Chemistry & Biology, Vol. 9, 519–531, April, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00126-6

The Biosynthetic Gene Cluster for the Antitumor Rebeccamycin: Characterization and Generation of Indolocarbazole Derivatives

César Sánchez,[1] Igor A. Butovich,[2] Alfredo F. Braña,[1] Jürgen Rohr,[2,3] Carmen Méndez,[1] and José A. Salas[1,3]
[1]Departamento de Biología Funcional e Instituto Universitario de Oncología del Principado de Asturias
Universidad de Oviedo
33006 Oviedo
Spain
[2]Department of Pharmaceutical Sciences
College of Pharmacy
Medical University of South Carolina
280 Calhoun Street
P.O. Box 250140
Charleston, South Carolina 29425

Summary

Rebeccamycin, a halogenated natural product of the indolocarbazole family, is produced by *Saccharothrix aerocolonigenes* ATCC39243. Several rebeccamycin analogues, which target DNA topoisomerase I or II, have already entered clinical trials as anticancer drugs. Using as a probe an internal fragment of *ngt*, a *Saccharothrix aerocolonigenes* gene encoding an indolocarbazole *N*-glycosyltransferase, we isolated a DNA region that directed the biosynthesis of rebeccamycin when introduced into *Streptomyces albus*. Sequence analysis of 25.6 kb revealed genes for indolocarbazole core formation, halogenation, glycosylation, and sugar methylation, as well as a regulatory gene and two resistance/secretion genes. Heterologous expression of subsets of these genes resulted in production of deschloro-rebeccamycin, 4′-demethyl-deschloro-rebeccamycin, and deschloro-rebeccamycin aglycone. The cloned genes should help to elucidate the molecular basis for indolocarbazole biosynthesis and set the stage for the generation of novel indolocarbazole analogues by genetic engineering.

Introduction

Since its discovery in 1977, the indolocarbazole family of natural products has received great attention from both organic chemists and biomedical researchers. The reason for this interest resides in the novel structures and wide range of biological activities (antibacterial, antifungal, and antitumor) that members of this family display [1]. Indolocarbazole natural products are defined by their characteristic structure containing either indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole (e.g., rebeccamycin), indolo[2,3-*a*]carbazole (e.g., tjipanazoles), or *bis*-indolylmaleimide (e.g., arcyriarubin) moieties. In particular, the indolopyrrolocarbazoles constitute a new class of antitumor drugs, which can be divided into two major groups depending on their mechanisms of action and structural features. One group consists of protein kinase inhibitors, such as staurosporine and K-252a. Usually, they contain a sugar moiety linked to both indole nitrogens of the indolocarbazole core. The second group consists of DNA-damaging agents, exemplified by rebeccamycin and AT2433, that act on DNA topoisomerase I or II. Most members of this group contain a sugar moiety attached by a β-glycosidic linkage to only one of the indole nitrogens of the aglycone. It has been shown that the carbohydrate plays a crucial role in the activity of these compounds [2]. Presently, three protein kinase inhibitors (UCN-01, CGP 41251, CEP-751) and two DNA-damaging agents (NB-506, NSC655649) have already entered clinical trials for their use against several types of tumors [3].

Rebeccamycin (Figure 1) is an antitumor antibiotic produced by the actinomycete *Saccharothrix aerocolonigenes* ATCC39243 [4, 5]. It shows antibacterial activity against several Gram-positive bacteria, including *Staphylococcus aureus* and *Streptococcus faecalis*. Rebeccamycin also inhibits the growth of some tumor cell lines and displays activity against several types of tumors implanted in mice [5]. There is a considerable effort placed on the design and synthesis of new rebeccamycin analogs and derivatives with enhanced antitumor activity [6]. From a different point of view, rebeccamycin has also attracted interest because it is a halogenated natural product [7].

Despite the interest focused on indolocarbazoles, very little is known about the biochemical and genetic aspects of their biosyntheses. As far as we know, no complete set of genes for the biosynthesis of an indolocarbazole natural product has been reported. A patent application has claimed the cloning of some genes needed for the last steps in the biosynthesis of the staurosporine sugar moiety in *Streptomyces longisporoflavus* DSM10189 [8]. More recently, the cloning and sequencing of a gene, *ngt*, encoding an indolocarbazole *N*-glycosyltransferase from *Saccharothrix aerocolonigenes* ATCC39243 has been reported [9].

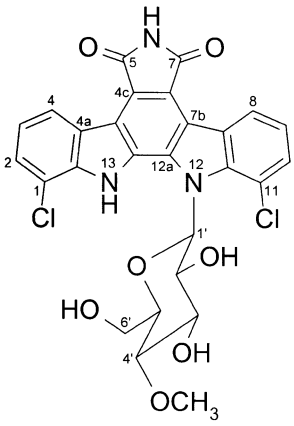
Here we report the cloning and sequencing of the complete gene cluster encoding rebeccamycin biosynthesis from *Saccharothrix aerocolonigenes* ATCC39243. We also show that the identified set of genes is necessary and sufficient for rebeccamycin production by heterologous expression in *Streptomyces albus*. Finally, we report the production of rebeccamycin derivatives in *S. albus* when transformed with selected subsets of the identified genes.

Results and Discussion

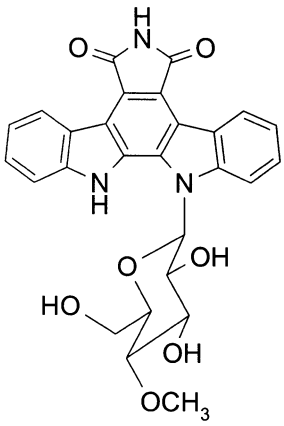
Isolation of the DNA Region Surrounding the *ngt* Gene

Recently, the *Saccharothrix aerocolonigenes* ATCC39243 *ngt* gene, encoding an *N*-glycosyltransferase able to introduce a D-glucose moiety into indolocarbazoles

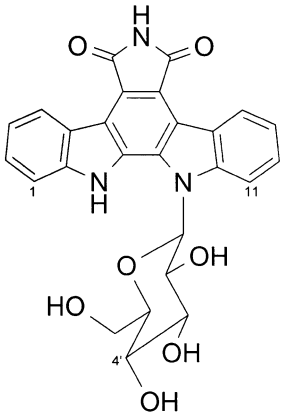
[3]Correspondence: jasf@sauron.quimica.uniovi.es (J.A.S., for molecular-biology communications) rohrj@musc.edu (J.R., for chemical communications)



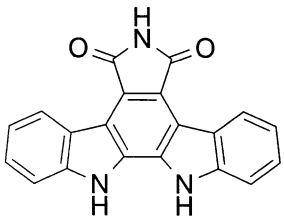
1: Rebeccamycin



2: Deschloro-rebeccamycin



3: 4'-Demethyl-deschloro-rebeccamycin



4: Deschloro-rebeccamycin aglycon

Figure 1. Structure of Rebeccamycin and Derivatives

J-104303 and 6-*N*-methylarcyriaflavin C, has been cloned and sequenced [9]. No other open reading frame (ORF) was reported in the sequenced region, encompassing 1898 nucleotides (nt). Given the known precedent that antibiotic biosynthetic genes commonly occur as a cluster in antibiotic-producing actinomycetes, we set out to isolate the DNA flanking *ngt* in order to identify the rebeccamycin biosynthetic gene cluster. A *Saccharothrix aerocolonigenes* ATCC39243 genomic library was constructed in the *Escherichia coli-Streptomyces* shuttle vector pKC505 [10]. This library was screened with a probe from an internal fragment of *ngt*, resulting in the isolation of several overlapping cosmids (10A4, 14E8, 17A12, and 24B2). It has been previously shown by PCR amplification that the genome of *Saccharothrix aerocolonigenes* ATCC39243 contains at least one $FADH_2$-dependent halogenase gene [11]. In order to investigate if this putative halogenase gene was part of the rebeccamycin biosynthetic gene cluster, we analyzed the isolated cosmids by PCR with degenerate oligoprimers designed to amplify halogenase genes. A PCR product of the expected size was amplified only from cosmids 14E8 and 17A12. The sequencing of these PCR products confirmed that they were identical to the sequence previously reported [11].

Production of Rebeccamycin in a Heterologous Host

In order to determine if any of the isolated cosmids harboured the complete rebeccamycin gene cluster, we independently introduced the cosmids into the heterologous host *Streptomyces albus* J1074 and analyzed transformants for production of rebeccamycin or any biosynthetic precursor. We routinely use *S. albus* J1074 as a host for heterologous gene expression and metabolite production because this strain shows a nice, dispersed growth, a very low background concerning aromatic metabolite production (as detected by HPLC), and a good efficiency of protoplast transformation. Two of the cosmids (14E8 and 17A12) conferred the ability to produce a compound not produced by the host strain *S. albus* J1074. The identity of this compound with rebeccamycin was demonstrated as follows. First, it showed antibacterial activity in growth inhibition tests performed on *Micrococcus luteus*. Second, it had the same relative mobility as authentic rebeccamycin in

Table 1. ^{1}H NMR data of rebeccamycin (**1**), deschlororebeccamycin (**2**), 4′-demethyl-deschlororebeccamycin (**3**), and deschlororebeccamycin aglycon (**4**) in d_6-DMSO at 400 MHz

	Compound			
	1	**2**	**3**	**4**
Position	δ, Multiplicity (*J*/HZ)	δ, Multiplicity (*J*/HZ)	δ, Multiplicity (*J*/HZ)	δ, Multiplicity (*J*/HZ)
1-H	–	9.05 d (8)	9.04 d (8)	8.99 d (8)
2-H	7.66 d (8)	7.33 t (8)	7.33 t (8)	7.34 t (8)
3-H	7.44 t (8)	7.54 t (8)	7.50-7.57 t (8)[a]	7.54 t (8)
4-H	9.06 d (8)	7.68 d (8)	7.65 d (8)	7.80 d (8)
N6-H	11.34 s	11.37 s	11.06 s	10.96 s
8-H	9.24 d (8)	7.92 d (8)	7.93 d (9)	7.80 d (8)
9-H	7.44 t (8)	7.54 t (8)	7.50-7.57 dd (9,8)[a]	7.54 t (8)
10-H	7.72 d (8)	7.33 t (8)	7.33 t (8)	7.34 t (8)
11-H	–	9.13 d (8)	9.12 d (8)	8.99 d (8)
N12-H	–	–	–	11.77 s
N13-H	10.66 s	11.57 s	11.63 s	11.77 s
1′-H	6.91 d (9)	6.26 d (9)	6.24 d (9)	–
2′-H	3.69 ddd (9,9,6)	3.52 ddd (9,9,5)	3.50 ddd (9,9,5)[b]	–
2′-OH	5.0 d (6)	4.93 d (5)	4.87 d (5)	–
3′-H	3.56 ddd (9.5,9,6)	3.69 ddd (9.5,9,6)	3.54 ddd (9,9,5)[b]	–
3′-OH	5.39 d (6)	5.24 d (6)	5.09 d (5)	–
4′-H	3.64 dd (9.5,9.5)	3.74 dd (9.5,9.5)	3.96 ddd (9.5,9.5,5)	–
4′-OH	–	–	5.34 d (5)	–
4′-OCH_3	3.59 s	3.62 s	–	–
5′-H	3.82 dt (9.5,5)	3.85 ddd (9.5,5,4)	3.78 ddd (9,4,2)	–
6′-H_2	3.94 dd (6,5)	3.93 dd (11,4) 3.98 dd (11,5)	3.93 dd (11,2)[c] 4.04 dd (11,4)	–
6′-OH	5.27 t (6)	6.10 dd (5,4)	5.96 t (4)	–

δ is given in ppm relative to internal TMS.
[a] Complex, overlapping signals; *J* not exactly determinable.
[b] Overlapping signal assignments are interchangeable.
[c] Broad signal.

HPLC analysis under different chromatographic conditions. Third, it showed the characteristic rebeccamycin absorption spectrum with maxima at 238 and 316 nm. Fourth, the compound was purified and studied by high-resolution EI-MS, yielding a main peak with a mass of 569.0753 corresponding to rebeccamycin with theoretical calculated molecular mass of 569.0757 (for $C_{27}H_{21}N_3O_7Cl_2$), and a minor peak with a mass of 393 corresponding to rebeccamycin aglycone. Similar MS results were previously reported for purified rebeccamycin [4]. Finally, the structure of the product was confirmed to be that of rebeccamycin by NMR analysis as described below.

Rebeccamycin (**1**, Figure 1) was analyzed by 1D ^{1}H and 2D ^{1}H, ^{13}C NMR spectroscopy. The ^{13}C, ^{1}H HSQC (heteronuclear single quantum coherence) experiment and a 2D ^{13}C, ^{1}H HMBC experiment allowed the detection and assignment of almost all carbons of the molecule (Table 1). For instance, a distinctive proton signal with δ 11.36 couples to the carbons with δ 170.1 (C5 and C7), 123.3 (C7a), and 121.3 (C4c) and apparently belonged to the N bound 6-H of rebeccamycin. The cross peaks at 10.69/137.8, 10.69/130.4, and 10.69/118.3 tied the proton at N-13 to the corresponding carbons 13a, 12b, and 4b of the pyrrole ring. An aromatic proton, part of an ABC system, with δ 9.09 (4-H) was strongly coupled to the carbons with δ values of 137.8 (C13a) and 127.8 (C4a), and the ^{1}H signal with δ 9.27 (8-H) gave intense cross peaks with carbons possessing δ 138.2 (C11a) and 130.7 (C7c).

The signals, typical for a sugar moiety (Tables 1 and 2, C1′ to C6′′, 1′-H to 6′-H_2) and in line with the observations that the nitrogen atom at the 12 position lacks a proton and that carbons 11a and 12a, adjacent to N-12, are coupled to a proton of C1′ of this glycoside residue, confirm the presence of the sugar moiety and its *N*-glycosidic linkage to N-12. The NMR data are close to those reported earlier [4]. Thus, the combination of HSQC, HMBC, and DQCOSY NMR spectra along with the MS data clearly showed the compound at hand to be rebeccamycin (**1**; Figure 1).

Rebeccamycin production levels in these *S. albus* recombinant strains were several fold greater than those obtained with *Saccharothrix aerocolonigenes* ATCC39243 grown in the same conditions. Cosmid 14E8, one of the cosmids able to direct rebeccamycin biosynthesis, was chosen for further study.

It is known that rebeccamycin has antibacterial activity against some Gram-positive bacteria and that it produces a weak inhibition of growth against some *Streptomyces* spp. [12]. Based on the frequent linkage between biosynthetic and resistance genes, we decided to determine whether the cloned DNA contained rebeccamycin resistance determinants. We checked the effect of exogenously added rebeccamycin on the growth of *S. albus* J1074/14E8 and on that of the control strain *S. albus* J1074/pKC505. Paper disc diffusion assays showed that growth of *S. albus* J1074/pKC505 was totally inhibited by 100 μg rebeccamycin, whereas *S. albus* J1074/14E8 was not affected by amounts of as much as 200 μg.

Table 2. ^{13}C NMR data of rebeccamycin (**1**)[a], deschloro-rebeccamycin (**2**), 4′-demethyl-deschloro-rebeccamycin (**3**), and deschloro-rebeccamycin aglycon (**4**) in DMSO-d_6 at 100.6 MHz

	Compound			
	1[a]	**2**	**3**	**4**
Position	δ	δ	δ	δ
1	117.0	124.4	125.1	124.1
2	130.4	120.4[b]	121.0[b]	120.1
3	123.1	126.9[b]	127.6[b]	126.6
4	124.1	112.2	112.8[b]	111.9
4a	127.8	121.4[b]	122.1[b]	121.4
4b	118.3	118.6[b]	119.1[b]	115.4
4c	121.3	121.1[b]	121.3[b]	119.7
5	170.1	171.0	171.7	171.1
7	170.1	171.0	171.7	171.1
7a	123.3	119.4[b]	120.1[b]	119.7
7b	119.2*	116.9[b]	117.6[b]	115.4
7c	130.7	121.1[b]	121.7[b]	121.4
8	124.7	111.7	112.5[b]	111.9
9	123.1	126.8[b]	127.4[b]	126.6
10	128.1	120.6[b]	121.7[b]	120.1
11	117.0	124.4	125.1	124.1
11a	138.2	142.1	142.8	140.1
12a	129.5[c]	128.2	129.0	128.9
12b	130.4	129.6	130.4	128.9
13a	137.8	140.7	141.4	140.1
1′	85.2	84.1	85.1	–
2′	72.9	73.1	73.2	–
3′	78.0	76.3	77.3	–
4′	79.7	77.3	68.2	–
4′-OCH_3	60.7	60.0	–	–
5′	80.8	77.1	79.2	–
6′	60.4	58.5	58.9	–

δ is in ppm relative to internal TMS. Assignments were made with the help of couplings observed in the HSQC and HMBC spectra.
[a] Data from HSQC and HMBC spectra.
[b] Assignments are interchangeable within each group.
[c] Not observed by the indirect detection methods, taken from the literature ([4], 90 MHz) for comparison reasons.

Analysis of the Rebeccamycin Biosynthetic Gene Cluster

The complete DNA sequence of the insert contained in cosmid 14E8 was determined to consist of 25,681 nt (the sequence was deposited at the EMBL Nucleotide Sequence Database with accession number AJ414559). Sequence analysis revealed the presence of 16 complete ORFs and two incomplete ORFs, whose organization is shown in Figure 2A. Database comparisons allowed us to propose functions for most of these ORFs (Table 3). Genes apparently involved in rebeccamycin biosynthesis seem to be organized in four transcriptional units: *rebG*, *rebODCPM*, *rebRFUH*, and *rebT*. In operons *rebODCPM* and *rebRFUH*, genes are translationally coupled, with *rebM*, whose start codon is only 26 nt away from *rebP* stop codon, as the only exception. Each of these polycistronic transcriptional units ends with an inverted repeat (nt 17,412–17,447, and nt 23,872–23,927), which could lead to stem-loop secondary structures at the mRNA and transcription termination.

Genes Probably Involved in Formation of the Indolocarbazole Core

Based on sequence analysis and database comparison, we propose a group of four genes, *rebO*, *rebD*, *rebC*, and *rebP*, to be involved in indolocarbazole core formation. Some experimental evidence for that assumption is also presented in this paper (see below). Previous reports have shown that the indolocarbazole core of both rebeccamycin and staurosporine is biosynthesized from two units of L-tryptophan, with the carbon skeleton incorporated intact [13, 14]. Moreover, indolepyruvic acid has been identified as an intermediate of rebeccamycin biosynthesis [15]. In agreement with this, the *rebO* gene product belongs to a family of flavin-containing amine oxidases, including L-amino acid oxidase AIP from the fish *Scomber japonicus* [16] and putative amine oxidase CC1091 from *Caulobacter crescentus* [17]. The protein encoded by *rebO* could catalyze the first step in indolocarbazole biosynthesis, the oxidative deamination of L-tryptophan to yield indolepyruvic acid. Very recently, it has been shown that feeding either 5-fluorotryptophan or 6-fluorotryptophan to cultures of *Saccharothrix aerocolonigenes* induces the production of novel fluorinated analogs of rebeccamycin [18]. This suggests that RebO may accept different tryptophan derivatives for oxidative deamination.

The exact nature of intermediates and reactions between indolepyruvic acid (or 7-chloro-indolepyruvic acid) and rebeccamycin aglycone is unknown, but a pathway including condensation of two tryptophan-derived units, several oxidations, and two decarboxylations seems plausible. Next we describe and discuss genes *rebD*, *rebC*, and *rebP*, probably responsible for these biosynthetic steps.

rebD encodes a protein that showed significant similarities to two database proteins, VioB from *Chromobacterium violaceum* [19, 20] and hypothetical protein SC9A4.17 from *Streptomyces coelicolor* (accession number CAC01644) [21]. VioB is involved in early steps of the biosynthesis of violacein, a blue pigment whose carbon skeleton originates from two molecules of L-tryptophan accompanying decarboxylation [19, 20]. It has been speculated that VioB is a multifunctional enzyme that catalyzes a 1,2-indole shift as well as a condensation reaction between the two tryptophan (or tryptophan-derived) units to generate the violacein pyrrole ring [19]. We propose that RebD performs a similar function in rebeccamycin biosynthesis by condensing two tryptophan-derived units (probably indolepyruvic acid or its 7-chlorinated form) to yield the first *bis*-indole intermediate. However, important functional differences between violacein and indolocarbazole biosynthetic enzymes must exist. The non-indolic nitrogen in violacein exclusively originates from one of the tryptophan units that is incorporated intact; the other tryptophan unit somehow loses its α-amino group and suffers an intramolecular 1,2-indole shift [20]. On the other hand, the origin of the non-indolic nitrogen in indolocarbazoles remains mysterious; it appears that the side-chain nitrogen in tryptophan is cleaved and not incorporated into staurosporine [22].

As mentioned above, some genes needed for biosynthesis of the staurosporine sugar moiety in *S. longisporoflavus* DSM10189 have been previously identified [8]. These authors reported 10 kb of contiguous DNA sequence, deposited as three fragments with accession numbers A60304, A60301, and A60305. However, a de-

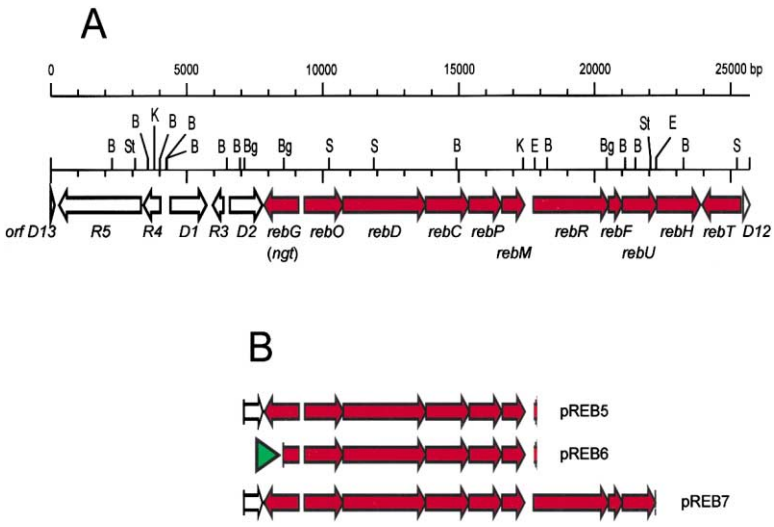


Figure 2. The Rebeccamycin Gene Cluster and Its Heterologous Expression

(A) Restriction map and genetic organization of the *Saccharothrix aerocolonigenes* ATCC39243 DNA region that includes the rebeccamycin biosynthetic gene cluster (in red). Proposed functions for individual ORFs are summarized in Table 1. Sites for restriction enzymes are abbreviated as follows: B, BamHI; Bg, BglII; E, EcoRI; K, KpnI; S, ScaI; and St, StuI.

(B) Schematic representation of inserts contained in plasmids pREB5, pREB6, and pREB7. A green arrow indicates the *ermE**p promoter included in pREB6.

tailed analysis of the complete DNA sequence was not reported, and they identified only the following ORFs in their patent application (but not in their database entry): (1) in accession number A60304—gene 2 (putative sugar *N*-methyltransferase), gene 5 (putative sugar aminotransferase), and three ORFs with no proposed function (genes 1, 3, and 4); (2) in accession number A60301—gene 1 (sugar *O*-methyltransferase) and gene 2 (putative sugar 3,5-epimerase); and (3) in accession number A60305—no ORFs reported. Because we expected that rebeccamycin and staurosporine gene clusters would share some equivalent genes for indolocarbazole core biosynthesis, we decided to perform a detailed sequence analysis of the staurosporine DNA sequence. This analysis revealed a 5′-truncated ORF, previously not reported, at an end of the staurosporine sequence

Table 3. Deduced Functions for Genes in the Rebeccamycin Gene Cluster

Gene	Amino Acids	Closest Similar Protein (% Identity/Similarity), Accession Number	Proposed Function
orfD13	44[a]	–	unknown
orfR5	1003	regulatory protein AfsR from *Streptomyces coelicolor* (36/49), P25941	regulatory protein
orfR4	210	putative dipeptidase SCC53.19 from *S. coelicolor* (52/63), CAB93448	D-alanyl-D-alanine dipeptidase
orfD1	472	hypothetical protein SC6E10.10 from *S. coelicolor* (31/48), CAB51964	secreted esterase
orfR3	133	–	unknown
orfD2	395	hypothetical protein CC0823 from *Caulobacter crescentus* (53/66), AAK22808	unknown
rebG/ngt	421	probable glycosyltransferase from *Deinococcus radiodurans* (45/57), F75587	*N*-glycosyltransferase
rebO	473	L-amino acid oxidase AIP from *Scomber japonicus* (29/44), CAC00499	L-amino acid oxidase
rebD	1013	VioB from *Chromobacterium violaceum* (34/47), AAD51809	*bis*-indole formation
rebC	529	2.4-dihydroxybenzoate monooxygenase from *Sphingomonas* sp. (32/46), CAA51370	FAD-containing monooxygenase
rebP	397	cytochrome P450 YjiB from *Bacillus subtilis* (37/53), O34374	P450 heme-thiolate protein
rebM	273	methyltransferase from *Amycolatopsis mediterranei* (50/66), AAC01738	methyltransferase
rebR	923	transcriptional activator NysRI from *Streptomyces noursei* (25/35), BAB50206	regulatory protein
rebF	170	putative FMN:NADH oxidoreductase Gra-orf34 from *S. violaceoruber* (39/52), CAA09661	flavin reductase
rebU	426	putative integral membrane ion antiporter from *A. orientalis* (40/54), CAB45049	integral membrane transporter
rebH	530	tryptophan halogenase PrnA from *Pseudomonas chlororaphis* (55/72), AAD46360	$FADH_2$-dependent halogenase
rebT	473	putative antibiotic antiporter FrnF from *S. roseofulvus* (44/62), AAC18101	integral membrane transporter
orfD12	81[a]	hypothetical transcriptional regulator from *M. tuberculosis* (43/63), Q10810	regulatory protein

[a] Incomplete ORF.

(nt 1–332 of accession number A60304). The deduced product of this ORF showed similarity to the C termini of *C. violaceum* VioB (48% identity, 57% similarity) and of RebD (63% identity, 70% similarity). The presence of a *vioB*/*rebD* homolog in the violacein, staurosporine, and rebeccamycin gene clusters reinforces the idea of its involvement in the condensation of two tryptophan-derived units.

Two of the identified genes, *rebC* and *rebP*, could be involved in oxidative conversion of the first *bis*-indole intermediate into the indolocarbazole core. Analysis of *rebC* gene product showed an N-terminal FAD binding domain [23] and clear similarities to a family of FAD-containing, NAD(P)H-dependent monooxygenases acting on aromatic compounds [24]. The highest similarities were found to 2,4-dihydroxybenzoate monooxygenase DxnD from *Sphingomonas* sp. [25] and to putative polyketide hydroxylase SchC from *Streptomyces halstedii* [26]. DxnD exhibits NADH oxidation activity with 2,4-dihydroxybenzoate and is believed to catalyze decarboxylating monooxygenation to yield hydroxyquinol as a product [25]. In our above-mentioned analysis of the staurosporine DNA sequence, we also found a previously unreported ORF (nt 1845–2122 of accession number A60301, plus nt 1–1373 of accession number A60305) whose deduced product showed similarities to the same family of monooxygenases, including *Sphingomonas* sp. DxnD (32% identity, 47% similarity) and RebC (63% identity, 71% similarity). According to this, a monooxygenase such as RebC could catalyze decarboxylative monooxygenations, as shown for *Sphingomonas* sp. DxnD [25]. On the other hand, *rebP* gene product is homologous to P450 heme-thiolate proteins [27]. The highest similarities were found to a putative cytochrome P450 YjiB from *Bacillus subtilis* (accession number O34374) and a probable cytochrome P450 hydroxylase from *S. coelicolor* (accession number T36526). In our analysis of the staurosporine DNA sequence, we found that the deduced product of accession number A60304 "gene 1" (nt 378–1655), of unreported function, was also similar to P450 heme-thiolate proteins, including *B. subtilis* YjiB (32% identity, 51% similarity) and RebP (51% identity, 60% similarity). It is worth mentioning that some P450 enzymes are able to form new C-C or C-O bonds between aromatic moieties by phenol oxidative coupling [28, 29]. One may hypothesize that, in an analogous way, RebP could catalyze oxidative ring closure of a *bis*-indole intermediate to form the indolocarbazole core.

Genes Probably Involved in Indolocarbazole Modification

After (or during) formation of the indolocarbazole core, additional reactions must occur to arrive at the final rebeccamycin molecule. These would include glycosylation and halogenation. A gene encoding an indolocarbazole *N*-glycosyltransferase, *ngt*, has been previously identified from *Saccharothrix aerocolonigenes* [9]. A comparison of the 1898 nt DNA sequence reported for *ngt* (accession number AB023953) [9] and the corresponding sequence determined by us (complementary to nt 7484–9390 of accession number AJ414559) revealed several discrepancies in *ngt*/*rebG* coding sequence and also in flanking sequences. In particular, the deduced products of *ngt* and *rebG* differ in eight amino acid residues, and RebG includes six additional residues at its C terminus. Some of the discrepancies in the *ngt*/*rebG* flanking sequences also affect *orfD2* and *rebO*, genes not previously defined, and make them unfunctional in the published *ngt* sequence [9]. After rechecking and confirming our sequence, we find it plausible that these discrepancies originate from sequencing errors in the published *ngt* sequence. So, from now on we use the name *rebG* for the corrected sequence of the previously known *ngt* gene. RebG belongs to "family 1" of NDP-sugar glycosyltransferases [30]. The highest similarities were found to a probable glycosyltransferase from *Deinococcus radiodurans* (accession number F75587) and zeaxanthin glucosyl transferase CrtX from *Synechocystis* sp. (accession number S74500). It has been previously shown that *rebG* (*ngt*) encodes an *N*-glycosyltransferase that, when expressed in *S. lividans*, is able to introduce a D-glucose moiety into indolocarbazoles J-104303 and 6-*N*-methylarcyriaflavin C [9]. We propose that the physiological function of RebG in *Saccharothrix aerocolonigenes* is the catalysis of an *N*-glycosidic bond between a nucleotide-activated D-glucose and the rebeccamycin indolocarbazole core (Figure 3).

A methylation step is required to yield the 4-*O*-methyl-D-glucose moiety present in rebeccamycin. The deduced product of *rebM* is similar to S-adenosylmethionine-dependent methyltransferases, including a methyltransferase from *Amycolatopsis mediterranei* (accession number AAC01738) and MitM from *Streptomyces lavendulae* [31]. RebM is also similar to sugar methyltransferases deduced from the reported staurosporine DNA sequence: *O*-methyltransferase encoded by gene 1 of accession number A60301, nt 845–1684 (36% identity, 53% similarity), and putative *N*-methyltransferase encoded by gene 2 of accession number A60304, nt 1747–2553 (43% identity, 58% similarity). RebM could catalyze a methylation at the 4-hydroxy position of a D-glucose moiety, probably after the glycosylation step during rebeccamycin biosynthesis (Figure 3).

Chlorination in the rebeccamycin pathway is probably carried out by the *rebH* gene product. RebH resembles $FADH_2$-dependent halogenases [7], including tryptophan halogenase PrnA from *Pseudomonas* spp. and *Myxococcus fulvus* [32] and putative tryptophan halogenases from *Caulobacter crescentus* [17]. PrnA requires $FADH_2$, O_2, and chloride ions to regioselectively convert tryptophan into 7-chlorotryptophan. $FADH_2$ is provided by a flavin reductase that reduces FAD with the help of NADH and that can be substituted by the corresponding enzyme from other organisms [7]. Rebeccamycin is chlorinated precisely at indole positions equivalent to that of 7-chlorotryptophan. This halogenation reaction could be catalyzed by RebH, either on tryptophan or on a later intermediate during rebeccamycin biosynthesis (Figure 3, see below). Chloride ions may be substituted by bromide ions for RebH halogenation because bromoindolocarbazole compounds can be isolated from cultures of *Saccharothrix aerocolonigenes* supplemented with potassium bromide [33].

The deduced product of *rebF* belongs to a family of

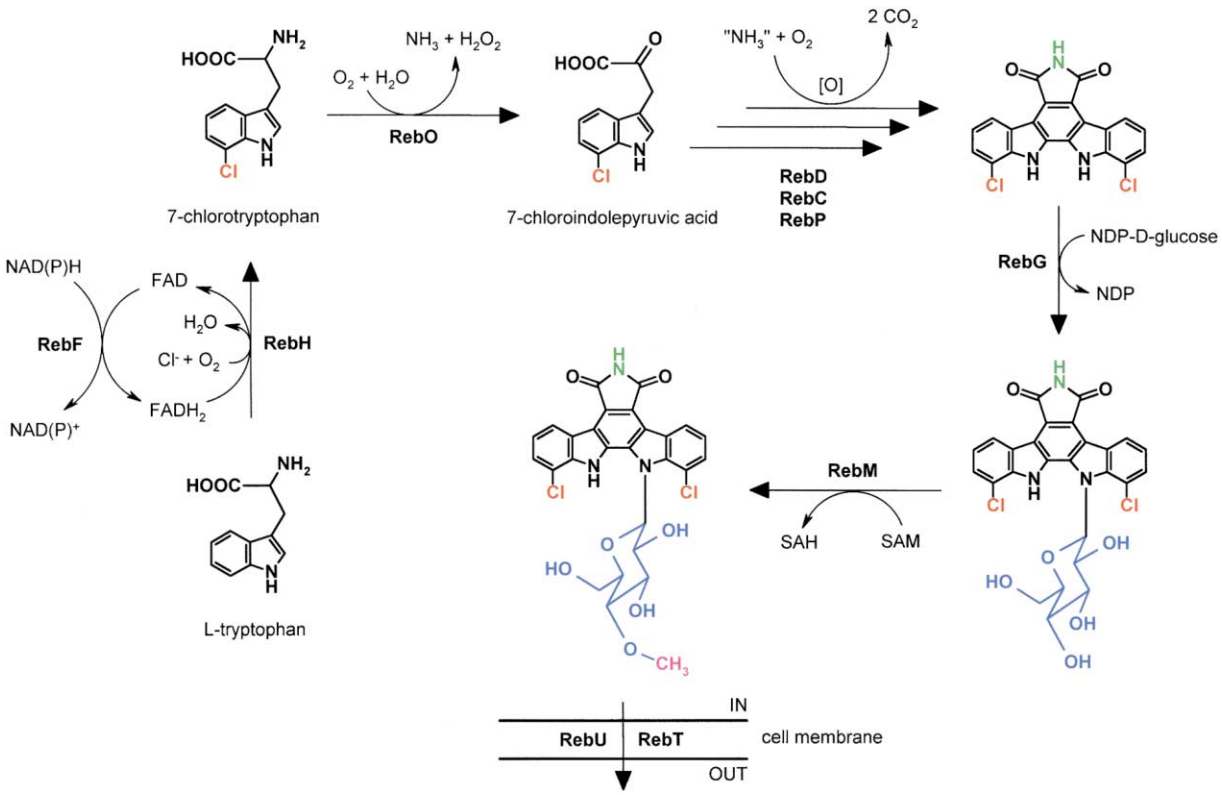


Figure 3. Proposed Biosynthetic Pathway for Rebeccamycin

The relative order of some reactions, such as the halogenation step, may well be different from the one presented here.

flavin:NAD(P)H reductases, the majority of which are part of two-component flavin-diffusible monooxygenase systems [34]. Highest resemblances were found to ActVB homologs from *S. violaceoruber* (Gra-orf34) [35], *S. roseofulvus* (FrnH) [36] and *S. coelicolor* (ActVB) [37], and to StyB protein from *Pseudomonas fluorescens* [38]. ActVB (= ActI-ORF6) is a flavin:NADH oxidoreductase that participates in the last step of actinorhodin biosynthesis, a symmetrical dimerization of two benzoquinone units by a phenolic oxidative coupling. ActVB probably supplies reduced FMN to another as-yet-unknown enzyme directly involved in oxidative chemistry [37]. On the other hand, StyB is a part of stirene monooxygenase, a two-component system (StyAB) responsible for the transformation of stirene to epoxystirene [38]. We propose that RebF and RebH form a two-component halogenase system, in which RebF supplies the reduced, diffusible flavin that RebH needs to function (Figure 3). Very recently, a pair of *rebF-rebH* homologs has been found to be linked in the genome of *S. albogriseolus*, producer of the halogenated metabolite thienodolin (Corina Schmid, personal communication). However, an additional role for RebF in supplying reduced flavin to other rebeccamycin enzymes (RebO, RebC) cannot be ruled out.

Genes Probably Involved in Resistance and Regulation

Two genes, *rebU* and *rebT*, could participate in rebeccamycin resistance and/or secretion. The deduced product of *rebU* is similar to Na^+/H^+ exchange membrane proteins, which contain 10–12 transmembrane regions at the N terminus and a large cytoplasmic region at the C terminus [39]. Characterized members of this family function as antiporters of Na^+ (or K^+) and H^+ and play a key role in maintaining cellular pH and other processes. RebU was found to be most similar to a putative integral membrane ion antiporter from *Amycolatopsis orientalis* (accession number CAB45049) and putative antibiotic transporter AviJ from *S. viridochromogenes* [40]. Very recently, a similar protein (ComF) has been found encoded in the complestatin biosynthetic gene cluster of *S. lavendulae* [41].

The *rebT* gene product, the second candidate for rebeccamycin resistance and/or secretion, belongs to the major facilitator family of integral membrane transporters, responsible for antibiotic or antiseptic efflux with the aid of transmembrane electrochemical gradients [42]. Highest similarities were detected to putative antibiotic antiporter FrnF from *S. roseofulvus* [36] and to putative efflux protein EncT from *S. maritimus* [43]. Plasmid pCS006, which contained *rebT*, was introduced in *S. albus* and was found to confer resistance to rebeccamycin.

Concerning regulation, we found that *rebR* gene product belongs to subfamily LAL of regulatory proteins of the LuxR family, involved in ATP-dependent transcriptional activation [44]. LAL proteins contain an N-terminal ATP binding domain and a C-terminal LuxR-type DNA binding domain. RebR shows low end-to-end similarities to LAL proteins, including transcriptional activator

NysRI from *Streptomyces noursei* [45]. We propose that RebR could function as a transcriptional activator of the expression of rebeccamycin biosynthetic genes, and some experimental evidence (see below) points toward this idea.

Genes Probably Not Involved in Rebeccamycin Biosynthesis

Results from database searching were not significant enough for us to propose functions for *orfD13*, *orfR3*, and *orfD2*. On one hand, deduced gene products of *orfD13* (truncated) and *orfR3* did not have any homologs in the databases, so their possible functions are unknown. On the other hand, OrfD2 is similar to conserved hypothetical proteins from several bacteria, including CC0823 from *Caulobacter crescentus* [17] and SCF41.20c from *S. coelicolor* [21]. It also shows weak, limited similarities to putative 2-hydroxyhepta-2,4-diene-1,7-dioate isomerases such as BH2000 from *Bacillus halodurans* [46], but at the moment we do not feel this is enough to propose a function for OrfD2.

Deduced products of another group of ORFs (*orfR5*, *orfR4*, and *orfD1*) showed similarities to proteins of known functions, but these proteins are apparently unrelated to rebeccamycin biosynthesis. OrfR5 resembles AfsR, a global regulatory protein conditionally required for secondary metabolism in *S. coelicolor* [47]. OrfR4 is similar to putative dipeptidase SCC53.19 from *S. coelicolor* [21] and to *Synechocystis* sp. confirmed D-alanyl-D-alanine dipeptidase SynVanX, probably having a role in cell-wall turnover [48]. Finally, the deduced product of *orfD1* is probably a secreted protein with a signal peptide cleaved between amino acids 47 and 48 (as predicted by the SignalP program [49]) and resembles hypothetical protein SC6E10.10 from *S. coelicolor* [21] and fusidic-acid esterase FusH, an extracellular enzyme from *S. lividans* [50].

At the right end of the sequenced region (Figure 2A), *orfD12* (incomplete) encodes the N-terminal part of a protein homologous to MarR family regulatory proteins, many of which respond to phenolic compounds [51]. Highest similarities were found to hypothetical transcriptional regulators Rv2887 from *Mycobacterium tuberculosis* [52] and SC1A4.04 from *S. coelicolor* [21]. The involvement of OrfD12 in gene regulation for rebeccamycin biosynthesis is unclear because rebeccamycin genes are efficiently expressed from cosmid 14E8, in which *orfD12* is truncated.

Production of Rebeccamycin Derivatives in a Heterologous Host

In order to confirm some of the gene functions deduced by sequence analysis and to produce some rebeccamycin biosynthetic intermediates, we constructed several plasmids containing fragments of cosmid 14E8 (Figure 2B). These multicopy plasmids were introduced into *S. albus* J1074 and, after cultivation of these recombinant strains in the presence of thiostrepton (resistance marker in the vector), extracts of the new strains were analyzed by HPLC for production of indolocarbazole derivatives. The first plasmid, pREB5, contained genes *rebG*, *rebO*, *rebD*, *rebC*, *rebP*, and *rebM*, with their own promoter and/or regulatory sequences. On the basis of the gene functions present in pREB5, it was expected that the plasmid would direct the biosynthesis of deschloro-rebeccamycin. However, transformants *S. albus* J1074/pREB5 did not produce any detectable indolocarbazole (Figure 4A). Because this could have been a consequence of some missing regulatory factor causing poor expression of the cloned genes, two new plasmids, pREB6 and pREB7, were constructed (Figure 2B). Plasmid pREB6 contained genes *rebO*, *rebD*, *rebC*, *rebP*,

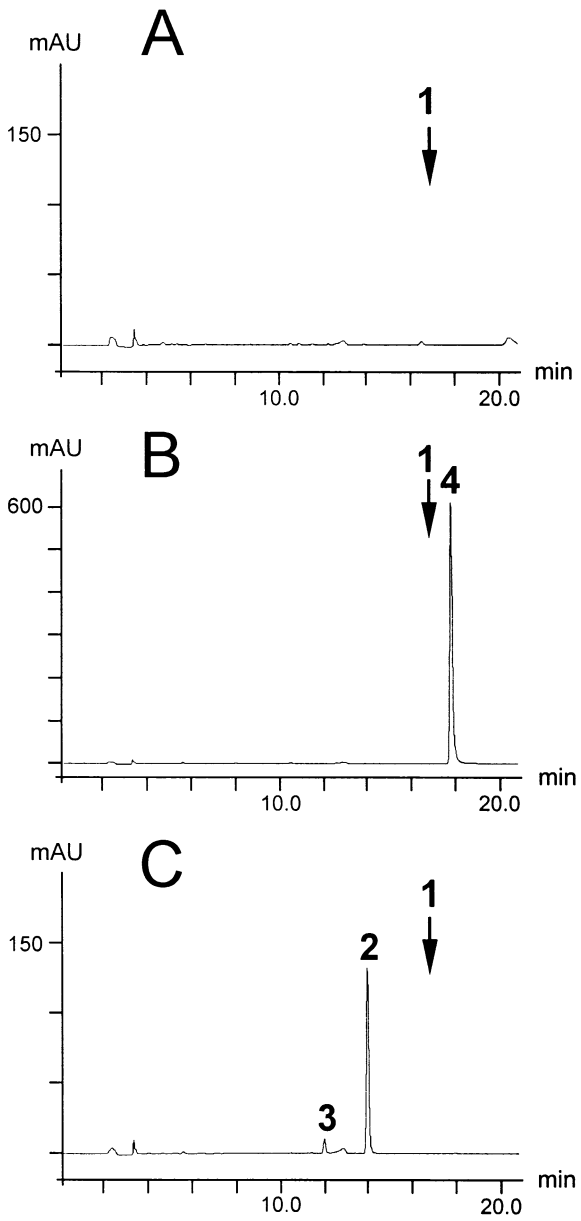


Figure 4. Heterologous Expression of Selected Subsets of Genes from the Rebeccamycin Cluster

HPLC analysis of extracts from (A) *S. albus* J1074/pREB5 (the control *S. albus* J1074/pEM4 showed an identical chromatogram), (B) *S. albus* J1074/pREB6, and (C) *S. albus* J1074/pREB7. The elution time of rebeccamycin, used as a standard and previously purified from *Saccharothrix aerocolonigenes*, is indicated by an arrow. Chemical structures of compounds 1–4 are shown in Figure 1.

and *rebM*, under the control of the strong and constitutive promoter *ermE**p [53], so it was expected to direct the biosynthesis of nonchlorinated rebeccamycin aglycone. Plasmid pREB7 included genes *rebG*, *rebO*, *rebD*, *rebC*, *rebP*, *rebM*, *rebR*, *rebF*, and a 3′-truncated *rebU*, and we expected it could produce deschloro-rebeccamycin. HPLC analysis (Figures 4B and 4C) of extracts from transformants *S. albus* J1074/pREB6 and *S. albus* J1074/pREB7 indeed revealed three new products with characteristic indolocarbazole absorption spectra.

These new biosynthetic products were purified by preparative HPLC and characterized by high-resolution MS and NMR, confirming them as being deschloro-rebeccamycin aglycone (from *S. albus* J1074/pREB6), deschloro-rebeccamycin, and demethyl-deschloro-rebeccamycin (the last two products are from *S. albus* J1074/pREB7), respectively (Figure 1).

The putative aglycone showed a main peak with an m/z value of 325.0845, which corresponds to the calculated molecular mass of the nonchlorinated, deglycosylated rebeccamycin derivative, 325.0851. Interestingly, this ion was also present in the spectra of all the other products analyzed in this study. The ^{13}C NMR experiment showed only ten signals, as expected because of the symmetry of the molecule (Table 1, compound 4), and none of the typical carbohydrate carbons with δ in the range of 60–90 was present in the spectrum of this compound. The HSQC and HMBC experiment allowed the assignments of most carbons; however, the assignments of C1 and C4 and those of C2 and C3 may be interchanged. The integrals of the ^{1}H NMR spectrum along with the MS data clearly reveal the symmetry of the molecule because each signal corresponds to two protons, except the one of the central N6-H. The presence of the new doublet at δ 8.96 (d, 2H, 1-H, and 11-H), which is part of an ABCD system, proves that the aromatic rings contain here four adjacent protons, which proves that no chlorine atoms are any longer attached to the 1 and 11 positions. Thus, the NMR data also confirm the proposed structure, deschloro-rebaccamycin aglycone (**4**, Figure 1).

The second compound gave a molecular ion with a high-resolution m/z ratio of 501.1543, which is in agreement with the molecular mass of deschloro-rebeccamycin (calculated for $C_{27}H_{23}N_3O_7$: 501.1536). The broadband-decoupled ^{13}C NMR spectrum of the compound shows all the expected signals; most of the assignments were deduced from the comparison with rebeccamycin and from the 2D homonuclear and heteronuclear correlation experiments (HSQC, HMBC, and DQCOSY, Table 1). As indicated from the NMR data, a 4-O-methyl glucose residue was clearly present in the structure of the metabolite. The data show in particular that the compound contains only two nitrogen bound protons (N13-H and N6-H); the latter shows—as in all compounds described here—a strong coupling with the characteristic carbonyl signal at $\delta_C \approx 171$ in the HMBC spectrum. Also, the usual cross peaks of N13-H with C12b/C13a were observed, as well as the couplings of C1′-H of the 4-O-methyl glucose moiety with C12a and C11a. The protons attached at C1 and C11 ruled out the presence of chlorine atoms at the corresponding positions. These data combined with the MS data prove the structure of deschloro-rebeccamycin (**2**, Figure 1).

The third isolated compound was a minor rebeccamycin derivative that had a demethylated sugar residue and also lacked both chlorine atoms. This follows from a high-resolution EI-MS experiment, in which the compound gave a molecular ion with m/z ratio of 487.1378, and the prominent aglycone peak at m/z 325 found also in the MS spectra of all the other analyzed products. The high-resolution mass of 487.1378 is in agreement with the calculated mass for $C_{26}H_{20}N_3O_7$ of 487.1380. Unlike rebeccamycin, the compound lacks both chlorine atoms because both the 1-H and the 11-H signals are present in the ^{1}H NMR spectrum. Also obvious from the ^{1}H NMR spectrum was that the O-methyl group in the 4′ position is missing. Instead, a new hydroxyl group at δ 5.34 (4′-OH) was found in close proximity to 4′-H (δ 3.96), as indicated by the H,OH coupling observable in the ^{1}H- and H,H-DQCOSY spectrum. A coupling between 4′-OH and C4′ (δ_C 68.2) was also detected in the HMBC spectrum. All other signals are very similar to those of compound **2**. Also, the ^{13}C NMR spectrum (Table 1, compound **3**) of this third new compound is very similar to the one of deschloro-rebeccamycin; the only differences are that the signal of C4′ (δ 68.2) is shifted upfield approximately 10 ppm, in agreement with the a missing OCH_3 group in this position, and that there is no signal from an OCH_3 group. These data in combination with the high-resolution MS data prove that this third compound is indeed 4′-demethyl-deschloro-rebeccamycin (**3**). The occurrence of a minor rebeccamycin derivative that is glycosylated and not methylated may seem illogical because plasmid pREB7 included the putative methyltransferase gene *rebM*. However, similar results have been previously obtained when *Saccharothrix aerocolonigenes* ATCC39243 was fed with unnatural substrates; bioconversion of indolocarbazole aglycone J-104303 yielded the corresponding glycosylated (but not methylated) compound [9], and feeding either 5-fluorotryptophan or 6-fluorotryptophan resulted in production of the corresponding fluorinated 4′-demethyl-deschloro-rebeccamycins or a mixture of both methylated and nonmethylated derivatives [18].

The fact that deschloro-rebeccamycin was produced by a strain harbouring pREB7, but not by a strain carrying pREB5, suggests that transcription of *rebODCPM* probably needs to be activated by the *rebR* product. In the case of pREB6, the strong promoter *ermE**p overcomes this problem because of its constitutive expression. However, *rebG* transcription might not need to be activated by RebR because *rebG* (*ngt*) was expressed in *S. lividans* without any additional promoters [9], although this could also be explained by read-through transcription from vector promoter sequences. The structures of the isolated rebeccamycin derivatives reinforce the functions proposed for the *reb* genes from sequence analysis and database comparisons.

Other laboratories have previously used chemical synthesis or hemisynthesis to obtain compounds structurally identical to nonchlorinated rebeccamycin aglycone and deschloro-rebeccamycin [6]. These authors analyzed the biological activities of these compounds in antimicrobial and antiproliferative tests, as well as in inhibition assays toward protein kinases C (PKC) and A (PKA) and topoisomerases I (Topo I) and II (Topo II) [6,

12]. No inhibition of Topo II was detected with any of the two compounds, as is the case for rebeccamycin itself. Nonchlorinated rebeccamycin aglycone did not show any antimicrobial activity against the different microorganisms tested, including two *Streptomyces* spp. [12]. This result may explain why strain *S. albus* J1074/pREB6 is able to produce this compound and survive, despite the fact that no rebeccamycin resistance or transporter gene is included in pREB6. In the same report, dechlorinated rebeccamycin showed no growth-inhibitory effect against the microorganisms tested, with the only exception being a very weak activity toward one of the two *Streptomyces* spp. assayed [12]. Strain *S. albus* J1074/pREB7 may be protected from the weakly toxic effects of deschloro-rebeccamycin by a 3′-truncated *rebU*, included in pREB7, which encodes an almost complete putative transporter (lacking only 13 amino acids) that could participate in metabolite secretion/resistance. Interestingly, when compared to rebeccamycin, dechlorinated rebeccamycin showed a higher inhibitory activity toward Topo I and similar in vitro antiproliferative activity against murine B16 melanoma and P388 leukemia cells [6]. From the study of a range of glycosylated and nonglycosylated rebeccamycin semi-synthetic derivatives, it has been concluded that the presence of the chlorine atoms on the indole residues is detrimental to both PKC and Topo I inhibitory activities [6]. In this respect, we have shown in this work that the expression of some rebeccamycin biosynthetic genes allows the formation of three nonchlorinated rebeccamycin derivatives that are potentially useful as lead compounds for further drug development. It is also worth mentioning that nonchlorinated rebeccamycin aglycone has comparatively poor antiproliferative and anti-Topo I activities, but it shows some inhibition toward PKC and PKA (whereas neither rebeccamycin nor dechlorinated rebeccamycin inhibit protein kinases). In this sense, it seems that the sugar residue attached to the indolocarbazole chromophore is critical for the drug ability to interfere with Topo I as well as for the formation of intercalation complexes with the DNA [6]. Furthermore, nonchlorinated rebeccamycin aglycone is structurally identical to arcyriaflavin A, a natural product isolated from the slime mold *Arcyria denudata* [54] and the marine ascidian *Eudistoma* sp. [55]. Recently, it has been reported that arcyriaflavin A is a potent inhibitor of human cytomegalovirus replication in cell culture, which could eventually lead to a new class of drugs [56, 57].

With the present work, the utility of manipulating the rebeccamycin biosynthetic genes to produce a variety of indolocarbazole compounds with different biological properties becomes clear. Further experiments are now in progress to produce novel indolocarbazole derivatives and to achieve a better understanding of the biochemical processes involved in rebeccamycin biosynthesis.

Significance

Several indolocarbazoles derived from natural products are presently undergoing phase I/II clinical trials for anticancer chemotherapy, based on their activity as inhibitors of DNA topoisomerases or protein kinases. Additionally, a great interest is focused on the synthesis of a growing number of new indolocarbazole analogs with promising anticancer, antiviral, antifungal, antihypertensive, or neuroprotective activities. Complementary to the efforts in chemical synthesis, the genetic manipulation of genes governing indolocarbazole biosynthesis offers a promising alternative for preparation of these compounds. This work represents the first characterization of a complete gene cluster governing the biosynthesis of an indolocarbazole metabolite, rebeccamycin. Our results are consistent with precursor incorporation studies showing that rebeccamycin is biosynthetically derived from two units of L-tryptophan, one D-glucose, and one L-methionine in *Saccharothrix aerocolonigenes*. Several genes encoding enzymes with unusual or poorly understood activities have been identified, among them those involved in *bis*-indole/indolocarbazole formation, *N*-glycosylation, and halogenation. In addition, genes responsible for indolocarbazole core formation are of special interest because they might be useful as probes for identifying related natural-product biosynthetic genes from other organisms. The availability of the *reb* gene cluster should facilitate future attempts, using combinatorial biosynthesis and rational metabolic pathway engineering, to produce novel indolocarbazoles with improved, or even new, therapeutic activities. The advantage of having this information has already been demonstrated through heterologous expression of (1) the complete gene cluster, which provided a several-fold increase in rebeccamycin production and (2) selected subsets of genes, which allowed us to obtain three different rebeccamycin-related compounds.

Experimental Procedures

Bacterial Strains, Culture Conditions, and Vectors

Saccharothrix aerocolonigenes ATCC39243, *Streptomyces albus* J1074 [58], *Escherichia coli* XL1-Blue [59], *E. coli* ED8767 [60], and *Micrococcus luteus* ATCC1024 were used in this work. Vectors pKC505 [10], pWHM3 [61], and pEM4 [62] have been described previously, and LITMUS 28 (New England BioLabs) was from a commercial source. Shuttle vector pUWL201 was obtained from U. Wehmeier and W. Piepersberg (Wuppertal, Germany). For sporulation, *Saccharothrix aerocolonigenes* and *S. albus* were routinely grown for 7 days at 30°C on agar plates containing either A medium [63] or Bennett's agar [64]. When antibiotic selection of transformants was needed, 100 μg/ml ampicillin, 25 μg/ml thiostrepton (5 μg/ml in liquid media), 25 μg/ml apramycin, or 20 μg/ml tobramycin was used.

DNA Manipulation

Total DNA isolation, plasmid DNA preparations, restriction endonuclease digestions, ligations, and other DNA manipulations were performed according to standard procedures for *E. coli* [65] and for *Streptomyces* [64]. A genomic library of *Saccharothrix aerocolonigenes* was constructed in pKC505 according to literature procedures [10] and screened with a probe made from the *N*-glycosyltransferase gene, *ngt* [9]. This probe was obtained by polymerase chain reaction (PCR) according to standard procedures. Total DNA from *Saccharothrix aerocolonigenes* with oligoprimers CS003 (5′-*TAGAA*TTCATCGAACCCGCGGCC-3′, altered sequence in italics, EcoRI underlined) and CS004 (5′-*TATAAGCT*TCGGCTGCCA GCGCTC-3′, altered sequence in italics, HindIII underlined), de-

signed to amplify a DNA fragment encompassing nt 552–1146 from the published *ngt* sequence, was used [9]. Digoxigenin labeling of DNA probes, Southern analysis, hybridization, and detection were performed according to literature procedures [65] and manufacturer recommendations (Boehringer Mannheim). Transformation of *S. albus* protoplasts followed procedures routinely used for *Streptomyces* [64].

PCR amplification of halogenase-encoding regions was performed with degenerate oligoprimers trpst+ (5′-TATCGGATCCGG STGGACCTGGRASATYCC-3′, S = (C, G), R = (A, G), Y = (C, T), BamHI site underlined) and trpst- (5′-AGTTGGTACCGSGCSGCG TASAKGAAGTA-3′, K = (G, T), KpnI site underlined), as previously described [11].

Construction of Plasmids

Cosmid 14E8 was partially digested with BglII and EcoRI, and fragments were subcloned in LITMUS 28, yielding (among others) plasmids pREB1, pCS002, pCS003, and pCS004. One of the resulting plasmids, pREB1, contained a BglII-EcoRI DNA fragment encompassing nt 7,119–17,783 of the sequenced region. This 10.6 kb DNA fragment was obtained from pREB1 as a SpeI-EcoRI fragment and transferred into the XbaI-EcoRI sites of pWHM3, yielding plasmid pREB5. Plasmid pCS002 contained a BglII-EcoRI DNA fragment encompassing nt 8,562–17,783 of the sequenced region. This 9.2 kb DNA fragment was obtained from pCS002 as a BglII-EcoRI fragment and transferred into the BamHI-EcoRI sites of pEM4, yielding plasmid pREB6. Plasmids pCS003 and pCS004 contained, respectively, an EcoRI-BglII DNA fragment (nt 17,783–20,425) and a BglII-EcoRI DNA fragment (nt 20,425–22,241). These two DNA inserts were obtained from pCS003 and pCS004 as EcoRI-BglII and BglII-EcoRI fragments, respectively, and ligated to EcoRI-digested pREB5, resulting in plasmid pREB7.

Plasmid pCS006 was generated by cloning an EcoRI-XbaI DNA fragment from cosmid 14E8 into vector pUWL201. This fragment, which included *rebT*, spanned from nt 22,241 to the end of the insert (nt 25,681) plus 500 bp from pKC505.

DNA Sequencing and Analysis

DNA sequencing was performed on double-stranded DNA templates in pUC18 with the dideoxynucleotide chain termination method [66] and the Cy5 AutoCycle Sequencing Kit (Pharmacia Biotech). An ALF-express automatic DNA sequencer (Pharmacia) was used to sequence both DNA strands with primers supplied in the kit or with internal oligoprimers (17-mer). Computer-aided database searching and sequence analysis were carried out with the University of Wisconsin Genetics Computer Group software [67] and the BLAST program [68].

Tests for Rebeccamycin Antibacterial Activity

Growth inhibition tests on *Micrococcus luteus* were performed to show antibacterial activity of the rebeccamycin present in extracts from *S. albus* transformants. For this, aliquots of acetone extracts were added to paper discs placed on agar plates of TSB (Oxoid) at half nutrient concentration; the plates were preseeded with *M. luteus*. After 2 hr at 4°C, incubation was carried out at 37°C overnight. Rebeccamycin antibacterial activity was also tested against *S. albus* J1074/14E8 and the control *S. albus* J1074/pKC505. For this, different amounts of rebeccamycin (dissolved in acetone) were added to paper discs placed on plates of Bennett's agar (containing 25 μg/ml apramycin) preseeded with the corresponding *S. albus* strain. After 2 hr at 4°C, incubation was carried out at 30°C for 4 days.

HPLC Analysis

Detection of rebeccamycin and related compounds was performed by HPLC in a reversed-phase column (Symmetry C18, 4.6 × 250 mm, Waters), with acetonitrile and 0.1% trifluoroacetic acid in water as solvents. A linear gradient from 20% to 75% acetonitrile in 20 min, at a flow rate of 1 ml/min, was used. Detection and spectral characterization of peaks were performed with a photodiode array detector and Millennium software (Waters), and bidimensional chromatograms were extracted at 316 nm.

Isolation of Rebeccamycin and Related Compounds

Spores of strains *S. albus* J1074/14E8, *S. albus* J1074/pREB6, and *S. albus* J1074/pREB7 were inoculated in TSB medium (Oxoid) and incubated for 24 hr at 30°C and 250 rpm. Each seed culture was used to inoculate (at 2.5%, v/v), eight 2 liter Erlenmeyer flasks containing 400 ml of R5A medium (described as "modified R5 medium" in [63]). After incubation for 5 days in the above conditions, the cultures were centrifuged (12,000 rpm, 30 min). In the cultures of strain *S. albus* J1074/pREB7, rebeccamycin-related compounds were found both in the supernatants and in the pellets, whereas in the other two strains these compounds were largely associated with the pellets; therefore, their supernatants were discarded. The pellets corresponding to each strain were extracted with 400 ml acetone and shaken for 2 hr, after which the suspensions were centrifuged and the organic extract was evaporated in vacuo. The *S. albus* J1074/pREB7 culture supernatant was filtered and applied to a solid-phase extraction cartridge (Sep-Pak Vac 35 cc, Waters). The retained material was eluted with methanol and water, with a linear gradient from 0% to 100% methanol in 1 hr at 10 ml/min. Fractions were taken every 5 min and analyzed by HPLC. The rebeccamycin-related compounds produced by this strain eluted in the fraction taken between 40 and 45 min, which was evaporated in vacuo.

For purification, the material extracted from pellets of each strain was redissolved in 5 ml of a mixture of DMSO and acetone (50:50). The extract from strain *S. albus* J1074/14E8 was chromatographed in a μBondapak C18 radial compression cartridge (PrepPak Cartridge, 25 × 100 mm, Waters) via isocratic elution with acetonitrile and water (55:45) at 10 ml/min. The extract from strain *S. albus* J1074/pREB6 was chromatographed in a semipreparative column (XTerra RP18, 7.8 × 300 mm, Waters) via isocratic elution with methanol and water (90:10) at 3 ml/min. The material extracted from either the pellets or the broth of strain *S. albus* J1074/pREB7 was chromatographed on the semipreparative column mentioned above and eluted with acetonitrile and water (50:50) in isocratic conditions. The compounds collected in every case after multiple injections were dried in vacuo and finally lyophilized.

MS and NMR Methods

NMR spectra were recorded in d6-DMSO on a Varian Inova 400 instrument at 400 MHz for 1H and 100.6 MHz for 13C by the use of 1D spectra and 2D homo- and heteronuclear correlation experiments (1H, 13C, H,H-COSY, HSQC, and HMBC).

Electron ionization mass spectrometry (EI-MS) was used to determine the mass of the purified compounds. The high-resolution electron ionization (HR-EI) mass spectra were acquired at the University of South Carolina, Department of Biochemistry and Chemistry facilities in Columbia, SC with a VG70SQ double focusing magnetic sector MS instrument.

Acknowledgments

The authors wish to thank K.H. van Pée, S. Zehner, and C. Schmid for providing us primers for halogenase genes and for facilitating information prior to publication. We thank the U.S. Department of Defense for support of J.R. et al. and Obra Social Cajastur for financial support to C.S. We acknowledge Drs. W. Cotham and M. Walla (University of South Carolina) for providing the high-resolution MS spectra.

Received: January 11, 2002
Revised: February 12, 2002
Accepted: February 14, 2002

References

1. Gribble, G.W., and Berthel, S.J. (1993). A survey of indolo[2,3-a]carbazoles and related natural products. In Studies in Natural Products Chemistry, vol. 12. (Amsterdam: Elsevier Science), pp. 365–409.
2. Bailly, C., Qu, X., Graves, D.E., Prudhomme, M., and Chaires, J.B. (1999). Calories from carbohydrates: energetic contribution of the carbohydrate moiety of rebeccamycin to DNA binding

and the effect of its orientation on topoisomerase I inhibition. Chem. Biol. *6*, 277–286.

3. Akinaga, S., Sugiyama, K., Akiyama, T. (2000). UCN-01 (7-hydroxystaurosporine) and other indolocarbazole compounds: a new generation of anti-cancer agents for the new century? Anticancer Drug Des. *15*, 43–52.
4. Nettleton, D.E., Doyle, T.W., Krishnan, B., Matsumoto, G.K., and Clardy, J. (1985). Isolation and structure of rebeccamycin—A new antitumor antibiotic from *Nocardia aerocolonigenes*. Tetrahedron Lett. *26*, 4011–4014.
5. Bush, J.A., Long, B.H., Catino, J.J., Bradner, W.T., and Tomita, K. (1987). Production and biological activity of rebeccamycin, a novel antitumor agent. J. Antibiot. *40*, 668–678.
6. Prudhomme, M. (2000). Recent developments of rebeccamycin analogues as topoisomerase I inhibitors and antitumor agents. Curr. Med. Chem. *7*, 1189–1212.
7. van Pée, K.H. (2001). Microbial biosynthesis of halometabolites. Arch. Microbiol. *175*, 250–258.
8. Schupp, T., Engel, T.N., Bietenhader, J., Toupet, C., and Pospiech, A. (1997). Staurosporin biosynthesis gene clusters. WIPO Patent No. WO9708323.
9. Ohuchi, T., Ikeda-Araki, T.A., Watanabe-Sakamoto, A., Kojiri, K., Nagashima, M., Okanishi, M., and Suda, H. (2000). Cloning and expression of a gene encoding N-glycosyltransferase (*ngt*) from *Saccharothrix aerocolonigenes* ATCC39243. J. Antibiot. *53*, 393–403.
10. Richardson, M.A., Kuhstoss, S., Solenberg, P., Schauss, N.A., and Nagaraja, R.N. (1987). A new shuttle cosmid vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a *Streptomyces ambofaciens* library. Gene *61*, 231–241.
11. Zehner, S., Wynands, I., Schmid, C., Tölzer, S., Kuschel, M., Barth, G., and van Pée, K.H. (2000). Design of PCR-primers for the detection of halogenase-encoding regions in bacteria. 100th General Meeting of the American Society for Microbiology, Los Angeles, CA, USA.
12. Sancelme, M., Fabre, S., and Prudhomme, M. (1994). Antimicrobial activities of indolocarbazole and bis-indole protein kinase C inhibitors. J. Antibiot. *47*, 792–798.
13. Pearce, C.J., Doyle, T.W., Forenza, S., Lam, K.S., and Schroeder, D.R. (1988). The biosynthetic origins of rebeccamycin. J. Nat. Prod. *51*, 937–940.
14. Meksuriyen, D., and Cordell, G.A. (1988). Biosynthesis of staurosporine. 2. Incorporation of tryptophan. J. Nat. Prod. *51*, 893–899.
15. Lam, K.S., Forenza, S., Doyle, T.W., and Pearce, C.J. (1990). Identification of indolepyruvic acid as an intermediate of rebeccamycin biosynthesis. J. Ind. Microbiol. *6*, 291–294.
16. Jung, S.K., Mai, A., Iwamoto, M., Arizono, N., Fujimoto, D., Sakamaki, K., and Yonehara, S. (2000). Purification and cloning of an apoptosis-inducing protein derived from fish infected with *Anisakis simplex*, a causative nematode of human anisakiasis. J. Immunol. *165*, 1491–1497.
17. Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., et al. (2001). Complete genome sequence of *Caulobacter crescentus*. Proc. Natl. Acad. Sci. USA *98*, 4136–4141.
18. Lam, K.S., Schroeder, D.R., Veitch, J.M., Colson, K.L., Matson, J.A., Rose, W.C., Doyle, T.W., and Forenza, S. (2001). Production, isolation and structure determination of novel fluoroindolocarbazoles from *Saccharothrix aerocolonigenes* ATCC 39243. J. Antibiot. *54*, 1–9.
19. August, P.R., Grossman, T.H., Minor, C., Draper, M.P., MacNeil, I.A., Pemberton, J.M., Call, K.M., Holt, D., and Osburne, M.S. (2000). Sequence analysis and functional characterization of the violacein biosynthetic pathway from *Chromobacterium violaceum*. J. Mol. Microbiol. Biotechnol. *2*, 513–519.
20. Ruhul Momen, A.Z.M., and Hoshino, T. (2000). Biosynthesis of violacein: intact incorporation of the tryptophan molecule on the oxindole side, with intramolecular rearrangement of the indole ring on the 5-hydroxyindole side. Biosci. Biotechnol. Biochem. *64*, 539–549.
21. Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D.A. (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol. Microbiol. *21*, 77–96.
22. Yang, S.W., and Cordell, G.A. (1997). Origin of nitrogen in the indolocarbazole unit of staurosporine. J. Nat. Prod. *60*, 788–790.
23. Schreuder, H.A., van der Laan, J.M., Swarte, M.B., Kalk, K.H., Hol, W.G., and Drenth, J. (1992). Crystal structure of the reduced form of p-hydroxybenzoate hydroxylase refined at 2.3 Å resolution. Proteins *14*, 178–190.
24. Gatti, D.L., Palfey, B.A., Lah, M.S., Entsch, B., Massey, V., Ballou, D.P., and Ludwig, M.L. (1994). The mobile flavin of 4-OH benzoate hydroxylase. Science *266*, 110–114.
25. Armengaud, J., Timmis, K.N., and Wittich, R.M. (1999). A functional 4-hydroxysalicylate/hydroxyquinol degradative pathway gene cluster is linked to the initial dibenzo-p-dioxin pathway genes in *Sphingomonas* sp. Strain RW1. J. Bacteriol. *181*, 3452–3461.
26. Blanco, G., Pereda, A., Brian, P., Méndez, C., Chater, K.F., and Salas, J.A. (1993). A hydroxylase-like gene product contributes to synthesis of a polyketide spore pigment in *Streptomyces halstedii*. J. Bacteriol. *175*, 8043–8048.
27. Degtyarenko, K.N., and Archakov, A.I. (1993). Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS Lett. *332*, 1–8.
28. Stadler, R., and Zenk, M.H. (1993). The purification and characterization of a unique cytochrome P-450 enzyme from *Berberis stolonifera* plant cell cultures. J. Biol. Chem. *268*, 823–831.
29. Pelzer, S., Süssmuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G., and Wohlleben, W. (1999). Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. Antimicrob. Agents Chemother. *43*, 1565–1573.
30. Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. *326*, 929–942.
31. Mao, Y., Varoglu, M., and Sherman, D.H. (1999). Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564. Chem. Biol. *6*, 251–263.
32. Hammer, P.E., Burd, W., Hill, D.S., Ligon, J.M., and van Pée, K.H. (1999). Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. FEMS Microbiol. Lett. *180*, 39–44.
33. Lam, K.S., Schroeder, D.R., Veitch, J.M., Matson, J.A., and Forenza, S. (1991). Isolation of a bromo analog of rebeccamycin from *Saccharothrix aerocolonigenes*. J. Antibiot. *44*, 934–939.
34. Galán, B., Díaz, E., Prieto, M.A., and García, J.L. (2000). Functional analysis of the small component of the 4-hydroxyphenylacetate 3-monooxygenase of *Escherichia coli* W: a prototype of a new flavin:NAD(P)H reductase subfamily. J. Bacteriol. *182*, 627–636.
35. 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.
36. Bibb, M.J., Sherman, D.H., Omura, S., and Hopwood, D.A. (1994). Cloning, sequencing and deduced functions of a cluster of *Streptomyces* genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. Gene *142*, 31–39.
37. Kendrew, S.G., Harding, S.E., Hopwood, D.A., and Marsh, E.N.G. (1995). Identification of a flavin:NADH oxidoreductase involved in the biosynthesis of actinorhodin. J. Biol. Chem. *270*, 17339–17343.
38. Beltrametti, F., Marconi, A.M., Bestetti, G., Colombo, C., Galli, E., Ruzzi, M., and Zennaro, E. (1997). Sequencing and functional analysis of styrene catabolism genes from *Pseudomonas fluorescens* ST. Appl. Environ. Microbiol. *63*, 2232–2239.
39. Dibrov, P., and Fliegel, L. (1998). Comparative molecular analysis of Na+/H+ exchangers: a unified model for Na+/H+ antiport? FEBS Lett. *424*, 1–5.

40. Weitnauer, G., Muhlenweg, A., Trefzer, A., Hoffmeister, D., Sussmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U., and Bechthold, A. (2001). Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of *Streptomyces viridochromogenes* Tu57 and production of new antibiotics. Chem. Biol. *8*, 569–581.
41. Chiu, H.T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T., and Khosla, C. (2001). Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc. Natl. Acad. Sci. USA *98*, 8548–8553.
42. Marger, M.D., and Saier, M.H., Jr. (1993). A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem. Sci. *18*, 13–20.
43. Piel, J., Hoang, K., and Moore, B.S. (2000). Natural metabolic diversity encoded by the enterocin biosynthesis gene cluster. J. Am. Chem. Soc. *122*, 5415–5416.
44. De Schrijver, A., and De Mot, R. (1999). A subfamily of MalT-related ATP-dependent regulators in the LuxR family. Microbiology *145*, 1287–1288.
45. 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.
46. Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., et al. (2000). Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res. *28*, 4317–4331.
47. Floriano, B., and Bibb, M.J. (1996). afsR is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). Mol. Microbiol. *21*, 385–396.
48. Lessard, I.A., Pratt, S.D., McCafferty, D.G., Bussiere, D.E., Hutchins, C., Wanner, B.L., Katz, L., and Walsh, C.T. (1998). Homologs of the vancomycin resistance D-Ala-D-Ala dipeptidase VanX in *Streptomyces toyocaensis*, *Escherichia coli* and *Synechocystis*: attributes of catalytic efficiency, stereoselectivity and regulation with implications for function. Chem. Biol. *5*, 489–504.
49. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. *10*, 1–6.
50. von der Haar, B., Walter, S., Schwapenheuer, S., and Schrempf, H. (1997). A novel fusidic acid resistance gene from *Streptomyces lividans* 66 encodes a highly specific esterase. Microbiology *143*, 867–874.
51. Sulavik, M.C., Gambino, L.F., and Miller, P.F. (1995). The MarR repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. Mol. Med. *1*, 436–446.
52. 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.
53. Schmitt-John, T., and Engels, J.W. (1992). Promoter constructions for efficient secretion expression in *Streptomyces lividans*. Appl. Microbiol. Biotechnol. *36*, 493–498.
54. Pindur, U., Kim, Y.S., and Mehrabani, F. (1999). Advances in indolo[2,3-a]carbazole chemistry: design and synthesis of protein kinase C and topoisomerase I inhibitors. Curr. Med. Chem. *6*, 29–69.
55. Horton, P.A., Longley, R.E., McConnell, O.J., and Ballas, L.M. (1994). Staurosporine aglycone (K252-c) and arcyriaflavin A from the marine ascidian, *Eudistoma* sp. Experientia *50*, 843–845.
56. Slater, M.J., Cockerill, S., Baxter, R., Bonser, R.W., Gohil, K., Gowrie, C., Robinson, J.E., Littler, E., Parry, N., Randall, R., et al. (1999). Indolocarbazoles: potent, selective inhibitors of human cytomegalovirus replication. Bioorg. Med. Chem. *7*, 1067–1074.
57. Slater, M.J., Baxter, R., Bonser, R.W., Cockerill, S., Gohil, K., Parry, N., Robinson, E., Randall, R., Yeates, C., Snowden, W., et al. (2001). Synthesis of N-alkyl substituted indolocarbazoles as potent inhibitors of human cytomegalovirus replication. Bioorg. Med. Chem. Lett. *11*, 1993–1995.
58. Chater, K.F., and Wilde, L.C. (1980). *Streptomyces albus* G mutants defective in the SalGI restriction-modification system. J. Gen. Microbiol. *116*, 323–334.
59. Bullock, W.O., Fernández, J.M., and Short, J.M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with β-galactosidase selection. Biotechniques *5*, 376.
60. Murray, N.E., Brammar, W.J., and Murray, K. (1977). Lamboid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. *150*, 53–61.
61. Vara, J., Lewandowska-Skarbek, M., Wang, Y.G., Donadio, S., and Hutchinson, C.R. (1989). Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). J. Bacteriol. *171*, 5872–5881.
62. Quirós, L.M., Aguirrezabalaga, I., Olano, C., Méndez, C., and Salas, J.A. (1998). Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*. Mol. Microbiol. *28*, 1177–1185.
63. Fernández, E., Weissbach, U., Sánchez Reillo, C., Braña, A.F., Méndez, C., Rohr, J., and Salas, J.A. (1998). Identification of two genes from *Streptomyces argillaceus* encoding two glycosyltransferases involved in the transfer of a disaccharide during the biosynthesis of the antitumor drug mithramycin. J. Bacteriol. *180*, 4929–4937.
64. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical *Streptomyces* genetics. (Norwich, UK: The John Innes Foundation).
65. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
66. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.
67. Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. *12*, 387–395.
68. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Accession Numbers

The complete DNA sequence of the insert contained in cosmid 14E8 has been deposited at the EMBL Nucleotide Sequence Database with accession number AJ414559.