Chemistry & Biology, Vol. 10, 431–441, May, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S1074-5521(03)00092-9

A Complete Gene Cluster from *Streptomyces nanchangensis* NS3226 Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin

Yuhui Sun,[1,3] Xiufen Zhou,[1,2] Hui Dong,[4] Guoquan Tu,[3] Min Wang,[4] Bofei Wang,[4] and Zixin Deng[1,2,*]

[1]Bio-X Life Science Research Center
Shanghai Jiaotong University
Shanghai 200030
[2]Huazhong Agricultural University
Wuhan 430070
[3]Jiangxi Agricultural University
Nanchang 330045
[4]Chinese National Human Genome Center at Shanghai
Shanghai 201203
China

Summary

The PKS genes for biosynthesis of the polyether nanchangmycin are organized to encode two sets of proteins (six and seven ORFs, respectively), but are separated by independent ORFs that encode an epimerase, epoxidase, and epoxide hydrolase, and, notably, an independent ACP. One of the PKS modules lacks a corresponding ACP. We propose that the process of oxidative cyclization to form the polyether structure occurs when the polyketide chain is still anchored on the independent ACP before release. 4-*O*-methyl-L-rhodinose biosynthesis and its transglycosylation involve four putative genes, and regulation of nanchangmycin biosynthesis seems to involve activation as well as repression. In-frame deletion of a KR6 domain generated the nanchangmycin aglycone with loss of 4-*O*-methyl-L-rhodinose and antibacterial activity, in agreement with the assignments of the PKS domains catalyzing specific biosynthetic steps.

Introduction

Many pharmaceutically important natural products are produced by modular type I polyketide synthases (PKS) [1]. Each condensation cycle usually needs a module consisting of a β-ketoacyl synthase (KS) domain for condensing the next carboxylic acid onto the growing polyketide chain, an acyltransferase (AT) domain for loading extension units, an acyl carrier protein (ACP) domain for retention of the growing polyketide chain on the PKS, and optionally, β-ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains determining the reduction state of the incorporated extender unit. Release and, sometimes, cyclization of the polyketide chain to form a macrolide ring depends on a thioesterase (TE). The flexibility of the assembly line for polyketide biosynthesis has enabled many successful manipulations to be made to generate novel polyketide derivatives by changing the number of modules, their specificities toward carboxylic acids, or by inactivating or inserting domains with reductive activities [2–4]. The discovery of genes encoding enzymes for biosynthesis of more diverse polyketides (see [5, 6], for recent examples) will facilitate further targeted design of unnatural natural compounds.

Polyether antibiotics are a group of polyketides that affect cation transport in mitochondria. They often inhibit gram-positive bacteria, including mycobacteria, and fungi. They are also used as growth promotants in ruminants. They adopt a cyclic conformation by concentrating oxygen functions at their center where they complex a suitable cation, e.g., sodium in the case of dianemycin [7, 8]. The branched alkyl groups on the outer surface make the molecule lipid soluble so that cations can be conducted across membranes. The bis-spiroacetal ring system [9, 10], unique for this group of compounds, suggests novel enzymatic steps in the biosynthetic pathway. DNA sequence of a partial gene cluster (103,450 bp) involved in the biosynthesis of a polyether antibiotic monensin in *Streptomyces cinnamonensis*, which is quite similar to nanchangmycin cluster, has become available online (accession number AF440781).

A region of ~132 kb of the genome of the producer of nanchangmycin, *S. nanchangensis* NS3226, was implicated in nanchangmycin biosynthesis by bioassay and HPLC after targeted gene replacement [11]. The structure of nanchangmycin suggests involvement of at least 70 kb of type I PKS-encoding DNA, enough for the 14 modules needed for biosynthesis of its polyketide backbone before polycyclic ether bond formation. Synthesis of the nanchangmycin polyketide chain would start with malonyl-CoA, and proceed by condensation of four malonyl-CoA and ten methylmalonyl-CoA extender units. We describe an analysis of the complete nanchangmycin biosynthetic gene cluster and generation of a mutant strain by gene disruption and replacement that produced a novel nanchangmycin derivative with the structure expected from gene assignments based on sequence analysis.

Results and Discussion

The Nanchangmycin Gene Cluster and Its PKS Genes

Four pHZ1358-derived cosmids putatively covering the entire region involved in nanchangmycin biosynthesis [11] were randomly subcloned into *Escherichia coli* and sequenced. Computer-assisted analysis of the DNA sequence (132,544 bp) of the cloned region led to identification of the genes shown in Figure 1 and listed in Table 1. Eleven ORFs (*nanA1–nanA11*) encoding numerous typical type I PKS subunits were identified. The amino acid (aa) sequences of the deduced products encoded by these genes were analyzed by comparing them to those of known type I PKSs. The predicted functional features of the NanA1–NanA11 proteins are shown in Figure 2. A separate module of active sites would be

*Correspondence: zxdeng@sjtu.edu.cn

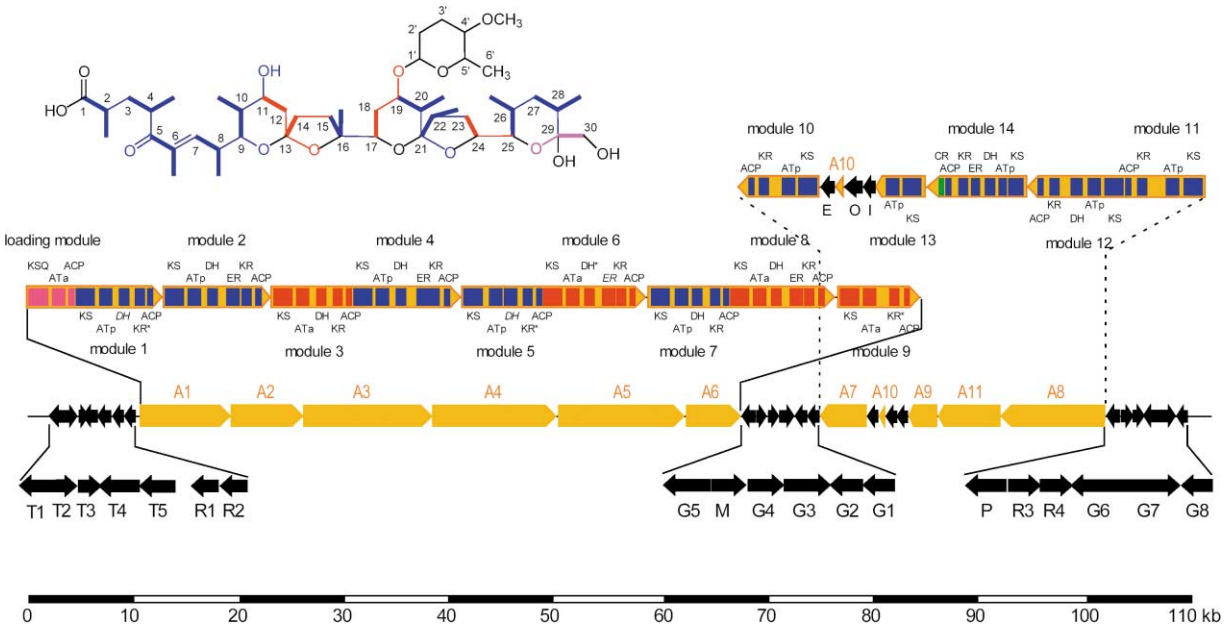

Figure 1. Nanchangmycin, the Chemical Structure and Genetic Organization of Its Gene Cluster

Top: nanchangmycin [44], a polyether ionophore antibiotic with similar structure to dianemycin. Bold lines indicate the building units used by the PKS for polyketide biosynthesis using acetate (29–30) as a starter followed by condensation steps involving four malonyl and ten methylmalonyl extender units. The carbon atoms of the deoxysugar 4-*O*-methyl-L-rhodinose attached at C-19 are labeled 1′–6′. Bottom: organization of the *nan* gene cluster of *S. nanchangensis* NS3226. The ORFs in orange represent the modular PKS genes, whose organization is detailed above, with purple and red representing the domains of the modules responsible for the loading and extension of a malonate unit (whose specific AT is labeled as ATa), respectively; blue for the extension by a methylmaolonate unit (whose specific AT is labeled as ATp); and green for the CR domain. Black-arrowed ORFs are non-PKS genes, whose proposed functions are detailed in the text and in Table 1. Asterisks indicate inactive domains; two presumably silent DH domains, and one silent ER domain are shown in italics.

used for each cycle of polyketide chain extension, as for modular PKSs for macrolide biosynthesis [2].

Fourteen condensation steps should be required for production of the carbon skeleton of nanchangmycin, and in agreement with this, there are 14 PKS modules distributed among 11 ORFs. Two modules are contained in each of five ORFs (*nanA1, nanA3, nanA4*, *nanA5*, and *nanA8*), and another five ORFs (*nanA2, nanA6, nanA7*, *nanA9*, and *nanA11*) each carry one module. Interestingly, one ORF (*nanA9*) carries an incomplete module, with only KS and AT domains, but an independent ORF carrying a single ACP (*nanA10*) was found downstream of *nanA9*, separated by two non-PKS genes presumably required for nanchangmycin biosynthesis. Genes encoding this type of ACP, and a type I PKS lacking an ACP domain, were reported in the genome of *S. avermitilis*, but for an unidentified compound; it is not known whether they are functional [12].

NanA10 is a 104 aa protein resembling the ACPs of type II PKSs [2], which would serve as an independent ACP for module 13. The possibility of its independence as a separate ORF from the rest of module 13, and its sandwiching between *nanI* (encoding a protein similar to ketosteroid isomerase), *nanO* (encoding a putative epoxidase), and *nanE* (encoding a putative epoxide hydrolase), which are presumably required for the isomerization and unusual oxidative cyclization steps required to produce the polycyclic ether structure (see below), suggest that further PKS elongation catalyzed by module 14 might take place after the polyether structure is formed, while the polyketide chain is still anchored on the NanA10 ACP [13].

The NanA1 protein seems to constitute two modules, with the N-terminal one most likely representing a loading module involved in initiation of nanchangmycin polyketide biosynthesis. The N-terminal loading module of NanA1 starts with a KS domain (KSQ), followed by AT and ACP domains, similar to those in the presumed loading modules of the macrolides tylosin [14], spiramycin and niddamycin [15], and the polyether monensin [13]. Similarly to many other type I PKSs (exceptions are the separate loading modules for biosynthesis of the polyene macrolides nystatin NysA [6] and pimaricin PIMS0 [16]), the loading module in NanA1 is fused to the first condensing module. The presence of a glutamine (Q) residue in the active site is likely to be associated with the decarboxylase activity shown by KSQ domains, which is required for chain initiation after loading a malonyl-CoA unit, to yield the required acetate starter unit [17].

Three KR domains, in modules 1, 5, and 9, appear to be inactive, with two changes in the four conserved active site residues (Figure 3), and the DH domain in module 6 of NanA4 also has a change in the conserved active site residues and therefore must be inactive. As a consequence, the apparently functional DH1, DH5, and ER6 domains would be silent.

The AT domains in the loading module and modules 3, 6, 8, and 9 display features characteristic of AT domains recognizing malonate extenders (ATa), and all other

Table 1. Deduced Functions of ORFs in the Nanchangmycin Biosynthetic Gene Cluster

Polypeptide (ORF)	Amino Acid No.	Proposed Function						
Nan A1	2902	PKS						
Lading module		KSQ	ATa				ACP	
Module 1		KS	ATp	*DH*		KR[a]	ACP	
Nan A2	2223	PKS						
Module 2		KS	ATp	DH	ER	KR	ACP	
Nan A3	4032	PKS						
Module 3		KS	ATa	DH		KR	ACP	
Module 4		KS	ATp	DH	ER	KR	ACP	
Nan A4	3956	PKS						
Module 5		KS	ATp	*DH*		KR[a]	ACP	
Module 6		KS	ATa	DH[a]	*ER*	KR	ACP	
Nan A5	3979	PKS						
Module 7		KS	ATp	DH		KR	ACP	
Module 8		KS	ATa	DH	ER	KR	ACP	
Nan A6	1665	PKS						
Module 9		KS	ATa			KR[a]	ACP	
Nan A7	1646	PKS						
Module 10		KS	ATp			KR	ACP	
Nan A8	3455	PKS						
Module 11		KS	ATp			KR	ACP	
Module 12		KS	ATp	DH		KR	ACP	
Nan A9	802	PKS						
Module 13		KS	ATp					
Nan A10	104	PKS (ACP)						
Nan A11	2187	PKS						
Module 14		KS	ATp	DH	ER	KR	ACP	CR
Nan E	290	Epoxidase hydrolase						
Nan G1	302	Glucose-1-phosphate:TTP thymidylyl transferase						
Nan G2	331	dTDP-D-glucose-4,6-dehydratase						
Nan G3	434	NDP-D-glucose-3,4-dehydratase and perosamine synthase						
Nan G4	346	NDP-D-glucose-4,6-dehydratase, NDP-D-glucose-4-epimerase, NDP-D-glucose-4-reductase						
Nan G5	460	Glycosyl transferase						
Nan G6	524	Glycosyl transferase						
Nan G7	611	Glycosyl transferase						
Nan G8	353	Glycosyl hydrolase						
Nan I	313	Ketosteroid isomerase						
Nan M	305	Methyltransferase						
Nan O	478	Epoxidase						
Nan P	423	Cytochrome P450						
Nan R1	241	Regulatory protein						
Nan R2	253	Regulatory protein						
Nan R3	335	Transcription regulator						
Nan R4	313	Transcription regulator						
Nan T1	456	Integral membrane transport protein						
Nan T2	271	ABC transporter						
Nan T3	233	Two-component response regulator						
Nan T4	399	Chemoreceptor						
Nan T5	396	Two-component sensor histidine kinase						

Italic type, silent domains.
[a] Inactive domains.

modules would have AT domains recognizing methylmalonate (ATp) (Figure 2). This agrees with the expected chemistry of the elongation steps.

Overall, the PKS genes are split into two groups: *nanA1–nanA6* (encoding the loading module and modules 1–9) are transcribed from left to right, while *nanA7–nanA11* (encoding extension modules 10–14) and the set of non-PKS genes *nanI, nanO*, and *nanE* are transcribed divergently. The two sets of PKS genes are separated by a group of six post-polyether modification genes involved in biosynthesis of 4-*O*-methyl-L-rhodinose (see below).

Regulatory Genes

Putative regulatory genes lie on both sides of the gene cluster (Figure 1). The *nanR1* and *nanR2* genes lie upstream of the PKS loading module. Their deduced products both resemble transcriptional activators of antibiotic biosynthetic pathways. The highest scores were with pathway-specific regulatory proteins TylS [18] (38% identity for NanR1 and 36% for NanR2) and TylT [18] (40% identity for NanR1 and 38% for NanR2) from *Streptomyces fradiae*. These transcriptional activators constitute a family of *Streptomyces* antibiotic regulatory proteins (SARPs) [19]. Their N termini contain amino acid

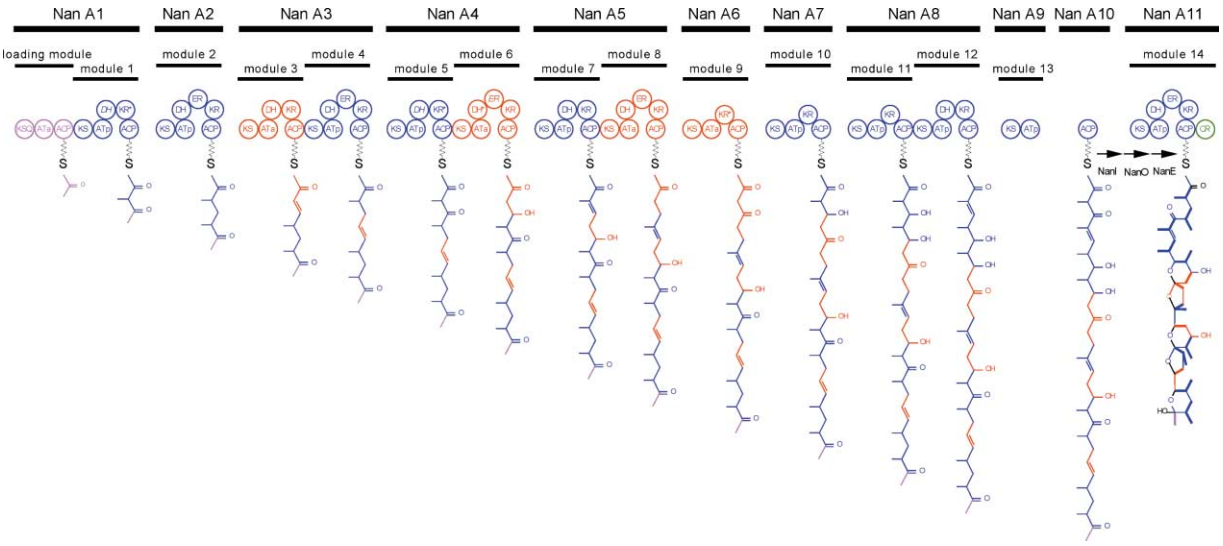

Figure 2. Model for Polyether Biosynthesis of Nanchangmycin in *S. nanchangensis* NS3226

Each circle represents an enzymatic domain in the PKS polypeptide. The meaning of colors, asterisks, and italics conform to Figure 1.

motifs that resemble the DNA binding domain of the C terminus of the *E. coli* activator OmpR. These conserved motifs are also very clear in NanR1 and NanR2. SARPs are believed to interact with specific sequences upstream of *Streptomyces* antibiotic biosynthetic genes to activate transcription. Characteristic SARP binding sites consisting of three multiple tandem repeat sequences between 16 and 19 bp in length were found upstream of *nanT3, nanT5*, and *nanR2*.

Downstream of *nanR1* and *nanR2* lies an obvious two-component signal transduction system, consisting of NanT5 as a putative sensor histidine kinase (21% identity to *Lactococcus lactis* MG1363 [20]) and NanT3, a putative response regulator protein (24% identity to *Pseudomonas solanacearum* VsrD [21]); both had greatest similarity to members of nitrate/nitrite pathway regulators). Between these two genes is *nanT4*, transcribed in the same orientation as *nanT5*, which encodes a putative chemoreceptor. This is reminiscent of bacterial transmembrane signaling, which is mediated by bacterial chemoreceptors by recognizing specific chemicals and regulating a noncovalently associated histidine kinase [22–24]. In the case of nanchangmycin, which most likely serves as an ionophore for sodium, the ligand may bind to the external domain of the membrane-spanning receptor to generate a transmembrane signal that modulates kinase activity inside the cell and, in turn, regulates nanchangmycin production. A putative integral membrane transport protein, NanT1 (48% identity with that of *S. argillaceus* [25]), and a putative ABC transporter, NanT2 (37% identity with spermidine/putrescine ABC transporter of *Vibrio cholerae* [26]), lie at the very left of the *nan* cluster.

At the right side of the PKS gene cluster lies a putative repressor gene (*nanR3*) whose product resembles (29% identity) a transcriptional regulator of the LacI family from many different bacteria and many other transcriptional regulators that control nucleotide biosynthesis [27], catabolic pathways [26, 28], and sugar transport. Next to *nanR3* is *nanR4,* encoding another transcriptional regulator homologous with the AraC family of activators and similar to that of *S. lividans* [29] (32% identity) and many other *Streptomyces*, *Mycobacterium*, *Sinorhizobium*, and *Agrobacterium* species.

Putative Deoxysugar (4-*O*-Methyl-L-Rhodinose) Biosynthesis Genes

The cluster contains five genes presumably involved in biosynthesis and attachment of the deoxysugar moiety (Figure 1; Table 1). NanG5 shows considerable homology to glucosyl transferases from *S. venezuelae* [30] (42% identity). Enzymes of the UDP-glycosyltransferase family are involved in eliminating potentially toxic xenobiotics by glycosylation [31]. NanG5 probably represents a glycosyltransferase that attaches the deoxysugar moiety to the nanchangmycin aglycone at C-19. Two genes encoding a glucose-1-phosphate:TTP thymidylyl transferase (NanG1) and a dTDP-D-glucose-4,6-dehydratase (NanG2) are most likely involved in the early stages of 6-deoxysugar biosynthesis, which are well conserved. NanG1 would catalyze activation of D-glucose-1-phosphate to dTDP-D-glucose and NanG2 would convert dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose. Additionally, multiple homologies were detected with NDP-D-glucose-4,6-dehydratase, NDP-D-glucose-4-epimerase, and NDP-D-glucose-4-reductase for NanG4, and with NDP-D-glucose-3,4-dehydratase for NanG3. These enzymatic functions are just sufficient for biosynthesis of L-rhodinose, as reported for the sugar moieties of granaticin [32] and urdamycin [33]. The sugar moiety (4-*O*-methyl-L-rhodinose) for nanchangmycin biosynthesis needs another function (4-*O*-methylation), which seems to be encoded by *nanM*. The proposed deoxyhexose pathway (4-*O*-methyl-L-rhodinose) is outlined in Figure 4.

Modification Genes

Downstream of *nanA9*, which would encode the KS and AT domains of module 13, are four ORFs, *nanI*, *nanO*,

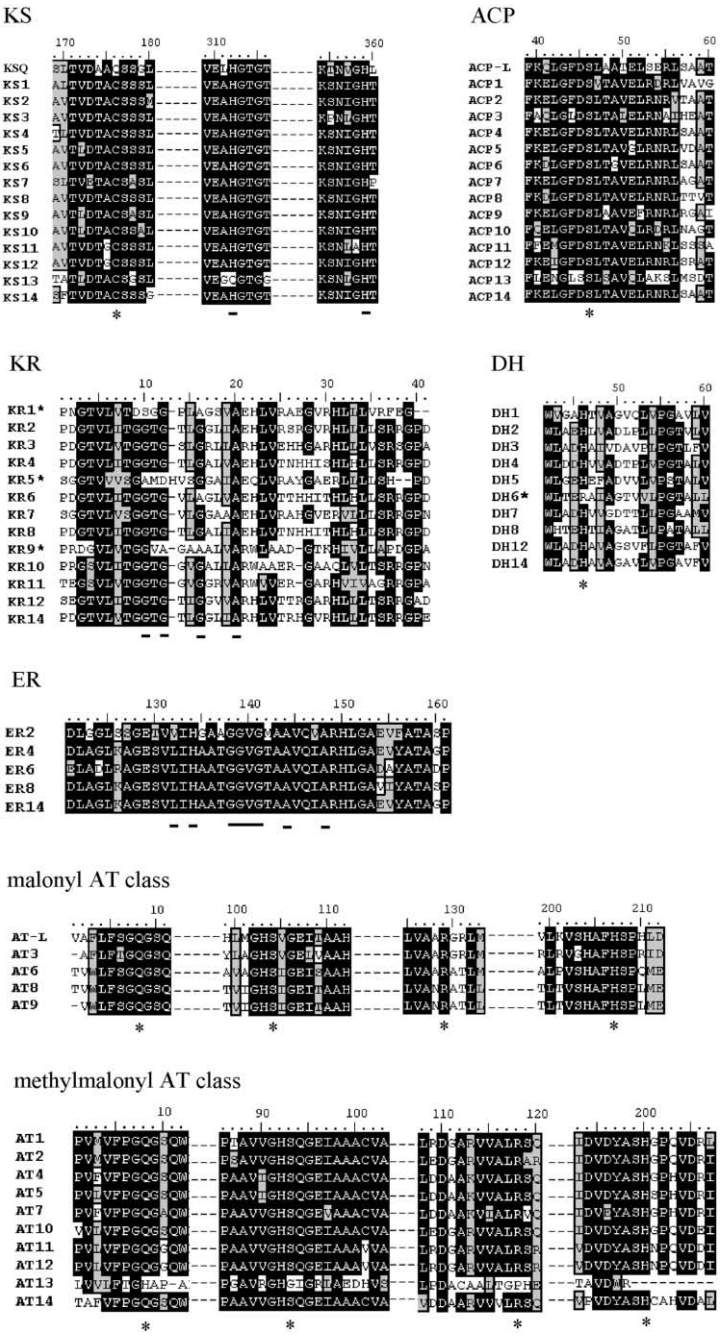

Figure 3. Alignment of Active Site Sequences of Grouped Domains from *nan* PKSs

Only the regions containing the proposed active sites are shown. The numbers correspond to the order in the amino acid sequence at the top of each group. Active site residues are marked with asterisks. Conserved motifs important for the function of each domain are underlined.

nanA10 (for the ACP), and *nanE,* transcribed in the same direction as *nanA9* and *nanA7*. The *nanO* and *nanE* genes would encode an epoxidase (~20% identity to those of many prokaryotes and eukaryotes [34]) and an epoxide hydrolase [35] (38% identity to that of *Caulobacter crescentus*), respectively. NanO is therefore a candidate to carry out two epoxidations of a diene intermediate, and NanE is a candidate for catalyzing the subsequent ring opening of a di-epoxide (Figure 5).

Another predicted enzyme, NanI, shows 50% identity with the Δ^5-3-ketosteroid isomerase from *Pseudomonas testosteroni,* which catalyzes conversion of a Δ^5-3-keto- into a Δ^4-3-ketosteroid [36, 37]. Two such homologs were also discovered in monensin A biosynthesis [13] and were implicated in the *E* to *Z* interconversion of an activated double bond during polyketide chain synthesis (via an extended enolate ion). A similar role could also be assumed in nanchangmycin biosynthesis, although the exact position of *E* to *Z* interconversion remains obscure.

A gene immediately to the right of the PKS gene cluster (*nanP*) encodes a protein resembling cytochrome P450 monooxygenases (41% identity to cytochrome P450 of *Mycobacterium tuberculosis* [38]). This protein is presumably involved in oxidation of the methyl group at C-30 (Figure 5).

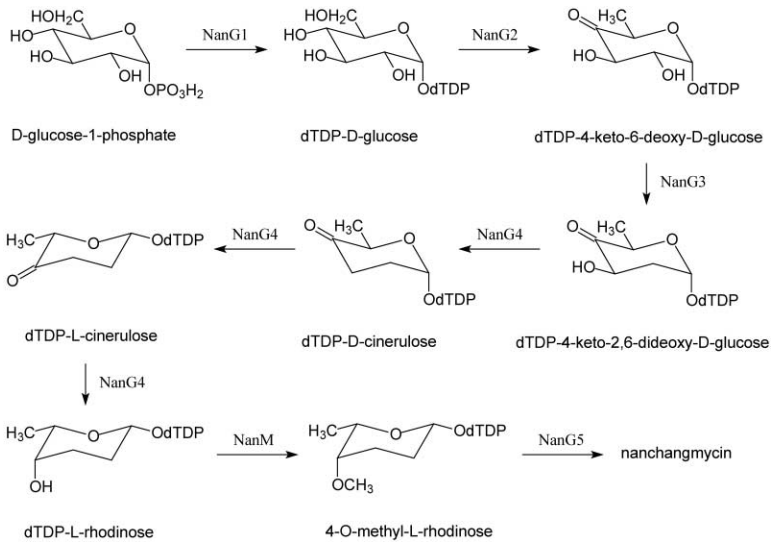

Figure 4. Proposed Deoxyhexose (4-*O*-methyl-L-rhodinose) Biosynthetic Pathway

Release of the Polyether Chain

Which enzyme is responsible for release of the aglycone polyether (or polyketide) chain? A search for a putative thioesterase-like gene, which could be involved in chain release of the mature aglycone, was not successful. However, a conserved domain showing homology with a family of proteins related to a large superfamily of metalloenzymes, including phosphotriesterases involved in nucleotide metabolism, was detected at the very end of module 14 (the last module) of the PKS (Figures 1 and 2). A member of this family, TatD [39, 40], a DNase with esterase activity, gives the highest sequence identity (28.8%), suggesting that this domain (tentative named CR for *c*hain *r*elease) may serve to release the polyketide chain, which has already formed a polyether structure. The predicted CR functionality differs from the thioesterases for release of macrolide polyketide chains and would obviously lack cyclization ability.

Limits of the Cluster

Sequencing of an additional 20 kb DNA to the right of *nanR4,* encoding a putative AraC family transcriptional activator, revealed another two genes (*nanG6* and *nanG7*) encoding glycosyltransferase homologs with up to 21% amino acid identity to that of *E. coli* [41] and 23% amino acid identity to that of *Bradyrhizobium japonicum* [42]. They are unlikely to be involved in attachment of the deoxysugar at C-19 because they are not clustered with the other genes necessary for deoxysugar biosynthesis, but with another gene (*nanG8*) that resembles a glucosyl hydrolase encoding a putative secreted endoglucanase, CenA, of *Cellulomonas fimi* (41% identity) [43]. We propose that these three genes lie outside the *nan* cluster and are possibly concerned with primary metabolism. Further away from the right side of *nanG8* lie several putative genes for nitrate reductase, NTP pyrophosphohydrolase, and alcohol dehydrogenase etc. (our unpublished data) for which no role in nanchangmycin biosynthesis could be assigned. The above observations suggest that *nanG6* is the right border of the gene cluster.

Located second from the leftmost gene, *nanT1*, of the cluster is *nanT2* (Figure 1), which encodes a polypeptide with 37% identity to a spermidine/putrescine ABC transporter of *Vibrio cholerae* [26] and thus might be involved in ATP-dependent efflux of nanchangmycin. DNA extending to the left of *nanT1* (Figure 1) identified a putative gene with 81.2% identity to a family of highly conserved CN (carbon-nitrogen)-hydrolases that break carbon-nitrogen bonds. This cannot be assigned to nanchangmycin biosynthesis because nanchangmycin does not contain nitrogen. The possibility that this gene is outside the left border of the *nan* cluster was proved by a gene disruption experiment: when the ORF carrying the putative CN-hydrolase was disrupted via an internal 572 bp SmaI fragment (see Experimental Procedures), the engineered mutants (SYH35-4 and SYH35-11) still produce nanchangmycin at a level similar to that of wild-type NS3226, as determined by bioassay as well as by LC-MS/MS (data not shown).

Model for Nanchangmycin Biosynthesis

Nanchangmycin biosynthesis (Figure 5) would start with loading of malonyl-CoA onto the NanA1 protein and, after decarboxylation, would proceed by condensation of nine methylmalonyl-CoA and four malonyl-CoA extender units by the NanA1-NanA10 PKS subunits. Oxidative cyclization seems to be initiated before the PKS chain is released, prior to the last condensation to complete the carbon chain [15]. This would involve a possible double-bond isomerization, resulting in an *E-Z* interconversion at an unknown position, by the putative NanI, two epoxidations of a diene intermediate by an epoxidase (NanO), and subsequent ring opening of a di-epoxide by a putative epoxide hydrolase (NanE). The process of oxidative cyclization is proposed to occur when the polyketide is still anchored on Nan10, an obvious type II ACP. After cleavage of the mature polyether chain from the PKS complex by an unusual CR domain of NanA11, it would form an uncyclized aglycone. The next step is probably accomplished by the putative NanP monooxygenase, which hydroxylates at C-30 (Figures 1 and 5). Biosynthesis of 4-*O*-methyl-L-rhodinose pre-

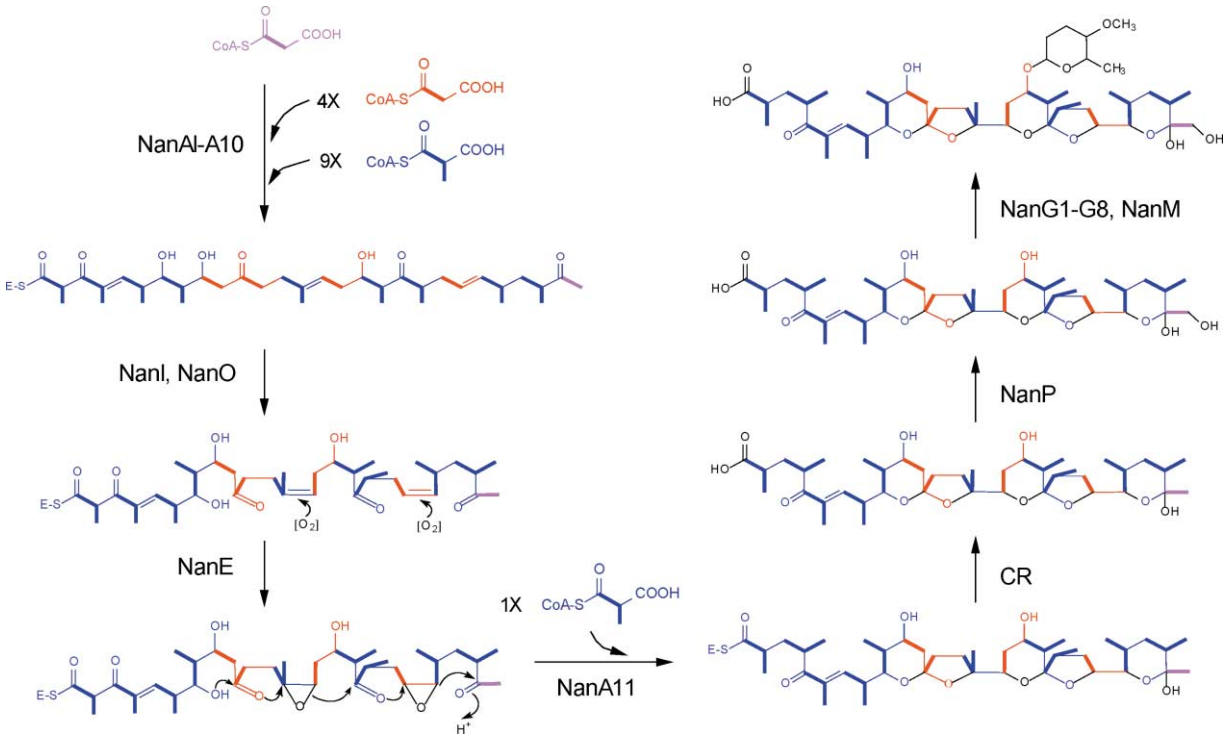

Figure 5. Proposed Biosynthetic Pathway and Enzymatic Functionalities Involved in Nanchangmycin Biosynthesis
Colors conform to Figure 1.

sumably starts with D-glucose-1-phosphate, which is converted to 4-*O*-methyl-L-rhodinose as summarized in Figure 4, and then attached to the aglycone by the NanG5 glycosyl transferase.

Obviously, the model presented above is mostly based on sequence information. The functions of many of the genes, especially those of regulation, polyether formation, chain release, and post-polyether modification, will have to be verified by ongoing experiments on inactivation or replacement of the genes concerned, and structural analysis of nanchangmycin derivatives produced by the corresponding engineered mutants, as exemplified below.

Generation of a Novel Nanchangmycin Aglycone by In-Frame Deletion of the KR6 Domain

The KR6 domain of module 6 in NanA4 is assumed to be responsible for the appearance of a hydroxy group at the C-19 position of nanchangmycin (Figure 6, top). Thus, deletion of KR6, to leave a keto group at C-19, would abolish attachment of 4-*O*-methyl-L-rhodinose to the aglycone (Figure 6, bottom). To prove this hypothesis, an engineered mutant strain (SYH28 in Figure 6) with an in-frame deletion of KR6 from the wild-type NS3226 was created (as described in Experimental Procedures and outlined in Figure 6). When five independent SYH28 isolates were tested for antibacterial activity against *Bacillus cereus* 1126, no activity (Figure 6, bottom plate without inhibition zone) corresponding to nanchangmycin (Figure 6, top plate with inhibition zone) could be detected. The structural change in the compound was as expected: when a methanol extract of the SYH28 culture was subjected to LC-MS/MS analysis, no compound corresponding to nanchangmycin (theoretical mass 866.12 versus MS1 *m/z* 866.0 in Figure 6) was seen, but a compound corresponding to nanchangmycin aglycone (theoretical mass 735.93 versus MS1 *m/z* 735.2 in Figure 6) was revealed. The chemical structures of nanchangmycin and its aglycone were further confirmed by their respective fragmentation patterns after loss of one (MS2) and two (MS3) H_2O molecules. The correlation of a mass difference with the molecular weight of the lost sugar moiety (4-*O*-methyl-L-rhodinose) unambiguously demonstrated the precise functionality of the KR6 domain in the specified catalytic step.

Significance

Polyether antibiotics are a group of natural compounds that have received little combined genetic and biochemical attention. Our analysis of the nanchangmycin biosynthetic pathway genes from *S. nanchangensis* NS3226 enabled us to propose a model for nanchangmycin biosynthesis, which seems to be the first example of complete DNA sequence analysis of a polyether antibiotic biosynthetic gene cluster. Identification of 67 constituent active sites (excluding 4 that are presumably inactive and 3 that are silent), several putative regulatory genes, functionalities unique for the oxidative cyclization to form the polyether structure during polyketide elongation, and a putative novel enzyme domain for polyether chain release will expand our understanding of the biosynthesis of this important group of anticoccidial agents and increase

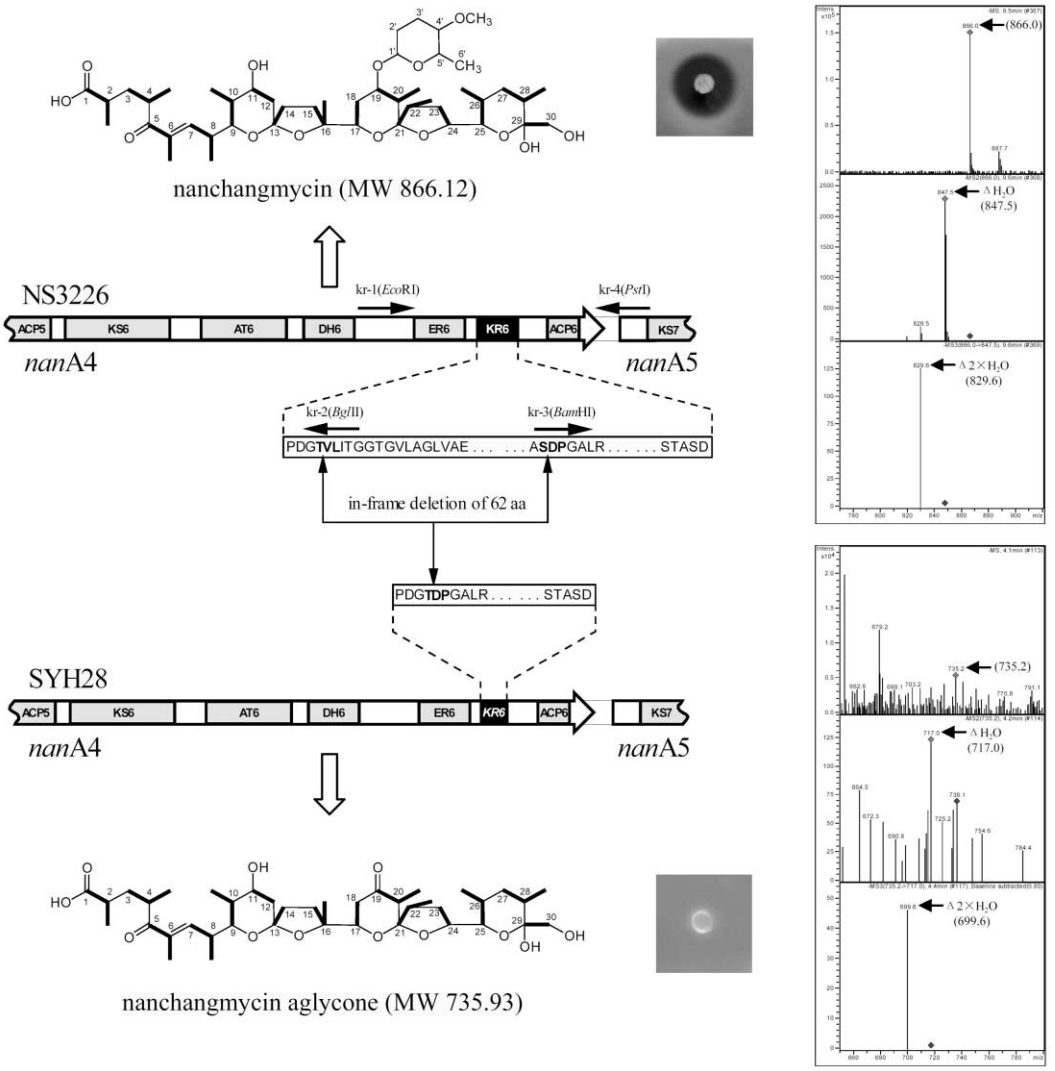

Figure 6. In-Frame Deletion of KR6 Resulted in Production of a Nanchangmycin Derivative without a Sugar Moiety

The upper part shows the region flanking KR6 (shown in black) in the chromosome of the wild-type strain NS3226, the inhibition against *Bacillus cereus* 1126, and the LC-MS/MS fragmentation (loss of H_2O, ΔH_2O) patterns (MS1-MS3 in top, middle, and bottom frames, respectively) of its produced antibiotic nanchangmycin. The lower part shows, for the engineered mutant SYH28, lack of inhibition against *Bacillus cereus* 1126 and the LC-MS/MS fragmentation (loss of H_2O) patterns of its produced nanchangmycin aglycone. The enlarged KR6 section in the middle shows changes before and after in-frame deletion of the 62 aa. The horizontal solid arrows indicate the region corresponding to the synthetic PCR primers (kr-1, 2 for the left and kr-3, 4 for the right arms flanking the deleted region) used for constructing the pHZ1358-derived vector [11] for the in-frame deletion. The restriction sites in parentheses represent engineered specific sequences at the end of PCR primers so as to favor cloning of the PCR products. The amino acid residues in bold highlight the in-frame joining after the expected deletion to form an inactivated KR6 (italics).

our ability to interconvert or modify a wider range of polyketide structures to obtain a greater diversity of macrolides, polyenes, and polyethers, and to enhance the yield of nanchangmycin (and its derivatives) by genetic rather than chemical means.

Experimental Procedures

Bacterial Strains, Plasmids, Culture Techniques, and Media

S. nanchangensis NS3226 [44], the wild-type producer for nanchangmycin and meilingmycin, was used for nanchangmycin isolation, bioassay, and generation of mutant strains by targeted gene disruption and replacement. DH5α (F$^-$, *rec*A, *lac*Z, ΔM15) [45] was used as *E. coli* host. pHZ1358 is a cosmid for constructing the NS3226 genomic library [11], and pUC18 [46] was used as vector for DNA sequencing.

Culture techniques and media for vegetative growth, sporulation, and antibiotic production of *S. nanchangensis* were as described [11]. Luria-Bertani medium was used for *E. coli* propagation.

DNA Sequencing and Analysis

DNA sequencing was performed using a set of four cosmids, 3B4, 19B4, 5H3, and 3C5, which overlapped each other by at least 5 kb, from 17 overlapping cosmids in contig A [11] covering the whole nanchangmycin biosynthetic gene cluster. The insert of each cosmid DNA was cut out with DraI and purified using a Plasmid Maxi kit (Qiagen), and sonicated with a 550 Sonic Dismembrator (Fisher Scientific). DNA fragments of 1.6–2.0 kb were recovered from 0.7%

low melting agarose gel using a Geneclean II reagent kit (Bio 101, Inc), and subcloned into pUC18. For automated sequencing, plasmid DNA templates were prepared by alkaline lysis using the Prep 96 Plasmid Kit (Qiagen). Sequencing reactions were carried out with BigDye Terminator Cycle Sequencing kits (Applied Biosystem Division, Perkin Elmer). The sequences of custom-designed sequencing primers were 5′-GTAAAACGACGGCCAGT-3′ (forward) and 5′-GCGGATAACAATTTCACACAGG-3′ (reverse). Sequence reads were obtained from 377 DNA Sequencers (PE/ABD). Some gaps were filled using PCR amplifications as detailed in [47].

Sequence contig assembly and base editing utilized the Phred/Phrap/Consed package [48, 49]. The sequence data were analyzed with the Frame-Plot 2.3 online program [50]. DNA and deduced protein sequence homology searches were performed using BLAST [51–53] and FASTA [54].

Creation of an In-Frame Deletion of the KR6 Domain by Targeted Gene Replacement and Determination of the Left Boundary of the *nan* Gene Cluster by Gene Disruption

For creation of in-frame deletion of KR6 by targeted gene replacement, two pairs of oligonucleotide primers, kr-1 (5′-GGAGGGG GAATTCGGTGAGC-3′, EcoRI site underlined) and kr-2 (5′-GTGCCACCGGTGATCAGATCTGTGCCGTCG-3′, BglII site underlined), and kr-3 (5′-GCCGCGGACGCCTCGGATCCGGGGCCCTG-3′, BamHI site underlined) and kr-4 (5′-CAGGCCCTGCAGATCCCAGC-3′, PstI site underlined), were used to amplify the DNA fragment flanking the left and right sides of the KR6 domain, respectively. Two amplified products were digested with EcoRI and BglII for the former to regenerate a 1419 bp fragment, and BamHI and PstI for the latter to regenerate a 1422 bp fragment and ligated with pUC18 digested with EcoRI and PstI. The recombinant fragment (1419 + 1422 bp) was recovered from a resultant plasmid pJTU302 after EcoRI and PstI digestions and cloned into the corresponding sites of pIJ2925 [46] so that it could be regenerated as a single BglII fragment to be finally cloned into the unique BamHI site of pHZ1358 [11], giving rise to pJTU304 as a vector used for gene replacement. The insert in pJTU304 was proved to have only a 186 bp expected deletion encoding 62 aa residues by PCR amplification using primers kr-cp1 (5′-GCGATCCCGAACAGCTGGCC-3′) and kr-cp2 (5′-CTGTCCGGGGTTGCCGAGGG-3′) flanking both sides of the deleted region and by DNA sequencing of the insert DNA in pJTU304.

pJTU304 was transferred by conjugation from *E. coli* ET12567 carrying a RP4 derivative pUZ8002 [55] into strain NS3226. Thiostrepton-sensitive ($Thio^S$) colonies were counterselected from the initial $Thio^R$ exconjugants after one round of nonselective growth. 5 $Thio^S$ exconjugants were proved to have the expected deletions among 28 $Thio^S$ exconjugants, distinguishable from NS3226 revertants by PCR amplification using kr-cp1 and kr-cp2 primers and by DNA sequencing of the PCR products.

For determination of the left boundary of the *nan* gene cluster by gene disruption, a 572 bp SmaI fragment internal to the ORF to the left of the putative boundary of the *nan* gene cluster (*nanT1*), encoding a hypothetical CN-hydrolase, was cloned into the unique HpaI site of pHZ1358 [11]. When the resultant plasmid, pJTU335, was transferred by conjugation from *E. coli* ET12567 carrying a RP4 derivative pUZ8002 [55] into strain NS3226, 2 of the 11 $Thio^R$ exconjugants after three rounds of nonselective growth (SYH35-4 and SYH35-11) were found to have pJTU335 stably integrated into the NS3226 chromosome by homologous recombination via the 572 bp SmaI fragment, as determined by Southern hybridization.

Antibiotic Bioassay and Structural Characterization of Nanchangmycin and Its Derivative Compounds

Antibiotic bioassay of *S. nanchangensis* NS3226 and its derivative SYH28, SYH35-4, were as described [11]. Samples for structural characterization were prepared by extraction of a 7-day-old fermentation culture with 2 vol of methanol for 8 hr, followed by centrifugation at 12,000 rpm for 10 min. The liquid phase was applied to LC-MS/MS (Agilent 1100 series LC/MSD Trap system) after passing through a 0.45 μm filter. The LC was operated as a gradient flow at a rate of 0.8 ml/min using a Waters XTerra RP18 (3.9 × 150 mm, 5 μm) column at 25°C. Eluent A was MilliQ deionized water, and eluent B was pure acetonitrile (Merck). The initial conditions were 50% down to 5% A and 50% up to 95% B during 0–15 min, then 5% A and 95% B during 15–20 min. The iontrap mass spectrometer was operated with the electrospray ionization source in the negative ion mode. Drying gas flow was 10 l/ml, and nubulizer pressure was 40 psi. Drying gas temperature was 350°C.

Acknowledgments

We thank Prof. Sir. D.A. Hopwood, FRS for critical reading of the manuscript and many valuable comments, and Prof. Ben Shen for many helpful discussions. This work received support from the National Science Foundation of China, the Ministry of Science and Technology, and the Shanghai Municipal Council of Science and Technology.

Received: January 26, 2003
Revised: March 17, 2003
Accepted: April 14, 2003
Published: May 16, 2003

References

1. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675–679.
2. Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. Chem. Rev. *97*, 2465–2497.
3. Hutchinson, C.R., Kennedy, J., Park, C., Kendrew, S., Auclair, K., and Vederas, J. (2000). Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. Antonie Van Leeuwenhoek *78*, 287–295.
4. Rawlings, B.J. (2001). Type I polyketide biosynthesis in bacteria (part B). Nat. Prod. Rep. *18*, 231–281.
5. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., and Omura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. USA *96*, 9509–9514.
6. Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R., Valla, S., and Zotchev, S.B. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. Chem. Biol. *7*, 395–403.
7. Steinrauf, L.K., Czerwinski, E.W., and Pinkerton, M. (1971). Comparison of the monovalent cation complexes of monensin, nigericin, and dianemycin. Biochem. Biophys. Res. Commun. *45*, 1279–1283.
8. Czerwinski, E.W., and Steinrauf, L.K. (1971). Structure of the antibiotic dianemycin. Biochem. Biophys. Res. Commun. *45*, 1284–1287.
9. Allen, P.R., Brimble, M.A., and Prabaharan, H. (2001). Synthesis of the bis-spiroacetal moiety of the polyether antibiotic CP44, 161. J. Chem. Soc. Perkin Trans. 1 *4*, 379–389.
10. Brimble, M.A., and Fares, F.A. (1999). Synthesis of bis-spiroacetal ring systems. Tetrahedron *55*, 7661–7706.
11. Sun, Y., Zhou, X., Liu, J., Bao, K., Zhang, G., Tu, G., Kieser, T., and Deng, Z. (2002). *Streptomyces nanchangensis*, a producer of the insecticidal polyether antibiotic nanchangmycin and the antiparasitic macrolide meilingmycin, contains multiple polyketide gene clusters. Microbiology *148*, 361–371.
12. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., et al. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. USA *98*, 12215–12220.
13. Leadlay, P.F., Staunton, J., Oliynyk, M., Bisang, C., Cortes, J., Frost, E., Hughes-Thomas, Z.A., Jones, M.A., Kendrew, S.G., Lester, J.B., et al. (2001). Engineering of complex polyketide biosynthesis: insights from sequencing of the monensin biosynthetic gene cluster. J. Ind. Microbiol. Biotechnol. *27*, 360–367.
14. Fouces, R., Mellado, E., Diez, B., and Barredo, J.L. (1999). The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organization of the left region. Microbiology *145*, 855–868.

15. Kakavas, S.J., Katz, L., and Stassi, D. (1997). Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. J. Bacteriol. *179*, 7515–7522.
16. Aparicio, J.F., Colina, A.J., Ceballos, E., and Martin, J.F. (1999). The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes. J. Biol. Chem. *274*, 10133–10139.
17. 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.
18. Bate, N., Butler, A.R., Gandecha, A.R., and Cundliffe, E. (1999). Multiple regulatory genes in the tylosin biosynthetic cluster of *Streptomyces fradiae*. Chem. Biol. *6*, 617–624.
19. Wietzorrek, A., and Bibb, M. (1997). A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol. Microbiol. *25*, 1181–1184.
20. O'Connell-Motherway, M., Fitzgerald, G.F., and van Sinderen, D. (1997). Cloning and sequence analysis of putative histidine protein kinases isolated from *Lactococcus lactis* MG1363. Appl. Environ. Microbiol. *63*, 2454–2459.
21. Huang, J., Carney, B.F., Denny, T.P., Weissinger, A.K., and Schell, M.A. (1995). A complex network regulates expression of eps and other virulence genes of Pseudomonas solanacearum. J. Bacteriol. *177*, 1259–1267.
22. Falke, J.J., and Hazelbauer, G.L. (2001). Transmembrane signaling in bacterial chemoreceptors. Trends Biochem. Sci. *26*, 257–265.
23. Falke, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A., and Danielson, M.A. (1997). The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. Annu. Rev. Cell Dev. Biol. *13*, 457–512.
24. Chervitz, S.A., and Falke, J.J. (1996). Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. Proc. Natl. Acad. Sci. USA *93*, 2545–2550.
25. Lombo, F., Brana, A.F., Mendez, C., and 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.
26. Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature *406*, 477–483.
27. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496–512.
28. Henkin, T.M., Grundy, F.J., Nicholson, W.L., and Chambliss, G.H. (1991). Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* lacI and galR repressors. Mol. Microbiol. *5*, 575–584.
29. Chen, C.W., Yu, T.W., Chung, H.M., and Chou, C.F. (1992). Discovery and characterization of a new transposable element, Tn4811, in *Streptomyces lividans* 66. J. Bacteriol. *174*, 7762–7769.
30. Xue, Y., Zhao, L., Liu, H.W., and Sherman, D.H. (1998). A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. Proc. Natl. Acad. Sci. USA *95*, 12111–12116.
31. Radominska-Pandya, A., Czernik, P.J., Little, J.M., Battaglia, E., and Mackenzie, P.I. (1999). Structural and functional studies of UDP-glucuronosyltransferases. Drug Metab. Rev. *31*, 817–899.
32. Ichinose, K., Bedford, D.J., Tornus, D., Bechthold, A., Bibb, M.J., Revill, W.P., Floss, H.G., and Hopwood, D.A. (1998). The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tu22: sequence analysis and expression in a heterologous host. Chem. Biol. *5*, 647–659.
33. Faust, B., Hoffmeister, D., Weitnauer, G., Westrich, L., Haag, S., Schneider, P., Decker, H., Kunzel, E., Rohr, J., and Bechthold, A. (2000). Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in *Streptomyces fradiae* Tu2717. Microbiology *146*, 147–154.
34. Kosuga, K., Hata, S., Osumi, T., Sakakibara, J., and Ono, T. (1995). Nucleotide sequence of a cDNA for mouse squalene epoxidase. Biochim. Biophys. Acta *1260*, 345–348.
35. Grant, D.F., Storms, D.H., and Hammock, B.D. (1993). Molecular cloning and expression of murine liver soluble epoxide hydrolase. J. Biol. Chem. *268*, 17628–17633.
36. Kim, S.W., Joo, S., Choi, G., Cho, H.S., Oh, B.H., and Choi, K.Y. (1997). Mutational analysis of the three cysteines and active-site aspartic acid 103 of ketosteroid isomerase from *Pseudomonas putida* biotype B. J. Bacteriol. *179*, 7742–7747.
37. Massiah, M.A., Abeygunawardana, C., Gittis, A.G., and Mildvan, A.S. (1998). Solution structure of delta 5-3-ketosteroid isomerase complexed with the steroid 19-nortestosterone hemisuccinate. Biochemistry *37*, 14701–14712.
38. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., III, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537–544.
39. Wexler, M., Sargent, F., Jack, R.L., Stanley, N.R., Bogsch, E.G., Robinson, C., Berks, B.C., and Palmer, T. (2000). TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. J. Biol. Chem. *275*, 16717–16722.
40. Holm, L., and Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins *28*, 72–82.
41. Bastin, D.A., and Reeves, P.R. (1995). Sequence and analysis of the O antigen gene (rfb) cluster of *Escherichia coli* O111. Gene *164*, 17–23.
42. Bhagwat, A.A., and Keister, D.L. (1995). Site-directed mutagenesis of the beta-(1->3),beta-(1->6)-D-glucan synthesis locus of *Bradyrhizobium japonicum*. Mol. Plant Microbe Interact. *8*, 366–370.
43. Wong, W.K., Gerhard, B., Guo, Z.M., Kilburn, D.G., Warren, A.J., and Miller, R.C., Jr. (1986). Characterization and structure of an endoglucanase gene cenA of *Cellulomonas fimi*. Gene *44*, 315–324.
44. Ouyang, L., Wan, S., Tu, G., Chen, X., Zhen, Y., Gao, Y., and Xie, X. (1984). A new species of *Streptomyces* producing insecticidal antibiotics. Chin. J. Microbiol. *24*, 195–199.
45. Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. *166*, 557–580.
46. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical *Streptomyces* Genetics (Norwich, UK: John Innes Foundation).
47. Zhang, J., Voss, K.O., Shaw, D.F., Roos, K.P., Lewis, D.F., Yan, J., Jiang, R., Ren, H., Hou, J.Y., Fang, Y., et al. (1999). A multiple-capillary electrophoresis system for small-scale DNA sequencing and analysis. Nucleic Acids Res. *27*, e36.
48. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. *8*, 175–185.
49. Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. Genome Res. *8*, 195–202.
50. Ishikawa, J., and Hotta, K. (1999). FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. FEMS Microbiol. Lett. *174*, 251–253.
51. Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., and Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. *29*, 2994–3005.
52. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

53. Altschul, S.F., and Koonin, E.V. (1998). Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases. Trends Biochem. Sci. *23*, 444–447.
54. Pearson, W.R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. *183*, 63–98.
55. MacNeil, D.J., Occi, J.L., Gewain, K.M., MacNeil, T., Gibbons, P.H., Ruby, C.L., and Danis, S.J. (1992). Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. Gene *115*, 119–125.

Accession Numbers

The complete DNA and deduced protein sequences of the nanchangmycin biosynthetic gene cluster reported in this paper have been deposited in GenBank under the accession number AF521085.