Tight linkage of glnA and a putative regulatory gene in Rhizobium leguminosarum

S.Colonna-Romano, A.Riccio, M.Guida, R.Defez, A.Lamberti, M.Iaccarino*, W.Arnold¹, U.Priefer¹ and A.Pühler¹

International Institute of Genetics and Biophysics, CNR, Via Marconi 10, 80125, Naples, Italy and ¹Lehrstuhl fur Genetik, Universitat Bielefeld, Postfach 8640, 4800 Bielefeld 1, FRG

Received December 29, 1986; Revised and Accepted February 12, 1987

ABSTRACT

<u>Rhizobium leguminosarum</u>, biovar viceae, strain RCC1001 contains two glutamine synthetase activities, GSI and GSII. We report here the identification of <u>glnA</u>, the structural gene for GSI. A 2 kb fragment of DNA was shown to complement the Gln phenotype of <u>Klebsiella pneumoniae glnA</u> mutant strains. DNA sequence analysis revealed an open reading frame (ORF) of 469 codons specifying a polypeptide of 52,040 daltons. Its deduced amino acid sequence was found to be highly homologous to other glutamine synthetase sequences. This ORF was expressed in <u>Escherichia coli</u> minicells and the corresponding polypeptide reacted with an antiserum raised against GSI. Upstream of <u>glnA</u> we found an ORF of 111 codons (ORF111) preceded by the consensus sequence for an <u>ntrA</u>-dependent promoter. Minicells experiments showed a protein band, with a molecular weight in good agreement with that (10,469) deduced from the nucleotide sequence. On the basis of homology studies we discuss the possibility that the product of ORF111 is equivalent to the P_{II} protein of <u>E.coli</u> and plays a similar role in regulation of nitrogen metabolism.

INTRODUCTION

<u>Rhizobium</u> bacteria use ammonia for growth in the free-living state, but in the <u>Rhizobium</u>-legume symbiosis the nitrogen fixing bacteroids export all ammonia produced in the nitrogen fixing process to the plant fraction of the symbiotic nodule (1). Thus, the enzymes for ammonia assimilation need to be regulated differently in the two bacterial states. Another peculiarity of <u>Rhizobiaceae</u> is the presence of two glutamine synthetases (GS; EC 6.3.1.2): GSI, similar to the GS of enteric bacteria, regulated by adenylylation and relatively heat stable; and GSII, heat labile and not known to be modified after translation (2,3,4,5,6). Biochemical studies (7) have shown that GS plays a central role in the regulation of nitrogen metabolism, while genetical studies (8,9) indicate that regulation of the <u>glnA</u> gene in <u>Enterobacteriaceae</u> is very complex. A coordinated nitrogen control system similar to that of enteric bacteria (10) has not been described in <u>Rhizobium</u> spp.. We started a study of the GS activities of <u>R.leguminosarum</u> biovar <u>viceae</u>, strain RCC1001, in order to better understand nitrogen assimilation and its regulation in this species (11). In this paper we report the DNA sequence of <u>glnA</u>, the structural gene for GSI, and the comparison of the deduced polypeptide sequence with other known sequences. We also report the sequence of an open reading frame upstream of <u>glnA</u> coding for a protein which appears to play a regulatory role, probably equivalent to that of the P_{II} protein, which is involved in the regulation of GS adenylylation and <u>glnA</u> expression in <u>Escherichia coli</u> (12,13).

MATERIALS AND METHODS

Strains, plasmids and media

Abbreviations used are: Ap (Ampicillin), Km (kanamycin), Cm (chloramphenicol), Tet (tetraciclin).

Strains used were <u>E.coli</u> HB101 (14), JM83 (15), DS998 (16); <u>Klebsiella</u> <u>pneumoniae</u> wild type (17), UNF1827 and UNF1838 (18); <u>R.leguminosarum</u> strain LPR1105, a rifampicin resistant derivative of RCC1001 (19); <u>Agrobacterium</u> <u>tumefaciens</u> strain LBA2715 containing the <u>R.leguminosarum</u> symbiotic plasmid (pSym) (20).

Plasmids used were: pMMB34 (21); pSVB20, pSVB23, pSVB24, pSVB25 (pUC8 derivatives, Ap^{r} ; W.A. et al., in preparation); pACYC184 (22); p7D9 (11), Gln⁺, Km^r, containing 27 kb of <u>R.leguminosarum</u> DNA inserted in the cosmid vector pMMB34; pMG10 (11), Gln⁺, Cm^r, containing 6.5 kb of <u>R.leguminosarum</u> DNA and 3.4 kb of pMMB34 DNA inserted in the vector pACYC184.

Media used were: TY (14); PA (14) supplemented with 0,1% glucose; minimal citrate (23). When needed, glutamine was used at a concentration of 200 ug/ml.

DNA manipulation

Most procedures used were according to Maniatis et al.(14). The rapid isolation of plasmid DNA was performed as described (24). Plasmid DNA was prepared according to Davis et al.(25). pSym DNA was prepared from strain LBA2715 as described (26). Hybridization experiments were all carried out at 60°C. DNA sequencing was performed with minor modifications (W.A. et al., in preparation) of the method of Maxam and Gilbert (27).

Protein synthesis in minicells

Strain DS998 was transformed with specific plasmids and grown in minimal



Figure 1. (a): restriction map of the <u>glnA</u> region and the DNA left in different Bal31 deletions. The black bar indicate the location of the <u>EcoRI</u> fragment cross-hybridizing to <u>K.pneumoniae</u> <u>glnA</u> (11). The dotted lines indicate DNA (about 3 kb) omitted from the figure. Only the <u>glnA</u> region-proximal end of the deletions is indicated in the figure. The boundary of pl0-1 and p30-11 was confirmed by DNA sequencing (see Fig.2). The boundary of pE-9 is between the <u>EcoRI</u> and <u>PstI</u> restriction sites (striped bar); that of pD-8 is between the <u>SmaI</u> and <u>BamHI</u> sites. (b): sequencing strategy and location of the thre ORFs deduced by computer analysis. Vertical arrows indicate putative ribosome binding sites. P shows the position of a sequence homologous to the consensus for an <u>ntrA</u>-dependent promoter. A black bar indicates the <u>SalI</u> fragment used as a probe in the hybridization experiments of Fig.5.

medium supplemented with thiamine (10 ug/ml), casaminoacids (0,5%) and the appropriate antibiotic. Minicells were purified (28), divided into aliquots and stored at -80° C. After preincubation for 30 min, incorporation of either 35 S-methionine, 3 H-valine or 3 H-leucine was performed for 60 min (28). Immunoprecipitation was performed as described (29).

TCGAGATETE 10 TATTORNAL DOCAMAATA OCTUACTITI TODETTIOOC ALCTOTTITI GATOORAGAC AATOTTITIC CACOLEGUAAA COACAATACAA TIOCOCOCAA ACCOCCOCTI 130 TTERCHEMA TECHNOTETT TTTTECHEMA CATEGOTTE ANTTECTE CATATOTICE CATATOTICE COMPERSION ANTICICE ATMACHEME ANTICENCE ALASTITUCE 230 ATC ANA ANG ATC GAN GOC ATC ATT ANG COT TTC ANG CTC CAC GAN GTC AGG MOC COT TCA GGA GTC GOT GTC GAG GGT ATC MOG GTC AGG 340 mot lyn iye ile giu als ile ile lyn yre pae lyn lam amp giu val arg eer pro eer giy val giy iem gin giy ile thr wel thr GAN GOC ANG GOT TTO GOC GOT CAO ANG GOC GAC AGG GAL GTO TAO GOC GOL GOC GAL TAO GTO GAT TTO CTO GOC ANG GTA AAA GTO 430 giv ala iya giy pasa giy arg gia iya giy kis the giu ion tyr arg giy ala giu tyr vel val any pasa low pro iya vel CAG GTT GTA CTG GOC GAC GAG AAT GOC GAA GOC CTG ATT GAA GOC ATC GOC AAG GOC GAC GAG ACC GOC GOC ATC GOC GAC GGA AAG ATC 370 glw val vel lee ala amp giw aam ala giw ala val ile giw ala ile arg lya ala ala gia thr giy arg ile giy amp giy lya ile CLACTITICCC TTO GTO TOO AND CTO CAN CAN CTO ATT ATT COO ATO COO ACO GOO CAN ACO GOO ATT CAT GOO ATO TOA phe val see ann val giv giv vel tie arg tie arg tir giv giv the giv tie ann als tie and AATOGOGCAA ACTOOCTTAC 616 ATC<u>OCHOCHA</u> ACTACTTA ATC CCG ACC CCA ACC CAA ATT CTC AAC CAO ATC AAC CAC GAA CAC CTA AAC TTC 718 met als the als ear gle ile los lys glo ile lys glo ann amp val lys plo CETCATORCE CACCENCACE GTCATCOCAN סדר לאז רדים כמד דדים אמר השבי כמא מצב מכא אמר מדים להם לאז ידים אמר אדים לא מעור ביו ביו לא מאר האל האל הידים ל ידים אמים ואיים הידים אמר לאז מאר הידים לא הידים הידים הידים האל הידים האל הידים הידים הידים הידים הידים הידים ה ETE ATE THE GAE GOE THE TOE ATE GOE GOE TOE AND GOE ATE AND GOE ATE CAE ATE OTE CTE ATE COE CAE ACE GOE GAE ACE GTE CAT wal not plue any gly ser per ile gly gly try lys als ile and glu ser any met wal low met pro any the glu the wal his ATC 898 GAG COG THE THE OCA CAG FOG ACC ATG ONE ATE OTE TOE CAG ATE CHE GAT COG OTE TOE COE GAG COE TAIT AAC COE CAT COG COE GAG AND PTO POR PAR ALL SIL AND TOE ATE SIL ACC GOC ANG ANG GOC GAA GOC TAC CTC ANG GOA TOC GOC ATC GOC GAT AGG GTT TTC GTC GOC GOA GOC GAA TTC TTC GTC TTC the also by by also gin also typ loss by also see giy the giy any the val phe wal giy pro gin also gin phe phe val phe CAC 1078 ACC 1168 CAG GTO AND TAC AND GOD GAT OUT TAC ANT ADD GOD TTO AND CTO GAT TOO ADD GAA CTO GOD TOO AND CAD GAD AND ADD WE ly who ly ty lys als app pro by ann the gly pho lys los app por the glu low pro sor ann any app the app ty glu AGE ANG ETE GGE CAT GGE GGE GTE ANG GGE GGE TAG THE GGE GTE GGE GGE GTE GAG AGE GGE CAG ATE GTE TOG GAA ATE ETE 1358 gly som lon gly kis arg pro arg val iya giy giy tyr pika pro val pro yra any sar ala gin any mat arg sar gin mat AGG GTE GTE TOE GAG ATG GOC GTE GTE GTE GAG AAG GAT CAG GAT GAG GTE GOE GOE GOE GAG GAG GAT AGG GTE GTE GOE AAG GOE The well bes een allo met ally wel wel wel and alle line his his alle alle alle alle alle his anne the low wel and anne alle GAC 1348 GAN GTC GGT ΑΤΟ ΑΜΟ ΤΤΟ ΑΜΟ ΑΤΟ CMO ΑΤΟ ΤΑΟ ΑΤΑ ΤΑΟ ΕΤΟ ΕΛΕ Ο ΕΛΕ Ο ΕΧΟ ΟΧΟ ΤΑΤ ΘΟΟ ΑΜΟ ΑΝΟ ΘΟΟ ΑΟ ΤΤΟ ΑΤΟ ΕΟΟ Ι438 βία ίστι βία βία ματά πρόμα μα στις βία δία την Έργα την αναί δια βία και δια απα εία την βία την δια τά την πρό AAG CCG ATC TAT 1528 TOO GAG AGO TGO CTO TTT TAT ATO GGO GGO ATO ATO AAG GAT GGO AAG GGO ATO AAG GGT TTO AGO AAT GGO TGO AGO AAG 1618 mer giu mer cym lew pèn tyr ile gly gly ile ile lym bin als lym als ile new als pèn thr ann pro ser thr ann OCC OCT TOC TAC AME COT TTC OTC OCC OCT TAC GAA OCA OTT GTC CTC OCC TAT TOC OCC GAC COC MAC COC TOC COC TOC COC ATT COC ATT COC TTC 1708 ser tyr lyn arg len wal pre gly tyr gla ala pre val len len ala tyr eer ala arg een arg eer ala ser Cys arg ile pre pbe ANG COC ANG COC ANA COC CTC CAC CTC CAC TTC COC CAT COC ANG COC ANG COC TAT CTC ACC COC ATG CTG ATG CTC ATG CT OGC TOC GGC CTC GAC GGC ATC ANG ANG ANG ATC CAT CCC GGC ANG GGC ATU GAC ANG GAT CTC TAG GAT CTG GGG GGG ANG GAA TTG ANG ANG ATC 1888 gly low any gly ile lys een lys ile kis yre gly lys ele met any lys any low tyr any low yre yro lys glu low lys lys ile COD ACC CTC TCC GOC ACC TTC CGC CAA GCA CTC CAA ACC CTC GAC AAC CAC GCC ACC TTC CTC ACC GCC GCC GCC GTC TTC GAC GAC pro thr wel cys gly ser low arg gin als low gin ser low any lys say arg lys pix low thr als gly gly wel pixe say any GAT 1978 CAG ATE GAT GOE TTE ATE CAG CTO ANG ATG GET CAG CTO ATG GET ITE GAA ATG AGG CGE CAT COE GTE GAA TAG GAG ATG TAG gin ile amp ala pho ile gin low iys met ala gin wel met arg pho gin met the pro him pro val gim tyr amp met tyr tyr TCG 2068 TODOCTOTE ANCAGAMAA COOCCTOC COCCOCTTC COCCOCTTC CONTOCACE TCATOGACCE CACTACACCE CTURALACE 2184 OCC TAA

TOGACTOCIC CATAATTITO TOCTTANATC CAT

Figure 2. DNA sequence and deduced amino acid sequence of ORF111 and ORF469 from the <u>XhoI</u> site to the <u>ClaI</u> site of Fig.1. A consensus sequence for an <u>ntrA-dependent</u> promoter is boxed, while putative ribosome binding sites are underlined. An asterisk (*) shows the left boundary of pl0-1 and p30-11 deletions.

RESULTS

Localization of the glnA region on pMG10.

We have previously shown (11) that cosmid p7D9 (see Materials and Methods) complements the Gln phenotype of <u>K.pneumoniae</u> strains UNF1827 (<u>glnA</u>) and UNF1838 (Δ <u>glnA-ntrC</u>). A subclone of p7D9, pMG10, complementing both strains, is used in this paper. In Fig.la we show part of pMG10 DNA containing the 1.3 kb fragment cross-hybridizing to <u>K.pneumoniae glnA</u> (11).

We generated deletions of pMGlO by digestion at the single <u>Xho</u>I site shown in Fig.la and treatment with Bal31 nuclease. After ligation and transformation into strains UNF1827 and UNF1838 we analyzed 24 clones, 7 of

R.1. HATAJSEJILKQIK - ENDVKFVDLRFTDRRASLQHVTHDVVCVDE L.C. H - SLAJEHVLTHLR - EHEVKFVDLRFTDTKGKEOHVTIPABQVMA A.b. H - SDISKVFDLIKEHDVKVVDLRFTDPRGKLHTAQHVSTTDE Ana. HTTPQEVLKREQ - DEKIELTDLKFTDTVGTVQHLTLYQWQIDE R.1. DM FADGOVHFDGSSIGGWKAINESDHVLNPDTETVHHDPFPAQS E.C. EFFEEGRHFDGSSIGGWKGINESDHVLNPDASTAVIDPFFADS A.D. DVFEDGINFDGSSIAGWKAINESDHILQLDPTTAVHDPFSADS ADA. SSFSGGVFFDGSSIAGWKAINESDHITHVLDPNTAVIDPF K.1. TH VIIYCDILDPYSCEAYHRDPRCTARKAEAYLKASGIGDTYFY E.C. TLIIRCDILEPGTLQGYDRDPRSHSKRAEDYLRSTGIADTYL A.b. TLHILCDYYEPSTGQPYARCPRCIAKAAEKYHASAGIADTAYF Ana. TLSIYCSIKEPRTCEWYHRCPRVIAGKATDYLVST<u>CIGDTA</u>F R.1. GPEAEFFVFDDVKYKADPTNTGFKLDSTELPSHDDTDYETG--E.C. GPEPEFFDDDIFFGSSISGSHVAIDDIEGAWNSSTQYEGG--A.b. GPEAEFFVFDDVKFKVEMNKVSYEFDSEEGPTSDKOTBDGKP Ann. GPEAEFFTIFDSARFAQNANEGYTFLDSVEEGFABSGKEGTAD--...... R.1. (N L G H R P R V K G G Y F P V P P V D S A Q D M R S E M L T V L SE M G Y V V E K H H E.C. B K G H R P A V K G G Y F P V P P V D S A Q D I R S E M C L V M E Q M G L V V E A H A.D. N L G H R P G V K G G Y F P V A P V D S G S D L R A R H L S Y L A E H G V P V E K H H Ana. N L A Y K P R P K E G Y P P V S P T D S P Q D I R T E M L I T M A K L G V P T E K H H R.1. HEVAAA-QHELGIKFDTLVRNADKHQIYKYVHQVANAYGKTA E.C. HEVATAGQNEVATRFNTHTKADETQIYKYVVHNVAHRJGKTA A.b. HEVAAS-QHELGIKFDTLVRTGDNHQTYKYVVHNVAHRJGKTA Ana. HEVAAS-QHELGIRFEGKLIEAADWLHIYKYVIKNVAKKYGKT R.1. TFMPKPIFGDNGSGNBVHQSIWKGGKPTFAGDEYAGLSESCLF E.C. TFMPKPMFGDNGSGNHCHB<u>MSL</u>SKMGVNLFAGDKYAGLSEQALY A.b. TFMPKPVFGDNGSGNHHRQSIWKEGQPLFAGMQYADLSELALY Ana. TFMPKPIFGDNGSGNHCKQSIWKDCKPLFAGDQYAGLSEMGLY R. 1. YIGGIIKHAKAINAFTNPSTNSYKRLVPGYEAPVLLAYSARNR E.C. YIGGVIKHAKAINALANPTTNSTKRLVPGYEAPVLLAYSARNR A.b. YIGGIIKHAKALNAFTNPTTNSYKRLVPGYEAPVLLAYSARNR Ana. YIGGILKHAFALTATTNPSTNSYKRLVPGYEAPVNLAYSQGNR R.1 SASCRIP-FCSNPKAKRVEVRFPDPTANPYLAFAAHLHAGLDC E.c. SASIRIP-VVSJPKARRIEVRFPDPAANPYLCFAALLMAGLDC A.b. SASCRIP-VVASPKGKRVEVRFPDPSANPYLAFAALLHAGLDG Ama. SASTRIPLSCTNPKAKRLEFRCPDATSNPYLAFAAHLCAGTDG R.1. IKHKIHPGKAMDKDLYDLPPKELKKIPTVCGSLRGALESLDKD E.C. IKNKIHPGEAMDKWLYDLPPEEAKEIPOVGGSLEGAL¥ELDKD A.b. IGNKIHPGEAMDKWLYDLPAELAKVPPTVCGSLEAL¥EAL Ama. IKNKIHPGKPLDKWIYELSPELAKVPSTPGSLEKALUEAL 00000 R.1. RKFI.TAGGVFDDDDQIDAFIELKMA-EVMRFENTPHPVEVDHYY E.C. REFLKACGVFIDEAIDAYIAILRRE-EODRVRHTPHPVEFELYV A.b. SAFLQKGDVFTKDN1ESYIDLRTE-ELLAFETMPHPLEVKNYY A.a. HAFLTDTGVFTEDFIQNWJDYKLANEVKQNQLRPHPYEFSTYY R.1. SA * E.c. SV * A.b. SV *

Ana WV+

Figure 3. Homology of the deduced <u>R.leguminosarum</u> GSI sequence to other GS sequences. R.l.= <u>R.leguminosarum</u>; E.c.= <u>E.coli</u>; A.b.= <u>A.brasilense</u>; Ana.= <u>Anabaena</u>. Sequences assumed to correspond to the N-domain and C-domain (33) are indicated by (*) and (o) respectively.

which were Gln^+ and 17 were Gln^- . Restriction analysis showed that the <u>XhoI</u> site was deleted in all clones. We searched for the presence of the surrounding sites in order to construct a restriction map of all clones and the <u>Gln</u>⁺(pl0-1) and Gln⁻(p30-11) clones shown in Fig.la were chosen because their end-points, confirmed by sequencing (see below Fig.2), are the closest definition of the left-side border of the glnA region.

We used clone plO-1 to generate a set of new deletions by digestion at the single <u>Bst</u>EII site and treatment with Bal31 nuclease. Restriction analysis of 40 clones carrying deletions showed that the <u>Eco</u>RI site to the right of <u>Cla</u>I is present in all Gln^+ clones and absent in all $Gln^$ clones. In Fig.la a Gln^+ clone (pE-9) and a Gln^- clone (pD-8) are shown. <u>R.leguminosarum</u> DNA left in pE-9 is 2 kb long and therefore about 2 kb of DNA are sufficient to suppress the Gln^- phenotype of strains UNF1827 and UNF1838.

DNA sequencing of the glnA region.

The sequence of <u>R.leguminosarum</u> DNA from the <u>XhoI</u> site to the <u>ClaI</u> site indicated in Fig.1 is presented in Fig.2. A computer search for open reading frames longer than 100 codons resulted in the three ORFs shown in Fig.1b. A sequence with good homology to an <u>ntrA</u>-dependent promoter (9) is found at position 174-190 followed by a putative ribosome binding site at position 235-240 and an ORF of 111 codons (ORF111). At the end of this ORF there is no obvious sequence suggesting rho-independent transcription termination and at position 650-655 there is a new putative ribosome binding site followed by an ORF of 469 codons (ORF469). Downstream of this, we do not see any sequence suggesting rho-independent transcription. On the opposite strand another ORF of 517 codons is present (see Fig.1b) starting with an Arg codon at position 2215 and ending at position 662.

DNA sequencing of appropriate subclones shows that p10-1 is deleted up to nucleotide 347 and that p30-11 is deleted up to nucleotide 750, as indicated in Fig.2.

We compared the deduced aminoacid sequence of the ORFILL product with the protein sequences present in the January 1986 data bank using the Micro-Genie program (30) and found no significant homology. However, it was pointed out to us by dr. M. Merrick that codons 47 to 57 of ORFILL are strikingly similar to the sequence of a peptide of the P_{II} protein of <u>E.coli</u> (12):

> P_{II}: -Gly-Ala-Glu-Tyr-Met-Val-Asp-Phe-Leu-Pro-Lys-ORFIII: -Gly-Ala-Glu-Tyr-Val-Val-Asp-Phe-Leu-Pro-Lys-



Figure 4. SDS-PAGE of minicells extracts containing different plasmids. Panel A: S-methionine labeling of minicells containing: (a): vector pACYCl84; (b): pMGl0; (c): pl0-l; (d): p30-l1; (e): pE-9; (f): pD-8. Panel B H-valine (a and b) or H-leucine (c) labeling of: (a): pACYCl84; (b) and (c): pMGl0. The 58, 33 and 12 markers are explained in the text. Lanes (a) show the CAT band of pACYCl84 present also in the other lanes. Molecular weights of insert-specific bands were interpolated from the molecular weight of commercially available C-standards. The experiments shown were performed with 10% (Panel A) and 20% (Panel B) polyacrylamide.

Since the complete amino acid sequence of the P_{II} protein of <u>E.coli</u> is not published, we could not compare other portions of the two proteins. The molecular weight of the P_{II} protein of <u>E.coli</u> is 11,000 to 13,400 (31), similar to that of the deduced amino acid sequence of ORF111 (10,469). The comparison of the amino acid composition of the P_{II} protein of <u>E.coli</u> (31) and of the ORF111 product shows a striking similarity. We conclude that the product of ORF111 might be equivalent to the <u>E.coli</u> P_{TT} protein.

We also searched for homology of the deduced amino acid sequence of ORF469 to other protein sequences present in the data bank and found a significant value only with <u>Anabaena</u> sp. strain 7120 GS. We introduced in the data bank the deduced protein sequences of <u>E.coli</u> (32) and <u>Azospirillum brasilense</u> sp.7 (33) GSs and that of <u>Bradyrhizobium japonicum</u> GSII (6). Homology of ORF469 is 56%, 63% and 68% with <u>Anabaena</u>, <u>E.coli</u> and <u>A.brasilense</u> GSs, respectively, and most of it is in boxes of identical amino acids, as shown in Fig.3, thus suggesting that this ORF codes for <u>R.leguminosarum</u> GSI. Homology with <u>B.japonicum</u> GSII is 15%. The molecular weight of the deduced amino acid sequence of ORF469 is 52,040, in good agreement with that (60,000)

of the GS purified from UNF1827(pMG10) (A. Fuggi and R.D., manuscript in preparation). The molecular weight of the deduced amino acid sequences of <u>Anabaena</u>, <u>E.coli</u> and <u>A.brasilense</u> GSs are 53,265, 51,814 and 51,917 respectively.

A recent paper (34) reports the crystallographic structure of <u>Salmonella</u> <u>typhymurium</u> GS and defines an N-domain formed by 92 N-terminal amino acids and a C-domain formed by 5 stretches of amino acids (123-137; 208-228; 263-276; 343-361 and 392-400), participating in the conformation of the active site. We measured homology on portions of the GS sequences which we assume to be analogous to the above mentioned domains (see Fig.3). We found that N-domain sequences of <u>Anabaena</u>, <u>E.coli</u> and <u>A.brasilense</u> GSs give values of homology to <u>R.leguminosarum</u> GSI slightly lower than those found for the complete protein (497, 582 and 552 respectively), while C-domain sequences show higher values (722, 692 and 822 respectively). A more detailed analysis shows that 4 of the C-domain stretches are highly homologous while the 5th one (residues 392-400) is very conserved in <u>R.leguminosarum</u>, <u>E.coli</u> and <u>A.brasilense</u> GSs and different in <u>Anabaena</u> GS.

Using the computer program reported above (30) we analyzed the secondary structure of 110 N-terminal and of 110 C-terminal residues of the 4 protein sequences. No significant homology was found among the secondary structures of the C-terminal sequences while the N-terminal sequences showed homology except in the case of Anabaena (data not shown).

The ORF on the opposite strand starts at least 80 codons upstream of the Arg codon, as shown by DNA sequencing not reported in this paper. Also <u>E.coli</u> and <u>A.brasilense</u> sequences present an extended ORF on the opposite strand, which is absent in the <u>Anabaena</u> sequence. The codons of these three ORFs are on a reading frame coinciding (with opposite polarity) with the reading frame of the respective GSs. Analysis of the deduced amino acid sequence in the data bank showed no significant homology of the <u>R.leguminosarum</u> ORF to other proteins.

Expression of pMG10 and its deletion derivatives in minicells.

In order to demonstrate the expression of the ORFs identified by DNA sequencing, minicells experiments were carried out. Plasmid pMG10 and the deletion derivatives described above were introduced into strain DS998 by transformation and minicells were purified. After labeling with 35 S-methionine and SDS-PAGE, we observed, in addition to vector-coded bands, a band specific of the insert of pMG10 (Fig.4A). This band, 58,000 in molecular weight, is also present in the Gln⁺ deletion derivatives pl0-1 and pE-9, is absent in the Gln⁻ deletion p30-11, while it is reduced in



Figure 5. Hybridization of a ${}^{32}P$ -pSym DNA fragment (left panel) and P-glnA DNA (right panel) to total DNA (lane t) and pSym DNA (lane p). The probe pSym DNA is a 1.8 kb EcoRI fragment isolated from plasmid pl085 (40); glnA DNA is a Sall fragment of pMGl0 indicated in Fig.lb and in the text. 5 ug of total DNA and 1 ug of pSym DNA (a plasmid preparation from strain LBA 2715), were digested with EcoRI and used in each experiment. The molecular weight of the hybridizing band is interpolated from a A _HindIII marker (not shown).

molecular weight (33,000) in the Gln⁻ deletion pD-8. We conclude that the latter is a truncated polypeptide. Therefore, this experiment shows that the insert of pMG10 codes for a protein, 58,000 in molecular weight, that the direction of transcription is left to right as indicate in Fig.lb and that this protein is necessary for the Gln⁺ phenotype either in strain UNF1827 or UNF1838. The additional band, approximately 28,000 in molecular weight (lanes b,c and d), appears to be coded by the DNA deleted in pE-9 and it is not relevant to this study.

A protein A-Sepharose 4CL column was treated with a polyclonal antiserum raised in rabbit against pure GSI obtained from strain UNF1827 (p7D9) (A.Fuggi and R.D., manuscript in preparation). This column retained the 58,000 molecular weight protein produced in 35 S-methionine labeled minicells. After elution with the appropriate buffer (see Materials and Methods), SDS-PAGE showed a single band comigrating with that of pMG10 (data not shown). We conclude that this band is GSI, coded by ORF469, which we call <u>glnA</u>.

In the experiment of Fig.4A we could not see any band corresponding to the size predicted from ORF111. Since this ORF contains only one methionine at its amino-terminal end (Fig.2) we repeated a minicell experiment using either 3 H-valine or 3 H-leucine and found a band, 12,000 to 14,000 in molecular weight, in addition to the 58,000 band (Fig.4B). These experiments demonstrate that ORF111 and <u>glnA</u> are expressed in minicells, while the ORF present on the opposite strand is not expressed.

The glnA region is not located on the symbiotic plasmid

A <u>SalI</u> fragment of pMGl0, shown in Fig.lb, was used as a probe and hybridized to <u>Eco</u>RI digests of <u>R.leguminosarum</u> total and pSym DNA. We observed hybridization only to total DNA, as shown in Fig.5. If, instead, the probe originates from <u>R.leguminosarum</u> pSym DNA, there is hybridization to both DNAs. Therefore the insert of pMGl0 is not located on pSym. The size of the hybridizing fragment is that (1.3 kb) of the <u>Eco</u>RI fragment of pMGl0 shown in Fig.la. Since total DNA was digested with <u>Eco</u>RI, this experiment shows collinearity between the <u>Eco</u>RI restriction sites of the insert of pMGl0 and those of total DNA.

DISCUSSION

Free-living <u>Rhizobium</u> bacteria assimilate ammonia for growth, but the nitrogen fixing bacteroids export to the plant fraction of the symbiotic nodule all ammonia produced. That is, when nitrogenase activity becomes derepressed ammonia assimilation is blocked, probably by repression and inhibition of the GSs present in the nodule (2,4). At the same time a nodule-specific plant GS is derepressed (35). The study of regulation of <u>R.leguminosarum glnA</u> is important to understand the physiology of ammonia utilization in the symbiosis.

We report in this paper the sequence of the <u>glnA</u> gene and of a contiguous gene which, at the amino acid level, shows homology to the P_{II} protein of <u>E.coli</u> (31). These two genes are not located on the symbiotic plasmid (pSym) as shown in Fig.5.

The DNA of pMG10 required to suppress the Gln phenotype of UNF1827 and of UNF1838 has been localized with the help of Bal31 deletions and restricted to about 2 kb. This DNA region overlaps with the fragment cross-hybridizing to <u>K.pneumoniae glnA</u> DNA (11) as indicate in Fig.1a. The DNA sequence of this region, reported in Fig.2, contains an ORF at position 665 to 2071 (ORF469), the deduced amino acid sequence of which is highly homologous to that of <u>E.coli</u>, <u>Anabaena</u> and <u>A.brasilense</u> GSs (Fig.3), and poorly homologous to <u>B.japonicum</u> GSII. Expression of ORF469 in minicells (Fig.4) reveals the presence of a protein, 58,000 in molecular weight, that specifically reacts with an antiserum against <u>R.leguminosarum</u> GSI. The molecular weight of GSI is 60,000 (A.Fuggi and R.D., in preparation), while the deduced molecular weight of the protein encoded by the ORF at position 665-2071 is 52,040. We conclude that ORF469 corresponds to <u>glnA</u>, the structural gene for CSI.

R.leguminosarum GSI can be adenylylated in vivo, at least in K.pneumoniae, because pure GSI, or GSI activity in crude extracts of UNF1827(pMG10), is partially adenylylated (A.Fuggi and M.G., unpublished results). Therefore, R.leguminosarum, E.coli and A.brasilense GSs can be adenylylated, while Anabaena GS cannot (36). We compared the 5 amino acid stretches of the C-domain (34) in the four protein sequences of Fig.3 and found Anabaena-specific differences only in the 5th stretch (residues 392 to 400, containing the Tyr target of adenylylation). An analysis of the secondary structure of the four proteins in this region, using the Micro-Genie computer program (30) shows that these are all different and therefore the correlation between lack of adenylylation and secondary structure around the target Tyr, previously proposed (32), is not confirmed. Anabaena-specific differences by this type of analysis were only found in the N-domain (34). If this is related to the lack of adenylylation should be confirmed by independent evidence.

The ORF on the opposite strand of <u>glnA</u> does not show a band in minicell experiments and we found no significant homology of its deduced amino acid sequence to other sequences in the data bank. An extended ORF is present also on the opposite strand of <u>E.coli</u> and <u>A.brasilense glnA</u> DNA, but not in the case of <u>Anabaena</u>. As shown under Results the reading frame of these three ORFs is coincident, but with opposite polarity, with that of GS. Although this is certainly striking and suggests a functional or evolutionary role (37) we did not investigate its significance further on.

Plasmid pMGl0 codes not only for GSI, but also for a protein expressed in minicells with a molecular weight from 12,000 to 14,000. This molecular weight is in good agreement with that (10,469) of the deduced amino acid sequence of ORFIII (nucleotides 251 to 584). As indicated under Results the product of ORFIII might be equivalent to the <u>E.coli</u> P_{II} protein. If so, while <u>glnB</u>, the supposed structural gene for P_{II}, is unlinked to <u>glnA</u> in <u>K.pneumoniae</u> (38) and in <u>E.coli</u> (10) it might be contiguous to it in <u>R.leguminosarum</u>. A <u>glnB</u>-like gene appears to be contiguous to <u>glnA</u> also in <u>A.brasilense</u> and <u>B.japonicum</u>. In fact, 350 bp upstream of <u>A.brasilense glnA</u> (33) the published sequence starts with 75 bp coding for 25 residues, 20 of which are identical to residues at the -COOH end of the ORFI11 product. In the case of <u>B.japonicum</u> (5), 336 bp upstream of <u>glnA</u> the published sequence starts with 180 bp coding for 60 residues, 45 of which are identical to residues at the -COOH end of the ORFI11 product.

We previously reported evidence (39) that DNA in the region coding for ORF111 is responsible for inhibition of the growth observed in the presence of nitrate. In fact, when UNF1827 carrying either p7D9 or pMG10 is grown in nitrate, there was no complementation of the Gln phenotype. Although still obscure, this phenomenon suggests that GSI might be repressed or adenylylated when <u>K.pneumoniae</u> containing pMG10 is grown in nitrate. UNF1827(p10-1) grows in nitrate indicating that the DNA deleted in pMG10 to generate p10-1 is required for the nitrate effect. This observation suggests a regulatory role, or an interference with a <u>K.pneumoniae</u> regulatory circuit, caused by the product of ORF111.

In Fig. 2 we show a sequence with good homology to an <u>ntrA</u>-dependent promoter at position 174-190. It is tempting to conclude that <u>glnA</u> is transcribed from this promoter in pMG10, since <u>R.leguminosarum</u> DNA in this plasmid is inserted with opposite orientation to the Tet promoter of the pACYC184 vector. However, the putative <u>ntrA</u>-dependent promoter is deleted in the p10-1 (Gln⁺) clone (Fig.la) and therefore a promoter more proximal to <u>glnA</u> might exist. Indeed, at position 564-597 we find a sequence homologous to the promoter sequence of <u>B.japonicum glnA</u> (5). Experiments are in progress to identify transcription initiation(s) in this region.

Aknowledgements

We thank dr. M. Merrick for bringing to our attention the similarity between the sequence of ORF111 and that of a peptide of protein P_{II} ; F. Graziani, C. Malva and their collaborators for discussion and advice; A. Downie for supplying plasmid pl085; C. Moscatelli for help with the use of the computer; A. Covarrubias and C.Elmerich for unpublished information. We also thank C. Migliaccio for technical assistance. S.C.R. is a fellow of Pierrel SpA. Research work supported by CNR, Italy. Special grant I.P.R.A. Subproject 1, Paper n°1187. Also partially supported by Pierrel SpA.

* To whom correspondence should be addressed.

REFERENCES

- Bergersen, F.J. & Turner, G.L. (1967). Biochim. Biophys. Acta. 141,507-515.
- Darrow, R.A., Crist, D., Evans, W.R., Jones, B.L., Keister, D.L. & Knotts, R.R (1981). In Gibson, A.H.& Newton, E. (eds). Current perspective in Nitrogen fixation, Australian Academy of Science pp 182-185.
- Somerville, J.E. & Kahn, M.L. (1983). J. Bacteriol. 156,168-176.
- 4. Donald, R.G.K. & Ludwig, R.A. (1984). J. Bacteriol. 158, 1144-1151.
- Carlson, T.A., Guerinot, M.L. & Chelm, B.K. (1985). J. Bacteriol. 162,698-703.
- 6. Carlson, T.A. & Chelm, B.K. (1986). Nature 322, 568-570.
- Stadtman,E.R., Shapiro,B.M., Kingdom,H.S., Woolfolk,C.A. & Hubbard,J.S. (1986). Advances in Enzyme Regulation 6,257-289.
- 8. Reitzer, F.J. & Magasanik, B (1986). Cell, 45,785-792.
- 9. Dixon, R.A.D. (1984). Nucl. Acid Res. 12,7811-7830.
- 10. Magasanik, B (1982). Ann. Rev. Gen. 16,135-168.
- 11. Filser, M.K.K., Moscatelli, C., Lamberti, A., Vincze, E., Guida, M.,
- Salzano, G. & Iaccarino, M. (1986). J. General Microbiol. 132,2561-2569.
- 12. Rhee,S.G., Chock,P.B. & Stadtman,E. (1985). In Freedman,R.B. & Hawkins,H.C.(ed), The enzymology of post-traslational modification of proteins, Academ. Press Vol II pp. 273-397.
- Ninfa,A.J. & Magasanik,B. (1986). Proc. Natl. Acad. Sci.(USA) 83,5809-5913.
- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). In Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, N.Y.
- 15. Vieira, J. & Messing, J. (1982). Gene, 19,259-268.
- 16. Dougan, G. & Sherratt, D. (1977). Mol. Gener. Genet. 151, 151-160.
- Streicher, S.L., Shanmugam, K.T., Ausubel, F., Morandi, C., Goldberg, R.B. (1974). J. Bacteriol. 120,815-821.
- Espin,G., Alvarez-Morales,A., Cannon,F., Dixon,R. & Merrick,M. (1982). Mol. Gen. Genet. 186,518-524.
- Hooykaas, P.J.J., Klapwijk, P.M., Nuti, M.P., Schilperpoort, R.A. & Roersch, A. (1977). J. General Microbiol. 98,477-484.
- Hooykaas, P.J.J., Snijdewint, F.G. & Schilperpoort, R.A. (1982). Plasmid 8,73-82.
- Frey, J., Bagdasarian, M., Feiss, D., Franklin, F.C. & Deshusses, J. (1983). Gene 24,288-308.
- 22. Chang, A.C.Y. & Cohen, S.N. (1978). J. Bacteriol. 134,1-41.
- 23. Vogel, H.J. & Bonner, D.M. (1956). J. Biol. Chem. 218, 97-106.
- Priefer, U. (1984). In A. Pühler & K.N. Timmis (eds), Adv. Mol. Genet. pp 12-25.
- 25. Davis, R.L. & Davidson, N. (1984). Mol. Cell. Biol. 4,358-367.
- Cannon,F.C. (1980). In F.J.Bergersen(ed), Methods for Evaluating Biological Nitrogen Fixation, pp 367-413.
- 27. Maxam, A. & Gilbert, W. (1977). Proc. Natl. Acad. Sci. (USA) 74,560-564.
- Roozen,K.J. Fenwick,R.G. & Curtiss, R.III (1971). J.Bacteriol. 107,21-33.
- Feunten, J., Kress, M., Gardes, M. & Monier, R. (1978) Proc. Natl. Acad. Sci.(USA) 75,4455-4459.
- 30. Queen, C. & Korn, L.J. (1984). Nucl. Acid Res. 12,581-599.
- Adler, S.P., Purich, D. & Stadtman, E.R. (1973). J. Biol. Chem. 250,6264-6272
- 32. Colombo, G. & Villafranca, J. (1986). J. Biol. Chem. 261,10587-10591.
- 33. Bozouklian, H. & Elmerich, C. (1986). Biochimie 68, 1181-1187.

- Almassy, R.J., Janson, C.A., Hamlin, R., Xuong, N-H. & Eisenberg, D. (1986). Nature 33, 304-310.
- 35. Cullimore, J.V. & Miflin, B. (1984). J. Exper. Bot. 35, 581-587.
- 36. Tumer, N.E., Robinson, S.J. & Haselkorn, R. (1983). Nature 306, 337-342.
- 37. Tramontano, A., Scarlato, V., Barni, N., Cipollaro, M., Franzé, A.,
- Macchiato, M.F., & Cascino, A. (1984). Nucl. Acid Res. 12,5049-5059. 38. Leonardo, J.M. & Goldberg, R.B. (1980). J. Bacteriol. 142,99-110.
- Colonna-Romano, S., Defez, R. Guida, M;, Iaccarino, M., Lamberti, A., Riccio, A., Fuggi, A., Arnold, W., Priefer, U., Puhler, A., (1986).
- Proceedings of the III International Symposium on the Molecular Genetics of Plant Microbe Interaction, Montreal, Canada.
- Downie, J.A., Ma, Q.S., Knight, C.D., Hombrecher, G. & Johnston, A.W. (1983). EMBO Journal 2,947-952.