Cloning and Sequencing of the β-Lactam Hydroxylase Gene (cefF) from Streptomyces clavuligerus: Gene Duplication May Have Led to Separate Hydroxylase and Expandase Activities in the Actinomycetes

STEVEN KOVACEVIC AND JAMES R. MILLER*

Department of Molecular Genetics Research, Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, Indiana 46285

Received 3 August 1990/Accepted 5 October 1990

The deacetylcephalosporin C synthetase (hydroxylase) gene from Streptomyces clavuligerus has been cloned and sequenced. The open reading frame codes for a protein with an M_r of 34,584. The hydroxylase gene (cefE) is closely linked to the epimerase gene (cejD) and the expandase gene (cejH) is transcribed in the opposite orientation. The hydroxylase and expandase genes are 59 and 71% identical at the amino acid and DNA levels, respectively. cefE and cefF may have arisen from a gene duplication in the actinomycetes.

The production of cephalosporins in both the fungi and the actinomycetes is characterized by the expansion of the five-membered ring of penicillin N to the six-membered ring of deacetoxycephalosporin C. This reaction is catalyzed by the enzyme deacetoxycephalosporin C synthetase (expandase). The next step in the pathway is the hydroxylation of deacetoxycephalosporin C to deacetylcephalosporin C catalyzed by deacetylcephalosporin C synthetase (hydroxylase). In Cephalosporium acremonium, these two steps are carried out by a single enzyme (2, 14, 15). However, in Streptomyces clavuligerus, these reactions have been shown to be catalyzed by separate enzymes (4). Expandase and hydroxylase belong to the same class of enzymes, characterized as α-ketoglutarate-linked dioxygenases. The C. acremonium and the S. clavuligerus expandase genes have been cloned and sequenced (7, 14). S. clavuligerus expandase and C. acremonium expandase-hydroxylase show a high degree of similarity (7, 10), making it difficult to predict which amino acid residues are involved in substrate binding or to recognize functional domains responsible for the two activities. By characterizing the S. clavuligerus hydroxylase gene, we hoped to gain further insight into the regions required for activity or substrate binding.

Amino acid sequence (residues 2 to 21) obtained from purified S. clavuligerus hydroxylase (20) was used to design a single, 60-base oligonucleotide by using streptomycetes codon bias (16). The DNA probe was made to the sense strand and differed from the derived DNA sequence at residues 7 (Pro [CCG]), 17 to 19 (Gly [GGC] Ala [GCC] Asp [GAT]), and 21 (Glu [GAG]) (Fig. 1). Cosmid DNA that was previously characterized and contains the epimerase and expandase genes (6) was subjected to DNA blot analysis (12). The 60-base oligonucleotide hybridized to a region approximately 3.0 kb upstream of epimerase (Fig. 2). (It is now recognized that the DNA probe used to isolate the expandase gene [7] also cross hybridized to the 6-kbp BamHI fragment containing the hydroxylase gene [10].) DNA sequencing was performed with an Applied Biosystems model 370A automated DNA sequencer. Areas of ambiguity were resequenced with custom primers with the TAQ-TRAK (Promega Biotec) sequencing kit. The sequence of the open reading frame is shown in Fig. 3.

After site-directed mutagenesis to incorporate an NcoI restriction site at the initiation codon for the hydroxylase open reading frame, the gene (cefF) was cloned in an Escherichia coli expression vector as previously described (6). The plasmid, pOW399, was used for the high-level expression of hydroxylase in E. coli JM109. Cultures of E. coli JM109(pOW399) were induced for protein production and cell extracts were assayed for the presence of hydroxylase and expandase activity (2, 7, 20). E. coli JM109 (pOW399) cell extracts contained hydroxylase activity, whereas E. coli JM109 has no hydroxylase activity. Interestingly, the S. clavuligerus hydroxylase, either native or recombinant in origin, has low (ca. 2%) expandase activity. Moreover, the S. clavuligerus expandase, first thought to have no hydroxylase activity, has the same low level of hydroxylase activity (20).

Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis demonstrated the overproduction (>10% of the total cell protein) of a polypeptide with an M_r of 36,000 (data not shown). The corresponding open reading frame would code for a polypeptide with a molecular weight of 34,584.

The similarities between the S. clavuligerus expandase and hydroxylase and the identities among all three dioxygenases are shown in Fig. 4. At the amino acid level, hydroxylase and expandase are 59% identical (74% using conservative amino acid changes); DNA sequences are 71% identical (data not shown). Several regions not only are conserved among the dioxygenases but have conserved motifs among other β-lactam biosynthetic enzymes as well (5). One motif is located around Cys-100 (14), analogous to a Cys residue implicated in isopenicillin N synthetase activity (13). However, hydroxylase contains a Ser residue at this position which may result in lower expandase activity. We are presently testing this possibility by site-directed mutagenesis. Another motif is present in the conserved region from residues 179 to 205 (188 to 214 in Fig. 4). These two regions also have the highest amino acid conservation between the two expandases (7).
Four β-lactam genes from *S. clavuligerus* have now been cloned and sequenced (6–8; this report); all map within a 25-kbp region, and other β-lactam genes are located in this cluster (9, 11, 18). Moreover, the expandase and epimerase genes are transcribed as part of an operon that may contain other β-lactam genes (6). We anticipated that the hydroxylase gene would be part of this operon, but the hydroxylase gene is located upstream from the epimerase gene and transcribed in the opposite direction. The sequences upstream and downstream of the hydroxylase gene bear no similarities to the sequences flanking the epimerase and expandase genes (data not shown).

It is interesting to speculate how the genes for these closely related proteins have evolved. The similarity among the β-lactam gene products (specifically, isopenicillin N synthetase) led to the hypothesis that the β-lactam genes were horizontally transferred from the procaryotes to the fungi approximately 370 million years ago (3, 10). Because the expandase and hydroxylase genes from *S. clavuligerus* are so similar, they likely arose from a common ancestor. We favor the idea that initially a single gene was present that coded for an enzyme with both activities similar to the C. *ac crem o ni um* cefE gene product. Following gene duplication, one of the genes evolved to have better expandase activity; the other evolved to have better hydroxylase activity. Some β-lactam-producing bacteria may still only have one gene. Mapping and DNA sequencing data have shown that cefD does not immediately follow the epimerase gene (cefD) in *Streptomyces lipmanii*, yet regions normally flanking cefE are conserved (19). Also, *Flavobacterium* sp. strain SC 12,154, a gram-negative bacterium, produces deacetoxycephalosporin C (17) and may not have cefE, as judged by the mapping data of Smith et al. (18).

When the proposed transfer of the biosynthetic genes to the fungi occurred, a single expandase-hydroxylase gene may have been present in the epimerase-coding factor. This would explain the bifunctionality of the *C. ac crem o ni um* enzyme. Alternatively, separate expandase and hydroxylase genes might have been transferred. After transfer, a gene fusion which gave rise to the bifunctional enzyme occurred in *C. ac crem o ni um*. These alternatives could be readily tested by determining whether *C. ac crem o ni um* contains another gene. If a vestigial gene is present in *C. ac crem o ni um*, both expandase and hydroxylase genes would have been transferred.

Our mapping, DNA sequencing, and transcript data for the *S. clavuligerus* biosynthetic genes (6) conflict with the data of Smith et al. (18), who reported that the expandase gene is upstream of the epimerase gene (i.e., cefE is between cefD and pcbC). The location of the hydroxylase gene upstream from epimerase and the presence of low-level expandase activity might explain these discrepancies. Nevertheless, the complete set of β-lactam biosynthetic genes may soon be available to enhance the development of higher-producing strains and to aid the study of regulation of these important genes of secondary metabolism (1, 3).

**Nucleotide sequence accession number.** These data have been submitted to GenBank and have been assigned accession number M37186.

We thank W.-K. Yeh and J. Dotzlaw for providing us with...
FIG. 4. Amino acid comparison of S. clavuligerus hydroxylase (DACS), S. clavuligerus expandase (DAOC8) (7), and C. acremonium expandase-hydroxylase (DAOC8/DACS) (14). A consensus sequence showing identical residues among the three proteins is also given.

N-terminal protein sequence prior to publication and for performing expandase and hydroxylase assays.

REFERENCES


19. Weigel, B. Personal communication.

20. Yeh, W.-K. Personal communication.