Nucleotide Sequence of Escherichia coli pyrG Encoding CTP Synthetase*

Manli Weng, Christopher A. Makaroff†, and Howard Zalkin

From the Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907

The amino acid sequence of Escherichia coli CTP synthetase was derived from the nucleotide sequence of pyrG. The derived amino acid sequence, confirmed at the N terminus by protein sequencing, predicts a subunit of 544 amino acids having a calculated Mr of 60,300 after removal of the initiator methionine. A glutamine amide transfer domain was identified which extends from approximately amino acid residue 300 to the C terminus of the molecule. The CTP synthetase glutamine amide transfer domain contains three conserved regions similar to those in GMP synthetase, anthranilate synthase, p-aminobenzoate synthase, and carbamoyl-P synthetase. The major 5' end of pyrG mRNA was localized to a position approximately 48 base pairs upstream of the translation initiation codon. Translation of the gene eno, encoding enolase, is initiated at 89 base pairs downstream of pyrG. The pyrG-eno junction is characterized by multiple mRNA species which are ascribed to monocistronic pyrG and/or eno mRNAs and a pyrG eno polycistronic mRNA.

CTP synthetase is a glutamine amidotransferase that catalyzes the terminal reaction in the de novo pathway for pyrimidine nucleotide synthesis: UTP + ATP + glutamine \rightarrow CTP + ADP + P\textsubscript{i} + glutamate. Similar to other glutamine amidotransferases (1), NH\textsubscript{3} can replace glutamine in which case the products are CTP, ADP, and P\textsubscript{i}. Escherichia coli CTP synthetase is a complex regulatory enzyme that exhibits both positive and negative cooperative effects and is subject to allosteric activation by GTP (2). Activation by GTP results in increased enzyme activity. These effects are ascribed to monocistronic pyrG and/or eno mRNAs and a pyrG eno polycistronic mRNA.

EXPERIMENTAL PROCEDURES

RESULTS

Subcloning pyrG—Plasmid pNP1519 contains pyrG in a 4.5-kb PstI fragment of E. coli DNA cloned in pBR322 (Fig. 1). pyrG was subcloned into vector pUC8 as shown in Fig. 1. Plasmid pMW1 was obtained by ligation of a mixture of BamHI fragments from pNP1519 into the BamHI site of pUC8. Selection for pyrG\textsuperscript{+} was by functional complementation of pyrG in strain JP646. Plasmid pMW5 was constructed by ligating the 2.6-kb SalI-PstI segment of E. coli DNA from pMW1 into the SalI and PstI polylinker sites in pUC8. Further subcloning indicated that DNA at the BamHI and KpnI sites in pMW5 was essential for pyrG function.

DNA Sequence—The DNA sequence of pyrG was initially determined using fragments isolated from plasmid pMW5. Real, HpaII, or TaqI digests of plasmid pMW5 or of the SalI

* This work was supported by United States Public Health Service Grant GM24658. This is Journal Paper No. 10,525 from the Purdue University Agricultural Experiment Station. The costs of publication of this article were defrayed 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.
† Recipient of National Institutes of Health Predoctoral Training Grant GM07211.

1 The abbreviations used are bp, base pair; kb, kilobase pair.
2 Portions of this paper (including "Experimental Procedures" and Figs. 1, 2, 4, 5, and 6) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9550 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-3563, cite the authors, and include a check or money order for $4.40 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.
PstI insert were ligated into M13mp18 or M13mp19 (Fig. 2A). In addition, specific subfragments were isolated from digests that were obtained using restriction enzymes having 6-bp recognition sequences (Fig. 2B). Finally, the exonuclease III procedure (26) was employed to obtain a set of overlapping sequences from NruI-BamHI and BamHI-PstI segments of the cloned DNA (Fig. 2C). The DNA sequence shown in Fig. 3 extends from 11 bp upstream of the NruI site to the downstream PstI site at nucleotide 2442. The entire sequence was determined on both DNA strands from overlapping fragments.

The derived amino acid sequence of CTP synthetase is shown in Fig. 3. The protein chain of 545 amino acid residues has a calculated molecular weight of 60,450. At nucleotides 2074-2076, an ATG initiates an open reading frame that extends 173 codons to the 3' end of the cloned E. coli DNA. By screening protein data banks, the downstream sequence was found to be homologous with that of yeast enolase. Thus, E. coli eno is 89 bp downstream from pyrG.

CTP Synthetase—Enzyme was purified to homogeneity from cells bearing plasmid pHMW5. A single stained protein band was obtained by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. By comparison with proteins of known molecular weight, the CTP synthetase subunit had an estimated molecular weight of 60,000 (Fig. 4).

The 5' end of pyrG mRNA was mapped by the nuclease S1 procedure (29). A 462-nucleotide NruI-BamHI DNA probe was used which extends from nucleotides 12-473 (Figs. 2 and 3). The probe was labeled with [α-32P]dCTP by primer extension or the double-stranded fragment was isolated and 5' end-labeled with [γ-32P]ATP and poly-nucleotide kinase. The results of nuclease S1 mapping are shown in Fig. 5. Two protected fragments of coding strand DNA were obtained (Fig. 5A, lane 2). Noncoding strand DNA did not anneal to RNA and was completely digested (Fig. 5A, lane 4). To confirm that the two transcripts extended into the pyrG coding sequence, nuclease S1 mapping was repeated using the 5' end-labeled NruI-BamHI probe. Fig 5B shows that the same two protected fragments were obtained. The size of the major protected fragment is about 175 nucleotides, and the minor one approximately 255 nucleotides. More precise mapping was obtained by using a DNA sequencing ladder as a size standard (Fig. 5C). These results confirm those obtained with restriction fragment size standards. Corresponding sites for transcription initiation are overlaid in Fig. 3.

The 3' end of pyrG mRNA was mapped with a KpnI-PstI DNA probe that extends from nucleotides 1663-2442 (Figs. 2 and 3). The results of nuclease S1 mapping are shown in Fig. 6. The major products obtained from the coding strand were undigested probe and a fragment of approximately 400 nucleotides (Fig. 6, lane 2). Minor fragments of approximately 420, 530, and 650 nucleotides were also obtained. The same pattern of fragments was obtained when the nuclease S1 concentration was increased 3-fold (data not shown). The noncoding strand DNA probe did not anneal to RNA and was completely digested (Fig. 6, lane 4). These results indicate that there are multiple species of pyrG and eno mRNA. An mRNA that anneals and fully protects the probe is suggestive of polycistronic pyrG eno mRNA.

**DISCUSSION**

The nucleotide sequence of E. coli pyrG was determined in order to extend our analysis of the relationship of glutamine amidotransferase structure to function. Recent sequence analyses indicate that different amidotransferases contain one of two distinct glutamine amide transfer domains (8-12). In all amidotransferases, a glutamine amide transfer domain is combined with various arrangements with a domain that catalyzes an NH₃-dependent biosynthetic reaction. This combination endows glutamine amidotransferases with the capacity to catalyze a glutamine-dependent as well as an NH₃-dependent biosynthetic reaction, both in vitro and in vivo (13, 14). Both types of glutamine amide transfer domain utilize an active site cysteine to form a covalent glutaminy intermediates for catalysis of amide transfer (13, 14). Amidophosphoribosyltransferase (11) and glucosamine-6-P synthase (12) have a highly conserved amino acid transfer domain of approximately 190 amino acids that is characterized by an N-terminal active site cysteine. The second type of glutamine amide transfer domain, in GMP synthetase (30), carbamoyl-P synthetase (7, 10), anthranilate synthase (5), and p-aminobenzoate synthase (9), shown in Fig. 7, has three conserved segments. The active site cysteine in segment 2 is usually at a position approximately 85 to 90 amino acids from the N terminus of the domain.

The alignment in Fig. 7 localizes the CTP synthetase glutamine amide transfer domain and establishes its similarity to that in carbamoyl-P synthetase, GMP synthetase, anthranilate synthase, and p-aminobenzoate synthase. Using current nomenclature, the CTP synthetase glutamine amide transfer domain is trpG-related (8). In GMP synthetase, anthranilate synthase component II, and p-aminobenzoate synthase subunit II, the homologous trpG-related glutamine amide transfer domain (8) is initiated at the N-terminal residue of the protein chain. By noting that the first block of conserved sequence occurs 46-54 residues from the beginning of the domain in the three preceding enzymes, we estimate that the CTP synthetase glutamine amide transfer domain begins approximately at amino acids 292 to 300. Amino acid residues 1 to approximately 300 should contribute the structure needed for catalyzing the NH₃-dependent reaction. The glutamine amide transfer domain is fused onto the C-terminal end of the enzyme. Likewise, in carbamoyl-P synthetase, the glutamine amide transfer domain is fused onto the C-terminal end of a protein chain, except that the function of the N-terminal 185-amino-acid segment is unknown.

The conservation of amino acids in region 2 is sufficiently high to predict that the conserved cysteine, residue 379 in CTP synthetase, functions to form the covalent glutaminy intermediate as has been shown for Cys-84 in anthranilate synthase component II (14). Likewise, CTP synthetase His-315 in region 3 is implicated in the proton transfer that is required for ionization of Cys-379 (15). Whereas previous experiments have provided evidence for catalytic roles of cysteinyl and histidyl side chains in regions 2 and 3, respectively, there is no evidence bearing on the possible role of region 1 in glutamine amide transfer.

In a previous analysis of the pattern for fusion of the glutamine amide transfer domain to other protein chains, a model was proposed to explain the evolution of glutamine amidotransferases from primitive NH₃-dependent enzymes (8). According to this model, after duplication, genes encoding...
Fig. 3. Nucleotide sequence of *E. coli* pyrG, flanking regions, and translated amino acid sequence of CTP synthetase. The translated sequence of an N-terminal segment of enolase is downstream of CTP synthetase. Underlined sequences included: pyrG ribosome binding site, *em* ribosome binding site. Regions corresponding to the major and minor mRNA 5' ends are overlined and numbered (I) and (2), respectively.
FIG. 3—continued.
a glutamine amide transfer domain translocated adjacent to a glutamine amide transfer domain in several microorganisms, it was proposed that trpG-related gene fusions occur in carbamoyl-P aminobenzoate synthase subunit II (PABS) (9), carbamoyl-P synthetase (CPS), and CTP synthetase (CTPS). The numbers between dashes indicate the number of amino acids from the N terminus (NH2) between segments, and to the CO2 terminus (CO2H) or to the end of the domain (/). The numbering system at the top, as used previously (8), counts all positions, including gaps, from the start of the domain.

Fig. 7. Alignment of amino acids in three conserved segments of the glutamine amide transfer domain in E. coli GMP synthetase (GMPS) (8, 30), anthranilate synthase component II (AS II) (5), p-aminobenzoate synthase subunit II (PABS) (9), carbamoyl-P synthetase (CPS) (7), and CTP synthetase (CTPS). The numbering system at the top, as used previously (8), counts all positions, including gaps, from the start of the domain.

pyrG eno mRNA annealed to the probe and protected against nuclease S1 digestion. An alternative possibility, not presently excluded, is that two overlapping monocistronic pyrG and eno RNA molecules can anneal to the probe forming a tripartate mRNA having a calculated molecular weight of approximately 400 nucleotides should correspond either to a pyrG transcript having a 3' end at approximately nucleotide 2059 or an eno transcript having a 5' end at approximately nucleotide 2040. Likewise, the minor mRNA species of 420, 530, and 650 nucleotides either terminate distal to pyrG or initiate upstream of eno. Further experiments are required to determine whether multiple pyrG and eno mRNA molecules arise from transcription termination after pyrG and transcription initiation prior to eno or whether a primary pyrG eno mRNA undergoes processing. Since pyrG expression appears to be constitutive (3), there are no obvious regulatory barriers to a polycistronic pyrG eno transcriptional unit.

Acknowledgments—We thank James Fiesen for providing pyrG plasmid pNF1519 and E. coli strain JF649, Mark Hermodson for sequencer analyses, and Dan Ebbie for insightful discussions.

REFERENCES
1. Buchanan, J. M. (1973) Adv. Enzymol. Related Areas Mol. Biol. 39, 91-183
2. Koshland, D. E., Jr., and Levitzki, A. (1974) The Enzymes 10, 539-559
3. O'Donovan, G., and Neuhard, J. (1970) Bacteriol. Rev. 34, 278-343
4. Kawamura, M., Keim, P. S., Goto, Y., Zalkin, H., and Heinrikson, R. L. (1983) J. Biol. Chem. 258, 469-468
5. Nichols, B. P., Miozzari, G. F., van Cleemput, M. B., Bennett, G. N., and Yanofsky, C. (1980) J. Mol. Biol. 142, 505-517
6. Kaplan, J. B., and Nichols, B. P. (1983) J. Mol. Biol. 168, 451-468
7. Piette, J., Nyunoya, H., Lusty, C. J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N., and Piérard, A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4134-4138
8. Zalkin, H., Argos, P., Narayanan, S. V. L., Tiedeman, A. A., and Smith, J. M. (1985) J. Biol. Chem. 260, 3536-3540
9. Kaplan, J. B., Merkel, W. K., and Nichols, B. P. (1985) J. Mol. Biol. 183, 327-340
10. Werner, M., Feller, A., and Piérard, A. (1985) Eur. J. Biochem. 146, 371-381
11. Tsu, Y., Zalkin, H., van Cleemput, M., Yanofsky, C., and Smith, J. M. (1982) J. Biochem. 257, 3525-3531
12. Walker, J. E., Gay, N. J., Saraste, M., and Eberle, A. N. (1984) Biochem. J. 224, 799-815
13. Mänttäli, P., and Zalkin, H. (1984) J. Biol. Chem. 259, 14230-14236
14. Paluh, J. L., Zalkin, H., Betsch, D., and Weith, H. L. (1985) J. Biol. Chem. 260, 1889-1894
15. Amuro, N., Paluh, J. L., and Zalkin, H. (1985) J. Biol. Chem. 260, 14844-14849
16. Fiesen, J. D., An, G., and Fül, N. P. (1978) Cell 15, 1187-1197

GMPS  NH2—53 — I I L S G G P — 19 — V P V F G V C Y G M T A R M L G — 79 — G V Q F H P E — 23 — /
AS II  NH2—54 — L M L S P G P — 19 — L P I I G I C L G H Q A I V E A Y G — 70 — G F Q F H P E — 33 — /
PABS  NH2—66 — I V I S P G P — 19 — L P I L G V C L G H Q A M A Q A F G — 73 — G V Q F H P E — 17 — CO2H
CPS  NH2—236— I F L S N G P — 20 — I P V F G I C L G H Q L L A A S G — 68 — S F G H P E — 27 — CO2H
CTPS  NH2—345— A I L V P G G — 20 — I P Y L G I C L G M O V A L I B Y A — 120 — A C O F H P E — 28 — CO2H
**EXPERIMENTAL PROCEDURES**

Plasmids, Strains and Nucleic Acid - E. coli strain JM101 (recombinant genotype, pyrG pyrC purA recA) and plasmids pBR322 (14) were provided by James Miller, University of Toronto. Plasmids were isolated from transformants by the alkaline lysis method (17). This procedure was modified for the preparation of plasmid DNA by the phenol/chloroform method (18). Similar plasmid DNA was isolated from strain JM101, grown with uracil, by complementation of the chromosomal pyrG mutation.

**Engine Purification** - CTP synthetase was purified to electrophoretic homogeneity by the method of Anderson (19). Crude extracts of plasmid-bearing strain JM101 contained approximately 50-fold elevated levels of CTP synthetase (20), thus allowing the pheny1-Sepharose and hydroxylapatite chromatographic steps to be omitted. CTP synthetase activity was determined at 23°C by the method of Low and Pardee (20) using the assay mixture specified by Anderson (19). The enzyme was purified essentially according to the method of Anderson (19). The enzyme was dialyzed against 25 mmol/liter Tris/HCl, pH 7.5, 0.5 mol/liter KCl, 0.1 mol/liter mercaptoethanol, 2 mmol/liter NADH, 0.2 mmol/liter uridine triphosphate, in the presence of 1% ampho, and stored at -80°C. CTP synthetase predominates in the purified enzyme by amino acid analysis. An extinction coefficient, E 260 nm of 8.33 was determined. The specific activity of the purified enzyme was a purifying process, assayed at 23°C. A unit of activity corresponded to the production of 1 umol CTP/min at 23°C.

**Nucleotide Sequence of E. coli pyrG Encoding CTP Synthetase**

Supplementary Material to Nucleotide Sequence of E. coli pyrG Encoding CTP Synthetase

**Fig. 1.** E. coli plasmid DNA. Restriction maps are drawn approximately to scale. The symbols are: [P] EcoRI (3613 bp) cutting DNA; open box, E. coli (3613 bp) cutting DNA; solid line, pBR322; dashed line, plasmid DNA. Restriction sites are: B, BamHI; E, EcoRI; H, HindIII; K, KpnI; N, NotI; P, PstI; S, SalI.

29. Berk, A. J., and Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75**, 1274-1278

30. Tiedeman, A., Smith, J. M., and Zalkin, H. (1985) *Biochem. Biophys. Res. Commun.* **126**, 8676-8679

31. Schechtman, M. G., and Yanofsky, C. (1983) *J. Mol. Appl. Genet.* **2**, 83-99

32. Zalkin, H., Paluh, J. L., van Cleemput, M., Myoe, W. S., and Yanofsky, C. (1984) *J. Biochem. Biophys. Res. Commun.* **126**, 3985-3992

33. Nyunoya, H., and Lusty, C. J. (1984) *J. Biochem. Biophys. Res. Commun.* **159**, 9790-9798

34. Levitzki, A., Stallcup, W. B., and Koshland, D. E., Jr. (1971) *Biochemistry* **10**, 3371-3378

35. Holland, M. J., Holland, J. P., Thrill, G. P., and Jackson, K. A. (1981) *J. Biol. Chem.*** **256**, 1385-1395

36. Bachmann, B. J., and Low, K. B. (1980) *Microbiol. Rev.* **44**, 1-56

37. Lopata, M. A., Solnner-Webb, B., and Cleveland, D. W. (1985) *Mol. Cell. Biol.* **5**, 2842-2846

**Fig. 4.** Sequencing strategy for pyrG. The pyrG coding sequence is indicated as a solid box. Numbering is from a start site just upstream of the unique BglII site. Arrows indicate the extent of sequence obtained from each cloning fragment. Three sets of clones were used: A, fragments obtained using BglII, HaeIII, KpnI, B, fragments obtained from restriction sites shown on the map, C, obtained by exonuclease III digestion.
**Fig. 6.** Nuclease S1 mapping of the 3' end of pyrG mRNA. DNA probes were synthesized by primer extension. The 3'-terminal first-pass fragment was cloned into M13mp18 and M13mp19 to give coding and noncoding strand probes, respectively. Lane 1, MspI-digested pBR322 size standard; lane 2, RNA-protected fragments of coding strand probe; lane 3, undigested coding strand probe; lane 4, non-coding strand probe incubated with RNA and digested with nuclease S1; lane 5, undigested non-coding strand probe. Arrows point to major protected fragments, circles mark minor protected fragments.

**Fig. 5.** Nuclease S1 mapping of the 5' end of pyrG mRNA. (A) DNA probes synthesized by primer extension. The KpnI/BamHI fragment was cloned into M13mp18 and M13mp19. Primer extension using McfI yielded coding strand DNA complementary to mRNA; M13mp18 yielded noncoding strand DNA. Hybridization reactions contained approximately 0.02 pmol DNA probe and 79 μg RNA. Hybridization was at 54°C for 12 h. Lane 1, MspI-digested pBR322 size standard; lane 2, probe, RNA, 0.4 pmol per ml nuclease S1; lane 3, probe, RNA, 1.1 pmol per ml nuclease S1; lane 4, undigested coding strand probe; lane 5, non-coding strand probe incubated with RNA and digested with nuclease S1; lane 6, undigested non-coding strand probe. (B) KpnI/BamHI probe 5' end labeled. Hybridization reactions contained approximately 0.3 pmol DNA probe and 79 μg RNA. Lane 1, size standard; lane 2, probe, RNA, 700 units per ml nuclease S1; lane 3, probe, RNA, 1,400 units per ml nuclease S1; lane 4, undigested probe. Lanes 1 and 4 were exposed for 4 h at -20°C with an intensifying screen; lanes 2 and 3 for 32 h at -70°C with an intensifying screen. (C) Nuclease S1-digested fragments from (A) electrophoresed alongside a dideoxy sequencing ladder of the KpnI/BamHI fragment. The arrows mark the positions of two nuclease S1-digested fragments. The lower major band corresponds to an mRNA 5' end at approximately 263 in the nucleotide sequence. The upper minor band corresponds to an mRNA 5' end at approximately 222 in the nucleotide sequence.