Nucleotide Sequence of the *Acinetobacter calcoaceticus* trpGDC Gene Cluster

Jeffrey B. Kaplan, Paul Goncharoff, Anita M. Seibold, and Brian P. Nichols

Department of Biological Sciences, University of Illinois at Chicago

A plasmid library of *Acinetobacter calcoaceticus* HindIII fragments was constructed, and clones that complemented an *Escherichia coli* pabA mutant were selected. Plasmids containing a 3.9-kb fragment of *A. calcoaceticus* DNA that also complemented *E. coli* trpD and trpC−(trpF+) mutants were obtained. We infer that complementation of *E. coli* pabA mutants was the result of the expression of the amphibolic anthranilate-synthase/p-aminobenzoate-synthase glutamine-amidotransferase gene and that the plasmid insert carried the entire trpGDC gene cluster. In *E. coli* minicells, the plasmid insert directed the synthesis of polypeptides of 44,000, 33,000, and 20,000 daltons, molecular masses that are consistent with the reported molecular masses of phosphoribosylanthranilate transferase, indoleglycerol-phosphate synthase, and anthranilate-synthase component II, respectively. A 3,105-bp nucleotide sequence was determined. Comparison of the *A. calcoaceticus* trpGDC sequences with other known trp gene sequences has allowed insight into (1) the evolution of the amphibolic trpG gene, (2) varied strategies for coordinate expression of trp genes, and (3) mechanisms of gene fusions in the trp operon.

Introduction

In all organisms investigated to date, ranging from gram-negative prokaryotes to lower-level eukaryotes, the five reactions that convert chorismate to tryptophan have not been found to differ, but the structure and organization of the genes involved in tryptophan biosynthesis vary considerably (Crawford 1975). *Acinetobacter calcoaceticus* contains seven separate genes that encode the enzymes of tryptophan biosynthesis (Twarog and Liggins 1970; Sawula and Crawford 1973), and these genes lie in three unlinked clusters (Sawula and Crawford 1972). The *Escherichia coli* trp operon, on the other hand, contains five contiguous genes encoding the tryptophan-biosynthetic enzymes (Yanofsky et al. 1981). Two gene fusions have occurred in the evolution of the *E. coli* trp operon, with the result that two of the *E. coli* genes encode bifunctional polypeptides (Creighton 1970; Greishaber and Bauerle 1972).

In certain organisms (*A. calcoaceticus* [Sawula and Crawford 1972], *Bacillus subtilis* [Kane et al. 1972], and *Pseudomonas acidovorans* [Buvinger et al. 1981]), tryptophan biosynthesis is "interlocked" with the biosynthesis of the vitamin dihydrofolate because the glutamine amidotransferase (GAT) subunit of anthranilate synthase (AS; EC 4.1.3.27) is amphibolic, that is, it also acts as the GAT subunit of

1. Key words: tryptophan genes, *Acinetobacter calcoaceticus*, gene rearrangements, gene fusions.

Address for correspondence and reprints: Brian P. Nichols, Department of Biological Sciences, University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680.

© 1984 by The University of Chicago. All rights reserved.
0737-4038/84/0106-0002$02.00
a second enzyme, p-aminobenzoate synthase (PABS). AS and PABS use identical substrates, chorismate and glutamine, and produce very similar products—glutamate, pyruvate, and either anthranilate (o-aminobenzoate) or p-aminobenzoate (PABA), the latter being incorporated into dihydrofolate.

We report here the nucleotide sequence of the \textit{A. calcoaceticus trpGDC} gene cluster. \textit{TrpG} encodes the amphibolic GAT subunit of AS and PABS. Since the nucleotide and amino acid sequences of genes encoding the pathway-specific GAT subunits—\textit{pabA}, \textit{trpG}, and \textit{trp}(\textit{G})\textit{D}—in several enterobacterial species are known (Nichols et al. 1980; Tso et al. 1980; Kaplan and Nichols 1983), determination of the primary structure of an amphibolic subunit might assist in the elucidation of the functional and evolutionary origins of pathway-specific genes. \textit{TrpD} and \textit{trpC} encode anthranilate phosphoribosyltransferase (PRTase; EC 2.4.2.18) and indoleglycerol phosphate synthase (InGPS; EC 4.1.1.48), respectively. In several species of the Enterobacteriaceae, these activities lie on bifunctional polypeptides. In \textit{E. coli}, for example, PRTase and AS GAT activities are encoded by a single gene, \textit{trpD} (Greishaber and Bauerle 1972; Miozzari and Yanofsky 1979). For clarity, we will follow the suggestion of Crawford (1975) and refer to the fused gene as \textit{trp(G)D}. Similarly, in \textit{E. coli} InGPS lies on the same polypeptide with phosphoribosylanthranilate isomerase (PRAI) and is encoded by \textit{trpC} (Creighton 1970). PRAI is encoded by \textit{trpF} in \textit{A. calcoaceticus}, and we will refer to the fused \textit{E. coli} gene as \textit{trpC(F)}. Comparisons of the separate and fused gene sequences can aid in the determination of the minimum amount of information necessary to encode each individual function as well as assist in the elucidation of events that may have resulted in fused genes.

Materials and Methods

Bacteriophages, Strains, and Strain Construction

The bacterial strains used in this work are listed in table 1. Bacteriophages M13mp8 and M13mp9 are those described by Messing and Vieira (1982) and were obtained from New England Biolabs (Beverly, Mass.). Bacteriophage P1kc was obtained from C. Yanofsky, and transductions were performed as described by Miller (1972). \textit{Escherichia coli} BN100 was constructed by transducing a P1kc lysate made from \textit{E. coli} AB3292 (Huang and Pittard 1967) into \textit{E. coli} W3110 \textit{Δ}{\textit{trpLD102}} and selecting streptomycin-resistant colonies. Resistant colonies were then screened for PABA auxotrophy. \textit{Escherichia coli} BN105 (\textit{trpR}, \textit{minA}, \textit{minB}, \textit{leu}, \textit{thi}, \textit{lacY}, \textit{ara}, \textit{gal}, \textit{malA}, \textit{xyl}, \textit{azi}, \textit{tsx}, \textit{tonA}, and \textit{rpsL}) was constructed by transduction of a P1kc lysate made from \textit{E. coli} W3110 \textit{trpR Δ}{\textit{trpEA2 ina2}} into \textit{E. coli} P678-54 and isolation of 5-methyltryptophan-resistant, threonine-independent colonies.

\textit{DNA manipulations}.—Restriction endonucleases were either prepared in this laboratory according to published procedures or purchased from commercial distributors (New England Biolabs; P-L Biochemicals, Milwaukee; and Boehringer Mannheim Biochemicals, Indianapolis). Calf intestinal phosphatase was obtained from Boehringer Mannheim. Restriction endonuclease digests were performed according to the commercial supplier's recommendations. Ligations were performed at DNA concentrations of 30–60 \textmu g/ml in 10 mM Tris–HCl (pH 7.4), 10 mM MgCl\textsubscript{2}, 10 mM 2-mercaptoethanol, and 0.5 mM ATP for 1–4 h at 14 °C. Transformation of \textit{E. coli} was carried out as described by Mandel and Higa (1970). \textit{Acinetobacter calcoaceticus} DNA was prepared according to Marmur (1966), except...
that DNA precipitated in ethanol was collected by centrifugation. Plasmid DNA was prepared by the method of Birnboim and Doly (1979).

Plasmid constructions.—Acinetobacter calcoaceticus DNA was prepared, and a plasmid library was constructed by ligating 20 μg of HindIII-digested A. calcoaceticus DNA with 10 μg of pBR322 treated with HindIII and phosphatase. A portion of the ligated DNA representing 3 μg of vector was transformed into E. coli LE392 (r-m+), and ampicillin-resistant colonies were selected. Approximately 7,000 ampicillin-resistant colonies were pooled, and the plasmid DNA was isolated. Two micrograms of the amplified plasmid library was transformed into E. coli BN100, and ampicillin-resistant, PABA-independent colonies were isolated.

Protein synthesis in E. coli minicells.—Plasmids containing A. calcoaceticus trp genes were transformed into E. coli BN105, and ampicillin-resistant colonies were isolated. Isolation of minicells and labeling of proteins with 35S-methionine (300 Ci/mmol; Amersham, Arlington Heights, Ill.) were performed as described by Matsumura et al. (1977).

DNA-sequence analysis.—DNA sequences were determined by the procedure of Maxam and Gilbert (1980) or Sanger et al. (1977). DNA fragments were end-labeled either with T4-polynucleotide kinase (P-L) and (γ-32P)ATP (>2,000 Ci/mmol; Amersham) or with E. coli DNA polymerase I (Klenow fragment;
Boehringer Mannheim) and (α-32P)dCTP (>800 Ci/mmol; Amersham) (Maxam and Gilbert 1980). Alternatively, restriction endonuclease–generated DNA fragments were cloned into M13mp8 or M13mp9 RF I. Recombinant phages were transformed into E. coli JM103, and single-stranded DNA was prepared as described by Messing and Vieira (1982). The polyacrylamide gel/urea electrophoresis system described by Sanger and Coulson (1978) was used for resolving DNA fragments. DNA sequences were analyzed in part by computer programs (Queen and Korn 1980; S. Weaver, B. Nichols, and C. Hutchison III, unpublished).

Results

Construction and Characterization of Recombinant Plasmids containing the Acinetobacter calcoaceticus trpGDC Gene Cluster

Plasmid libraries of A. calcoaceticus DNA were transformed into BN101 and selected for ampicillin resistance and PABA independence. Approximately 50 such colonies were obtained. Plasmid DNA prepared from six of the PABA-independent colonies obtained from the HindIII library was cleaved with HindIII. Agarose-gel electrophoresis showed that, in addition to the 4.3-kb pBR322-HindIII fragment, all of the plasmids contained a 3.9-kb HindIII fragment. Four of the six plasmids contained a small additional HindIII fragment. Retransformation of Escherichia coli BN101 with the two smaller plasmids resulted in a high proportion of PABA-independent colonies. One plasmid, designated pBN79, was selected for further characterization.

Plasmid pBN79 was mapped with several restriction endonucleases. The sites are indicated in figure 1. Since SalI cleaved the A. calcoaceticus DNA into approximately equal parts, the two SalI-HindIII DNA fragments were cloned separately into the SalI and HindIII sites of pBR322. Each was then transformed into E. coli BN101, and ampicillin-resistant colonies were tested for PABA independence. The plasmid containing the larger (2,350-bp) HindIII-SalI fragment (pBN78) was found to complement the E. coli pabA mutation, whereas the plasmid containing the smaller (1,750-bp) fragment (pMB466) did not.

It has been shown that the product of the A. calcoaceticus trpG gene, that is, AS Component II (AS Coll), does not complement an enterobacterial AS Component.
Therefore, in order to confirm that transformation to PABA independence was the result of cloning the *A. calcoaceticus trpG* gene and not the result of anomalous expression of a normally cryptic *pabA* gene, complementation tests were performed to test for the presence of other *trp* genes known to be linked to *A. calcoaceticus trpG*.

Complementation tests for the presence of *A. calcoaceticus trpD* were performed by transforming the plasmids into *E. coli* BN100. The tryptophan requirement of this strain cannot be circumvented by anthranilate unless the PRTase (*trpD*) function is supplied by the plasmid. Only pBN79 gives rise to colonies on medium containing anthranilate and PABA. Since neither pBN78 nor pMB466 is *trpD*+, separation of the *HindIII* fragment at the *SalI* site must either interrupt the *A. calcoaceticus trpD* gene or cause the loss of its expression in some other manner.

The test for complementation by *A. calcoaceticus trpC* was carried out in *E. coli* W3110 *trpC55*, a strain that contains a missense mutation in the first half of *trpC(F)*. This mutation causes the loss of InGPS activity without loss of PRAI activity. The strain harboring either pBN79 or pMB466 can grow on minimal medium, indicating the presence of *A. calcoaceticus trpC* on both of these plasmids. Plasmid pBN78 does not confer the ability to grow on minimal medium. The *trpC* gene of the *A. calcoaceticus trpGDC* cluster must therefore be located to the right of the *SalI* site (as indicated in fig. 1).

Growth-rate experiments indicate that *A. calcoaceticus trpG* and *trpD* complement *E. coli pabA* and *trpD* mutants as effectively as exogenous PABA or tryptophan (data not shown). *Acinetobacter calcoaceticus trpC*, on the other hand, complements the *E. coli trpC55* mutation slightly less effectively than exogenous tryptophan. These complementation data show that PABA independence is conferred on an *E. coli pabA* mutant by an *A. calcoaceticus* gene that lies in or near the *trpGDC* cluster on a 3.9-kb *HindIII* DNA fragment. Since only three polypeptides are synthesized from this DNA fragment (see below), we conclude that PABA independence is conferred by *trpG*. In addition, the data indicate that *trpG* and *trpC* are separated by a *SalI* site that inactivates *trpD*.

Analysis of Polypeptides Encoded by the *A. calcoaceticus* DNA Fragment

The polypeptides encoded by the three plasmids were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis of 35S methionine–labeled proteins produced by *E. coli* minicells. Plasmids pBN79, pBN78, and pMB466 were transformed into *E. coli* BN105, a minicell-producing strain containing a defective *trp* aporepressor. The results are shown in figure 2. Plasmid pBN79 efficiently directs the synthesis of three major polypeptides in addition to the pBR322-encoded β-lactamase. The molecular masses of 44,000, 33,000, and 20,000 daltons are consistent with the molecular masses of 37,543, 30,203, and 21,779 daltons predicted from DNA sequence analysis (see below). Molecular masses estimated by gel permeation chromatography for PRAI, InGPS, and AS CoII were 42,500, 30,000, and 14,000, respectively (Twarog and Liggins 1970; Sawula and Crawford 1973). The difference in the estimates for AS CoII may be due to the different analytical methods employed. The origin of the faint bands beneath the largest polypeptide is unknown but may be a result of artifactual translation initiation at internal methionine codons of *trpD*.
Acinetobacter trpGDC Sequence

![Autoradiogram](image)

**Fig. 2.**—Autoradiogram of an SDS-polyacrylamide gel showing 35S methionine-labeled proteins synthesized from *Escherichia coli* minicell plasmids pBN79 (lane a), pBN78 (lane b), and pMB466 (lane c).

Plasmids pBN78 and pMB466 each encode one major polypeptide in addition to β-lactamase, and the results are consistent with the complementation data described above. Plasmid pBN78 directs the synthesis of only the smallest polypeptide, that is, the trpG-gene product, and pMB466 directs the synthesis of the 30,000-dalton trpC-gene product. The largest polypeptide, the product of trpD, is absent from both the pBN78 and pMB466 lanes, confirming that cleavage at the SalI site inactivates the gene.

**Nucleotide-Sequence Determination of the *A. calcoaceticus* trpGDC-Gene Cluster**

The nucleotide sequence of 3,105 bp containing the *A. calcoaceticus* trpGDC gene cluster was determined by the methods of Sanger et al. (1977) and Maxam and Gilbert (1980). Restriction fragments from pBN78 and pMB466 were either end-labeled (Maxam and Gilbert procedure) or cloned into M13mp8 or M13mp9 (Sanger procedure). One restriction fragment from pBN79 was used to determine the sequence across the SalI site. All bases were determined on both strands of the DNA, and, with one exception, all restriction sites used for sequencing were overlapped. The exception was the HpaII site in trpG. We are confident that no information was lost, since this site occurs within a region of the DNA sequence that is highly conserved in several related GAT sequences. The nucleotide and the predicted amino acid sequences of the trpGDC gene cluster are presented in figure 3.

The three trp genes lie in proximity on the chromosome. Only three nucleotides (a termination codon) lie between trpG and trpD, and 13 nucleotides lie between trpD and trpC. The insufficient number of nucleotides for individual regulatory sequences between the genes suggests a polycistronic transcriptional unit, a suggestion that is in agreement with previous reports of the coordinate regulation of the trpGDC cluster (Cohn and Crawford 1976). The adjacent termination and initiation codons of *A. calcoaceticus* trpG and trpD are identical with those found between
Acinetobacter trpGDC Sequence

Leu Ile Val Ala Cys Leu Ser Asp Gln Gln Leu Val Glu Val His
TTG ATT GCT GCT TGT TGT TCT GAT CAG CCA CTC CAA CAG AGG TCA AAA ACT GCA TTT GAA TAT GAT CTT CAG GTT GCT GAA GTC GAT
2270 2280 2290 2300

Leu Ile Val Ala Cys Leu Ser Asp Gln Gln Leu Val Glu Val His
TTG ATT GCT GCT TGT TGT TCT GAT CAG CCA CTC CAA CAG AGG TCA AAA ACT GCA TTT GAA TAT GAT CTT CAG GTT GCT GAA GTC GAT
2270 2280 2290 2300

Leu Ile Val Ala Cys Leu Ser Asp Gln Gln Leu Val Glu Val His
TTG ATT GCT GCT TGT TGT TCT GAT CAG CCA CTC CAA CAG AGG TCA AAA ACT GCA TTT GAA TAT GAT CTT CAG GTT GCT GAA GTC GAT
2270 2280 2290 2300

Leu Ile Val Ala Cys Leu Ser Asp Gln Gln Leu Val Glu Val His
TTG ATT GCT GCT TGT TGT TCT GAT CAG CCA CTC CAA CAG AGG TCA AAA ACT GCA TTT GAA TAT GAT CTT CAG GTT GCT GAA GTC GAT
2270 2280 2290 2300

Fig. 3.—DNA and predicted amino acid sequences of the trpGDC gene cluster and flanking regions.

E. coli thrB and thrC (Cossart et al. 1981) and are reminiscent of the overlapping termination and initiation codons observed in the E. coli trp operon between trpE and trp(G)D (Nichols et al. 1980) and between trpB and trpA (Platt and Yanofsky 1975; Nichols and Yanofsky 1979). The overlapping termination and initiation codons in the E. coli trp operon have been implicated in the coordinate translation of the adjacent genes (Oppenheim and Yanofsky 1980; Aksoy et al. 1984), and it is possible that the adjacent termination and initiation codons play a similar role in the coordinate expression of A. calcoaceticus trpG and trpD.

The three trp genes can be easily recognized by their similarity to both E. coli trp(G)D and the 5'-terminal portion of E. coli trpC(F). Figure 4 shows the similarity between the aligned A. calcoaceticus and E. coli nucleotide sequences. The sequence similarity is evident along the entire length of the gene cluster and is strongest within the coding regions of the genes. Similarities are notably reduced around intercistronic regions and at the beginning and end of the A. calcoaceticus gene cluster. The intermittent nature of strong similarity suggests conservation of structurally and functionally critical amino acid sequences and divergence of expression-related signals and less important amino acid sequences.

Amino Acid-Sequence Alignments

Comparisons of the predicted amino acid sequences of A. calcoaceticus AS CoII, PRTase, and InGPS with the homologous sequences from other organisms are shown in figures 5, 6, and 7, respectively. Alignments are manual, and gaps have been inserted to increase similarity among the sequences. The three A. calcoaceticus amino acid sequences show 36%–39% similarity to the E. coli amino acid sequences (AS CoII, 37%; PRTase, 36%; InGPS, 39%), however, A. calcoaceticus AS CoII is much more similar to the E. coli PABS CoII (57%) than to AS CoII. The higher similarity may explain why A. calcoaceticus trpG functions in E. coli PABA synthesis but does not function in anthranilate synthesis.
FIG. 4.—Nucleotide-sequence similarity between *Acinetobacter calcoaceticus* trpGDC and the equivalent portion of the *Escherichia coli* trp operon. Each point represents the similarity between 40-nucleotide-long sequences. The positions of the *A. calcoaceticus* genes are indicated by the arrows along the horizontal axis.

**Discussion**

**Comparison of Amino Acid Sequences**

Alignment of the three AS GAT amino acid sequences with the *Escherichia coli* PABS GAT sequence underscores the common origin of the AS GAT and PABS GAT subunits (Kaplan and Nichols 1983). Strong similarity occurs in several regions throughout the sequences. One of these regions, at residues 74–83, contains a cysteine residue (position 79) that has been determined to be essential for catalytic activity. A glutamine analog covalently reacts with this cysteine residue in the *Pseudomonas putida* (Kawamura et al. 1978) and *Serratia marcescens* (Tso et al. 1980) AS GAT subunits. Another region, centered in the area of residue 103, is rather less conserved among the sequences but contains a lysine residue shown to be essential for activity (Bower and Zalkin 1982). Arginine replaces the essential lysine residue in the *Acinetobacter calcoaceticus* sequence. The specific roles of the other conserved sequences are not known, but their conservation across generic boundaries suggests critical roles in either substrate binding or catalysis.

The *A. calcoaceticus* trpG product is most closely related to the *P. putida* trpG product and least related to the *E. coli* trp(G) product. These relationships correlate

![Alignment of predicted amino acid sequences of *Acinetobacter calcoaceticus* anthranilate synthase (AS) CoII (row 1), *Pseudomonas putida* AS CoII (Kawamura et al. 1978) (row 2), *Escherichia coli* p-aminobenzoate synthase CoII (Kaplan and Nichols 1983) (row 3), and the equivalent portion of *E. coli* AS CoII (Nichols et al. 1980) (row 4). Numbering refers only to the *A. calcoaceticus* sequence. Identical residues in all sequences are boxed.](image-url)
with intrageneric AS subunit exchange experiments. The *A. calcoaceticus* and *P. putida* GAT subunits can be exchanged without appreciable loss of function, but neither of the AS CoI subunits can function with an enterobacterial AS CoII (Sawula and Crawford 1973; Queener et al. 1973); however, at least two (*A. calcoaceticus* AS CoI and *E. coli* PABS CoII) of the three most similar sequences shown in figure 5 do function in PABA synthesis with *E. coli* PABS CoI in vivo. Because of the observed sequence similarity between *A. calcoaceticus* trpG and *E. coli* trpG,
it would not be surprising to find that \textit{P. putida} AS CoII functioned similarly, although it has not yet been determined whether \textit{P. putida} trpG encodes an amphibolic GAT subunit.

It is not obvious from inspection of the amino acid sequences why the \textit{A. calcoaceticus} GA1 subunit does not function amphibolically in \textit{E. coli}. Evidence suggests that the N-terminal portion of \textit{E. coli} AS CoII interacts with the AS CoI subunit (Greishaber and Bauerle 1972). Whether the same is true of the PABS subunit is unknown, however, if most of the remainder of the sequence is conserved for other functions, then by analogy with AS CoII, it is likely that the N-terminal region of the PABS GAT subunit also specifies interaction with the large PABS subunit. Since the \textit{A. calcoaceticus} GAT cannot complement the \textit{E. coli} AS CoI but behaves instead as a PABS-specific GAT subunit, it is likely that the amino acid differences observed in \textit{E. coli} trp(G) in this region account in part for the functional differences in subunit specificity. Subunit-specificity determinants would also be expected to occur on the CoI subunits of AS and PABS, and according to observed complementation patterns, it is likely that in the subunit-interaction areas \textit{A. calcoaceticus} AS CoI will show similarities to \textit{A. calcoaceticus} PABS CoI that are not found between \textit{E. coli} AS CoI and \textit{E. coli} PABS CoI.

The predicted amino acid sequences of \textit{A. calcoaceticus} and \textit{E. coli} PRTase are compared in figure 6. The \textit{E. coli} sequence shown begins within the fusion region between the trp(G) and trpD portions of the sequence. (The monofunctional PRTase-encoding sequence of \textit{S. marcescens} begins at the position equivalent to \textit{A. calcoaceticus} residue 3.) Similarity is quite strong in the central region of the polypeptide, whereas the N- and C-terminal portions show less similarity. The diversity at the N and C termini may be due to either a structural or a functional adaptation of the \textit{E. coli} sequence to the fusion of trp(G) in conjunction with the evolution of the internal \textit{E. coli} trpP, promoter (Morse and Yanofsky 1968; Horowitz et al. 1982; Horowitz and Platt 1982) or to an entirely different origin of this DNA region.

The amino acid sequence of the \textit{A. calcoaceticus} trpC product aligns with the N-terminal portion of the fused \textit{E. coli} trpC(F) product (Christie and Platt 1980). As in the PRTase sequences, the extent of similarity is strongest in the central portion of the sequence and is somewhat lower toward the termini. The product of the \textit{E. coli} trpC(F) gene is 450 amino acids long and catalyzes both the InGPS and PRAI activities. Analysis of proteolytic fragments has determined that the two activities lie on independently folding domains and that the InGPS activity is confined to the 289 N-terminal residues (Kirschner et al. 1980). Comparison of the \textit{E. coli} trpC(F) sequence with the monofunctional \textit{A. calcoaceticus} trpC sequence suggests that the InGPS activity is encoded within the first 260 amino acid residues and that the domain containing PRAI activity lies in the remaining 190 amino acids. \textit{Saccharomyces cerevisiae} TRP1 (equivalent to trpF) encodes InGPS on a separate polypeptide chain. The first recognizable similarity to the \textit{E. coli} sequence occurs at positions 15–19 of TRP1, matching \textit{E. coli} positions 266–271 as numbered in figure 7 (Tschumper and Carbon 1980). This latter sequence, KVCGLT, also overlaps the last three amino acids of the \textit{A. calcoaceticus} InGPS sequence. While it has been shown that the \textit{E. coli} trp(G)D fusion is the result of modification and translation through an ancestral intercistronic region (Miozzari and Yanofsky 1979), this is probably not the same mechanism that resulted in the fusion of trpC and trpF. We suggest instead that the trpC-trpF fusion occurred via a deletion between
separate but adjacent \( \text{trpGDC} \) genes with end points within each of the ancestral genes. This hypothesis is supported by the observation that the “fusion peptide” between the GAT and PRTase activities lies near the surface of the molecule and is accessible to proteases (Li et al. 1974), whereas the junction between the fused InGPS and PRAI activities is buried within the protein core (Kirschner et al. 1980).

Codon and Amino Acid Usage in the \( A. \text{calcoaceticus trpGDC} \) Gene Cluster

The choice of specific codons within a codon family differs between \( A. \text{calcoaceticus trpGDC} \) and the equivalent region of the \( E. \text{coli trp} \) operon (table 2). As seen in many prokaryotes, there is an avoidance of the use of such codons as ATA(Ile), CTA(Leu), CGR(Arg), and AGR(Arg) (Grantham et al. 1981). These preferences correlate with insignificant amounts of the cognate tRNAs in \( E. \text{coli} \) (Ikemura 1981), and it may be that the same is true in other prokaryotes. Other \( A. \text{calcoaceticus} \) codon usage trends that are similar to those of \( E. \text{coli} \) are preferred use of ATT (Ile), AAA (Lys), and GGY (Gly). Among the notable differences in \( A. \text{calcoaceticus} \) codon usage, however, are the following: (1) CTG (Leu) is not the preferred Leu codon; rather, TTA and TTG represent one-half of the Leu codons; (2) TCT and TCA (Ser) are preferred to TCG; (3) CCT (Pro) is preferred to CCG; and (4) in all of the two-codon families, the member ending in T or A is used from two to six times as often as is the alternative.

It has been observed that within the \( \text{trp} \) genes of the Enterobacteriaceae, codon usage is influenced by the genomic G + C content of the organism and that the third positions of the codons are the most influenced (Nichols et al. 1980, 1981). The same effect is seen in the \( A. \text{calcoaceticus trpGDC} \) gene cluster, albeit in the opposite direction: the G + C content of \( A. \text{calcoaceticus} \) is 42%, but in the third positions of the codons it is only 35%. As mentioned above, this is very noticeable in the two-codon families, but it is also obvious in some of the four-codon families (e.g., Ser and Pro).

In addition to the influence exerted on the third position of the codon, the genomic G + C content may also affect the choice of amino acids. That is, within a family of amino acids, those that are encoded by A/T-rich codons are used more frequently in \( A. \text{calcoaceticus} \). For example, in the (aliphatic) hydrophobic class of amino acids, Ile (ATH) is used more frequently and Ala (GCX) used less frequently than in \( E. \text{coli} \). A similar statement can be made concerning the choice between the basic amino acids Lys (AAR) and Arg (CGX + AGR). The true significance of this observation awaits further documentation of the amino acid composition of orthologous proteins from additional organisms with varied genomic G + C content.

Conclusion

The nucleotide sequence of the \( A. \text{calcoaceticus trpGDC} \) gene cluster emphasizes the fundamental similarity of the enzymes involved in tryptophan metabolism yet also illustrates the variety of genomic arrangements and regulatory differences coordinating their expression. The G-D-C gene order is common to gram-negative bacteria, but the amphibolic nature of \( \text{trpG} \) and the separateness of the \( \text{trpGDC} \) cluster are not. Transcription of the \( A. \text{calcoaceticus trpGDC} \) cluster is coordinated with the expression of \( \text{trpE} \), but the \( \text{trpFBA} \) cluster is regulated differently (Cohn and Crawford 1976). In the absence of other results, we cannot identify potential
Table 2
Codon Usage in *Acinetobacter calcoaceticus trpGDC* and in the Equivalent Portion of the *Escherichia coli trp* Operon

<table>
<thead>
<tr>
<th>CODON</th>
<th>A. CALCOACETICUS TRP</th>
<th>E. COLI TRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>D</td>
</tr>
<tr>
<td>TTT (Phe)</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>TTC (Phe)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TTA (Leu)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>TTG (Leu)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>CTT (Leu)</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>CTC (Leu)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CTA (Leu)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CTG (Leu)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>ATT (Ile)</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>ATC (Ile)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>ATA (Ile)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATG (Met)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>GTT (Val)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GTC (Val)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>GTA (Val)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>GTG (Val)</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>TCT (Ser)</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>TCC (Ser)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>TCA (Ser)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>TCG (Ser)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CCT (Pro)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CCC (Pro)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CCA (Pro)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CCG (Pro)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>ACT (Thr)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>ACC (Thr)</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>ACA (Thr)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ACG (Thr)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>GCT (Ala)</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>GCC (Ala)</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>GCA (Ala)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>GCG (Ala)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>TAT (Tyr)</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>TAC (Tyr)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TAA (End)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TAG (End)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAT (His)</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>CAC (His)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CAA (Gln)</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>CAG (Gln)</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>AAT (Asn)</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>AAC (Asn)</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>AAA (Lys)</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>AAG (Lys)</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
transcriptional signals (either promoters or operators) by inspection of the 185-bp sequence flanking trpG. Similarly, we do not see evidence of secondary structures or leader peptide-coding regions that might reflect an attenuation mechanism of transcriptional regulation; however, the coupling of translation by the propinquity of termination and initiation signals of adjacent genes may rival gene fusion as a method of coordinating expression. We have pointed out above that the termination and initiation codons of trpG and trpD are adjacent to one another, and we have also noticed that the termination codon of trpD overlaps a nine-base sequence that closely resembles the Shine and Dalgarno (Escherichia coli) sequence (Shine and Dalgarno 1974). Although translational coupling has not been directly demonstrated in such arrangements, it is possible that termination of translation within or very near to an initiating environment influences the efficiency of the subsequent initiation event.

The evolution of an amphibolic trp/pab GAT subunit can be envisioned by several mechanisms, involving the duplication of one (CoI) or two (CoI + CoII) genes (with one of the duplicated CoII genes being deleted in the latter case). While the sequence data do not support one mechanism more than another, the similarity of the trp-linked amphibolic subunit to E. coli pabA rather than to E. coli trp(G) suggests that the pabA-like sequence may more closely resemble the ancestral GAT sequence. This is further supported by the fact that the trp-dedicated GAT sequences of lower eukaryotes also closely resemble the E. coli pabA sequence (Schectman and Yanofsky 1983; H. Zalkin, personal communication).

Gene fusions are commonly encountered in evolutionary comparisons of trp-gene arrangements. The fusion junctions that have thus far been characterized at
the nucleotide-sequence level by comparisons of fused and separate genes include enterobacterial trpG-trpD (Miozzari and Yanofsky 1979; Nichols et al. 1980), Neurospora crassa TRP1 (equivalent to trpG-trpC-trpF) (Schechtman and Yanofsky 1983), and Saccharomyces cerevisiae TRP5 (equivalent to trpA-trpB) (Zalkin and Yanofsky 1981). In each case, the fusion regions are characterized by the presence of a "linking peptide" in the fused gene that did not previously exist in either of the separate genes (Yanofsky, personal communication). We report here the first complete nucleotide sequences of separate trpD and trpC genes. In conjunction with previous sequence data, we propose that an alternative fusion mechanism involving an in-phase deletion with end points within adjacent but separate genes may be responsible for joining trpC and trpF. The Bacillus subtilis trp cluster contains separate but adjacent trpC and trpF genes, and determination of the intercistronic region may provide further information regarding this hypothesis.

Acknowledgments

We thank the members of BioS 466 (Procaryotic Molecular Biology Laboratory)—G. Alberts, P. Balandyk, D. Baro, J.-D. Chen, S. Hahn, B. Howard, J.-W. Lee, P. Lysakowski, J. Malakooti, M. Sussman, H. Weidmann, and C. Wilke—for determining the sequence of the Acinetobacter calcoaceticus DNA fragment contained in pMB466. W. K. Merkel for assisting with the complementation studies, I. Crawford for providing the A. calcoaceticus strains and for helpful advice and discussions, and C. Yanofsky and H. Zalkin for communicating information prior to publication. This work was supported by U.S. Public Health Service grant AI 18639 and by the Department of Biological Sciences of the University of Illinois at Chicago.

LITERATURE CITED


Acinetobacter trpGDC Sequence


NICHOLS, B. P., G. F. MIOZZARI, M. VAN CLEEMPUT, G. N. BENNETT, and C. YANOFSKY.


WALTER M. FITCH, reviewing editor

Received January 16, 1984; revision received April 16, 1984.