Amino Acid Sequence of *Escherichia coli* Glutamine Synthetase Deduced from the DNA Nucleotide Sequence*

(Received for publication, March 14, 1986)

Giovanna Colombo and Joseph J. Villafranca

From the Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Glutamine synthetase is encoded by the glnA gene of *Escherichia coli* and catalyzes the formation of glutamine from ATP, glutamate, and ammonia. A 1404-base pair fragment from a cDNA containing the glnA structural gene for *E. coli* glutamine synthetase has been sequenced. An open reading frame of 1404 base pairs encodes a protein of 468 amino acid residues with a calculated molecular weight of 51,814. With few exceptions, the amino acid sequence deduced from the DNA sequence agreed very well with the amino acid sequences of several peptides reported previously. The secondary structure predicted for the protein has -36% of the residues in α-helices which is in agreement with calculations of ~39% based on optical rotatory dispersion data. Comparison of the amino acid sequences of glutamine synthetase from *E. coli* (468 amino acids) and *Anabaena* (473 amino acids) (Turner, N. E., Robinson, S. T., and Haselkorn, R. (1983) Nature 306, 337-342) indicates that 260 amino acids are identical and 80 are of the same type (polar or non-polar) when aligned for maximum homology. Several homologous regions of these two enzymes exist, including the sites of adenyllylation and oxidative modification, but the regulation of each enzyme is different.

In *Escherichia coli*, glutamine, a central intermediate in nitrogen metabolism, is synthesized from ATP, glutamate, and ammonia in a divergent manner dependent on reaction catalyzed by glutamine synthetase (EC 6.3.1.2). The enzyme is a dodecamer (double hexagon), and each subunit contains a catalytic site. In vivo, *E. coli* glutamine synthetase is regulated both at the biosynthetic level (Ginsburg and Stadtman, 1973) and at the level of enzymatic activity by a sophisticated covalent modification cascade mechanism (Stadtman and Ginsburg, 1974). The latter involves the covalent modification (adenyllylation) of a single tyrosyl residue/subunit. A unique peptide containing the covalently bound adenyl group has been isolated and sequenced (Heinrikson and Kingdon, 1974). The sites of adenyllylation and oxidative modification were located both at the biosynthetic level (Ginsburg and Stadtman, 1973) and at the level of enzymatic activity by a sophisticated covalent modification cascade mechanism (Stadtman and Ginsburg, 1974). The latter involves the covalent modification (adenyllylation) of a single tyrosyl residue/subunit. A unique peptide containing the covalently bound adenyl group has been isolated and sequenced (Heinrikson and Kingdon, 1974).

Classical chemical modification studies have shown that the integrity of lysine and arginine residues is required for enzyme activity. The ATP analog, 5′-p-fluorosulfonylbenzoylvalenasin, inactivates the enzyme (1 eq/subunit), and its point of irreversible attachment to *E. coli* glutamine synthetase has been located near the amino terminus at lysine 47 (Foster et al., 1981; Pinkofsky et al., 1984). An arginine residue at the ATP binding site (Colanduoni and Villafranca, 1985a) and 2 lysine residues, one at the glutamate binding site and the other at the ATP binding site (Colanduoni and Villafranca, 1985b), also seem to be involved in enzymatic activity. Glutamine synthetase is also inactivated by oxidation of a single histidine residue/subunit (Levine et al., 1978; Levine, 1983), and a pentapeptide containing this histidine residue has been sequenced (Levine, 1984). This oxidation makes the enzyme susceptible to proteolytic degradation.

Previous genetic studies indicate that the glnA gene of *E. coli* codes for glutamine synthetase, and a 1404-base pair fragment of *E. coli* DNA that contains glnA has been cloned in pBR322 (Backman et al., 1981; Chen et al., 1982). We report here the complete nucleotide sequence of the glnA structural gene. A single open reading frame of 1404 nucleotides encoding glutamine synthetase was found. The deduced amino acid sequence is compared with previously published amino acid sequences of several peptide fragments and the sequence of *Anabaena* glutamine synthetase (Turner et al., 1983). Our future goals are to identify amino acid residues involved in the activity and regulation of the enzyme by using site-directed mutagenesis.

**EXPERIMENTAL PROCEDURES**

**RESULTS AND DISCUSSION**

DNA Sequence and Derived Amino Acid Sequence—The plasmid DNA that harbors the glnA *E. coli* gene for glutamine synthetase is called pglnA. This plasmid was constructed by Backman et al. (1981) and Chen et al. (1982) by cloning a 2800-base pair ClaI digest of the *E. coli* genomic DNA into the ClaI site of pBR322.

Using standard techniques (Maniatis et al., 1982) we determined the orientation of the *E. coli* genomic DNA insert (Fig. 1). Position 4150 of the *E. coli* DNA is next to the unique EcoRI site of pBR322 and position 7010 lies next to the unique HindIII site of pBR322. Backman et al. (1981) had reported that glnA maps approximately between position 6000 and 4500 of the *E. coli* genome and that transcription of glnA is right to left (Fig. 1, top).

The nucleotide sequence of a 1922-nucleotide DNA fragment was determined (Fig. 2). This sequence represents approximately 69% of the *E. coli* DNA insert. The DNA se-

---

* Portions of this paper (including "Experimental Procedures" and Figs. 1 and 4) are presented in miniprint at the end of the paper. The abbreviation used is: NaDodSO4, sodium dodecyl sulfate. Miniprint is easily read with the aid of a standard magnifying glass. Full-size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 86M-0815, cite the authors, and include a check or money order for $1.20 per set of photocopies. Full-size photocopies are also included in the microfilm edition of the Journal that is available from Waverley Press.

**Vol. 261, No. 23, issue of August 15, pp. 10587-10591, 1986**

**Printed in U.S.A.**

**THE JOURNAL OF BIOLOGICAL CHEMISTRY**

© 1986 by The American Society of Biological Chemists, Inc.
sequence contains a single open reading frame of 1404 nucleotides, and the translation product contains 468 amino acid residues (Fig. 2). The sequence of about 65% of glnA was determined for both strands (see Fig. 1, bottom). Only one strand was sequenced for the amino-terminal and carboxyl-terminal regions. For these regions the nucleotide assignment was unambiguous, and the protein sequence was also known (see below). The 5' flanking region of the open reading frame contains an AGT initiation codon, a Shine-Dalgarno ribosome binding site (AGGAG), and at least one inverted repeat (Fig. 2, underlined). Also, the sequence TATcCTT' (upper case denotes matches with E. coli consensus sequence, Rosenberg and Court [1979]) and lower case denotes mismatches) can be found for the −10 region, and TTGCA for the −35 region (Fig. 2, underlined). The 1404-nucleotide open reading frame ends with two termination signals (TAAG, TAG). The open reading frame reported in Fig. 2 codes for a protein of 468 amino acid residues with a calculated molecular weight of 51,814. This value is in good agreement with the value of 49,000−50,000 estimated from sedimentation equilibrium experiments (Hunt and Ginsburg, 1972) and from disc gel electrophoresis (Ginsburg, 1972). The amino acid composition derived from the DNA sequence was compared to a previously reported amino acid analysis of glutamine synthetase (Orr et al., 1981). For most residues the values are in close agreement (Table I).
The sequences of several peptides of glutamine synthetase have previously been reported and were very useful in confirming the DNA derived sequence. 1) The sequence of the amino-terminal region (amino acids 1-26) was determined by automated Edman degradation by Orr et al. (1981), that of amino acid residues 9-48 was reported by Pinkofsky et al. (1984), and that of amino acid residues 49-58 by Heinrikson. Our deduced sequence is in perfect agreement with the peptide sequences even though the corresponding region of the nucleotide sequence (residues 1001-1161) was determined for only one strand. 2) A 38-residue peptide which corresponds to the protease-sensitive region of glutamine synthetase has been sequenced by Monroe et al. (1984). We found an almost exact match between the peptide sequence and our deduced amino acid sequence from amino acid 165-202 (Fig. 2). The only mismatch is at Cys-196 which was reported by Monroe et al. (1984) to be a Ser. Our DNA sequence through this region was determined from both strands (nucleotides 1253-1392 in Fig. 2) and 1529-1668 in Fig. 1). 3) The DNA derived sequence Met-His-Cys-His-Met (Fig. 2, amino acid residues 268-272) matches perfectly with the reported amino acid sequence of the pentapeptide which contains the His residue that is modified by oxidation of glutamine synthetase with ascorbate, Fe, or O$_3$H$_2$O. 4) Two peptides that are part of the carboxy-terminal region of the enzyme are also in excellent agreement with the DNA derived sequence. Heinrikson and Kingdon (1971) determined the sequence of a 21-amino acid tryptic peptide containing the covalently bound adenyl group. This peptide corresponds to the sequence between amino acids 386 and 405 (Fig. 2). Adenylylation of the enzyme occurs at tyrosine 397. Perfect matches at amino acids 286-292, 393-396, and 403-405 occur. While Heinrikson and Kingdon reported Lys-Asp at positions 393-394, we find Asp-Lys (Fig. 2). They also reported a Gly between Glu-402 and Glu-403 which we do not find. However, a revised amino acid sequence of this region of the protein as well as that containing Cys-196 (see above) shows our DNA derived sequence to be correct. Finally a perfect match exists for the carboxy-terminal 14 amino acids (Lahiri et al., 1972; R. Heinrikson, personal communication in Hanau et al., 1983). We have also found very good agreement between our deduced sequence with that of several cyanogen bromide peptides which were sequenced by automated Edman degradation. There were mismatches at eight positions, however. For five of the amino acids, our data resulted from sequencing both strands of DNA giving us confidence in these results. The other three amino acids deduced from analysis of one strand of DNA were Ser-83, Thr-84, and Lys-429. These were identified as Glx, Arg, and Glx by amino acid analysis. Several different overlapping DNA fragments from one strand were sequenced to arrive at our assignments. Because there are major differences in the codons for Ser versus Glx and Thr versus Arg, misassignment by us at positions 83 and 84 is unlikely. However, for position 429 (Lys versus Glx) some ambiguity may still exist.

Secondary Structure Prediction—The secondary structure of E. coli glutamine synthetase as predicted by the method of Chou and Fasman (1978) using a 4-residue window was evaluated. The prediction was that 36% of the residues should form $\alpha$-helices and 26% should be in the $\beta$ sheet configuration. The former value was consistent with an $\alpha$-helix content of $\sim$39% calculated from optical rotatory dispersion curves (Wulff et al., 1967; Hunt and Ginsburg, 1972).

Sequence Homology of Glutamine Synthetase from E. coli and Anabaena—In order to elucidate important functional domains, the protein sequence derived in this paper for E. coli glutamine synthetase is compared to the sequence from Anabaena published by Turner et al. (1983). Gaps have been introduced in both sequences to maximize alignment of homologous areas (Fig. 3). The comparison between these 2 glutamine synthetases is of interest because at the level of enzyme activity their regulation differs. In enteric bacteria (E. coli) enzyme activity is switched off by adenylylation of a
single tyrosyl residue/subunit. This covalent modification mechanism does not exist in cyanobacterium (Anabaena). The latter bacterium appears to regulate glutamine synthetase activity by feedback inhibition mediated by several amino acids (Orr and Haselkorn, 1981). The product of the cloned glnA gene of Anabaena is not adenylated in an E. coli strain that is capable of fully adenylating the product of the cloned E. coli glnA gene (Fisher et al., 1981). The Anabaena enzyme has 473 amino acids and the E. coli enzyme 468; 260 amino acids are identical in both enzymes (Fig. 3, boxed residues) and 80 residues are of the same type, i.e. non-polar, uncharged, or charged (Fig. 3, dashed-line boxed residues). In the E. coli enzyme the site of adenylation is Tyr-397 while in the Anabaena enzyme this residue corresponds to Tyr-401. In both enzymes a stretch of 8 identical amino acids (residues 387-394 in Anabaena and residues 382-390 in E. coli) precedes the tyrosine residues. The sequence around this tyrosine is Pro-Leu-Asp-Amp-Ile-Tyr-Glu-Leu-Glu for Anabaena glutamine synthetase and Ala-Met-Asp-Lys-Asn-Leu-Pro-Glu-Glu for the E. coli enzyme. Two differences are evident in this 13-amino acid stretch. First, following the tyrosine residue there is a Glu in the Anabaena enzyme while there is an Asp in the E. coli enzyme. There is not a Lys in the equivalent position of the enzyme while there is an Asp in the equivalent position of the Anabaena enzyme. Two differences are evident in this 13-amino acid stretch. First, following the tyrosine residue there is a Glu in the Anabaena enzyme while there is an Asp in the E. coli enzyme. This difference could alter the recognition site for adenyltransferase. Second, the presence and location of the proline residues could orient the polypeptide chain differently in each enzyme. In order to test this hypothesis we conducted a Chou-Fasman analysis of the carboxyl-terminus region of the Anabaena glutamine synthetase. In the E. coli enzyme the prediction is that Tyr-397 is preceded by a small stretch of α-helix and a turn. Following Tyr-397 a Pro-Pro stretch would determine a turn into a long stretch of α-helix. On the other hand, in the Anabaena glutamine synthetase Tyr-401 is predicted to be preceded by a random coil stretch and a turn and following Tyr-401 there would be a short α-helical stretch. It is possible that these differences result in kinks of the peptide chain such that Tyr-401 in Anabaena glutamine synthetase is buried while Tyr-397 in E. coli enzyme is exposed to the solvent and available for reaction with adenyltransferase.

A major difference in the amino-terminal region of E. coli and Anabaena glutamine synthetase is at Lys-47 of the E. coli enzyme. There is not a Lys in the equivalent position of Anabaena glutamine synthetase; instead there is a Val at position 48. Because both enzymes are clearly homologous, it is possible that Lys-47, which is at the ATP binding site (Foster et al., 1981) in the E. coli enzyme, does not play a catalytic role but is only involved indirectly in the binding of ATP. The prediction is that Anabaena glutamine synthetase would not be inactivated by 5'3'-fluorosulfonylbenzoyladenosine but should still bind ATP. Preliminary studies verify this prediction.

The oxidatively modified His-269 of E. coli glutamine synthetase has a counterpart in the Anabaena enzyme in His-272 (Fig. 3). Indeed the residues are Met-His-Cys-His-Met in E. coli (Levine, 1984), and Met-His-Cys-His-Gln in Anabaena. It is quite possible that Anabaena glutamine synthetase could be oxidized by the mixed function oxidases as is the E. coli enzyme.

Sequence Homology of Glutamine Synthetase from E. coli and Salmonella typhimurium—The nucleotide sequence corresponding to the amino terminus and 5' flanking region as well as the nucleotide sequence corresponding to the carboxyl terminus of the glnA gene from E. coli and S. typhimurium (Hanau et al., 1983) have a high degree of homology as shown in Fig. 4. After the initiation codon (ATG) a 39-base pair fragment codes for 13 identical amino acids. The DNA sequence in this region has only 3 differences that do not alter the amino acid residues. Towards the carboxyl terminus the enzymes are identical as is the encoding DNA. Only 3 differences in a 66-base pair long fragment occur at the DNA level which do not affect the amino acid sequence. The stop signal (TAA) is also identical in both DNAs. In the 5' flanking region, the Shine-Dalgarno sequence is AGGAG for E. coli and GGAG for Salmonella. In Salmonella there is a BgII recognition site in the upstream region which is not present in the E. coli genome. In the E. coli genome there is a BgIII (unique site in the 2500-base pair E. coli insert in pgln6, see Fig. 1) which is downstream from the ATG initiation codon. There is yet no information for this region of the DNA for Salmonella. There is no homology in the 3' flanking region of both genes.

In conclusion, we have used a recombinant plasmid containing glnA from E. coli to sequence the subunit of glutamine synthetase. The size and amino acid composition of the polypeptide encoded by glnA match those of the isolated enzyme. The N terminus and C terminus for the E. coli and Salmonella enzymes are identical and they both differ somewhat with the enzyme from the cyanobacterium Anabaena.

Acknowledgments—We are indebted to Dr. Linda S. Herndon for sharing her expertise in DNA sequencing and to Dr. C. Robert Matthews for use of equipment during the initial stages of this work. We are also grateful to Dr. R. Heinrikson for the unpublished amino acid sequence of peptide fragments of E. coli glutamine synthetase, to Dr. Allen Phillips for his assistance in the computer analysis of the sequence, and to Nels Pedersen for his assistance with the computer analysis of the secondary structure prediction.

REFERENCES


3 J. A. Colanduoni and J. J. Villafranca, unpublished results.
Restriction endonuclease map and strategy for determining the sequence for E. coli glnA. The box represents the 2800 bp insert of the E. coli chromosome containing glnA. The lines next to the box represent PBR322. The numbering is for the right (710) and left (4150) end of the E. coli DNA insert as well as for the BamHI site (4810) in that of Backman et al., 1981. The glnA coding region is indicated by the arrow at the top of the figure. Recognition sites for restriction endonucleases that occur not more than three times in the E. coli genomic insert are indicated. The arrows below the map denote sequencing direction and the length of the arrow is proportional to the number of nucleotides determined. All fragments were sequenced twice, and each time DNA from a different plasmid preparation was used.

Amino-terminus

S. typhimurium

C-terminus

E. coli

Homology of the amino-terminus and carboxy-terminus of glutamine synthetase from E. coli and Salmonella typhimurium. In the DNA sequence the nucleotides that are different or missing are in boxes. In the 5' flanking region of the DNA the nucleotides have been aligned to maximize the number of identical residues. The Shine-Dalgarno sequences, initiation (ATG) and termination codons (TAA) are underlined. APsiI recognition site is overlined.