Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699

Paul R August[1*†], Li Tang[2*], Yeo J Yoon[2], Sandra Ning[1], Rolf Müller[1‡], Tin-Wein Yu[1], Matthew Taylor[1], Dietmar Hoffmann[1‡], Chun-Gyu Kim[1§], Xiaohong Zhang[1], C Richard Hutchinson[2] and Heinz G Floss[1]

Background: The ansamycin class of antibiotics are produced by various Actinomycetes. Their carbon framework arises from the polyketide pathway via a polyketide synthase (PKS) that uses an unusual starter unit. Rifamycin (rif), produced by *Amycolatopsis mediterranei*, is the archetype ansamycin and it is medically important. Although its basic precursors (3-amino-5-hydroxy benzoic acid AHBA, and acetic and propionic acids) had been established, and several biosynthetic intermediates had been identified, very little was known about the origin of AHBA nor had the PKS and the various genes and enzymes that modify the initial intermediate been characterized.

Results: A set of 34 genes clustered around the *rifK* gene encoding AHBA synthase were defined by sequencing all but 5 kilobases (kb) of a 95 kb contiguous region of DNA from *A. mediterranei*. The involvement of some of the genes in the biosynthesis of rifamycin B was examined. At least five genes were shown to be essential for the synthesis of AHBA, five genes were determined to encode the modular type I PKS that uses AHBA as the starter unit, and 20 or more genes appear to govern modification of the polyketide-derived framework, and rifamycin resistance and export. Putative regulatory genes were also identified. Disruption of the PKS genes at the end of *rifA* abolished rifamycin B production and resulted in the formation of P8/1-OG, a known shunt product of rifamycin biosynthesis, whereas disruption of the *orf6* and *orf9* genes, which may encode deoxysugar biosynthesis enzymes, had no apparent effect.

Conclusions: Rifamycin production in *A. mediterranei* is governed by a single gene cluster consisting of structural, resistance and export, and regulatory genes. The genes characterized here could be modified to produce novel forms of the rifamycins that may be effective against rifamycin-resistant microorganisms.

Addresses: [1]Department of Chemistry, Box 351700, University of Washington, Seattle, WA 98195-1700, USA. [2]School of Pharmacy, University of Wisconsin, 425 North Charter Street, Madison, WI53706, USA.

Present addresses: [†]MDS-Panlabs, Inc., 11804 North Creek Parkway, S. Bothell, WA 98011-8805, USA. [‡]Institut fuer Pharmazeutische Biologie der Universitaet, Nußallee 6, D-53115 Bonn 1, Germany. [§]Department of Chemical Engineering B-419, Inje University, Obang-dong 607, Kimhae, Kyongnam 621-749, Korea.

*Contributed equally to this work.

Correspondence: C Richard Hutchinson and Heinz G Floss
e-mail: crhutchi@facstaff.wisc.edu and floss@chem.washington.edu

Key words: 3-amino-5-hydroxy benzoic acid, combinatorial biosynthesis, gene disruption, modular type I polyketide synthase, shikimate pathway

Received: **29 December 1997**
Revisions requested: **14 January 1998**
Revisions received: **19 January 1998**
Accepted: **20 January 1998**

Published: **13 February 1998**

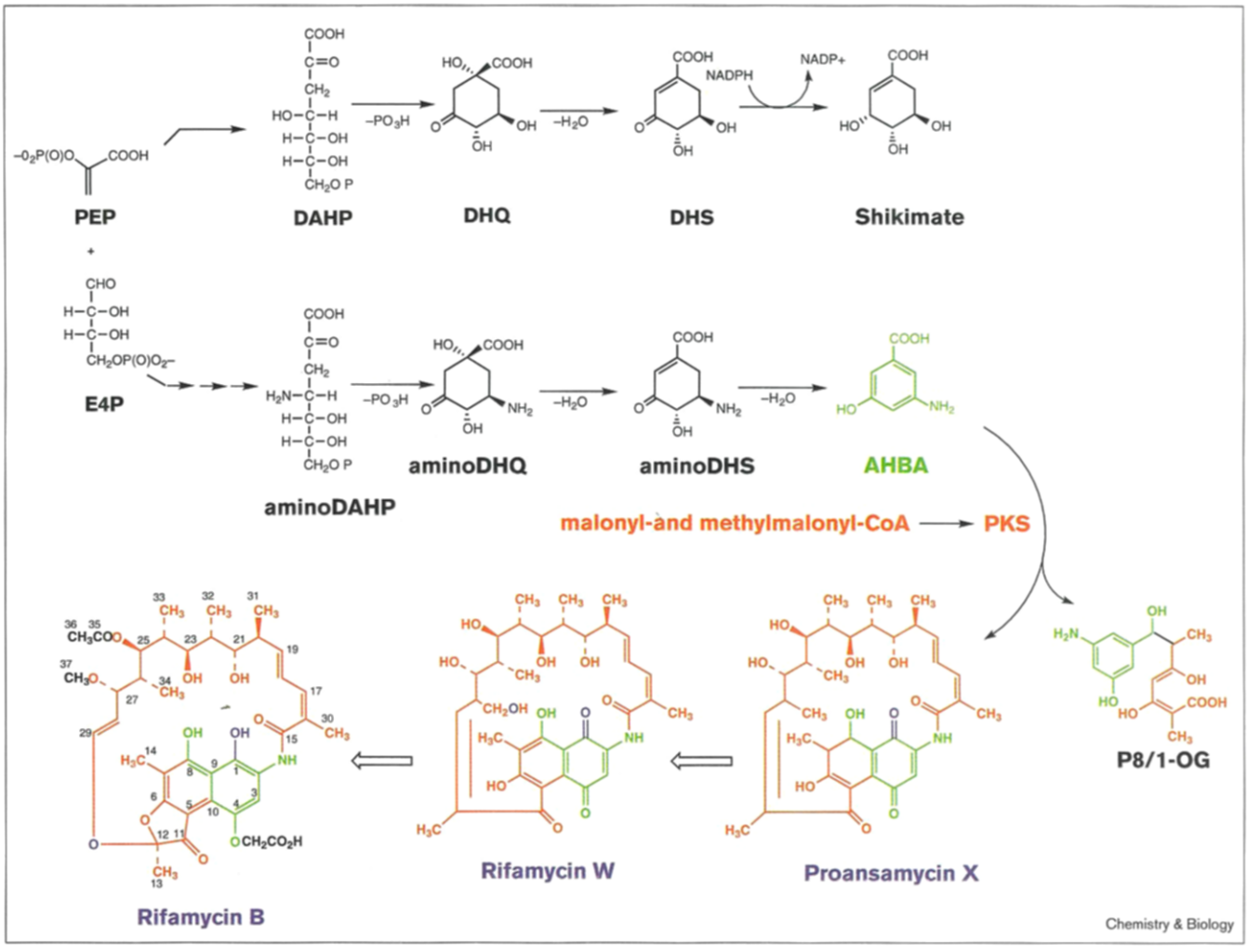
Chemistry & Biology February 1998, **5**:69–79
http://biomednet.com/elecref/1074552100500069

Introduction

The ansamycins are a prominent class of antibiotics produced by various Actinomycetes. Structurally they are characterized by a macrocycle composed of a benzenic or naphthalenic chromophore bridged by an aliphatic ansa chain that terminates at the chromophore in an amide linkage. Structural features that distinguish the various ansamycin antibiotics include the macrocycle chain length, chromophore type and degree of post-synthetic modification. Examples of ansamycins include the ansamitocins, geldanamycins, streptovaricins and naphthomycins [1]. Rifamycin B (Figure 1) [2,3], produced by *Amycolatopsis mediterranei*, is one of the most notable members of the ansamycin family. It has been used clinically in a synthetically modified form called rifampicin since the mid 1960s, and it is still one of the first-line therapies effective in the treatment of tuberculosis and other mycobacterial infections [4].

Feeding experiments using labeled precursors have shown that the ansa chain of rifamycin is biosynthesized from propionate and acetate units [5,6] in a manner that suggests assembly on a multifunctional type I polyketide synthase (PKS) [7]. The starter unit for this polyketide assembly, called the mC_7N unit, is part of the chromophore and is derived from 3-amino-5-hydroxybenzoic acid (AHBA) [8]. AHBA, according to recent cell-free experiments, arises from a new metabolic route, the aminoshikimate pathway, that parallels the early steps of the shikimate pathway (Figure 1) [9]. The aminoshikimate pathway leads to the

Figure 1



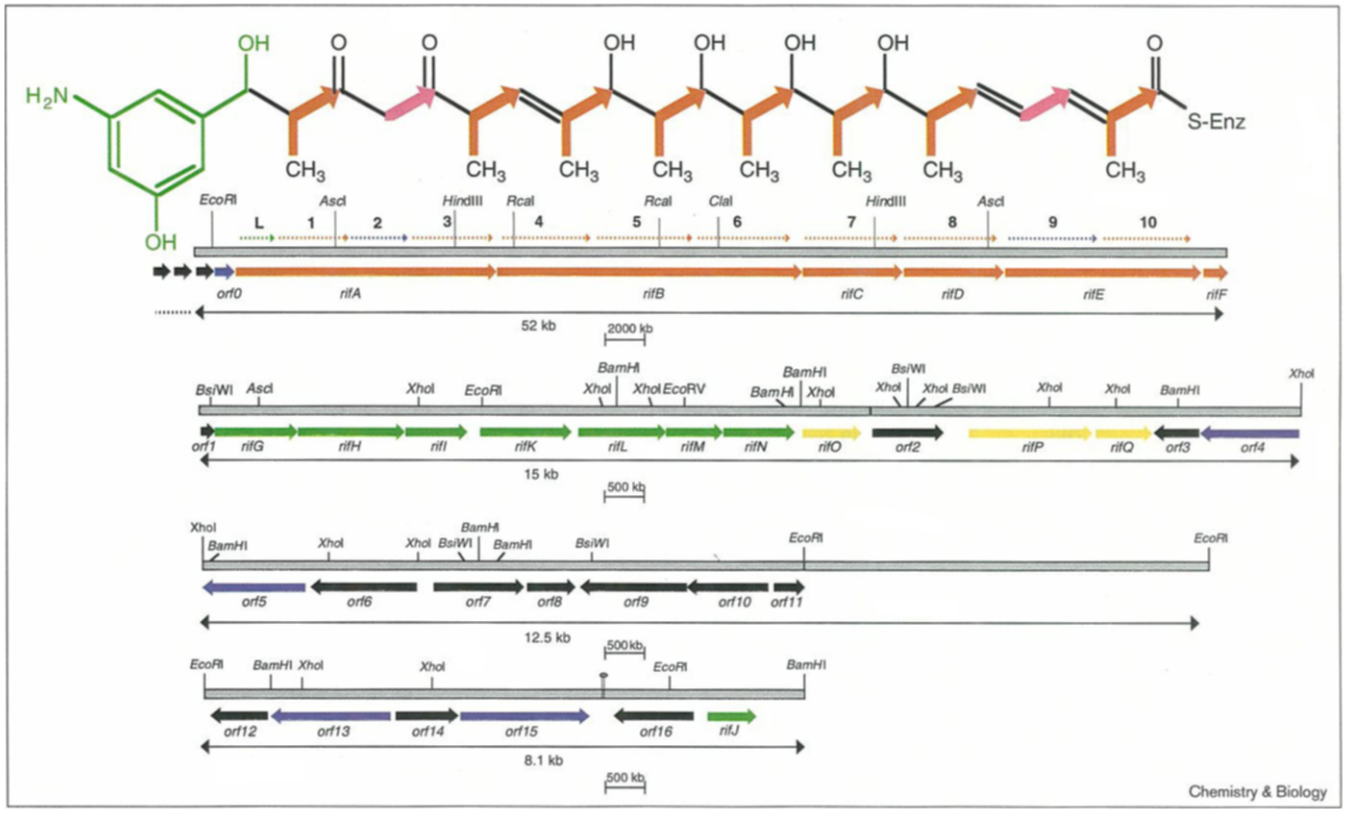
Comparison of the early shikimate pathway with the alternate aminoshikimate pathway as proposed by Kim *et al.* [9] and its connection with rifamycin B biosynthesis. Color coding indicates the portions of rifamycin B resulting from AHBA (green), malonyl-CoA and methylmalonyl-CoA (red), molecular oxygen (blue), and other sources (black). DAHP, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate; DHQ, dehydroquinate; DHS, dehydroshikimic acid; PEP, phosphoenolpyruvate; E4P, erythrose-4-phosphate.

formation of 5-deoxy-5-amino-3-dehydroshikimic acid (aminoDHS) which is then aromatized to AHBA by a novel pyridoxal phosphate enzyme, AHBA synthase. Cloning of the gene encoding this enzyme, *rifK*, and demonstration that this gene is essential for rifamycin production [10] then provided a probe that allowed the isolation of the rifamycin biosynthetic gene cluster. The sequence and functional analysis of this cluster is the subject of the present report.

Results

Kim *et al.* [10] established that the AHBA synthase gene product is essential for the biosynthesis of the mC_7N unit of rifamycin. Given the precedent for the clustering of related biosynthetic genes in Actinomycetes [11], we proceeded to isolate and sequence DNA flanking the *rifK* gene in order to identify the rifamycin biosynthetic gene cluster (Figure 2). Approximately 90 kilobases (kb) of DNA were sequenced: one region was obtained from the primary cosmid clones pFKN108 and pFKN105; when pFKN108 containing part of the PKS genes was found to be prone to deletions and rearrangements, the PKS genes were then obtained as several overlapping clones from DNA mini-libraries made in plasmid vectors. Computer-assisted analysis of the nucleotide sequence revealed open reading frames (ORFs) homologous to genes involved in the shikimate pathway of plants and bacteria, various putative modifying and regulatory genes as well as five large ORFs coding for a modular type I PKS (Figure 2). Genes sequenced and the actual or putative functions of some of their gene products are listed in Table 1.

Figure 2



Genetic organization of the rifamycin biosynthetic gene cluster. For comparison, the acyl chain of the rifamycin polyketide is placed over the respective modules (the loading domain L and modules 1–10) believed to be involved in the incorporation of the appropriate extender units, indicated by thick red (for propionate-derived) or pink (for acetate-derived) arrows, with the AHBA starter unit shown in green. Groupings of the restriction map are based upon biosynthetic functions (see text). Genes are colored according to proven or likely functions: AHBA biosynthesis (green), PKS (red), oxygenation by CYP450 enzymes (blue) and regulation or resistance (yellow).

Nucleotide sequence of the rifamycin mC_7N unit biosynthetic genes

Proximate to the previously characterized *rifK* gene [10], there are a number of genes that form a sub-cluster and potentially an operon that appears to be responsible for the biosynthesis of the mC_7N polyketide starter unit (Figure 2). A strong transcription terminator sequence immediately follows *rifO*, which therefore seems to be the last gene in a transcription unit starting at least with *rifK* and possibly earlier. The products of several genes in this sub-cluster are highly homologous to proteins involved in the shikimate biosynthetic and the quinate utilization pathways of bacteria, plants and fungi. RifG shows a high degree of similarity to *Escherichia coli* and eukaryotic dehydroquinate synthases. 3-Dehydroquinate synthase catalyzes the cyclization of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) into 3-dehydroquinate, the second step in the shikimate pathway [12]. RifG has the highest degree of similarity to a hypothetical ORF from *Actinoplanes teichomyceticus* followed by significant identity with AroB of *E. coli*, *Salmonella enterica* serovar *typhimurium*, and AroM, the pentafunctional polypeptide of *Emericella nidulans* [13–15].

The gene product of *rifH* shows a high degree of similarity to the DAHP synthases (DAHPS) of plants and a DAHPS involved in phenazine biosynthesis. A similarity of 61.8%, 60.0% and 59.6% and identity of 33.8%, 37.4% and 37.6% was observed with PhzC of *Pseudomonas fluorescens*, AroG of *Lycopersicon esculentum* and AroG of *Arabidopsis thaliana*, respectively. The proteins encoded by *rifG* and *rifH* are proposed to be involved in the formation of aminoDAHP and its further cyclization to aminoDHQ (Figure 1). The *rifI* gene appears to encode a shikimate or quinate dehydrogenase. It shows 63.5% similarity and 41.6% identity with the *aroE* gene product of *Pseudomonas aeruginosa* which catalyzes the reversible reduction of 3-dehydroshikimate to shikimate. The presence of the *rifI* gene is surprising, as this step or the analogous interconversion of aminoDHS and aminoshikimate should not be required for AHBA formation. The product of *rifK* is the previously described AHBA

Table 1

Deduced functions of ORFs in the rifamycin biosynthetic gene cluster.

Polypeptide (gene)	Amino acids	Sequence similarity	Proposed function
Orf0(*orf0*)	397	ChoP *Streptomyces* sp. [55]	Cytochrome P450 monooxygenase
RifA (*rifA*)	4,735	EryA *S. erythraea* [22]	PKS domains:
Loading domain			CoA ligase ACP
Module 1			KS AT DH* KR ACP
Module 2			KS AT^m ACP
Module 3			KS AT KR* ACP
RifB (*rifB*)	5,060	EryA *S. erythraea*	PKS domains:
Module 4			KS AT DH KR ACP
Module 5			KS AT DH* KR ACP
Module 6			KS AT *DH* KR ACP
RifC (*rifC*)	1,763	EryA *S. erythraea*	PKS domains:
Module 7			KS AT *DH* KR ACP
RifD (*rifD*)	1728	EryA *S. erythraea*	PKS domains:
Module 8			KS AT *DH* KR ACP
RifE (*rifE*)	3,413	EryA *S. erythraea*	PKS domains:
Module 9			KS AT^m DH KR ACP
Module 10			KS AT DH KR ACP
RifF (*rifF*)	260	E.C. 2.3.1.5 *G. gallus* [30]	Amide synthase (N-acetyl transferase)
Orf1 (*orf1*)	62	–	Unknown
RifG (*rifG*)	336	AroB *E. coli K12* [14]	Aminodehydroquinate synthase
RifH (*rifH*)	441	AroG *S. lycopersicum* [56]	AminoDAHP synthase
RifI (*rifI*)	263	AroE *Synechocystis sp.* [41]	Aminoquinate dehydrogenase
RifK (*rifK*)	388	AHBAS *S. collinus* (Genbank Z54208) [10]	AHBA synthase
RifL (*rifL*)	360	Pur10 *S. ablboniger* [16]	Oxidoreductase
RifM (*rifM*)	232	CbbzP *A. eutrophus* [18]	Phosphatase
RifN (*rifN*)	303	XylR *Synechocystis sp.* [41]	Kinase
RifO (*rifO*)	261	LmbE *M. leprae* (Genbank 699274)	Putatively regulation
Orf2 (*orf2*)	310	Orf8 *A. eutrophus* [57]	Esterase
RifP (*rifP*)	525	Ptr *S. pristinaespiralis* [31]	(antibiotic efflux/resistance)
RifQ (*rifQ*)	242	actII-Orf1 *Streptomyces coelicolor* [32]	Transcriptional repressor
Orf3 (*orf3*)	166	–	Unknown
Orf4 (*orf4*)	403	SoyC *S. griseus* [58]	Cytochrome P450 monooxygenase
Orf5 (*orf5*)	421	OrfA *S. thermotolerans* [59]	Cytochrome P450 monooxygenase
Orf6 (*orf6*)	435	RfbH *Salmonella enterica* [60]	dNTP-hexose dehydratase
Orf7 (*orf7*)	381	Orf5 *S. violaceoruber* [61]	dNTP-hexose glycosyl transferase
Orf8 (*orf8*)	206	StrM *S. griseus* [62]	dNTP-hexose 3,5 epimerase
Orf9 (*orf9*)	432	YokM *B. subtilis* (Genbank AF006665)	Aminotransferase
Orf10 (*orf10*)	330	RdmF *S. purpurascens* [63]	Oxidoreductase
Orf11 (*orf11*)	–	Mer *M. thermoautotrophicum* [64]	Reductase
Orf12 (*orf12*)	252	Orf5 *S. fradiae* [43]	Thioesterase
Orf13 (*orf13*)	500	EryF *S. erythraea* [65]	Cytochrome P450 monooxygenase
Orf14 (*orf14*)	272	Erg6 *Synechocystis sp.* [41]	Methyltransferase
Orf15 (*orf15*)	590	Orf2 *R. fascians* [66]	Transketolase
Orf16 (*orf16*)	351	SubC *S.* griseolus [67]	Cytochrome P450 monooxygenase
RifJ (*rifJ*)	163	AroD *A. pleuropneumoniae* [68]	Aminodehydroquinate dehydratase

Potential PKS enzymatic activities are abbreviated as follows: ACP, acyl carrier protein; KS, β-ketoacyl-ACP synthase; AT^m, acyltransferase utilizing malonyl-CoA as an extender unit (the other ATs use methylmalonyl-CoA); KR, β-ketoacyl-ACP reductase; DH, β-hydroxy-acyl-thioester dehydratase. Enzymatic activities listed for the rif PKS proteins are colinear with the sequence of each ORF as listed from left to right and top to bottom. Asterisked and italicized domains indicate that the assigned enzymatic activity is possibly non-functional or that an apparent functional domain does not correspond to the predicted final polyketide product respectively (see text).

synthase involved in the conversion of aminoDHS into AHBA [10].

The enzyme encoded by *rifL* shows a high degree of similarity at the amino terminus to the product of *purl0* which is involved in the biosynthesis of the aminoglycoside antibiotic puromycin [16]. The gene product of *purl0*, which has been implicated in the oxidation of a hydroxyl to a carbonyl group, in turn is similar to the glucose–fructose oxidoreductase of *Zymomonas mobilis*. This enzyme oxidizes glucose to gluconolactone and reduces fructose to sorbitol [17]. RifL also shows a high degree of amino acid identity to Orf10 of the rifamycin biosynthetic gene cluster, which is adjacent to a gene that codes for a putative pyridoxal phosphate (PLP)-dependent transaminase. Significantly, *rifL* homologs are located immediately downstream of all the AHBA synthase genes so far isolated: two AHBA synthase genes from *Streptomyces collinus*, presumably involved in naphthomycin and ansatrienin biosynthesis, one from *Actinosynnema pretiosum* (E. Leistner, unpublished observations) and one from the mitomycin biosynthetic gene cluster of *Streptomyces lavendulae* (D.H. Sherman, unpublished observations). The juxtaposition of these oxidoreductase genes with genes coding for PLP-dependent proteins may be an indication of coordinated catalytic activity of the products of these adjacent genes.

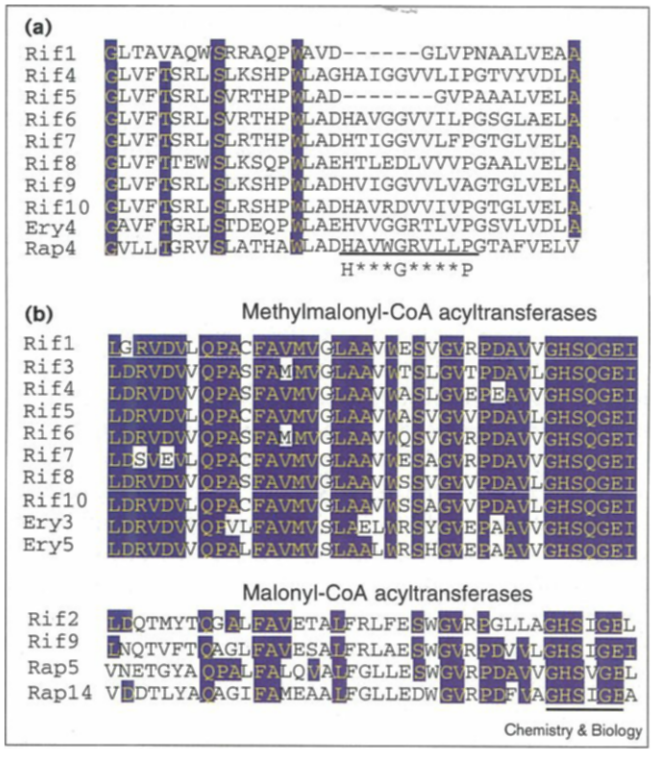
The product of *rifM* is similar to a class of phosphatases in the CBBY family that appear to be involved in glycolate oxidation, specifically converting 2-phosphoglycolate to glycolate in the Calvin carbon reduction cycle [18]. A phosphoglycolate phosphatase gene, *gph*, has been found in *E. coli*; the enzyme that is encoded by *gph* catalyzes the dephosphorylation of 2-phosphoglycolate but its physiological role is unclear [19]. Downstream of *rifM* is the gene designated *rifN* which codes for a protein that has sequence similarity to a glucose kinase from *Streptomyces coelicolor* A3(2) that is involved in glucose catabolite repression [20]. At the amino terminus of RifN is the amino acid sequence Asp–Val–Gly–Gly–Thr (DVGGT) which is essentially identical to the motif found in the *S. coelicolor* enzyme and is believed to constitute part of the ATP-binding site in mammalian, yeast and bacterial hexokinases [20]. Rifamycin biosynthesis, however, is not subject to catabolite repression by glucose (R.M., M.T. and H.G.F., unpublished observations) and the observed homology for the gene products of *rifM* and *rifN* may not be indicative of their functions. Both *rifM* and *rifN*, as well as *rifL*, were shown by gene inactivation experiments to be required for AHBA formation (R.M. and H.G.F., unpublished observations).

Notably absent from the cluster of genes surrounding *rifK* is a gene that has homology to either type I or type II dehydroquinate (DHQ) dehydratases, a candidate to code for an aminoDHQ dehydratase. As earlier enzymatic work had shown that the primary metabolic DHQ dehydratase of *A. mediterranei* does not accept aminoDHQ as a substrate [21], a separate enzyme catalyzing this dehydration must exist. A candidate gene, *rifJ*, encoding a protein that has homology to type II DHQ dehydratases was finally located about 30 kb downstream from *rifK*. Despite its remote location, this gene was shown, by insertional inactivation, to be involved in AHBA formation (X.Z. and H.G.F., unpublished observations).

Nucleotide sequence of the rifamycin polyketide synthase genes

Upstream of the genes for AHBA synthesis lies a large region encoding the five PKS genes (Figure 2), as initially deduced from the results of Southern analysis with segments of the *eryA* PKS genes [22]. The five proteins encoded by *rifA–rifE* contain ten separate modules, typical of type I PKSs [7], the active sites of which largely resemble those found in fatty acid synthases. RifA contains at the amino terminus the loading domain for AHBA, which consists of an acyl-CoA ligase linked to an acyl carrier protein (ACP), and modules 1–3; RifB contains modules 4–6, RifC contains module 7, RifD contains module 8 and RifE contains modules 9 and 10. Module 2, which only has the minimally essential PKS domains (β-ketoacyl-ACP reductase, acyltransferase and ACP), lacks the large block of presumed structural protein that lies between the acyltransferase and the β-ketoacyl-ACP reductase domains of all the other modules. The β-hydroxyacyl-thioester dehydratase (DH) domains do not all correspond to the pattern expected for the proposed structure of the PKS product drawn in Figure 2 above the genetic map: modules 6 and 7 of RifB and RifC appear to contain functional DH domains (Figure 3a), yet neither the hypothetical product of the PKS nor rifamycin W and rifamycin B (Figure 1) have double bonds at the positions corresponding to the segments made by modules 6 and 7 of the PKS; DH activity in modules 9 and 10 of RifE is predicted to be required for the biosynthesis of these three compounds, but the DH domains of these two modules do not display a close alignment with the active sites of the functional DH motifs of rapamycin [23] and erythromycin [22] (Figure 3a). This is also true for the DH domain of module 8 which is not likely to be functional because it has large deletions in regions at the carboxyl and amino termini that are highly conserved in the functional DH domains. In contrast, the DH domains in modules 1 and 5 appear to be inactive and the one in module 4 is active, consistent with biosynthetic expectations (Figure 3a). All other structural features of the RifA–RifE proteins are consistent with existing precedents [22–25], including the clear differentiation of the acyltransferase domains specific for the malonyl-CoA or methylmalonyl-CoA substrates [23,24] (Figure 3b), the seemingly inactive β-ketoacyl-ACP reductase domain in module 3, and the presence of single module proteins [25,26]. Finally, the juxtaposition of the

Figure 3



A comparison of active-site sequences from various PKSs.
(a) Comparison of the active-site sequences of DH domains from the eight rifamycin (rif) PKS modules (shown in Figure 2) with DH domains with those from two modules of the erythromycin (ery4) [22] and rapamycin (rap4) [23,24] PKSs. The consensus sequence for the probable active site amino acids is shown beneath the rap4 sequence. **(b)** Comparison of the active-site sequences of the ten acyltransferase domains from the rif PKS with representatives from the erythromycin (ery3 and ery5) and rapamycin (rap5 and rap14) PKSs. The active-site region is underlined.

orf0 CYP450 and *rifA–rifE* genes may indicate that the Orf0 protein hydroxylates C-1 before or during formation of the naphthoquinone nucleus of proansamycin X (Figure 1). A function for three other *orf*s sequenced upstream of *orf0* could not be deduced by sequence comparisons with the available databases.

Several mutants were constructed by disrupting the PKS genes at the end of different modules and each strain failed to produce rifamycin B and its known precursors (L.T., Y.J.Y. and C.R.H., unpublished observations). The mutant truncated at the end of RifA produced the triketide P8/1-OG as the major product (Figure 1), first identified as a shunt product of a rifamycin non-producing strain [27]. The formation of P8/1-OG supports the enzymatic functions predicted for RifA and, together with the results of oxygen labeling studies conducted by Rickards and coworkers [28], suggests that proansamycin X (Figure 1) could be an early precursor of rifamycin instead of proansamycin A [29]. Mutants with disruptions in other rifamycin PKS genes produce tetraketide, pentaketide, hexaketide and octaketide intermediates of the polyketide assembly (L.T. and C.R.H., unpublished observations).

Located between the modular polyketide synthase genes and the AHBA cluster is an ORF designated *rifF*. The gene product of *rifF* shows a high degree of similarity to arylamine N-acetyl-transferases [30]. These enzymes have been shown to detoxify arylamines by acetylating the amino group with acetyl-CoA as acetyl donor. The acetylated product is subsequently degraded via catabolic pathways. On the basis of the analogous chemistry involved, RifF is proposed to catalyze the release of the completed polyketide from the PKS by intramolecular amide formation to generate the macrocyclic lactam structure.

Resistance and regulatory genes

Without exception, all antibiotic biosynthetic gene clusters in Actinomycetes contain genes that confer resistance to the product on the producing organism. Some of the resistance mechanisms involve modification of the secondary metabolite or the target site (this is the case for the *A. mediterranei* S699 *rpoB* gene product, a RNA polymerase subunit; T.L. and C.R.H., unpublished observations), whereas others operate by removing the secondary metabolite from the intracellular target by transporting it out of the cell. The gene product of *rifP* appears to be involved in rifamycin resistance in *A. mediterranei* in the latter fashion. RifP is 66.1% identical and 81.2% similar to an integral membrane protein, Ptr, from *Streptomyces pristinaesprialis* that confers resistance to pristinamycins I and II and rifampicin in *Streptomyces lividans* [31]. Ptr employs the transmembrane electrochemical proton gradient in an antiporter mode to drive the efflux of intracellular antibiotics. It is reasonable to suggest, on the basis of the high degree of amino acid conservation, that RifP operates similarly to export rifamycin from *A. mediterranei*.

Down regulation of secondary metabolic biosynthetic gene expression is essential for well-balanced primary metabolism (growth). One gene that may be involved in transcriptional regulation of the rifamycin biosynthetic gene cluster is *rifQ*. RifQ shows a high degree of similarity to a transcriptional repressor encoded by the *actII–orf1* gene of *S. coelicolor* [32]. Experimental evidence indicates that *actII*–Orf1 represses the expression of the *actII–orf2* actinorhodin efflux protein gene until stationary phase is reached [33]. Presumably RifQ operates similarly in autoregulating the expression of a *rifPQ* operon.

Computer-assisted database comparisons failed to reveal any homology to the small ORF designated *orf1* downstream of *rifF*. The diminutive length of *orf1* suggests that it could encode a transcriptional regulatory protein, but there is no evidence for this in the sequence data.

RifO, which has a high degree of similarity to the *lmbE* gene product from *Mycobacterium leprae* and *Streptomyces lincolnensis*, may also be involved in regulation of rifamycin biosynthesis. The product of the *lmbE* gene in *S. lincolnensis* has been suggested, without experimental evidence, to be involved in the biosynthesis of the propyl-proline precursor of lincomycin [34]. On the other hand, RifO is similar to 2′,3′-cyclic nucleotide 2′-phosphodiesterases. B factor (3′-(1-butylphosphoryl)adenosine) has previously been shown to stimulate rifamycin B synthesis at extremely low concentrations [35,36]. It is therefore conceivable that RifO controls the levels of B factor.

Rifamycin-modifying genes

The biosynthesis of rifamycins proceeds through two clearly distinct stages: assembly of a polyketide from the AHBA starter unit and malonyl-CoA or methylmalonyl-CoA chain extension units to give proansamycin X, followed by elaboration of proansamycin X into first rifamycin W, then rifamycin B and, potentially, other rifamycins. The cluster contains a number of genes that appear to be involved in this further elaboration of the rifamycin structures from proansamycin X. In addition to *orf0*, *orfs 4*, *5*, *13* and *16* also code for proteins highly homologous to CYP450 monooxygenases that may be involved in some of the oxidative modification steps that occur at C-34a and in the carbon chain cleavage between C-12 and C-29 or naphthoquinone formation. The *orf11* gene product shows similarity to the coenzyme F420-dependent N^5N^{10}-methylenetetrahydro-methanopterin reductase (MER) from *Methanobacterium thermoautotrophicum*, a member of the kingdom Archea [37]. This protein is also similar to the products of a *Mycobacterium tuberculosis* gene of unknown function and the *lmbY* gene of *S. lincolnensis* involved in lincomycin biosynthesis [34]. *Orf2* codes for another protein that is highly homologous to a *Mycobacterium* gene product of unknown function [37]. This protein and Orf2 share a high degree of identity with esterases, particularly with lipases.

Curiously, located upstream of *orf5* is a set of genes (*orfs 6*, *7*, *8*, *9* and *10*) that appear to encode the formation of an amino-deoxyhexose nucleotide. It is predicted that Orf6 catalyzes reductive 3-deoxygenation of a deoxyhexose nucleotide, Orf8 catalyzes epimerization from the D to the L series, Orf9 catalyzes amination of a keto group to the aminosugar and Orf7 catalyzes transfer of the sugar moiety to an aglycone. *Orf10* encodes an oxidoreductase similar to RifL and like *rifL* is associated with a gene encoding a PLP/PMP-containing enzyme (PMP, pyroxamine phosphate). There is no obvious role for Orfs 6, 7, 8, 9 and 10 in the biosynthesis of rifamycin B which is not glycosylated. Consistent with this, inactivation of *orf6* and *orf9* had no effect on rifamycin B production (L. Heide, R.M. and H.G.F., unpublished observations). The isolation of glycosylated rifamycin derivatives, the tolypomycins, has been reported from another organism, *Streptomyces tolypophorus*, that also produces rifamycin B. The sugar moiety attached to C-4 of the tolypomycins, a 4-amino-2,3-dideoxy-L-rhamnose [38], could indeed require enzymes like the ones encoded by *orfs 6–10* for the late stages of its biosynthesis. Functionally, these genes may be silent in *A. mediterranei* S699 because no glycosylated rifamycin has so far been isolated from this culture, or they may be expressed at such low levels that their biosynthetic product has escaped detection.

Upstream of *orf13* there seems to be a transcriptional unit comprised of two genes (*orf14* and *orf15*) that encode proteins with high degrees of homology to methyl transferases and transketolases, respectively. The *orf14* gene product shows 34% identity and 55% similarity to EryG that catalyzes the methylation of the 3-OH group of the mycarosyl moiety in erythromycin biosynthesis [39]. Orf14 is therefore proposed to be responsible for the *O*-methylation at C-27 of rifamycin W during rifamycin B biosynthesis. Orf15 strongly resembles transketolases from different organisms (52% identity and 62% similarity to *cbbT* of *Xanthobacter flavus* [40], 42% identity and 55% similarity to *tkt*A of *Synechocystis sp.* [41]). A second copy of this primary metabolic gene from the pentose phosphate pathway could serve to increase the production levels of erythrose-4-phosphate (E4P) and subsequently also of AHBA and rifamycin B. Frost and Draths [42] have shown in *E. coli* that overexpression of transketolases can increase carbon flow through the common pathway to E4P.

Located between *orf15* and *orf16* are two inverted repeats; these two stem loops could function as terminators for the transcription of *orf14* and *orf15*, and *orf16*, respectively. Finally, downstream of *orf13* we find an almost complete ORF (*orf12*) that extends into an unsequenced region and shows strong homology to various thioesterase gene products, such as Orf5 from the tylosin gene cluster of *S. fradiae* [43] and the thioesterase domain in the last module of the erythromycin PKS [22]. The function of the thioesterase domain in the erythromycin PKS evidently is the release of the completed polyketide from the enzyme as the macrocyclic lactone, but thioesterase genes have also been found separate and distal from PKS genes in other antibiotic biosynthesis gene clusters, such as the tylosin gene cluster, and their functions are at present not known.

Discussion

In the work reported here we have sequenced a region of DNA from *A. mediterranei* spanning about 95 kb that is involved in the biosynthesis of the rifamycin antibiotics. The sequence is contiguous except for a 5 kb unsequenced region on one side of the cluster and includes the entire sequence of a type I PKS with the correct number of modules required for the assembly of the rifamycin polyketide backbone from two acetate and

eight propionate units. Inactivation of genes on either fringe of the region suggests that the rifamycin biosynthetic gene cluster extends over the entire 95 kb region, and possibly even beyond this region.

The cluster includes at least seven, possibly eight, genes involved in the formation of the polyketide starter unit AHBA. This number is larger than we had expected on the basis of the model for AHBA biosynthesis (Figure 1) which predicts only four or five genes [9]. In addition, the location of a key biosynthetic gene (*rifJ*) almost 30 kb away from the AHBA synthase gene was entirely unexpected. It implies that the intervening genes are all still part of the rifamycin biosynthetic gene cluster, despite the finding that some of them do not seem to be essential for the formation of rifamycin B itself. Thus, it appears that the biosynthesis of the AHBA starter unit may be more complicated than previously anticipated.

The sequence data confirm the notion that the mC_7N unit biosynthetic pathway diverges at a very early stage from the primary metabolic shikimate pathway, probably in the very first step(s). The homologies identified, however, give no clue as to the source of the nitrogen or the process by which it is introduced into aminoDAHP. It was originally proposed to come from the amide nitrogen of glutamine by action of an amidohydrolase associated with the DAHP synthase [44]. No homology to known amidohydrolases has been seen in the entire cluster, however. Yet, the end-to-end similarity of RifH with DAHP synthases from plants, except for a plastid-targeting leader sequence, suggests that additional enzymatic functions are required to form aminoDAHP. The presence of the quinate/shikimate dehydrogenase-homologous gene (*rifI*) in the cluster is rather curious. How such an enzyme would fit into the current model of AHBA biosynthesis is unclear. One may speculate that RifI reoxidizes C-3 of aminoquinate or aminoshikimate when aminoDHQ or aminoDHS has been aberrantly reduced. This reoxidation reaction would direct aminoquinate or aminoshikimate away from catabolic pathways and back into the mC_7N unit biosynthetic manifold. There is at present no evidence to support this hypothesis, however.

The relationship between the number and deduced functions of the five rifamycin PKS genes and the structure of rifamycin W and B are convincing evidence for the proposed involvement of proansamaycin X as the intermediate produced directly from the product of the rifamycin PKS (Figure 2). Similar to the erythromycin and rapamycin PKSs, the rifamycin PKS genes appear to be colinear with respect to the functions in their products required to build the rifamycin carbon skeleton. Yet this PKS has a novel loading domain specifying an aromatic starter unit and lacks a conventional thioesterase off-loading domain at the end of RifE, which typically catalyzes polyketide chain termination in macrolide formation. RifF may substitute for the thioesterase by displacing the thioester linkage of the assembled polyketide chain to the last module of the PKS with the amino group of the starter unit in forming the unique macrolactam ring of rifamycin. A thioesterase gene has been located almost 30 kb downstream of *rifE*, but it seems doubtful that this gene functions in normal polyketide chain termination. A type I PKS without a thioesterase domain is also present in the rapamycin biosynthetic gene cluster [23], in which chain termination of the polyketide has been postulated to occur by transfer to the amino group of an enzyme-bound pipecolic acid, catalyzed by the *rapP* gene product [45], and cyclization by subsequent displacement of the pipecolic acid thioester linkage to the enzyme by a hydroxyl group of the polyketide chain. RapP bears no similarity to RifF, however.

There appear to be more modifying genes in the cluster than are probably required to assemble rifamycin B (e.g., the cluster of deoxyhexose biosynthesis genes or more CYP450 genes than seem necessary) whereas some required genes (e.g., an acyltransferase gene to acetylate the 25-OH group) have not been located yet. Some of the missing genes may be located in the 5 kb gap between *orf11* and *orf12*, which is presently being sequenced, or adjacent to either end of the sequenced region. The presence of additional genes not required to produce rifamycin B may reflect the fact that all present rifamycin-B-producing *A. mediterranei* cultures, presumably including strain S699, are apparently derived from the same parent [6], which originally produced a broad range of different rifamycins [46]. It was then found that addition of barbital to the fermentation greatly enhances rifamycin B titers and suppresses the formation of the other components [47], and subsequent strain improvement gave cultures, like S699, that produce predominantly rifamycin B even in the absence of barbital. It seems likely, therefore, that strain S699 still contains the genetic information to produce a large number of different rifamycins, but that many of these genes are either silent or expressed at very low levels.

Structural modification of macrolide antibiotics has proven to impart significant improvements in oral bioavailability and more persistent serum/tissue levels of the molecules [48,49]. Such changes can increase their effectiveness and thus they have the potential to slow the emergence of antibiotic resistance and can lead to the discovery of molecules with improved antibacterial spectra. In recent years, success has been achieved in manipulating the structures of naturally produced secondary metabolites by applying modern methods of molecular biology [7,50]. Clearly, it could be profitable to apply molecular genetic methods to alter the structure of rifamycins with the aim of generating new analogs capable of overcoming emerging resistance to rifampicins. The multitude of potential modifying genes discovered in this work offers ample opportunities to alter

functionalities introduced after polyketide assembly by the inactivation of such genes. And the availability of the complete sequence of the rifamycin PKS sets the stage for the manipulation of the polyketide backbone structure itself by selective elimination, addition or replacement of specific enzyme functions or sets of enzyme functions in this PKS by genetic engineering.

Significance

The naturally occuring rifamycins and the semisynthetic antibiotics derived from them represent a major class of medically important drugs as well as being polyketides with a unique mode of biosynthesis in a genus of the Actinomycetes that has not been well-studied genetically. 3-Amino-5-hydroxybenzoic acid (AHBA), the aromatic amino acid starter unit, is assembled by an unusual variation of the shikimate pathway that appears to be considerably more complex than originally thought. This issue can now be investigated in detail through experiments using the genes we have implicated to be involved in the aminoshikimate pathway. Similarly, characterization of the rifamycin polyketide synthase (PKS) genes enables a study of the role of the novel amide synthase gene in the PKS mechanism to learn if the mechanism of ansamacrolide ring formation is analogous to the one used to make the macrolactone rings of the erythromycins. Of more importance, however, is the potential for modifying the genes we have characterized using combinatorial biosynthesis and targeted mutagenesis in attempts to produce novel forms of the rifamycins that may have enhanced activity or potency, especially against the rapidly emerging rifamycin-resistant microorganisms.

Materials and methods

General procedures

The *E. coli* strain XL-1 Blue (Stratagene) and plasmids pUC19 and pGEM3zf(+) (Promega, Madison, WI) were used in this work. The *A. mediterranei* strain S699, which produces predominantly rifamycin B, was a gift from Professor Giancarlo Lancini (Lepetit Research Laboratory, Geranzano, Italy). Growth of *A. mediterranei* was as described elsewhere [10].

E. coli was grown on LB agar medium. Growth media were supplemented with the appropriate antibiotics at the following concentrations: Agar plate; carbenicillin, 100 mg/ml; Luria Broth; Carbenicillin, 50 mg/ml. Competent *E. coli* XL-1 Blue cells were prepared and transformed according to Sambrook *et al.* [51].

Isolation and purification of *A. mediterranei* genomic DNA was performed according to Hopwood *et al.* [52] using the Kirby protocol. Isolation of plasmid DNA from *E. coli* was performed using Quiagen (Hilden, Germany) ion exchange columns. Restriction enzymes were obtained from Gibco/BRL, and digestions were performed according to the manufacturer's instructions.

Reagent grade chemicals were purchased from Sigma Chemical Co., or from EM Science. [^{35}S]dATPα-S (5,000 Ci/mmol) and [α-^{32}P]dATP (3,000 Ci/mmol) were obtained from Amersham Corp.

Double-stranded DNA sequencing was performed using the chain termination method, (Sequenase kit, version 2.0), deazaguanosine dinucleotides (United States Biochemical) and [α-^{35}S]dATP. Additional sequencing was performed using the Perkin Elmer Amplitaq Dye-terminator sequencing system and run on a Applied Biosystems 373A or 377 automated DNA sequencer. Alignment of sequence contigs was performed using the DNA Star program Seqman. DNA sequence analysis was performed using the Intelligenetics, Inc., the Wisconsin Genetics Computer Group software programs versions 7.0 and 8.1 [53], and the NCBI world wide web blast server.

Gene library preparation

A. mediterranei genomic DNA was partially digested with *Sau3AI* and fragments in the 30–40 kb size range were cloned into *Bam*HI/*Hpa*I digested vector pOJ446 as described by Bierman *et al.* [54]. A 2.3 kb *Xho*I fragment carrying the AHBA synthase gene [10] was used to probe this cosmid library. Southern hybridization and restriction enzyme mapping identified 12 clones that contained the AHBA synthase gene and jointly extended over a 69 kb region with the AHBA synthase gene in the center. Cosmid DNA flanking the AHBA synthase gene in both directions was sequenced as described. In addition, the 1.8 kb *Bam*HI fragment containing a portion of the PKS gene from the flanking region of the AHBA synthase gene cluster was used as the starting point to probe several restriction enzyme digests of *A. mediterranei* genomic DNA. Mini-libraries of size-fractionated genomic DNA cloned in pUC19 or pGEM3zf(+) were screened to provide multiple coverage over a contiguous 55 kb region of the chromosome. Clone junctions or small gaps that were not overlapped were filled by the PCR. The entire region was sequenced a minimum of two times on both strands.

Construction of rif mutants by gene disruption or replacement

This was accomplished by inserting a hygromycin resistance gene into an ORF (all cloning steps were carried out with the *E. coli* plasmid vectors described above in an *E. coli dcm dam* strain) and, following electroporative transformation of *A. mediterranei*, allowing that construct to undergo homologous recombination with the genome. The initial, hygromycin resistant clone produced by single crossover recombination was grown non-selectively to isolate the antibiotic sensitive recombinant resulting from a second crossover through which the vector was excised and the *rif* gene was replaced with its disrupted copy. Typically, 10 to 20 recombinants were isolated in the initial disruption step and by screening several hundred clones grown without hygromycin in the medium, two to four gene replacement clones were obtained. An example of the use of this method to produce the *rifK* mutant is given in [10].

Accession numbers

The nucleotide sequence of the rifamycin genes described in this paper is available from GenBank and related databases under the accession numbers AF040570 and AF040571.

Acknowledgements

We thank Chris Reeves at MDS-Panlabs for use of an automated DNA sequencer in this research program. This work was supported by grants (AI2026 to H.G.F., AI38947 to C.R.H.) and fellowships (GM17457 to P.R.A.) from the National Institutes of Health.

References

1. Rinehart, K.L. Jr. & Shields, L.S. (1976). Chemistry of the ansamycin antibiotics. *Fortschr. Chem. Org. Naturst.* **33**, 231-273.
2. Brufani, M., Fedeli, W., Giacomello, G. & Vaciago, A. (1964). X-ray analysis of the structure of rifamycin B. *Experientia* **20**, 339-342.
3. Oppolzer, W. & Prelog, V. (1973). Über die konstitution und die konfiguration der rifamycine B, O, S und SV. *Helv. Chim. Acta* **56**, 2287-2314.
4. Sepkowitz, K. A., Rafalli, J., Riley, L., Kiehn, T.E. & Armstrong, D. (1995). Tuberculosis in the AIDS era. *Clinical Microbiol. Revs.* **8**, 180-199.
5. Chiao, J.S., Xia, T.H., Mei, B.G., Jin, Z.K. & Gu, W.L. (1995). Rifamycin SV and related ansamycins. In: *Genetics and Biochemistry of Antibiotic Production*, L. C. Vining, C. Stuttard, Eds., Butterworth-Heinemann, Boston, pp. 477-498.

6. Lancini, G. & Cavalleri, B. (1997). Rifamycins. In *Biotechnology of Antibiotics*, 2nd Ed., W. R. Strohl, Ed., Dekker, New York, pp. 521-549.
7. Katz, L. & S. Donadio. (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* **47**, 875-912.
8. Ghisalba, O. & Nüesch, J. (1981). A genetic approach to the biosynthesis of the rifamuycin chromophore in *Nocardia mediterranei*. III. Isolation and identification of 3-amino-5-hydroxybenzoic acid as a direct precursor of the seven carbon amino starter-unit. *J. Antibiot.* **34**:64-71.
9. Kim, C.-G., *et al.*, & Floss, H.G. (1996). Biosynthesis of 3-amino-5-hydroxybenzoic acid, the precursor of mC_7N units in ansamycin antibiotics. *J. Amer. Chem. Soc.* **18**, 7486-7491.
10. Kim, C.-G., Yu, T.W., Fryhle, C., Handa, S. & Floss, H.G. (1998). 3-Amino-5-hydroxybenzoic acid (AHBA). synthase, the terminal enzyme in the formation of the precursor of mC_7N units in rifamycin and related antibiotics. *J. Biol. Chem.*, in press.
11. Hopwood, D.A. (1988). Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. *Proc. Roy. Soc. London B* **235**, 121-138.
12. Dewick, P.M. (1990). The biosynthesis of shikimate metabolites. *Nat. Prod. Repts* **7**, 165-189.
13. Sosio, M., Lorenzetti, R., Robbiati, F. & Denaro, M. (1991). Nucleotide sequence to a teicoplanin resistance gene from *Actinoplanes teichomyceticus*. *Biochim. Biophys. Acta, Gene Struct. Expr.* **1089**, 401-403.
14. Millar, G., & Coggins, J.R. (1986). the complete amino acid sequence of 3-dehydroquinate synthase of *Escherichia coli* K12. *FEBS Letters*, **200**, 11-17.
15. Charles, I.G., Keyte, J.W., Brammar, W.J. & Hawkins, A.R. (1985). Nucleotide sequence encoding the biosynthetic dehydroquinase function of the penta-functional *arom* locus of *Aspergillus nidulans*. *Nucleic Acids Res.* **13**, 8119-8128.
16. Tercero, J.A. , Espinosa, J.C.R., Lacalle, R.A. & Jiménez, A. (1995). The biosynthetic pathway for the aminoglycoside antibiotic puromycin, as deduced from the molecular analysis of the *pur* cluster of *Streptomyces alboniger*. *J. Biol. Chem.* **271**, 1579-1590.
17. Loos, H., Sahm, H. & Sprenger, G.A. (1993). Glucose-fructose oxidoreductase, a periplasmic enzyme of *Zymomonas mobilis*, is active in its precursor form. *FEMS Microbiol. Lett.* **107**, 293-298.
18. Schaferjohann, J. , Yoo, J.G., Kusian, B., & Bowien, B. (1993). The *cbb* Operons of the facultative chemoautotroph *Alcaligenes eutrophus* encode phosphoglycolate phosphatase. *J. Bacteriol.* **175**, 7329-7340.
19. Lyngstadaas, A., Løbner-Olesen, A. & Boye, E. (1995). Characterization of three genes in the *dam*-containing operon of *Escherichia coli*. *Mol. Gen. Genet.* **247**, 546-554.
20. Angell, S., Schwartz, E. & Bibb, M.J. (1992). The glucose kinase gene of *Streptomyces coelicolor* a3(2): its nucleotide sequence, transcriptional analysis and role in glucose repression. *Mol. Microbiol.* **6**, 2833-2844.
21. Ning, S. (1996). Studies on the biosynthesis of rifamycin by *Amycolatopsis mediterranei*. Ph.D. thesis, University of Washington.
22. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675-679.
23. Schwecke, T., *et al.*, & Leadlay, P. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl Acad. Sci. USA* **92**, 7839-7843.
24. Haydock, S.F. , *et al.*, & Leadlay, P.F. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374**, 246-248.
25. MacNeil, D.J., Occi, J.L., Gewain, K.M. & MacNeil, T. (1994). Correlation of the avermectin polyketide synthase genes to the avermectin structure: implications for designing novel avermectins. *Annu. NY Acad. Sci.* **721**, 123-132.
26. Ruan, X., Stassi, D., Lax, S. & Katz, L. (1997). A second type-I PKS gene cluster isolated from *Streptomyces hygroscopicus* ATCC 29253, a rapamycin-producing strain. *Gene* **203**, 1-9.
27. Ghisalba, O., Fuhrer, H., Richter, W.J. & Moss, S. (1981). A genetic approach to the biosynthesis of the rifamycin-chromophore in *Nocardia mediterranei*. III. Isolation and identification of an early aromatic ansamycin-precursor containing the seven-carbon amino start-unit and three initial acetate/propionate-units of the ansa chain. *J. Antibiot.* **34**, 58-63.
28. Anderson, M.G., Monypenny, D., Rickards, R.W. & Rothschild, J.M. (1989). Biosynthetic origins of the oxygen atoms in the ansamycin antibiotics rifamycin B, O & S. *J. C. S. Chem. Commun.* **1989**, 311-313.
29. Ghisalba, O., Auden, J.A.L., Schupp, T. & Nüesch, J. (1984). The rifamycins: Properties, biosynthesis and fermentation. In *Biotechnology of Industrial Antibiotics* (E. Vandamme, ed), pp. 281-327, Marcel Dekker, Basel.
30. Ohsako, S., Ohtomi, M., Sakamoto, Y., Uyemura, K. & Deguchi, T. (1988). Arylamine N-acetyltransferase from chicken liver II. Cloning of cDNA and expression in Chinese hamster ovary cells. *J. Biol. Chem.* **263**, 7534-7538.
31. Blanc, V., Salah-Bey, K., Folcher, M. & Thompson, C.J. (1995). Molecular characterization and transcriptional analysis of a multidrug resistance gene cloned from the pristinamycin-producing organism, *Streptomyces pristinaespiralis*. *Mol. Microbiol.* **17**, 989-999.
32. Fernández-Moreno, M.A., Caballero, J.L., Hopwood, D.A. & Malpartida, F. (1991). The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell* **66**, 769-780.
33. Caballero, J.L., Malpartida, F. & Hopwood, D.A. (1991). Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture. *Mol. Gen. Genet.* **228**, 372-380.
34. Peschke, U., Schmidt, H., Zhang, H., & Piepersberg, W. (1995). Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11. *Mol. Microbiol.* **16**, 1137-1156.
35. Kawaguchi, T., Asahi, T., Satoh, T., Uozumi, T. & Beppu, T. (1984). B-factor, an essential regulatory substance inducing the production of rifamycin in a *Nocardia* sp. *J. Antibiot.* **37**, 1587-1595.
36. Kawaguchi, T., Azuma, M., Horinuchi, S. & Beppu, T. (1988). Effect of B-factor and its analogues on rifamycin biosynthesis in *Nocardia* sp. *J. Antibiot.* **41**, 360-365.
37. Smith, D.R., *et al.*, & Reeve, J. (1997). complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135-7155.
38. Kishi, T., Yamana, H., Muroi, M., Harada, S., Asai, M., Hasegawa, T. & Mizuno, K. (1972). Tolypomycin, a new antibiotic. iii. Isolation and characterization of tolypomycin Y. *J. Antibiot.* **25**, 11-15.
39. Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortes, J. & Leadlay, P.F. (1991). Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythrea*: sequence similarities between EryG and a family of *S*-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* **230**, 120-128.
40. Van den Bergh, E.R.E., *et al.*,, & Meijer, W.G. (1996). Primary structure and phylogeny of the Calvin cycle enzymes transketolase and fructose bisphosphate aldolase of *Xanthobacter flavus*. *J. Bacteriol.* **178**, 888-893.
41. Kaneko, T., *et al.*, & Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109-136.
42. Frost, J.W. & Draths, K.M. (1995). Biocatalytic syntheses of aromatic compounds from D-glucose: renewable microbial sources of aromatic compounds. *Annu. Rev. Microbiol.* **49**, 557-579.
43. Merson-Davies, L.A. & Cundliffe, E. (1994). Analysis of five tylosin biosynthetic genes from the *tylIBA* region of the *Streptomyces fradiae* genome. *Mol. Microbiol.* **13**, 349-355.
44. Floss, H.G. & Beale, J.M. (1989). Studies on the biosynthesis of antibiotics. *Angew. Chem. Int. Ed. Engl.* **28**, 146-177.
45. Konig, A., *et al.*, & Leadlay, P.F. (1997). The pipecolate-incorporating enzyme for the biosyntheiss of the immunosuppressent rapamycin. *Eur. J. Biochem.* **247**, 526-534.
46. Margalith, P. & Pagani, H. (1961). Rifomycin. XIII. Fermentation and production of rifomycin complex. *Appl. Microbiol.* **9**, 320-324.
47. Margalith, P. & Pagani, H. (1961). Rifomycin. XIV. Production of rifomycin B. *Appl. Microbiol.* **9**, 325-334.
48. Kirst, H.A. & Sides, G.D. (1989). New directions for macrolide antibiotics: pharmacokinetics and clinical efficacy. Antimicrob. *Agents Chemother.* **33**, 1419-1422.
49. Hardy, D.J., Swanson, R.N., Shipkowitz, N.L., Freiberg, L.A., Lortey, P.A & Clement, J.J. (1991). *In vitro* activity and *in vivo* efficacy of a new series of 9-deoxo-12-deoxy-9,12-epoxyerythromycin A derivatives. *Antimicrobial Agents Chemother.* **35**, 922-928.
50. Hutchinson, C.R. (1997). Antibiotics from genetically engineered microorganisms. In: *Biotechnology of Antibiotics*, 2nd Ed., W.R. Strohl, Ed., Dekker, New York, pp. 659-682.
51. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* 2nd edn., Cold Spring Harbor Laboratory Press, NY.
52. Hopwood, D.A., *et al.*, & Schrempf, H. (1985). Genetic manipulation of *Streptomyces*, a laboratory manual. The John Innes Foundation, Norwich, UK.

53. Devereux, J., Haeberli, P & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.
54. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. & Schoner, B.E. (1992). plasmid cloning vectors for the conjugal transfer of DNA from *E. coli* to *Streptomyces* spp. *Gene,* **116**, 43-49.
55. Horii, M., Ishizaki,T., Paik, S.-Y., Manome,T. & Murooka,Y. (1990). An operon containing the gene for cholesterol oxidase and a cytochrome P450-like protein from *Streptomyces* sp. *J. Bacteriol.* **172**, 3644-3653.
56. Gorlach, J., *et al.,,* & Amrhein, N. (1993). Differential expression of tomato (*Lycopersicon esculentum* L.) genes encoding shikimate pathway isoenzymes. I. 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. *Plant Mol. Biol.* **23**, 697-706.
57. Valentin, H.E., Zwingmann, G., Schonebaum, A. & Steinbuchel, A. (1995). Metabolic pathway for biosynthesis of poly (3-hydroxybutyrate-co-4-hydroxybutyrate) from 4-hydroxybutyrate by *Alcaligenes eutrophus. Eur. J. Biochem.* **227**, 43-60.
58. Trower , M.K., Lenstra, R., Omer, C., Buchholz, S.E. & Sariaslani, F.S. (1992). Cloning, nucleotide sequence determination and expression of the genes encoding cytochrome P-450soy (*soyC*) and ferredoxinsoy (*soyB*) from *Streptomyces griseus. Mol. Microbiol.* **6**, 2125-2134.
59. Arisawa, A., Tsunekawa, H., Okamura, K. & Okamoto, R. (1995). Nucleotide sequence analysis of the carbomycin biosynthetic genes including the 3-O-acyltransferase gene from *Streptomyces thermotolerans. Biosci. Biotechnol. Biochem.* **59**, 582-588.
60. Jiang, X.M., Neal, B., Santiago, F., Lee, S. J., Romana, L.K. & Reeves, P.R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Mol. Microbiol.* **5**, 695-713.
61. Bechthold, A., Sohng, J.K., Smith, T.M., Chu, X. & Floss, H.G. (1995). Identification of *Streptomyces violaceoruber* Tü22 genes involved in the biosynthesis of granaticin. *Mol. Gen. Genet.* **248**, 610-620.
62. Pissowotzki, K., Mansouri, K. & Piepersberg, W. (1991). Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes strELMB2N. *Mol. Gen. Genet.* **231**, 113-123.
63. Niemi, J., Ylihonko, K., Hakala, J., Parssinen, R., Kopio, A. & Mantsala, P. (1994). Hybrid anthracycline antibiotics: production of new anthracyclines by cloned genes from *Streptomyces purpurascens* in *Streptomyces galilaeus. Microbiol.* **140**, 1351-1358.
64. Vaupel, M. & Thauer, R.K. (1995). Coenzyme F420-dependent N^5,N^{10}-methylenetetrahydromethanopterin reductase (Mer) from *Methanobacterium thermoautotrophicum* strain Marburg. Cloning, sequencing, transcriptional analysis, and functional expression in *Escherichia coli* of the *mer* gene. *Eur. J. Biochem.* **231**, 773-778.
65. Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. & McAlpine, J.B. (1991). An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea. Science* **252**, 114-117.
66. Crespi, M., Vereecke, D., Temmerman, W., Van Montagu, M. & Desomer, J. (1994). The *fas* operon of *Rhodococcus fascians* encodes new genes required for efficient fasciation of host plants. *J. Bacteriol.* **176**, 2492-2501.
67. Omer, C.A., *et al.,,* & O'Keefe, D.P. (1990). Genes for two herbicide-inducible cytochromes P-450 from *Streptomyces griseolus. J. Bacteriol.* **172**, 3335-3345.
68. Lalonde, G., O'Hanley, P.D., Stocker, B.A. & Denich, K.T. (1994). Characterization of a 3-dehydroquinase gene from *Actinobacillus pleuropneumoniae* with homology to the eukaryotic genes *qa-2* and *QUTE. Mol. Microbiol.* **11**, 273-280.