Nucleotide Sequence of the Genes for Tryptophan Synthase in *Pseudomonas aeruginosa*[1]

Ayele Hadero and Irving P. Crawford
Department of Microbiology, University of Iowa

We have determined the DNA sequence of the two adjacent genes for the α and β chains of tryptophan synthase in *Pseudomonas aeruginosa,* along with 34 5′-flanking and 799 3′-flanking base pairs. The gene order is *trpBA* as predicted from earlier genetic studies, and the two cistrons overlap by 4 bp; a ribosome binding site for the second gene is evident in the coding sequence of the first gene. We have also determined the location of three large deletions eliminating portions of each gene. A detailed comparison of the deduced *P. aeruginosa* amino acid sequence with those published for *E. coli, Bacillus subtilis,* and *Saccharomyces cerevisiae* shows much similarity throughout the β and most of the α subunit. Most of the residues implicated by chemical modification or mutation as being critical for enzymatic activity are conserved, along with many others, suggesting that three-dimensional structure has remained largely constant during evolution. We also report the construction of a recombinant plasmid that overproduces a slightly modified α subunit from *P. aeruginosa* that can form a functionally effective multimer with normal *E. coli* β_2 subunit in vivo.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/992267 by guest on 21 August 2022

Introduction

Tryptophan synthase, the last enzyme in the tryptophan synthetic pathway, is a well-studied example of a complex enzyme with interdependent subunits. The overall reaction that it catalyzes can be divided into two half-reactions: (1) indole-3-glycerol phosphate → indole + D-glyceraldehyde-3-phosphate; and (2) indole + L-serine → L-tryptophan. Half-reaction (1) is accomplished by an active site localized in the α subunit of the bacterial $\alpha_2\beta_2$ tetramer; half-reaction (2) is dependent on an active site in the β chain that binds the pyridoxal phosphate cofactor for the enzyme and is active only when self-associated to a β_2 dimer (Yanofsky and Crawford 1972; Miles 1979). In *Saccharomyces* and *Neurospora,* in contrast to the situation in prokaryotes, green algae, and plants, the genes for the two subunits are fused, forming a single polypeptide having N-terminal and C-terminal portions resembling in sequence the prokaryotic α and β chains; this fusion product dimerizes to perform both reactions (1) and (2) (Tsai et al. 1974; Matchett and DeMoss 1975; Zalkin and Yanofsky 1982). In several gram-negative bacteria and *Bacillus subtilis* the genes for the tryptophan synthase subunits are adjacent and transcribed in the order *trpB* (β chain)–*trpA* (α chain) (Crawford 1975). This order is reversed in the fungal fusion product, where the segment homologous to the α chain is N-terminal.

In the enteric bacteria, where they have been best studied, the α and β polypeptides are each organized into two independently folding domains, separable after limited

1. Key words: tryptophan synthase, *Pseudomonas aeruginosa,* nucleotide sequence.

Address for correspondence and reprints: Dr. Irving P. Crawford, Department of Microbiology, University of Iowa, Iowa City, Iowa 52242.

Mol. Biol. Evol. 3(3):191–204. 1986.
© 1986 by The University of Chicago. All rights reserved.
0737-4038/86/0303-3301$02.00

proteolysis and capable of reassociating in solution (Crawford et al. 1978; Higgins et al. 1979; Miles 1979).

Comparative sequence information for bacterial tryptophan synthase is available only for *E. coli,* for its close relative *Salmonella typhimurium* (Nichols and Yanofsky 1979; Crawford et al. 1980), and for *B. subtilis* (Henner et al. 1984). In these cases the *trpB* and *trpA* genes are found in five- or six-gene *trp* operons in association with genes for earlier enzymes in the tryptophan pathway. As yet there is no sequence information from other bacteria in which the *trp* genes are dispersed to three or four different, well-separated locations, often showing independent regulation.

It has long been known that in *Pseudomonas putida* and *P. aeruginosa* the *trpB* and *trpA* genes are separate from the structural genes for the other enzymes of the pathway and are regulated by induction rather than by repression (Crawford and Gunsalus 1966; Gunsalus et al. 1968; Calhoun et al. 1973). The relevant chromosomal segment from *P. aeruginosa* PAC174 has recently been cloned. It has been shown that the tryptophan synthase genes are transcribed in the order *trpB–trpA* and that a regulatory gene producing a trans-acting product involved in their induction lies just upstream from them (Hedges et al. 1977; Manch and Crawford 1981, 1982). We report here the nucleotide sequence of *P. aeruginosa trpB* and *trpA,* along with some of the flanking sequence, confirming homology of the genetically separate form of the enzyme in *Pseudomonas* with the operonic form found in the enteric bacteria and *Bacillus.*

Downloaded from https://academic.oup.com/mbe/article/3/3/191/985255 by guest on 20 August 2022

Material and Methods

The construction of the plasmid pZAZ167 used as the source of DNA for sequencing has been described (Manch and Crawford 1981). A detailed restriction map of the distal two-thirds of the 4.5-kb insert of *Pseudomonas aeruginosa* chromosomal DNA in this pBR322 derivative is shown in figure 1. Plasmid DNA was purified as earlier described (Manch and Crawford 1981) by CsCl–ethidium bromide equilibrium centrifugation. After digestion with commercial restriction endonucleases, fragments to be sequenced were electroeluted from acrylamide gels and end labeled with ^{32}P by one of three methods: 5′ labeling with T4 polynucleotide kinase and [γ-^{32}P]ATP (Maxam and Gilbert, 1980), 3′ labeling with terminal transferase and [α-^{32}P] 2′-3′-dideoxyadenosine triphosphate (Yousaf et al. 1984), or 3′ fill-in labeling with the large fragment of *E. coli* DNA polymerase I and the appropriate radioactive deoxynucleotide triphosphate (Maxam and Gilbert 1980). When the procedure used resulted in both strands being labeled, the strands were separated as described earlier (Crawford et al. 1980). Sequencing reactions were carried out according to the procedure of Maxam and Gilbert (1980), then developed by urea–polyacrylamide gel electrophoresis according to the method Sanger and Coulsen (1978). The resulting data were analyzed with the aid of the PCS computer program (Lagrimini et al. 1984).

Plasmid construction by ligation of acrylamide gel–purified fragments to cleaved and phosphatase-treated vectors, followed by transformation into competent *E. coli* cells and selection of antibiotic-resistant transformants, was done according to the procedures detailed by Maniatis et al. (1982).

Results

Sequence of the Structural Genes

Figure 1 shows the fragments whose sequences were determined in assembling the 2,839-bp-long sequence of figure 2. The sequence of both strands was obtained—

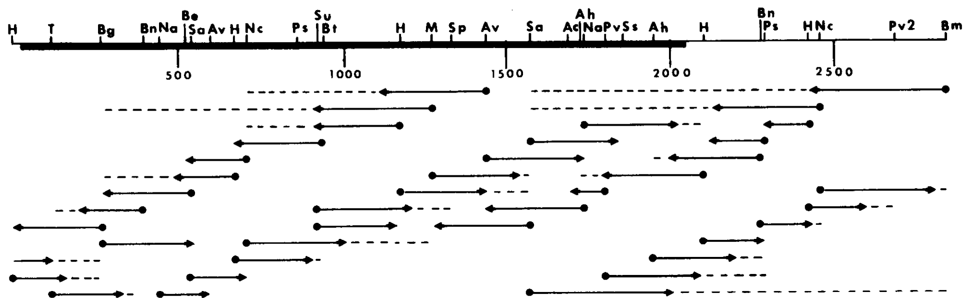

FIG. 1.—Restriction map of the sequenced *Pseudomonas aeruginosa* DNA. The location of the *trpB* and *trpA* genes is indicated by a heavy line. The direction of transcription is from left to right. Below the restriction map are indicated the fragments used to obtain the sequence. Solid lines indicate regions of the fragments where the sequence was readable. Black dots indicate the labeled end, and arrowheads the direction of sequencing. Restriction enzyme abbreviations: Ac, *Acc*I; Ah, *Aha*II; Av, *Ava*II; Be, *Bst*EII; Bg, *Bgl*II; Bn, *Ban*I; Bm, *Bam*HI; Bt, *Bst*NI; H, *Hinf*I; M, *Mlu*I; Na, *Nar*I; Nc, *Nco*I; Ps, *Pst*I; Pv, *Pvu*I; Pv2, *Pvu*II; Sa, *Sal*I; Sp, *Sph*I; Ss, *Sst*II; Su, *Sau*96I; and T, *Taq*I.

except for the final 10 bp, where only the coding strand was available—and all restriction sites were overlapped at least once. Given the size of the smallest labeled fragments, we believe that our coding strand sequence ends ~7 bp from the *Bam*HI site defining the end of the *Pseudomonas* chromosomal DNA insert in pZAZ167.

From the orientation and approximate location of the *trpB* and *trpA* genes determined earlier (Manch and Crawford 1982), it was easy to recognize the corresponding open reading frames, which have been translated in figure 2. The *trpB* gene begins at bp 35, preceded by an acceptable ribosome binding site (GGAG) at bp 25–28. The first stop codon in this reading frame is TGA at bp 1,238–1,240, which shares its first two bases with the *trpA* gene start codon at bp 1,237. *TrpA* ends with a TGA stop codon at bp 2,038–2,040. Codon usage in the two *Pseudomonas* structural genes is given in table 1. Figures 3 and 4 illustrate these deduced *Pseudomonas* polypeptides' extensive amino acid sequence similarity to the *trpB* and *trpA* gene products of *E. coli* and *Bacillus subtilis,* as well as to the B and A domains of the fused tryptophan synthase gene in *Saccharomyces cerevisiae.* From this we infer that the sequences are homologous.

The 4-base cistron overlap between *trpB* and *trpA* and *P. aeruginosa* (ATGA) is intermediate between the 1-base overlap seen in *E. coli* and its close relatives (TGATG) and the 8-base overlap found in *B. subtilis* ATGTTTAA). In all three cases the *trpA* start codon is preceded by an acceptable ribosome binding site, in this case AGGAG at bp 1,224–1,228 comprising the glutamine and glutamate codons near the end of *trpB.* On the grounds of similarity of structure it would not be unreasonable to assume that translational coupling occurs between the two *Pseudomonas* genes as it does for the corresponding enteric bacterial ones (Das and Yanofsky 1984), but the critical experiments, especially those involving frameshift mutants and ablation of the ribosome binding site, remain to be done.

Although little protein structural information exists for *P. aeruginosa* tryptophan synthase, the enzyme from its close relative *P. putida* (Palleroni et al. 1973) has been studied to some extent. For the β subunit C-terminal alanine was identified (Maurer and Crawford 1971). (The C-terminal residue is isoleucine in *E. coli* and valine in *B. subtilis.*) The molecular weight of the *P. putida* β_2 dimer found by acrylamide gel

Downloaded from https://academic.oup.com/mbe/article/3/2/191/1086555 by guest on 20 August 2022

FIG. 2.—DNA sequence of *trpB* and *trpA* from *Pseudomonas aeruginosa,* along with 34 bp of 5′-flanking and 799 bp of 3′-flanking sequences. Only the coding strand is shown. Deduced amino acid residues for the *trpB* and *trpA* gene products are shown above the DNA sequence, with stop codons indicated by asterisks. The residue numbering is determined by the alignments shown in figs. 3 and 4, not the consecutive

electrophoresis and ultracentrifugation, 86,000 ± 2,000 (Maurer and Crawford 1971), is in close agreement with that calculated from the *P. aeruginosa* sequence, 87,002. The β-subunit amino acid composition also shows good agreement with the deduced sequence. The 23-residue pyridoxyl peptide from *P. putida* shows near identity with the sequence at bp 278–346, the sole replacements being valine for isoleucine at positions 12 and 19 and arginine for lysine in position 23 of the peptide.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

sequence; location of the gaps introduced to increase similarity is shown by ( ) in the residue-numbering line. The location of areas of potential secondary structure in the mRNA is shown by under- or overlining near the 3′ end of *trpA*. The observed end points of three deletions in plasmid pZAZ101 are indicated by the designations Δ2, Δ9, and Δ14.

For the α subunit the *P. putida* results include the molecular weight (29,100 ± 1,500 [Enatsu and Crawford 1971]; 28,385 by calculation from the DNA sequence) as well as the sequence of the first 50 residues determined by Edman degradation (Crawford and Yanofsky 1971). This partial sequence is given in figure 4. With the exception of amidated residues, which are often difficult to identify as phenylthio-

Table 1
Codon Usage in *Pseudomonas aeruginosa* Tryptophan Synthase

Body	Third Base			
	T	C	A	G
TT	0	23	0	4
CT	0	14	0	46
AT	0	37	0	18
GT	1	23	1	22
Subtotal	1	97	1	90
TA	2	15	0	0
CA	4	20	3	21
AA	1	17	3	19
GA	3	37	18	26
Subtotal	10	89	24	66
TC	1	8	0	5
CC	0	7	0	23
AC	1	32	0	4
GC	0	53	3	23
Subtotal	2	100	3	55
TG	0	8	2	2
CG	0	33	1	5
AG	1	9	0	1
GG	4	57	3	6
Subtotal	5	107	6	14

NOTE.—GC content: total = 68.0%; first base = 68.2%; second base = 43.6%; and third base = 92.2%.

hydantoin derivatives following Edman degradation, there are only eight replacements, none in highly conserved residues of the α subunit.

Sequence Distal to *TrpA*

We sequenced 799 bp beyond *trpA*. On the strand coding for *trpB* and *trpA* there is a 128-codon open reading frame extending from an ATG at bp 2,225 to bp 2,611. It does appear to have a suitably positioned ribosome binding site. In a different reading frame, open to the end of the sequence, there is an ATG codon at bp 2,646 separated by 5 bp from a possible ribosome binding site (GAGG). The codon usage in these two possible genes shows a bias for G or C in the third position similar to that shown by the *trpB* and *trpA* genes. On the complementary strand can be found several short open reading frames, but none with well-positioned start codons or ribosome binding sequences. Which, if any, of these putative genes is actually translated in *P. aeruginosa* remains to be ascertained.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

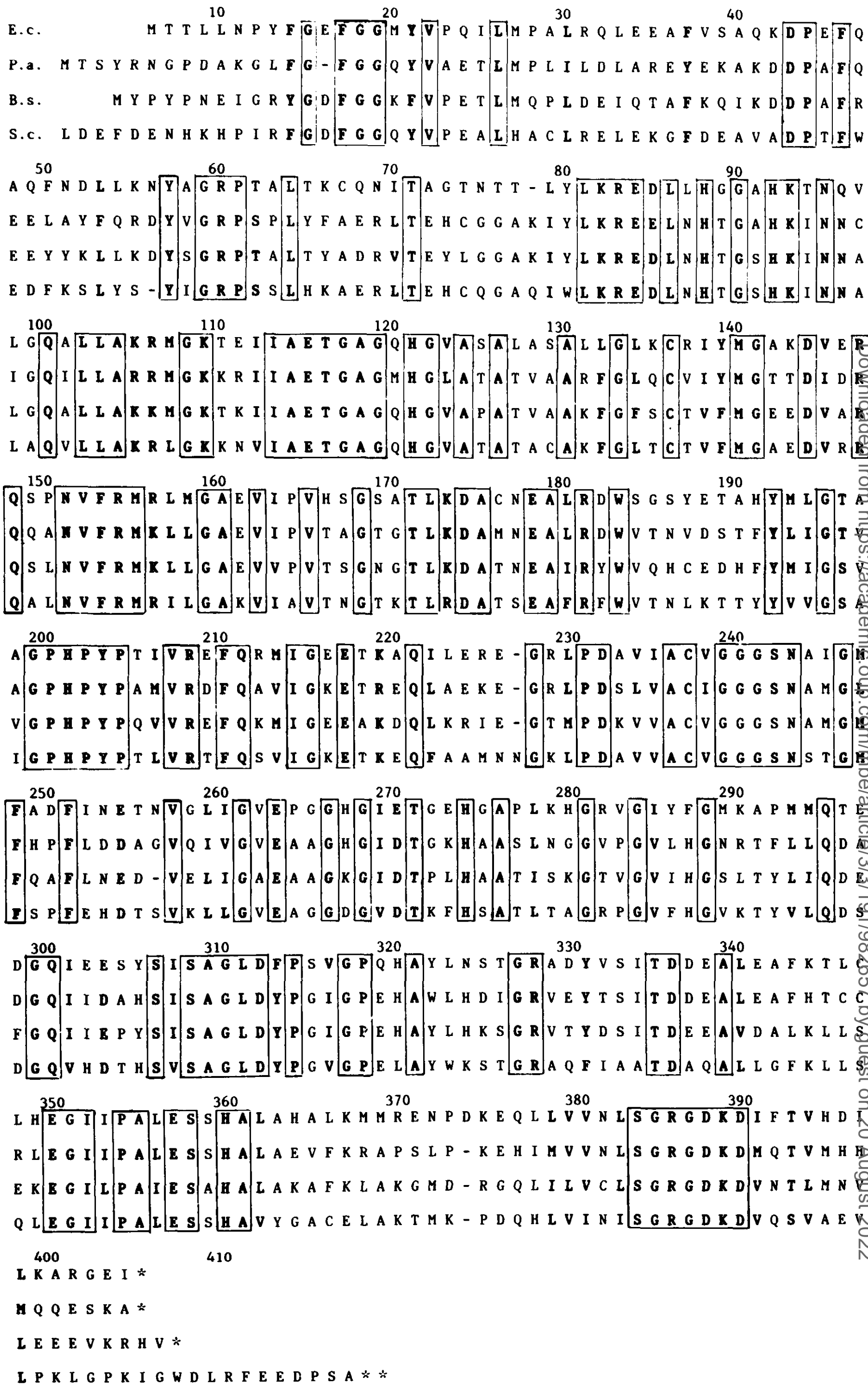

FIG. 3.—Sequence comparison for the β chain of tryptophan synthase in *E. coli*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and the distal 60% of the fused tryptophan synthase polypeptide in *Saccharomyces cerevisiae*. Single-letter abbreviations are used, with dashes indicating gaps introduced to improve alignment. Positions with identical residues are boldface and boxed; those with similar residues are boldface. Similarity groupings: D = E; F = Y; N = Q; K = R; S = T; and I = L = V = F = M.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

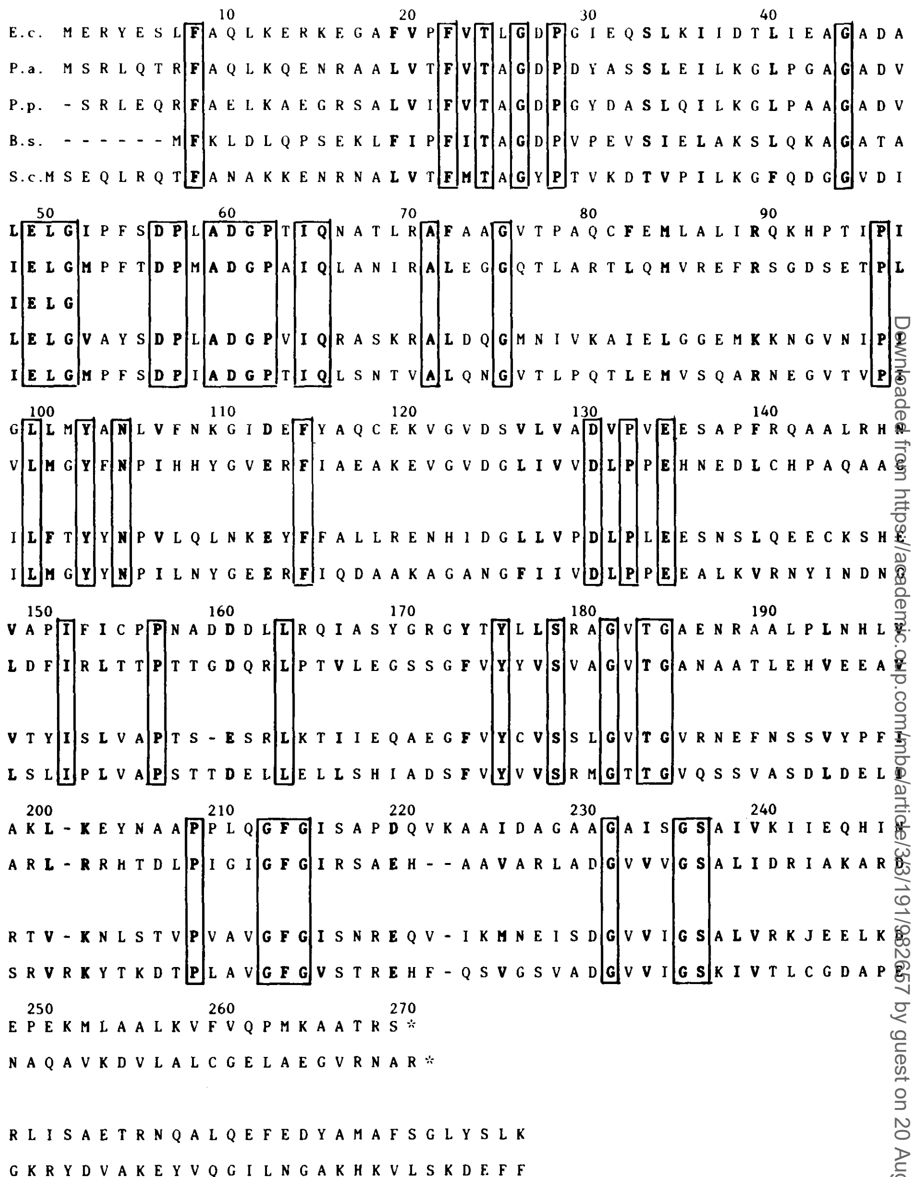

FIG. 4.—Sequence comparison for the α chain of tryptophan synthase in *E. coli, Pseudomonas aeruginosa, Bacillus subtilis,* and the proximal 40% of the fused tryptophan synthase polypeptide in *Saccharomyces cerevisiae.* The initial 50 residues of the *P. putida* α chain are also included in the comparison. Symbols are as in fig. 3.

The *trpA* gene is expected to be near the end of the transcript of the tryptophan synthase gene pair in *P. aeruginosa,* a position like that of its homologue in the *E. coli trp* operon. We searched the entire 799-bp post-*trpA* region for areas of secondary structure resembling *trp t,* the rho-independent terminator in *E. coli* (Wu and Platt 1978). (*Trp t* is more effective as a terminator in vitro than in vivo. It is followed by a rho-dependent terminator, *trp t′,* that is responsible for most of the termination in vivo [Wu et al. 1981], but the characteristic structure of *trp t*—i.e., a stable stem and

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

loop followed by a run of T's—makes it a more recognizable feature.) The two strongest stem-loop structures found are located in tandem just downstream from the *trpA* stop codon between bp 2,061 and 2,132 (fig. 2). The sequence AATTTT following the second stem, marked b-b′ in figure 2, is quite suggestive of a terminator, especially in view of the high GC content of *Pseudomonas* DNA, but the significance of this entire structure is still speculative. Two weak possible stem-loop structures are also apparent in the vicinity of the a-a′ stem-loop. These are marked c-c′ and d-d′ below the sequence in figure 2. They would be mutually exclusive with a-a′, of course.

Sequence of Deletions Lacking Tryptophan Synthase

Previously a variant plasmid containing the *trpB* and *trpA* genes released from their normal inducible regulation and transcribed constitutively from the pBR322 Tet promoter was synthesized and designated pZAZ101 (Manch and Crawford 1981). This plasmid contains a unique *Mlu*I site at position 1,325 in the *trpA* gene. In an attempt at in vitro mutagenesis at this location we isolated several plasmids containing large deletions. When transformed into *E. coli* cells lacking tryptophan synthase, these plasmids conferred the ampicillin resistance but not the indole utilization conferred by normal pZAZ101. Since deletions are uncommon among spontaneous tryptophan auxotrophs of *Pseudomonas,* we studied three of these with care. Two of them proved to have deletions consistent with the DNA manipulations employed, since they terminate within the *Mlu*I recognition sequence. The other was more interesting.

DNA sequencing showed absences of 679 bp in the largest of these deletions, termed pZAZ101Δ14; 428 bp in the next largest, pZAZ101Δ2; and 315 bp in the smallest one, pZAZ101Δ9. The locations of the deletion end points are shown in figure 2. Deletions Δ14 and Δ2 extend from different sites in the *trpB* gene (bp 721 and bp 898, respectively) to the same position in *trpA* (bp 1,326, the second base in the *Mlu*I recognition site). This result could be reconciled with the ligation of a fortuitous, nuclease-caused lesion in *trpB* to the polymerase-filled *Mlu*I site in *trpA,* an outcome compatible with the experimental manipulations. The Δ9 deletion must have had a different origin, however, and examination of its sequence shows that it begins and ends somewhere within the 7-bp direct repeats found within positions 1,132–1,138 in *trpB* and 1,444–1,450 in *trpA*. This may have been the result of an illegitimate recombination event occurring during the growth of the plasmid. Each of the three deletions spans the *trpB-trpA* junction and alters the reading frame, which accounts for the phenotype observed.

Construction of an α Subunit Overproducer

The fortuitous location of a *Pst*I site involving codons 4 and 5 of *trpA* suggested the possibility of fusing the protein with a short segment of the *lac* α peptide produced in the pUC series of expression vectors designed by Vieira and Messing (1982). Inspection of the sequence of pUC8 shows that insertion of a 1,089-bp *Pst*I fragment (*Pst*I sites located at bp 1,246 and 2,338) from pZAZ167 in the correct orientation should result in a polypeptide having the initial (Met) Ser Arg residues of the *trpA* gene product replaced by a (Met) Thr Met Ile Thr Asn Ser Arg Gly Ser Val Asp undecapeptide. The construction was accomplished by ligating the appropriate acrylamide gel–eluted *Pst*I fragment of pZAZ167 to pUC8 cut with *Pst*I and treated with calf intestinal phosphatase (Maniatis et al. 1982) to discourage self-ligation. After transformation into *E. coli* JM83, white colonies on β-galactosidase indicator plates (Vieira and Messing 1982) were screened for plasmids with inserts of the correct size

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

in either orientation. When one of each orientation was transformed into an *E. coli* strain whose *trpA* gene had been deleted (IC151 *his pro tonB*Δ*trpA4*), only the plasmid with its insert in the predicted orientation produced ampicillin-resistant colonies able to grow without the addition of tryptophan to the medium. Growth without tryptophan was clearly slower than that occurring in its presence, and the cells accumulated large amounts of indoleglycerol, indicating some deficiency in tryptophan synthase activity. The significance of this interspecific subunit complementation is discussed below.

Discussion

Gene Structure

The level of amino acid sequence conservation found along the entire length of the tryptophan synthase subunits from three diverse prokaryotic and one eukaryotic species (figs. 3 and 4) is considerable: 35.5% of the residues in *trpB* and 15.6% of those in *trpA* are absolutely conserved. Although at present there does not appear to be a consensus in the literature concerning similar, frequently interchangeable residues, one possible collocation results in an increase in similarity to 51.8% for *trpB* and 31.2% for *trpA* (see legend to fig. 3). To make these *trpB* comparisons we increased similarity via the introduction of single-codon gaps at positions 16 for *Pseudomonas aeruginosa,* 56 for *Saccharomyces cerevisiae,* 78 for *E. coli,* 256 for *Bacillus subtilis,* 227 for all except *S. cerevisiae,* and 374 for all except *E. coli.* For the *trpA* sequences we introduced single-codon gaps at positions 159 for *B. subtilis,* 201 for all except *S. cerevisiae,* and 222 for *B. subtilis* and *S. cerevisiae,* as well as a two-codon gap at 221–222 for *P. aeruginosa;* all positions are in the distal half of the gene.

The greater evolutionary conservation of sequence in the β chain compared with the α chain of the same molecule has been noted earlier in immunological (Rocha et al. 1972) and sequence (Crawford et al. 1980) comparisons of the two subunits in several enteric bacterial species. The present result shows that this trend persists over much wider evolutionary distances.

There is one gene pertinent to the tryptophan pathway for which more comparative information exists than for *trpA* and *trpB*—i.e., that for the paralogous glutamine amidotransferase subunits of anthranilate and p-aminobenzoate synthases (Kaplan et al. 1985). In that comparison larger gaps of three and six residues are required at two locations to improve sequence similarity, which also seems to be localized more to six or seven distinct regions of the approximately 200-residue molecule.

The results reported here imply that the secondary structure of tryptophan synthase may be conserved throughout during evolution, with the possible exception of the C-terminal 30 residues of the α subunit, where little sequence similarity is found. We have subjected these sequences to several secondary-structure prediction programs (data not shown) and noted considerable similarities through the β and much of the α chain for all four organisms. Given the empirical nature of these predictions, however, and the fact that the three-dimensional structure of the molecule is not yet available for any species, we have deferred a more critical analysis of this matter. Instead, in view of the extensive information on catalytically essential residues in the *E. coli* enzyme (Yanofsky and Crawford 1972; Miles 1979), we will discuss their conservation in the other three species.

The β_2 subunit of the *E. coli* enzyme can be cleaved by limited trypsin digestion into two independently folding domains, F1 and F2. This cleavage can occur at any of three basic residues numbered 279, 282, and 290 in figure 3. None is preserved as a lysine or arginine residue in all four species. It is interesting, however, that the three

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982257 by guest on 20 August 2022

glycine residues in this "linking" peptide are absolutely conserved and that residue 285 is maintained as either isoleucine or valine, facts suggesting that the sequence in this region of the polypeptide is not entirely free to change.

Most of the catalytically important residues of the β_2 subunit, including the cofactor binding site, are in the F1 domain. The lysine at position 93 that forms a Schiff's base with pyridoxal phosphate is conserved, as are both nearby histidine residues at 88 and 92, one of which is known to play an important role in catalysis (Miles 1979). An essential cysteine residue at 237 is also conserved. Although the cysteine at position 136 is conserved too, it has not been shown that its function is essential. The N-ethylmaleimide-reactive cysteine at 176, which has not been directly implicated in catalysis, has not been conserved. An arginine at 154 has been shown by Tanizawa and Miles (1983) to interact with the carboxyl group of the substrate L-serine; it is conserved and lies within a very highly conserved region extending from position 151 to position 160. Inactivating mutations occur at position 122 (glycine to glutamate; L. Eberle and I. P. Crawford, unpublished results) and near the C-terminus at position 382 (lysine to aspartate; Cotton and Crawford 1972); both of these residues are completely conserved. The tryptophan content of the *trp* operon genes in *E. coli* (Yanofsky et al. 1981) and *B. subtilis* (Henner et al. 1984) is quite low, which is logical for a protein set that is expressed only in response to tryptophan deprivation. The single tryptophan residue in the β subunit of these organisms is conserved in *P. aeruginosa* and *S. cerevisiae* (position 183), implying a possibly critical role for it. The tryptophan residue at position 322 in the *P. aeruginosa trpB* gene product, on the other hand, is a tyrosine in the other three organisms.

In the *E. coli* α subunit the arginine at position 179 is thought to interact with the phosphate of the substrate indoleglycerol phosphate. It is not conserved in *P. aeruginosa* and *B. subtilis,* and it is not obvious which basic residue in the sequence plays this role. The arginine at 188 is the site of cleavage by trypsin to yield two domains, α1 and α2 (Higgins et al. 1979); this residue is not conserved. Missense mutations have been observed at the following positions: Phe 22, Glu 49, Tyr 175, Leu 177, Thr 183, Gly 212, Gly 235, and Ser 236 (Yanofsky and Crawford 1972). These residues have all been conserved in the other species, except for Leu 177, for which valine seems to be fungible. There is no obvious similarity in the α-chain sequences beyond position 240, and the point where the role of the *S. cerevisiae* polypeptide changes from that of an α-chain homologue to that of a linker to the β-chain homologue cannot be established by inspection.

Codon usage in these two *P. aeruginosa* genes is skewed toward those codons having G or C in the third position (table 1). The sole exception to this generalization, and the amino acid responsible for nearly one-third of the total codons ending in A or T, is glutamine, where GAA is used 18 of 44 times. That the average GC content of these structural genes is 68%, while that of the genome as a whole is reported to be 65% (Shapiro 1968), suggests that regulatory regions might have GC contents < 65%, which is not unexpected. Further elimination of codons ending in A or T obviously could make only a small additional increase in total GC content, and the shift from first-base T- to first-base C-containing leucine codons and from first-base A- to first-base C-containing arginine codons has already largely transpired. The total lysine content of the two subunits is only 22 residues, compared to 31 for *E. coli,* 44 for *S. cerevisiae,* and 45 for *B. subtilis* (GC contents are 50%, 40%, and 42%, respectively), with an opposite bias in arginine content, suggesting that, in positions where only a basic residue is called for, arginine codons, with their higher GC content, have been preferred. It will be interesting to ascertain the mechanism by which some bacteria,

Downloaded from https://academic.oup.com/mbe/article/3/3/.../992667 by guest on 24 August 2024

such as *Micrococcus lysodeikticus* and certain actinomycetes (Shapiro 1968), attain GC contents as high as 75%.

The striking difference in codon usage between *P. aeruginosa* and *E. coli* does not result in problems in high-level expression from plasmids containing strong *E. coli* promoters. The marked codon bias in *Pseudomonas* appears to be a result of a series of AT-to-GC substitutions occurring chiefly in the third-codon position, a process whose beginnings can be seen when comparing different strains of *E. coli* (Milkman and Crawford 1983) and closely related enteric bacteria (Nichols and Yanofsky 1979; Crawford et al. 1980). A correlate of this is that the extensive amino acid conservation observed in the four very diverse microbes under discussion, especially that seen for the *trpB* gene product, indicates a large proportion of highly constrained residues, whether these constraints be for participation in the active site (as in many of the catalytically critical residues singled out above) or to maintain a necessary three-dimensional conformation.

Heterologous Subunit Complementation

If, as we would argue, the subunits of tryptophan synthase have conserved a particular polypeptide conformation during evolution, it should come as no great surprise that, when the α subunit of *P. aeruginosa* is copiously produced in *E. coli* cytoplasm, it can complement the *E. coli* β_2 subunit and yield a functional enzyme molecule. This is, however, contradictory to information published earlier (Enatsu and Crawford 1971), which indicated that purified α and β_2 subunits from *E. coli* and *P. putida* were ineffective in complementing each other in heterospecific combinations. There are three possible reasons for the divergent results. The earlier work was done with *P. putida* rather than with *P. aeruginosa* subunits, and the latter may be compatible with *E. coli*'s subunits while the former may not. Because the two fluorescent pseudomonads are similar in many ways, including the unusual regulation and disposition of their tryptophan synthase genes, this seems to us unlikely, but it cannot be ruled out, in view of the amino acid sequence differences shown in figure 4. A second explanation is that in the earlier experiments the assays were done in vitro and the interaction of dissimilar subunits may occur more easily in the cytoplasmic milieu than in the test tube. That possibility also has not been ruled out, but the third hypothesis—that the heterologous subunit affinity is low but not zero and that the effective molar concentration of the subunits within the cell is much higher than was attained in the earlier in vitro experiments—seems to us most probable.

Flanking Sequences

The nature of the transcription initiation site in the region ahead of *trpB* has not been investigated in this study for several reasons. Little is known about *Pseudomonas* promoters other than that they may differ considerably from those of *E. coli* (Mermod et al. 1984). This one in particular functioned poorly when first brought into *E. coli* cytoplasm and was subjected to a protracted and ill-defined selection process to increase its activity (Hedges et al. 1977). It is quite possible that the actual promoter now being used in *E. coli* is different in sequence and even in location from the original one.

The existence of a sequence resembling *trp t* of *E. coli* just downstream from *trpA* in *P. aeruginosa* suggests that this structure may have been conserved during evolution, but the fact that in *E. coli* most transcription does not stop at *trp t* and that its main function may be to retard exonucleolytic degradation from the 3′ end (Wu et al. 1981) makes an assignment of function to the *P. aeruginosa* structure hazardous.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/993667 by guest on 20 August 2022

Until means are developed to assess the transcriptional response of *Pseudomonas* RNA polymerase to specific sequences, the resemblance of the region between bp 2,111 and 2,138 in figure 2 to a rho-independent terminator in *E. coli* must remain an observation without proved significance. The identity of the initial 16S ribosomal RNA sequences in *Pseudomonas* and the enteric bacteria, however, puts the identification of translational initiation sequences on much firmer ground. The location of Shine-Dalgarno sequences ahead of both the *trpB* and *trpA* initiation codons is quite obvious, and the relationship of the *trpB* stop and *trpA* start codons is reminiscent of, though not identical to, that found in *E. coli* and *Salmonella typhimurium.* It encourages the prediction that there should be translational coupling between the two genes in *Pseudomonas,* as there is in the enteric bacteria.

Acknowledgment

This research was supported by grant AI20279 from the National Institutes of Health.

LITERATURE CITED

CALHOUN, D. H., D. L. PIERSON, and R. A. JENSEN. 1973. The regulation of tryptophan biosynthesis in *Pseudomonas aeruginosa.* Mol. Gen. Genet. **121**:117–132.

COTTON, R. G. H., and I. P. CRAWFORD. 1972. Tryptophan synthetase β_2 subunit: application of genetic analysis to the study of primary structure. J. Biol. Chem. **247**:1883–1891.

CRAWFORD, I. P. 1975. Gene rearrangements in the evolution of the tryptophan synthetic pathway. Bacteriol. Rev. **39**:87–120.

CRAWFORD, I. P., M. DECASTEL, and M. GOLDBERG. 1978. Assignment of the ends of the β-chain of *E. coli* tryptophan synthase to the F1 and F2 domains. Biochem. Biophys. Res. Commun. **85**:309–316.

CRAWFORD, I. P., and I. C. GUNSALUS. 1966. Inducibility of tryptophan synthetase in *Pseudomonas putida.* Proc. Natl. Acad. Sci. USA **56**:717–724.

CRAWFORD, I. P., B. P. NICHOLS, and C. YANOFSKY. 1980. Nucleotide sequence of the *trpB* gene in *Escherichia coli* and *Salmonella typhimurium.* J. Mol. Biol. **142**:489–502.

CRAWFORD, I. P., and C. YANOFSKY. 1971. *Pseudomonas putida* tryptophan synthetase: partial sequence of the α subunit. J. Bacteriol. **108**:248–253.

DAS, A., and C. YANOFSKY. 1984. A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res. **12**: 4757–4768.

ENATSU, T., and I. P. CRAWFORD. 1971. *Pseudomonas putida* tryptophan synthetase. J. Bacteriol. **108**:431–438.

GUNSALUS, I. C., C. F. GUNSALUS, A. M. CHAKRABARTY, S. SIKES, and I. P. CRAWFORD. 1968. Fine structure mapping of the tryptophan genes in *Pseudomonas putida.* Genetics **60**:419–435.

HEDGES, R. W., A. E. JACOB, and I. P. CRAWFORD. 1977. Wide ranging plasmid bearing the *Pseudomonas aeruginosa* tryptophan synthase genes. Nature **267**:283–284.

HENNER, D. J., L. BAND, and H. SHIMOTSU. 1984. Nucleotide sequence of the *Bacillus subtilis* tryptophan operon. Gene **34**:169–177.

HIGGINS, W., T. FAIRWELL, and E. W. MILES. 1979. An active proteolytic derivative of the α subunit of tryptophan synthase: identification of the site of cleavage and characterization of the fragments. Biochemistry **18**:4827–4835.

KAPLAN, J. B., W. K. MERKEL, and B. P. NICHOLS. 1985. Evolution of glutamine amidotransferase genes: nucleotide sequences of the *pabA* genes from *Salmonella typhimurium, Klebsiella aerogenes* and *Serratia marcescens.* J. Mol. Biol. **183**:327–340.

LAGRIMINI, L. M., S. T. BRENTANO, and J. E. DONELSON. 1984. A DNA sequence analysis package for the IBM personal computer. Nucleic Acids Res. **12**:605–614.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/982657 by guest on 20 August 2022

MANCH, J. M., and I. P. CRAWFORD. 1981. Ordering tryptophan synthase genes of *Pseudomonas aeruginosa* by cloning in *Escherichia coli*. J. Bacteriol. **146**:102–107.

———. 1982. Genetic evidence for a positive-acting regulatory factor mediating induction in the tryptophan pathway of *Pseudomonas aeruginosa*. J. Mol. Biol. **156**:67–77.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MATCHETT, W., and J. A. DEMOSS. 1975. The subunit structure of tryptophan synthase form *Neurospora crassa*. J. Biol. Chem. **250**:2941–2946.

MAURER, R., and I. P. CRAWFORD. 1971. Properties and subunit structure of the B component of *Pseudomonas putida* tryptophan synthetase. Arch. Biochem. Biophys. **144**:193–203.

MAXAM, A. M., and W. GILBERT. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.

MERMOD, N., P. R. LEHRBACH, W. REINEKE, and K. N. TIMMIS. 1984. Transcription of the TOL plasmid toluate catabolic pathway operon of *Pseudomonas putida* is determined by a pair of coordinately and positively regulated overlapping promoters. EMBO J. **3**:2461–2466.

MILES, E. W. 1979. Tryptophan synthase: structure, function, and subunit interaction. Adv. Enzymol. Relat. Areas Mol. Biol. **49**:127–186.

MILKMAN, R., and I. P. CRAWFORD. 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. Science **221**:378–380.

NICHOLS, B. P., and C. YANOFSKY. 1979. Nucleotide sequences of *trpA* of *Salmonella typhimurium* and *Escherichia coli*: an evolutionary comparison. Proc. Natl. Acad. Sci. USA **76**: 5244–5248.

PALLERONI, N. J., R. KUNISAWA, R. CONTOPOULOU, and M. DOUDOROFF, 1973. Nucleic acid homologies in the genus *Pseudomonas*. Int. J. Syst. Bacteriol. **23**:333–339.

ROCHA, V., I. P. CRAWFORD, and S. E. MILLS. 1972. Cooperative immunological and enzymatic study of the tryptophan synthase β_2 subunit in the *Enterobacteriaceae*. J. Bacteriol. **111**:163–168.

SANGER, F., and A. R. COULSON. 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.

SHAPIRO, H. S. 1968. Distribution of purines and pyrimidines in deoxynucleic acids. Pp. H31–H36 *in* H. A. SOBER, ed. Handbook of biochemistry: selected data for molecular biology. CRC, Cleveland.

TANIZAWA, K., and E. W. MILES. 1983. L-serine binds to arginine-148 of the β_2 subunit of *Escherichia coli* tryptophan synthase. Biochemistry **22**:3594–3603.

TSAI, H., C. Y. YANG, and J. H. J. TSAI. 1974. The subunit structure of *Neurospora* tryptophan synthase: a reappraisal. Biochem. Biophys. Res. Commun. **61**:1332–1339.

VIEIRA, J., and J. MESSING. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.

WU, A. M., G. E. CHRISTIE, and T. PLATT. 1981. Tandem termination sites in the tryptophan operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:2913–2917.

WU, A. M., and T. PLATT. 1978. Transcription termination: nucleotide sequences at 3′ end of the tryptophan operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **75**:5442–5446.

YANOFSKY, C., and I. P. CRAWFORD. 1972. Tryptophan synthetase. Pp. 1–31 *in* P. D. BOYER, ed. The enzymes. Vol. **7**. 3d ed. Academic Press, New York.

YOUSAF, S. I., A. R. CARROLL, and B. E. CLARKE. 1984. A new and improved method for 3′-end labeling DNA using [α-^{32}P]ddATP. Gene **27**:309–313.

ZALKIN, H., and C. YANOFSKY. 1982. Yeast gene TRP5: structure, function, regulation. J. Biol. Chem. **257**:1491–1500.

WALTER M. FITCH, reviewing editor

Received September 3, 1985; revision received December 5, 1985.

Downloaded from https://academic.oup.com/mbe/article/3/3/191/1026574 by guest on 20 August 2022