Volume 13 Number 16 1985 Nucleic Acids Research

Analysis of the 5′ regulatory region of the gene for δ-aminolevulinic acid synthetase of *Rhizobium meliloti*

Sally A.Leong, Peter H.Williams[1] and Gary S.Ditta[2]

USDA-ARS, Department of Plant Pathology, University of Wisconsin, Madison, WI 53706, USA, [1]Department of Genetics, University of Leicester, Leicester, UK, and [2]Department of Biology B-022, University of California, La Jolla, CA 92093, USA

Received 1 April 1985; Revised and Accepted 23 July 1985

ABSTRACT

Transcriptional regulation of the δ-aminolevulinic acid synthetase gene of Rhizobium meliloti was investigated under conditions of normal vegetative growth and during symbiosis with the legume host alfalfa. S1 nuclease mapping and DNA sequence analysis indicated that transcription originates from two sites separated by 238 base pairs. A deletion analysis of the putative promoter regions P1 and P2, corresponding to the proximal and distal RNA start sites, was carried out with Bal-31 nuclease. Promoter function was monitored as β-galactosidase activity after fusing the deletions to lac Z and introducing them into Rhizobium on a broad host range plasmid. The data obtained suggest that both regions function equivalently as promoters. The DNA sequences of P1 and P2 show considerable homology in the region between -35 and the start of transcription. Both contain a -35 region that is analogous to the consensus E. coli promoter sequence, while the -10 region is dissimilar. No resemblance was found between either P1 or P2 and the promoter regions of genes under general nitrogen control.

INTRODUCTION

A major soluble protein component of nitrogen-fixing nodules is leghemoglobin, a hemoprotein that is thought to regulate oxygen tension in the nodule and thus provide sufficient oxygen to the bacteroids for active respiration (1). While apoleghemoglobin is clearly a plant gene product (2), the origin of the heme moiety is less certain. Several studies have suggested that leghemoglobin heme is a bacterial product. We have previously reported the identification and partial characterization of the structural gene for δ-aminolevulinic acid (δ-ALA) synthetase, the first enzyme in the heme biosynthetic pathway of R. meliloti (3). This enzyme catalyzes the condensation of glycine and succinyl CoA to form δ-ALA. We now present evidence for the occurrence of tandem transcriptional initiation sites for this gene. DNA sequence analysis of the corresponding upstream regions indicated that they contain shared and unique features when compared to the consensus promoter of E. coli and promoters for symbiotic genes of R. meliloti. No preferential gene expression from either of these sequences was

© IRL Press Limited, Oxford, England.

observed in mature alfalfa nodules. A preliminary account of this work has been presented elsewhere (4).

MATERIALS AND METHODS

Published procedures that were used without modification are referenced in the text and are not summarized here. The methods used for extraction of RNA (5), isolation of plasmid DNA (3), S1 nuclease analysis (5), and DNA sequence determination (6) have been described. Additional experimental details are provided in the figure legends.

Bacterial Strains and Plasmids. *Escherichia coli* strains were HB101 [*pro*, *leu*, *thi*, *lac*Y, Str^r, *endo* I^-, *rec*A, $r_k^- m_k^-$] and TB1, an *hsr* *hsm*$^+$ derivative of JM83 (7) constructed by Dr. T. Baldwin (personal communication). *Rhizobium meliloti* strains were 102F34 (8) and its recA::Tn5 derivative (9).

The low copy number broad host range cloning vectors pRK291, pGD499, and pGD500 have been described (10). pXLGD4 was constructed by cloning the 2.0 Kb *Bam* HI-*Hind* III fragment that encodes the 5' half of the *R. meliloti* δ-ALA synthetase gene into pGD499 (10). pGD23 was constructed by inserting the 0.9 Kb *Hind* III fragment that carries the CAT (chloramphenicol trans-acetylase) gene cartridge (11) into the *Hind* III site upstream of δ-ALA synthetase in pSL5. The latter consists of a 2.8 Kb *Sal* I fragment, that carries the δ-ALA synthetase gene of *R. meliloti* cloned into the *Sal* I site of pUC9. The purpose of the CAT gene was to permit sufficient Bal-31 digestion without destroying the integrity of the plasmid vector.

β-Galactosidase Assay.

(a) Vegetative cells: Aliquots of overnight cultures of *R. meliloti* grown in yeast mannitol broth (13) plus tetracycline were assayed at room temperature with the SDS-$CHCl_3$ technique described by Miller (14).

(b) Bacteroids: Thirty to fifty nodules from 10 different 5-week-old-plants were pooled and homogenized in 0.05 M Tris (pH 8.0) + 0.25 M mannitol at 0°C. After allowing debris to settle, the upper 3/4 of the homogenate, consisting primarily of bacteroids, was removed and aliquots were taken for analysis as described above.

RESULTS

Low Resolution S1 Nuclease Mapping. Extensive Tn5 insertion mutation analysis had previously localized the δ-ALA synthetase gene in a 1.4 Kb region of a 4.3 Kb *Bam* HI fragment of *R. meliloti* DNA (3). Low resolution S1

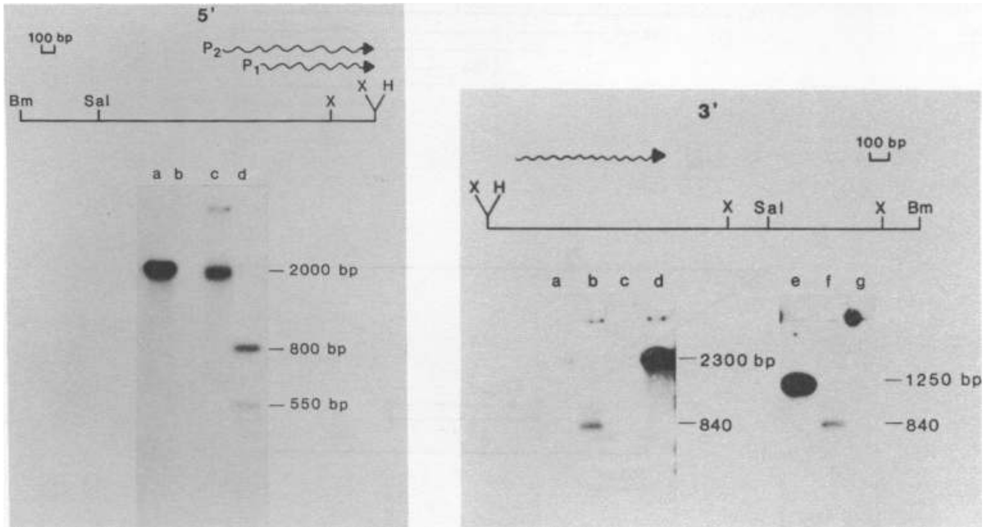

Figure 1. Low resolution S1 mapping. Hybridizations containing 10,000 cpm of the doubly end-labeled fragment were carried out in the presence or absence of 20 μg R. meliloti vegetative cell RNA in 80% formamide, 50 mM PIPES, pH 7, 0.4 M NaCl, 1 mM EDTA at 56°C (5' end) or 57°C (3' end) for 8 to 16 hours. Single-stranded nucleic acid was digested with 100 U S1 nuclease and the protected fragments were sized on 5% acrylamide-8 M Urea gels. 5' end analysis was carried out with the 2.0 Kb Bam HI-Hind III fragment labeled by treatment with T4 polynucleotide kinase and γ^{32}P-ATP. a. -RNA, -S1; b. -RNA, +S1; c. +RNA, -S1; d. +RNA, +S1. 3' end analysis was carried out using the 2.3 Kb Bam HI-Hind III fragment (a,b,c,d) or the 1.25 Kb Xho I-Xho I (e,f,g)fragment labeled by treatment with the Klenow fragment and α ^{32}P-deoxynucleoside triphosphates. a. Bam HI-Hind III fragment marker; b. +RNA, +S1, c. -RNA, +S1; d. +RNA, -S1; e. Xho I-Xho I fragment marker; f. +RNA, +S1; g. -RNA; +S1. Restriction endonuclease cleavage sites are indicated by: Bm, Bam HI; Sal, Sal I; X, Xho I, and H, Hind III.

nuclease mapping was carried out to define the 3' and 5' ends of the δ-ALA synthetase mRNA and to verify the genetic coordinates of the gene. The presence of two RNAs, differing by approximately 250 bp in length, was observed when several 5' end-labeled fragments were used as probes. Data obtained for the doubly end-labeled 2.0 Kb Bam HI-Hind III fragment carrying the 5' half of the gene are shown in Figure 1. Identical results were observed for the 1.5 Kb Hind III-Sal I fragment labeled uniquely at the Hind III end (data not shown), which implies that both RNAs have the same polarity. The relative intensity of the two protected fragments was found to vary from experiment to experiment. Both protected fragments were also observed when total nodule RNA was employed (data not shown). In contrast, a

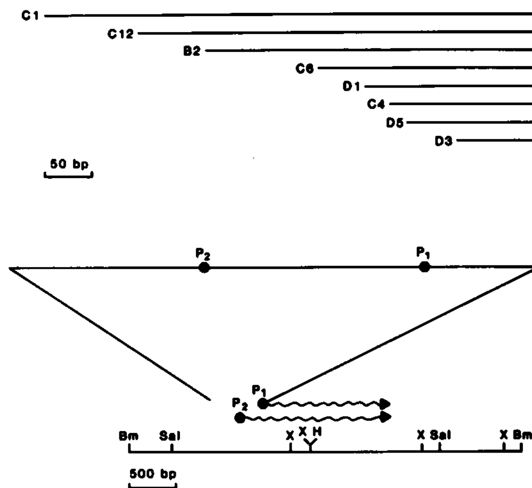

Figure 2. Promoter deletion map. The relative position of the Bal-31 nuclease generated deletions in the 5' noncoding region of the δ-ALA synthetase gene is indicated in the enlarged view of the map. The solid bars denote the DNA that remains after deletion. Restriction endonuclease cleavage sites are indicated by: Bm, BamHI; Sal, Sal I; X, Xho I; and H, Hind III.

single major 3' end was observed when either the doubly 3' end-labeled 1.25 Kb Xho I-Xho I fragment or the 2.3 Kb Bam HI/Hind III fragment were utilized (figure 1). The limits of the gene as determined by this method are consistent with the Tn5 insertion data and suggest that the δ-ALA synthetase gene is not part of a larger operon. The data also suggest that expression of δ-ALA synthetase is controlled by two promoters, designated P1 and P2. It should be noted, however, that we have not isolated any Tn5 insertion mutants in the region between P1 and P2.

Promoter Deletion Analysis. To assess the functional expression of the δ-ALA synthetase gene from each putative promoter, a series of deletions were generated in the 5' noncoding region by Bal-31 nuclease treatment of pGD23, linearized at a unique Pst I site upstream of δ-ALA synthetase. The deleted products were recircularized following the addition of Bgl II linkers. Individual Bgl II-Hind III fragments were then subcloned into pGD499, a broad host range promoter probe vector (10). The resulting plasmids were mobilized into R. meliloti (recA::Tn5) and β-galactosidase activity was determined during vegetative and symbiotic growth. As shown in Figure 2 and Tables 1 and 2, both upsteam sequences were employed

Table 1
Vegetative β-Galactosidase Activity
of Promoter-Deleted Clones

Clone	β-Galactosidase Activity[1]		
	Average[2]	Net[3]	%[4]
pXLGD4	714	578	97
C1	732	596	100
C12	692	556	93
B2	486	350	59
C6	474	338	57
D1	403	267	45
C4	442	306	51
D5	223	87	15
D3	156	20	3
pGD500	136	-0-	-0-

1 β-Galactosidase activity was assayed in R. meliloti 102F34 recA::Tn5 as described in Methods.
2 Average of 3 determinations.
3 Activity remaining after correction for transcription originating from vector sequences (pGD500).
4 Percent of the activity observed for deletion C1.

equivalently to activate expression of lacZ. Deletion D5, which is estimated to extend to within 20 bp of P1, eliminated the majority of expression both vegetatively and symbiotically.

DNA Sequence Analysis. The DNA sequence for 718 bp was determined in the region surrounding the δ-ALA synthetase promoters (Figures 3 and 4). The RNA start sites were localized by electrophoresis of the S1 nuclease-protected fragments next to a sequencing ladder (Figure 5). Two major transcription start sites separated by 238 bp were identified as before. In the case of the upstream 5' end P2, three additional minor start sites were observed at 27, 31, and 36 bp 5' to this sequence (data not visible in the reproduction of Figure 6).

The DNA sequences for P1 and P2 are compared in Figure 6. Both contain similar -35 regions that show good homology with the consensus -35 region of

Table 2
Symbiotic β-Galactosidase Activity
of Promoter-Deleted Clones

Clone	β-Galactosidase Activity[1]		
	Average[2]	Net[3]	%[4]
pXLGD4	(88)	(82)	(99)
C1	110	100	100
C12	105	95	95
B2	(47)	(41)	(49)
C6	67	57	57
D1	54	44	44
C4	32	22	22
D5	(5)	(-0-)	(-0-)
D3	(6)	(-0-)	(-0-)
pGD500	10	-0-	-0-

1 β-Galactosidase activity was assayed in R. meliloti 102F34 recA::Tn5 as described in Methods.
2 Average of 2 determinations, except for values in parentheses, which are the result of a single determination. In the latter case, percentages are based on a value of 89 for C1 and 6 for pGD500. Each determination represents a different group of nodules.
3 Activity remaining after correction for transcription originating from vector sequences (pGD500).
4 Percent of the activity observed for deletion C1.

E. coli promoters (15). In contrast, the -10 regions differ considerably from that of E. coli. These data are compatible with the inability of the cloned δ-ALA synthetase gene to effectively complement an E. coli hemA mutant when expression is completely dependent on the Rhizobium promoter(s) (3). Moreover, no β-galactosidase activity for the P1/P2-lac fusion pXLGD4 in E. coli was previously observed (10). Interestingly, both P1 and P2 show a 5 bp conserved sequence around -25. Based on the homology of P1 and P2 with each other, the region which may constitute the -10 equivalent for P1 and P2 has been underlined. Neither the sequence of P1 nor P2 bears any resemblance to the nif promoters of R. meliloti (16).

Four small open reading frames (ORFs) of 126 (-223), 93 (-190), 60 (-157), and 111 (-98) bp initiating with Met were found between P1 and P2.

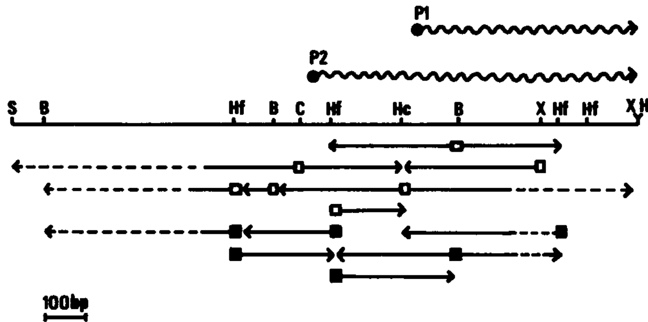

Figure 3. DNA sequencing strategy. The DNA sequence was determined using the procedure of Maxam and Gilbert. DNA fragments were labeled with T4 polynucleotide kinase (5'); T4 DNA polymerase (3'), or the Klenow fragment (3'). Open squares indicate 3' end-labeled fragments and closed squares indicate 5' end-labeled fragments. Restriction endonuclease cleavage sites are indicated by: S, Sal I; B, Bg I; Hf, Hinf I; C, Cla I; Hc, Hinc II; and X, Xho I.

The first three of these are in the same frame and have potential upstream Shine-Delgarno sequences while the 111 bp ORF appears to lack one. This last open reading frame crosses the P1 upstream region and ends 14 bp after the apparent start of transcription. A single ORF starting with Met at 13-21 bp after P1 and extending to the end of the reported sequence was also observed. This ORF is also preceded by a potential ribosome binding site (GGAG) and is presumed to be the coding sequence for δ-ALA synthetase. The codon usage for the 76 predicted N-terminal amino acids of this latter ORF is similar to that determined for the R. meliloti nif H gene (17). One ORF was found to extend across the entire sequence of the opposite DNA strand.

DISCUSSION

Analysis of the physical limits of the δ-ALA synthetase gene revealed two widely separated transcriptional start sites. One initiation site was located 13 bp upstream of the putative translational reading frame for the synthetase coding sequence, while the second start site was situated 248 bp further upstream. A number of ORFs were found in the region between P1 and P2, but all terminate prior to the AUG start codon of the synthetase. Thus the occurrence of overlapping genes or preproteins is unlikely. The possibility that these ORFs encode functional gene products cannot be excluded, in which case δ-ALA synthetase would be part of a multicistronic operon. Although considerable secondary structure can be identified in the leader regions, features characteristic of bacterial attenuators were not observed. The ORFs for 126 bp and 111 bp contained three and five glycine

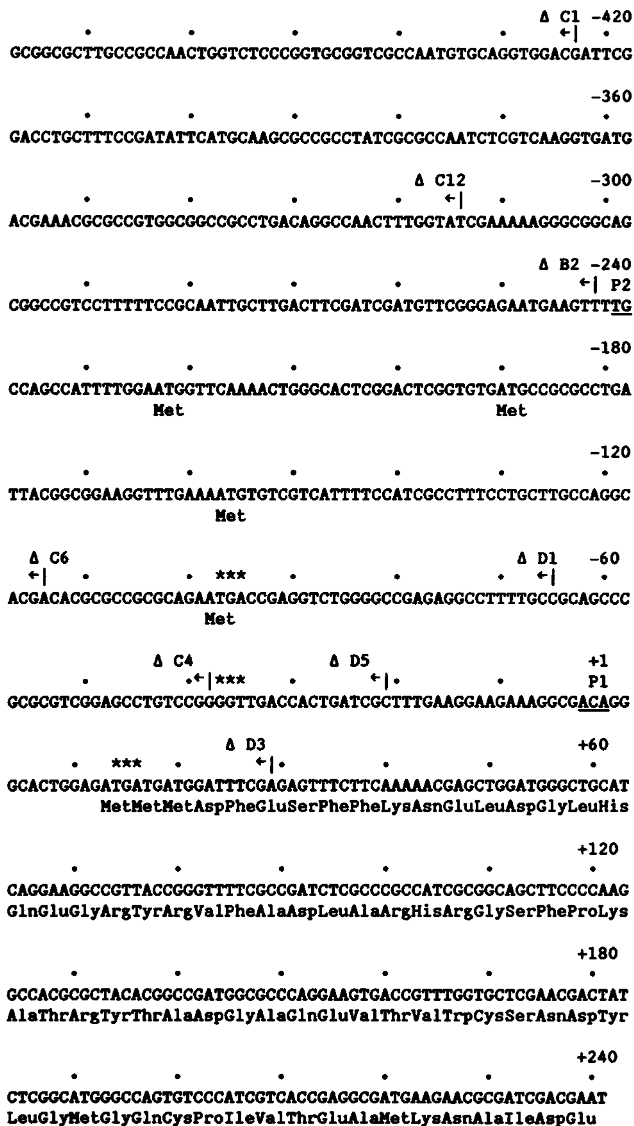

Figure 4. DNA sequence of the 5' region of the δ-ALA synthetase gene of *R. meliloti* 102F34. The sequence of the antisense strand is presented. The RNA start sites P1 and P2 are underlined. The sequence is numbered relative to P1 which is at coordinate +1. The amino acid sequence of the putative coding region for the δ-ALA synthetase gene is indicated below the respective DNA sequence. Asterisks indicate translational termination sites. The approximate locations of the Bal 31-generated deletions are indicated by arrows.

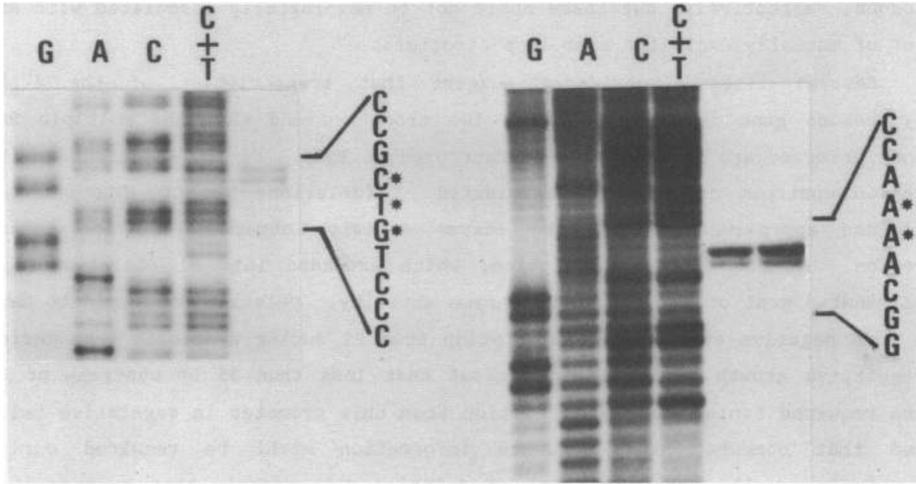

Figure 5. High resolution S1 nuclease mapping of the δ-ALA synthetase transcripts. Hybridizations were carried out as described in figure 1 using a singled-stranded 440 bp _Bgl_ I-_Bgl_ I fragment for P1 and a double-stranded 125 bp _Hinf_ I-_Bgl_ I fragment for P2, both asymmetrically labeled at the downstream 5' end. The location of these fragments relative to P1 and P2 is indicated in Figure 3. Hybridizations were carried out with 20 μg total _Rhizobium_ RNA at 56°C for P1 and with 40 μg and 80 μg total RNA at 40°C for P2. The left panel shows the S1 nuclease-protected fragments observed for P1 compared with the respective ladder of chemically cleavaged DNA on a 12% acrylamide-8M urea gel. The right panel shows the data that were obtained for P2 on a 20% acrylamide-8M urea gel.

```
          -40         -30         -20         -10
                                                       . .
P2   GCAATTGCTTGACTTCGATCGATGTTCGGGAGAATGAAGTTTTG
                                                      . . .
P1   GTCCGGGGTTGACCACTGATCGCTTTGAAGGAAGAAAGGCGACAG

P1/P2          TTGAC                        GG(G/A)AGAA(T/A)G

E. coli        TTGACA                         TATAAT

nif                         CTGGPAPP     TTGCA
                                y yu
```

Figure 6. Comparison of the δ-ALA synthetase P1 and P2 sequences with the consensus promoter of _E. coli_ and _nif_ genes. The RNA start sites, which are indicated by dots, have been corrected by taking into account the difference between chemically and S1-cleaved DNA (12).

codons, respectively, but these could not be meaningfully associated with any set of mutually exclusive stem-loop structures.

Several lines of evidence suggest that transcription of the δ-ALA synthetase gene is initiated from two promoters and that the multiple RNA ends observed are not simply precursor product RNAs. First, deletions in the region upstream of P1 which eliminated P2 (deletions B2, C6, D1, and C4) yielded approximately half the enzyme activity observed for the intact region. A more extensive deletion, which extended into P1 (deletion D5), eliminated most of the remaining enzyme activity. Deletion C4 seemed to have a more negative effect on transcription from P1 during symbiosis than during vegetative growth, which would suggest that less than 35 bp upstream of P1 are required for optimal transcription from this promoter in vegetative cells and that somewhat more sequence information might be required during symbiotic growth. The end point of deletion C12 suggests that no more than 75 bp are required in cis for maximal expression from P2. These findings are comparable with those obtained for the R. meliloti nif HDK operon in which only a short stretch of DNA (30 bp) is needed for near optimal promoter activity during symbiosis (18). Second, DNA sequence analysis of P1 and P2 indicated a high degree of homology between their respective -10 and -35 regions as well as at -25. This structural similarity is consistent with the functional equivalence that was observed for these sequences in the lac fusion analysis. Third, no evidence for a processed form of the larger transcript, i.e. a 3' end that mapped upstream of P1, was obtained (data not shown). However, such processed RNAs could have been rapidly degraded. Attempts were made to specifically label the true 5' end of the message with guanylyl transferase (19), but the low abundance of these RNAs did not permit their detection above the background in S1 mapping experiments (data not shown).

Numerous examples of multiple promoters for single genes have been reported in procaryotes and eucaryotes. In general, such promoters are both structurally and functionally distinct. For example, the tandem promoters for the gln A gene from Anabaena (20) and Klebsiella pneumoniae (21) are differentially regulated. In both cases the more upstream promoter is E. coli-like while the downstream promoter resembles activatable nitrogen fixation gene promoters. Likewise, the closely linked but structurally unique promoters for the crystal toxin from Bacillus thuringiensis are differentially expressed during sporulation (22), and the overlapping promoters of the E. coli galactose operon show differential responses to

control by the cyclic AMP-cyclic AMP receptor protein complex (23). A similar phenomenon has been noted for a number of eucaryotic genes (24,25, 26,27,28,29,30). For many of these examples, however, the distinction between differential origins of transcription and differential RNA processing has not been clearly established.

In the case of the δ-ALA synthetase gene, no evidence for structural or functional specialization of the P1 and P2 sequences was observed. Both regions seemed to function equivalently and simultaneously, during vegetative growth and under steady state nodule conditions, as determined by lac fusion analysis. However, since no attempts were made to examine the relative activities of P1 and P2 early in nodule establishment, the possibility has not been eliminated that one of these promoters may undergo transient activation or repression. Also, no discrimination was made between the mature and immature bacteroid forms normally found in alfalfa nodules, nor were attempts made to change the oxygen or iron availability to broth-grown cells. Hybridization levels in the S1 mapping experiments were generally supportive of data from gene fusion studies.

The presence of two functional promoters may simply be a mechanism to enhance gene expression. This mode of improving gene expression is commonly employed in the highly expressed bacterial ribosomal protein and rRNA operons (31,32). The transcriptional activity of the δ-ALA synthetase gene is significant. In vegetative cells, the β-galactosidase activity observed for the P1/P2-lac fusions was notably high. Although an absolute comparison with bacteroid activity cannot be made, we have observed that the level of β-galactosidase activity from ALA synthetase-lac fusions reported here is approximately 4% that observed for comparable lac fusions involving the highly active, symbiotically specific nif promoters P1 and P2 (C. H. Kim, unpublished observations). Such data are also in approximate agreement with results from RNA hybridization studies which have indicated ALA synthetase expression to be 10% that observed for nif H in nodules (B. Lewis, unpublished observations). This is the first promoter sequence reported for a gene that is not known to be under general nitrogen control in *R. meliloti*. These moderately strong, constitutive promoters should be useful in experiments involving the manipulation and analysis of rhizobial genes that are developmentally regulated.

ACKNOWLEDGEMENT

This work was supported by grant PCM 82-04730 from the National Science Foundation, Training Grant 784-821-24182-3 from the National Institutes of

Health, and funds from the Medical Research Council and Wellcome Trust Fund for Travelling Fellowships.

REFERENCES

1. Dilworth, M. J., and Appleby, C. A. (1979) in A Treatise on Dinitrogen Fixation, Hardy, R. W. F., Bottomley, F., and Burns, R. C., eds., Sections I and II, pp. 691-764, John Wiley and Sons, New York.
2. Sullivan, D., Brisson, N., Goodchild, B., Verma, D. P. S., and Thomas, D. Y. (1981) Nature 291, 516-518.
3. Leong, S. A., Ditta, G. S., and Helinski, D. R. (1982) J. Biol. Chem. 257, 8724-8730.
4. Leong, S. A., and Ditta, G. S. (1984) Abstr. 2nd International Symposium on the Plant Bacterial Interaction, Ithaca, New York.
5. Corbin, D., Barran, L., and Ditta, G. (1983) Proc. Natl. Acad. Sci. U.S.A.80, 3005-3009.
6. Maxam, A. M., and Gilbert, W. (1980) Meth. Enzymol. 65, 449-560.
7. Vieira, J., and Messing, J. (1982) Gene 19, 259-268.
8. Ditta G., Stanfield, S., Corbin, D., and Helinski, D. R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7347-7351.
9. Better, M., and Helinski, D. R. (1983) J. Bacteriol. 155, 311-316.
10. Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X., Finlay, D. R., Guiney, D., and Helinski, D. (1985) Plasmid 13, 149-153.
11. Close, T. J., and Rodriquez, R. L. (1982) Gene 20, 305-316.
12. Green, M. R., and Roeder, R. G. (1980) Cell 22, 231-242.
13. Vincent, J. M. (1970) A Manual for the Practical Study of Root Nodule Bacteria, IBP Handbook No. 15, Blackwell, Oxford.
14. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory.
15. Rosenberg, M., and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
16. Ausubel, F. M. (1984). Cell 37, 5-6.
17. Torok, I., and Kondorosi, A. (1981) Nuc. Acids Res. 9, 5711-5723.
18. Better, M., Ditta, G. S., and Helinski, D. R. (1985) EMBO (in press).
19. Kassavetis, G. A., and Geiduschek, E. P. (1982) EMBO J. 1, 107-114.
20. Tumer, N. E., Robinson, S. J., and Haselkorn, R. (1983) Nature 306, 337-341.
21. Dixon, R. (1984) Nucl. Acids Res. 12, 7811-7829.
22. Wong, A. C., Schnepf, H. E., and Whiteley, H. R. (1983) J. Biol. Chem. 258, 1960-1967.
23. Majumdav, A., and Adhya, S. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 6100-6104.
24. Young, R. A., Hagenbuchle, O., and Schibler, U. (1981) Cell 23, 451-458.
25. Benyajati, C., Spoerel, N., Haymerle, H., and Ashburner, M. (1983) Cell 33, 125-133.
26. Allan, M., Lanyon, G., and Paul, J. (1983) Cell 35, 187-197.
27. Ley, T. J., and Nienhuis, A. W. (1983) Nucl. Acids Res. 112, 1041-1048.
28. Sarokin, L., and Carlson, M. (1984) Mol. Cell. Biol. 4, 2750-2757.
29. Rose, M., and Bostein, D. (1983) J. Mol. Biol. 170, 883-904.
30. Buchman, A. R., Fromm, M., and Berg, P. (1984) Mol. Cell. Biol. 4, 1900-1914.
31. Burton, Z. F., Gross, C. A., Watanabe, K. K., and Burgess, R. R. (1983) Cell 32, 335-349.
32. Stewart, G. C., and Bott, K. F. (1983) Nucl. Acids Res. 11, 6289-6299.