Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5754–5757, August 1987
Cell Biology

Expression of a complete soybean leghemoglobin gene in root nodules of transgenic *Lotus corniculatus*

(symbiotic nitrogen fixation/plant transformation/RNA processing/Lbc_3 synthesis and stability)

JENS STOUGAARD, TORBEN E. PETERSEN, AND KJELD A. MARCKER

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

Communicated by Diter von Wettstein, April 15, 1987 (received for review March 26, 1987)

ABSTRACT **The complete soybean leghemoglobin *lbc*$_3$ gene was transferred into the legume *Lotus corniculatus* using an *Agrobacterium rhizogenes* vector system. Organ-specific expression of the soybean gene was observed in root nodules formed on regenerated transgenic plants after infection with *Rhizobium loti*. The primary transcript was processed in the same way as in soybean nodules and the resulting mRNA was translated into Lbc_3 protein. Quantitative determination of the Lbc_3 protein in nodules of transgenic plants indicated that the steady-state level of the soybean protein is comparable to that of endogenous *Lotus* leghemoglobin.**

Studies of plant gene expression have mainly used convenient reporter gene sequences linked to 5′ promoter regions in chimeric gene constructions. The chloramphenicol acetyltransferase (CAT) and neomycin phosphotransferase coding sequences have been extensively used to analyze the regulation of the ribulose-1,5-biphosphate carboxylase small subunit gene *rbcSss3.6* (1), the chalcone synthase gene *chs* (2), the alcohol dehydrogenase gene *Adh-1* (3), and the leghemoglobin gene *lbc*$_3$ (4). It is however also of interest to study the behavior of complete genes transferred into a heterologous plant since aberrant RNA processing and translation may result. Ineffective splicing of transcripts from genes of monocot plants (5) or mammalian genes (6) in a dicot plant have been observed, indicating that precursor mRNA processing may vary in plants. Functional expression of a gene will furthermore depend on efficient translation and for some proteins also depend on posttranslational modifications that ensure the stability or activity of the polypeptide. Several complete plant genes have been transferred to a heterologous genetic background and the expression has been studied in callus or regenerated plants. The soybean heat shock protein gene *hs6871* (7) and the zein seed storage protein gene *Z4* (8) are both correctly transcribed in sunflower callus, as is the wheat chlorophyll a/b binding protein gene *whAB1.6* in petunia (9). None of these genes contains intervening sequences, but precursor mRNA molecules from the intron containing pea ribulose biphosphate carboxylase small subunit genes *rbcS-E9* (10), *rbcS-3A*, and *rbcS-3C* (11) are also processed to correct-size mRNA transcripts in tobacco or petunia. The primary transcripts of these genes are thus correctly processed in the new background. Storage protein genes from bean phaseolin (12) and soybean conglycinin (13) are expressed in seeds of transgenic tobacco and petunia, respectively. Substantial amounts of both proteins are synthesized and subsequently stored in the seeds.

We have transferred a complete gene encoding the soybean root nodule-specific leghemoglobin, *lbc*$_3$, polypeptide into the genome of *Lotus corniculatus*. The gene has three introns interrupting the coding sequence in codon 32, between codons 68 and 69, and between codons 103 and 104 (14). Expression of the gene in soybean root nodules is detectable at the level of precursor mRNA (15) and at the protein level (16). We have found that the soybean Lbc_3 protein accumulates in root nodules of transgenic *Lotus* plants to a concentration comparable to that of the endogenous *Lotus* Lb proteins.

MATERIALS AND METHODS

DNA Manipulations. The complete *lbc*$_3$ gene was subcloned from the genomic clone λGmLb3 (17) as a 3.5-kilobase (kb) *Msp* I fragment (HH). The 2.5-kb *lbc*$_3$ fragment (HX) was generated from the unique *Xba* I site in the HH 5′ region. Both fragments were end-filled with Klenow polymerase and *Bam*HI sites were generated by ligating to the Klenow polymerase end-filled *Bam*HI site of pBR322. The *lbc*$_3$ HH and HX constructs were subsequently cloned into the "intermediate integration vectors" pAR1 and pAR2 (18) as shown in Fig. 1, using standard methodology (19).

Plant Transformation and Regeneration. The intermediate integration vectors pAR10 and pAR26 were conjugated into the *Agrobacterium rhizogenes* C58C1pRi15834 rifampicin-resistant (rif^R) strain by the method of Van Haute *et al.* (20). Cointegrates in the T_L-DNA were selected on plates with rifampicin (100 μg/ml), spectinomycin (100 μg/ml), and streptomycin (100 μg/ml) and checked by Southern analysis as described in Stougaard *et al.* (18). Transformed *L. corniculatus* roots were generated by wound site infection and plants were regenerated according to Petit *et al.* (21).

RNA Techniques. Total RNA was extracted, oligo(dT)-purified, and analyzed in transfer blots as outlined in Marcker *et al.* (15) and Stougaard *et al.* (4).

Isoelectric Focusing. Protein was extracted from root nodules by the method of Fuchsman and Appleby (16). The isoelectric focusing system LKB 1804 111 was used according to the manufacturer's instructions.

Protein Sequencing. The Lbc_3 protein was electroeluted from the isoelectric focusing gel using the method of Hunkapiller *et al.* (22). Total Lb protein was purified on a Superose 12 column (Pharmacia) in 50 mM NH_4HCO_3. The N-terminal amino acid sequences were determined on an Applied Biosystems 470A protein sequencer.

RESULTS

Plant Transformations. Two different versions of the complete soybean *lbc*$_3$ gene were transferred into *Lotus* using the *A. rhizogenes* transformation system previously described (18, 21). The *lbc*$_3$ HH with a 2-kb 5′ region and the *lbc*$_3$ HX carrying a 1-kb 5′ region were separately integrated into the T_L-DNA segment of the *A. rhizogenes* pRi15834 plasmid by means of the intermediate integration vectors pAR1 and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: CAT, chloramphenicol acetyltransferase.

pAR2 (18) (Fig. 1). Transformed roots from wound site infections of *L. corniculatus* were subsequently propagated *in vitro* and plants were regenerated (21). Fully regenerated plants were inoculated with the *Rhizobium loti* NZP2037 strain. Plant lines recovered from the transformation and regeneration procedure were analyzed for the presence of the lbc_3 gene in Southern blots, using a full-length soybean *lba* cDNA as a probe (Fig. 2). Endogenous *L. corniculatus lb* genes are also recognized by this probe, but the transformed line contains the additional hybridizing (3.5 kb) *Bam*HI and (6 kb) *Hin*dIII fragments originating from the lbc_3 HH plasmid pAR10 carried by this line (Fig. 1). The hybridization pattern of the endogenous fragments suggests that *L. corniculatus lb* genes are homologous to soybean *lb* genes.

RNA Processing in Root Nodules. We have previously reported the organ-specific expression of a chimeric soybean gene, lbc_3 *5'3'-CAT*, in root nodules of *L. corniculatus* (4). Expression of this gene was controlled at the level of RNA (4). Total RNA was therefore extracted from leaf plus stem, root, and nodule tissues of a *Lotus* line transformed with the complete lbc_3 HH gene. The RNA was purified by oligo(dT) chromatography and subsequently subjected to RNA transfer blot analysis. The soybean *lba* cDNA probe was initially used to probe the presence of mRNA corresponding to the transferred lbc_3 HH gene. The lbc_3 mRNA does, however, not separate from the endogenous *Lotus* Lb mRNA that also hybridizes to the Lba probe (4, 21). It was therefore not possible to unambiguously demonstrate the presence of soybean lbc_3 mRNA in the transgenic nodules. An lbc_3-specific probe originating from the third intron (15) was therefore used to detect lbc_3 precursor RNA molecules (Fig. 3). Transcription of the soybean lbc_3 gene is only detectable in root nodules of transgenic *Lotus* plants and the gene is therefore organ-specifically expressed. RNA processing seems furthermore to generate the same precursors (A and B, Fig. 3) as seen in soybean root nodules. The short distance between intermediates is due to the short length of the first (120 nucleotides) and second (100 nucleotides) lbc_3 introns (15).

Leghemoglobin lbc_3 Synthesis in Transgenic Root Nodules. Root nodules from *Lotus* lines carrying the soybean lbc_3 HH

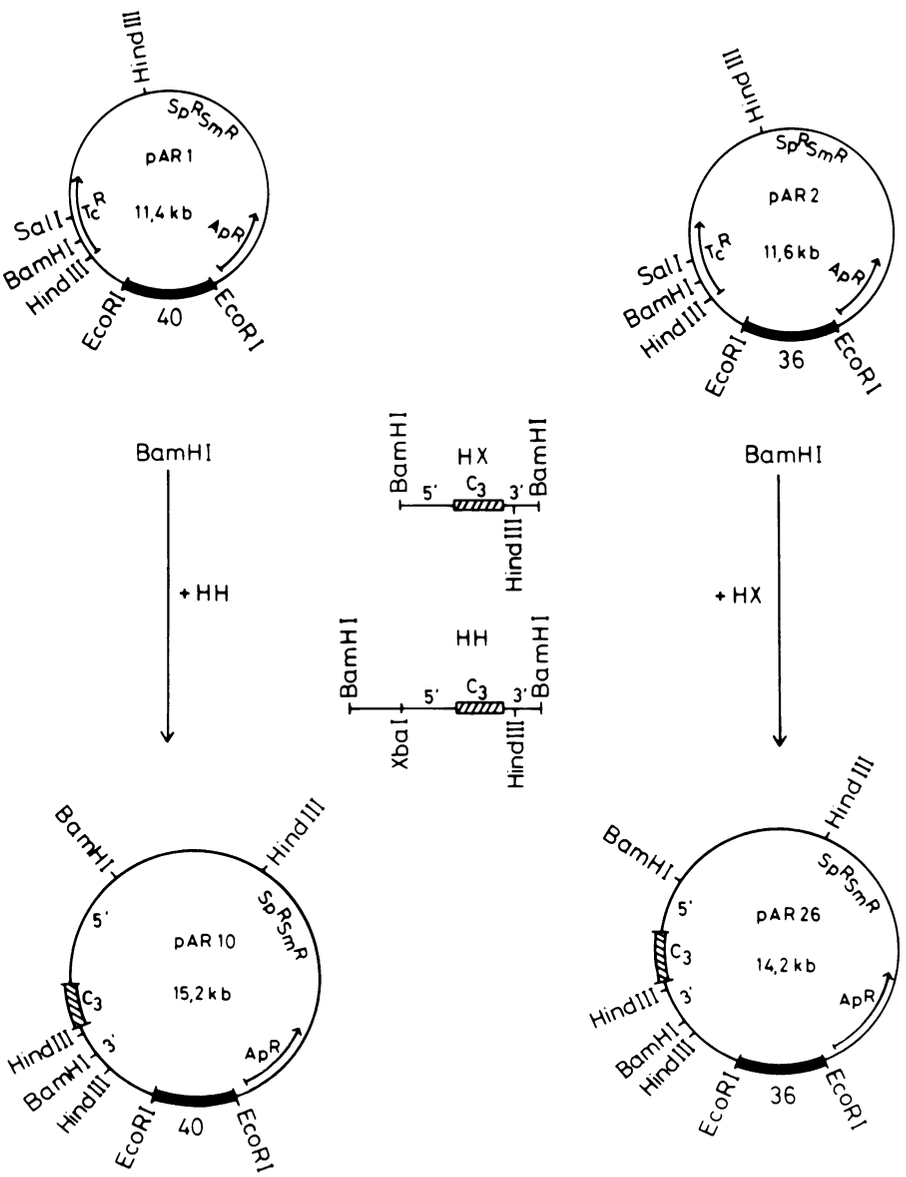

FIG. 1. Cloning of the complete soybean lbc_3 gene into intermediate integration vectors. Two different versions of the lbc_3 gene, HH and HX, were subcloned from the original λGmLb3 vector (17) and transferred to *L. corniculatus*. The lbc_3 HH carrying a 2-kb 5' region corresponds to the 3.5-kb *Msp* I fragment of λGmLb3. The shorter lbc_3 HX carrying a 1-kb 5' region corresponds to the 2.5-kb *Msp* I/*Xba* I fragment of λGmLb3. *Bam*HI sites were created at the termini of both fragments before subcloning. Ap^R, ampicillin-resistance; Tc^R, tetracycline-resistance; Sp^R, spectinomycin-resistance; Sm^R, streptomycin-resistance.

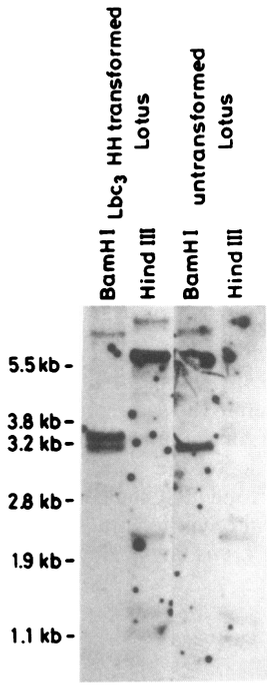

FIG. 2. Southern analysis of transgenic *L. corniculatus* plants carrying the soybean lbc_3 HH gene. Total DNA was extracted from leaves and digested with *Bam*HI and *Hin*dIII restriction endonucleases, respectively, and the cDNA corresponding to the Lba polypeptide was used as probe. The first two lanes on the left demonstrate the presence of the expected 3.5-kb *Bam*HI and 6-kb *Hin*dIII fragments from the lbc_3 HH gene generated from pAR10 in the HH transformed plant. The additional hybridizing fragments originate from endogenous *lb* sequences as shown in the two lanes on the right.

or HX genes were analyzed for leghemoglobin polypeptides. Soluble protein was extracted from nodules of transformed and untransformed plants by the method of Fuchsman and Appleby (16). In this procedure the Lb proteins are present as ferric nicotinate complexes allowing their separation according to isoelectric points. The four leghemoglobins in soybean root nodules—Lbc_1, Lbc_2, Lbc_3, and Lba—have isoelectric points between 4 and 5 (16), and consequently an acrylamide/Ampholine system creating a pH gradient from 4 to 5.5 was used for separating the Lb proteins. Untransformed *Lotus* has four dominant proteins separating in this system (Fig. 4 *Left*). One protein (Lb_I) migrates as the soybean Lbc_1 with an isoelectric point of 4.66 (16), while the other proteins migrate toward the pH 5.5 boundary. The lbc_3 transformed lines contain, in addition to these proteins, a protein migrating to the soybean Lbc_3 position (Fig. 4 *Right*). Protein from this band was electroeluted from the gel and analyzed by amino acid sequencing. The sequence obtained (12 cycles) is identical to the known N-terminal sequence of the soybean Lbc_3 protein. *Lotus* leghemoglobins from untransformed plants were purified on a Superose 12 column and applied onto the sequenator. Fig. 5 shows the N-terminal sequence obtained (22 cycles) on the total *Lotus* Lb fraction together with the N-terminal sequence of soybean Lbc_3. There are several differences between the *Lotus* and soybean Lbs and it is therefore concluded that the additional protein in the transformed plants is indeed soybean Lbc_3. The amount of Lbc_3 in transgenic root nodules was determined by quantitating the sequence reactions on total purified Lb from the Lbc_3 HH transformed lines shown in Fig. 4 *Right*. In this way it was found that soybean Lbc_3 comprised 15–20% of the total Lb sequences present.

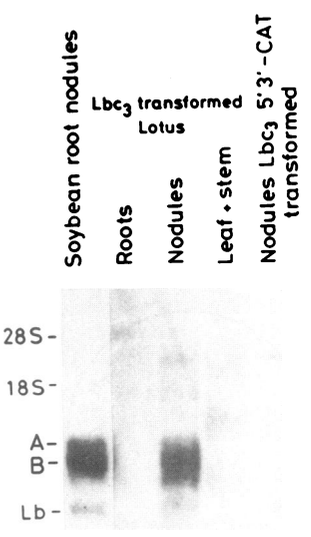

FIG. 3. Processing of soybean lbc_3 precursor RNA in *L. corniculatus*. An intron-specific lbc_3 probe originating from the third intron was hybridized to oligo(dT)-purified poly(A)$^+$ RNA. A and B denote two of the mRNA precursors. The leftmost lane contains RNA from 20-day-old soybean root nodules. The centermost lanes (left to right) contain RNA isolated from roots, nodules, or leaves and stem of a lbc_3 transformed *Lotus* line. The rightmost lane contains RNA from nodules of lbc_3 5′3′-CAT-transformed *Lotus*.

DISCUSSION

Root nodules formed on legumes after infection with *Rhizobium* contain a range of specialized proteins. At least 30 polypeptides are specifically synthesized in root nodules (23, 24). The predominant nodule-specific proteins are the leghe-

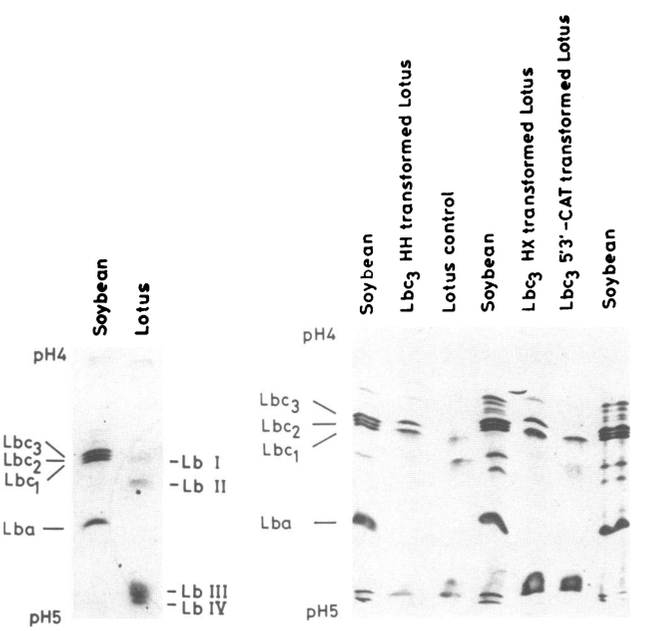

FIG. 4. Isoelectric focusing of soluble protein from root nodules of lbc_3 HH, HX transformed, and untransformed *L. corniculatus*. (*Left*) Nodule proteins from soybean and untransformed *Lotus* separated in the pH 4–5.5 range. The soybean Lba, Lbc_1, Lbc_2, and Lbc_3 proteins are indicated. The putative *Lotus* Lbs are numbered I, II, III, and IV. Protein was stained with Coomassie blue. (*Right*) Nodule protein from soybean, lbc_3 HH, and lbc_3 HX transformed *Lotus* separated in the pH 4–5.5 range and stained with Coomassie blue. Notice the Lbc_3 band in the HH and HX lines.

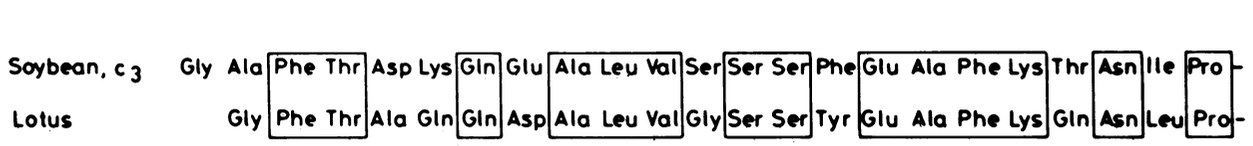

FIG. 5. N-terminal amino acid sequences of total leghemoglobin from *L. corniculatus* and the soybean Lbc_3 protein present in transgenic *Lotus* plants.

moglobins, which comprise up to 20% of total protein in the nodule. Soybean nodules contain four different leghemoglobins—Lba, Lbc_1, Lbc_2, and Lbc_3 (16)—and the corresponding genes have been isolated and characterized (14, 25).

This study describes the transfer and organ-specific expression of a complete soybean leghemoglobin gene in *L. corniculatus*. Expression of the lbc_3 gene in transgenic plants is regulated at the level of RNA, in agreement with results obtained using a chimeric lbc_3 *5'3'-CAT* gene (4). Processing of the corresponding precursor RNA molecules in the transgenic plants apparently proceeds through the same splicing intermediates as in soybean. However, the rate of processing may differ between the two plant species since different concentrations of splicing intermediates were observed in the RNA transfer blot analysis. The high level of soybean Lbc_3 protein present in transgenic root nodules indicates that Lbc_3 is stable and that the lbc_3 gene is expressed at a level comparable to that found in soybean, where Lbc_3 represents ≈15% of the Lb fraction. Although gene dosage and position effects might influence the quantitative results, it is concluded that the functional expression of the soybean lbc_3 gene in the *L. corniculatus* background is quite efficient. The comparable amounts of Lbc_3 protein in the HH and HX transformed lines indicate that 1 kb of the lbc_3 5′ promoter region is sufficient for root nodule-specific expression. Since the level of RNA was not determined, the similar amounts of Lbc_3 protein might also be due to posttranscriptional control mechanisms. *L. corniculatus* root nodules contain four dominant red/brown proteins visible in isoelectric focusing gels of the pH 4–5.5 range. We have positively identified the polypeptide migrating as the soybean Lbc_1 as a leghemoglobin, Lb_1 with the N-terminal sequence shown in Fig. 5. The other three proteins are most likely leghemoglobins, indicating together with the result of the Southern analysis (Fig. 1) that the *Lotus* Lbs are encoded in the *Lotus* genome as a small family of genes.

We thank Astrid Kühle, Dorte Abildsten, and Ida Thøgersen for excellent technical assistance. This research was supported by European Economic Community Contract BAP-0173-DK, the Danish FTU Programme for Research and Technology, De Danske Sukkerfabrikker A/S, and Aarhus Universitets Forskningsfond.

1. Herrera-Estrella, L., Van den Broeck, G., Maenhaut, R., Van Montagu, M., Schell, J., Timko, M. & Cashmore, A. (1984) *Nature (London)* **310,** 115–120.
2. Kaulen, H., Schell, J. & Kreuzaler, F. (1986) *EMBO J.* **5,** 1–8.
3. Ellis, J. G., Llewellyn, D. J., Dennis, E. S. & Peacock, W. J. (1987) *EMBO J.* **6,** 11–16.
4. Stougaard, J., Marcker, K. A., Otten, L. & Schell, J. (1986) *Nature (London)* **321,** 669–674.
5. Keith, B. & Chua, N.-H. (1986) *EMBO J.* **5,** 2419–2425.
6. Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M. A. & Matzke, A. J. M. (1986) *Plant Mol. Biol.* **6,** 347–357.
7. Schöffl, F. & Baumann, G. (1985) *EMBO J.* **4,** 1119–1124.
8. Matzke, M. A., Susani, M., Binns, A. N., Lewis, E. D., Rubenstein, I. & Matzke, A. J. M. (1984) *EMBO J.* **3,** 1525–1531.
9. Lamppa, G., Nagy, F. & Chua, N.-H. (1985) *Nature (London)* **316,** 750–752.
10. Nagy, F., Morelli, G., Fraley, R. T., Rogers, S. G. & Chua, N.-H. (1985) *EMBO J.* **4,** 3063–3068.
11. Fluhr, R. & Chua, N.-H. (1986) *Proc. Natl. Acad. Sci, USA* **83,** 2358–2362.
12. Sengupta-Gopalan, C., Reichert, N. A., Barker, R. F., Hall, T. C. & Kemp, J. D. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3320–3324.
13. Beachy, R. N., Chen, Z.-L., Horsch, R. B., Rogers, S. G., Hoffmann, N. J. & Fraley, R. T. (1985) *EMBO J.* **4,** 3047–3053.
14. Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. Ø., Paludan, K. & Marcker, K. A. (1982) *Nucleic Acids Res.* **10,** 3487–3494.
15. Marcker, A., Lund, M., Jensen, E. Ø. & Marcker, K. A. (1984) *EMBO J.* **3,** 1691–1695.
16. Fuchsman, W. H. & Appleby, C. A. (1979) *Biochim. Biophys. Acta* **579,** 314–324.
17. Bojsen, K., Abildsten, D., Jensen, E. Ø., Paludan, K. & Marcker, K. A. (1983) *EMBO J.* **2,** 1165–1168.
18. Stougaard, J., Abildsten, D. & Marcker, K. A. (1987) *Mol. Gen. Genet.* **207,** 251–255.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Van Haute, E., Joos, H., Maes, M., Warren, G., Van Montagu, M. & Schell, J. (1983) *EMBO J.* **2,** 411–417.
21. Petit, A., Stougaard, J., Kühle, A., Marcker K. A. & Tempé, J. (1987) *Mol. Gen. Genet.* **207,** 245–250.
22. Hunkapiller, M. W., Lujan, E., Ostrander, F. & Hood, L. E. (1983) *Methods Enzymol.* **91,** pp. 227–236.
23. Legocki, R. P. & Verma, D. P. S. (1980) *Cell* **20,** 153–163.
24. Govers, F., Gloudemans, T., Moerman, M., Van Kammen, A. & Bisseling, T. (1985) *EMBO J.* **4,** 861–867.
25. Hyldig-Nielsen, J. J., Jensen, E. Ø., Paludan, K., Wiborg, O., Garrett, R., Jørgensen, P. & Marcker, K. A. (1982) *Nucleic Acids Res.* **10,** 689–701.