### 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.Arnold<sup>1</sup>, U.Priefer<sup>1</sup> and A.Pühler<sup>1</sup>

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

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

#### ABSTRACT

Rhizobium leguminosarum, biovar viceae, strain RCC1001 contains two plutamine synthetase activities, GSI and GSII, We report here the glutamine synthetase activities, GSI and GSII. We report here the identification of glnA, the structural gene for GSI. A 2 kb fragment of DNA was shown to complement the Gln phenotype of Klebsiella pneumoniae glnA mutant strains. DNA sequence analysis revealed an open reading frame (ORF) of 469 codons specifying a polypeptide of 52,040 daltons. Its deduced amlno acid sequence was found to be highly homologous to other glutamine synthetase sequences. This ORF was expressed in Escherichia coli minicells and the corresponding polypeptide reacted with an antiserum raised against GSI. Upstream of glnA we found an ORF of 111 codons (0RF111) preceded by the consensus sequence for an ntrA-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  $P_{xx}$  protein of E.coli and plays a similar role in regulation of nitrogen metabolism.

#### INTRODUCTION

Rhizobium bacteria use ammonia for growth in the free-living state, but in the Rhizobium-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 Rhizoblaceae 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 glnA gene in Enterobacterlaceae is very complex. A coordinated nitrogen control system similar to that of enteric bacteria (10) has not been described in Rhizobium spp..

We started a study of the GS activities of R. leguminosarum biovar viceae, 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 glnA, 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 glnA coding for a protein which appears to play a regulatory role, probably equivalent to that of the  $P_{TT}$  protein, which is involved in the regulation of GS adenylylation and glnA expression in Escherichia coll (12,13).

#### MATERIALS AND METHODS

## Strains, plasmids and media

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

Strains used were E.coli HB101 (14), JM83 (15), DS998 (16); Klebslella pneumoniae wild type (17), UNF1827 and UNF1838 (18); R.leguminosarum strain LPR11O5, a rifampicin resistant derivative of RCC1OO1 (19); Agrobacterium tumefaciens strain LBA2715 containing the R.leguminosarum symbiotic plasmid (pSym) (20).

Plasmids used were: pMMB34 (21); pSVB20, pSVB23, pSVB24, pSVB25 (pUC8 derivatives, Ap<sup>r</sup>; W.A. et al., in preparation); pACYC184 (22); p7D9 (11),  $Gln<sup>+</sup>$ , Km<sup>r</sup>, containing 27 kb of <u>R.leguminosarum</u> DNA inserted in the cosmid vector pMMB34; pMG10 (11),  $G\ln^+$ ,  $Cm^T$ , containing 6.5 kb of R. leguminosarum DNA and 3. A kb of pMMB34 DNA inserted in the vector pACYC184.

Media used were: TY (14); PA (14) supplemented with O,1Z 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 glnA region and the DNA left In different Bal31 deletions. The black bar indicate the location of the EcoRI** fragment cross-hybridizing to K.pneumoniae glnA (11). The dotted lines **indicate DNA (about 3 kb) omitted from the figure. Only the glnA region-proximal end of the deletions Is indicated in the figure. The boundary of plO-1 and p30-ll was confirmed by DNA sequencing (see Fig.2). The boundary of pE-9 is between the EcoRI and PstI restriction sites (striped bar); that of pD-8 is between the Smal and BamHI 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 ntrA-aependent promoter. A black** bar indicates the Sall fragment used as a probe in the hybridization **experiments of Fig.5.**

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

TODACATCTO 10 TATTORCACT TOCCAMATA OCTEACTETT TODETTIDEE ATCTOTETTE GATOORAGAC AATUTETTEC CACCIUTECO OCCAOCAMA COACATAGAA ITTOCCOCCAA ACOCOCCUTT 130 TTCHCHCH TOUNCETETT TTTCCHCH CATCOATTE HARTERING CATATORICA TOATATOOC COMOCOCAN AATTOTECE ATACHOCHC MODERNIC RATTTEETC 230 .<br>ATC AAA AAC ATC GAA GOD ATC ATT AAC GOT TTC AAC CTC GAAC GAA GTC AGO AGO COT TCA GGA GTC GOT CTC CAG GOT ATC AGO GTC AGO<br>met lyn lyn ile giw als ile ile lyn pro pho lyn lem amp giw wal arg aer pro mer gly val gly lem gin e<br>Awd conf free door corr two aad door due adda afte had door deal door daa had dre dre dan rre choi door aad dr<br>Uys giy phas giy seg gis lys giy kis the gis lees tyr seg giy sis gis tyr wal val say phas lees pre lys val l CAA GCC<br>Els als .<br>ENG STT CTA CTG GOC CAC CAC ANT DOG CAN GOS GTC ATT CAN GOG ATC COC ANG GOC GOG CAS ACC GOC COC ATC GOC CAC DEA ANG ATC 370<br>Els val val les als app els aps als els als val ils els als ils are lys als als els thr att ils .<br>TOC AAC GTC GAA GAG GTT ATC GOC ATC GGC ACC GGC GAG AGC GGC ATT GAT GCC ATC TGA.<br>Per am val giw giw yol tie erg tie erg thy giv giw thr giv tie am als tie med. COACTTROOC AATOECOCAA ACTOOCTTAC 616 ATC<u>ONACH</u>A ACTACTTA ATC CCC ACC OCA ACC CAA ATT CTC AAC CAO ATC AAC CAO AND CAC TTA AAC TTC 718<br>set als thi als ear als il ile ile all pa an and an app will be a **CETCATOGOC** CACCTACACC OTCATCOCAA THE ATC TTC GAC GOOD TTC TOG ATC GOOD COD TOG AAG GOOD ATC AAG GAG TOG GAG ATG GTC GTG ATG GOOD GAA AGG GTG GAT ATG BYB<br>Yal met plus ago aly met met ils aly aly ttp lys als ils ago alw met sam met wal law met yrs ago the a .<br>And led TTC TTC OCA CAD TOG ACK ATH OTC ATO DE TOG GAD ATO CTC GAT COS OTC TOG GOD GAD OCC TAT AAC OGD GAT GOD<br>And bre box who als als ser the met ballie ball one and the lew ewe see ball ser gly gle als tyr am argused p .<br>ACC COC AMG ANG COC GAA GOC TAC CTC AMG GCA TCC GOC ATC GOC GAT ACG GTT TTC GTC GOC GAG GOC GAA TTC TTC GTC TTC GAC 1078<br>Cht als lys lys als gle sis tyr lew lys als ser gly ile gly say the val phe wal gly pre gle als gle .<br>GAC GTC ANG TAC ANG GCC GAT CCT TAC AAT ACC GGC TTC ANG GTC GAT TOO ACC GAA GTG COG TOG AAC GAC ANG GAT TAC GAC ACC 1168<br>Ang wal lys tyt lys als ago pto tyt and the gly pho lys low ago out the glu low pto set ago ago the .<br>On the tre cod car one cod one ofte and cod cod the tre cod city one cod give and and cod car and cot too can<br>Sily and low sit his see from are vall its sit sit tre play for vall from the vall and made all site and site CAC 1348 CAA CTC COT AT ANG TTC ANG ATC CAG ATC TAC ANA TAC GTC CTC CAG CTC GOC ANG COC TAT COC ANG ACC COC ACC TTC ATC<br>Elw lew ely lie lys who lys met six lie tyr lys tyr yel wil his six yai als anm als tyr six the six the met ste AAG COD ATC TIT COD CAR AND COD TOG ONE ONE CAR COLORED TOTAL TO DAMO OOD AND COD AND TIT OCD OOD AND EAT 1332<br>Iya pro ile moo aly ame and aly may aly met his val his aim eer ile tre ire aly aly to to co the she als the tr OOC OUT CT? TOO GAG AGC TGC CTG TIT TAT ATC GGC GGC ATC ATC AAG CAT GOC AAG OOC ATC AAC OOT TTC AGC AAT COG TOG AGG AAG 1618<br>sis gly lee ser gle ser tys lee ples tyr ile gly gly ile ile lys his gla lys gla ile see gla ples .<br>TOC TAC AME COT TTC OTC CO2 GCT TAC GAA BCA CCT OTG CTG CTG OCC TAT TO2 GCC CGC AAC CGC TGG GCC TO2 FDC DGC ATT CGC<br>ser tyr lys arg low wal pro gly tyr glu ala pro val low low als tyr mer als arg som arg ser als met cys  $TIC$  1708 OGC TOC AAC COD AAC GOD AAA GED GTD GAG GTD COD TTD COD GAT COD AAD OO AAT COD TAT COD TAT CTD GOD TTD GOD OOD ATG<br>Ely met aan ers lys als lys att val els wal att mee ers aan ers the als aan ers tyt low als phe als als met  $\frac{1}{2}$  and  $\frac{1}{2}$  and  $\frac{1}{2}$ ood did dad ood ahd aad aad ahd dat doo ood aad ood ahu dad aan dah did tid dan die oos oos aan daa tite aad a<br>giy lee aan giy ila iya aes iya ile his pro giy iya ala set aan iya aan lee tyr aan lee pro yro iya giw lee iya  $\frac{GM}{4H}$  1979 COD ACC GTC TOO GOD AGO TTG COD CAA GOA OT CAA AGO CTC GAC AAG CAO QOD AAG TTC CTC AGO GOD GOD GOD GTC TTC GAC GAC<br>pro the wel gra giy out low ang gim ain low giw out low may lys say ang lys pho low the sim giy giy wel pho CAO ATC GAT GOD TTC ATC GAO CTO AAG ATG GUT GAG CTO ATG GUT TTC GAA ATG AGO GOD GAT GOD GTC GAA TAC GAC ATG TAC TAT<br>Ein ile aam ala phe ile sim low lys met als gig wel met arg phe gig met the pro his pro val gin tyr anp me TCC 2068 

TODACTOCTC CATAATTTTE TOCTTAAATC CAT

Figure 2. DNA sequence and deduced amino acid sequence of ORFIII and ORF469 from the XhoI site to the ClaI site of Fig.1. A consensus sequence for an ntrA-dependent promoter is boxed, while putative ribosome binding sites are  $\overline{\text{underline}}$ . 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<sup>-</sup> phenotype of K.pneumoniae strains UNF1827 (glnA) and UNF1838 ( $\Delta$  glnA-ntrC). 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 K.pneumoniae glnA  $(11)$ .

We generated deletions of pMG10 by digestion at the single XhoI 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

ATT DRESS TO CARLO CERTIFICATIVE SPHATER DE LE CONFERENCE DE L'ANGERIA DE L'ANG **K.1.** THE VI V C D I LIDFIV SIG HIAT RE D P R GILARE A EIAT LIK A SIG I G D T V PIV E.<br>E.C. TIL RILLING D I LIGIT I LIQ CITORE D R GILARE A EIAT LIK S TIG ITAD T VIE P<br>Ana. [2] L S I V C D I V E PIS TIG O PID R GILARE A  $\overline{\phantom{a}}$ R.1. GPEAEFFVFDDVEJKEADPTHTGFKLOSITELLEJSHOLTDYEJTG--<br>E.c. GPEDEFFLFDDULLFGSSISGSHVAIDDIECAWMSSITOTECICI--<br>A.b. GPEAEFFVFDDVEJKESSISGSHVAIDDIECAWMSSITOTECICI--<br>Ana. GPEAEFFTIEDSARFAQHANECYTFLLDSVEJGAWMSGEBGPADS-- $\overline{\circ\circ\circ}$ R.1. NLGHRPRVKCCVPPVPPVDSAQDNESEHLJTVLSEHCVVVEKHE<br>E.c. NKGHRPAVKCCYPPVPPVDSAQDIRSEHCLVMEQHCLVVVEKHE<br>A.b. NLGHRPCVKCCYPPVPPVDSAQDIRSEHCLVMEQHCLVVVEKHB<br>Ana. NLGAVKPRFKSCTPPVSPJDSPGDIRAEHLLTNAKTCCVPFLKKBB R.1. HEVAAA - QHELGIKFDTLVRHADKHOITKYVVHOVAHAYGKTA<br>E.C. HEVALAGONELGIKFDTLVRHADE IQITKYVVHHVAHEFGKTA<br>A.b. HEVAAS - QHELGIKFDTLVRITGDHHOITKYVVHHVAHEFGKTA<br>Ana. HEVATGGQGELGFREGKLIEAADULLHIYKYVIKHVAHATYGKTYV R.1. TFHPKPIFGDNGSGHHVHQSIWKGGKPIFAGDETAGLSESCLP<br>E.c. TFHPKPHFGDNGSGHHCHHSLSKHGGHLFAGDETAGLSEQALT<br>A.b. TFHPKPIFGDNGSGHHCHQSIWKEGGFLFAGHQTADLSELALT<br>Ana. TFHPKPIFGDNGSGHHCHQSIWKDOKPLFAGDQTAGLSEHGLY R. L. YIGGIIKHAKAINAFTHPSTHSYKRLVPGYEAPVLLAYSARHE<br>E.C. YIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVLLAYSARHE<br>A.B. YIGGIIKHAKALNAFTNPTTNSYKRLVPGYEAPVLLAYSARHE<br>Ana. YIGGILIKHAFTALITIAITTNPSTNSYKRLVPGYEAPVILLAYSOGINE R.1 SASCRIP - FGSNPRAKRVEVRTPDFTANPYLATAANLHAGLDG<br>E.C. SAS[IRIP - VVSJS]PRAFIRIEVRTPDFTANPYLATAANLHAGLDG<br>A.b. SASCRIP - YVASJPRGKRVEVRTPDFSANPYLATAANLHAGLDG<br>Ana. SASTIRIPLSGTLATAARIHETRICPDTATSMPYLATAANLGAGTIDG R.1. IKBK1HPGKANDK DLTDLPPKELKKIPTVCGSLEGALESLDKD.<br>A.b. IQNKIHPGEANDK HLTDLPFAEELAKK PQVGGSLEGALESLDKD.<br>A.b. IQNKIHPGEANDK HLTDLPFAEELAKK PTVGGSLEGALESLDKAND.<br>Ana. IKNKIHPGKPLDK HTYELSEDEELAKVESTPGSLELAKLUSLUKAND  $0.0000$  $R.1.$   $S \uparrow A \uparrow B.$ <br>  $R.c.$   $S \uparrow V \uparrow A.$ <br>  $A.b.$   $S \uparrow V \uparrow A$ 

Ana

Figure 3. Homology of the deduced R. leguminosarum GSI sequence to other GS sequences. R.1.= R.leguminosarum; E.c.= E.coli; A.b.= A.brasilense; Ana. - Anabaena. Sequences assumed to correspond to the N-domain and C-domain  $(33)$  are indicated by  $(*)$  and  $(o)$  respectively.

which were  $G\ln \frac{1}{\pi}$  and 17 were  $G\ln \frac{1}{\pi}$ . Restriction analysis showed that the Xhol 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  $Gln<sup>+</sup>(p10-1)$  and  $Gln<sup>-</sup>(p30-11)$  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 pl0-1 to generate a set of new deletions by digestion at the single BstEII site and treatment with Bal31 nuclease. Restriction analysis of 40 clones carrying deletions showed that the EcoRI site to the right of ClaI is present in all  $Gln<sup>+</sup>$  clones and absent in all  $Gln$ clones. In Fig. la a  $Gln^+$  clone (pE-9) and a  $Gln^-$  clone (pD-8) are shown. R.leguminoaarum DNA left in pE-9 is 2 kb long and therefore about 2 kb of DNA are sufficient to suppress the Gin phenotype of strains UNF1827 and UNF1838.

# DNA sequencing of the glnA region.

The sequence of R. leguminosarum DNA from the Xhol site to the Clal site indicated in Fig.l is presented in Fig.2. A computer search for open reading frames longer than 100 codons resulted in the three ORFs shown in Fig.lb. A sequence with good homology to an ntrA-dependent promoter (9) is found at position 174-190 followed by a putative rlbosome 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 termination. On the opposite strand another ORF of 517 codons is present (see Fig.lb) starting with an Arg codon at position 2215 and ending at position 662.

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

We compared the deduced aminoacid sequence of the ORF111 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 ORF111 are strikingly similar to the sequence of a peptide of the  $P_{TT}$  protein of E.coll (12):

> P<sub>TT</sub>: -Gly-Ala-Glu-Tyr-Met-Val-Asp-Phe-Leu-Pro-Lys-ORF111: -Gly-Ala-Glu-Tyr-Val-Val-Asp-Phe-Leu-Pro-Lys-



Figure 4. SDS-PAGE of minicells extracts containing different plasmids. Panel A: S-T-bob theorem and the minicelis extracts containing different plasmins. Panel<br>A: S-methionine labeling of minicells containing: (a): vector pACYC184;<br>(b): pMG10; (c): p10-1; (d): p30-11; (e): pE-9; (f): pD-8. Panel B<br> (c): pMG10. The 58, 33 and 12 markers are explained in the text. Lanes (a) show the CAT band of pACYC184 present also in the other lanes. Molecular weights of insert-specific bands were interpolated from the molecular weight<br>of commercially available <sup>1</sup>C-standards. The experiments shown were performed with 10% (Panel A) and 20% (Panel B) polvacrylamide.

Since the complete amino acid sequence of the  $P_{TT}$  protein of E.coli is not published, we could not compare other portions of the two proteins. The molecular weight of the  $P_{TT}$  protein of E.coli 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  $P_{TT}$  protein of E.coli (31) and of the ORFIII product shows a striking similarity. We conclude that the product of ORFIII might be equivalent to the E.coli P<sub>TI</sub> 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 Anabaena sp. strain 7120 GS. We introduced in the data bank the deduced protein sequences of E.coli (32) and Azospirillum brasilense sp.7 (33) GSs and that of Bradyrhizobium japonicum GSII (6). Homology of ORF469 is 56%, 63% and 68% with Anabaena, E.coli and A.brasilense 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 R. leguminosarum GSI. Homology with B.japonicum 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 amlno acid sequences of Anabaena, E.coll and A.brasilense GSs are 53,265, 51,814 and 51,917 respectively.

A recent paper (34) reports the crystallographic structure of Salmonella typhymurlum 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 Anabaena, E.coli and A.brasilense GSs give values of homology to R.legumlnosarum GSI slightly lower than those found for the complete protein (49Z, 58Z and 55Z respectively), while C-domain sequences show higher values (72%, 69% and 82% respectively). A more detailed analysis shows that 4 of the C-domain stretches are highly homologous while the  $5<sup>th</sup>$ one (residues 392-400) is very conserved in R.leguminosarum, E.coll and A.brasilense GSs and different in Anabaena 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-terninal 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 E.coli and A.brasilense sequences present an extended ORF on the opposite strand, which is absent in the Anabaena 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 R. leguminosarum ORF to other proteins.

## Expression of pMGlO 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 pMGlO and the deletion derivatives described above were introduced into strain DS998 by transformation and minicells were purified. After labeling with S-methionine and SDS-PAGE, we observed, in addition to vector-coded bands, a band specific of the insert of pMGlO (Fig.4A). This band, 58,000 in molecular weight, is also present in the Gln<sup>+</sup> deletion derivatives pl0-l and pE-9, is absent in the Gln<sup>n</sup> 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);<br>glnA DNA is a Sall fragment of pMG10 indicated in Fig.1b 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.1b and that this protein is necessary for the Gln<sup>+</sup> 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 35<sub>S-methionine</sub> 58,000 molecular weight produced  $In$ protein 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 glnA.

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.AB). These experiments demonstrate that 0RF111 and glnA 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 Sail fragment of pMGlO, shown in Fig.lb, was used as a probe and hybridized to EcoRI digests of R. leguminosarum total and pSym DNA. We observed hybridization only to total DNA, as shown in Fig.5. If, instead, the probe originates from R.legumlnosarum pSym DNA, there Is hybridization to both DNAs. Therefore the insert of pMGlO is not located on pSym. The size of the hybridizing fragment is that (1.3 kb) of the EcoRI fragment of pMGlO shown in Fig.la. Since total DNA was digested with EcoRI, this experiment shows colinearity between the EcoRI restriction sites of the insert of pMGlO and those of total DNA.

#### DISCUSSION

Free-living Rhizobium 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 R.legumlnosaram glnA is important to understand the physiology of ammonia utilization in the symbiosis.

We report in this paper the sequence of the glnA gene and of a contiguous gene which, at the amino acid level, shows homology to the  $P_{TT}$ protein of E.coll (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<sup>-</sup> 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 K.pneumoniae glnA DNA (11) as indicate in Fig.la. 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 E.coli, Anabaena and A.brasllense GSs (Fig.3), and poorly homologous to B.japonlcum 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 R.leguminosarum 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 0RF at position 665-2071 is 52,040. We conclude that ORF469 corresponds to glnA, the structural gene for GSI.

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  $5<sup>th</sup>$  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 glnA 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 E.coli and A.brasilense glnA DNA, but not in the case of Anabaena. 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.

Flasmid pMGlO 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 0RF111 (nucleotides 251 to 584). As indicated under Results the product of ORF111 might be equivalent to the  $E.\text{coll}$   $P_{1T}$  protein. If so, while  $glnB$ , the supposed structural gene for  $P_{TT}$ , is unlinked to  $glnA$  in K.pneumoniae (38) and in E.coli (10) it might be contiguous to it in R. leguminosarum. A glnB-llke gene appears to be contiguous to glnA also in

**A.brasllenee and B.japonlcum. In fact, 350 bp upstream of A.brasllense glnA (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 0RF111 product. In the case of B.japonicum (5), 336 bp upstream of glnA the published sequence starts with 180 bp coding for 60 residues, 45 of which are identical to residues at the -COOH end of the 0RF111 product.**

**We previously reported evidence (39) that DNA in the region coding for 0RF111 is responsible for inhibition of the growth observed In the presence of nitrate. In fact, when ONF1827 carrying either p7D9 or pMGlO is grown in nitrate, there was no complementation of the Gin phenotype. Although still obscure, this phenomenon suggests that GSI might be repressed or adenylylated when K.pneumoniae containing pMGlO is grown in nitrate. UNF1827(plO-l) grows in nitrate indicating that the DNA deleted in pMGlO to generate plO-1 is required for the nitrate effect. This observation suggests a regulatory role, or an interference with a K.pneumoniae regulatory circuit, caused by the product of ORF111.**

**In Fig. 2 we show a sequence with good homology to an ntrA-dependent promoter at position 174-190. It is tempting to conclude that glnA is transcribed from this promoter in pMGlO, since R. leguminosarum DNA in this plasmid is inserted with opposite orientation to the Tet promoter of the pACYC184 vector. However, the putative ntrA-dependent promoter is deleted in the pl0-l (Gin ) clone (Fig.la) and therefore a promoter more proximal to glnA might exist. Indeed, at position 564-597 we find a sequence homologous to the promoter sequence of B.japonlcum glnA (5). Experiments are in progress to identify transcription initiation(s) in this region.**

## **Aknowledgement s**

**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<sub>TT</sub>; F. Graziani, **C. Malva and their collaborators for discussion and advice; A. Downie for supplying plasmld plO85; C. Moscatelli for help with the use of the computer; A. Covarrubias and C.Elmerich for unpublished information. He 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

- 1. Bergersen, F.J. & Turner,G.L. (1967). Biochim. Biophys. Acta. 141,507-515.
- 2. Darrov.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.
- 
- 3. 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.<br>5. Carlson,T.A., Guerinot,M.L. & Chelm,B.K. (1985). J. Bacteriol.
- 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.<br>7. Stadtman.R.R., Shapiro.B.M., Kingdom.H.S., Woolfolk.C.
- 7. Stadtman.E.R., Shapiro,B.M., Kingdom,H.S., Hoolfolk.C.A. & Hubbard.J.S. (1986). Advances in Enzyme Regulation 6,257-289.
- 8. Reitzer, F.J. & Magasanik, B (1986). Cell, 45, 785-792.<br>9. Dixon, R.A.D. (1984). Nucl. Acid Res. 12.7811-7830.
- 9. Dixon, R.A.D. (1984). Nucl. Acid Res. 12,7811-7830.<br>10. Magasanik. B (1982). Ann. Rev. Gen. 16.135-168.
- 10. Magasanik.B (1982). Ann. Rev. Gen. 16,135-168.
- 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.
- 13. Ninfa.A.J. & Magasanik.B. (1986). Proc. Natl. Acad. Sci.(USA) 83,5809-5913.
- 14. 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.<br>17. Streicher, S.L., Shanmugam, K.T., Ausubel, F., Morandi.C., Goldber
- 17. Streicher.S.L. , Shanmugam,K.T. , Ausubel.F., Morandi.C, Goldberg,R.B. (1974). J. Bacteriol. 120,815-821.
- 18. Espin.G., Alvarez-Morales,A., Cannon,F., Dixon.R. & Merrick.M. (1982). Mol. Gen. Genet. 186,518-524.
- 19. Hooykaas.P.J.J., Klapwijk.P.M., Nuti.M.P., Schilperpoort,R.A. & Roersch.A. (1977). J. General Microbiol. 98,477-484.
- 20. Hooykaas.P.J.J., Snijdewint,F.G. & Schilperpoort,R.A. (1982). Plasmid 8,73-82.
- 21. 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.
- 24. Priefer,0.(1984). In A.Pilhler & K.N.Timmis(eds) , Adv. Mol. Genet, pp 12-25.
- 25. Davis,R.L. & Davidson,N. (1984). Mol. Cell. Biol. 4,358-367.
- 26. 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.<br>28. Roozen, K.J. Fenwick, R.G. & Curtiss, R.III (1971). J.Bacteriol.
- 28. Roozen.K.J. Fenwick.R.G. & Curtiss.R.Ill (1971). J.Bacteriol. 107,21-33.
- 29. 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.
- 31. 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.
- 34. Almassy.R.J., Janson.C.A., Hamlin.R., Xuong,N-H. & Elsenberg.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.,
- 38. Macchlato.M.F., & Casclno.A. (1984). Nucl. Acid Res. 12,5049-5059. Leonardo,J.M. & Goldberg,R.B. (1980). J. Bacterlol. 142,99-110.
- 39. Colonna-Romano, S., Defez, R. Guida, M;, Iaccarino, M., Lamberti, A., Rlccio.A., Fuggl.A., Arnold,W., Prlefer.U., Puhler.A., (1986). Proceedings of the III International Symposium on the Molecular Genetics of Plant Microbe Interaction, Montreal, Canada.
- 40. Downie.J.A., Ma.Q.S., Knight,CD., Hombrecher,G. & Johnston,A.W. (1983). EMBO Journal 2,947-952.