The Nucleotide Sequence of the Structural Gene for *Escherichia coli* Tryptophanyl-tRNA Synthetase*

(Received for publication, January 25, 1982)

Carol V. Hall§, Magda vanCleempuut, Karl H. Muench¶, and Charles Yanofsky†

From the §Department of Biological Sciences, Stanford University, Stanford, California 94305 and the ¶Division of Genetic Medicine, Departments of Medical and Biochemistry, University of Miami School of Medicine, Miami, Florida 33101

The complete nucleotide sequence of *trpS*, the structural gene for *Escherichia coli* tryptophanyl-tRNA synthetase, was determined using a plasmid carrying the structural gene. From the single open reading frame of correct length and orientation we deduced an amino acid sequence consistent with the amino acid composition of the purified protein. In addition, previously sequenced peptides representing 52% of the protein were readily aligned with regions of the deduced sequence.

The deduced amino acid sequence of the *E. coli* enzyme is 60% homologous with the sequence of the enzyme from *Bacillus steaerothermophilus*. Using currently available procedures we predicted the secondary structure for the enzyme from each organism and compared these structures to those of the two aminoacyl-tRNA synthetases whose three-dimensional structures have been determined. We used a convenient plasmid recombination procedure to map the regional locations of missense mutations within *trpS* that have characteristic effects on the properties of the enzyme.

**Bacterial Strains and Plasmids**—The *E. coli* K12 strains used in this study (Fig. 4) have been described previously (11, 12). The construction of pCH6 and pCH17 have also been described (6).

**Plasmid Construction**—The four plasmids used to map the *trpS* mutants are all derived from pCH6, an 8-kb plasmid which contains *trpS* and its regulatory region cloned into plasmid pBR322 (6) (Fig. 4). The regulatory region and proximal half of *trpS* were deleted to obtain pCH20 by digestion of pCH6 with *Bam* H1 and *Bgl* II followed by recircularization of the vector-containing fragment. The distal half of *trpS* was deleted by digestion of pCH6 with *Bgl* II and *Eco* RI, filling in the cohesive ends with DNA polymerase I and the four deoxynucleoside triphosphates and recircularization of the blunt ends to obtain pCH27. To construct pCH20, a plasmid containing the proximal segment of *trpS*, pCH6 was digested with *Sal* I followed by *Hinc* II. A 700-bp *Sal* I/*Hinc* II fragment containing the regulatory region and first 120 bp of *trpS* was isolated from a 5% polyacrylamide gel. This fragment was then ligated to pBR322 digested with *Sal* I and *Pvu* II. pCH30, which contains the portion of *trpS* from bp 124-762, was constructed by first purifying the *Hinc* II/*Sau* 96I fragment, which spans this region in *trpS*. The *Sau* 96I staggered ends were filled in with DNA polymerase I and the four deoxynucleoside triphosphates to generate a blunt end fragment, which was ligated into the *Pvu* II site of pBR322.

To obtain each of the above plasmids, the ligation mixture was used to transform *E. coli* strain W3110. Colonies were selected on rich medium containing ampicillin (40 μg/ml). Plasmid DNA from individual ampicillin-resistant (Ap') colonies was isolated and subjected to restriction analysis to characterize the desired plasmid.

**Mapping trpS Mutants**—Each of the *trpS* strains in Fig. 4 was transformed separately with pCH20, pCH27, pCH29, or pCH30. Ap' colonies were selected on rich medium containing ampicillin. Several individual colonies were scraped from the plate, and each was resuspended in 0.5 ml of minimal medium (13). Serial dilutions were then plated on either rich medium containing ampicillin or selective medium (minimal medium supplemented with 0.5% casein hydrolysate) containing ampicillin to determine the frequency of tryptophan prototrophs arising by recombination. Reversion frequencies for each *trpS* recipient strain were determined in control experiments.

**RESULTS AND DISCUSSION**

A 30-kb ColE1 plasmid, pSC13-42, has been shown previously to contain the structural gene for *E. coli* tryptophanyl-tRNA synthetase (6). We subcloned a 2.8-kb *Sal* I/*Eco* RI fragment derived from this plasmid which retains *trpS* complementing activity into pBR322. The resulting plasmid, pCH17, was the source of DNA for sequence determinations. After identifying the location and orientation of *trpS* on pCH17 (6), we isolated overlapping restriction fragments which span the 1-kb region containing the gene (Fig. 1). The restriction fragments were sequenced on both strands using the Maxam and Gilbert technique. Fragments were first end labeled with [*α-32P]*ATP and polynucleotide kinase followed either by restriction with a second enzyme or by strand separation to obtain singly end labeled fragments.

Using the *trpS* promoter to define the 5' end of the coding

---

*This research was supported by grants from the United States Public Health Service (GM09738), The National Science Foundation (PCM77-24333), and the American Heart Association (68-015). The costs of publication of this article were deferred in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Predoctoral trainee of the United States Public Health Service.

¶ Career Investigator of the American Heart Association.

1 The abbreviations used are: kb, kilobase; bp, base pairs.
region, we found a single open reading frame of the predicted length within the sequenced 1-kb region; from it we deduced the amino acid sequence which is presented in Fig. 2. We confirmed the identity of the coding region by the following three criteria. First, in E. coli B 32 independent tryptic peptides with 211 amino acid residues have been sequenced (21). Of these peptides, 26 with 175 amino acid residues are identical or homologous to peptide sequences deduced for the tryptophanyl-tRNA synthetase of E. coli K12. As shown in Fig. 2, the 26 peptides are scattered throughout the deduced sequence; therefore, no nucleotides have been added or deleted to alter the reading frame. Furthermore, the NH₂-terminal amino acid sequence of the protein unambiguously defines the start of the gene. Second, the amino acid composition and molecular weight calculated from the deduced amino acid sequence are in good agreement with those determined experimentally (Table I). Finally, the amino acid sequence of the tryptophanyl-tRNA synthetase from Bacillus stearothermophilus (14) can be aligned with the deduced E. coli sequence with an overall homology of 60% (Fig. 2).

**Codon Usage**—From nucleotide sequence comparisons of prokaryotic protein structural genes two patterns of codon usage have been identified (15-17). One pattern appears to be characteristic of genes expressed at low to moderate levels. These genes contain most of the codons but show a preference for codons of the major isoaccepting species of certain tRNAs. In the second pattern, which is seen in highly expressed genes, notably the ribosomal protein genes and outer membrane protein genes, codons for the major isoaccepting tRNA species are used exclusively. The explanation for the difference is not known, but it has been suggested that the use of codons for the minor isoaccepting tRNA species would reduce the rate of translation.

The pattern of codon usage in trpS (Table II) shows the same strong preference for CGPy(Arg), GGPy(Gly), and ATPy(Ile) that is found in the highly expressed genes (16, 18). The frequency is 100% in each case, and consequently the codons CGPu(Arg), AGPu(Arg), GGPu(Gly), and ATPu(Ile) are the only codons used.

---

**Fig. 1.** Restriction endonuclease sites and restriction fragments used to determine the nucleotide sequence of trpS. The arrows indicate the direction and length of sequence determined from each restriction site. The bar represents the coding region of trpS; nucleotide pair 1 is the A of the AUG start codon.

**Fig. 2.** Nucleotide and amino acid sequence of E. coli tryptophanyl-tRNA synthetase. Below are shown only those amino acid differences or deletions (X) in the B. stearothermophilus amino acid sequence. The bars above the nucleotide sequence span sequenced tryptic peptides of the E. coli B tryptophanyl-tRNA synthetase (21).
are absent. The strong preference for GAA(Glu) (80%) in the trpS sequence also resembles that found in highly expressed genes (18). Although the preferences for CUG(Leu) (70%) and CCG(Pro) (71%) are not as pronounced in trpS as in the highly expressed genes, there is a greater bias toward the use of these two codons than found in moderately expressed genes (50% combined (19)).

Despite the preference in trpS for codons of the major isoaccepting tRNA species, the estimated number of molecules of tryptophanyl-tRNA synthetase in the cell is approximately 10-fold lower than the number of ribosomes (5, 20). Thus, while codon usage in trpS would appear to support approximately 10-fold lower than the number of ribosomes (5, 20). However, recent investigations have revealed that the trpS sequence also resembles that found in highly expressed genes (19).

Comparison of Amino Acid Sequences—Although E. coli and B. stearothermophilus are only distantly related prokaryotes, their tryptophanyl-tRNA synthetases possess regions of extensive amino acid sequence homology (21). When the two sequences are aligned for maximum homology two-thirds of the residues are conserved. The conserved residues are not distributed randomly; rather, regions with 80-90% homology are scattered among more poorly conserved segments, with the NH2-terminal portions of the proteins more homologous overall than the COOH-terminal portions (Fig. 2). The E. coli amino acid sequence is seven residues longer than the B. stearothermophilus sequence. The additional residues can best be accounted for as internal additions within the E. coli protein. In Fig. 2 we have indicated the seven residues of the E. coli sequence which, when deleted, result in maximum homology between the two sequences.

Key Residues in Tryptophanyl-tRNA Synthetase—The four cysteine residues in E. coli tryptophanyl-tRNA synthetase are clustered in the NH2-terminal third of the protein, at positions 37, 40, 70, and 97 (Fig. 2). Three of these residues are conserved in the B. stearothermophilus sequence; the fourth, the cysteine residue at position 70, is replaced by a valine residue. In the E. coli enzyme, previous studies (4, 22)

revealed a single cysteine residue whose reaction with 5,5'-dithiobis(2-nitrobenzoate) totally abolished enzyme activity. Subsequent treatment with dithiothreitol completely restored activity, with concomitant release of a single 5-carboxy-4-nitrothiophenol molecule per cysteine residue. The inactivation was totally prevented by the presence of tryptophan, ATP, and Mg2+. The precise stoichiometry, reversibility, and substrate protection indicate that this cysteine residue may be at the active site. The cysteine residue at position 70 is the most probable candidate for the required cysteine, because the B. stearothermophilus enzyme, lacking only cysteine residue 70 of the four in the E. coli enzyme, is unreactive with and unaffected by 5,5'-dithiobis(2-nitrobenzoate). The thiol of cysteine residue 70 obviously plays no direct catalytic role, because replacement of the cysteine by valine in the B. stearothermophilus enzyme is inconsequential. Although cysteine residue 70 is not conserved in B. stearothermophilus, there is otherwise total homology between the two sequences in this region (residue numbers 65 to 74 in Fig. 2).

Secondary Structure Prediction—In view of the extensive homology between the amino acid sequences of tryptophanyl-tRNA synthetase from E. coli and B. stearothermophilus we expect the secondary and tertiary structures of these enzymes to be similar (23, 24). We used the rules developed by Garnier et al. (10) and a computer program provided by R. Stroud (Dept. of Biochemistry, Univ. of California at San Francisco) to predict the secondary structure of each protein.

From the results of the computer analysis we predict that for the E. coli protein, 55% of the residues exist in α helices and 35% in β strands; for the B. stearothermophilus protein 44% of the residues are in α helices and 30% in β strands. The predicted secondary structures are drawn in Fig. 3. Approximately 80% of the secondary structure is invariant (indicated by boxes in Fig. 3). The conserved secondary structures are not due to regions of high amino acid homology because amino acid conservation in these areas is no higher than for the protein overall. The invariant secondary structure regions suggest that tryptophanyl-tRNA synthetase is composed of a region of alternating α helices and β strands followed by an α helical segment in the COOH-terminal third of the protein.

Comparison of the amino acid sequence of tryptophanyl-tRNA synthetase with sequences available for other synthetases reveals no significant homologies (2, 25). However, certain similarities are evident in the predicted secondary structure of tryptophanyl-tRNA synthetase when compared with the observed secondary structures of two synthetases.
Structure of trpS

Fig. 3. Predicted secondary structure for tryptophanyl-tRNA synthetase from B. stearothermophilus (top) and E. coli (bottom). Conserved secondary structures are in boxes. ---, coil or turn; , , -helical residue; \( \beta \), \( \beta \)-sheet residue.

Tertiary structure analysis of tyrosyl-tRNA synthetase (a dimer composed of identical 45,000-dalton subunits) and a fully functional monomeric fragment of methionyl-tRNA synthetase (64,000 daltons) have identified a homologous structure within the NH\( \_\)terminal portion of each protein that is thought to contain the ATP binding site (26, 27). Analogous structures have been found in other classes of enzymes that bind nucleotides such as dehydrogenases (28) and kinases (29, 30). These structures, so-called Rossmann folds, are composed of approximately 150-200 amino acid residues organized in four or more strands of parallel \( \beta \) pleated sheet with inter-spersed \( \alpha \) helices. The predicted NH\( \_\)terminal invariant secondary structure of tryptophanyl-tRNA synthetase bears a close resemblance to these nucleotide binding folds.

Mapping of Tryptophanyl-tRNA Synthetase Mutants—In order to identify the different functional regions of the protein and to relate these regions to the predicted secondary structure of the protein we have begun to map the sites of trpS mutations. These mutants were initially isolated as tryptophan auxotrophs and subsequently mapped at the trpS locus (10). Previous characterization of the mutant proteins showed that some have 10-fold reduced affinity for tryptophan while others are labile in the absence of high levels of tryptophan (10). To map the mutations we constructed a series of plasmids lacking various portions of trpS; pCH29 contains only the distal half of trpS as well as downstream sequences while pCH27 contains the proximal half of trpS and upstream sequences, pCH29 contains only the first 124 bp of trpS and sequences upstream, and pCH30 contains an internal 538-bp fragment from bp 124-762 (Fig. 4). This set of plasmids containing segments spanning the entire trpS gene was used to determine the region within which each trpS mutation is located. Each mutant strain was transformed separately with pCH27, pCH29, pCH30, and pCH30.

<table>
<thead>
<tr>
<th>Strain ( ^a )</th>
<th>Phenotype ( ^b )</th>
<th>In Vivo Complementation ( ^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>trpS10100</td>
<td>tryptophan affinity</td>
<td>+</td>
</tr>
<tr>
<td>trpS9969</td>
<td>reduced</td>
<td>-</td>
</tr>
<tr>
<td>trpS9897</td>
<td>protein labile</td>
<td>-</td>
</tr>
<tr>
<td>trpS10330</td>
<td>without Mg(^{2+}), ATP</td>
<td>+</td>
</tr>
<tr>
<td>trpS10343</td>
<td>and tryptophan</td>
<td>-</td>
</tr>
<tr>
<td>trpS003</td>
<td>in vitro</td>
<td>+</td>
</tr>
<tr>
<td>trpS003</td>
<td>not known</td>
<td>-</td>
</tr>
</tbody>
</table>

\( ^a \) All W3110 derivatives.

\( ^b \) All strains are tryptophan auxotrophs with less than 10% wild-type tryptophanyl-tRNA synthetase activity (10, 31).

\( ^c \) Indicates that Trp\(^+\) prototrophs appeared at a frequency greater than 10\(^{-4}\).

\( ^d \) Indicates that Trp\(^+\) prototrophs appeared at a frequency less than 10\(^{-4}\).

Fig. 4. Mapping of trpS mutants. The chromosomal insert on pCH6 is shown at the top of the figure. Fragments which were subcloned are shown immediately below.

the four plasmids selecting initially for ampicillin resistance. Individual colonies were analyzed to determine the frequency of tryptophan prototrophs arising by recombination between the wild type trpS region present on the plasmid and the
chromosomal segment containing the trpS mutation. The detection of Trp+ recombinants indicates that the plasmid trpS region contains the wild type segment corresponding to the site of the trpS mutation. In this manner we localized seven mutations to three regions within trpS (Fig. 4). Mutations trpS969, trpS10330, and trpS903 are located between bp 124 and 509; trpS10169, trpS10110, and trpS10345 lie between bp 609 and 762. trpS987 does not give recombinants with either pCH20 or pCH27 but does with pCH30. This suggests that the trpS987 lesion is located in the vicinity of the mutation and that the failure to detect recombinants with pCH20 or pCH27 is due to reduced recombination when the mutation is close to an end of the trpS region on the plasmid. All seven mutations examined are located in the N-terminal two-thirds of the gene. Although we cannot as yet draw clear correlations between the approximate sites of mutation and mutant phenotype, the precise localization of the mutations by DNA sequence analysis may provide clues to the location of particular binding and catalytic sites within the protein molecule.

Acknowledgments—We wish to thank Robert Stroud and Janet Moore for their invaluable comments, suggestions, and help with the computer program, and Kasper Kirschner for critically reading this manuscript.

REFERENCES
30. Evens, P. R., and Hudson, P. J. (1979) Nature 279, 500-504
The nucleotide sequence of the structural gene for Escherichia coli tryptophanyl-tRNA synthetase.
C V Hall, M vanCleemput, K H Muench and C Yanofsky


Access the most updated version of this article at http://www.jbc.org/content/257/11/6132

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/257/11/6132.full.html#ref-list-1