ABSTRACT  Coronavirus acid (CFA) is the polyketide component of the phytotoxin coronatine, a virulence factor of the plant pathogen Pseudomonas syringae. Our current knowledge of polyketide biosynthesis largely is based on the analysis of polyketide synthases (PKSs) in actinomycetes and other Gram-positive bacteria. Consequently, the cloning and characterization of the CFA biosynthetic gene cluster will contribute significantly to our knowledge of polyketide synthesis in Pseudomonas. In this report, we describe two genes in the CFA biosynthetic gene cluster that encode PKSs that are structurally and functionally similar to the multifunctional PKSs, which catalyze the synthesis of macrolide antibiotics. The CFA PKS genes were overproduced in Escherichia coli and shown to cross-react with antisera made to a modular PKS involved in erythromycin synthesis. A scheme for CFA biosynthesis is presented that incorporates the activities of all proteins in the CFA PKS. In this report a gene cluster encoding a pseudomonal polyketide has been completely sequenced and the deduced gene functions have been used to develop a biosynthetic scheme.

Polyketides constitute a huge family of structurally diverse natural products including antibiotics, chemotherapeutic compounds, and antiparasitics. The biosynthesis of polyketides is similar to the assembly of fatty acids (1). The latter begins with the condensation of acetyl-CoA (starter unit) and malonyl-CoA (chain extender unit), which is catalyzed by a ketosynthase (KS), resulting in an enzyme-bound β-keto thiolester. The keto group is then fully reduced in three subsequent steps: (i) reduction to a hydroxyl group by a ketoreductase (KR), (ii) dehydration to an unsaturated thiolester by a dehydratase (DH), and (iii) reduction of the unsaturated thiolester to a saturated fatty acid by an enoyl reductase (ER). The intermediates in this process usually are bound to an acyl carrier protein (ACP) in the form of a thiolester, and acyl transferases (ATs) are required for the loading of the starter and extender units. By varying chain length, acyl-CoA substrates, reductive modifications, and stereochemistry, nature has produced a vast diversity of polyketide natural products.

Historically, polyketide synthases (PKSs) have been classified as type I or II systems based on their relatedness to fatty acid synthases (FASs). Type I FASs, which are characteristic of fungi and vertebrates, are multifunctional enzymes containing catalytic sites for AT, ACP, KS, KR, DH, and ER activities. Type II FASs, which are characteristic of bacteria and plants, contain each catalytic site on a separate protein unit (1). The first example of a fungal PKS, 6-methyl salicylic acid synthase, most closely resembled the animal type I FAS multifunctional proteins (1), whereas the first bacterial PKS to be studied (coding for an aromatic polyketide produced by Streptomyces) had a type II structure and consisted of multiple proteins (1). Surprisingly, sequence analysis of the PKS for macrolide polyketides produced by actinomycetes revealed the presence of a “type I” organization; however, unlike the fungal PKSs, the bacterial “type I” enzymes are used nonreiteratively. In a macrolide PKS, each domain is dedicated to a distinct step in the biosynthetic reaction sequence, and a group of domains that catalyzes one cycle of condensation and reduction is designated as a module. The most extensively characterized modular PKS in a bacterium is 6-deoxyerythronolide B synthase (DEBS), an enzyme involved in the synthesis of the macrolide antibiotic erythromycin (2). The activities catalyzed by the domains within a modular PKS often are apparent in the structure of the growing polyketide chain; consequently, nucleotide sequencing has become a predictive tool for deducing the biosynthetic route for these compounds. Macrolide polyketides have a fairly reduced structure and often contain structurally related domains (3, 4). In contrast, polyketides synthesized by “type II” PKSs generally result in aromatic polyketides, and biosynthesis is catalyzed by a multienzyme complex in which individual enzymes may be used reiteratively (5, 6).

Most of the research on polyketide synthesis in bacteria has focused on compounds synthesized by Streptomyces or other actinomycetes; however, Pseudomonas spp. synthesize a number of polyketides including the phytotoxin coronatine (COR) (7), mupirocin (pseudomonic acid) (8), pyoluteorin (9), and 2,4-diacetylphloroglucinol (10). The phytotoxin COR consists of two distinct moieties, the polyketide coronafacic acid (CFA) and a cyclopropyl amino acid designated coronamic acid (CMA) (11). COR is produced by several pathovars (pvs.) of Pseudomonas syringae (12), a plant pathogen that causes disease in many commercially important plants worldwide. Although some disease occurs in the absence of COR, the phytotoxin substantially enhances the virulence of P. syringae (12). Although the mode of action of COR is not completely understood, the phytotoxin is thought to function as a molecular mimic of the octadecanoid-signaling pathways of higher plants (13). Unlike many phytotoxins, COR is not broadly

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424

PNAS is available online at www.pnas.org.

Biosynthesis of the Pseudomonas polynketide coronafacic acid requires monofunctional and multifunctional polynketide synthase proteins

VIDHYA RANGASWAMY*†, SAO JIRALERSPONG†‡, RONALD PARRY‡§, AND CAROL L. BENDER*§

*Department of Entomology and Plant Pathology, 110 Noble, Research Center, Oklahoma State University, Stillwater, OK 74078; and †Department of Chemistry, Rice University, Houston, TX 77005.

Edited by Salih J. Wakil, Baylor College of Medicine, Houston, TX, and approved October 22, 1998 (received for review August 20, 1998)
antimicrobial; instead, its toxicity is directed primarily at plant hosts.

COR biosynthesis has been intensively studied in *P. syringae glycinea* PG4180, where the 32-kb COR gene cluster contains two distinct regions encoding the structural genes for CMA and CFA biosynthesis (12) (Fig. 1A). Nucleotide sequence analysis revealed that the CMA region contained four genes designated *cmaA, cmaB, cmaT*, and *cmaU*, and further analysis indicated that CMA is produced via a thiophosphate mechanism by a modified peptide synthetase (refs. 14 and 15; J. Patel and R.P., unpublished data). A regulatory region controlling CFA and CMA biosynthesis (12) (Fig. 1A), sequenced, and shown to contain three genes designated *corP, corS*, and *corR* (16). The translational products of *corP* and corR were related to response regulators that function as members of two-component regulatory systems, and corR showed similarity to histidine protein kinases that function as environmental sensors (16).

With the exception of pyoluteorin, where partial sequence analysis has revealed the presence of a type I PKS (17), detailed reports on the organization of polyketide gene clusters in *Pseudomonas* are lacking. Partial sequencing of the CFA gene cluster and overexpression studies previously indicated that CFA biosynthesis required monofunctional proteins (18–20). However, in the present work, we show that CFA biosynthesis also requires two “type I” PKS modules, with each module comprising a separate protein. These proteins were overproduced in *Escherichia coli* and shown to be immuno logically related to DEBS. A model for the synthesis of CFA is presented that incorporates the activities of all proteins encoded by the CFA gene cluster.

**EXPERIMENTAL PROCEDURES**

**Materials.** *P. s. glycinea* strains were cultured routinely on King’s medium B (21) or mannitol-glutamate (MG) medium (22) at 28°C. *P. s. glycinea* N9 produces both COR and CFA and was derived from the wild-type PG4180 (23). The following COR–CFA–Tn5 mutants of *P. s. glycinea* were also utilized: A1, A2, A25, A28, B2, B21, E9, D5, and G2 (23–25). *E. coli* strains were grown under standard conditions (26), and antibiotics were added to media in the following concentrations: 100 μg/ml ampicillin, 25 μg/ml kanamycin, and 25 μg/ml tetracycline. Standard techniques were used for DNA manipulations (26), and constructs were transformed into either *E. coli* DH5α (ΔlacZΔM15) or BL21 (F*ompT lacZYA-argF*U169) DNA templates for sequencing reactions were purified using Qiagen or Wizard (Promega) columns.

**DNA Sequencing and Analysis.** Automated DNA sequencing was provided by the Oklahoma State University Recombinant DNA/Protein Resource Facility and the Molecular Genetics Core Facility, University of Texas–Houston Medical School by using standard Applied Biosystems sequencing techniques (Perkin–Elmer). Subclones were constructed in pZerO2.1 (Invitrogen), pBluescript (Stratagene), or pUC19 (27), and the nucleotide sequence was derived by using the T3/T7 or M13 universal/reverse primers. Gaps were filled by generating sequence from internal primers, and contiguous sequences were assembled by using SEQUENCER 3.0 (Gene Codes, Ann Arbor, MI). Sequence data were aligned and homology searches were performed by using the BLAST (28) and FASTA (29) programs. Further analyses were conducted by using the University of Wisconsin Genetics Computer Group (UWGGC) and DNASTAR sequence analysis packages or the Protein Resource Facility and the Molecular Biology Department.

**Protein Analysis.** *P. s. glycinea* strains were cultured in Hoitink–Sinden medium optimized for COR production (30) at 18°C for 120 h at 250 rpm. Cells were harvested from 1 ml of culture, lysed as described previously (31), and total cellular proteins were separated on SDS/PAGE gels containing 6% polyacrylamide. Separated proteins were electrophoretically transferred onto Immobilon-P membranes (Millipore) and visualized by Western blotting by using DEBS 2 antiserum (32).

Proteins were overproduced in *E. coli* as described previously (20). Proteins were labeled with [35S]methionine and [35S]cysteine, separated on 6% SDS/PAGE gels, blotted, and exposed to x-ray film.

**RESULTS AND DISCUSSION**

Previous sequencing of the 5′ end of *srl* fragment 4 (Fig. 1A) revealed a 230-bp region with relatedness to the *ery* (17) genes involved in the synthesis of erythromycin (20). This prompted us to investigate whether there was any antigenic similarity between the proteins involved in CFA biosynthesis and DEBS. Since very large enzymes (>200 kDa) are associated with modular PKSs, we analyzed the cellular proteins of *P. s. glycinea* N9 (COR–CFA–) and several COR–CFA– mutants for the presence of large proteins when grown under conditions optimal for COR production. SDS/PAGE analyses of total cellular proteins in N9 revealed two large proteins of approximately 220 and 280 kDa when this strain was grown under conditions that stimulated COR production (Fig. 2A, lane 1). These proteins were not present when cells were grown in rich medium or at 28°C, conditions unfavorable for COR production. Mutant D5, which maps at the 3′ end of the CFA gene cluster in *cfaB* (20), produced the 220- and 280-kDa proteins (Fig. 2A, lane 4). However, the CFA– mutants B2 and B21, which mapped within Srl fragment 5, produced only the 280-kDa protein (Fig. 2A, lanes 2 and 3). Mutants E9 and G2, which mapped to *cfa5*, produced neither protein (Fig. 2A, lanes 5 and 6), presumably because of the polarity of Tn5 on downstream genes. Mutants A1, A2, A25, and A28, which mapped within Srl fragment 6, also failed to synthesize the two large proteins associated with CFA synthesis (data not shown). The antigenic relatedness of the 220- and 280-kDa proteins to DEBS was investigated in immunoblotting experiments using DEBS 2 antiserum. The 220- and 280-kDa proteins in N9 and D5 (Fig. 2B, lanes 1 and 4) and the 280-kDa protein in B2 and B21 (Fig. 2B, lanes 2 and 3) showed immunoreactivity to the DEBS 2 antiserum. There was no reactivity with similar-sized proteins in E9, G2 (Fig. 2B, lanes 5 and 6), A1, A2, A25, or A28 (data not shown). These observations suggested that both the 220- and 280-kDa proteins are modular PKSs encoded by DNA largely present in the *srl* fragment 5 and 6, respectively. Although the exact location of the start site was unknown, we reasoned that the 280-kDa protein was encoded by DNA in Srl fragment 6. This hypothesis was tested by subcloning a 12-kb HindIII/EcoRI fragment encompassing Srl fragment 6 into pBluescript SK(+) and KS(−) to form pCfa6A and pCfa6I, respectively (Fig. 1B). When induced with isopropyl β-D-thiogalactoside, *E. coli* BL21(pCfa6A) cells contained a 280-kDa protein (Fig. 3, lane 2), which was absent from BL21 cells containing pCfa6I (Fig. 3, lane 3) and BL21[pBluescript SK(+)] (Fig. 3, lane 1). In a parallel experiment, the 6.0-kb Srl/BamHI fragment containing Srl fragment 5 was subcloned in pBluescript SK(+) and KS(−) to form pCfa7A and pCfa7I, respectively (Fig. 1B). After induction, *E. coli* BL21(pCfa7A) cells contained a 220-kDa protein (Fig. 3, lane 4), which was absent from BL21(pCfa7I) cells (data not shown) and from BL21 containing the other constructs used in this experiment (Fig. 3, lanes 1–3). These results confirmed that pCfa6A and pCfa7A contained ORFs encoding the 280- and 220-kDa proteins, respectively.

The 14.2-kb region beginning at the BamHI site separating fragments 6 and 7 and extending to the end of Srl fragment 5 (see shaded region, Fig. 1A) was sequenced, and the deduced organization of this region is shown in Fig. 1C. The most notable feature was the presence of two large ORFs designated...
cfa6 (8.0 kb) and cfa7 (6.2 kb). Each ORF had a potential ribosome binding site 9–12 bp upstream of the translational start site. The authenticity of the ORFs was validated by codon preference analysis with the UWCG program CODONPREFERENCE and a Pseudomonas codon usage table. The predicted size of the translational products of cfa6 and cfa7 were 284 and 221 kDa, respectively. Sequence analysis using the BLASTX and FASTEMBL programs revealed a high degree of similarity to DEBS, and each gene contained one PKS module. The spatial location of the AT, ACP, KS, AT, DH, ER, KR, and ACP domains in Cfa6 and the KS, AT, DH, KR, ACP, and thioesterase (TE) domains in Cfa7 are shown in Fig. 1C. The catalytic sites in the ACP, DH, KS, and KR domains of Cfa6 and Cfa7 were closely related to those previously identified in the DEBS and rapamycin synthase modules (3, 33). The ER domain in Cfa6 was closely related to the ER domain in module 4 of DEBS 2, whereas Cfa7 contained a TE domain at the C terminus that retained the active site residues conserved in the TE domain of DEBS 3 (2). The N-terminal AT and ACP domains of Cfa6 resembled the analogous loading domain of DEBS 1, module 1. The internal AT domain in Cfa6 contained a motif related to that in AT domains which incorporate propionate (34); however, it is likely that this domain is involved in the introduction of butyrate, and a similar observation has been reported for the niddamycin PKS (4). The AT domain in Cfa7 is similar to that observed in ATs which incorporate acetate (34).

Complementation analysis and primer extension experiments previously demonstrated that the CFA region was organized as a single 19-kb transcriptional unit (35). The 5’ end of the transcript contained six discrete ORFs including cfl and cfa5 showed relatedness to enzymes that activate carboxylic acids by adenylation (acyl-CoA ligases) (18, 19). The protein products of cfa1, cfa2, and cfa3 showed similarity to ACP, DH, and KS, respectively (19). The function of cfa4 could not be predicted from database searches. Interestingly, the following genes were coupled translationally: cfa2-cfa3, cfa3-cfa4, and cfa4-cfa5, ensuring equimolar production of the resulting gene products (19).

In the present study, we show that SsrI fragments 5 and 6 contain two large ORFs, cfa6 and cfa7, and that the transla-
tional products show strong similarity to DEBS. Cfa6 contains all domains required for extension of the polyketide by butyrate followed by complete reduction, whereas Cfa7 contained the domains needed for extension of the polyketide by acetate followed by reduction and dehydration. Two remaining ORFs in the CFA cluster mapped downstream of cfa7 (Fig. 1C). The translational product of cfa8 showed similarity to crotoyl-CoA reductases from Streptomyces spp. (20). Crotoyl-CoA reductase catalyzes one step in the conversion of acetoacetyl-CoA to butyryl-CoA, and the latter product, as ethylenediamine is also part of the biosynthetic gene cluster encoding the polyketide tylosin (36), and the overexpression of ccr in Saccharopolyspora erythraea has been utilized to produce novel derivatives of erythromycin (37). The recruitment of a ccr gene into the CFA gene cluster may reflect the requirement of butyryl-CoA as a precursor for DEBS synthesis. It is also important to note that cfa7 and cfa8 are coupled translationally (20). The protein product of cfa9 showed relatedness to TEs involved in the synthesis of gramicidin, tyrocidine, and tylosin (20). Analysis of a cfa9 mutant indicated that this gene is dispensable for CFA and COR production but may increase the release of enzyme-bound proteins from the COR pathway (20).

Precursor feeding studies with 13C-labeled substrates demonstrated that CFA is a novel polyketide synthesized from C-2 and C-3 of pyruvate, three acetate units, and one butyrate unit (7). Further studies indicated that the pyruvate used for CFA biosynthesis is converted to α-ketoglutarate before incorporation into CFA (38). Since C-1 of pyruvate is not incorporated into CFA, C-1 of α-ketoglutarate must be lost at some stage in CFA biosynthesis. The most reasonable way for this loss to occur would be by decarboxylation of α-ketoglutarate before it is incorporated into CFA. Precedent for this transformation is provided by the occurrence of a thiamine-dependent α-ketoglutarate decarboxylase in the Gram-positive bacterium Leuconostoc oenos (39). The decarboxylation of α-ketoglutarate produces succinic semialdehyde, which presumably would need to be converted into a CoA ester before it could be used as the starter unit for polyketide assembly. The isolation of 2-[1-oxo-2-cyclopenten-2-ylmethyl]butanoic acid (CPE) from the fermentation broth of COR producers has been reported (40). This suggests that a cyclopentenone ring may be formed early in the biosynthetic pathway leading to CFA. These considerations allow the formulation of a working hypothesis for the role of most of the genes present in the CFA gene cluster (Fig. 4).

The hypothetical pathway begins with conversion of succinic semialdehyde into its CoA ester. This process might require one of the two ligase genes (Cfl or Cfa5) located in the CFA gene cluster or it might utilize a gene that maps outside of the COR biosynthetic region. Cfl (ACP), Cfa3 (KS), and Cfa4 then are postulated to create enzyme-bound 2-carboxy-2-cyclopentenone, which subsequently is dehydrated by Cfa2 (DH) to give enzyme-bound 2-carboxy-2-cyclopentenone (CPC). CPC is assigned the role of a cyclase by default since none of the ORFs in the Cfl-Cfa5 region exhibits a cyclase signature motif. A noteworthy feature of this portion of the pathway is that the ACP, KS, and DH would be used noniteratively, in contrast to proteins involved in the type II bacterial PKSs and FASs. Furthermore, these proteins would catalyze the potential synthesis of a unique starter, possibly CPC. This sequence of events is reminiscent of the synthesis of norsoloronic acid, in which hexanoate is utilized as a starter unit (41).

The next stage in the pathway could proceed in one of two ways. The first would involve the release of free CPC from Cfa1 (ACP) with the aid of Cfa9 (TE), followed by conversion of CPC to its CoA ester using CF1 or Cfa5. The CoA ester of CPC then could be loaded onto Cfa6, which possesses a loading domain similar to that found at the beginning of DEBS I. Alternatively, CPC could be transferred directly from Cfa1 to Cfa6. Cfa6 contains all of the domains required for the extension of CPC by a butyrate unit followed by complete reduction at the β-keto ester position to give enzyme-bound CPC. The presence of trace amounts of CPC in the fermen-
tation broth of *P. s. glycinea* could be explained by the release of CPE from Cfa6 by the TE encoded by cfa9. Transfer of enzyme-bound CPE from Cfa6 to Cfa7 would allow completion of CFA biosynthesis. Cfa7 possesses all domains required for the extension of CPE by malonate followed by reduction and dehydration of the β-keto ester; it also contains a C-terminal TE domain to unload free CFA. Cfa7 does not contain a region that can be assigned a cyclase function. The presence of enone functionality in CPE suggests that the formation of the six-membered ring of CFA probably occurs via a Michael addition rather than an aldol condensation. This, in turn, suggests that the KR and DH domains of Cfa7 may operate after the cyclization since the β-keto ester would be the preferred substrate for the Michael addition. The final stage in COR biosynthesis would involve the formation of an amide bond between CFA and CMA. It seems likely that this process is catalyzed by CII or Cfa5.

In summary, we have described an example of a polyketide biosynthetic pathway that requires both monofunctional and multifunctional PKS proteins and have presented a model for CFA and COR biosynthesis that correlates well with the enzymatic activities implied by sequence analysis. More detailed investigations of CFA biosynthesis using precursor incorporation experiments and assays of overproduced proteins are underway.

We kindly thank H. von Döhrn for providing antisera to DEBS 2 and M. Ullrich for assistance with the immunoblotting experiments. C.B. acknowledges support from the Oklahoma Agricultural Experiment Station and National Science Foundation Grants MCB-9603618.
and MCB-9807774. R.P. acknowledges grant support from the National Science Foundation (MCB-9807774) and the Robert A. Welch Foundation (C-729).