Alanylclavam Biosynthetic Genes Are Clustered Together With One Group of Clavulanic Acid Biosynthetic Genes in *Streptomyces clavuligerus* $^\dagger$ $^\S$

Nathan J. Zelyas, Hui Cai, Thomas Kwong, and Susan E. Jensen*

Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Received 16 May 2008/Accepted 7 October 2008

*Corresponding author. Mailing address: Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada. Phone: (780) 492-0830. Fax: (780) 492-9234. E-mail: susan.jensen@ualberta.ca.

$^\S$ Supplemental material for this article may be found at http://jb.asm.org/.


*Streptomyces clavuligerus* produces at least five different clavam metabolites, including clavulanic acid and the methionine antimetabolite, alanylclavam. In vitro transposon mutagenesis was used to analyze a 13-kb region upstream of the known paralogue gene cluster. The paralogue cluster includes one group of clavulamic acid biosynthetic genes in *S. clavuligerus*. Twelve open reading frames (ORFs) were found in this area, and mutants were generated in each using either in vitro transposon or PCR-targeted mutagenesis. Mutants with defects in any of the genes orfA, orfB, orfC, or orfD were unable to produce alanylclavam but could produce all of the other clavams, including clavulanic acid. orfA encodes a predicted hydroxymethyltransferase, orfB encodes a YigF/YER057c/UK114-family regulatory protein, orfC encodes an aminotransferase, and orfD encodes a dehydrodrugase. All of these types of proteins are normally involved in amino acid metabolism. Mutants in orfC or orfD also accumulated a novel clavam metabolite instead of alanylclavam, and a complemented orfC mutant was able to produce trace amounts of alanylclavam while still producing the novel clavam. Mass spectrometric analyses, together with consideration of the enzymes involved in its production, led to tentative identification of the novel clavam as 8-OH-alanylclavam, an intermediate in the proposed alanylclavam biosynthetic pathway.

The biochemistry and genetics behind the production of β-lactam antibiotics by microbes have long been an area of active research due to the clinical importance of these compounds. *Streptomyces clavuligerus* is a bacterium capable of producing a number of β-lactam metabolites, including, most notably, the β-lactamase inhibitor clavulamic acid (9). Although clavulamic acid has only weak antimicrobial activity, it inhibits serine β-lactamase enzymes, which are responsible for much of the penicillin resistance that is seen in clinical settings. When used in combination with conventional β-lactam antibiotics, clavulamic acid provides a powerful tool to combat infections by antibiotic-resistant pathogens. Understanding the complete biosynthetic pathway of clavulamic acid could lead to the design of new strains of *S. clavuligerus* able to produce higher levels or more potent variants of clavulamic acid.

In addition to clavulamic acid, *S. clavuligerus* also produces the following other β-lactam metabolites that are structurally similar to clavulamic acid: clavam-2-carboxylate, 2-formyloxyclavam, 2-hydroxymethylclavam, 2-carboxyoxazolidinclavam, and alanylclavam (10, 42) (Fig. 1). All of these β-lactam compounds, including clavulamic acid, are called clavams by virtue of containing a 5R stereochemistry, the other clavams mentioned above have a 5S configuration and are thus called the 5S clavams. The β-lactamase inhibitory activity displayed by clavulamic acid is associated with its 5R stereochemistry (3), and therefore the 5S clavams exhibit antifungal or antibacterial activities but do not have β-lactamase inhibitory properties (10, 42).

Clavulamic acid and the 5S clavams share the same early steps in their biosynthesis. However, once clavaminic acid is synthesized, the pathway splits into two branches, one specific for clavulamic acid production and the other for 5S clavam production (13) (Fig. 1). Two reactions from the late stages of the pathway leading to clavulamic acid have been characterized (2, 21, 35), while only one intermediate leading specifically to the 5S clavams, 2-carboxymethylideneclavam, has been identified (51).

Three unlinked gene clusters containing genes involved in clavam biosynthesis have been identified in *S. clavuligerus* (52). These are the clavulamic acid, the clavam, and the paralogue gene clusters. The clavulamic acid gene cluster contains genes involved in the early steps required for the biosynthesis of clavaminic acid, as well as genes specific to the late steps of clavulamic acid production (17, 21, 23, 29, 31, 40). The clavam gene cluster carries primarily genes required for the late steps of 5S clavam biosynthesis but also includes *cas1* encoding clavamine synthase, an early biosynthetic enzyme. *cas1* is a paralogue of the *cas2* gene from the clavulamic acid gene cluster (34, 53). Except for *cas1*, paralogues for all of the other early biosynthetic genes of the clavulamic acid gene cluster are found in the paralogue gene cluster (22, 53). The paralogue gene cluster also contains genes involved specifically in the late steps of 5S clavam biosynthesis (22, 51, 53).

Five genes required specifically for the late steps of 5S clavam biosynthesis have been described. In the clavam gene cluster, *cvm1, cvm2*, and *cvm3* are necessary for the production...
of all 5S clavams (34, 51). cvm1 encodes a putative aldo-keto reductase, cvm2 shows similarity to isomerases, and cvm5 encodes a predicted monooxygenase. The roles of these three genes in 5S clavam biosynthesis are uncertain, but Cvm5 may catalyze a Baeyer-Villiger-type oxidation leading to 2-formyl-oxymethylclavam production. Two other genes, c6p and c7p, are found in the paralogue gene cluster and are also required for the production of all the 5S clavams (51). c6p encodes an aminotransferase that is predicted to catalyze the first step from clavaminic acid toward the 5S clavams, while c7p encodes a probable transcriptional regulator of 5S clavam biosynthesis.

While the activities of these gene products allow an outline of a pathway to the 5S clavams to be envisioned, it is apparent that additional activities are required to account for the diversity of the 5S clavam products formed.

In an effort to identify additional genes involved in 5S clavam biosynthesis, a region upstream of the known paralogue gene cluster was sequenced, and genes found in this area were subjected to mutagenesis. In vitro transposon mutagenesis was used to facilitate both DNA sequence analysis and mutagenesis, thus establishing the utility of this technique in *S. clavuligerus*. Analysis of the mutant phenotypes revealed four genes required specifically for alanylclavam production and suggested a putative biosynthetic pathway for alanylclavam.

**MATERIALS AND METHODS**

**Bacterial strains, plasmids, cosmids, and culture conditions.** Bacterial strains used in this study are listed in Table S1 in the supplemental material. *Escherichia coli* strains were grown in Luria broth (LB) or on LB agar plates at 37°C (45). Strains carrying resistance determinants were grown in medium supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml), apramycin (Apr; 50 μg/ml), or chloramphenicol (25 μg/ml), as appropriate.
Strains of *S. clavuligerus* were grown on maltose-yeast-extract-malt extract agar (50) or ISP4 (Difco) agar plates at 28°C. Liquid cultures used for determining antibiotic production levels were grown as previously described (21). Briefly, seed cultures were prepared by inoculating 10⁵ spores from glycerol stocks into 25-ml amounts of Trypticase soy broth supplemented with 1% (wt/vol) soluble starch. These cultures were incubated for 40 h and then inoculated at 2% (vol/vol) into soy flour medium (21). Culture supernatants were collected after 72 and 96 h for assay of metabolite titers. All *S. clavuligerus* liquid cultures were grown at 28°C and shaken at 250 rpm. Mutant strains were grown at least three times to determine metabolite-producing capabilities. Antibiotic-resistant strains of *S. clavuligerus* were grown with Apr (25 μg/ml) or thiostrepton (2.5 to 10 μg/ml).

**DNA manipulations.** Plasmid and cosmids DNA was isolated from *E. coli* using a standard technique (25). Plasmid and cosmids DNA was introduced into *S. clavuligerus* through intergeneric conjugation from *E. coli* ET12567(pUZ8002) as previously described (25). Exconjugants were isolated on AS-1 plus 10 mM MgCl₂ medium (4) agar using nalidixic acid (25 μg/ml) to select the donor *E. coli* strain, and an appropriate antibiotic was used to select for exconjugants. Both Southern analysis and mycelial PCR were used to confirm mutations in *S. clavuligerus*. PCRs were typically carried out using mycelia from 40-h seed cultures that had been washed twice with 10.3% sucrose, suspended to their original volume in dimethyl sulfoxide, and then added directly to PCR mixtures to give a final dimethyl sulfoxide concentration of 5% (vol/vol). Oligonucleotide primers used in this study are listed in Table S2 in the supplemental material.

**Mutant generation.** In vitro transposon mutagenesis. In vitro transposon mutagenesis reactions were carried out according to the manufacturer’s specifications (Epicentre, Madison, WI). Briefly, Tn5062 (carrying Apr<sup>e</sup>, ery<sup>r</sup>, and orf<sup>T</sup>) was liberated from pQM5062 (8) by restriction digestion and combined with cosm<sup>d</sup>G9, transposase, and reaction buffer. The reaction was incubated at 37°C for 2 h and then terminated by the addition of a stop solution and heating at 70°C for 10 min. Reaction mixtures were introduced into *E. coli* Transformax EC100, and transformants were selected on Apr containing media.

**DNA sequence analysis.** Cosmid 6G9 carries the paralogue gene cluster and its flanking regions. Within 6G9, an 11-kb EcoRI fragment comprises much of the region upstream of the paralogue cluster, and this was targeted for sequencing. Mutant forms of 6G9 carrying individual transposon insertions in the 11-kb EcoRI fragment were identified via restriction analysis. Thirty such mutant cosmids were introduced into *E. coli* transformants carrying different restriction sites. This construct, pSET-AT, was then conjugated into an *S. clavuligerus* end of the paralogue gene cluster of *S. clavuligerus* was performed using a standard technique (25). Plasmid and cosmids DNA was introduced into *S. clavuligerus* through intergeneric conjugation from *E. coli* ET12567(pUZ8002) as previously described (25). Exconjugants were isolated on AS-1 plus 10 mM MgCl₂ medium (4) agar using nalidixic acid (25 μg/ml) to select the donor *E. coli* strain, and an appropriate antibiotic was used to select for exconjugants. Both Southern analysis and mycelial PCR were used to confirm mutations in *S. clavuligerus*. PCRs were typically carried out using mycelia from 40-h seed cultures that had been washed twice with 10.3% sucrose, suspended to their original volume in dimethyl sulfoxide, and then added directly to PCR mixtures to give a final dimethyl sulfoxide concentration of 5% (vol/vol). Oligonucleotide primers used in this study are listed in Table S2 in the supplemental material.

**Results**

In vitro transposon mutagenesis of *S. clavuligerus*. The paralogue gene cluster of *S. clavuligerus* includes genes involved in the early stages of biosynthesis of all clavam metabolites as well as genes specific for the production of 5S clavam metabolites. In order to analyze the region of the *S. clavuligerus* chromosome located upstream of the paralogue cluster, in vitro transposon mutagenesis of cosm<sup>d</sup>G9 was undertaken. The resulting mutant cosmids were introduced into *E. coli*, and thousands of clones carrying Tn5062 in 6G9 were obtained. Among these, 156 mutant clones were subjected to restriction analysis to determine the Tn5062 insertion point. 6G9 can be digested into the following six EcoRI fragments: 8.2-kb fragment (pWE15) and 20-kb, 11-kb, 5.7-kb, 2.8-kb, and 100-bp fragments derived from the genomic DNA insert. The relative positions of the 5.7- and 2.8-kb fragments were already known from prior sequencing, and Southern analysis revealed that the 11-kb fragment (Fig. 2) carries the region of interest lying upstream of the paralogue gene cluster (data not shown). Thirty of the 156 mutant 6G9 cosmids carried Tn5062 in the 11-kb upstream fragment, as determined by restriction analysis. The number of Tn5062 insertions in the different EcoRI fragments is indicated in Table 1. Only 134 cosmids are represented in the table because 22 of the mutant cosmids were not readily interpretable by analysis of EcoRI restriction digests. For the 134 mutant cosmids, frequencies of insertion correlate well with the size of each fragment.

---

**Construction of pSET-AT-orf<sup>T</sup>:** The 5' gene from pDA504 was inserted as a 11-kb BglII/BamHI fragment into the BamHI site of pSET152 to give pSET-AT. A 1.9-kb AgeI/EcoRI fragment containing orf<sup>T</sup> and its upstream region was isolated from cosm<sup>d</sup>G9 and inserted into pSET-AT after passage through pSS1/180 to pick up compatible restriction sites. This construct, pSET-AT-orf<sup>T</sup>, was then conjugated into an *S. clavuligerus* orf<sup>T</sup>:Tn5062 mutant.

**Bioassays.** Clavulanic acid production was detected using a disc-diffusion bioassay with *Klebsiella pneumoniae* ATCC 29665 as the indicator organism (34). Alanylcavam and 2-hydroxymethylcavam were detected using a disc-diffusion bioassay with *Bacillus* sp. ATCC 27860 as the indicator organism (42).

**HPLC and LC-MS analyses.** Culture filtrates were analyzed for clavulanic acid and 5S clavams by high-performance liquid chromatography (HPLC) after imidazole derivatization using a BondElute C<sub>18</sub> column (100 by 8.0 mm, 10 μm; Phenomenex, Torrance, CA) under previously described conditions (34). Culture filtrates containing novel metabolites were further analyzed using LC-mass spectrometry (MS) as previously described (21).

---

**Analysis of the novel clavam.** Five-milliliter amounts of culture supernatants from both wild-type and orf<sup>D</sup> mutants of *S. clavuligerus* grown in soy flour medium for 96 h were acidified by the addition of 20 μl of 5 N HCl and then centrifuged for 5 min at 13,000 x g to remove precipitated matter. Supernatants were passed through SepPak Classic C<sub>18</sub> cartridges (Waters Scientific, Milford, MA), and the eluant (unretained fraction) containing the clavam metabolites was immediately neutralized with NaOH. Samples (0.5 ml) of each neutralized eluant were dried to dryness in a SpeedVac, dissolved in 100 μl of water, and derivatized by the addition of 25 μl of imidazole reagent (7). After 15 min at 21°C, 5-μl samples of each were analyzed by HPLC on an X Terra column (0.21 by 10 cm; Waters Scientific) operating under isocratic conditions in 0.1% acetic acid adjusted to pH 3.7 with NH₄OH (96%), methanol (4%) at 0.2 ml/min with detection at 311 nm and on-line analysis of effluents by electrospray MS using a ZMD-2 single quadrupole spectrometer (Waters Scientific).

**Nucleotide sequence accession number.** The nucleotide sequence of this region of the *S. clavuligerus* chromosome has been appended to the existing file in GenBank for the paralogue gene cluster under accession number AY426768.

---

**RESULTS**

In vitro transposon mutagenesis of *S. clavuligerus*. The paralogue gene cluster of *S. clavuligerus* includes genes involved in the early stages of biosynthesis of all clavam metabolites as well as genes specific for the production of 5S clavam metabolites. In order to analyze the region of the *S. clavuligerus* chromosome located upstream of the paralogue cluster, in vitro transposon mutagenesis of cosm<sup>d</sup>G9 was undertaken. The resulting mutant cosmids were introduced into *E. coli*, and thousands of clones carrying Tn5062 in 6G9 were obtained. Among these, 156 mutant clones were subjected to restriction analysis to determine the Tn5062 insertion point. 6G9 can be digested into the following six EcoRI fragments: an 8.2-kb fragment (pWE15) and 20-kb, 11-kb, 5.7-kb, 2.8-kb, and ~100-bp fragments derived from the genomic DNA insert. The relative positions of the 5.7- and 2.8-kb fragments were already known from prior sequencing, and Southern analysis revealed that the 11-kb fragment (Fig. 2) carries the region of interest lying upstream of the paralogue gene cluster (data not shown). Thirty of the 156 mutant 6G9 cosmids carried Tn5062 in the 11-kb upstream fragment, as determined by restriction analysis. The number of Tn5062 insertions in the different EcoRI fragments is indicated in Table 1. Only 134 cosmids are represented in the table because 22 of the mutant cosmids were not readily interpretable by analysis of EcoRI restriction digests. For the 134 mutant cosmids, frequencies of insertion correlate well with the size of each fragment.
Sequencing and analysis of the region upstream of the paralogue gene cluster. The 30 6G9 cosmids carrying Tn5062 in the 11-kb EcoRI fragment were used as templates to sequence bidirectionally outward from each transposon insertion point. Gaps in the resulting assembled sequence were filled by sequencing from custom primers. The end of the 2.8-kb EcoRI segment that abuts the 11-kb EcoRI fragment was also sequenced to bridge the gap between the known paralogue gene cluster and the newly sequenced region. Taken together, this provided 13,003 bp of new sequence upstream of the paralogue gene cluster.

Examination of the sequence revealed 12 new ORFs (Fig. 2; Table 2). Of these, four encode products with similarities to proteins involved in amino acid metabolism. orfA encodes a putative serine hydroxymethyltransferase, an enzyme typically involved in the reversible conversion of serine to glycine (46). orfA is also notable in that it contains a TTA codon as the 30th codon of the 391-codon ORF. The product encoded by orfC strongly resembles aminotransferases, which catalyze the transfer of amino groups from amino acids to oxoacids, such that the ketone oxygen is replaced by an amino group (30). orfD encodes a protein similar to threonine dehydratases. Threonine dehydratases (also called threonine deaminases) catalyze the dehydration and deamination of threonine or serine to yield 2-ketobutyrate or pyruvate, respectively (15).

Finally, the protein encoded by orfJ shows similarity to dihydrololate reductases, which carry out the reduction of dihydrololate to tetrahydrololate (48). This reaction is an essential step in the production of glycine, purines, and deoxymyridine phosphate.

Two of the new ORFs are similar to amino acid transport proteins. Proteins encoded by orfE and orfF have similarity to AzLD and AzIC, respectively. In Bacillus subtilis, AzIC and AzLD cause 4-azaleucine resistance when overproduced and are believed to be involved in branched-chain amino acid transport (5).

orfB, orfL, and orfJ code for proteins which show similarities to regulatory proteins. OrfB, based on its amino acid sequence, belongs to the YigF/YER057c/UK114 family of proteins. In eukaryotes, proteins in this family have been implicated in RNase activity, mitochondrial genome maintenance, and isoleucine biosynthesis (33, 26, 37). In prokaryotes, this family of proteins has been found to be involved in the repression of purine and isoleucine biosynthesis (14, 43). OrfL appears to be a LysR-type transcriptional regulator. LysR-type transcriptional regulators are found in many prokaryotic systems. In fact, ClaR, a protein that positively regulates clavulanic acid biosynthesis in S. clavuligerus, is itself a LysR-type protein (38, 41). OrfL shows similarity to the ArsR family of transcriptional repressors, proteins that regulate expression of arsenical resistance operons (12).

The three other ORFs in the region upstream of the paralogue gene cluster do not have any commonalities with the other ORFs or with each other. The protein encoded by orfG shows very limited similarity to protein kinases, the orfH protein product is similar to multidrug resistance transporters, and orfK codes for a protein with limited similarity to activators of Hsp90 ATPase.

Generation and analysis of mutants. To determine whether any of the newly discovered ORFs play a role in the biosynthesis of S5 clavams or other metabolites produced by S. clavuligerus, mutants with defects in each of the ORFs were obtained. Mutants in orfA, orfC, orfD, orfI, orfJ, orfK, and orfL were generated by conjugating cosmid 6G9 derivatives carrying Tn5062 in the ORF of interest into S. clavuligerus. Recombination between the mutant cosmids and the chromosome allowed gene conversion to occur, thus transferring the mutations to the chromosomal copy of the gene. However, even after repeated attempts, no orfE, orfF, orfG, or orfH mutations could be obtained via this route. Although this initially suggested that these ORFs may be essential for cellular functions, PCR-targeting mutagenesis was subsequently performed successfully to generate mutants (16). Since there were no mutant cosmids carrying a transposon insertion in orfB, the mutation of orfB was also carried out via PCR targeting. Thus, transposon-bearing mutants were generated in orfA, orfC, orfD, orfI, orfJ, orfK, and orfL, and mutants carrying an antibiotic resistance gene cassette replacing the ORF of interest were pre-
pared in orfB, orfE, orfF, orfG, and orfH. All mutations were confirmed through either Southern analyses or mycelial PCR. None of the mutants showed any morphological effects or dramatic changes in either cephamycin C or clavulanic acid production (data not shown). Most of the mutants also showed no changes in production of 5\textsuperscript{S}clavams. However, mutants with defects in orfA, orfB, orfC, and orfD were specifically defective in the production of alanylclavam (Fig. 3). Interestingly, analysis of culture filtrates of orfC and orfD mutants showed the presence of an additional clavam-type peak eluting just before the alanylclavam retention time on HPLC chromatograms, indicating that these mutants were producing a novel metabolite (Fig. 3).

Mass determination of the novel clavam. Underivatized clavam metabolites have poor retention on reverse-phase HPLC columns, and so to investigate the structure of this new metabolite, imidazole-derivatized clavams from wild-type and orfD culture supernatants were subjected to HPLC with on-

![Image of HPLC analysis](http://jb.asm.org/) Downloaded from http://jb.asm.org/ on August 15, 2017 by guest

**FIG. 3.** HPLC analysis of clavam metabolites produced by wild-type and mutant strains of *S. clavuligerus*. Wild type (A), orfA mutant (B), orfB mutant (C), orfC mutant (D), and orfD mutant (E) are shown. AC, alanylclavam; C2C, clavam-2-carboxylate; HMC, 2-hydroxymethylclavam; CA, clavulanic acid; NC, novel clavam. The faint dotted line shows the expected location of the novel clavam in the HPLC profile from the wild-type strain.

### Table 2. Newly sequenced ORFs from the region upstream of the paralogue cluster of *S. clavuligerus*

<table>
<thead>
<tr>
<th>ORF</th>
<th>Size (aa)a</th>
<th>Most similar protein, accession no., similarity/identity (%)b</th>
<th>Proposed function</th>
</tr>
</thead>
<tbody>
<tr>
<td>orfA</td>
<td>390</td>
<td>Serine hydroxymethyltransferase from <em>Methanocaldococcus jannaschii</em>, AAB99615, 32/47 over 380 aa</td>
<td>Hydroxymethyltransferase</td>
</tr>
<tr>
<td>orfB</td>
<td>126</td>
<td>YigF/YER057c/UK114 family protein from <em>Saccharopolyspora erythraea</em>, YP_001107461, 55/70 over 126 aa</td>
<td>Amino acid biosynthesis regulator</td>
</tr>
<tr>
<td>orfC</td>
<td>393</td>
<td>Aminotransferase from “Sordicella usitattus,” YP_824873, 37/53 over 344 aa</td>
<td>Aspartic/tyrosine/aromatic aminotransferase</td>
</tr>
<tr>
<td>orfD</td>
<td>351</td>
<td>Threonine dehydratase from an <em>Acidobacteria</em> bacterium, YP_591159, 39/56 over 314 aa</td>
<td>Threonine dehydratase</td>
</tr>
<tr>
<td>orfE</td>
<td>102</td>
<td>Hypothetical protein from <em>S. erythraea</em>, YP_001104250, 54/69 over 102 aa</td>
<td>AzlD-type amino acid permease</td>
</tr>
<tr>
<td>orfF</td>
<td>253</td>
<td>Branched-chain amino acid permease from <em>Rhodococcus</em> sp., 53/66 over 204 aa</td>
<td>AzlD-type amino acid permease</td>
</tr>
<tr>
<td>orfG</td>
<td>364</td>
<td>Hypothetical protein from <em>Streptomyces avermitilis</em>, YP_821993, 51/65 over 353 aa</td>
<td>Serine/threonine-type protein kinase</td>
</tr>
<tr>
<td>orfH</td>
<td>416</td>
<td>Transporter from <em>Streptomyces ambofaciens</em>, CAK50890, 45/60 over 397 aa</td>
<td>Aral-type arabinose efflux permease</td>
</tr>
<tr>
<td>orfI</td>
<td>356</td>
<td>LysR-type transcriptional regulator from <em>S. erythraea</em>, YP_001107820, 48/62 over 294 aa</td>
<td>Transcriptional regulator</td>
</tr>
<tr>
<td>orfJ</td>
<td>201</td>
<td>Dihydrofolate reductase-type protein from <em>Frankia alni</em>, YP_714368, 84/91 over 189 aa</td>
<td>Bifunctional deaminase-reductase</td>
</tr>
<tr>
<td>orfK</td>
<td>277</td>
<td>Hypothetical protein from <em>Frankia</em> sp., YP_481895, 36/49 over 240 aa</td>
<td>Uncharacterized conserved protein</td>
</tr>
<tr>
<td>orfL</td>
<td>229</td>
<td>ArsR-type transcriptional regulator from <em>Rhodococcus</em> sp., YP_707049, 66/82 over 195 aa</td>
<td>Transcriptional regulator</td>
</tr>
</tbody>
</table>

*a* aa, amino acids.

*b* Percent similarities/identities were determined using the online BLASTp program.
line electrospray MS analysis (Fig. 4). Analysis of wild-type culture supernatants showed an alanylclavam peak that gave the expected prominent mass ions of 269 (intact imidazole derivative) and 201 (fragmented derivative), whereas the novel clavam from orfD mutant culture supernatants gave a peak with mass ions of 285 (intact imidazole derivative) and 217 (fragmented derivative). These masses are consistent with the novel clavam having a structure of 8-OH-anylclavam. The HPLC profile of the wild-type culture supernatant shows a shoulder on the leading edge of the alanylclavam peak corresponding in retention time to that of 8-OH-anylclavam. That shoulder peak also shows low levels of mass ions at 285 and 217, indicating that 8-OH-alanylclavam may normally be present in wild-type culture supernatants but at low levels compared to those of alanylclavam. In contrast, the 8-OH-alanylclavam peak from the orfD mutant culture supernatant showed no appreciable levels of alanylclavam (Fig. 4A).

**Complementation of the orfC mutant.** Mutants with defects in either orfC or orfD accumulate the same novel clavam, suggesting either that the OrfC and OrfD proteins are both needed for a single reaction in alanylclavam synthesis or that the genes are transcribed together in an operon but that transposon insertion in the upstream orfC has polar effects on the expression of orfD. This latter possibility seemed most likely, since the orfC stop codon overlaps the orfD start codon, and this in turn suggested that the loss of OrfD alone could be responsible for accumulation of the novel metabolite.

To test whether these two genes are transcribed in an operon, the orfC mutant was complemented. orfC was inserted into pSET-AT, which confers thiostrepton resistance and integrates into the chromosome at the ΦC3I phage attachment site (6). orfC mutant strains carrying pSET-AT-orfC were fermented, and culture filtrates were analyzed (Fig. 5). Chromatograms revealed that the complemented mutants still produced primarily the novel clavam (retention time, 3.0 min), but a small peak (retention time, 3.3 min) consistent with the presence of alanylclavam was also seen (Fig. 5C). This shoulder peak is not found in the uncomplemented orfC mutant (Fig. 5B).

Complementation of the orfC mutant with pSET-AT-orfC was able to restore only trace levels of alanylclavam production, supporting the hypothesis that the orfC mutation was polar on orfD. To obtain additional evidence, we attempted to produce an orfC in-frame deletion mutant, which would not be polar on orfD. As a first step, a ΔorfC::apra mutant was created by PCR-targeted mutagenesis. HPLC analysis of culture filtrates from this mutant showed the same chromatographic profile as was seen in the orfC transposon mutant, with accumulation of the novel clavam (data not shown). However, when the ΔorfC::apra mutant was subjected to flp-mediated recombination to remove the Apr’ cassette and replace it with an 81-bp in-frame scar (16), no in-frame mutants were recovered, despite two independent attempts.

---

**FIG. 4.** MS analysis of the novel clavam accumulated by orfC and orfD mutants of S. clavuligerus. HPLC profile of orfD mutant of S. clavuligerus (A), mass spectrum of the novel clavam peak (retention time, 3.62 min) seen in panel A (B), HPLC profile of wild-type S. clavuligerus (C), and mass spectrum of the alanylclavam peak (retention time, 3.77 min) seen in panel C (D) are shown. Abbreviations for metabolite peaks are given in legend to Fig. 3.
glycine binding, it lacks five of the six amino acids involved in methylenetetrahydrofolate to serine and tetrahydrofolate. OrfD shows similarity to serine hydroxymethyltransferases, enzymes that catalyze the reversible interconversion of glycine and 5,10-methylenetetrahydrofolate (46). This suggests that OrfA may catalyze a reaction between glycine and a substrate other than 5,10-methylenetetrahydrofolate.

OrfB belongs to the YigF/YER057c/UK114 family of proteins (14, 26, 33, 37, 43). While these proteins have been implicated in many cellular functions in both eukaryotes and prokaryotes, involvement in regulation of isoleucine biosynthesis has been observed in a number of organisms. YigF from Salmonella enterica serovar Typhimurium is necessary for isoleucine biosynthesis in that it is required for normal IlvB (an aminotransferase directly involved in isoleucine biosynthesis) activity (14, 47). As well, members of the YigF/YER057c/UK114 family in Haemophilus influenzae and E. coli are capable of binding to α-keto acids, such as 2-ketobutyrate, the first intermediate of isoleucine biosynthesis (11, 39). This led to the formulation of a model in which YigF binds to an α-keto acid, possibly an intermediate in the isoleucine biosynthetic pathway such as 2-ketobutyrate, that would, in its unbound form, repress IlvB activity (47). By binding to the α-keto acid, YigF sequesters the metabolite and prevents it from inhibiting the aminotransferase, thus allowing isoleucine production to continue (47).

Examination of the OrfB amino acid sequence reveals that it contains all of the eight residues implicated in ligand binding (Y18, G32, Q33, F86, N90, P102, R104, and E118) as determined from the crystal structures of E. coli and human YigF/YER057c/UK114 family proteins (32, 55). Although C107 in TdcF (an E. coli family member) has also been identified as an important residue for 2-ketobutyrate binding, it is not as conserved as the other eight active site residues and is not found in either the human YigF/YER057c/UK114 family protein nor OrfB (11, 55). Therefore, OrfB is likely to be capable of binding to α-keto acids.

Analysis of the OrfC amino acid sequence shows that OrfC belongs to family I of the aminotransferases (20, 30). Family I contains aspartate, alanine, and aromatic amino acid aminotransferases (30), but OrfC contains only two of the six amino acids involved in substrate binding found in similar family I aminotransferases (24, 36, 44), suggesting that OrfC may catalyze aminotransferase reactions but use atypical substrates.

OrfD is similar to threonine and serine dehydratases. It contains all of the residues likely involved in substrate binding found in the E. coli biosynthetic threonine dehydratase (15), the S. enterica serovar Typhimurium threonine deaminase TdcB (49), and the rat liver serine dehydratase (56). For this reason, it is probable that OrfD catalyzes a reaction similar to that carried out by threonine and serine dehydratases, involving a substrate structurally analogous to threonine or serine.

From the apparent activities of OrfA, OrfB, OrfC, and OrfD, it is possible to envision a putative alanylclavam biosynthetic pathway. OrfA may use glycine as a substrate and fuse it to an as-yet-unobserved aldehyde-containing clavam intermediate in a reaction analogous to that catalyzed by serine hydroxymethyltransferase (Fig. 1). This would produce 8-OH-alanylclavam, which has the same mass as the novel clavam accumulated by orfC and orfD mutants. OrfB, by analogy to threonine deaminase, would then carry out the dehydration/deamination of 8-OH-alanylclavam to produce a clavam intermediate with a pyruvyl side chain attached at C2 (Fig. 1). Alanylclavam would finally be generated by transamination of 5,10-methylenetetrahydrofolate binding (46).

The known sequence of the paralogue gene cluster has been increased to 23,328 bp through the use of in vitro transposon mutagenesis, the first time this technology has been applied to S. clavuligerus. Within the 13 kb of new DNA sequence, 12 new ORFs were located. Using a combination of in vitro transposon mutagenesis and PCR-targeted mutagenesis, mutations were generated in each of the newfound ORFs.

For unknown reasons, Tn5062 could not be introduced into all of the genes on the chromosome. Mutant cosmids with Tn5062 inserted into orfE, orfF, orfG, and orfH were obtained, but transfer of these mutations into the S. clavuligerus chromosome was unsuccessful, despite repeated attempts. This suggested that the ORFs might be essential for S. clavuligerus survival, but PCR-targeted mutagenesis subsequently proved that not to be the case. The Tn5062 system is currently being used to mutate the entire Streptomyces coelicolor genome, and no indication of difficulties has been reported in that species (8, 18). Although the exact cause is not known, in vitro transposon mutagenesis using Tn5062 may not be ideal for saturation mutagenesis of S. clavuligerus, but it is still useful for DNA sequencing.

Of the 12 newly described ORFs, four were specifically required for alanylclavam biosynthesis: orfA, orfB, orfC, and orfD. Each of these ORFs encodes a protein with similarities to those involved in amino acid regulation or biosynthesis. OrfA shows similarity to serine hydroxymethyltransferases, enzymes that catalyze the reversible interconversion of glycine and 5,10-methylenetetrahydrofolate to serine and tetrahydrofolate. While OrfA contains three of the four amino acids involved in glycine binding, it lacks five of the six amino acids involved in
the keto group on the pyruvyl side chain through the action of OrfC (Fig. 1). In this putative pathway, the role of OrfB is regulatory in nature. If the pyruvylclavam intermediate (an α-keto acid) inhibits OrfC activity, analogous to the way α-keto acids inhibit aminotransferase activity in isoleucine biosynthesis, OrfB may function to sequester this intermediate, thereby allowing OrfC to remain active (Fig. 1). Although more detailed analyses will be required to confirm the identity of 8-OH-alanylclavam, the proposed structure is in agreement with the observed mass of the compound and is consistent with the proposed biosynthetic pathway.

The proposed pathway can also help to explain the observed phenotypes of the orfC and orfD mutants and of the complemented orfC mutant. In the orfD mutant, 8-OH-alanylclavam would be expected to accumulate, since its further conversion requires OrfD. Because orfC and orfD form an apparent operon, the polar effects of orfC disruption on orfD expression would also result in a buildup of 8-OH-alanylclavam in the orfC mutant. However, despite the operon arrangement, orfD may be expressed at low levels in the orfC mutants due to some level of transcription originating within the transposon or to the presence of a weak promoter located within orfC but after the point of Tn5062 insertion. In an orfC mutant, this low level of OrfD would generate correspondingly small amounts of the pyruvylclavam intermediate (not observed), but the main product would be 8-OH-alanylclavam. In the complemented orfC mutant, OrfD would remain at the same low level, and so even full restoration of OrfC would still result in 8-OH-alanylclavam as the main product, although low levels of alanylclavam would now also be produced. The fermentation profiles observed for these mutants and complemented strains are consistent with this proposed pathway and the transcriptional arrangement of the orfC and orfD genes. Interestingly, our unsuccessful attempts to generate an in-frame orfC mutant, which should have decoupled expression of orfD from that of orfC, suggests that a buildup of the predicted α-keto acid product of OrfD might be toxic to the cell.

Regulation of alanylclavam production is also of interest in view of the observation that orfA contains a TTA codon. This TTA codon represents only the second such codon found in any of the genes associated with production of β-lactam metabolites in S. clavuligerus. The only other TTA codon is found in ccaR, a gene that encodes a pathway-specific transcriptional regulator controlling cephamycin and clavulanic acid production. bldA is a rare leucyl tRNA for the translation of UUA codons in Streptomyces spp. (27). Genes containing TTA codons are typically strongly dependent on bldA for expression and are usually genes that are expressed during the later stages of growth, such as those involved in secondary metabolite production (28).

Examination of S. clavuligerus bldA mutants has revealed that they can produce all of the clavams except alanylclavam (unpublished results). orfA would make a logical target for regulation of alanylclavam biosynthesis by bldA, since it encodes the first dedicated enzyme of the proposed alanylclavam branch of the 5S clavam biosynthetic pathway, but additional studies will be required to confirm its involvement, since ccaR expression was found unexpectedly to proceed in a bldA mutant of S. clavuligerus, despite the presence of a TTA codon (54).

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). N.J.Z. was supported by a studentship from NSERC.

REFERENCES

24. Kamitori, S., A. Okamoto, K. Hirotsu, T. Higuchi, S. Kuramitsu, H.


