The complete nucleotide sequence of the *glnALG* operon of *Escherichia coli* K12

Juan Miranda-Ríos[1], Ray Sánchez-Pescador[2], Mickey Urdea[2] and Alejandra A.Covarrubias[1]*

[1]Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México, Apdo. Postal 565-A, Cuernavaca, Morelos, México and [2]CHIRON Corporation, 4560 Horton St., Emeryville, CA 94608, USA

Received September 11, 1986; Revised January 16, 1987; Accepted January 21, 1987

ABSTRACT

The nucleotide sequence of the E. coli glnALG operon has been determined. The glnL (ntrB) and glnG (ntrC) genes present a high homology, at the nucleotide and aminoacid levels, with the corresponding genes of Klebsiella pneumoniae (1, 2). The predicted aminoacid sequence for glutamine synthetase allowed us to locate some of the enzyme domains. The structure of this operon is discussed.

INTRODUCTION

Escherichia coli and other enteric bacteria are able to respond to nitrogen limitation by means of a complex global control system. The genes and operons subject to nitrogen control encode proteins that permit the utilization of various alternative nitrogen sources when ammonia is lacking in the growth medium (reviewed in refs. 3-5). Among the nitrogen-regulated (Ntr) operons is the complex glnALG operon, which contains glnA, the structural gene for glutamine synthetase, the primary ammonia assimilating enzyme, as well as two nitrogen regulation genes, glnL (ntrB) and glnG (ntrC) (6-11). Other nitrogen-regulated genetic systems include aminoacid transport components, degradative enzymes and nitrogenase with its associated factors (3,5).

The activation of transcription of Ntr (nitrogen regulated) genes during growth in a nitrogen-deficient medium requires the products of glnF (ntrA or rpoN), an RNA polymerase sigma subunit (σ^{60}) and of glnG (ntrC) (12, 13).

The expression of the glnALG operon is regulated by the glnG product (NR_I) at three promoters: two located upstream of glnA (glnAp1 and glnAp2) and a third one localized upstream of glnL (glnLp) (9,14-19). Initiation of transcription at glnAp1 is stimulated by the catabolite activator protein (CAP) charged with cyclic AMP and it is repressed by NR_I (18, unpublished results). Transcription from glnAp1 and glnLp occurs when cells are growing

under conditions of carbon deficiency and nitrogen excess, apparently, to maintain basal intracellular concentrations of glutamine synthetase (GS), of NR_I and of the glnL product (NR_{II}) (13,14,16,18, unpublished results).

Transcription initiated at glnLp is also subject to repression by NR_I (13,20,21). Nitrogen deprivation conditions result in high intracellular levels of GS; in this case transcription starts from glnAp2, which requires σ^{60} and NR_I (12,13,22).

It has been proposed that nitrogen excess converts NR_I to a form unable of activating transcription. The interconversion of the active and inactive NR_I forms is greatly stimulated by NR_{II}, which in turn receives an accurate assesment of the availability of cellular nitrogen from the products of glnB (P_{II}) and glnD (UTase) (8,18,20,23). Recently, it has been shown that the conversion of NR_I to a form capable of activating the initiation of transcription at glnAp2 involves the NR_{II}-catalyzed phosphorylation of NR_I (37). Furthermore, it has been suggested that NR_{II} can be converted by P_{II} to an NR_I-phosphate phosphatase (37). These two NR_{II} activities suggest the functioning of a cyclic cascade system responsible for the regulation of the GS synthesis in response to the nitrogen availability.

In this paper we report the complete nucleotide sequence of the E. coli glnALG operon (4320 bp). The nucleotide sequence of DNA fragments containing the glnA and glnL regulatory regions have already been reported (16,25,26). In this work we determined the sequence of 3488 bp needed to complete the entire operon. Analysis of our data show that the glnL and glnG products (NR_{II} and NR_I, respectively) from E. coli are highly homologous to those from K. pneumoniae (1,2). We also report the amino acid sequence of the enzyme glutamine synthetase as deduced from the nucleotide sequence obtained.

METHODS AND MATERIALS

Cloning and DNA sequence

DNA's carrying glnA, glnL or glnG sequences were derived mostly from plasmid pACR34 (25). Part of the glnL gene sequence was derived from plasmid pACR3 (25).

Restriction and DNA-modifying enzymes were obtained from commercial sources and used according to the manufacturers' instructions or as described by Maniatis et. al. (27).

Dideoxy sequencing reactions were carried out as described (28,29) using clones prepared in M13mp18 and M13mp19 vectors (30). ^{35}S-dATP or ^{32}P-dCTP were used as the labelled nucleotides.

The DNA sequencing strategy to complete the nucleotide sequence of the glnALG operon was as follow: the 276 bp EcoRI-EcoRI and the 698 bp EcoRI-BamHI fragments which are included in the glnA gene were cloned into M13 vectors to be sequenced in both orientations. The results obtained were confirmed by sequencing both chains of fragments EcoRI-BglII (317 bp), BglII-BamHI (376 bp) and the junction at the EcoRI sites. M13 derivatives containing the 477 bp ClaI-SalI fragment and the 2260 bp SalI-PstI fragment were used to obtain the glnL and glnG nucleotide sequences. The ClaI-SalI fragment was sequenced in both orientations and the information obtained was confirmed by sequencing HpaII subclones from this DNA fragment (the HpaII sites used are located at positions 298 and 528 in Fig. 3). Most of the nucleotide sequence of the SalI-PstI fragment (2260 bp) was obtained using synthetic oligonucleotides, ^{32}P-labelled at their 5'-end, as primers for the sequencing reaction. The primers used allowed us to sequence both chains of this SalI-PstI fragment. M13 clones containing the ClaI and SalI sites at positions 67 (Fig. 3) and 544 (Fig. 3) were used to confirm the sequence at the corresponding junctions.

RESULTS

Sequence of glnA

The physical map of the glnALG operon previously obtained in this laboratory, was used as a basis for the sequencing strategy which is shown in Fig. 1.

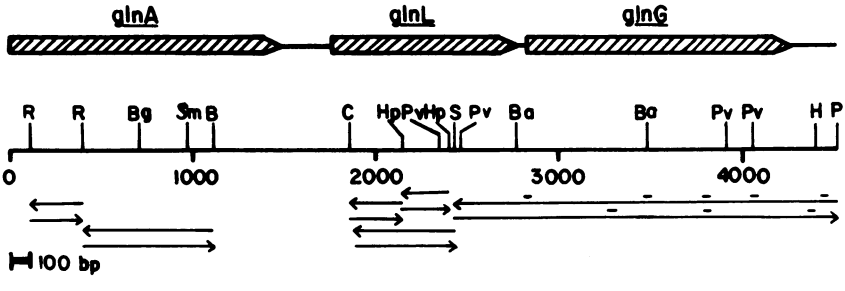

Fig. 1. The physical and genetic map of the E. coli complex glnALG operon. The precise location of some restriction sites is indicated. The hatched arrows indicate the localization, extension and direction of transcription of the genes forming this operon. The M13 clones used for sequencing are shown below the restriction map. The arrowheads indicate the sequencing direction. The small lines above the SalI-PstI M13 clones indicate the localization of the synthetic oligonucleotides used as primers in the sequencing reaction. R=EcoRI, B=BamHI, Ba=BalI, Bg=BglII, C=ClaI, H=HincII, S=SmaI, P=PstI, Pv=PvuII, Hp=HpaII.

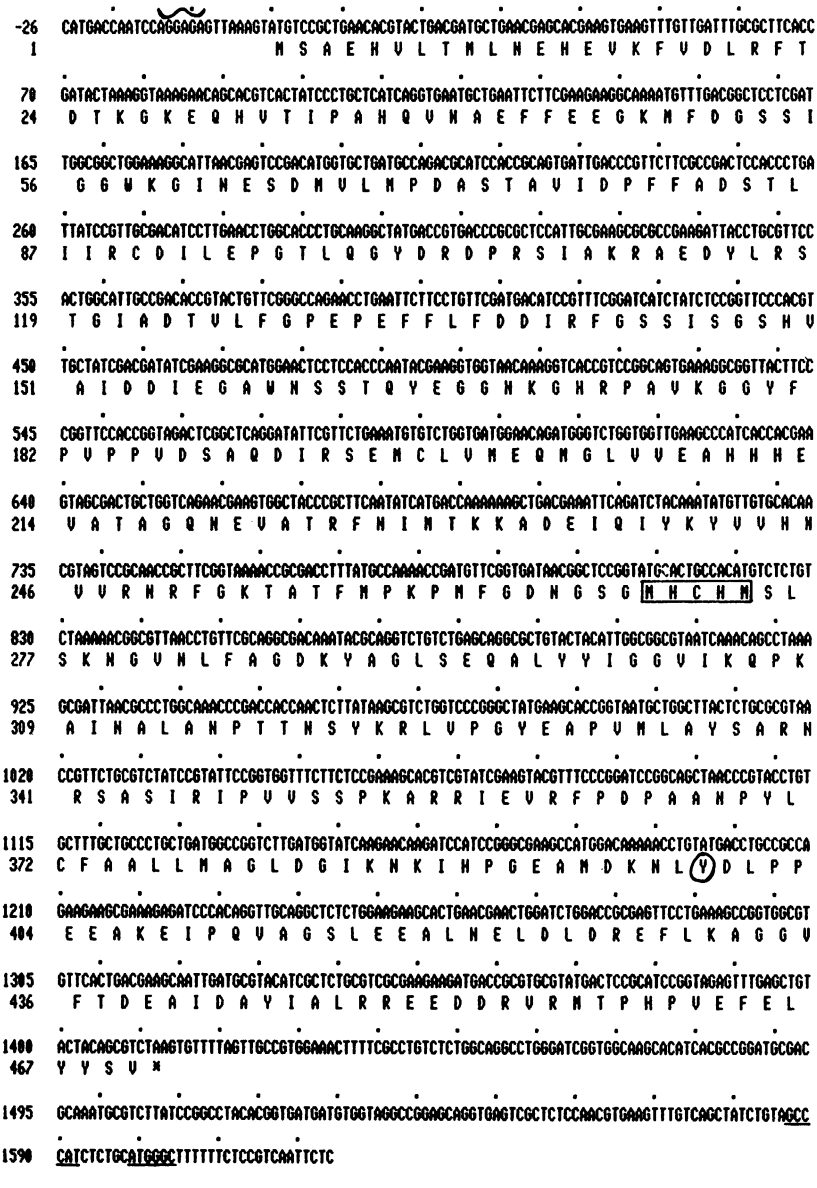

Fig. 2. The nucleotide sequence of the E. coli glnA gene. The initiation codon is in agreement with the reported GS N-terminal aminoacid sequence (31). The wavy line indicates the possible SD sequence. The circled tyrosyl residue is the one to which AMP is covalently bound (31,32). The squared aminoacid sequence corresponds to the peptide containing the oxidizable histidine of GS (33). The underlined sequence downstream the stop codon corresponds to a potential transcription termination signal.

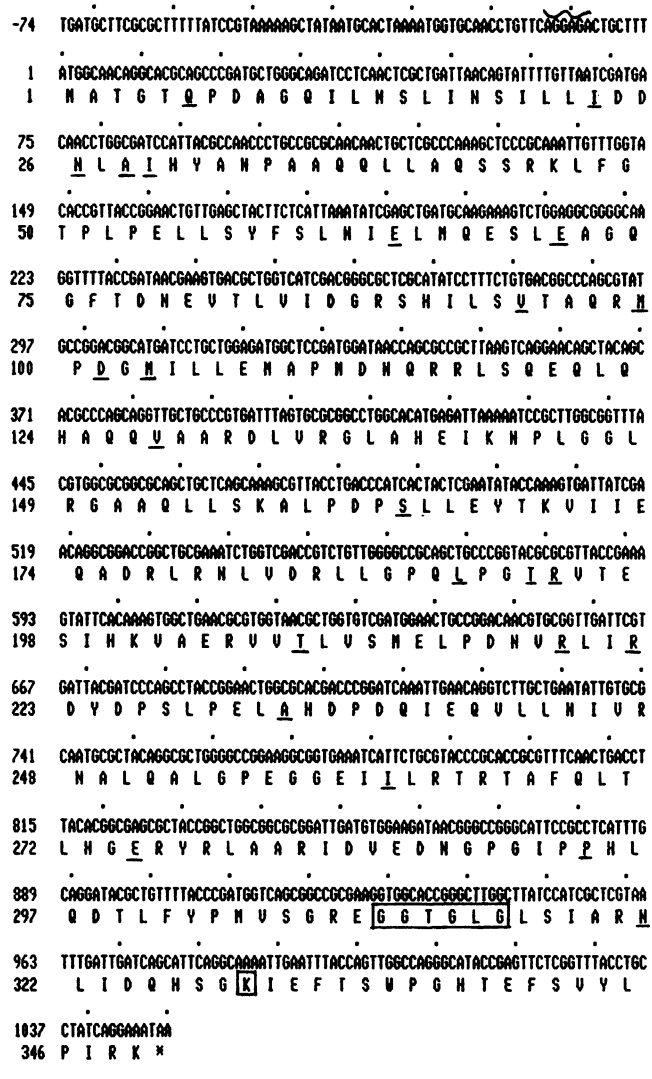

Fig. 3. The nucleotide sequence of the E. coli glnL gene. The wavy line indicates the possible SD sequence. The underlined aminoacids corresponds to those which are not homologous to the K. pneumoniae sequence. Numbers at the far left of the figure correspond to the positions of the nucleotides (up) and aminoacid residues (down). The squared aminoacid residues correspond to the glycine-rich region and the lysine highly conserved in all kinases described until now. The nucleotide sequence in this figure is contiguous to that shown in Fig. 2.

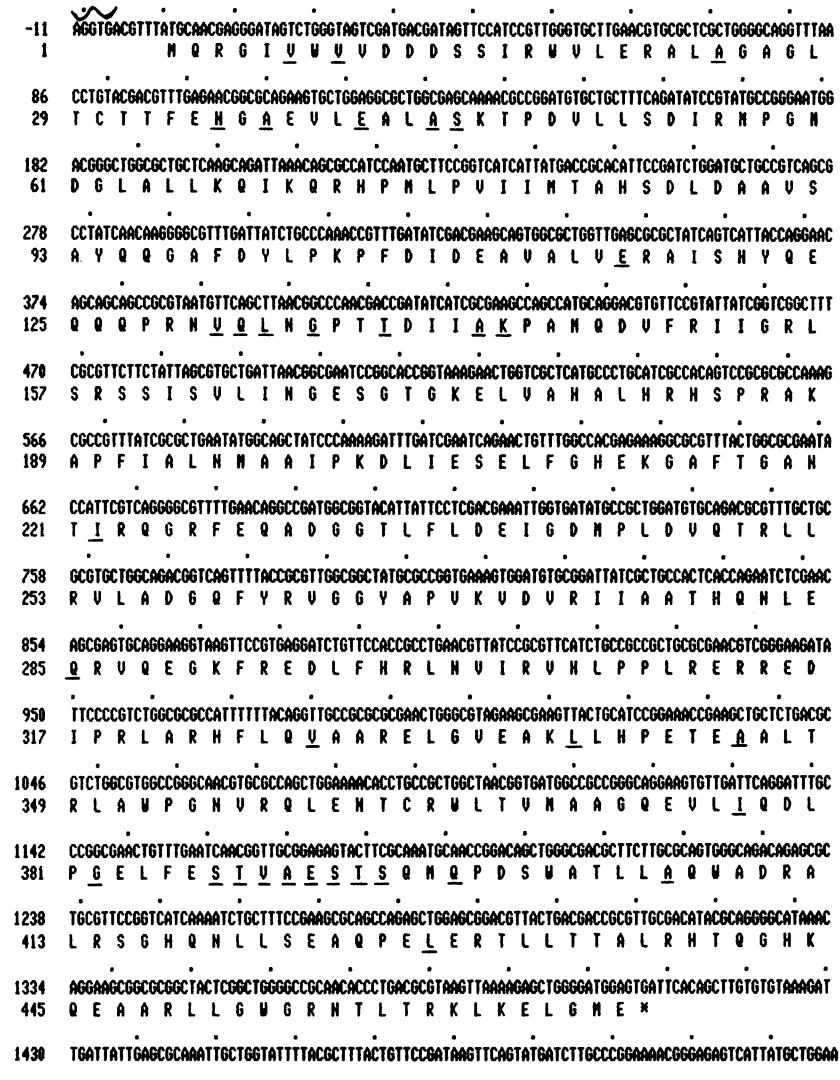

Fig. 4. The nucleotide sequence of the E. coli glnG gene. The signs are as indicated in Fig. 3. The nucleotide sequence in this Figure is contiguous to that shown in Fig. 3.

The nucleotide sequence containing the glnA and glnL control regions have already been reported. These sequences include the DNA regions from -26 bp to 130 bp and from 1,096 bp to 1,631 bp at the 5'-end and 3'-end of glnA, respectively (25,26) (Fig. 2). In this work we added 966 bp to complete the

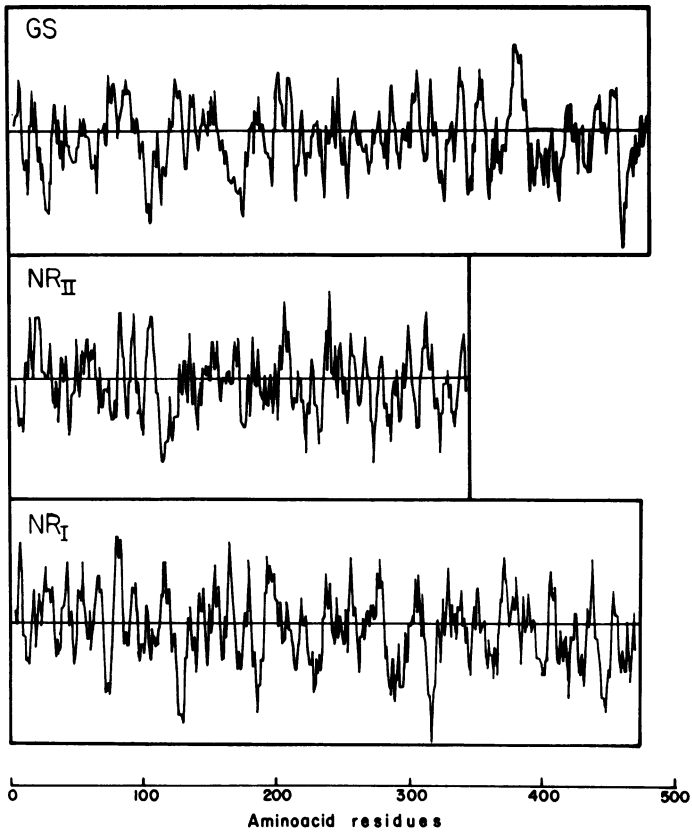

Fig. 5. Hydropathy profiles of the deduced aminoacid sequences of the E. coli GS, NR_{II} and NR_{I}.

sequence of the glnA gene. The amino acids sequence of the E. coli GS was deduced. According to this data the E. coli GS monomer is a protein of 470 amino acids with a predicted molecular weight of 52,222 daltons (Fig. 2).

Sequence of glnL and glnG.

The sequence downstream glnA revealed two extensive open-reading frames. The start codon of the first one is located 284 bp downstream the stop codon for glnA and it extends 1,050 bp. It encodes a protein of 349 amino acids with a predicted molecular weight of 38,647 daltons as corresponds to the glnL product or NR_{II} (8) (Fig.3).

The second open-reading frame start is 11 bp downstream of glnL stop codon and extends 1,408 bp. This encodes for a protein of 468 amino acids

with a molecular weight of 52,205 (Fig. 4). The predicted molecular weight for this protein corresponds to that determined for the glnG product or NR_I (8, 20).

The 11 bp region that separates the glnL and glnG genes does not show any obvious transcription termination signals, which confirms that these two genes are transcribed as a bicistronic message.

DISCUSSION

The DNA sequence obtained for the E. coli glnALG operon indicates that it occupies a region of aproximately 4,320 bp, from which 1,414 bp, 1,050 bp and 1,407 bp correspond to glnA, glnL and glnG coding regions, respectively. The deduced molecular weights of 52,222 for GS, 38,647 for NR_{II} and 52,205 for NR_I are in close agreement with the estimate of 55,000, 36,000 and 50,000, respectively, derived from the analysis of the polypeptides in SDS-PAGE electrophoresis (8, 20).

Figure 5 shows the plots of hydropathies of GS, NR_I and NR_{II}. The profiles of these plots do not show long stretches of hydrophobic or hydrophilic amino acids, as expected for globular-soluble proteins.

The deduced amino acid sequence for GS indicates that this protein is formed of 470 amino acids with the composition shown in Table 1.

It has been reported that mixed-function oxidation of GS from E. coli causes loss of catalytic activity. This inactivation correlates with the loss of one histidine residue per subunit (33). The peptide, MHCHM, which contains the oxidizable histidine and may form part of one of the well-studied cation-binding sites of GS, is located from residues 270 to 274 (Fig. 2). The loss of the catalytic activity upon oxidative modification is probably due to alteration of binding of divalent cations essential for activity. On this basis, it has been proposed that this peptide constitute part of the active site of the enzyme (33). This proposition is confirmed by the X-ray crystallography studies made by Almassy et al (38) which identify the GS active site in the electrodensity map. The distance between this peptide and the tyrosil residue to which AMP is covalently bound is of 125 amino acids (Fig. 2).This location is in agreement with the X-ray crystallography data and, as indicated by Almassy et al (38), it explains how the adenylyl group at Tyr399 could affect the active site by interacting with N-domain residues, restricting the structure or motion of part of the N-domain with respect to the C-domain.

The comparison of the predicted amino acid sequences between the

TABLE 1. AMINOACID COMPOSITION AND CODON USAGE

Amino acid	codon	glnA	glnL	glnG	Amino acid	codon	glnA	glnL	glnG
F	UUU	4	5	10	Y	UAU	5	1	3
	UUC	19	2	4		UAC	10	6	2
L	UUA	-	9	6	H	CAU	4	6	11
	UUG	1	8	5		CAC	10	4	4
	CUU	3	3	8					
	CUC	-	4	5	Q	CAA	2	7	6
	CUA	-	4	2		CAG	9	16	23
	CUG	28	25	33					
					N	AAU	2	6	5
I	AUU	12	13	10		AAC	18	8	7
	AUC	16	11	13					
	AUA	-	-	1	K	AAA	19	7	10
						AAG	5	0	5
M	AUG	7	8	11					
					D	GAU	13	12	17
V	GUU	8	3	7		GAC	19	8	8
	GUC	4	4	5					
	GUA	9	1	2	E	GAA	29	16	24
	GUG	12	10	15		GAG	7	6	10
S	UCU	12	1	2	C	UGU	1	-	1
	UCC	12	2	5		UGC	3	-	1
	UCA	1	3	3					
	UCG	2	4	2	W	UGG	2	1	7
	AGU	-	5	4					
	AGC	1	5	4	R	CGU	15	7	15
						CGC	9	11	14
P	CCU	4	4	-		CGA	-	1	3
	CCC	-	3	3		CGG	-	4	5
	CCA	6	2	6		AGA	-	-	1
	CCG	14	13	22		AGG	-	1	-
T	ACU	6	-	3	G	GGU	15	6	8
	ACC	12	9	9		GGC	19	12	15
	ACA	-	2	1		GGA	1	-	1
	ACG	1	6	1		GGG	1	8	8
A	GCU	12	5	8					
	GCC	9	6	10					
	GCA	13	2	7					
	GCG	9	13	24					

The underlined codons correspond to rare codons. Aminoacids are designated according to the one letter code.

Anabaena (34) and E. coli glutamine synthetase indicate that they share an homology of 53%. This homology increases up to 71% if we consider those amino acids which are functionally equivalent. The peptide MHCHM as well as

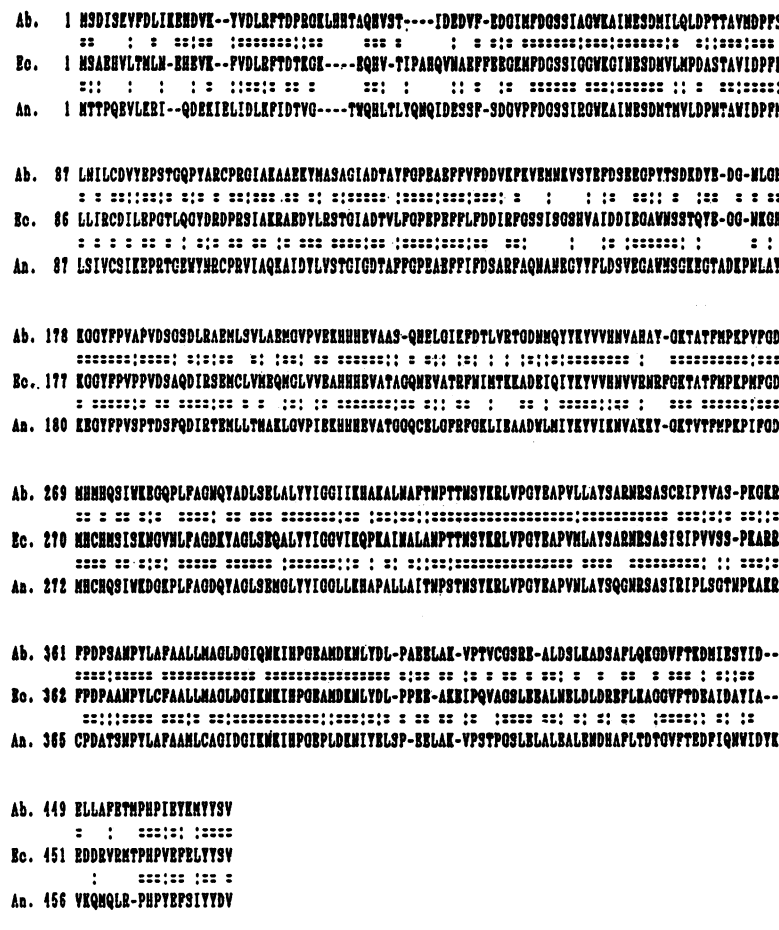

Fig. 6. Alignment of the E. coli (Ec.), A. brasilense (Ab.) and Anabaena (An) GS aminoacid sequence. = identical residues; : functionally equivalent residues.

the region around are highly conserved between the GS of Anabaena and E. coli. Also, the tyrosil residue (Tyr399) which is adenylylated in the E. coli GS is found in the Anabaena GS at a similar position. The region upstream of Tyr399 shows a very high degree of homology (Fig. 6). In the case of the Azospirillum brasilense GS these regions are also found highly conserved (42) (Fig 6). Although the MHCHM peptide in the E. coli GS in A. brasilense has the sequence MHMHQ, the oxidizable histidine residues are still present. Neither adenylylation nor mixed-function oxidation of GS have

been shown in A. brasilense; the conservation of the regions involved in those functions might imply that they in fact take place.

The comparison between the nucleotide sequence of the E. coli and K. pneumoniae glnL and glnG genes shows 81% and 82% homology, respectively, indicating that those genes are highly conserved in these organisms. This homology is higher than that expected according to DNA reassociation studies which indicate that the percent relatedness between E. coli and K. pneumoniae is 30-38% (40). The high conservation found for these two genes may be due to the different domains with very specific functions in both gene-products (NR_{II} and NR_{I}) (11,18,26,37,39). Also a very high homology (93% in both cases) is found between their amino acid sequences suggesting that these proteins (NR_{I} and NR_{II}) have identical roles in both organisms. Our analysis indicates that NR_{I} has a structural domain common to DNA-binding proteins, localized from residue 444 to 465 (Fig. 4), as it is the case for NR_{I} of K. pneumoniae (35) and Bradyrhizobium parasponiae (39). This observation is supported by experimental evidence which shows the ability of this protein for DNA-binding (18,26,41). In a recent work (35,39) it has been reported that the NR_{I} and NR_{II} proteins of B. parasponiae and K. pneumoniae share homology with other regulatory proteins. Many of these proteins function in pairs to regulate gene expression in response to environmental stimuli such as nutrient limitation (phoR/phoB), altered osmolarity (envZ/ompR) and plant exudate (virA/virG). As expected, this homology is also found in the E. coli NR_{I} and NR_{II} proteins, between residues 7 and 120 for NR_{I} (Fig. 4) and between residues 86 and 349 for NR_{II} (Fig. 3)

Although, the analysis of the amino acid sequences of the E. coli and K. pneumoniae (2) NR_{II} also shows and extensive homology to the DNA-binding domains of other site-specific DNA-binding proteins, there is not any experimental evidence which support these observations. Recently, it has been shown that NR_{II} controls the activity of NR_{I} by covalent modification. NR_{II} is and NR_{I}-kinase that can be converted by the glnB product (P_{II}) to an NR_{I}-phosphate phosphatase (37). It has been observed that all the protein kinases present a glycine-rich region followed by a lysine residue which may form part of the ATP binding site (43). We have found a region with these characteristics in the NR_{II} amino acid sequence between residues 306 and 331. This region is also observed in the C-terminal portion of other gene products homologous to NR_{II} which form part of two-component regulatory systems (39). This opens the question whether protein phosphorylation is a more common regulation mechanism in bacteria than is recognized at present.

As other complex operons in E. coli, the glnALG operon is formed by regulatory (glnL and glnG) and non-regulatory genes (glnA). In order to adjust the expression of genes in the nitrogen-regulated network the glnALG operon can be transcribed from two promoters upstream glnA and an internal promoter which allows the expression of the regulatory genes separately from glnA. The levels of the glnL and glnG products are always kept lower than those of the glnA product (18, unpublished results). The termination signal located between glnA and glnL (16, 26) is a way to keep these levels low when the three genes are transcribed from the same promoter. The analysis of the codon usage in the glnA, glnL and glnG coding reading frames indicates that glnL and glnG contains an unusually large number of rare codons, 26% for glnL and 20% for glnG (Table 1). The use of rare codons in these genes may be one of the mechanisms, at least in some physiological conditions, to mantain their products at low molar concentration compared with GS. This is in agreement with the finding that the regulatory genes usually contain higher percentages of rare codons (24%) than non-regulatory ones (12%) (36).

Special attention is being paid to the glnALG operon due to the major participation of NR_I and NR_{II} in ensuring harmonic regulation of nitrogen assimilation genes, and to the central role of GS in the nitrogen metabolism.

The availability of the complete nucleotide sequence of this operon will help to answer questions concerning the identification of specific domains in the proteins, their evolution and their relations with other components of their own and other regulatory networks.

ACKNOWLEDGMENTS

We thank Alejandro Garcíarrubio and Irene Castaño for valuable discussions; Enrique Merino for assistance with the computer work; Jaqueline Mazzari and Rosalba Sánchez for technical assistance; Regina Basurto and María Dolores Cuéllar for typing this manuscript. J. Miranda-Ríos was recipent of a Consejo Nacional de Ciencia y Tecnología (México) fellowship. This work was supported by research grants PCCBBNA-020413 from Consejo Nacional de Ciencia y Tecnología (México), and 17/85 from Fondo de Estudios e Investigaciones Ricardo J. Zevada.

*To whom correspondance should be addressed.

REFERENCES

1. Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. (1985) Nucl. Acids Res., 13, 4539-4555.

2. McFarlane, S.A. and Merrick, M. (1985) Nucl. Acids. Res., 21, 7591-7606.
3. Merrick, M.J. (1982) Nature (London) 297, 362-363.
4. Magasanik, B. (1982) Ann. Rev. Genet. 16, 135-16.
5. Gottesman, S. (1984) Ann. Rev. Genet. 18, 415-441.
6. Kustu, S., Burton, D., Garcia, E., McCarter, L. and McFarland N. (1979) Proc. Natl. Acad. Sci. USA. 76, 4576-4580.
7. Pahel, G. and Tyler, B. (1979) Proc. Natl. Acad. Sci. USA, 76, 4544-4548.
8. McFarland, N., McCarter, L., Artz, S. and Kustu, S. (1981) Proc. Natl. Acad. Sci. USA. 78, 2135-2139.
9. Pahel, G., Rothstein, D.M. and Magasanik, B. (1982) J. Bacteriol 150, 202- 213.
10. Guterman, S.K., Roberts, G and Tyler, B. (1982) J. Bacteriol. 150, 1314-1321.
11. MacNeil, T., MacNeil, D. and Tyler, B. (1982) J. Bacteriol. 150, 1302-1313.
12. Hirschman, J., Wong, P.-K., Sei, K., Keener, J. and Kustu, S. (1985) Proc. Natl. Acad. Sci. USA 82,7525-7529.
13. Hunt, T.P. and Magasanik, B. (1985). Proc. Natl. Acad. Sci. USA 82, 8453-8457.
14. Ueno-Nishio, S. Backman, K.C. and Magasanik, B. (1983) J. Bacteriol. 153, 1247-1251.
15. Krajewska-Grynkiewicz, K. and Kustu, S. (1984) Mol. Gen. Genet. 193, 135-142.
16. Ueno-Nishio, S., Mango, S., Reitzer, L.J. and Magasanik, B. (1984) J. Bacteriol. 160, 379-384.
17. Dixon, R. (1984). Nucl. Acids Res. 12, 7811-7830.
18. Reitzer, L.J. and Magasanik, B. (1985) Proc. Natl. Acad. Sci. USA. 82, 1979-1983.
19. Leon, P., Romero, D., Garciarrubio, A., Bastarrachea, F. and Covarrubias, A.A. (1985) J. Bacteriol. 164, 1032-1038.
20. Reitzer, L.J. and Magasanik, B. (1983) Proc. Natl. Acad. Sci. USA. 80, 5554-5558.
21. Hawkes, T., Merrick, M. and Dixon, R. (1985) Mol. Gen. Genet. 201, 492-498.
22. Merrick, M.J. and Gibbins, J.R. (1985) Nucl. Acids. Res. 13, 7607-7620.
23. Backman, K.C., Chen, Y.M., Ueno Nishio, S. and Magasanik, B. (1983) J. Bacteriol. 154, 516-519.
24. Bueno, R., Pahel, G. and Magasanik, B. (1985) J. Bacteriol. 164, 816-822.
25. Covarrubias, A.A. and Bastarrachea, F. (1983) Mol. Gen. Genet. 190, 171-175.
26. Rocha, M., Vazquez, M., Garciarrubio, A. and Covarrubias, A.A. (1985) Gene 37, 91-99.
27. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
28. Bankier, A.T. and Barrel, B.G. (1983) Techniq. in Nucl. Acid Biochem. 13508, 1-34.
29. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA. 80, 3963-3965.
30. Messing, J. (1983) Methods Enzymol. 101, 20-78.
31. Heinrikson, R.L. and Kingdon, H.S. (1971) J. Biol. Chem. 246, 1099-1106.
32. Shapiro, B.M. and Stadtman, E.R. (1968) J. Biol. Chem. 243, 3769-3771.
33. Farber, J.M. and Levine, R.L. (1986). J. Biol. Chem. 261, 4574-4578.
34. Fisher, R., Tuli, R. and Haselkorn, R. (1981). Proc. Natl. Acad. Sci. USA 78, 3393-3397.
35. Drummond, M., Whitty, P. and Wooton, J. (1986) EMBO J., 5, 441-447.
36. Konigsberg, W. and Godson, N. (1983) Proc. Natl. Acad. Sci. USA 80, 687-691.

37. Ninfa, A.J. and Magasanik, B. (1986) Proc. Natl. Acad. Sci. USA 83, 5909-5913.
38. Almassy, R.J., Janson, C.A., Hamli, R., Xuong, N-H, and Eisenberg, D. (1986) Nature 304-309.
39. Nixon, B.T., Ronson, C.W. and Ausubel, F.M. (1986) Proc. Natl. Acad. Sci. USA 83, 7850-7854.
40. Sanderson, K.E. (1976) Ann. Rev. Microbiol. 30, 327-349.
41. Ames Ferro-Luzzi, G and Nikaido, K. (1985) EMBO J., 4, 539-547.
42. Bozouklian, H. and Elmerich, C. (1986) Biochimie 68, 1181-1187.
43. Hunter, T. and Cooper, J.A. (1985) Ann. Rev. Biochem. 54, 897-930.