vector sequence. The spinosyns synthesized by each strain under normal fermentation conditions, and when fed the precursors spinosyn P and K, were determined (Table 3). Other submethylated factors (Fig. 1) could not be used in these tests because even the wild-type strain is unable to complete their methylation (C. Broughton, M.L.B. Huber, L.C. Creemer, H.A. Kirst and J.R. Turner, Abstracts of the Annual Meeting of the American Society for Microbiology, Dallas, TX, USA, May 5–9, 1991; H.A. Kirst, L.C. Creemer, M.C. Broughton, M.L.B. Huber and J.R. Turner, Abstracts of the Annual Meeting of the American Chemical Society, Anaheim, CA, USA, March 21–25, 1999). Unfortunately, these disruption experiments were complicated by pleiotropic effects, presumably due to transcriptional effects on downstream genes. Such problems could potentially be avoided by introducing in-frame deletions, but gene replacement technology in *S. spinosa* has only occasionally been successful in our hands. The mutant that contained vector DNA in *spnH* produced spinosyn factor P, which contains the 2′-*O*-methyl but lacks both the 3′- and 4′-*O*-methyl groups of tri-*O*-methylrhamnose. It failed to bioconvert spinosyn K to spinosyn A, also indicating a defect in 4′-*O*-methyltransferase activity. The mutants disrupted in *spnI* or *spnK* produced no spinosyns. The *spnI* mutant was able to convert factor K to

spinosyn A, indicating that it is proficient in the methylation of the 4′ position of rhamnose. It converted factor P, but only to factor J, demonstrating a deficiency in methylation of the 3′ position of rhamnose.

2.3.3. *Forosamine biosynthetic genes*

Gene disruption experiments identified three genes that have some function in forosamine biosynthesis or attachment. Each of the mutants disrupted in *spnN*, *spnO*, or *spnP* accumulated the pseudoaglycone (PSA) containing tri-*O*-methyl rhamnose, but lacking the dimethylamino sugar forosamine (Table 2). The *spnO* gene product showed highest sequence similarity to the *lanS* product of *Streptomyces cyanogenus* [31], and is homologous to 2,3-hexose dehydratases such as *gra*-ORF27 [27] whose enzymatic activity has been demonstrated [32]. The 2,3-enol product of this enzyme is expected to be stabilized by a 3′-ketoreductase equivalent to *gra*-ORF19 [32]. The *spnN* gene product showed high sequence similarity to *gra*-ORF26 (Table 3) and to oxidoreductase homologues from other secondary metabolite gene clusters, including *lanT* from *S. cyanogenus* [31], which has been proposed to function as a 3-ketoreductase [25]. We therefore propose that *spnO* and *spnN* encode the enzymes involved in 3′-deoxygenation during NDP-forosamine biosynthesis (Fig. 4).

The *spnQ*, *spnR* and *spnS* genes, which are clustered with *spnN* and *spnO*, also appear to participate in forosamine biosynthesis, based upon their DNA sequences (Table 3). The *spnQ* gene product showed very high sequence similarities to the *gra*-ORF23 gene product, a putative CDP-4-keto-6-deoxyglucose-3-dehydratase [27]. It was also closely related to the NDP-hexose 3,4-dehydratase homologues from other secondary metabolite biosynthetic pathways in actinomycetes, including landomycin [31] and the glycopeptide chloroeremomycin [33]. We therefore propose that *spnQ* encodes the 3,4-dehydratase activity involved in NDP-forosamine biosynthesis (Fig. 4). This would generate a 3,4-enol that could be stabilized by a ketoreductase, or possibly by an aminotransferase [34]. The latter could be encoded by *spnR*.

The *spnR* product showed high sequence similarity to *lmbS* from the lincomycin pathway of *Streptomyces lincolnensis* ([35]; Table 3), to *strS* from the streptomycin pathways of *Streptomyces glaucescens* and *Streptomyces griseus* [36], and to *tylB* from the tylosin pathway of *S. fradiae* [28]. All of these genes are believed to encode aminotransferases involved in aminosugar biosynthesis [25,28,34,36].

The *spnS* gene product showed high sequence similarity to the *desVI*-encoded *N*,*N*-dimethyltransferase involved in the biosynthesis of NDP-desosamine in *Streptomyces venezuelae* ([37]; Table 3). Like forosamine, desosamine contains a dimethylamino group. We propose that the *spnS* gene encodes a *N*,*N*-dimethyltransferase involved in NDP-forosamine biosynthesis (Fig. 4). From the postulated biosynthetic pathway to forosamine, it appears that the cluster of *spnN*, *spnO*, *spnQ*, *spnR*, and *spnS* genes is sufficient to encode all of the enzymatic functions required to convert NDP-4-keto-6-deoxyglucose to NDP-forosamine. This cluster also contains the forosaminyltransferase gene *spnP* (see below).

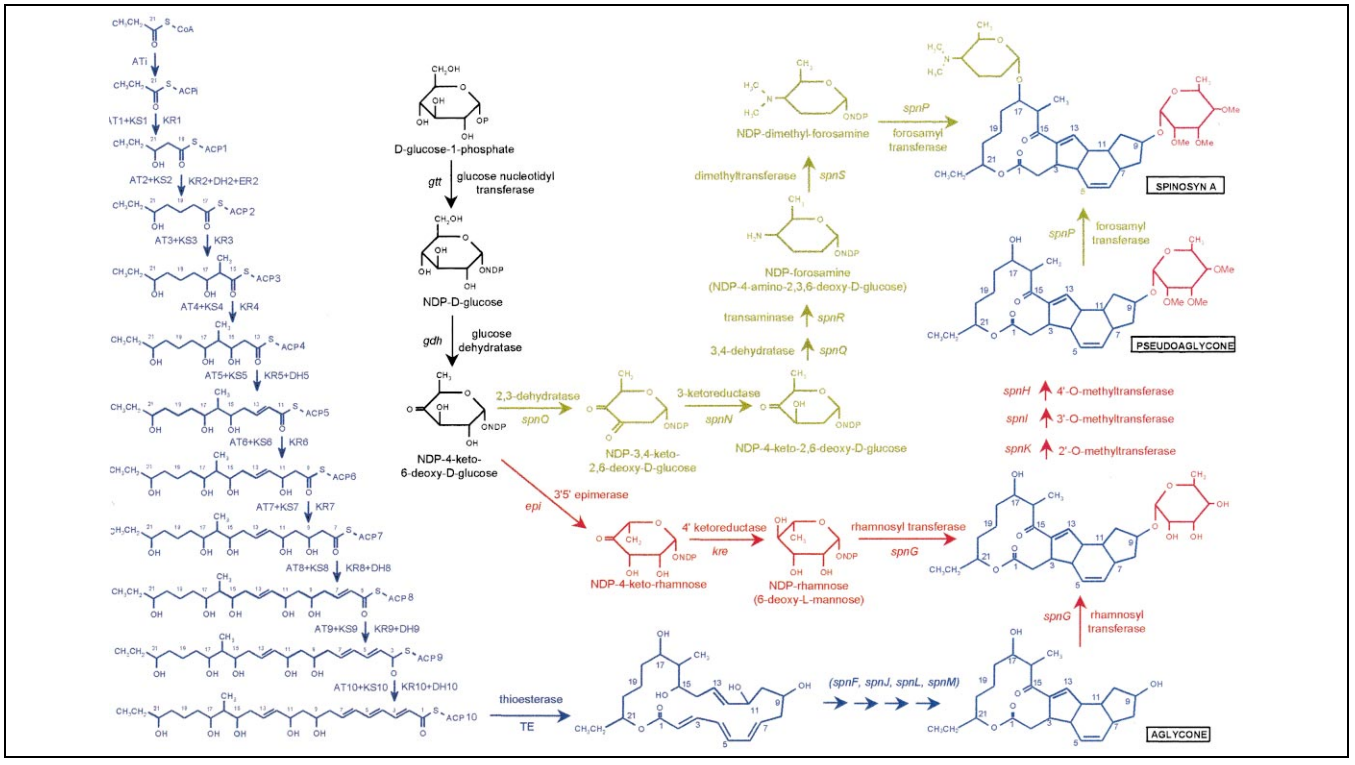


Fig. 4. Proposed biosynthetic pathway to spinosyn A.

2.3.4. Glycosyltransferase genes

The *spnG* gene product was most similar to glycosyltransferase enzymes, and showed highest sequence similarity overall to the *lanGT1* gene product from *S. cyanogenus* ([31]; Table 3). The putative *spnG* product contains the motif for glycosyltransferases [38,39], although it lacks the two adjacent histidine residues that are generally conserved (Table 5). However, the *Saccharothrix aerocolonigenes N*-glycosylation gene NGT also lacks these residues (Table 5). Disruption of *spnG* by insertional mutagenesis abolished the production of spinosad or related factors (Table 2). The recombinant was able to bioconvert PSA, but not AGL, to spinosyn A. This phenotype is consistent with a block in the addition of rhamnose to the aglycone. The lack of accumulation of the reverse PSA (spinosyn lacking rhamnose) is consistent with the suggestion that the addition of rhamnose is the obligatory first step in the conversion of aglycone to spinosyn (H.A. Kirst, L.C. Creemer, M.C. Broughton, M.L.B. Huber and J.R. Turner, Abstracts of the Annual Meeting of the American Chemical Society, Anaheim, CA, USA, March 21–25, 1999). This strict order of sugar additions is similar to that observed in tylosin biosynthesis, where the addition of mycaminose is the obligatory first step in the conversion of tylactone to the tri-glycosylated tylosin [40]. The lack of accumulation of aglycone in the *spnG*-disrupted mutant is noteworthy. It suggests that without subsequent glycosylation the synthesis of AGL is reduced, or its degradation is increased. A similar lack of tylactone accumulation in a *tylM2*-disrupted mutant of *S. fradiae* is due to less polyketide synthesis in the absence of glycosylation [39,41].

The *spnP* gene product showed highest sequence similarity to the desosaminyltransferase enzyme encoded by the *desVII* gene in *S. venezuelae* ([37]; Table 3). It also showed high sequence similarities to the desosaminyltransferase encoded by *eryCIII* of *S. erythraea* [26] and the mycaminosyltransferase encoded by *tylM2* of *S. fradiae* [41]. Like forosamine on spinosyns, desosamine and mycaminose are dimethylamino sugars that are attached to the macrolides pikromycin, erythromycin and tylosin. The *spnP* protein contains the characteristic motif for glycosyltransferases involved in secondary metabolite biosynthesis ([38,39]; Table 5). When the *spnP* gene was disrupted, the recombinant accumulated PSA (Table 2), which lacks forosamine. The mutant bioconverted aglycone to PSA, and therefore retained rhamnosyltransferase activity. We conclude that *spnP* encodes the forosaminyltransferase that converts PSA to spinosad.

Table 5
Sequence motif characteristic for glycosyltransferases involved in secondary metabolite biosynthesis[a]

TylM2	LLPTCSAVVHHGGAGTCFTATLNGLPQIVVA
TylN	LMPRTAAVVHHGGFGTTADAVRAGVPQVLVP
DesVII	LLPSCSAIIHHGGAGTYATAVINAVPQVMLA
GtfB	LFGRVAAVIHHGGAGTTHVAARAGAPQILLP
GtfD	LFPRVAAVIHHGSAGTEHVATRAGVPQLVIP
SpnP	LLESCSVIIHHGSTTTQETATVNGVPQLILP
SpnG	FLRTCELVICAGGSGTAFTATRLGIPQLVLP
NGT	LLPHVDAMVTNGGYGGVNTALAHGVPLVVAR
Consensus	LLP CAAhhHHGGAGT hAhh GhPQhhhP

[a]Amino acid sequences for TylM2 [41], TylN [61], DesVII [37], GtfB [62], GtfD [62], and NGT (accession AB023953) from positions ~283–312 were compared to those from SpnP and SpnG. 'h' indicates a hydrophobic residue.

2.3.5. Polyketide bridging genes

Four genes (*spnF*, *spnJ*, *spnL*, and *spnM*) are presumed to play a role in spinosyn biosynthesis because they are present within the cluster. When each of these genes was disrupted, the resulting strains failed to produce significant amounts of any detectable spinosyns (the *spnM* mutant produced a low level of spinosyn A, presumably because one of the truncated products retained sufficient activity to produce a leaky phenotype). All the mutants were able to bioconvert added AGL or PSA to spinosyn A, a phenotype that is consistent with a defect in polyketide cross-bridging reactions. However, the precise functions of these gene products are difficult to predict because the biochemical mechanism(s) of the cross-bridging reactions are not yet understood. We summarize the information from similarity searches below.

The *spnF* gene product showed highest sequence similarity to the *aveD* gene product of *Streptomyces avermitilis* ([9]; Table 3). The *aveD* gene encodes avermectin B 5-*O*-methyltransferase that methylates the C5 hydroxyl group on the polyketide. The *spnF* gene product also showed high sequence similarity to two proposed *O*-methyltransferase genes (*mitM* and *mitN*) involved in mitomycin biosynthesis in *Streptomyces lavendulae* [42], and to a macrolide *O*-methyltransferase (ORF14) involved in rifampin biosynthesis in *Amycolatopsis mediterranei* [43]. A mutant disrupted for *spnF* function expressed all three *O*-methyltransferases for the production of tri-*O*-methylrhamnose, and the *N*-methyltransferase for forosamine, because it converted AGL to spinosyn A. Therefore, the *spnF* gene does not encode any of the sugar methylating enzymes. Although the lack of production of any spinosyns by the *spnF* mutant is consistent with *spnF* playing a role in AGL formation, we cannot exclude the possibility that the mutant phenotype results from a polar effect on the downstream PKS genes. However, there is sufficient separation of the ORFs (116 nucleotides) for an independent *spnA* promoter.

The *spnL* gene was similar in many ways to *spnF*. Their putative translation products had 34% sequence identity, with two gaps. In both cases, the translation products showed highest sequence similarity to the *aveD* gene prod-

Table 6
Conserved motifs in S-adenosyl-methionine-dependent methyltransferases[a]

	Motif I	Motif II	Motif III
AveD	VLDVGCGSG	GSFDAAWALECLLH	VLRPGGRLAV
Orf14	LLDIGCGNG	ASFDVVWALESLHH	VLRPGGRLAL
MitM	VLDLGCGVG	ESFDAVIALESIIH	VLRPGGRLVL
MitN	VLDVGSGNG	GSFDACYAIESICH	VLRPGGRVTV
SnogM	VLDIGCGTG	DSFDAVWFFESIFH	VLRPGGRLAL
SpnF	LLDVGCGTG	N***A***FDAAWAM***Q***SLL***E***	VLKPGG***I***LGV
SpnL	L***F***DLGCGNG	G***F***F***Q***AAWAM***Q***SVV***Q***	IL***E***PGGRFVL
Consensus	VLD GCG G	SFDA WA ES H	VLRPGGR L

[a]Amino acid sequences for AveD [9], Orf14 [43], MitM [42], MitN [42], and SnogM (accession AF323753) were compared to those from SpnF and SpnL. The D in motif II is the only amino acid conserved in all S-adenosyl-methionine-dependent methyltransferases. Amino acids encoded by the *spn* genes that do not match the consensus are highlighted in bold italics.

uct (Table 3), and had high sequence similarity to the *O*-methyltransferase genes ORF14 from *A. mediterranei*, and *mitM* and *mitN* from *S. lavendulae* (29–38% identity). Mutants disrupted in either *spnF* or *spnL* were able to convert AGL to spinosyn A, indicating that neither gene encodes a deoxysugar methyltransferase activity. The phenotype is consistent with a role in AGL formation, but we cannot exclude the possibility that the spinosyn non-production phenotype results from a polar effect on a downstream gene (*spnM* in the case of *spnL*).

Since both *spnF* and *spnL* have highest sequence similarities, by BLAST analysis, to *O*-methyltransferase genes involved in secondary metabolite production, we looked for the three motifs that are conserved in S-adenosyl-methionine-dependent methyltransferases [44]. Table 6 shows motifs I, II, and III from the *spnF* and *spnL* gene products and five related *O*-methyltransferase enzymes from secondary metabolite-producing actinomycetes. Motif I from SpnF and SpnL has many of the conserved residues found in the related proteins. However, in motifs II and III, there were several substitutions in one or both of the *spn* gene products that were significantly different from the other proteins (bolded in Table 6). Most strikingly, SpnL lacks the very highly conserved aspartate residue in position 4 of motif II that is found in all methyltransferases analyzed previously [42,44]. These data beg the question: are *spnF* and *spnL* functional *O*-methyltransferase genes, or have one or both evolved some related function(s) involved in cross-bridging?

The *spnJ* gene product showed highest sequence similarity to the product of the *Streptomyces peucetius dnrW*, a gene of unknown function (Table 3). It also showed lower but significant sequence similarity to hexose oxidase from *Chondrus crispus* (accession U89770), including an FAD-binding domain. Furthermore, it is related to a number of methylene bridge forming reticuline:oxygen oxidoreductases (berberine bridge enzymes) from *Bacillus* (accession Z82044 and Z99108) and *Arabidopsis* (accession AC0044238). The ability of a disrupted mutant to convert AGL to spinosyn A indicates that *spnJ* disruption has not affected the downstream *spnK* gene (which is believed to encode a rhamnose *O*-methyltransferase activity). The inability of this mutant to produce AGL is therefore solely due to lack of *spnJ* function.

We presume that the *spnM* gene also plays an essential role in AGL formation because its disruption should not affect the downstream gene, which is transcribed from the opposite strand. The *spnM* gene product showed highest sequence similarity to a putative secreted protein (AL158057) from *Streptomyces coelicolor* (Table 3), a hypothetical protein Rv1592c from *Mycobacterium tuberculosis* [45], and a family of secreted lipases from *Candida albicans* [46]. We therefore examined the *spnM* gene product and the *S. coelicolor* hypothetical secreted protein (AL158057) for features associated with secreted lipases. Amino acid alignments by ClustalW [47] indicated that *S. coelicolor* AL158057 has all four conserved cysteine residues observed in the *C. albicans* lipases [46], but SpnM had none of the four (data not shown). *S. coelicolor* AL158057 has a potential signal sequence at its amino-terminus [48,49], but SpnM has a shorter amino-terminal extension without the characteristic features of a signal sequence. It does not therefore appear to be a secreted protein. Bacterial lipases [50] and the *C. albicans* lipases

Table 7
Conserved motif around the catalytic site serine in *C. albicans* family of lipases compared to spnM[a]

Lip5**	AKVVMWGYSGGSLASGWAAAL
Lip1	AKVAMWGYSGGSLASGWAAAL
Lip3	AQVAMWGYSGGTLAAGWAATL
Lip2	AKVALWGYSGGSLATGWAAAL
Lip10	AKVALWGYSGGSLATGWAISL
Lip6	AKVVLWGYSGGSFASGWAAVL
Lip7	VKTALLGYSYGAVASLWASIV
Mt	TPIGLWGYSGGGLASAWAAEA
Sc	SRVGLFGYSQGGGATAAAAEL
SpnM	CPVGIWGY***A***QGGQASAFAGEL
Consensus	V WGYSGG A WAA L

**Lip4, Lip8 and Lip9 have the same sequence as Lip5. The unique substitution for the active site serine encoded by *spnM* is highlighted in bold italics.
[a]Amino acid sequences for Lip1 to Lip10 [46], Mt [45], and Sc (accession AL158057) were compared to that from SpnM.

[46] have an active site serine in a conserved G-X-S-X-G motif. Table 7 shows this conserved sequence and the immediately flanking amino acids from SpnM and its closest homologues. SpnM has an alanine in place of the conserved serine in the active site. The lack of an active site serine strongly suggests that the *spnM* gene product does not function as a lipase, and raises the possibility that SpnM functions by binding a portion of the polyketide molecule containing the ester bond, but does not cleave the ester bond.

2.3.6. Genes flanking the spinosad cluster

Two genes to the left of *spnS* were sequenced and analyzed by insertional mutagenesis. The ORF-L15 gene product showed the highest sequence similarities to a putative short chain oxidoreductase from *S. coelicolor* (Table 3) and to the oxyacyl (ACP) reductase gene *atcC* from *Agrobacterium tumefaciens* (accession U59485). Disruption of ORF-L15 had no discernible effect on spinosad biosynthesis (Table 2). The ORF-L16 gene product showed highest sequence similarity to a putative *lysR* family transcriptional regulator from *S. coelicolor*, and to other members of the *lysR* family. Disruption of ORF-L16 had no discernible effect on spinosad biosynthesis, however.

To the right of the PKS genes are two ORFs, one of which was disrupted. ORF-R1 encodes a small protein (199 amino acids) that is similar to a gene of unknown function from *S. coelicolor* (Table 3). It was too small to be disrupted by the technology available, so we were unable to determine if it plays a role in spinosad biosynthesis. The ORF-R2 gene product showed highest sequence similarity to the putative *recD* gene product, exodeoxyribonuclease V, from *Chlamydia pneumoniae* (Table 3), *Chlamydia trachomatis* (accession AE001278), and *M. tuberculosis* (accession Z92772). Disruption of ORF-R2 had no effect on spinosad biosynthesis.

From these studies, we can tentatively conclude that ORF-L15, ORF-L16 and ORF-R2 are not involved in spinosad biosynthesis. Further partial sequencing to the right of *spnE* or to the left of *spnS* (Fig. 2) failed to identify, by sequence homologies, any additional candidate spinosad genes (including regulatory genes) in the vicinity of the cluster.

3. Discussion

We cloned and sequenced an 80 kb segment of the *S. spinosa* genome that encodes most of the biosynthetic functions required to assemble spinosad from primary metabolic precursors. Spinosyns share common features with other complex glycosylated polyketides such as erythromycin, tylosin and avermectin. They differ in that the spinosyn polyketide contains three intramolecular cross-bridges. Therefore, the key functions that might be found in the spinosyn gene cluster include those required for polyketide assembly, polyketide cross-bridging, NDP-sugar biosynthesis, glycosyltransfer and possibly regulatory functions.

Our studies identified the polyketide assembly functions in five large genes encoding subunits of a multifunctional type I PKS. The five genes are transcribed unidirectionally, starting from the loading module and apparently proceeding sequentially through all of the assembly steps, ending with a TE function for release and cyclization of the macrolide. Gene disruption studies demonstrated that at least three transcripts are involved in translation of the PKS genes. The precise linear correlation between the gene sequence and the sequence of biochemical reactions needed for polyketide assembly is similar to that observed in erythromycin [26], niddamycin [22], and spiramycin [51]. It differs from the organization in the avermectin [9] and rapamycin [10] pathways. We did not identify a gene for an editing TE similar to that observed in the tylosin cluster [52].

If all the PKS domains function in the same order as they are encoded in the chromosome, then the product of their combined activities should be a novel cyclic macrolactone (Fig. 4). A compound of this structure has not been reported, but only a few modifications would be required to convert it to the first known spinosyn intermediate, AGL. The C15-hydroxyl must be re-oxidized to generate a carbonyl group, the C11-hydroxyl must be eliminated, and bridges must be formed at C3–C14, C4–C12 and C7–C11. We speculate that the products of the *spnF*, *spnJ*, *spnL* and *spnM* genes may play a role in these processes. Intriguingly, at least two of these genes may have evolved to carry out different functions from their most closely related homologues. The *spnM* gene is highly related to known eukaryotic lipase genes from *C. albicans*, but its gene product has an alanine substitution for the active site serine involved in lipase activity. It also lacks a signal sequence and conserved cysteine residues believed to participate in disulfide bridges to stabilize secreted lipases. It is conceivable that SpnM binds the fatty acyl-ester portion of spilactone, which mimics an acyl glycerol structure, but does not catalyze the breakage of the ester bond. Instead, it may lock the polyketide in position to participate in cross-linking reactions. The *spnF* and *spnL* genes are clearly related to macrolide *O*-methyltransferase genes from other actinomycetes, but they have a number of amino acid substitutions that eliminate serine or histidine residues conserved in closely related *O*-methyltransferases, and have substitutions that eliminate charged residues in motifs II or III. In particular, *spnL* has a glutamine substitution for the very highly conserved aspartate in motif II of methyltransferases. This raises the possibility that the product of *spnL* (and possibly that of *spnF*) lacks *O*-methyltransferase activity and functions in some other way to facilitate the cross-linking reactions. Alternatively, SpnF and SpnL might catalyze a coupled methylation/demethylation that is involved in cross-linking. (A demethylation

step would be required since the AGL does not contain any *O*-methyl groups.) Methylesterification of a carboxyl group has been shown to activate a cyclization reaction in doxorubicin biosynthesis [53]. In this case, removal of the methyl group is proposed to occur after subsequent reduction and glycosylation reactions are completed. SpnJ might function to oxidize the hydroxyl at position 15 to the ketone present in spinosad (Fig. 4), completing the modification of spilactone to AGL.

Although these models are speculative, the data suggest that at least two of the genes (*spnM* and *spnL*) have evolved to maintain substrate-binding capacities while losing their original enzymatic activities to take on new functions in secondary metabolite biosynthesis. A precedent for this is provided by the KSQ domain present in some type I PKS starter modules [20]. KSQ enzymes have a conserved glutamine residue instead of the active site cysteine residue needed for KS activity. KSQ enzymes lack KS activity, but retain decarboxylase activity.

Assignment of functions by sequence similarities alone was most difficult for the genes involved in the cross-linking reactions. This is to be expected, since these types of enzymatic reactions have not been well studied in *S. spinosa*, or in other secondary metabolite-producing microorganisms. However, cross-bridging has been observed in other complex macrolides [2,11,12], so it will be informative to see if similar genes are found in the producers of these compounds.

NDP-sugar biosynthesis is interesting in *S. spinosa* in that none of the genes required for NDP-rhamnose biosynthesis is located in the spinosyn cluster. Furthermore, it appears that a single set of rhamnose genes is shared for cell wall biosynthesis and spinosyn biosynthesis (Madduri et al., submitted). The first two genes in rhamnose biosynthesis are also required for forosamine biosynthesis. The sharing of deoxysugar biosynthetic genes for primary and secondary metabolism differs from the general observation that complex polyketide pathways have dedicated sugar biosynthetic genes clustered with the other secondary metabolic genes. Understanding the mechanism by which the *S. spinosa* sugar biosynthetic genes are regulated during vegetative growth and fermentation will be an important question to address.

The DNA sequence and insertional mutagenesis analyses of spinosyn genes did not identify any good candidates for regulatory genes. The *lysR* homologue to the immediate left of the cluster was the most likely candidate, but insertional mutagenesis failed to demonstrate its role in spinosyn biosynthesis. We cannot rule out the possibility that this gene plays a role under different physiological conditions, but it seems more likely that it functions in another metabolic pathway. A similar lack of serious candidates for regulatory genes within a macrolide biosynthetic cluster was also observed in *S. erythraea*, the producer of erythromycin [26]. Presumably the key regulatory genes are unlinked to spinosyn biosynthetic genes. Other approaches, such as genomic sequence analysis and gene disruption, will be required to identify these genes.

4. Significance

Spinosad is an important new insecticidal macrolide that is relatively devoid of toxicity for mammals, plants and non-target insects. As such, it provides a novel chemical scaffold for further embellishment of this important activity. The cloning and sequence analysis of the spinosyn biosynthetic cluster provide the starting materials for the biosynthetic modification of polyketide assembly and glycosylation to produce novel derivatives of spinosad. It also provides genes that can be used in spinosad yield enhancement. The spinosyn molecule contains three intramolecular cross-bridges that could be carried out by novel enzymes encoded by genes in the spinosad cluster. The cloned genes provide the starting materials to study these unusual polyketide tailoring enzymes.

5. Materials and methods

5.1. Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 1. The *S. spinosa* mutants blocked in spinosyn A biosynthesis were induced by *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine mutagenesis [54], and were identified by fermentation and high performance liquid chromatography (HPLC) analysis [2].

5.2. Media, growth conditions, fermentations and bioconversions

E. coli cells were grown at 37°C in LB broth or on LB agar (Difco Laboratories, Detroit, MI, USA). *S. spinosa* strains were grown at 29°C in CSM broth [13]. *S. spinosa* transconjugants were selected on R6 medium [13], then maintained on BHI agar (Brain Heart Infusion, Difco). Apramycin (obtained from K. Merkel, Eli Lilly and Co.) was used as a selection agent at 100 μg/ml for *E. coli* and 50 μg/ml for *S. spinosa*. Ampicillin was added to 100 μg/ml, or kanamycin to 50 μg/ml, for selection of plasmid-containing *E. coli* cells.

For fermentations, *S. spinosa* strains were first grown for 3 days at 29°C in CSM broth. From this seed culture, 0.3 ml was inoculated into 25 ml of fermentation media similar to those described by Strobel and Nakatsukasa [55], and grown for another 10 days under the same conditions. For bioconversion studies, spinosyns were added to a final concentration of 100 μg/ml in the fermentation medium at the same time as the seed culture.

5.3. Analysis of spinosyn biosynthesis and bioconversion

Cultures were extracted with three volumes of acetonitrile, followed by centrifugation at 3500 rpm for 10 min. The liquid phase was passed through a 0.2 μm filter and analyzed by isocratic

HPLC in a Beckman Gold system (Beckman Instruments, Palo Alto, CA, USA) using a C-18 reverse-phase column (Waters Radial-Pak Cartridge, type 8NVC184m) with a Waters RCM 8×10 Module (Millipore Corp., Milford, MA, USA). The column was developed at a flow rate of 2 ml/min for 30 min with acetonitrile/methanol/2% ammonium acetate (42.5:42.5:15), and metabolites were monitored at a wavelength of 250 nm. Reference standards of aglycone and pseudoaglycone [56], and spinosyns B, P, H, J and K [2] were provided by Paul Handy (Dow AgroSciences), and dissolved at 100 μg/ml in methanol.

5.4. DNA cloning and sequence analysis

Standard methods for DNA isolation and manipulation were used [57,58]. Oligonucleotide primers for DNA sequencing and PCR amplifications were synthesized on a Model 394 DNA/RNA Synthesizer (Applied Biosystems Inc., Foster City, CA, USA). Plasmids were sequenced using ABI Prism Ready Reaction Cycle Sequencing kits, and analyzed on an ABI373 Automated DNA Sequencer (Applied Biosystems Inc.). Cosmids pRHB3E11 and pRHB2C10 were sequenced at SeqWright (Houston, TX, USA). PCR reactions with AmpliTaq polymerase were carried out in GeneAmp Thermocyclers, using the conditions recommended by the manufacturer (Perkin Elmer, Foster City, CA, USA). Southern hybridizations of genomic DNA were performed according to manufacturer's conditions, with digoxigenin-labeled probes prepared by PCR amplification (Boehringer Mannheim, Indianapolis, IN, USA).

5.5. Construction of S. spinosa recombinants

Plasmids were constructed using standard protocols [57,58]. Cosmid pOJ436 [17] containing large inserts of *S. spinosa* DNA [13], and plasmid pOJ260 [17] containing internal fragments of *spn* genes were introduced into *S. spinosa* strains by conjugation from *E. coli* S17-1 [59] and homologous recombination into the chromosome as described previously [13].

Acknowledgements

DNA sequencing was performed at Eli Lilly by Stan Burgett, Mike Greaney, and Pam Rockey, and at Dow AgroSciences by Lyn Wegrich and Beth Rubin-Wilson. We also thank Eric Cundliffe and Dick Hutchinson for fruitful discussions, Herb Kirst and Paul Handy for samples of spinosyn factors and related compounds, and Dave Berard and Paul Lewer for chemical and biological analyses of disrupted mutants.

References

[1] H.A. Kirst, K.H. Michel, J.W. Martin, L.C. Creemer, E.H. Chio, R.C. Yao, W.M. Nakatsukasa, L. Boeck, J.L. Occolowitz, J.W. Paschal, J.B. Deeter, N.D. Jones, G.D. Thompson, A83543A-D, unique fermentation-derived tetracyclic macrolides, Tetrahedron Lett. 32 (1991) 4839–4842.

[2] H.A. Kirst, K.H. Michel, J.S. Mynderse, E.H. Chio, R.C. Yao, W.M. Nakatsukasa, L. Boeck, J.L. Occolowitz, J.W. Paschal, J.B. Deeter, G.D. Thompson, Discovery, isolation, and structure elucidation of a family of structurally unique, fermentation-derived tetracyclic macrolides, in: D.R. Baker, J.G. Fenyes, J.J. Steffens (Eds.), Synthesis and Chemistry of Agrochemicals III, American Chemical Society, Washington, DC, 1992, pp. 214–225.

[3] F.P. Mertz, R.C. Yao, *Saccharopolyspora spinosa* sp. nov. isolated from soil collected in a sugar mill rum still, Int. J. Syst. Bacteriol. 40 (1990) 34–39.

[4] C.V. DeAmicis, J.E. Dripps, C.J. Hatten, L. Karr, Physical and biological properties of the spinosyns: novel macrolide pest-control agents from fermentation, in: P.A. Hedin, R.M. Hollingworth, E.P. Masler, J. Miyamoto, D.G. Thompson (Eds.), ACS Symposium Series 658: Phytochemicals for Pest Control, American Chemical Society, Washington, DC, 1997, pp. 144–154.

[5] T.C. Sparks, G.D. Thompson, H.A. Kirst, M.B. Hertline, L.L. Larson, T.V. Worden, S.T. Thibault, Biological activity of the spinosyns, new fermentation derived insect control agents, on tobacco budworm (Lepidoptera: Nocuidae) larvae, J. Econ. Entomol. 91 (1998) 1277–1283.

[6] G.D. Crouse, T.C. Sparks, Naturally derived materials as products and leads for insect control: the spinosyns, Rev. Toxicol. 2 (1998) 133–146.

[7] T.C. Sparks, G.D. Thompson, H.A. Kirst, M.B. Hertlein, J.S. Mynderse, J.R. Turner, T.V. Worden, Fermentation-derived insect control agents: the spinosyns, Methods Biotechnol. 5 (1999) 171–188.

[8] S. Donadio, M.J. Staver, J.B. McAlpine, S.J. Swanson, L. Katz, Modular organization of genes required for complex polyketide biosynthesis, Science 252 (1991) 675–679.

[9] H. Ikeda, T. Nomoniya, M. Usami, T. Ohta, S. Omura, Organization of the biosynthetic gene cluster of the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*, Proc. Natl. Acad. Sci. USA 96 (1999) 9509–9514.

[10] T. Schwecke, J.F. Aparicio, I. Molnar, A. Konig, L.E. Khaw, S.F. Haydock, M. Oliynnnk, P. Caffrey, J. Cortes, J.B. Lester, G.A. Bohm, J. Staunton, P.F. Leadlay, The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin, Proc. Natl. Acad. Sci. USA 92 (1995) 7839–7843.

[11] D. Cane, C.C. Yang, Nargenicin biosynthesis: late stage oxidations and absolute configuration, J. Antibiot. 38 (1984) 423–426.

[12] P.R. Graupner, S. Thornburgh, J.T. Mathieson, E.L. Chapin, G.M. Kemmit, J.M. Brown, C.E. Snipes, Dihydromaltophilin; a novel fungicidal tetramic acid containing metabolite from *Streptomyces* sp., J. Antibiot. 50 (1997) 1014–1091.

[13] P. Matsushima, M.C. Broughton, J.R. Turner, R.H. Baltz, Conjugal transfer of cosmid DNA from *Escherichia coli* to *Saccharopolyspora spinosa*: effects of chromosomal insertions on macrolide production, Gene 146 (1994) 39–45.

[14] P. Matsushima, R.H. Baltz, Transformation of *Saccharopolyspora spinosa* protoplasts with plasmid DNA modified in vitro to avoid host restriction, Microbiology 140 (1994) 139–143.

[15] R.H. Baltz, Genetic manipulation of antibiotic-producing *Streptomyces*, Trends Microbiol. 6 (1998) 76–83.

[16] M. Basinski, M. Bierman, B. Schoner, Cloning of polyketide biosynthetic genes using colony hybridization and PCR, Dev. Indust. Microbiol. 33 (1993) 237–241.

[17] M. Bierman, R. Logan, K. O'Brien, E.T. Seno, R.N. Rao, B.E. Schoner, Plasmid cloning vectors for the transfer of DNA from *Escherichia coli* to *Streptomyces* spp., Gene 116 (1992) 43–49.

[18] S.F. Altschul, W. Gish, W. Miller, W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 251 (1990) 403–410.

[19] W.R. Pearson, Rapid and sensitive sequence comparison with FASTP and FASTA, Methods Enzymol. 183 (1990) 63–98.

[20] C. Bisang, P.F. Long, J. Cortes, J. Westcott, J. Crosby, A.-L. Ma-

tharu, R.J. Cox, T.J. Simmpson, J. Staunton, P.F. Leadlay, A chain initiation factor common to both modular and aromatic polyketide synthases, Nature 401 (1999) 502–505.

[21] S.F. Haydock, J.F. Aparicio, I. Molnar, T. Schwecke, L.E. Khaw, A. Konig, A.F. Marsden, I.S. Galloway, J. Staunton, P.F. Leadlay, Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases, FEBS Lett. 374 (1995) 246–248.

[22] S.J. Kakavas, L. Katz, D. Stassi, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, J. Bacteriol. 179 (1997) 7515–7522.

[23] I. Molnar, T. Schupp, M. Ono, R.E. Zirkle, M. Milnamoow, B. Nowak-Thompspon, N. Engle, C. Toupet, A. Stratmann, D.D. Cyr, J. Gorlach, J.M. Mayo, A. Hu, S. Goff, J. Schmid, J.M. Ligon, The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90, Chem. Biol. 7 (2000) 97–109.

[24] B. Julian, S. Shah, R. Ziermann, R. Goldman, L. Katz, C. Khosla, Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*, Gene 249 (2000) 153–160.

[25] A. Trefzer, J.A. Salas, A. Bechthold, Genes and enzymes involved in deoxysugar biosynthesis, Nat. Prod. Rep. 16 (1999) 283–299.

[26] L. Katz, Manipulation of modular polyketide synthases, Chem. Rev. 97 (1997) 2557–2575.

[27] K. Ichinose, D.J. Bedford, D. Tornus, A. Bechthold, M.J. Bibb, W.P. Revill, H.G. Floss, D.A. Hopwood, The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tu22: sequence analysis and expression in a heterologous host, Chem. Biol. 5 (1998) 647–659.

[28] L.A. Mershon-Davies, E. Cundliffe, Analysis of five tylosin biosynthetic genes from the *tylIBA* region of the *Streptomyces fradiae* genome, Mol. Microbiol. 13 (1994) 349–355.

[29] N. Bate, E. Cundliffe, The mycinose-biosynthetic genes of *Streptomyces fradiae*, producer of tylosin, J. Indust. Microbiol. Biotechnol. 23 (1999) 118–122.

[30] R. Fouces, E. Mellado, B. Diez, J.L. Barredo, The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organization of the left region, Microbiology 145 (1999) 855–868.

[31] L. Westrich, S. Domann, B. Faust, D. Bedford, D.A. Hopwood, A. Bechthold, Cloning and characterization of a gene cluster from *Streptomyces cyanogenus* S136 probably involved in landomycin biosynthesis, FEMS Microbiol. Lett. 170 (1999) 381–387.

[32] G. Draeger, S.-H. Park, H.G. Floss, Mechanism of the 2-deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin, J. Am. Chem. Soc. 121 (1998) 2611–2612.

[33] A.M.A. van Wageningen, P.N. Kiekpatrick, D.H. Williams, B.R. Harris, J.K. Kershaw, N.J. Lennard, M. Jones, S.J. Jones, P.J. Solenberg, Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic, Chem. Biol. 5 (1998) 155–162.

[34] J.S. Thorson, S.F. Lo, H.-W. Liu, C.R. Hutchinson, Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, and amino sugar construction, J. Am. Chem. Soc. 115 (1993) 6993–6994.

[35] U. Peschke, H. Schmidt, H.-Z. Zhang, W. Piepersberg, Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11, Mol. Microbiol. 16 (1995) 1137–1156.

[36] W. Piepersberg, Molecular biology, biochemistry, and fermentation of aminoglycoside antibiotics, in: W. Strohl (Ed.), Biotechnology of Antibiotics, Marcel Dekker, Inc., New York, 1997, pp. 81–163.

[37] Y. Xue, L. Zhao, H.W. Liu, D.H. Sherman, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity, Proc. Natl. Acad. Sci. USA 95 (1998) 12111–12116.

[38] E. Cundliffe, S.A. Fish, V.T.W. Wilson, A. Gandecha, S.L. Large, Glycosyl transfer during tylosin production in *Streptomyces fradiae*, Dev. Indust. Microbiol. 34 (1997) 109–114.

[39] S.A. Fish, E. Cundliffe, Stimulation of polyketide metabolism in *Streptomyces fradiae* by tylosin and its glycosylated precursors, Microbiology 143 (1997) 3871–3876.

[40] R.H. Baltz, E.T. Seno, Genetics of *Streptomyces fradiae* and tylosin biosynthesis, Ann. Rev. Microbiol. 42 (1988) 547–574.

[41] A.A.R. Gandecha, S.L. Large, E. Cundliffe, Analysis of four tylosin biosynthetic genes from the *tylLM* region of the *Streptomyces fradiae* genome, Gene 184 (1997) 197–203.

[42] Y. Mao, M. Varoglu, D.H. Sherman, Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564, Chem. Biol. 6 (1999) 251–263.

[43] P.R. August, L. Tang, Y.J. Yoon, S. Ning, R. Muller, T.-W. Yu, M. Taylor, D. Hoffman, C.-G. Kim, X. Zhang, C.R. Hutchinson, H.G. Floss, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699, Chem. Biol. 5 (1998) 69–79.

[44] R.M. Kagen, S. Clarke, Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes, Arch. Biochem. Biophys. 310 (1994) 417–427.

[45] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglemeier, S. Gas, C.E. Barry, F. Tekaia III, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Conner, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.

[46] B. Hube, F. Stehr, M. Bossenz, A. Mazur, M. Kretschmar, W. Schafer, Secreted lipases of *Candida albicans*: cloning, characterization and expression analysis of a new gene family with at least ten members, Arch. Microbiol. 174 (2000) 362–374.

[47] J.D. Thompson, D.J. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[48] R.H. Baltz, Gene expression in recombinant Streptomyces, Bioprocess. Technol. 22 (1995) 309–381.

[49] M. Gilbert, R. Morosoli, F. Shareck, D. Kluepfel, Production and secretion of proteins by *Streptomyces*, Crit. Rev. Biotechnol. 15 (1995) 13–39.

[50] J.L. Arpigny, K.E. Jaeger, Bacterial lipolytic enzymes: classification and properties, Biochem. J. 343 (1999) 177–183.

[51] S. Kuhstoss, M. Huber, J.R. Turner, J.W. Paschal, R.N. Rao, Production of a novel polyketide through the construction of a hybrid polyketide synthase, Gene 183 (1996) 231–236.

[52] A.R. Butler, N. Bate, E. Cundliffe, Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*, Chem. Biol. 6 (1999) 287–292.

[53] S.G. Kendrew, K. Katayama, E. Deutch, K. Madduri, C.R. Hutchinson, DnrD cyclase involved in the biosynthesis of doxorubicin: purification and characterization of the recombinant enzyme, Biochemistry 38 (1999) 4794–4799.

[54] R.H. Baltz, Mutagenesis in *Streptomyces*, in: A.L. Demain, N.A. Soloman (Eds.), Manual of Industrial Microbiology and Biotechnology, American Society for Microbiology, Washington, DC, 1986, pp. 184–190.

[55] R.J. Strobel Jr., W.M. Nakatsukasa, Response surface methods for optimizing *Saccharopolyspora spinosa*, a novel macrolide producer, J. Ind. Microbiol. 11 (1993) 121–127.

[56] L.C. Creemer, H.A. Kirst, J.W. Pashal, Conversion of spinosyn A and spinosyn D to their respective 9- and 17-pseudoaglycones and their aglycones, J. Antibiot. 51 (1998) 795–800.

[57] D.A. Hopwood, M.J. Bibb, K.F. Chater, T. Kieser, C.J. Bruton, H.M. Kieser, D.J. Lydiate, C.P. Smith, J.M. Ward, H. Schrempf,

Genetic Manipulation of *Streptomyces*: a Laboratory Manual, John Innes Foundation, Norwich, 1985.

[58] T. Maniatis, E.F. Fritsch, J. Sambrook, Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press, Plainview, NY, 1982.

[59] R. Simon, U. Preifer, A. Puhler, A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram negative bacteria, Bio/Technology 1 (1983) 784–791.

[60] J.F. Aparacio, A.J. Colina, E. Ceballos, J.F. Martin, The biosynthetic cluster for the 26-membered ring polyene macrolide pimaricin, J. Biol. Chem. 274 (1999) 10133–10139.

[61] V.T. Wilson, E. Cundliffe, Characterization and targeted disruption of a glycosyltransferase gene in the tylosin producer, *Streptomyces fradiae*, Gene 214 (1998) 95–100.

[62] P.J. Solenberg, P. Matsushima, D.R. Stack, S.C. Wilkie, R.C. Thompson, R.H. Baltz, Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*, Chem. Biol. 4 (1997) 195–202.