Two Isozymes of Clavaminate Synthase Central to Clavulanic Acid Formation: Cloning and Sequencing of Both Genes from *Streptomyces clavuligerus*[†,‡]

E. Neil Marsh,[§] Margaret Dah-Tsyr Chang, and Craig A. Townsend*

Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218

Received June 4, 1992; Revised Manuscript Received August 13, 1992

ABSTRACT: Clavaminate synthase (CS) is an α-ketoglutarate-dependent oxygenase central to the biosynthesis of clavulanic acid, a potent inhibitor of β-lactamases. CS catalyzes the oxidative cyclization/desaturation of proclavaminic acid to clavaminic acid in a two-step process involving the intermediacy of dihydroclavaminic acid [Salowe, S. P., Krol, W. J., Iwata-Reuyl, D., & Townsend, C. A. (1991) *Biochemistry 30*, 2281–2292]. During the purification of CS to homogeneity from *Streptomyces clavuligerus* [Salowe, S. P., Marsh, E. N., & Townsend, C. A. (1990) *Biochemistry 29*, 6499–6508], two forms of the enzyme capable of carrying out the complete reaction having very similar molecular weights and kinetic properties were isolated by Mono-Q chromatography. The gene for each has been cloned, sequenced, and found to be significantly homologous (87% identity). The two genes so isolated, *cs1* and *cs2*, have open reading frames of 975 and 978 nucleotides, respectively, encoding proteins of M_r 35 347 and 35 774. These genes are located in different loci of the genome separated by >20 kbp. This separation is large for a natural product biosynthetic pathway in bacteria where gene duplication and limited divergence are typically observed to occur within narrower confines of a gene cluster. Sequence comparisons made between *cs1*/*cs2* and other genes encoding iron-dependent proteins involved in penicillin and cephalosporin biosynthesis in the same organism show minimal homology. Further sequence alignments made to other non-heme iron oxygenases reveal unexpected dissimilarity within the α-ketoglutarate-dependent class itself. The limited data available suggests evolutionary convergence among these proteins.

Streptomyces clavuligerus is unusual for producing two major groups of β-lactam antibiotics. The oxygen-containing ring systems of the important β-lactamase inhibitor clavulanic acid (**3**) (Howarth et al., 1976) and the clavam metabolites represented by **4** (e.g., R = COOH, CH_2OH, CH_2OCHO, CH_2CH_2OH) (Brown et al., 1979; Wanning et al., 1981) having the opposite ring-fusion geometry comprise one of these. The other is represented by the many members of the penicillin (e.g., **6**), cephalosporin (e.g., **7** and **8**), and cephamycin (7-α-methoxycephalosporin) families (Nagarajan et al., 1971). As shown in Scheme I, iron-dependent enzymes play central roles in the construction of the strained bicyclic skeleta of these molecules which are essential to their antibiotic activities. Of these the non-heme iron oxygenases, clavaminate synthase (CS)[1] (Elson et al., 1987; Salowe et al., 1990), deacetoxycephalosporin C synthase (DAOCS) (Jensen et al., 1985; Rollins et al., 1988; Dotzlaf & Yeh, 1989), and deacetylcephalosporin C synthase (DACS) (Jensen et al., 1985; Baker et al., 1991) all require ferrous ion, molecular oxygen, and α-ketoglutaric acid (α-KG) for activity (Hook et al., 1979). Here we describe experiments to clone and sequence CS, which was unexpectedly found to exist as two isozymes. The deduced amino acid sequences for the two *cs* genes have been compared to those recently published for DAOCS (Kovacevic et al., 1989) and DACS (Kovacevic & Miller, 1991) from *S. clavuligerus* to test for homology suggestive of evolution from a common ancestral synthase gene. Further comparisons have been made to other iron-dependent oxygenases.

Clavaminate synthase (CS) was first reported by scientists at Beecham as a single protein (Elson et al., 1987; Woroniecki et al., 1989). Independent purification to homogeneity (Salowe et al., 1990) revealed two active proteins each capable of carrying out the full transformation of proclavaminic acid (**1**) to clavaminic acid (**2**). These enzymes were quite similar in molecular weight (M_r 47 000 and 46 000 ± 2000 as judged by SDS–PAGE), K_m, and V_{max}, although the smaller was a slightly better catalyst of this reaction (by less than a factor of 2) based on its k_{cat}/K_m (Salowe et al., 1990). As the ratio between these proteins was observed to vary over several preparations, it was thought that the smaller of the two (CS2) was derived by proteolysis of the larger (CS1) during fermentation and/or initial cell breakage (Salowe et al., 1990). CS is one of a number of Fe(II)/α-KG-dependent oxygenases that share an apparently common mechanism of oxygen activation. In a first step, sequential kinetics are typically observed such that Fe(II), oxygen, α-KG, and substrate are bound, and then oxidative decarboxylation of α-KG is believed to occur to give carbon dioxide, succinate, and a reactive iron–oxygen species (Siegel, 1979; Hanauske-Abel & Günzler, 1982). In a second step, the substrate, whose binding at this point can be reversible (Salowe et al., 1991), reacts most commonly by insertion of oxygen into substrate C–H bonds to give hydroxylation with retention of configuration (Fujita et al., 1964; Townsend & Barrabee, 1984; Englard et al., 1985; Stubbe, 1985). In the absence of substrate, an uncoupled reaction frequently takes place in which the reactive iron–oxygen species is quenched by water to give hydrogen peroxide

[†] This work was supported by the National Institutes of Health Grant AI14937 and the Camille and Henry Dreyfus Foundation.

[‡] The nucleotide sequences in this paper have been submitted to GenBank under Accession Numbers LO6213 (*cs1*) and LO6214 (*cs2*).

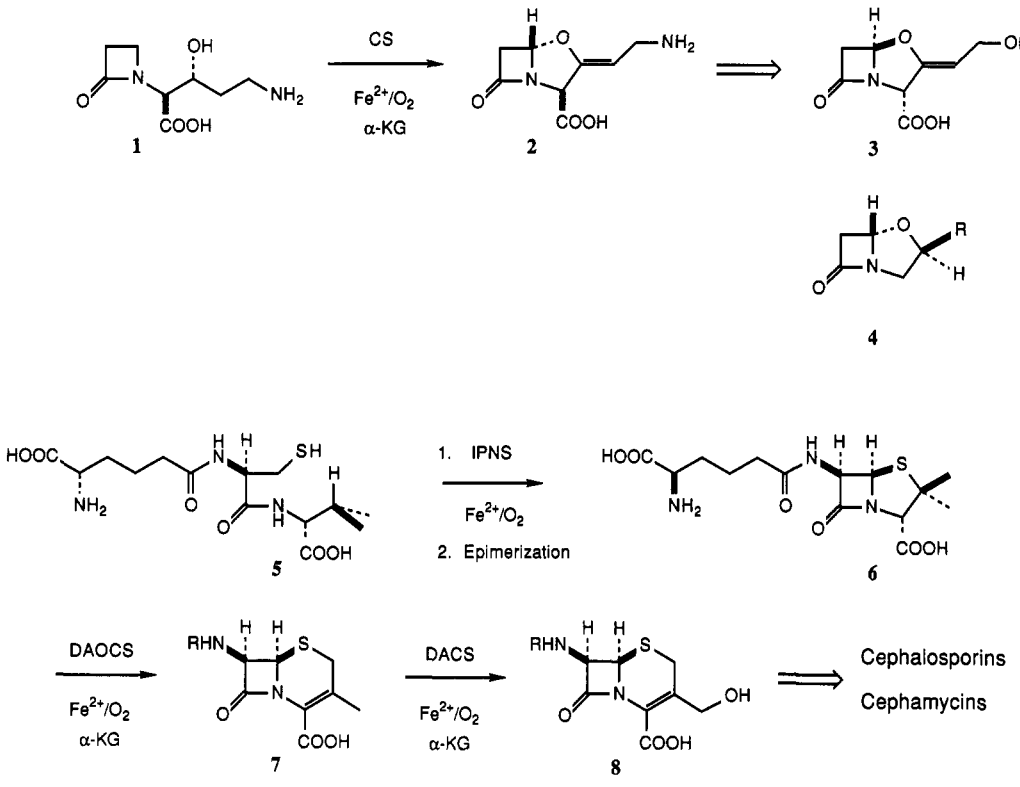
* Author to whom correspondence should be addressed.

[§] Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW U.K.

[1] Abbreviations: CS, clavaminate synthase; DAOCS, deacetoxycephalosporin C synthase; DACS, deacetylcephalosporin C synthase; α-KG, α-ketoglutaric acid (2-oxoglutaric acid); HPLC, high-performance liquid chromatography; IPNS, isopenicillin N synthase.

0006-2960/92/0431-12648$03.00/0

Scheme I



at a rate significantly below normal substrate turnover. CS has been shown to have all of these properties but one; unlike the hydroxylases it does not catalyze oxygen insertion from molecular oxygen into a substrate C–H bond during product formation. Instead, the conventional dioxygenase pathway is diverted to promote the insertion of *substrate* oxygen, again with retention of configuration, during the formation of clavaminic acid (Basak et al., 1990). In this overall transformation, therefore, molecular oxygen is reduced to water. This highly exothermic alternative is mechanistically linked to the otherwise thermodynamically unfavorable conversion of proclavaminic acid (**1**) to the significantly more strained clavaminic acid (**2**) (Salowe et al., 1991).

It is particularly noteworthy that this oxidative cyclization/desaturation process central to clavulanic acid (**3**) formation is most closely paralleled by the oxidative ring expansion of penicillin to cephalosporin, i.e., **7** → **8**, catalyzed by DAOCS. In the fungus *Cephalosporium acremonium* the ring expansion/hydroxylation reactions that relate penicillin N (**6**) to deacetylcephalosporin C (**8**) are carried out by a single enzyme (Scheidegger et al., 1984; Baldwin et al., 1987; Dotzlaf & Yeh, 1987). However, in the bacterium *S. clavuligerus*, two quite specialized α-KG-linked oxygenases can be separated, each catalyzing one reaction of the two-step process, but its complement with only about 2% efficiency (Baker et al., 1991; Kovacevic & Miller, 1991). In contrast, this degree of enzymic specialization is not seen for the two clavaminate synthases in the same organism. A great deal of sequence homology exists among the fungal DAOCS synthase/hydroxylase gene and the two bacterial DAOCS and DACS genes (Kovacevic & Miller, 1991). Apparent similarities among these genes and the *S. clavuligerus* isopenicillin N synthase (IPNS) gene also have been noted (Kovacevic & Miller, 1991). Thus one aim of our experiments was to determine whether the mechanistic parallels between DAOCS and CS were reflected in homology between the protein sequences of these enzymes. If an evolutionary relation existed between these two pathways, one would expect it to be most evident in these two enzymes.

EXPERIMENTAL PROCEDURES

Materials. Restriction endonucleases, calf alkaline phosphatase, T4 DNA ligase, and T4 polynucleotide kinase were purchased from Bethesda Research Laboratories. Endoproteinase Lys-C was from Boehringer, and modified T7 DNA polymerase (Sequenase) was supplied by United States Biochemical Corp. [γ-^{32}P]ATP and [α-^{35}S]dATP were obtained from Amersham International. Bluescript plasmid vectors and helper phage VCS M13 were supplied by Stratagene. Nylon 66 membranes were from Hoeffer Scientific. Oligonucleotides were custom synthesized on an Applied Biosystems 380B DNA synthesizer at the Protein/Peptide/DNA Facility, Department of Biological Chemistry, The Johns Hopkins Medical School, and purified by reverse-phase HPLC (Aquapore Octyl, 250 × 10 mm, 20 μm; Brownlee) using a buffer of 100 mM triethylammonium acetate and a 9–16% gradient of acetonitrile. A cosmid library of *S. clavuligerus* genomic DNA cloned in the vector pLAFR3 was the kind gift of Dr. Susan Jensen of the University of Alberta.

N-Terminal Sequence Determination of CS1 and Purification and Sequence Determination of Proteolytic Fragments of CS2. Clavaminate synthase was purified, and the two enzyme forms, CS1 and CS2, were separated as described previously (Salowe et al., 1990). CS1 was further purified and desalted by reverse-phase chromatography on a Vydac C_4 column equilibrated in 0.1% trifluoroacetic acid. The protein was eluted with an increasing gradient of acetonitrile containing 0.1% trifluoroacetic acid and then subjected to automated sequence analysis on an Applied Biosystems 470A gas-phase sequencer. CS2, after purification as for CS1, failed to give reliable sequence data, and so limited proteolysis of the protein was undertaken. CS2 (1 mg) was dialyzed against 0.5% NH_4HCO_3 at 4 °C for several hours to remove buffer salts before being lyophilized. The protein was then dissolved

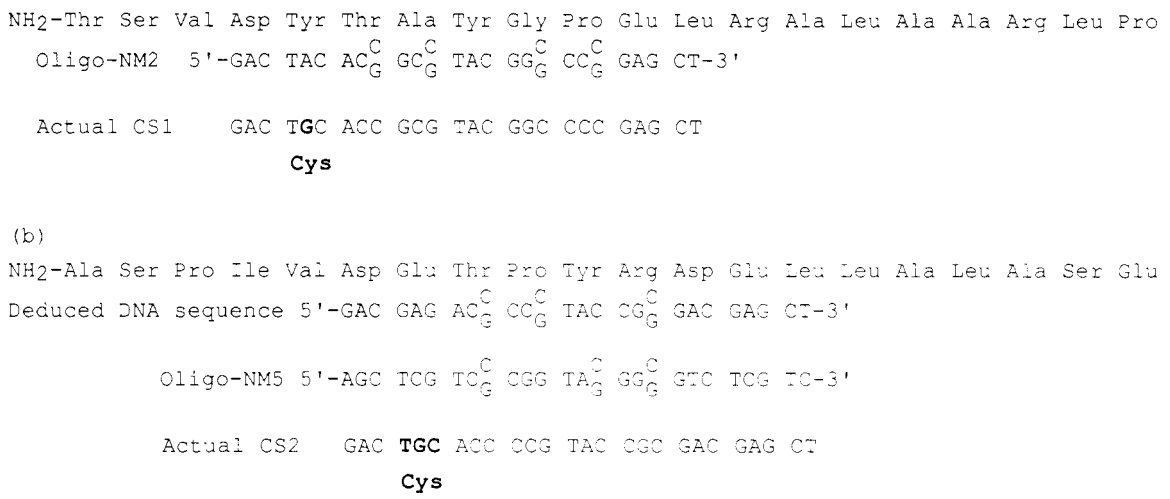


FIGURE 1: (a) N-Terminal amino acid sequence determined for CS1 in two trials. Partially degenerate 26-mer oligonucleotide probe NM2 was synthesized to the indicated portion using only G/C bases at positions of ambiguity. (b) Amino acid sequence obtained in two runs for a Lys-C proteolytic fragment of CS2. The deduced DNA sequence favoring G/C bases at codon positions of ambiguity for a portion of the peptide sequence is shown. Probe NM5 was synthesized as the reverse complement of this sequence.

in 50 μL of 0.4 M NH_4HCO_3 containing 8 M urea, dithiothreitol was added to 4 mM, and the mixture incubated at 50 °C for 15 min. Cysteinyl residues were then amidomethylated by addition of iodoacetamide to 8 mM at 25 °C for 15 min. The mixture was then diluted to 200 μL with water, and 10 μg of endoproteinase Lys-C was added. Digestion was allowed to proceed at 37 °C for 2 h, after which a further 10 μg of Lys-C was added and digestion continued overnight. The peptides were separated on a Vydac C_4 column using an increasing gradient of acetonitrile in the presence of 0.1% trifluoroacetic acid to elute them. Purified peptides were then subjected to automated sequencing as above.

For HPLC comparison of the proteolytic digests of CS1 and CS2, 1 mL (≈0.5 mg) samples of each from Mono-Q chromatography were lyophilized and amidomethylated as described above for CS2. The protein was divided into two 100-μL portions, and Lys-C endoproteinase was added (substrate:enzyme 50:1). After incubation at 37 °C for 24 h, the fragments were analyzed by HPLC on a Vydac C_4 column equilibrated with 0.1% trifluoroacetic acid and eluted using a linear gradient of 0–60% acetonitrile.

Screening of the S. clavuligerus Library. A library of *S. clavuligerus* genomic DNA fragments (20–25 kbp in size), ligated into the *Bam*H1 site of the cosmid vector pLAFR3 and packaged as *pseudo*phage, was transfected into *Escherichia coli* strain VCS 257 according to the supplier's protocol (Stratagene). Recombinant colonies were selected for by plating on Nylon 66 membranes overlying 2 × TY plates containing 12.5 μg/mL tetracycline. Colonies were replica-plated onto fresh nylon membranes which were then prepared for hybridization as described (Maas, 1983). Positive colonies were identified by hybridization to oligonucleotide probes (see Figure 1) which were end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase to a specific activity of approximately 10^8 cpm/μg (Maniatis et al., 1982). Hybridizations were performed using standard protocols (Ausabel et al., 1987) at 55 °C in a solution containing 2.5 × SSC buffer, 0.1% SDS, 1 × Denhardt's solution, and 0.05% $Na_4P_2O_7$ for 60 h. Final washing of the membranes was at 63 °C in 2.5 × SSC buffer for 30 min.

Nucleotide Sequencing and Analysis. Specific restriction fragments were subcloned into the plasmid vector pBS$^+$ by standard methods. Single-stranded DNA for sequencing was obtained from these subclones by rescue from plasmid-bearing cells with "helper" phage VCS M13 according to the supplier's instructions (Stratagene). Plasmid DNA for double-stranded sequencing was purified by the STET method (Sambrook et al., 1989) and denatured in 0.2 M NaOH and 0.2 mM EDTA. This solution was neutralized in 0.2 M NH_4OAc. DNA was sequenced by the chain-termination method of Sanger (Sanger et al., 1977), using either commercially available or custom-synthesized oligonucleotide primers and modified T7 DNA polymerase (Sequenase). Sequence data were compiled manually and analyzed using the DNA Strider program (Marck, 1988), and the codon usage was monitored using Brujene II (kindly provided by Professor C. R. Hutchinson, University of Wisconsin). Sequence comparisons were made using the BestFit, Gap, DotPlot, PileUp, and TFASTA programs available from the Genetics Computer Group (GCG) resident at the Pittsburgh Supercomputing Center using their default weighting factors, unless otherwise noted.

RESULTS

Clavaminate synthase from *S. clavuligerus* was resolved into two active protein components by Mono-Q (Pharmacia) chromatography, CS1 and CS2, of very similar molecular weights and kinetic properties (Salowe et al., 1990). The amino acid sequence was obtained of the N-terminus of the larger protein, CS1, by automated methods (Figure 1a, two trials). Similar attempts to obtain the N-terminal sequence of CS2 were unsuccessful. Consequently, the protein was digested with Lys-C endoproteinase, and the peptide fragments were separated by reverse-phase chromatography. Of three samples submitted for sequence analysis, only one gave reproducible data and credible stoichiometry (Figure 1b, two trials).

The two partial amino acid sequences were used to design oligonucleotide probes for hybridization experiments, each 26 bases in length. The degeneracy of these probes was minimized by recognizing that in *Streptomyces* structural genes G + C usage is preferred in the third position of codons typically with greater than 90% probability (Bibb et al., 1984, Wright & Bibb, 1992). With this bias in mind, partially degenerate probes NM2 and NM5 were synthesized by automated methods as shown in Figure 1. NM5 was prepared as the reverse complement of the deduced DNA sequence

with the idea that it might have been possible to use it subsequently with NM2 in a PCR experiment.

Approximately 1000 colonies from a cosmid library of *S. clavuligerus* DNA were screened using NM2 as a hybrid probe. Six positive colonies were identified and designated pCS1A to pCS1F. These were isolated and probed with NM5 to confirm the presence of the CS gene, but no hybridization was observed under stringent conditions. This was a surprising result because it was anticipated that CS2 was derived from CS1 by proteolysis, and hence the sequence for CS2 should be present within the gene for CS1. However, NM5 hybridized successfully to restriction digests of *S. clavuligerus* genomic DNA and, when used to screen the library, hybridized strongly to six different colonies, denoted pCS2A to pCS2F. These results were the first indication that CS1 and CS2 were the products of different genes, and they suggest that they may be separated by >20 kbp. As a further chemical test of dissimilarity, equal amounts of CS1 and CS2 were digested in parallel reactions with Lys-C endoproteinase and their C4 reverse-phase HPLC chromatograms compared. While some fragments appeared to have the same retention times, differing peak intensities and several nonsuperimposable peaks supported the view that these proteins were not the same (data not shown).

Plasmid DNA isolated from each of the two sets of clones identified by NM2 and NM5 was analyzed by endonuclease digestion with *Eco*RI and *Hin*dIII. Of these, pCS1A was found to contain an approximately 25 kbp insert which included a 7.4 kbp *Eco*RI fragment derived from a genomic restriction site and the polylinker of the vector. This fragment encompassed the binding site for NM2 and was cloned into pBS$^+$ (Stratagene) to give the construct pCS1A1 (Figure 2a). pCS2A was found to contain an approximately 28 kbp insert comprising in part a 8.4 kbp *Eco*RI fragment that hybridized to NM5. This restriction fragment was cloned into pBS$^+$ to give construct pCS2A1 (Figure 2b). Further restriction mapping established more precisely the position of the *cs1* and *cs2* genes within pCS1A1 and pCS2A1 and confirmed that the entire gene was present in each case.

The DNA sequences of the *cs1* and *cs2* genes were determined by both double- and single-stranded sequencing methods using pCS1A1 and pCS2A1 and subclones derived from these plasmids as templates. The sequence was determined from both strands over the entire coding region using "universal" and custom-synthesized oligonucleotide primers to extend the sequence. The sequencing strategy is outlined in Figure 2. *Cs1* comprises 325 codons (975 nucleotides) and encodes a protein of M_r 35 347, while *cs2* is 326 codons (978 nucleotides) and encodes a protein of M_r 35 774. Potential ribosome binding sites could be identified approximately 10 nucleotides upstream from the start codon for each gene. The reading frames for both genes were confirmed by analyzing the incidence of G + C usage in the third position of each codon (Bibb, 1984; Wright & Bibb, 1992). As expected, a high bias toward G or C (>90%) was seen throughout the coding region of each gene. Assigning the start codon for the *cs2* gene was aided by its extensive similarity (87% identity) to the *cs1* gene, for which the N-terminal sequence had been determined by protein sequencing. Comparison of the deduced amino acid sequence for the *cs2* gene with that of the peptide isolated from the Lys-C digest of CS2 revealed that, in fact, this peptide corresponded to the N-terminus of the protein. It was also apparent from the DNA sequence that cysteine-6 in *cs1* and cysteine-8 in *cs2* had been incorrectly assigned from the protein sequencing data (Figures 1 and 2). This

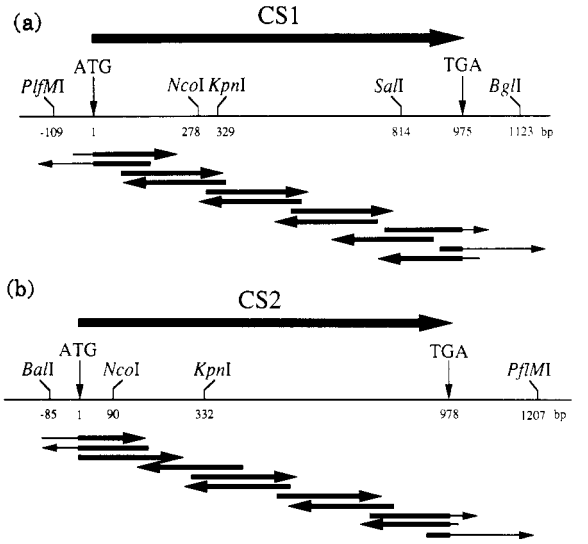


FIGURE 2: Partial restriction maps of *cs1* and *cs2* contained within the constructs (a) pCS1A1 and (b) pCS2A1, respectively. The sequencing strategy used for these genes and approximate probe locations and length of sequence obtained is shown by heavy arrows for each strand. The light arrows show the continued determination of sequence into flanking regions.

resulted in one mismatched nucleotide in NM2 and three mismatched nucleotides in NM5; fortunately, these errors did not prevent the oligonucleotides from hybridizing to their intended target sequences under the conditions used.

DISCUSSION AND CONCLUSIONS

Prior reports from this laboratory have described the resolution of clavaminate synthase (CS) into two active components, CS1 and CS2, by Mono-Q chromatography (Salowe et al., 1990). The N-terminal amino acid sequence was successfully obtained from CS1, but resort to enzymic proteolysis was necessary to secure reliable protein sequence data for CS2. Partially degenerate 26 bp oligonucleotide probes were designed to each peptide using only G or C bases at positions of ambiguity in the respective codons in keeping with the bias for these bases empirically observed in *Streptomyces* coding regions (Bibb, 1984; Wright & Bibb, 1992). To our surprise, each oligonucleotide recognized a different set of clones, suggesting that two genes had been identified and, hence, two distinct clavaminate synthases might be independently produced in *S. clavuligerus*. This conclusion was supported by differences in peak intensities and retention times in HPLC profiles of parallel Lys-C endoproteinase digestions of CS1 and CS2. The existence of two isoenzymes was further substantiated when the nucleotide sequences of both genes were compared and securely established when each was separately expressed as a catalytically active protein of the correct size as determined by SDS–PAGE (Townsend et al., unpublished).

Considerable attention was given to the DNA sequencing and codon usage of these two genes as earlier SDS–PAGE characterization of the highly purified CS1 and CS2 proteins had indicated molecular weights of 47 000 and 46 000 ± 2000, respectively (Salowe et al., 1990). While these values are substantially higher than those deduced from the gene sequences, a molecular weight of 38 000 ± 5000 was also determined earlier from CS1 retention on a calibrated Superose-12 column. This latter figure is more in keeping with the sequencing information. The region upstream of the initiation codon for each gene showed no sequence similarity with *Streptomyces* promotor sequences (Kovacevic et al., 1990;

(a)
```
-30  /   1                                          1   /  11
CGA GGA ACC AAG GAA AGG AAC CAC ACC ACC ATG ACC TCA GTG GAC TGC ACC GCG TAC GGC
                                        met thr ser val asp cys thr ala tyr gly
31   /  21                                         61  /  31
CCC GAG CTG CGC GCG CTC GCC GCC CGG CTG CCC CGG ACC CCC CGG GCC GAC CTG TAC GCC
pro glu leu arg ala leu ala ala arg leu pro arg thr pro arg ala asp leu tyr ala
91   /  41                                         121 /  51
TTC CTG GAC GCC GCG CAC ACA GCC GCC GCC TCG CTC CCC GGC GCC CTC GCC ACC GCG CTG
phe leu asp ala ala his thr ala ala ala ser leu pro gly ala leu ala thr ala leu
151  /  61                                         181 /  71
GAC ACC TTC AAC GCC GAG GGC AGC GAG GAC GGC CAT CTG CTG CTG CGC GGC CTC CCG GTG
asp thr phe asn ala glu gly ser glu asp gly his leu leu leu arg gly leu pro val
211  /  81                                         241 /  91
GAG GCC GAC GCC GAC CTC CCC ACC ACC CCG AGC AGC ACC CCG GCG CCC GAG GAC CGC TCC
glu ala asp ala asp leu pro thr thr pro ser ser thr pro ala pro glu asp arg ser
271  / 101                                        301 / 111
CTG CTG ACC ATG GAG GCC ATG CTC GGA CTG GTG GGC CGC CGG CTC GGT CTG CAC ACG GGG
leu leu thr met glu ala met leu gly leu val gly arg arg leu gly leu his thr gly
331  / 121                                        361 / 131
TAC CGG GAG CTG CGC TCG GGC ACG GTC TAC CAC GAC GTG TAC CCG TCG CCC GGC GCG CAC
tyr arg glu leu arg ser gly thr val tyr his asp val tyr pro ser pro gly ala his
391  / 141                                        421 / 151
CAC CTG TCC TCG GAG ACC TCC GAG ACG CTG CTG GAG TTC CAC ACG GAG ATG GCC TAC CAC
his leu ser ser glu thr ser glu thr leu leu glu phe his thr glu met ala tyr his
451  / 161                                        481 / 171
CGG CTC CAG CCG AAC TAC GTC ATG CTG GCC TGC TCC CGG GCC GAC CAC GAG CGC ACG GCG
arg leu gln pro asn tyr val met leu ala cys ser arg ala asp his glu arg thr ala
511  / 181                                        541 / 191
GCC ACA CTC GTC GCC TCG GTC CGC AAG GCG CTG CCC CTG CTG GAC GAG AGG ACC CGG GCC
ala thr leu val ala ser val arg lys ala leu pro leu leu asp glu arg thr arg ala
571  / 201                                        601 / 211
CGG CTC CTC GAC CGG AGG ATG CCC TGC TGC GTG GAT GTG GCC TTC CGC GGC GGG GTG GAC
arg leu leu asp arg arg met pro cys cys val asp val ala phe arg gly gly val asp
631  / 221                                        661 / 231
GAC CCG GGC GCC ATC GCC CAG GTC AAA CCG CTC TAC GGG GAC GCG GAC GAT CCC TTC CTC
asp pro gly ala ile ala gln val lys pro leu tyr gly asp ala asp asp pro phe leu
691  / 241                                        721 / 251
GGG TAC GAC CGC GAG CTG CTG GCG CCG GAG GAC CCC GCG GAC AAG GAG GCC GTC GCC GCC
gly tyr asp arg glu leu leu ala pro glu asp pro ala asp lys glu ala val ala ala
751  / 261                                        781 / 271
CTG TCC AAG GCG CTC GAC GAG GTC ACG GAG GCG GTG TAT CTG GAG CCC GGC GAT CTG CTG
leu ser lys ala leu asp glu val thr glu ala val tyr leu glu pro gly asp leu leu
811  / 281                                        841 / 291
ATC GTC GAC AAC TTC CGC ACC ACG CAC GCG CGG ACG CCG TTC TCG CCC CGC TGG GAC GGG
ile val asp asn phe arg thr thr his ala arg thr pro phe ser pro arg trp asp gly
871  / 301                                        901 / 311
AAG GAC CGC TGG CTG CAC CGC GTC TAC ATC CGC ACC GAC CGC AAT GGA CAG CTC TCC GGC
lys asp arg trp leu his arg val tyr ile arg thr asp arg asn gly gln leu ser gly
931  / 321                                        961 / 331
GGC GAG CGC GCG GGC GAC GTC GTC GCC TTC ACA CCG CGC GGC TGA
gly glu arg ala gly asp val val ala phe thr pro arg gly OPA
```

Petrich et al., 1992), and at present it is not clear whether the genes are transcribed independently or as part of a larger polycistronic messenger RNA. However, a potential ribosomal binding site could be identified within GGAAAGGAA at −19 to −11 bases from the start codon of the *cs1* gene. Analogously, the sequence GAAGGAG was found −17 to −10 bases upstream of the start codon for the *cs2* gene (Figure 3). These runs of purines are quite large and may account in part for the relatively high expression of these biosynthetic genes in vivo (Salowe et al., 1990). Interestingly, the last 15 bases of the *cs1* gene ending at the TGA stop codon are repeated just downstream out of frame. The role, if any, of this duplication is unclear.

Alignment of the *cs1* and *cs2* DNA sequences reveals a greater than 85% identity between the two (Gap and BestFit analyses). The homology is most extensive between the central portions of the genes, while somewhat greater variation is seen toward the N- and C-termini. This pattern of sequence variation was also observed in comparisons of the fungal and bacterial DAOCS/DACS genes [data not shown; see also Kovacevic and Miller (1991)]. When *cs1* and *cs2* were translated, their amino acid sequences showed an 82% identity and 87% similarity (Figure 4 and Table I). Given the extent of homology between the *cs1* and *cs2* genes, had the peptide fragment isolated in the proteolysis of CS2 been from the center of the protein, it is questionable whether discrimination between the two sets of *cs* clones would have been achieved. The isolation of the N-terminal peptide from the Lys-C digest of CS2 proved extremely fortunate as its sequence is sufficiently different from CS1 and the misidentified cysteine for the oligonucleotide probes designed from these sequences not to cross-hybridize. These results raise the possibility that isoenzymes may occur quite commonly in bacterial secondary metabolic pathways but that the strategies used to clone these genes have resulted in their so far having been overlooked.

Gene amplification in bacteria occurs relatively frequently and has been considered as the first step in divergent evolution (Rigby, 1974). Experiments in chemostats have shown that

(b)

```
-30  /   1                                   1   /  11
AGA ACC CAG TTG TGA AGG AGA CAT CGT GTC ATG GCC TCT CCG ATA GTT GAC TGC ACC CCG
arg thr gln leu OPA arg arg his arg val met ala ser pro ile val asp cys thr pro
31   /  21                                  61  /  31
TAC CGC GAC GAG CTG CTC GCG CTC GCC TCC GAG CTT CCC GAG GTG CCG CGC GCG GAC CTC
tyr arg asp glu leu leu ala leu ala ser glu leu pro glu val pro arg ala asp leu
91   /  41                                  121 /  51
CAT GGC TTC CTC GAC GAG GCG AAG ACG CTG GCC GCC CGT CTC CCG GAG GGG CTG GCC GCC
his gly phe leu asp glu ala lys thr leu ala ala arg leu pro glu gly leu ala ala
151  /  61                                  181 /  71
GCT CTC GAC ACC TTC AAC GCC GTG GGC AGC GAG GAC GGT TAT CTG CTG CTG CGC GGG CTG
ala leu asp thr phe asn ala val gly ser glu asp gly tyr leu leu leu arg gly leu
211  /  81                                  241 /  91
CCC GTC GAC GAC AGC GAG CTG CCC GAG ACG CCG ACC TCC ACC CCG GCC CCG CTG GAC CGC
pro val asp asp ser glu leu pro glu thr pro thr ser thr pro ala pro leu asp arg
271  / 101                                 301 / 111
AAG CGG CTG GTG ATG GAG GCC ATG CTC GCG CTG GCC GGC CGC CGG CTC GGT CTG CAC ACG
lys arg leu val met glu ala met leu ala leu ala gly arg arg leu gly leu his thr
331  / 121                                 361 / 131
GGG TAC CAG GAG CTG CGC TCG GGC ACG GTC TAC CAC GAC GTG TAC CCG TCG CCC GGC GCG
gly tyr gln glu leu arg ser gly thr val tyr his asp val tyr pro ser pro gly ala
391  / 141                                 421 / 151
CAC TAC CTG TCC TCG GAG ACC TCC GAG ACG CTG CTG GAG TTC CAC ACG GAG ATG GCG TAC
his tyr leu ser ser glu thr ser glu thr leu leu glu phe his thr glu met ala tyr
451  / 161                                 481 / 171
CAC ATC CTC CAG CCG AAC TAC GTC ATG CTG GCC TGC TCC CGC GCG GAC CAC GAG AAC CGG
his ile leu gln pro asn tyr val met leu ala cys ser arg ala asp his glu asn arg
511  / 181                                 541 / 191
GCG GAG ACG CTG GTC GGC TCG GTC CGC AAG GCG CTG CCC CTG CTG GAC GAG AAG ACC CGG
ala glu thr leu val gly ser val arg lys ala leu pro leu leu asp glu lys thr arg
571  / 201                                 601 / 211
GCC CGT CTC TTC GAC CGC AAG GTG CCC TGC TGC GTG GAC GTG GCC TTC CGC GGC GGG GTC
ala arg leu phe asp arg lys val pro cys cys val asp val ala phe arg gly gly val
631  / 221                                 661 / 231
GAC GAC CCG GGC GCG ATC GCC AAC GTC AAG CCG CTC TAC GGG GAC GCG AAC GAC CCG TTC
asp asp pro gly ala ile ala asn val lys pro leu tyr gly asp ala asn asp pro phe
691  / 241                                 721 / 251
CTC GGG TAC GAC CGC GAG CTG CTG GCG CCG GAG GAC CCC GCG GAC AAG GAG GCC GTC GCC
leu gly tyr asp arg glu leu leu ala pro glu asp pro ala asp lys glu ala val ala
751  / 261                                 781 / 271
CAT CTG TCC CAG GCG CTC GAC GAT GTG ACC GTC GGG GTG AAG CTC GTC CCC GGT GAC GTC
his leu ser gln ala leu asp asp val thr val gly val lys leu val pro gly asp val
811  / 281                                 841 / 291
CTC ATC ATC GAC AAC TTC CGC ACC ACG CAC GCG CGG ACG CCG TTC TCG CCC CGC TGG GAC
leu ile ile asp asn phe arg thr thr his ala arg thr pro phe ser pro arg trp asp
871  / 301                                 901 / 311
GGG AAG GAC CGC TGG CTG CAC CGC GTC TAC ATC CGC ACC GAC CGC AAT GGA CAG CTC TCC
gly lys asp arg trp leu his arg val tyr ile arg thr asp arg asn gly gln leu ser
931  / 321                                 961 / 331
GGC GGC GAG CGC GCG GGC GAC ACC ATC TCG TTC TCG CCG CGC CGC TGA
gly gly glu arg ala gly asp thr ile ser phe ser pro arg arg OPA
```

FIGURE 3: Nucleotide sequences of the (a, opposite page) *cs1* and (b, above) *cs2* genes. The corresponding amino acid assignments are indicated under each codon. Potential ribosome binding sites are underlined. The actual DNA sequence for the region of probe hybridization is shown by an overline for each gene. One incorrect nucleotide is evident in NM2 at Cys-6 in CS1, and three mismatches are evident in NM5 at Cys-8 in CS2 as shown in Figure 1.

bacteria grown on poorly metabolized carbon sources typically respond by duplication of catabolic genes rather than by mutations of the genes themselves (Rigby, 1974; Mortlock, 1982; Anderson & Roth, 1977; Eisenstark, 1977; Stark, 1984). An interesting exception to this pattern was found in an analogous metabolic selection for mandelate racemase in *Pseudomonas aeruginosa*. In the absence of mutagen, successful tranformants were observed to produce 30 times more racemase than wild-type. Neither gene duplication nor modification of the structural gene had occurred, but a single transversion 87 bp upstream from the initiation codon was discovered in each of three cases in a region thought to be the promoter (Tsou et al., 1989).

In secondary metabolic pathways, recent DNA sequencing studies have revealed that gene duplication is a prominent feature. Examples include the genes responsible for the biosynthesis of polyketide metabolites as actinorhodin, granaticin, tetracenomycin, daunorubicin (Hopwood & Sherman, 1990; Gramajo et al., 1991; Stutzman-Engwall et al., 1992) and 6-deoxyerythronolide B (Cortes et al., 1990; Donadio et al., 1991; Bevitt et al., 1992; Donadio & Katz, 1992), and nonribosomal bacterial peptides as, for example, cyclosporin (Lawen & Zocher, 1990; Kleinkauf & von Döhren, 1990). However, to our knowledge, this is the first report of a duplication event that leads to independent genes, potentially separated on the chromosome by >20 kbp, which are coexpressed as active enzymes in vivo.

Typically, genes involved in the biosynthesis of secondary metabolites in bacteria have been found clustered on the chromosome or a plasmid (Kirby & Hopwood, 1977; Martin

& Liras, 1989). For the present case of two independently synthesized *cs* gene products, it will be of interest to establish whether more of the biosynthetic pathway has been reproduced at these two locations in the genome or whether only this single gene has been duplicated. It might be thought that since *S. clavuligerus* also produces clavam metabolites of general structure **4**, one of the *cs* genes may have come from a separate metabolic pathway involved in the biosynthesis of these clavam products. Recent work, however, has established that clavam-2-carboxylate (**4**, R = COOH) and clavulanic acid (**3**), despite important structural and stereochemical differences, share a common biosynthetic route to the advanced intermediate proclavaminic acid (**1**) and perhaps beyond (Iwata-Reuyl & Townsend, 1992). Therefore, it might be expected that these pathways would share the agency of a single enzyme rather than two in accord with the general view that biosynthetic pathways have early steps in common and diverge only toward the end to give individual products.

Alternatively, the oxidative cyclization/desaturation of proclavaminate (**1**) to clavaminate (**2**) may be rate-limiting in vivo during the production of clavulanic acid. It may be that a second copy of this gene was generated to make this key protein available in larger amounts to enhance this conversion in keeping with the usual pattern of bacterial gene amplification. We have observed in wild-type *S. clavuligerus* that proclavaminic acid (**1**) accumulates on the order of 200 μM concentrations (Salowe and Townsend, unpublished) despite the presence of comparatively large amounts of CS (Salowe et al., 1990). The reasons for the inefficiency of this transformation in vivo are not fully clear but appear to stem in part from the relatively short half-life of CS itself owing to its frequency of oxidative self-inactivation (Salowe et al., 1990). The improvement of metabolic flux at a kinetically limiting step in the biosynthesis of a valuable secondary metabolite is a noteworthy rationale for gene duplication. Such a strategy has been successfully applied artificially in *C. acremonium* by a group at Eli Lilly to improve the in vivo conversion of penicillin N (**6**) to cephalosporin C by 15% (Skatrud et al., 1989).

CS and DAOCS/DACS have close coreactant and mechanistic parallels in the oxidative cyclization/desaturation reactions they catalyze. Their co-occurrence in *S. clavuligerus*, and certainly in other bacteria that co-produce clavams and penams/cephams, suggests that they might well be evolutionarily related. The genes encoding CS1 and CS2 are 87% identical, and those for DAOCS and DACS in *S. clavuligerus* are 71% identical (Kovacevic & Miller, 1991). The extent of the former homology is shown graphically in a DotPlot of the CS1 and CS2 genes (Figure 4a). Under identical window but lower stringency conditions, CS1 and DAOCS are compared in Figure 4b. The degree of identity at the DNA and protein levels drops to 44% and 22%, respectively (Gap analysis); and not more than five identical nucleotides or two amino acids position adjacent to each other. In accord with this observation, comparisons among CS1, CS2, DAOCS, and DACS show no significant correlation across the two β-lactam biosynthetic pathways. Both the CS1/CS2 enzymes and the DAOCS/DACS enzymes have arisen by gene duplication followed by limited divergence. In the case of DAOCS and DACS, this has resulted in the enzymes catalyzing separate steps in the oxidative ring expansion and hydroxylation of isopenicillin N to deacetylcephalosporin C. Each enzyme is still able, however, to catalyze the other step to a limited extent. In contrast, the fungus *C. acremonium* has retained both activities in a single homologous protein so

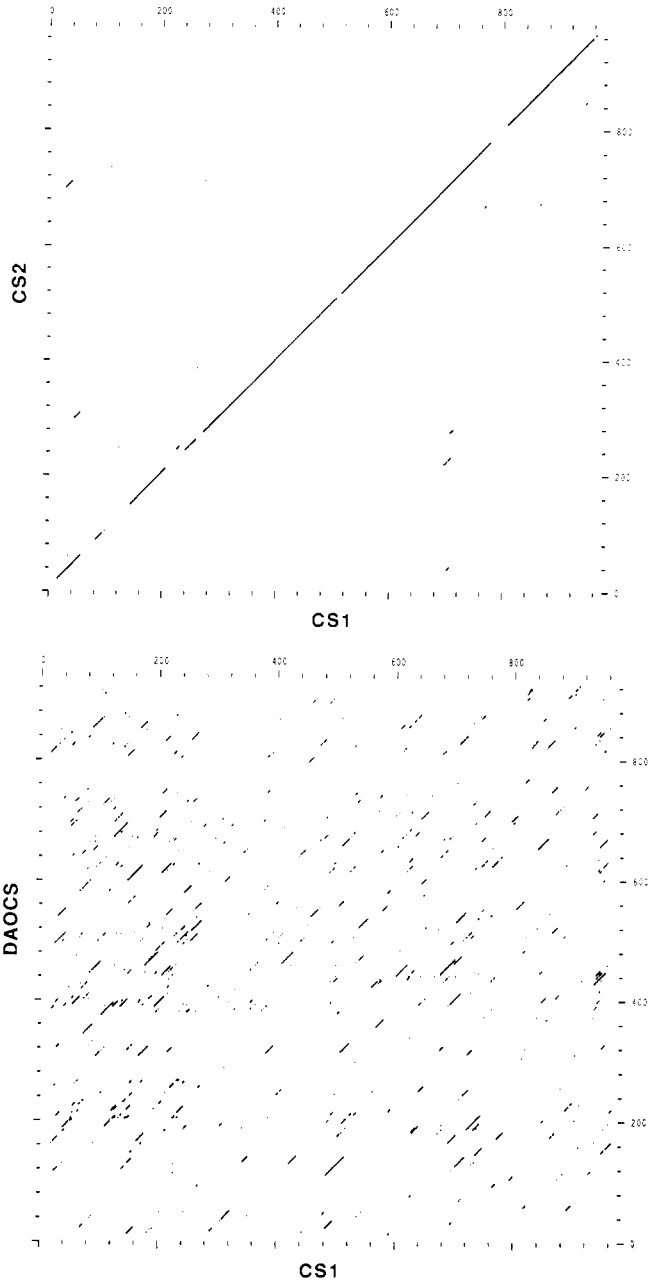


FIGURE 4: DotPlot comparisons between the genes encoding (a, top) CS1 and CS2 (window 25, stringency 18) and (b, bottom) CS1 and DAOCS (window 25, stringency 15).

that the significance of this duplication is unclear. It would be of interest to compare the net metabolic flux to cephalosporins imparted to each of these organisms by the presence of one enzyme or two functionally more specialized proteins.

Computer-aided sequence comparisons of CS1/CS2 and DAOCS/DACS from *S. clavuligerus* were made with other iron-dependent oxygenase enzymes. Thus, extensive homologies exist among enzymes apparently involved in ethylene formation from plants (see Table I). The ethylene-forming enzyme (EFE) from apple has been purified to homogeneity and found not to require α-KG but to utilize ascorbate stoichiometrically in the presence of Fe(II) and molecular oxygen [Dong et al., 1992a; see also Smith et al. (1992)]. It would appear that ascorbate substitutes for α-KG as an electron donor in the oxygen activation process catalyzed by this protein by a mechanism that is presumably representative of the other plant EFE enzymes [ethylene formation in bacteria appears to operate by a separate pathway (Chou & Yang, 1973; Fukuda et al., 1989)]. A preference for catalytic

amounts of ascorbate has been noted generally among α-KG-dependent oxygenases (England & Seifter, 1986), presumably to maintained the reduced ferrous oxidation state at the metal center, and in EFE enzymes from plants this preference evidently has been modified to recruit this relatively abundant plant carbohydrate to serve as a cofactor. On the other hand, perhaps those enzymes that require stoichiometric amounts of α-KG have evolved from simpler systems using ascorbate only, yet they retain vestigial affinity for ascorbate solely for its metal-reducing function.

Interestingly, the amino acid sequences of these enzymes, all quite probably involved in formation of the hormone ethylene, are somewhat distantly related to three plant enzymes that carry out more conventional hydroxylase reactions requiring α-KG as well as molecular oxygen and catalytic ferrous ion: hyoscyamine 6β-hydroxyase and two flavanone hydroxylases (H6H, Flav-3β Hy, and Flav-Hy; Table I). Indications of some of these similarities have been noted earlier (Matsuda et al., 1990) as well as an unexpected yet more distant relation to IPNS, an iron-dependent oxygenase that does not require α-KG or stochiometric ascorbate for activity (White et al., 1982). Limited homologies have also been suggested among IPNS and DAOCS/DACS in *C. acremonium* and *S. clavuligerus* (Kovacevic & Miller, 1991).

In contrast, the mammalian enzyme prolyl hydroxylase, which plays an important role in the maturation of collagen, appears unrelated to the oxygenases above. Even more surprising is that, while prolyl hydroxylase from human and chicken show strong homology (Prolyl Hy hum and Prolyl Hy chk; Table I), their companion enzyme functionally, lysyl hydroxylase (from chicken, Lysyl Hy chk; Table I), shows no obvious sequence relation. Finally, CS1 and CS2 in the present work reveal no significant homology to either of these collagen-related hydroxylases or to any of the plant and microbial enzymes of the first group above. Therefore, at least for the present limited data, it would appear that a minimum of four classes of α-KG-dependent oxygenases exist. The absence of homology among the four groups of enzymes requiring α-KG, the mechanistic complexity of oxygen activation and the diversity of the reactions catalyzed implies an impressive extent of functional convergence among these proteins.

Alternatively, these enzymes may tolerate very extensive mutation and yet retain the ability to carry out their synthetic tasks and adapt to a variety of mechanistic requirements. The evolutionary divergence in *S. clavuligerus* of CS1 and CS2 as well of DAOCS and DACS, though not great, could support this view. There is not enough sequence information presently available to distinguish these non-heme iron-dependent enzymes as a family of related genes or as an example of convergent evolution. It is interesting to note by comparison that cytochromes P-450 represent a genetic superfamily [for an overview, see Coon et al. (1992) and Nerbert and Gonzalez (1987)]. Each of these views about the origins of the α-KG-dependent enzymes and related proteins poses a number of interesting evolutionary and mechanistic questions.

Given their co-occurrence in *S. clavuligerus* and their mechanistic similarities, the α-KG-dependent enzymes central to clavulanic acid and cephalosporin formation present the closest chemical parallels currently identifiable between two β-lactam biosynthetic pathways. However, comparisons of the gene sequences for these proteins provide no convincing evidence that the two pathways are evolutionarily related. These comparisons and a growing body of chemical information point with increasing decisiveness to several separately evolved solutions to the task of β-lactam antibiotic synthesis in vivo

Table I: Amino Acid Sequence Comparisons (as Percent Identity) of Iron-Dependent Proteins Made Using Gap (GCG) Analysis

	CS2[a]	DACS[b]	DAOCS[c]	IPNS[d]	CARNATION[e]	EFE APPLE[f]	EFE PTOM13[g]	PAVOE3[h]	FLAV HY[i]	FLAV 3β HY[j]	H6H[k]	LYSYL CHK[l]	PROLYL HY HUM[m]	PROLYL HY CHK[n]
CS1[a]	82	21	21	14	18	19	18	16	15	16	17	22	17	20
CS2		18	14	19	17	22	17	17	13	20	16	19	22	17
DACS			58	24	19	20	18	18	22	21	19	24	16	16
DAOCS				25	17	20	20	25	20	20	21	25	20	19
IPNS					24	23	25	27	23	25	23	19	14	17
CARNATION						72	77	73	32	35	30	17	20	18
EFE APPLE							80	72	29	33	30	18	18	17
EFE PTOM13								76	31	36	31	19	19	16
PAVOE3									31	33	30	18	17	19
FLAV HY										72	30	13	14	18
FLAV 3β HY											31	21	17	15
H6H												20	22	23
LYSYL CHK													19	18
PROLYL HY HUM														89

PERCENT IDENTITY

0-25 26-40 41-65 66-100

[a] CS1 and CS2 from *S. clavuligerus* (this work); [b]DACS from *S. clavuligerus* (Kovacevic & Miller, 1991); [c]DAOCS from *S. clavuligerus* (Kovacevic et al., 1989); [d]IPNS from *S. clavuligerus* (Leskiw et al., 1988); [e]carnation senescence-related protein (Wang & Woodson, 1991); [f]EFE from apple (Dong et al., 1992b); [g]ripening-related protein from tomato (Holdsworth et al., 1987); [h]ripening-related protein from avocado (McGarvey et al., 1990); [i]flavanone-3-hydroxylase (Meldgaard, 1992); [j]flavanone-3β-hydroxylase (Britsch et al., 1992); [k]hyoscyamine 6β-hydroxylase (Matsuda et al., 1991); [l]lysyl hydroxylase from chicken (Myllylae et al., 1991); [m]prolyl hydroxylase from human (Helaakoski et al., 1989); [n]prolyl hydroxylase from chicken (Bassuk et al., 1989).

(Salowe et al., 1990). Although this behavior appears contrary to the increasingly appreciated importance of universal pathway steps in natural product biosynthesis, it underscores a determinant of perhaps wider importance than efficiency. The competitive advantage conferred on microorganisms by the ability to synthesize β-lactam antibiotics is sufficiently great that it has been achieved several times over by quite independent means. These biochemical successes have been preserved and occasionally shared. For example, *S. clavuligerus* contains at least two β-lactam antibiotic pathways, and *C. acremonium* is proposed to have acquired its ability to synthesize penicillin from bacteria (Weigel et al., 1988; Smith et al., 1990). It is notable that the breadth of sequence diversity among iron-dependent oxygenases and the flexibility of their catalytic chemistry has been similarly conscripted for independently developed but mechanistically related purposes in the biosyntheses of two classes of β-lactam antibiotics.

ACKNOWLEDGMENT

We are grateful to Dr. Shenbagamurthi of the Protein/Peptide/DNA Facility of the Johns Hopkins Medical School for obtaining amino acid sequence data from samples of purified CS1 and CS2. Oligonucleotide syntheses were carried out by Dr. Shenbagamurthi and J. Franklin of this laboratory and by S. Morse of the Department of Chemistry, whom we thank. We especially wish to acknowledge S. E. Jensen (University of Alberta) for providing a library of *S. cla-*

vuligerus DNA, and Z.-J. Chuang (Johns Hopkins) for his continuing advice. Gene and protein sequence comparisons were made at the Pittsburgh Supercomputer Center supported by the National Institutes of Health.

REFERENCES

Anderson, R. P., & Roth, J. R. (1977) *Annu. Rev. Microbiol. 31*, 473–505.

Ausabel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

Baker, B. J., Dotzlaf, J. E., & Yeh, W.-K. (1991) *J. Biol. Chem. 266*, 5087–5093.

Baldwin, J. E., Adlington, R. M., Coates, J. B., Crabbe, M. J. C., Crouch, N. P., Keeping, J. W., Knight, G. C., Schofield, C. J., Ting, H.-H., Vallejo, C. A., Thorniley, M., & Abraham, E. P. (1987) *Biochem. J. 245*, 831–841.

Basak, A., Salowe, S. P., & Townsend, C. A. (1990) *J. Am. Chem. Soc. 112*, 1654–1656.

Bassuk, J. A., Kao, W. W.-Y., Herzer, P., Kedersha, N., Seyer, J., DeMartino, J. A., Daugherty, B. L., Mark, G. E., III, & Berg, R. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7382–7386.

Bevitt, D. J., Cortes, J., Haydock, S. F., & Leadlay, P. F. (1992) *Eur. J. Biochem. 204*, 39–49.

Bibb, M. J., Findlay, P. R., & Johnson, M. W. (1984) *Gene 30*, 157–166.

Britsch, L., Ruhnau, B., & Forkmann, G. (1992) *J. Biol. Chem. 267*, 5380–5387.

Brown, D., Evans, J. R., & Fletton, R. A. (1979) *J. Chem. Soc., Chem. Commun.*, 282–283.

Chou, T. W., & Yang, S. F. (1973) *Arch. Biochem. Biophys. 157*, 73–82.

Coon, M. J., Ding, X., Pernecky, S. J., & Vaz, A. D. N. (1992) *FASEB J. 6*, 669–673.

Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., & Leadlay, P. F. (1990) *Nature 348*, 176–178.

Donadio, S., & Katz, L. (1992) *Gene 111*, 51–60.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., & Katz, L. (1991) *Science 252*, 675–679.

Dong, J. G., Fernandez-Maculet, J. C., & Yang, S. F. (1992a) *Proc. Natl. Acad. Sci. U.S.A. 20*, 9789–9793.

Dong, J. G., Olson, D., Silverstone, A., & Yang, S. F. (1992b) *Plant Physiol. 98*, 1530–1531.

Dotzlaf, J. E., & Yeh, W.-K. (1987) *J. Bacteriol. 169*, 1611–1618.

Dotzlaf, J. E., & Yeh, W.-K. (1989) *J. Biol. Chem. 264*, 10219–10227.

Eisenstark, A. (1977) *Annu. Rev. Genetics 11*, 369–396.

Elson, S. W., Baggaley, K. H., Gillett, J., Holland, S., Nicholson, N. H., Sime, J. T., & Woroniecki, S. R. (1987) *J. Chem. Soc., Chem. Commun.*, 1736–1738.

Englard, S., & Seifter, S. (1986) *Annu. Rev. Nutr. 6*, 365–406.

Englard, S., Blanchard, J. S., & Midelfort, C. F. (1985) *Biochemistry 24*, 1110–1116.

Fujita, Y., Gottlieb, A., Peterkofsky, B., Udenfriend, S., & Witkop, B. (1964) *J. Am. Chem. Soc. 86*, 4709–4716.

Fukuda, H., Kitajima, H., Fujii, T., Tazaki, M., & Ogawa, T. (1989) *FEMS Microbiol. Lett. 59*, 1–6.

Gramajo, H. C., White, J., Hutchinson, C. R., & Bibb, M. J. (1991) *J. Bacteriol. 173*, 6475–6483.

Hanauske-Abel, H. M., & Günzler, V. (1982) *J. Theor. Biol. 94*, 421–455.

Helaakoski, T., Vuori, K., Myllylae, R., Kivirikko, L., Pihlajaniemi, T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4392–4396.

Holdsworth, M. J., Bird, C. R., Ray, J., Schuch, W., & Grierson, D. (1987) *Nucleic Acids Res. 15*, 731–739.

Hook, D. J., Chang, L. T., Elander, R. P., & Morin, R. B. (1979) *Biochem. Biophys. Res. Commun. 87*, 258–265.

Hopwood, D. A., & Sherman, D. H. (1990) *Annu. Rev. Genet. 24*, 37–66.

Howarth, T. T., Brown, A. G., & King, T. J. (1976) *J. Chem. Soc., Chem. Commun.*, 266–267.

Iwata-Reuyl, D., & Townsend, C. A. (1992) *J. Am. Chem. Soc. 114*, 2762–2763.

Jensen, S. E., Westlake, D. W. S., & Wolfe, S. (1985) *J. Antibiot. 38*, 953–958.

Kirby, D., & Hopwood, D. A. (1977) *J. Gen. Microbiol. 98*, 239–252.

Kleinkauf, H., & von Döhren, H. (1990) *Eur. J. Biochem. 192*, 1–15.

Kovacevic, S., & Miller, J. R. (1991) *J. Bacteriol. 173*, 398–400.

Kovacevic, S., Weigel, B. J., Tobin, M. B., Ingolia, T. D., & Miller, J. R. (1989) *J. Bacteriol. 171*, 754–760.

Kovacevic, S., Tobin, M. B., & Miller, J. R. (1990) *J. Bacteriol. 172*, 3952–3958.

Lawen, A., & Zocher, R. (1990) *J. Biol. Chem. 265*, 11355–11360.

Leskiw, B. K., Aharonowitz, Y., Mevarech, M., Wolfe, S., Vining, L. C., Westlake, D. W. S., & Jensen, S. E. (1988) *Gene 62*, 187–196.

Maas, R. (1983) *Plasmid 10*, 296–298.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marck, C. (1988) *Nucleic Acids Res. 16*, 1829–1836.

Martin, J. F., & Liras, P. (1989) *Annu. Rev. Microbiol. 43*, 173–206.

Matsuda, J., Okabe, S., Hashimoto, T., & Yamada, Y. (1991) *J. Biol. Chem. 266*, 9460–9464.

McGarvey, D. J., Yu, H., & Christoffersen, R. E. (1990) *Plant Mol. Biol. 15*, 165–167.

Meldgaard, M. (1992) *Theor. Appl. Genet. 83*, 695–706.

Menssen, A., Höhmann, S., Martin, W., Schnable, P. S., Peterson, P. A., Saedler, H., & Gierl, A. (1990) *EMBO J. 9*, 3051–3057.

Mortlock, R. P. (1982) *Annu. Rev. Microbiol. 36*, 259–284.

Myllylae, R., Pihlajaniemi, T., Pajunen, L., Turpeenniemi-Hujanen, T., & Kivirikko, K. T. (1991) *J. Biol. Chem. 266*, 2805–2810.

Nagarajan, R., Boeck, L. D., Gorman, M., Hamill, R. L., Hoehn, M. M., Stark, W. M., & Whiney, J. C. (1971) *J. Am. Chem. Soc. 93*, 2308–2310.

Nebert, D. W., & Gonzalez, F. J. (1987) *Annu. Rev. Biochem. 56*, 945–993.

Rigby, P. W. J., Burleigh, B. D., & Hartley, B. S. (1974) *Nature 251*, 200–204.

Rollins, M. J., Westlake, D. W. S., Wolfe, S., & Jensen, S. E. (1988) *Can. J. Microbiol. 34*, 1196–1202.

Salowe, S., Marsh, E. N., & Townsend, C. A. (1990) *Biochemistry 29*, 6499–6508.

Salowe, S. P., Krol, W. J., Iwata-Reuyl, D., & Townsend, C. A. (1991) *Biochemistry 30*, 2281–2292.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, G., Nicklin, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5617–5621.

Scheidegger, A., Kienzi, M. T., & Niesch, J. (1984) *J. Antiobiot. 37*, 522–531.

Siegel, B. (1979) *Bioorg. Chem. 8*, 219–226.

Skatrud, P. L., Tietz, A. J., Ingolia, T. D., Cantwell, C. A., Fisher, D. F., Chapman, J. L., & Queener, S. W. (1989) *Biotechnology 7*, 477–485.

Smith, D. J., Burnham, M. K. R., Bull, J. H., Hodgson, J. E., Ward, J. M., Browne, P., Brown, J., Barton, B., Earl, A. J., & Turner, G. (1990) *EMBO J. 9*, 741–747.

Smith, J. J., Ververidis, P., & John, P. (1992) *Phytochemistry 31*, 1485–1494.

Stark, G. R., & Wahl, G. M. (1984) *Annu. Rev. Biochem. 53*, 447–491.

Stubbe, J. (1985) *J. Biol. Chem. 260*, 9972–9975.

Stutzman-Engwall, K. J., Otten, S. L., & Hutchinson, C. R. (1992) *J. Bacteriol. 174*, 144–154.
Townsend, C. A., & Barrabee, E. B. J. (1984) *J. Chem. Soc., Chem. Commun.*, 1586–1588.
Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Powers, V. M., & Kenyon, G. L. (1989) *Biochemistry 28*, 969–975.
Wang, H., & Woodson, W. R. (1991) *Plant Physiol. 96*, 1000–1001.
Wanning, M., Zähner, H., Krone, B., & Zeeck, A. (1981) *Tetrahedron Lett. 22*, 2539–2540.
Weigel, B. J., Burgett, S. G., Chen, V. J., Skatrud, P. L., Frolick, C. A., Queener, S. W., & Ingolia, T. D. (1988) *J. Bacteriol. 170*, 3817–3826.
White, R. L., John, E.-M. M., Baldwin, J. E., & Abraham, E. P. (1982) *Biochem. J. 203*, 791–793.
Woroniecki, S. R., Elson, S. W., & Baggaley, K. H. (1989) U.S. Patent 4795809.
Wright, F., & Bibb, M. J. (1992) *Gene 113*, 55–65.

Registry No. DNA (*Streptomyces clavuligerus* gene *cs1* plus flanks), 144269-47-6; DNA (*Streptomyces clavuligerus* gene *cs1*), 144269-45-4; DNA (*Streptomyces clavuligerus* gene *cs2* plus flanks), 144269-48-7; DNA (*Streptomyces clavuligerus* gene *cs2*), 144269-46-5; clavaminate synthase (*Streptomyces clavuligerus* isoenzyme 1 reduced), 144269-49-8; clavaminate synthase (*Streptomyces clavuligerus* isoenzyme 2 reduced), 144269-50-1; clavaminate synthase, 122799-56-8.