Volume 16 Number 1 1988 Nucleic Acids Research

Root nodule specific gene regulation: analysis of the soybean nodulin N23 gene promoter in heterologous symbiotic systems

Jan-Elo Jørgensen, Jens Stougaard, Anne Marcker and Kjeld A.Marcker

Department of Molecular Biology and Plant Physiology, University of Aarhus, C.F.Møllers Allé 130, DK-8000 Århus C, Denmark

Received October 22, 1987; Revised and Accepted December 2, 1987

ABSTRACT

The nodulin N23 gene promoter was analysed in transgenic plants using the chloramphenicol acetyltransferase (CAT) coding sequence as a reporter. A 5' flanking region of less than 1 kb was sufficient for the organ-specific expression of a chimeric N23-CAT-3'lbc_3 gene in root nodules formed on Lotus corniculatus and Trifolium repens after infection by their respective Rhizobium symbionts. Expression was regulated at the level of RNA in both species of transgenic plants. Promoter deletion analysis defined the 5' region required for high level expression and delimited two putative regulatory sequences involved in positive control of the N23 gene in L. corniculatus.

INTRODUCTION

Root nodules which develop on soybean roots after infection with the soil bacterium Bradyrhizobium japonicum contain approximately 30 specific polypeptides called "nodulins". (1,2,3). The majority of nodulins have not been assigned a biochemical function and so far only the leghemoglobins (lb) (4), the nodule specific forms of uricase (5,6) and sucrose synthase (7,8) have been characterised functionally. Recently a small gene family encoding nodulins N20, N22, N44 and N23 was characterized (9,10,11). The corresponding polypeptides contain putative metal binding domains that might be important for their structure or function (10). Nodulin N23 is associated with the peribacteroid membrane, while N22 is a cytosolic component. The N20 and N44 geneproducts could not be located with the immunological technique used (9). Gloudemans et al. (3) designated an early nodulin as N44; this gene and the N44 gene studied here are apparently not identical.

In growing nodules the nodulin genes are transcriptionally activated in a sequential manner, to meet the requirements of the developing organ (12). The N75, N44, N41, N38 nodulin genes are induced a few days after infection (3), then followed by the induction of the lbc genes, before finally the majority of the nodulin genes are transcribed (12). Induction of the lbc

genes exhibits a characteristic exponential kinetics. Thus the lbc genes are first transcriptionally active at a low level around 7-8 days after infection, then followed by a dramatic increase in the transcription rates around day 12. Concomitantly with the increase in lb transcription, the majority of the nodulin genes are induced.

To study this induction mechanism at the molecular level a transformation & regeneration scheme for the legume Lotus corniculatus was developed. (13,14). The soybean lbc_3 gene was transferred and nodule specifically expressed in transgenic Lotus plants indicating a common control mechanism for the induction of the lb genes in the various symbiotic associations (15,16). To monitor the developmental regulation of the differentially expressed nodulin genes a chimeric soybean N23-CAT-3'lbc_3 gene was transferred to Lotus corniculatus and Trifolium repens. In both types of transgenic plants the chimeric gene was expressed only in nodules formed after infection with the respective microsymbionts. Finally, DNA deletion analysis revealed regions in the N23 promoter important for the nodule specific expression of the chimeric N23-CAT-3'lbc_3 gene.

METHODS

Plasmid constructions

Standard techniques as compiled by Maniatis et al. (17) were used for DNA manipulations and DNA modifying enzymes were used according to manufacturers instructions. The N23-CAT-3'lbc_3 gene was constructed from lbc_3 5'3'-CAT 101 (18) by replacing the 5'lbc_3 with the N23 promoter region. The lbc_3 5'3'-CAT 101 gene derived from the previously described lbc_3 5'3'-CAT gene (18,15) has a unique BglII site 10 bp upstream of the ATG codon. A N23 5' (-915 to +66) was thus fused to the CAT sequence by ligation of the Klenow polymerase endfilled DdeI site at +66 to the filled in BglII site. The resulting N23-CAT-3'lbc_3 gene carrying the lbc_3 3' flanking region was SalI cloned into the "intermediate integration vector" pAR6. (14). The chimeric N23-CAT-3'OCS was constructed by replacing the NcoI/SphI fragment carrying the 3'lbc_3 (Fig. 2) with a NcoI/SalI (SphI and SalI endfilled) fragment carrying the 320 bp (PvuII/AhaIII) 3' region of the octopine synthase gene. (19,20). The 5' Bal31 deletion series was constructed in the pIV2 vector as described by Stougaard et al. (19).

Plant transformations

The Agrobacterium rhizogenes based transformation system decribed by Stougaard et al. (14,15) was used to transfer chimeric genes to L. corniculatus (Rodéo) and T. repens (Milkanova). Regenerated plants of L. corniculatus

were obtained by the procedure of Petit et al. (13). A short cut transformation procedure was used to generate transformed roots and root nodules on T. repens and L. corniculatus. Hairy roots generated as described in Petit et al (13) were used for rooting and nodulation after disinfection of the complete plantlet with claforan (500µg/ml). The normal root system was exised at the base of the stem. The R. loti strain NZP2037 (21) and R. trifolii strain Cl.F. Roth (Danish culture collection; Planteavlslaboratoriet, Lottenborgvej 24, Lyngby DK-2800) were used as inoculum on L. corniculatus and T. repens, respectively.

RNA analysis

Extraction and Northern blot analysis were as described by Marcker et al. (12) and Stougaard et al. (15). The CAT probe used was a selfligated mixture of two EcoRI fragments encompassing the entire CAT coding sequence of the lbc 5'3'-CAT gene. (18). The XhoI fragment of pHUb14-38 (22) was used as ubiquitin probe.

CAT assays

The activity of chloramphenicol acetyltransferase was determined according to Stougaard et al. (15).

RESULTS

Transcription of the N23 gene in soybean nodules

The N23 gene is a member of the nodulin gene family described by Sandal et al. (10), Jacobs et al. (9) and Sengupta-Gopalan et al. (11). This gene family is subdivided into two groups according to the divergence of their DNA sequences. One group contains the N20, N22 genes while the other contains the N44 and N23 genes. The previously described N17,10I cDNA complementary to the N44 mRNA (2) also recognizes the N23 gene and the corresponding mRNA. Accordingly radioactive labelled N17,10I cDNA was used to determine the induction kinetics of the N44 and N23 genes during nodule development. Fig. 1 shows the result of a Northern analysis using total RNA extracted from nodules in different developmental stages. The upper band corresponds to the transcript from the N44 gene while the lower band corresponds to the N23 transcript. The N23 and N44 genes are activated around 12 days after infection together with most other nodulin genes. Even prolonged exposure of the autoradiography does not allow detection of transcripts before day 12, while hybridization with the lb cDNA under similar conditions using the same filter clearly revealed lbc transcripts already at day 7-8 postinfection (12). The intensity of the N23 transcript signal at day sixteen indicates that the N23 gene has a strong promoter activity as also previously found, (10).

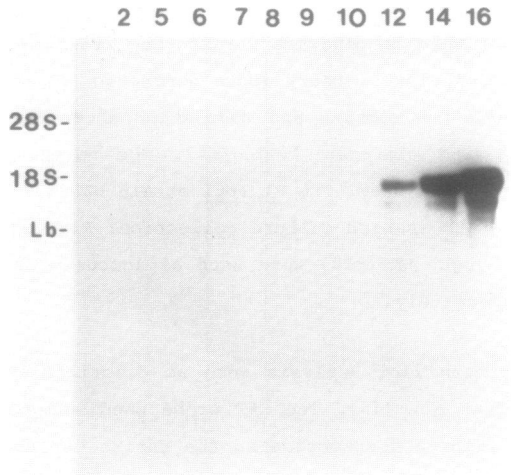

Fig. 1. Induction kinetics of the N23 gene in soybean root nodules. Total RNA (50µg) from developing root nodules was hybridized with the soybean N17,10I cDNA probe which recognizes the 1700 bp transcript of the N44 gene and the 1000 bp transcript from the N23 gene (2,10). The numbers above the lanes indicate the time in days after inoculation with the B. japonicum microsymbiont. The positions of the 28S, 16S rRNA and the lb mRNA (Lb) are indicated to the left. The lbc_3 gene transcript was in a previously performed hybridization with the lb cDNA probe detectable around day 7-8 postinfection, see fig. 3 Marcker et al. (12).

Construction of chimeric CAT genes

The chloramphenicol acetyltransferase (CAT) coding sequence was chosen as reporter gene for the analysis of the N23 promoter in two different constructions (Fig. 2). One chimeric gene N23-CAT-3'lbc_3 was constructed by substituting the lbc_3 promoter of the lbc_3 5'3'-CAT chimeric gene (18,15) with the N23 5' region from -915 to +66. The sequence of this region was determined in this laboratory and corresponds to the corrected version (23,10) of the first published sequence (24). The 3' lbc_3 flanking region of N23-CAT-3'lbc_3 was subsequently replaced by a 320 bp 3' region from the octopine synthase (OCS) gene resulting in the N23-CAT-3'OCS gene. Both chimeric genes were separately cloned into the SalI site of the "intermediate integration vector" pAR6. (14).

Transformation of plants

The chimeric N23-CAT-3'lbc_3 and N23-CAT-3'OCS genes carried by pAR6 were integrated into the TL-DNA of the A. rhizogenes strain C58C1 pRi15834 and transformed L. corniculatus plants were generated as described in Petit et al. (13) or by a shortcut transformation procedure generating transformed

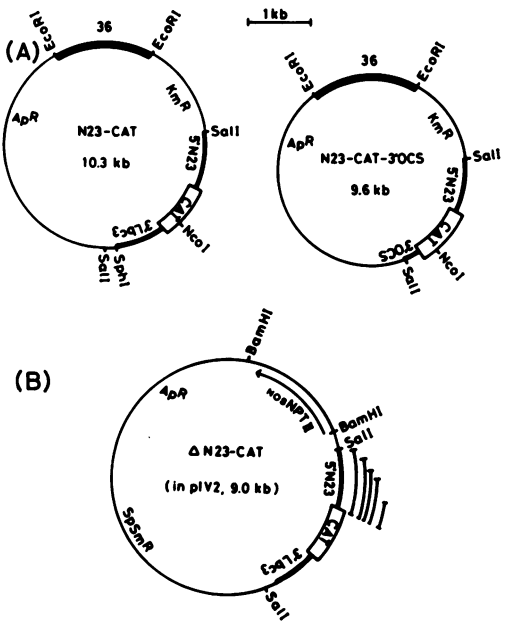

Fig. 2. Diagrams of chimeric gene constructions and plasmids used to study expression from the N23 promoter region. (A) The N23-CAT-3'lbc$_3$ and N23-CAT-3'OCS constructions used to analyse the 5' and 3' flanking regions. (B) Plasmid constructions ΔN23-CAT used in the 5' deletion analysis of the N23 promoter .

roots and root nodules on untransformed shoots. (See methods). The 5' DNA deletion series (see below) was transferred to plants with the A. rhizogenes AR1193 strain which accepts pBR322 plasmids into the TL-DNA. (14). The shortcut transformation procedure was used for expression of these constructs and for transformation of Trifolium repens.

Organ specific expression of the N23 promoter in L. corniculatus and T. repens

Expression of CAT activity under control of the N23 promoter was analysed in both the Lotus system and in Trifolium repens. Extracts from roots, nodules and leaves + stem pieces of completely regenerated Lotus plants were assayed using a standard procedure. Chloramphenicol acetyltransferase activity was detected in root nodules only and the N23-CAT-3'lbc$_3$ gene is therefore organ specifically expressed. (Fig. 3). Similar results were obtained in T. repens transformed with the N23-CAT-3'lbc$_3$ gene or the lbc$_3$ 5'3'-CAT gene. This indicates that the regulatory mechanism control-

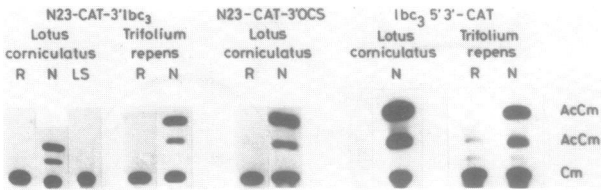

Fig. 3. Expression of the chimeric N23 and lbc_3 genes in tissues of L. corniculatus and T. repens. Equal amounts of root (R), nodule (N) or leaf + stem (LS) tissues were homogenized in a mortar and CAT activity measured by conversion of ^{14}C-chloramphenicol (Cm) to acetylated derivatives, (AcCm). The chimeric gene expressing the measured CAT activity is indicated above the respective lanes. Conversion of chloramphenicol in lanes 2,5 and 7 show the nodule specific expression of the N23 promoter.

ling nodulin gene expression is well conserved among the different symbiotic associations.

To further localize sequences responsible for organ-specific expression, the 3'lbc_3 flanking region of N23-CAT-3'lbc_3 was replaced with the 3'OCS region in the N23-CAT-3'OCS construction. This gene was also nodule-

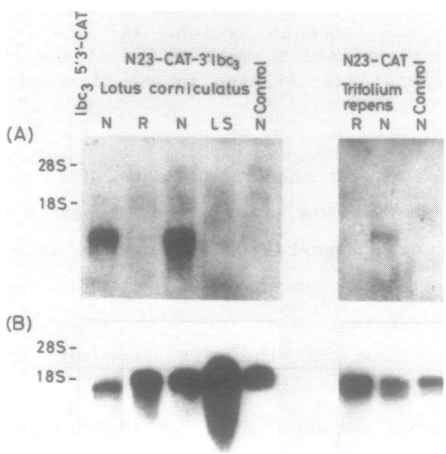

Fig. 4. Northern analysis of RNA from tissues of N23-CAT-3'lbc_3 transformed L. corniculatus and T. repens. Ten µg of total RNA extracted from root (R), nodule (N) or leaf + stem (LS) was hybridized. (A) Transcripts from the N23-CAT-3'lbc_3 gene visualized using a probe covering the CAT coding sequence. (B) Transcripts from the constitutively expressed ubiquitin genes visualized with a cDNA probe for the human ubiquitin gene. Lane 1 contains nodule RNA from a lbc_3 5'3'-CAT transformed Lotus line, Stougaard et al. (15), lane 5 control RNA from a lbc_3 transformed Lotus line, Stougaard et al. (16), lane 8 control RNA from nodules of untransformed T. repens.

specifically expressed in transgenic Lotus nodules. (Fig.3). The N23 promoter region must therefore carry sufficient regulatory elements for a high level of organ specific expression. The efficiency of transcription of the N23-CAT-3'lbc_3 gene was estimated by Northern analysis of total RNA extracted from roots, nodules and leaves plus stem pieces of transgenic Lotus corniculatus and T. repens plants. Total nodule RNA from a lbc_3 5'3'-CAT transformed Lotus line was used as a positive control (15). Fig. 4 shows the results of such an analysis with a probe covering of the entire CAT coding sequence. The intensity of the hybridization signal from N23-CAT-3'lbc_3 transformed Lotus nodules corresponds roughly to that of the the lbc_3- 5'3'-CAT transcripts and the two promoters thus seem of about equal strength in L.corniculatus. The N23-CAT-3'lbc_3 transcript is about 900 nucleotides as expected. The root nodule specific regulation of the N23 promoter is therefore at the level of RNA in both L.corniculatus and T. repens. Fig. 4b shows the control hybridization using a cDNA probe for the constitutively expressed ubiquitin genes (22). The size of the hybridizing mRNA corresponds to a trimer or tetramer ubiquitin polyprecursor. A tetramer form was constitutively expressed in barley leaves and roots, where no developmental control of this species was found (25). Different steady state levels of ubiquitin mRNA are present in different barley tissues. This is apparently also the case for different L.corniculatus and T.repens tissues, nevertheless the hybridization results clearly shows that there is no degradation of the applied mRNA and that the ubiquitin gene is constitutively expressed in all tissues examined.

Deletion analysis of the N23 promoter region

A Bal31 deletion series of the N23 promoter region was generated in order to identify DNA sequences required for the characteristic expression of the N23 gene. Deletions separated by approximately 50 bp were subcloned into the pIV2 vector (Fig. 2) and transgenic roots and nodules formed on Lotus using the shortcut procedure were subsequently analysed for CAT activity. Two to three individual transformants were tested for each deletion between -915 to -400; eight to ten transformants for each deletion between -344 to -15. The maximum variation in CAT activity in each series was found to be around 50% of the average values given in fig 5.

Deletion of 5' sequences until the -344 position had no dramatic effects on the level of CAT activity expressed from the N23 promoter. (Fig.5). Removal of a further 51 bp region to position -293 reduced the level of CAT to approximately 15% while removal of yet another 82 bp between -247 and

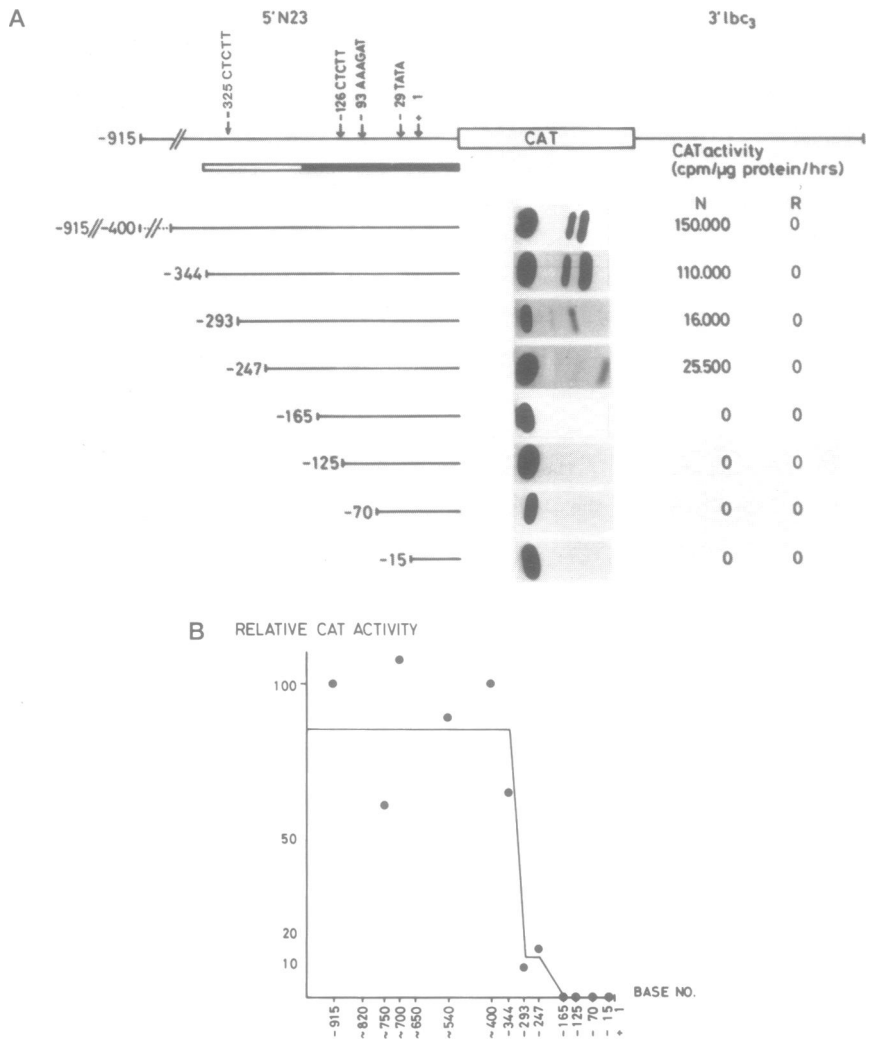

Fig. 5. 5' deletion analysis of the N23 promoter region. The diagram shows the deletions studied in transgenic root nodules of L. corniculatus. Endpoints are given as the nucleotide number relative to the cap site. The distal 5' deletions were determined by gel electrophoresis only. The position of TATA, AAAGAT and CTCTT boxes are shown relative to the transcription initiation site determined by Mauro et al. (24) at +1. Sequences conserved between the N20, N22, N44 and N23 genes are indicated with a closed bar; sequences homologous between N44 and N23 with an open bar. CAT activity in roots and root nodules is given in cpm/µg protein/hour in the column next to the schematic representation of deletions. The limit of detection is approximately 0.1% of the full CAT activity. (B) shows the two step CAT activity profile of the N23 promoter region.

-165 abolished CAT activity to below the level of detection. Two sequence elements potentially required for effective organ specific expression are thus delimited by the -344 to -293 and -247 to -165 deletions. Both sequences seem to function as positive control elements. The root nodule specific expression of the -247 deletion locates sequences sufficient for organ specificity within -247 to +66 region of the N23 5' region.

DISCUSSION

The nodule specific expression of the soybean *lbc$_3$* 5'3'-CAT gene in the heterologous symbiotic system formed between transgenic *Lotus corniculatus* and *Rhizobium loti* indicated that the molecular mechanism responsible for *lb* gene activation in legumes is to some extent conserved (15). The expression of the *lbc$_3$* 5'3'-CAT gene in transgenic *T. repens* nodules further supports this notion. Nodulin genes are however, differentially expresed during nodule development and a soybean nodulin N23 gene with an induction kinetic different from the *lbc$_3$* gene was therefore chosen for comparative studies.

The chimeric N23-CAT-3'*lbc$_3$* gene is also specifically expressed in the root nodules of transgenic *L. corniculatus* and *T. repens* formed after inoculation with their respective microsymbionts *Rhizobium loti* and *Rhizobium trifolii*. The Northern analysis of total RNA from transgenic *L. corniculatus* and *T. repens* further shows that the N23-CAT-3'*lbc$_3$* gene is controlled at the level of RNA as is the case for the N23 gene in soybean. The *lbc$_3$* and N23 promoters appear to be recognized equally well in *Lotus* since the steady state levels of the chimeric RNA's are of the same order of magnitude. Tentatively, it can be concluded that the molecular mechanisms responsible for activation of some of the nodulin genes are conserved to some extent in the various *Rhizobium* - legume associations. The apparent difference in steady state mRNA expressed form the N23-CAT-3'*lbc$_3$* gene in *L. corniculatus* and *T. repens* (Fig. 4A) suggests that regulatory components involved in nodule specific expression may interact with different efficiencies in the two heterologous systems. A generally lower expression level of individual nodulin genes in *T. repens* or a faster mRNA turnover may however also contribute to the observed difference. CAT activity expressed from N23-CAT-3'*lbc$_3$* in *T. repens* (Fig. 3) is about half the maximum level seen in *L. corniculatus* despite the fact that CAT mRNA is more than two fold lower in *T.repens*. This indicates that posttranscriptional events such as preferential translation or different protein stability are

also of importance in comparing expression levels of hetorologous genes in transgenic plants.

Replacement of the 3' $\underline{lbc_3}$ flanking region of the N23-CAT gene with the 3' OCS region does not have any significant effect on the level of CAT activity expressed from the N23 promoter. Assuming that the mRNA stability is not changed it can therefore be concluded that the N23 5' region carries sufficient promoter elements for a high level organspecific expression. The 5' deletion analysis located the essential part of the N23 promoter to within the -344 to +66 region. This short region contains $\underline{cis}$ elements sufficient for efficient organ-specific expression of the CAT coding sequence. Similar short promoter regions are known from the 35S promoter (26) and the β-conglycinin promoter (27). The 5' regions of the N20, N22, N44 and N23 genes are homologous up to position -188 after which the N20 and N22 sequences diverge separately (10). Homology between the N23 and N44 5' regions extends further until the -353 position of N23. The homologous 5' regions of the gene family are contained within the region of the N23 promoter sufficient for high level expression, while the putative positive element delimited by the -344 to -293 deletion endpoints is located outside the homologous region. Assignment of these regulatory sequences is based on the level of CAT activity expressed from 5' deletions assuming that the transcription initiation site is in all cases unchanged. Hybrid-promoter studies will now be used to define these elements further and to see whether they are organ-specific regulatory elements. Upstream elements with positive effect were also delimited in the chalcone synthase $\underline{chs}$ promoter (28), and the photoregulated $\underline{lhcp}$, $\underline{rbc}$SE9 promoters (29,30). Two sequences 5'AAAGAT 3' and 5'CTCTT 3' found in all nodule-specific genes examined so far, map around positions -93, -126, and -325 of the N23 5' region (10). The deletion analysis demonstrates that these sequences together with the TATA box are not sufficient to constitute an active promoter. One of the putative regulatory sequences identified here contains the conserved CTCTT sequences at -325. It is however, possible that these sequences are organ-specific elements or negative elements, as indicated for the $\underline{lbc_3}$ promoter (19). The compact arrangement of the N23 promoter is in contrast to the $\underline{lbc_3}$ promoter where the regulatory elements are widely spaced (19). This might reflect different regulatory control mechanisms which is as also suggested by the induction kinetics of the two genes, or merely indicate structural flexibility in nodule-specific promoters.

ACKNOWLEDGEMENTS

We would like to thank Astrid Kühle for excellent technical assistance. This research was supported by the EEC contract BAP-0173-DK, The Danish FTU programme, The Danish Biomolecular Engineering Programme, The Danish Biotechnology Programme and De Danske Sukkerfabrikker A/S.

REFERENCES

1. Legocki, R.P. and Verma D.P.S. (1980). Cell 20, 153-163.
2. Bojsen, K., Jensen, E.Ø., Lund, M., Sandal, N.N. and Marcker, K.A. (1985). In Analysis of the plant genes involved in Legume-Rhizobium symbiosis, p. 112-120. OECD Paris.
3. Gloudemans, T., de Vries, S., Bussink, H.-J., Malik, N.S.A., Franssen, H.J., Louwerse, J. and Bisseling, T. (1987). Plant Molec. Biol. 8, 395-403.
4. Dilworth, M.J. and Appleby, C.A. (1979) In Hardy R.W.F., Bottomely F. a and Burns R.C. (eds). A Treatise on Dinitrogen Fixation II. Wiley, NY, pp 691-763.
5. Christensen, T.M.I.E. and Jochimsen, B.U. (1983). Plant Physiol. 72, 56-59.
6. Bergmann, H., Preddie, E. and Verma, D.P.S. (1983). EMBO J. 2, 2333-2339.
7. Morell, M. and Copeland, L. (1985). Plant Physiol. 78, 149-154.
8. Thummler, F. and Verma, D.P.S. (1986). In Third International symposium on the Molecular genetics of Plant-Microbe Interactions. McGill University, Montreal. p. 182.
9. Jacobs, F.A., Zhang, M., Fortin, M.G. and Verma, D.P.S. (1987). Nucleic Acids Res. 15, 1271-1280.
10. Sandal, N.N., Bojsen, K. and Marcker, K.A. (1987). Nucleic Acids Res. 15, 1507-1519.
11. Sengupta-Gopalan, C., Pitas J.W., Thomson, D.V, Hofman, L.M., (1986). Molec. Gen. Genet. 203, 410-420.
12. Marcker, A., Lund, M., Jensen, E.Ø. and Marcker, K.A. (1984). EMBO J. 3, 1691-1695.
13. Petit, A., Stougaard, J., Kühle, A., Marcker, K.A. and Tempé, J. (1987). Molec. Gen. Genet. 207, 245-250.
14. Stougaard, J., Abildsten, D. and Marcker, K.A. (1987). Molec. Gen. Genet. 207, 251-255.
15. Stougaard, J., Marcker, K.A., Otten, L. and Schell, J. (1986). Nature 321, 669-674.
16. Stougaard, J., Petersen, T.E. and Marcker, K.A. (1987). Proc. Natl. Acad. Sci. U.S.A. 84, 5754-5757
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, N.Y.
18. Jensen, E.Ø., Marcker, K.A., Villadsen, I.S. (1986). EMBO J. 5, 843-847.
19. Stougaard, J., Sandal, N.N., Grøn, A., Kühle, A. and Marcker, K., EMBO J. in press.
20. Gielen, J., De Beuckeleer, M., Seurinck, J., Deboeck, F., De Greve, H., Lemmers, M., Van Montagu, M. and Schell, J. (1984). EMBO J. 3, 835-846.
21. Pankhurst, C.E. 1977. Can. J. Microbiol., 23, 1026-1033.
22. Wiborg, O., Pedersen, M.S., Wind, A., Berglund, L.E., Marcker, K.A. and Vuust, J., (1985). EMBO J. 4, 755-759.
23. Wong, S.L. and Verma, D.P.S. (1985). EMBO J. 4, 2431-2438.

24. Mauro, V.P., Nguyen, T., Katinakis, P. and Verma, D.P.S. (1985). Nucleic Acids. Res. 13, 239-249.
25. Gausing, K. and Barkardottir, R. (1986). Eur. J. Biochem. 158, 57-62.
26. Odell, J.T., Nagy, F. and Chua, N.-H. (1985). Nature 313, 810-812.
27. Chen, Z-L., Schuler, M.A. and Beachy, R.N. (1986). Proc. Natl. Acad. Sci. U.S.A. 83, 8560-8564.
28. Kaulen, H., Schell, J. and Kreuzaler, F. (1986). EMBO J. 5, 1-8.
29. Simpson, J., Timko, M.P., Cashmore, A.R., Schell. J., Van Montagu, M. and Herrera-Estrella L. (1985). EMBO J. 4, 2723-2729.
30. Fluhr, R., Kühlemeier, C., Nagy, F. nd Chua, N.-H. (1986). Science 232, 1106-1112.