Identification of a trpG-related Glutamine Amide Transfer Domain in Escherichia coli GMP Synthetase*

(Received for publication, July 20, 1984)

Howard Zalkin†, Patrick Argos‡, S. V. L. Narayana§, Amelia A. Tiedeman¶, and John M. Smith∥

From the Departments of †Biochemistry and §Biological Sciences, Purdue University, West Lafayette, Indiana 47907 and the ¶Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, Louisiana 71130

An improved method was developed to align related protein sequences and search for homology. A glutamine amide transfer domain was identified in an NH₂-terminal segment of GMP synthetase from Escherichia coli. Amino acid residues 1–198 in GMP synthetase are homologous with the glutamine amide transfer domain in trpG-D-encoded anthranilate synthase component II-anthranilate phosphoribosyltransferase and the related pabA-encoded p-aminobenzoyl synthase component II. This result supports a model for gene fusion in which a trpG-related glutamine amide transfer domain was recruited to augment the function of a primitive NH₂-dependent GMP synthetase. Sequence analyses emphasize that glutamine amide transfer domains are thus far found only at the NH₂-terminus of fused proteins. Two rules are formulated to explain trpG and trpG-related fusions. (i) trpG and trpG-related genes must have translocated immediately upstream of genes destined for fusion in order to position a glutamine amide transfer domain at the NH₂-terminus after fusion. (ii) trpG and trpG-related genes could not translocate adjacent to a regulatory region at the 5' end of an operon. These rules explain known trpG-like fusions and explain why trpG and pabA are not fused to trpE and pabB, respectively.

Alignment searches of GMP synthetase with two other enzymes that bind GMP, E. coli amidophosphoribosyltransferase and human hypoxanthine-guanine phosphoribosyltransferase, suggest a structurally homologous segment which may constitute a GMP binding site.

Glutamine amidotransferases are a family of enzymes that utilize the amide of glutamine in the biosynthesis of purine and pyrimidine nucleotides, amino acids, coenzymes, and an amino sugar (1, 2). These enzymes have a major role in utilization of assimilated nitrogen. Well characterized amidotransferases exhibit two distinct types of quaternary structure. Anthranilate synthase (EC 4.1.3.27) (3), PABA synthase (4), carbamoyl-phosphate synthetase (EC 6.3.5.5) (5), and glutamate synthetase (EC 1.4.1.13) (6, 7) have an αβ subunit composition while several glutamine amidotransferases including GMP synthetase (EC 6.3.5.2) (8–10), CTP synthetase (EC 6.3.4.2) (11), and amidophosphoribosyltransferase (EC 2.4.2.14) (12) are oligomers of identical subunits. Studies on anthranilate synthase (3) and carbamoyl-phosphate synthetase (5) have established a specific structure-function relationship for αβ-type glutamine amidotransferases. In anthranilate synthase one subunit, designated AS I, catalyzes an NH₂-dependent synthesis of anthranilate: chorismate + NH₂ → anthranilate + pyruvate. A second subunit, AS II, binds glutamine and provides glutamine amide transfer function to the AS I-AS II proteomer: chorismate + glutamine → anthranilate + pyruvate + glutamate. In some enteric bacteria (3, 13) and eukaryotic microorganisms (14, 15) AS II is covalently joined to other enzymes of tryptophan biosynthesis to yield multifunctional enzymes. Sequence analyses of several AS II enzymes indicate that multifunctional and multifunctional AS II proteins are homologous and define a structure required for glutamine amide transfer. A homologous glutamine amide transfer subunit, PABS II encoded by Escherichia coli pabA, functions in synthesis of p-aminobenzoyl glutamate (16).

In this paper we compare the recently completed amino acid sequence of GMP synthetase derived from E. coli gud2 with E. coli AS and PABS II sequences. This analysis indicates that an amino-terminal segment of GMP synthetase of approximately 200 residues contains a glutamine amide transfer domain that is homologous with AS II and PABS II. A second segment was identified in GMP synthetase which exhibits primary structural homology with regions in other enzymes having a GMP binding site.

MATERIALS AND METHODS

Computer Alignments—The protein sequences were correlated by comparing every possible span of length L residues in one protein with every such span in the second protein. Two scoring procedures were used. The first was based on the Dayhoff relatedness odds matrix (17–19) whose elements express relative weights with which amino acids have been allowed to substitute in aligned sequences of 71 protein families.

The second scoring procedure involved calculation of the mean correlation coefficient for each oligopeptide comparison for six physical characteristics which are thought to be primary forces directing protein folding (21): helix, sheet, and turn secondary structural conformational preferences; residue polarity; and two amino acid hydrophobicity measures, the hydration potential and surrounding hydrophobicity. The use of physical parameters has been previously discussed (20–23). The final search matrix was constructed by averaging the scores from the two procedures after subtracting the mean value from all elements of each search matrix and then scaling the respective elements such that the standard deviations of each matrix were made equal. The average score was placed in the final search matrix.

The abbreviations used were: PABA, p-aminobenzoic acid; AS I, anthranilate synthase component I; AS II, anthranilate synthase component II; PABS I, p-aminobenzoyl synthase component I; MBC/C, minimum base change per codon.

* This work was supported by United States Public Health Service Grants GM 24658, GM 27100, and AI 20068. This is Journal Paper 9968 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.

† The abbreviations used are: PABA, p-aminobenzoic acid; AS I, anthranilate synthase component I; AS II, anthranilate synthase component II; PABS I, p-aminobenzoyl synthase component I; MBC/C, minimum base change per codon.

‡ A. A. Tiedeman and J. M. Smith, unpublished information.
numbered according to the initial amino acid of each oligopeptide segment compared. The resulting matrix for proteins compared here had a lower noise level and indicated longer stretches for alignment than matrices calculated from either technique alone.

Search-matrix plots were made by attaching symbols to the matrix element values that fell within particular fractional standard deviation ranges. No matrix value was considered that was not 3.5 or greater; a value that should have a probability of $10^{-5}$ or less (17). At 3.5, the theoretical probability of such a matrix value occurring is less than $3 \times 10^{-5}$. A series of diagonally collinear, broken lines were easily detected by visual inspection through the use of the same symbol in all $L$ positions corresponding to the search segment length. If there was overlap for symbols in different standard deviation ranges, the symbol corresponding to the higher range was allowed to dominate. A window search length $L$ of 20-30 residues resulted in minimal noise.

Once the sequences had been matched, an assessment of the overall structural relatedness of the two proteins was calculated using mean correlation coefficients for all the aligned residues over the six physicochemical characteristics. A coefficient value equal to -1.0 would show an opposite trend in the aligned residue characteristics while 0.0 would point to a random relationship and 1.0, a perfect correlation (24).

Secondary Structure Prediction—Plots of the aligned sequence number (i.e. all positions counted irrespective of deletions or insertions) versus the conformational preference parameter (helix, $\beta$-strand, and turn (25)) for a given amino acid were first calculated for each protein sequence. Each of the structural potential curves was then smoothed such that every successive group of three points ($i$ to $i + 2$) was fitted by a least squares line and the value at ($i + 1$) replaced by that calculated from the line. This smoothing process was repeated successively for three cycles over each of the parametric plots. Finally, the smoothed curves for a given conformational potential (helix, sheet, and turn) were averaged over the aligned sequences which should yield a better prediction than that from any one sequence (26). Residues were assigned to one of the three secondary structural states if their mean potential was greater than 1.0, a neutral preference for an amino acid to be in a given state (27, 28). The structural type assigned was associated with the largest potential. Five successive values greater than 1.0 were required for helix initiation while three were used for strands or turns. If a Pro or Gly residue occurred within a helically predicted region past the fourth helical NH$_2$-terminal residue, the span was assigned instead as a turn segment. Pro and Gly often occur at the beginning of a helix but are rarely found in its interior or COOH-terminal region (25).  

RESULTS

Gene-enzyme relationships, subunit compositions, and reactions catalyzed by the glutamine amidotransferases from E. coli, discussed in this report are shown schematically in Fig. 1. The four enzymes summarized in Fig. 1 are PABA synthase, anthranilate synthase, GMP synthetase, and amidophosphoribosyltransferase. Although Fig. 1 shows only the reactions with glutamine, each enzyme also catalyzes an NH$_3$-dependent reaction in which NH$_3$ replaces glutamine. The black dots in PABS II, AS II, GMP synthetase, and amidophosphoribosyltransferase represent the active site cysteine that is required for the glutamine-dependent activity of each enzyme.

Glutamine Amide Transfer Domains—PABS II and AS II contribute glutamine amide transfer functions to PABA synthase and anthranilate synthase, respectively (Fig. 1). The PABS II and AS II sequences (13, 16, 29) define a glutamine amide transfer domain. Previous studies have established that anthranilate synthase and GMP synthetase exhibit similar mechanisms for glutamine amide transfer domain (3, 9). We have compared the entire GMP synthetase sequence of 525 amino acids to the 1857 amino acids of PABS II and to the 691 residues of bifunctional AS II-anthranilate phosphoribosyltransferase by the scoring procedure. About the first 200 NH$_2$-terminal residues of GMP synthetase showed excellent correlation in the search matrix with the AS II and PABS II sequences. The search matrix, showing homology between the NH$_2$-terminal segment of GMP synthetase and the AS II domain of AS II-anthranilate phosphoribosyltransferase, is presented in Fig. 2. The remaining domains of GMP synthetase and AS II-anthranilate phosphoribosyltransferase showed no high scores and no pattern of diagonally collinear lines over extended sequence regions. Even higher correlations were obtained for comparisons of AS II with PABS II and for PABS II with GMP synthetase. These results confirm the homology of AS II and PABS II sequences (16) and

![Glutamine Amidotransferase Gene-enzyme Relationships](image-url)
The mean correlation coefficients over six physical parameters for aligned residues

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Mean</th>
<th>No. of residues</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS II-PABS II</td>
<td>0.63</td>
<td>185</td>
<td>GAT</td>
</tr>
<tr>
<td>GMPS-PABS II</td>
<td>0.40</td>
<td>198</td>
<td>GMP</td>
</tr>
<tr>
<td>GMPS-AS II</td>
<td>0.33</td>
<td>189</td>
<td>GAT</td>
</tr>
<tr>
<td>APRT-HGPRT</td>
<td>0.44</td>
<td>105</td>
<td>GMP</td>
</tr>
<tr>
<td>GMPS-APRT</td>
<td>0.38</td>
<td>150</td>
<td>GMP</td>
</tr>
<tr>
<td>GMPS-HGPRT</td>
<td>0.28</td>
<td>95</td>
<td>GMP</td>
</tr>
<tr>
<td>ADH-GPD</td>
<td>0.12</td>
<td>118</td>
<td>NAD*</td>
</tr>
<tr>
<td>ADH-LDH</td>
<td>0.26</td>
<td>116</td>
<td>NAD</td>
</tr>
<tr>
<td>GPD-LDH</td>
<td>0.19</td>
<td>123</td>
<td>NAD</td>
</tr>
</tbody>
</table>

*Alignments of NAD binding domains are shown by Otto et al. (33).

The mean MBC/C was determined for all possible aligned five-residue spans and for all possible pairwise comparisons over the three glutamine amide transfer domains. A match involving a deleted residue was assigned a MBC/C value of 3. The mean MBC/C values were averaged over the three possible pairs for all triply aligned five-residue segments. The two lowest averages were 0.0 for aligned positions 183–187 and 0.4 for the span 88–92 (Fig. 3). The latter segment contains a conserved cysteine in all three proteins. This cysteine has been identified as an active site residue (30, 31) and is conserved in eight homologous AS II and PABS II sequences (15–16, 32). The sequences at position 183–187 encompass the same five residues in all three proteins and interestingly contain a histidine residue, a frequent constituent of active sites.

GMP Binding Domains—Alignment searches were also performed on GMP synthetase with amidophosphoribosyltransferase and human hypoxanthine-guanine phosphoribosyltransferase since the latter proteins also interact with GMP. An amidophosphoribosyltransferase sequence, approximately 120 residues in length, showed good match scores with segments in GMP synthetase and hypoxanthine-guanine phosphoribosyltransferase. Fig. 4 displays the search matrix between amidophosphoribosyltransferase and GMP synthetase where the two highest search values relate the COOH-terminal ends of both molecules. The resulting alignments of the segments in amidophosphoribosyltransferase, hypoxanthine-guanine phosphoribosyltransferase, and GMP synthetase are shown in Fig. 5.

Examination of the catalytic and allosteric functions of the three molecules shows GMP binding as the only common feature, thereby suggesting that the aligned regions share a GMP binding function. A search for the mean MBC/C over all possible residue spans between the three enzyme pairs showed two regions with the lowest mean; namely, segments 15–20 at 0.53 and 95–100 at 0.67. The residues in these spans are underlined in Fig. 5. These regions have conserved Lys, Asn, and Asp residues, all of which may be important for GMP binding or perhaps conserved for merely structural reasons.

Table I lists the mean correlation coefficient over six physical parameters for aligned residues in each pairwise comparison of the putative GMP binding domains. The correlation...
coefficients relate amidophosphoribosyltransferase and hypoxanthine-guanine phosphoribosyltransferase most closely (0.44) while GMP synthetase and hypoxanthine-guanine phosphoribosyltransferase are least related (0.28). It is noteworthy that similar calculations for the aligned sequences of domains in amidophosphoribosyltransferase chosen to indicate the standard deviation (e) fractions of the search were used to align the sequences.

The search window chosen was 25 amino acids. Symbols chosen to indicate the two highest values in the map are marked by arrows and are used to align the sequences.

There are important implications that can be derived from the structure of the trpG-related glutamine amide transfer domain in GMP synthetase. In particular, the domain is located at the NH3 terminus of GMP synthetase (Fig. 1). All known trpG fusions in bacteria (13), yeast (15), and fungi (14) position AS II at the NH2 terminus of a multifunctional enzyme. However, in all known trpG fusions, AS II is joined to an enzyme of tryptophan biosynthesis other than AS I. Why is the AS II glutamine amide transfer domain not fused to the subunit catalyzing the NH3-dependent reaction (AS I)? Clearly, the expected fusion of a glutamine amide transfer domain to a protein having NH3-dependent activity functions well in the case of GMP synthetase. Two rules may have governed the fusion of trpG and trpG-related glutamine amide transfer domains into enzymes. We suggest that (i) the glutamine amide transfer domain has to be at the NH2 terminus of a fused protein (Rule 1) and (ii) a newly evolved trpG function could not be translocated to a site upstream of trpE, proximal to the regulatory region (Rule 2). According to this proposal, the bacterial fusion of trpG-encoded glutamine amide transfer function to trpD in some enteric bacteria (13) reflects the requirements of translocation of a newly evolved and duplicated gene into an existing trp operon. In the case of the guaA operon, translocation of a trpG-related gene downstream of guaB could have positioned a glutamine amide transfer function at the NH2 terminus of GMP synthetase (guaA). This translocation is far removed from the upstream guaB regulatory region. Following translocation, gene fusion presumably enhanced the function of GMP synthetase. This proposal also explains why the glutamine amide transfer domain of PABA synthase is not fused to the subunit catalyzing the NH3-dependent reaction. pabA could not be translocated upstream of pabB, a position proximal to the regulatory region of pabB. We further propose that since trpG could not be translocated upstream of trpE, a trpE-trpG fusion was not possible, there was no strong selective pressure for other trpG fusions. Thus, in E. coli, trpG and the contiguous trpD were fused but in Serratia marcescens trpG and the contiguous trpD did not fuse. Although GMP synthetase, anthranilate synthase, and PABA synthase have homologous glutamine amide transfer domains, a similar structure was not identified in amidophosphoribosyltransferase, not withstanding that these enzymes

**Discussion**

Our analyses have identified functional domains for glutamine amide transfer and GMP binding that were likely incorporated into GMP synthetase by gene fusion. Nagano et al. (30) suggested that anthranilate synthase and other glutamine amidotransferases evolved by recruitment of glutamine amide transfer function to existing NH3-dependent enzymes. Recent experiments employing site-directed mutagenesis have verified that the NH3-dependent activities of anthranilate synthase (34) and amidophosphoribosyltransferase (35) can function in E. coli for tryptophan and purine nucleotide synthesis, respectively, when the NH3 concentration in the media is sufficiently high. The alignment in Fig. 3 indicates that the glutamine amide transfer domain in guaA-encoded GMP synthetase is homologous to corresponding sequences in anthranilate synthase (trpG) and PABA synthase (pabA). Thus, duplication and translocation of an ancestral gene provided trpG and pabA to augment the NH3-dependent activities encoded by trpE and pabB, respectively, and also provided glutamine amide transfer function to guaA. The capacity to utilize the amide of glutamine for biosynthesis instead of NH3 removes the dependency of cells upon environments high in NH3 (34, 35).

**Fig. 4.** Representation of the structural homology search matrix for GMP synthetase and amidophosphoribosyltransferase. The search window chosen was 25 amino acids. Symbols chosen to indicate the standard deviation (σ) of the search scores are: 4.0 σ ≤ S < 4.5 σ ( ); 4.5 σ ≤ S < 5.0 σ ( ); 5.0 σ ≤ S < 5.5 σ ( ). The two highest values in the map are marked by arrows and were used to align the sequences.
have certain common catalytic properties. They have distinct sites for glutamine and NH₄-dependent activities, utilize an active site cysteine for glutamine amide transfer, and exhibit glutaminase activity (12). Either a homologous glutamine amide transfer domain has diverged beyond recognition in amidophosphoribosyltransferase or some glutamine amidotransferases may have acquired glutamine amide transfer function from other ancestral sources. Based on Rule 2, above, we would not expect a trpG-related glutamine amide transfer domain in amidophosphoribosyltransferase because an NH₄-terminal position would have required translocation adjacent to the purF control region.

In all known AS II sequences (13–15, 32) and a PABS II sequence (16) an active site cysteine involved in formation of glutaminyl and glutamyl enzyme intermediates is conserved. This cysteine is also conserved in GMP synthetase at position 90 and thus identifies the active site cysteine in this enzyme. Interestingly, the cysteine at position 167 in GMP synthetase (Fig. 5) exhibited the greatest reactivity with iodoacetamide and was previously thought to be involved in glutamine amide transfer (36).

A basic residue at position 114 is conserved in all AS II and PABS II sequences and was suggested to participate along with the active site cysteine in glutamine amide transfer (37). However, position 114 is not conserved in GMP synthetase and therefore a possible catalytic role of Lys at position 114 in AS II is suspect.

GMP synthetase residues 200–525 should contain domains that function in NH₄-dependent synthesis of GMP. In vitro the NH₄-dependent activity of GMP synthetase was approximately 50–60% of the activity with glutamine (9). As noted in Fig. 5, GMP synthetase contains a region (residues 239–348) which is structurally conserved in amidophosphoribosyltransferase and human hypoxanthine-guanine phosphoribosyltransferase. Thus, residues 239–348 in GMP synthetase may provide a site for binding the substrate XMP which upon amination yields GMP. Since structural homology is maintained for an additional 46 residues in GMP synthetase and amidophosphoribosyltransferase, positions 120–175 (Fig. 5) may contribute to a common function. These two enzymes each have binding sites for NH₄ and an adenine nucleotide. Chemical modification or site-directed mutagenesis is required to provide direct evidence for predicted functions.

REFERENCES