

# Isolation and characterization of a light-inducible, organ-specific gene from potato and analysis of its expression after tagging and transfer into tobacco and potato shoots

Peter Eckes, Sabine Rosahl, Jeff Schell, and Lothar Willmitzer

Max Planck-Institut für Züchtungsforschung, D-5000 Köln 30, Federal Republic of Germany

**Summary.** The isolation and analysis of several cDNAs and of one genomic clone encoding ST-LS1, a single-copy gene from *Solanum tuberosum* with leaf/stem-specific, light-inducible expression, is described. The structure of the gene was determined by sequencing several overlapping partial cDNA clones as well as the genomic clone and by determining the transcription start site by RNA protection experiments. A “tagged” derivative of the gene, obtained by exon modification, was reintroduced into potato and into tobacco shoots using *Agrobacterium*/Ti-plasmid vector systems. The modified gene was expressed in both tobacco and potato shoots giving rise to an RNA of approximately 1,200 nucleotides which exceeded the length of the RNA made from the endogenous gene by the expected size of the “tag” (470 nucleotides). The level of expression of the modified gene varied substantially between independent transformants. A high proportion of the transformants (20%–40%) synthesized as much RNA from the added gene as from the resident gene. Expression of the transferred gene was light induced. Qualitatively and quantitatively the expression of the introduced gene was similar in a homologous (potato) and in a heterologous, but related, cellular background (tobacco).

**Key words:** *Agrobacterium* – Light induction – Position effect – Transformed plants – Gene isolation

## Introduction

The Ti-plasmids of *Agrobacterium tumefaciens* have been developed into an experimental system for the introduction of defined DNA sequences into plant cells. Chimaeric genes have been constructed combining bacterial coding sequences with different regulatory sequences. Until now most of these gene constructions have relied for their expression on Ti-plasmid-derived expression signals, such as the *nos* promoter (Bevan et al. 1983; Fraley et al. 1983; Herrera-Estrella et al. 1983a, b; De Block et al. 1984) or virus-derived sequences such as the promoters from the cauliflower mosaic virus (CaMV) directing the transcription of the 19 S or 35 S RNAs (Odell et al. 1985).

Gene transfers have been used to dissect functionally the DNA sequences involved in the regulation of expression of plant genes; thus, the promoter of the Ti-plasmid-en-

coded nopaline synthase gene (Shaw et al. 1984) and the CaMV-encoded promoter for the 35 S RNA (Odell et al. 1985) have been analysed in some detail. However, only a limited number of genes of plant origin have been studied thus far. Two of these encode storage proteins from bean (Sengupta-Gopalan et al. 1985) and soybean (Beachy et al. 1985). These were introduced into heterologous environments such as tobacco or petunia plants. In the case of the bean storage protein phaseolin and the soybean storage proteins, regulated (seed-specific) expression was observed, albeit at a lower level when compared to the expression of the resident gene in bean seeds (Beachy et al. 1985; Sengupta-Gopalan et al. 1985).

Another group of genes which has been retransferred to plants and analysed for expression comprises the small subunit of the ribulose biphosphate carboxylase (ssRubisco) from pea and the chlorophyll a/b binding protein isolated from pea, wheat or petunia. These genes have been reported to have light-inducible expression in transgenic plants and transformed callus culture (Broglie et al. 1984; Herrera-Estrella et al. 1984; Morelli et al. 1985; Lamppa et al. 1985; Simpson et al. 1985).

One common observation in these experiments as well as in an experiment involving a heat-shock gene (Schoeffl and Baumann 1985) is that the level of expression of the transferred alien genes was lower than that of the same gene in its endogenous situation. Another common feature was that the genes were transferred into a heterologous host, the only exception being an experiment where a construction using the 5' upstream region of a petunia gene encoding chlorophyll a/b binding protein was transformed back into petunia (Jones et al. 1985). Thus most of these experiments do not allow firm conclusions to be drawn about the reason(s) underlying the low level of expression of genes introduced into transgenic plants. The situation is further complicated by the fact that all the plant genes cited above are from multigene families.

Genes from potato displaying differential expression have recently been isolated (Eckes et al. 1985; Rosahl et al. 1986a, b; Sanchez-Serrano et al. 1986). As potato is a suitable host for *Agrobacterium*-mediated gene transfer experiments, derivatives of these genes can be introduced into both homologous (i.e. potato) as well as heterologous (e.g. tobacco) hosts allowing the assessment of the influence of host-specific factors on gene expression.

Here we describe the isolation and physical structure of a gene (ST-LS1) from potato (*Solanum tuberosum*),

which displays leaf/stem-specific, light-dependent and chloroplast-associated expression, controlled mainly at the level of transcription (Eckes et al. 1985). Experiments described below indicate that this gene is present as a single copy per haploid genome. It is therefore likely that the cloned gene is derived from a genomic copy which is actively expressed in the plant. This gene was "tagged" by exon modification (insertion of a DNA sequence into one of the exons of the gene), reintroduced into potato and tobacco shoots and its expression was analysed.

## Materials and methods

**Recombinant DNA techniques.** Standard procedures were used for recombinant DNA work (Maniatis et al. 1982).

**Isolation and analysis of nucleic acids.** Isolation of RNA and DNA from plants and their subsequent analysis by blot hybridization was performed according to Eckes et al. (1985).

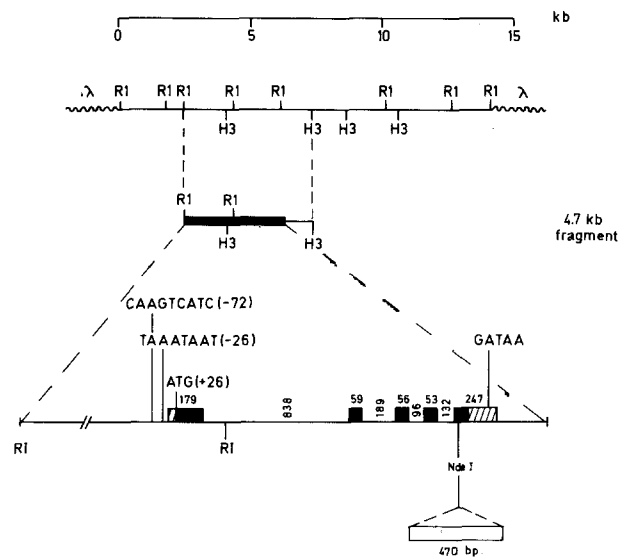
**Transfer of ST-LS1/470 into potato and tobacco shoots.** ST-LS1/470 was mobilized into the *Agrobacterium* acceptor strain C58C1 containing the Ti-plasmid pGV3851 (Zambryski et al. 1983) using triparental mating according to van Haute et al. (1983). Transformation of tobacco was performed using the leaf disc procedure of Horsch et al. (1985). For transformation of potato, surface-sterilised tuber slices (3 cm thickness) were put onto water agar (0.9%) containing 1 mg/l indole acetic acid and infected with C58C1 (pGV3851::ST-LS1/470). After incubation for 2 days at 23°C, the tuber slices were transferred to MS medium (Murashige and Skoog 1962) containing indole acetic acid (1 mg/l) and Claforam (Cefotaxime; 0.5 mg/ml). Shoots appearing were cut from the tuber disc and put on MS medium without hormones in the presence of 0.5 mg/ml Cefotaxime. All plant tissues were kept at 23°C using a dark/light rhythm of 8 h dark and 16 h light.

**RNA protection experiments.** Radioactively labelled RNA homologous to the *HindIII-EcoRI* or *HindIII-XbaI* fragments (cf. Fig. 3) was synthesized using appropriate SP6 vectors and SP6 RNA polymerase (Melton et al. 1984). Hybridization to total RNA (50 µg) isolated from leaves or tubers of potato was for 16 h at 45°C in a buffer containing 80% formamide, 40 mM Pipes, 0.4 M NaCl, 1 mM EDTA, pH 6.7. After hybridization single-stranded RNA was digested in an RNase digestion buffer containing 10 mM Tris, pH 7.5, 5 mM EDTA, 300 mM NaCl, 20 µg/ml RNase A and 1 µg/ml RNase T1 for 30 min at 30°C. Subsequently nucleic acids were precipitated and separated on a 6% polyacrylamide sequencing gel.

## Results

### Isolation and physical characterization of the ST-LS1 gene

Among several cDNA clones isolated from potato leaf poly(A)<sup>+</sup> RNA displaying leaf/stem-specific, light-inducible and chloroplast-associated expression (Eckes et al. 1985) the clone pCL 700 was chosen for further analysis. This decision was based mainly on the fact that pCL 700 hybridizes predominantly to a single band in a genomic blot



**Fig. 1.** Upper part, map of the lambda genomic clone gl700/5. Middle part, restriction map of the 4.7 kb *EcoRI-HindIII* fragment subcloned for sequencing purposes and used in gene transfer experiments. Lower part, structure of the ST-LS1 gene and its modified derivative, ST-LS1/470. Introns are indicated by thin lines, the number above the exons and introns indicate the size of the respective fragments. The *EcoRI* and *NdeI* recognition sites are indicated. The figure also illustrates the modified ST-LS1/470 gene which was obtained by inserting a 470 bp *PvuII* fragment of the TL-DNA of the octopine Ti-plasmid pTi ACH5 (Gielen et al. 1984) into the single *NdeI* site of the last exon. The *PvuII* fragment was first modified by adding an *NdeI* linker. R1, *EcoRI*; H3, *HindIII*

(Eckes et al. 1985) which makes it likely that only one copy of this gene is present per haploid genome.

**Analysis of cDNA clones.** A collection of cDNA clones derived from potato leaf poly(A)<sup>+</sup> RNA was screened for additional cDNA clones homologous to pCL 700. The screening of 5,000 colonies yielded 7 new cDNA clones, all of which were probably derived from the same gene since no nucleotide exchanges were observed between them by sequence analysis (data not shown). Analysis of two clones containing part of the poly(A) tail however, indicated that two different polyadenylation sites can be used (cf. Fig. 2). The use of more than one polyadenylation site has been observed for several other plant genes (Dean et al. 1986). The composite nucleotide sequence obtained from different cDNA clones is given in Fig. 2.

**Isolation and characterization of a genomic clone.** A genomic library established in the lambda phage replacement vector EMBL4 (Frischauf et al. 1983) from the nuclear genome of the haploid potato line AM 80/5793 (Rosahl et al. 1986b) was analysed using the partial cDNA clone pCL 700 as a probe. Screening of about  $8 \times 10^5$  plaques (representing about 4 genome complexities) yielded 5 lambda clones hybridizing strongly to the pCL 700 insert. Analysis of these clones by restriction and Southern blotting experiments indicated that all clones are derived from the same region of the potato genome (data not shown) which again is in agreement with the assumption that ST-LS1 represents a single copy gene.

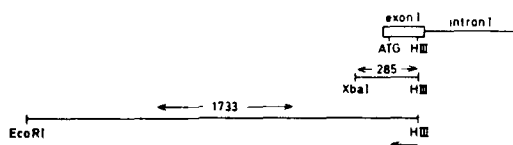
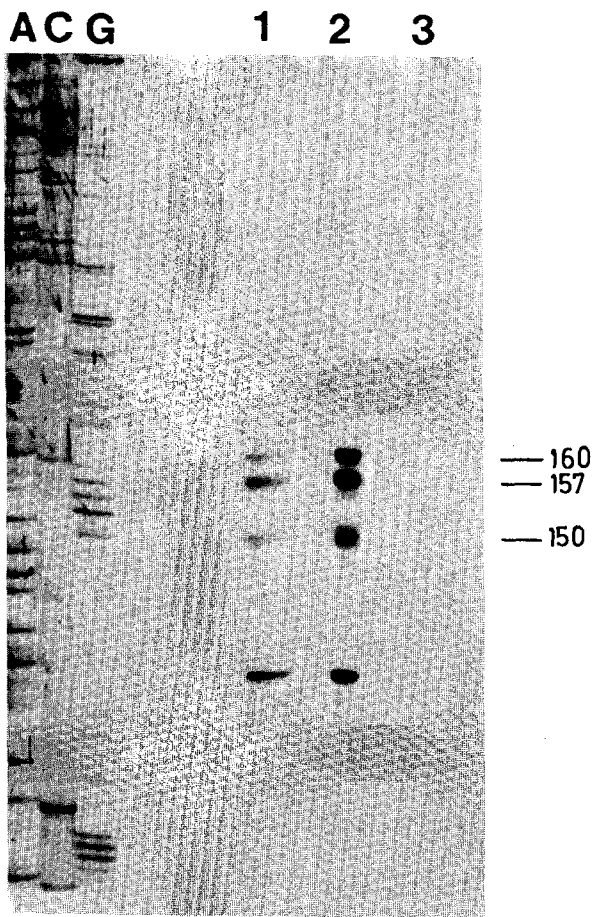
A partial restriction map of one of the lambda clones, gl700/5, is shown in Fig. 1. The 4.7 kb *EcoRI-HindIII* frag-

```

CCAAAGTTTTTGTGTTTCTATCCAGCGATGCTCTGCTACTGAAACATAATGCATATGGGGGGGGGGGGTGTGTTGATCTGTGGACAAGCTAATA 100
ATATCTAATTTGTGTCTGAATCACAATTTCCCATGATAAATAAAGCCAAAATTCGGCTTCTCACAACCTGATAGAATCTAATTTATGACTTTTATCA 200
AAAAAAGAAGAAGATAGAAATCTAATTTATGACATTTAATGGTACCATTCCITAAAAATTCGGATTCAACTAATCTTGGAGTTCATTTTCTCATAG 300
CTGCATTCATGCTGGACTTACTGTAATAGGTTGGATTCTACATCTTGGGAGAGTGATCTGCATTCACGCCCTCTGAAGCACAGGAGTCAGGTG 400
CATTTCAAATGATTACCCAATACAGACTATGGCAGCCGACCCATTAGATTAAATGGGAAAGAAACATCACTCCAATATTTTGAATCTATATA 500
TGGCTATCATACCATAGTACTGTATAGTACTGCTCCAGAAAGCAAAAATGTTGTTTTAGAAATTAATCTGGTTGCAACCGTATCAGTGTCTTGAACATT 600
ATCAAGGCCAAAATACATAATGGTCCCTGAAATCAACTCTAGTTTATTGAAATGATGTCTATTGAGCACCTATATGTGTTGCCAGTATTGCCATGCTTA 700
ACTTAGCATCTGTAAACAGTGTATGCTGGATGTGGAATCGAACATGGAGTTTTTCATGTTGCTTAGAAGTATATGCTTTCAGTGTAGCAGTATCGG 800
AAGGTAATACACAGTTTAAACACCAAAAAGTACAGAGAAATAAGGAGGAATTAATCTTTGGATTTCATTGCTGTTCTGTGGAAATGCTGAACCTGTAA 900
CCAGCTAGGTGAAATGAATCAACTACTTACTTCAAATGTTTAAAGTCCATTTTGGTTAAGAGAAATGAGTTCGCTCTAGCTTTGAAGATGATT 1000
TAGGTTCAAGGACTAATAATTTACAGCAATATAGACCTAGATCCATTGATAGGGGGGGTTCCTAATGTTTTGGCCTTGAATTCGTGTTTACATTGA 1100
TTGACTGAATGAGTGTGCTCTTTCTATATGATCTGGGATAATCTTTTTCTCATGAGCTAACTTAAGTTAGGATATGAGCTAACTGGTCTCCAACAGGC 1200
AACACCATAGAGGAACCAATCACTTTTTTCTTTTGTCTCTGCTTTCGTTGCAGGCTGGACTTAGTAAATAGAAAGGAAAATGGCAAAAATCTATAA 1300
GAAACAAGTCAATCTCTGGTCCCTTAATAATCAAGAAAACCAATGAAATGAACTCCAAAGTCCACAAAATCAAGATAAGGCAAAAGGAAAATAATTTGTA 1400
GTAAGAATGCTATATATTGATAATTTGGTAAATAATATATAGAGAGCAATCCAGAGAAAGGGTCCACCTCTCAACTTGTAAATCACAAATTCCAAGT 1500
CATCACTAATTATTATTAAGGGACCAATGCAAAAGTGAANTAATAATTCATACTAAGTAGTGAGAGCAAAAGAAAGAAAAGCTATAGGAGAAAATGG 1600
M A
CAAGCAGAGTAATGAGCTCAATGAGCCTAAAGCCAACTTTCACCTTTGGAGAAAACATCACTGAAAGGGCTTCCATCACTTCTGATGCTTCTTCTCTT 1700
CAAGCAGAGTAATGAGCTCAATGAGCCTAAAGCCAACTTTCACCTTTGGAGAAAACATCACTGAAAGGGCTTCCATCACTTCTGATGCTTCTTCTCTT
S T V H S S L S L K P T F T L E K T S V K G L P S L A R S S S S F
CAAAGTGTGGCTACTGCTGTCAAGAGCTTAAAGCTGACAAACCTATGGTATGTAATTTTCTATTCTTGGCACCATTAAATTAAGTATAGAGTTAAC 1800
CAAAGTGTGGCTACTGCTGTCAAGAGCTTAAAGCTGACAAACCTATGGTATGTAATTTTCTATTCTTGGCACCATTAAATTAAGTATAGAGTTAAC
K V V A S G V K K L K T D K P Y G
TTAGATCCGAAAGCAATATACAAATATTACATTTACTGTAGTTTTAATTTGCTTAGTAGTAATTTATGATCTCCATGTTACTATTAGAATTCCTT 1900
GTGATTAAGTTTTAAGTCTGACCTTAAATACTATATGAAATATTTTATGTTGCGATGATATATTTGCTCAAACCTTGGAAAAGGTGCATATTT 2000
ACTGATGTAGTAATGACATGACTTAAATCTATCATCTGTGTATTTTGGGGCTATATTTATACGCTTGTGTTAGCAACCGCTCTGATTTTCTCTTGA 2100
TGTTAATTTAGCTCTAGTTTACATGAGAGGATTCAGGTCTCAGAATACTTAAGTAAAAAGTTCAAATTTACCITAGCAACTTTGACTATGGTTAA 2200
CCAATTAATTTAGCTGTAGTCTAATACTCTCAGATGAGCTATGTTAGGGTCAAGGGCTAATAAAGCAATATGTTAACACAAGGATATGAATATCT 2300
GCTAAGTATAGCGGAAGATAAAATTAAGATAATGCTATAGAGAGAGAGAGAGAGAGAGTAGAGGCTGTTTTAATCCTTGTCTCTCAACTCTTIGAT 2400
GAAAGCAGCTCGGTATACCAAGCATCCGATATTAGTAGGATACGAGAAAGGGCCACAGCCTAAGAGATGCTGATGAGACAATCTATCTAATGCAAAATA 2500
TGCTATATATTTGAAAAACACAGATAAAATATTCTAATGCAAAATTAATACTACCTGCTGATTATGAAGTGGGCTTTGTTACTAGGAATTAATGGAA 2600
GAAATTAATGGAA
I N G S
GCATGGCCTTGAAGAGATGGCGTTGATGGCTCAGCCAGGAAGCCAAAGTTTTGTTTCTGCTTCTACCTTTGATATATATAAATTAATGATTAATTAAT 2700
GCATGGCCTTGAAGAGATGGCGTTGATGGCTCAGCCAGGAAGCCAAAGTTTTGTTTCTGCTTCTACCTTTGATATATATAAATTAATGATTAATTAAT
M A L R D G V D A S G R K P K
AGTAATAATAATTTCAAATATTTTTCAAAATAAAGAAATGTAGTATATAGCAATTCCTTTTCTGTAGTTTATAAGTGTATATTTAATTTATAAC 2800
TTTTCTAATATAGCCAAAACATCGTGTATGTTAGGAAAAGGGTGTACCAATATGTTGACAAATATGGAGCTAATGTTGATGGATACAGGTTTGTTA 2900
GGAAAGGGTGTACCAATATGTTGACAAATATGGAGCTAA.GTTGATGGATACAG
G K G V Y Q Y V D K Y G A N V D G Y S
AAAACCTCTAAATTTGTTAATGTAATGCTTCACTAATAAGAAAATGAAACAAATAATCTAATCTTTGTGTGATTTTGAAAATGCAGTCCCATCTACAA 3000
TCCCATCTACAA
P I Y N
CACAGATGAGTGTGCTCCAAAGTGGTGTATGCTATGTTGGAGGTATGACAAATTTACTCGAATCTCCCTTTTAACTCGAACTATGATATACACAACAAG 3100
CACAGATGAGTGTGCTCCAAAGTGGTGTATGCTATGTTGGAGGTATGACAAATTTACTCGAATCTCCCTTTTAACTCGAACTATGATATACACAACAAG
T D E M S P S G D V Y V G G
TTAATAATTAAGCTGACTCATTTTGAATCTACTGACTCTAGATCCTGATTCACACATGTAATAAATGCGAGTACTACTGGCTTAGCCATATGGGGCC 3200
GTCCACTGGCTTAGCCATATGGGGCC
T T G L A T M A V
TGACCTGGTTGGTATCTTCCAGCAAGTGTCTCTCTGTTGTACAAACAAGTCTTTGGCAGATAGATGTTATCTCT.GTGTGTACTATTTAAAGTTG 3300
TGACCTGGTTGGTATCTTCCAGCAAGTGTCTCTCTGTTGTACAAACAAGTCTTTGGCAGATAGATGTTATCTCTGTTACTATTTAAAGTTG
T L V G I L A G G A L L V Y N S A L A Q
TATTCAGCTGGCAGTATCTAATGTTTTTCAAAAATCTTCAAAAATTTCTTGGCATCTTGATTAATATGTAATGCTGTGCTTTACTGTGGATAAT 3400
TATTC.....CATGTAATTTTTTCAAAAATCTTCAAAAATTTCTTGGCATCTTGATTAATATGTAATGCTGTGCTTTACTGTGGATAAT
TACACCTAATTTGCTGTTTAAAGTCTTATAAATCTTAATCTTTAATTTAAGTCTGATTGCTTCCATGAGACTCCACACAGTTATAGGCTAATC 3500
TACACCTAATTTGCTGTTT
CAAGATTTGAAGTGAAGGTTCCGCAAGCATGCAAAATTAATACAAATAATATATGAGTTCACAGCAAAATTTAGTAGATCTTAAATGATAGATA 3600
GCTAATGGGATAATAATACCAAGCAACTAGATGGTACTTCAATGCAAAATGTTACAAAGCAAAATCTTCATTTGCTCTGCGCAAGAAATCATCAAAA 3700

```

**Fig. 2.** Nucleotide sequence and deduced amino acid sequences of the ST-LS1 coding region of gl700/5 and comparison with the composite nucleotide sequence of several cDNAs homologous to pL 700. Consensus sequences (CAT box, TATA box and polyadenylation signals) are *underlined*. The transcriptional start point as determined by RNA protection experiments is indicated by an *arrow* as are the two polyadenylation sites. Motifs homologous to the enhancer core sequence are *boxed*



**Fig. 3.** Determination of the transcriptional start site by RNA protection experiments. The lower part of the figure shows the region of clone gl700/5 containing the first exon and part of the first intron of the ST-LS1 gene. For the RNA protection experiments an RNA homologous to the *HindIII-EcoRI* or *HindIII-XbaI* fragments was synthesized using appropriate SP6 vectors and SP6 RNA polymerase (Melton et al. 1984). After hybridization to potato RNA, the hybrids were treated with RNase A and RNase T1 and separated on a 6% acrylamide sequencing gel using the sequence reactions of a defined clone as size markers (lanes A, C, G); the sizes of the three largest protected fragments are given in nucleotides. Lane 1, leaf poly(A)<sup>+</sup> RNA from potato protected by an RNA synthesized from the *HindIII-EcoRI* fragment; lane 2, leaf poly(A)<sup>+</sup> RNA from potato protected by an RNA synthesized from the *HindIII-XbaI* fragment; lane 3, tuber poly(A)<sup>+</sup> RNA from potato protected by an RNA synthesized from the *HindIII-XbaI* fragment

ment containing the complete coding sequence of ST-LS1 as well as 5' upstream and 3' downstream sequences was subcloned and characterized by sequencing. This 4.7 kb fragment also contains a 3.0 kb *HindIII* fragment which is responsible for the prominent hybridization signal seen in genomic hybridization blots using pcL 700 as a probe (Eckes et al. 1985). For all gene transfer experiments, the

4.7 kb *EcoRI-HindIII* clone was used after being tagged by exon modification (cf. Fig. 1 and below).

The sequence of the genomic ST-LS1 gene with about 1,600 nucleotides in front and 300 nucleotides downstream of the coding region is shown in Fig. 2 with the corresponding data obtained for the cDNA clones.

**Determination of the transcription start site.** Northern blot analysis using different fragments of the lambda clone gl700/5 indicated that the 5'-end of the RNA coding part of the gene is located in the 1.7 kb *EcoRI-HindIII* fragment X (cf. Fig. 1). In order to determine the transcription start site more accurately, RNA protection experiments were performed using the SP6 system (Melton et al. 1984). The results are shown in Fig. 3. The longest fragment protected has a size of 160 nucleotides and defines a transcription initiation site as indicated in Fig. 2. In addition to this long fragment there are other minor fragments of shorter length which are probably due to premature termination of the SP6 polymerase (Melton et al. 1984). No protected RNA is seen when RNA from potato tubers is used for hybridization (lane 3) thus demonstrating the specificity of the assay.

**Structure of the ST-LS1 gene.** Comparison of genomic and cDNA sequences (cf. Fig. 2) reveals the presence of four introns of varying size (cf. Fig. 1 for a more schematic view of the gene). The exon-intron junctions are all in agreement with the GT-AG rule. The longest open reading frame allowed by the exons starts with the ATG at position +26 and ends at position +439 which would result in a protein of 138 amino acids. Whether this sequence actually encodes a protein in plant cells is not known. The deduced amino acid sequence does not show any similarity to published protein sequences. The 3' untranslated region of the gene contains two sequences which can serve as poly(A) addition sites. In both cases the poly(A) addition site is preceded directly by four T residues. In addition the sequences GATTAA and GATAA are found 24 and 26 nucleotides in front of the poly(A) addition sites: these sequences have a high homology to the consensus sequence GATAA or TATAA found 20–30 nucleotides in front of the poly(A) addition sites in most plant or animal genes (Messing et al. 1983). In the 5' upstream region of the gene the sequences TAAATAAT (TATA box) and CAAGTCATC (CAT box) are found at positions -26 and -72. In addition the 5' upstream region contains four motifs of seven nucleotides which have a striking homology to the enhancer core sequence (cf. Fig. 2) first found in animal retroviruses (Weiher et al. 1983).

One unexpected finding is the fairly high number of differences between the nucleotide sequences of the cDNAs and of the genomic clone. Indeed nine nucleotide exchanges and an insertion of ten nucleotides are observed in the genomic clone relative to the cDNA sequence. This number is far too high to be explained by errors due to the reverse transcriptase during cDNA cloning. We have no explanation for this observation and realize that it to some extent contradicts the evidence listed above indicating that ST-LS1 is a single copy gene. However functional analysis (see below) shows that this gene is highly expressed in transformed shoots thus clearly proving that the isolated genomic clone represents an active gene. One possible explanation might be that the cDNAs were isolated from a diploid potato plant whereas the genomic clone was isolated from a hap-

loid potato plant. The haploid line is derived from a recent crossing involving the diploid line as one of the parents; it is therefore possible that the cDNAs originated from a different allele of the ST-LS1 gene to the genomic clone. A similar unexpected high number of differences between a genomic and a cDNA clone has recently been described for the sucrose synthase gene from maize (Werr et al. 1985).

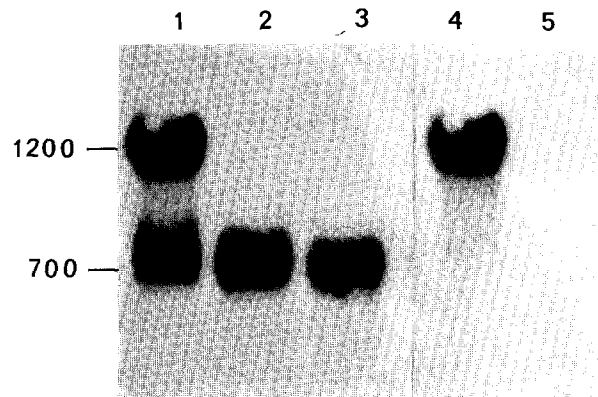
*Analysis of the expression of a modified derivative of the ST-LS1 gene in transformed potato and tobacco shoots*

**Tagging of the ST-LS1 gene.** One of the main problems with regard to the assay of the expression of a defined gene in an homologous host is the distinction between the expression of the transferred gene and that of the resident gene already present in the plant. In most cases this problem is solved by constructing a chimaeric gene consisting of regulatory sequences of the gene to be analysed and a coding region for a readily assayable product such as the chloramphenicol acetyltransferase (Gorman et al. 1982).

A different strategy was used in the experiments described here. The ST-LS1 gene was tagged by inserting, via *NdeI* linkers, a 470 bp *PvuII* fragment of gene 2 of the T-DNA of the octopine Ti-plasmid pTi ACH 5 (cf. Gielen et al. 1984) into the unique *NdeI* site of the last exon (Fig. 1). This DNA segment was chosen because it does not show any homology to potato or tobacco DNA (data not shown). In addition the RNA made from the coding strand of this fragment can be maintained stably in plant cells as a part of RNA 2 of octopine tumours (Willmitzer et al. 1983). Successful expression of the chimaeric gene in transformed plant tissue should result in a chimaeric RNA 470 nucleotides longer than the RNA made from the resident gene. This chimaeric RNA can be specifically probed for using the 470 bp fragment inserted into the gene. On the other hand both the RNA transcribed from the resident gene and that from the transferred gene can be assayed simultaneously using part of the coding region of the gene as a probe. Finally, this exon modification would not affect potential regulatory sequences in the 5' or 3' vicinity of the coding region or in the intact introns (Gillies et al. 1983).

**Introduction of the modified gene into potato and tobacco.** In order to transfer the modified ST-LS1/470 gene into plants it was first inserted into the Ti-plasmid pGV3851 (Zambryski et al. 1983) by homologous recombination. The presence of the modified ST-LS1/470 gene in the T-DNA region was confirmed by Southern blot analysis of total *Agrobacterium* DNA (data not shown). Subsequently discs of potato tuber or leaf discs of tobacco were infected with this *Agrobacterium* strain. Shoots appearing as a consequence of the action of gene 4 in pGV-3851 (Joos et al. 1983) were screened for nopaline synthase activity. In each case several independently obtained shoots were processed for expression analysis.

**The modified ST-LS1/470 gene produces an abundant RNA of about 1,200 nucleotides in transformed potato shoots.** To test whether the chimaeric construction ST-LS1/470 was expressed in transformed potato shoots, RNA was isolated from two transformed and two untransformed potato shoots, pooled and analysed for transcripts homologous

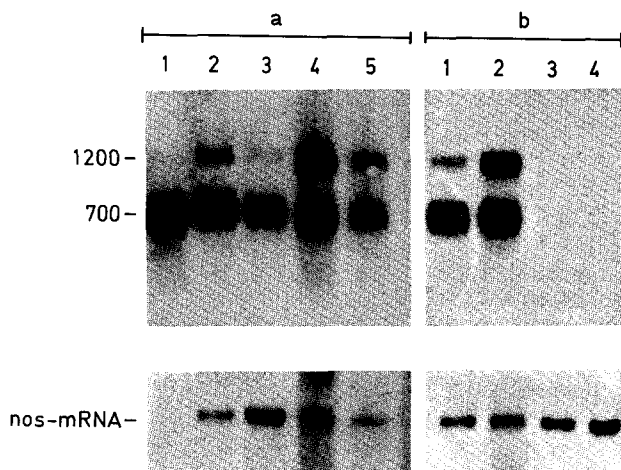


**Fig. 4.** Northern blot analysis of transformed and untransformed potato shoots for the presence of ST-LS1- and ST-LS1/470-derived fragments. Poly(A)<sup>+</sup> RNA (2 µg) isolated from two potato shoots transformed with ST-LS1/470 (lanes 1, 4), or with a derivative in which the 5'-terminal *EcoRI-HindIII* part of the gene had been deleted (lane 3), and from untransformed potato shoots (lanes 2, 5) was separated on 1.5% denaturing formaldehyde gels. Hybridization was against either the 470 bp fragment (lanes 4, 5) or a cDNA containing part of the coding region of the ST-LS1 gene (lanes 1-3) after transfer onto nitrocellulose filters

to the 470 bp fragment in a northern blot experiment (Fig. 4). Using the 470 bp fragment as a probe a band of approximately 1,200 nucleotides is visible only with RNA isolated from the ST-LS1/470 transformed potato shoots (Fig. 4, lane 4). No RNA hybridizing to the 470 bp fragment is visible in lane 5 containing RNA isolated from an untransformed control. The size of the RNA hybridizing to the 470 bp fragment indicates that this RNA is made from the chimaeric gene and is correctly spliced and processed. In addition the production of this RNA is dependent on 5' upstream regions of the ST-LS1 gene, since the deletion of the 1.7 kb *EcoRI-HindIII* fragment X containing the 5' upstream sequences as well as the 5' proximal part of the coding region (cf. Fig. 1) does not lead to the formation of chimaeric RNA in transformed potato shoots (Fig. 4, lane 3).

Figure 4 (lanes 1-3) shows the autoradiogram of the same Northern blot after hybridization with a clone containing part of the coding region of the ST-LS1 gene. In all three lanes an RNA of 700 nucleotides (Eckes et al. 1985) derived from the resident gene is observed. These data therefore demonstrate that a chimaeric RNA is synthesized from the modified ST-LS1/470 gene in transformed potato shoots.

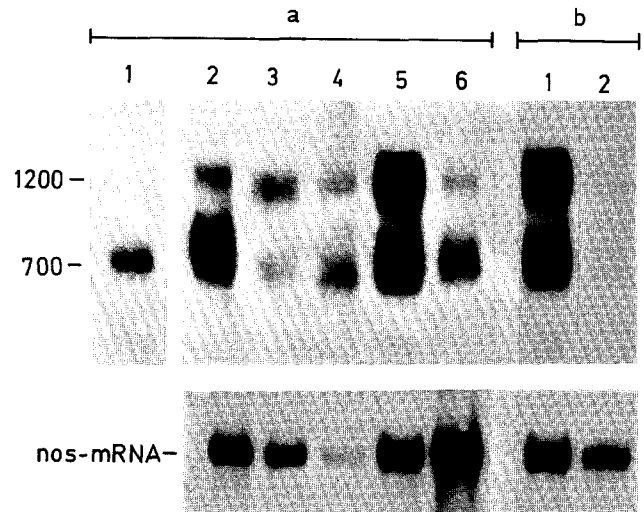
**The relative expression of the transferred ST-LS1/470 gene as compared to the expression of the endogenous gene varies significantly in independent transformants.** The experimental results shown in Fig. 4 were obtained by pooling the RNA from two independently transformed potato shoots. In order to determine whether the level of expression varies in independent transformants, RNA from five independently transformed potato shoots was analysed by northern blot experiments. The data are shown in Fig. 5a, upper part. All transformants display two RNAs differing in size by about 470 nucleotides when probed with the coding sequence of the ST-LS1 gene. However the ratio of the amounts of both RNAs varies significantly, from about equal (Fig. 5a, lane 4) to differences of 50- to 100-fold



**Fig. 5. a** Variability in the level of expression of the transferred ST-LS1/470 gene compared to the expression of the resident gene in independent transgenic potato shoots. Total RNA (100 µg) isolated from leaves of five independent ST-LS1/470 transformants (lanes 1–5) was separated by gel electrophoresis and transferred onto nitrocellulose. It was analysed for RNA derived from the resident ST-LS1 gene and from the added ST-LS1/470 derivative (upper part) by using a partial cDNA clone containing part of the coding region of ST-LS1 as a probe. The lower part of the figure shows the same blot after probing for RNA derived from the nopaline synthase gene. **b** Expression of ST-LS1/470 is light inducible in transgenic potato shoots. Total RNA (100 µg) isolated from transgenic ST-LS1/470 shoots of potato was probed for ST-LS1- or ST-LS1/470-derived RNA (cf. a). RNA was isolated from clonal shoots kept for 10 days in the dark (lanes 3, 4, upper part) or put back into light for 2 days (lanes 1, 2, upper part). The same blot was subsequently probed for RNA derived from the co-transferred nopaline synthase gene (lower part)

(Fig. 5a, lanes 1 and 3). Southern blot analysis of the DNA of the transformed potato shoots showed no correlation between the level of expression and the copy number of the transferred genes (data not shown). Another explanation for this variation is random integration of transferred genes, putting them under the influence of different neighbouring plant sequences. If this were the only reason for the variation, the level of expression of genes cotransferred with the ST-LS1/470 gene, such as the nopaline synthase gene carried by the vector pGV3851, should vary in the same manner. In order to test this, the same RNA blot was hybridized to a probe specific for the nopaline synthase gene. The results shown in Fig. 5a, lower part, demonstrate that the relative expression of the nopaline synthase gene does not always parallel that of the ST-LS1/470 gene. This therefore indicates that position effect cannot be an exclusive explanation for the variation observed.

*Expression of the ST-LS1/470 gene is light dependent.* The above results demonstrate that the DNA sequences contained in ST-LS1/470 are sufficient to determine high-level expression of the transferred gene in transgenic potato shoots. It has been shown previously that expression of the resident gene is light-dependent, chloroplast-associated and organ (leaf/stem)-specific (Eckes et al. 1985). To test whether the expression of the modified gene is similarly controlled, the influence of light was analysed. RNA was isolated from several independent transformants which had been kept in the dark for 5 days (Fig. 5b, upper part, lanes



**Fig. 6. a** Variation in the level of expression of the ST-LS1/470 gene in leaves of transformed tobacco shoots. Total RNA (100 µg) isolated from leaves of non-transformed tobacco plants (lane 1) and from five independent clonal transgenic tobacco shoots (lanes 2–6) was analysed for expression of the resident tobacco gene homologous to ST-LS1 as well as for the expression of the modified potato gene ST-LS1/470 (upper part) as in Fig. 5a. The lower part shows the same blot after hybridizing with a nopaline synthase probe. **b** Expression of ST-LS1/470 is light inducible in transgenic tobacco shoots. Total RNA (100 µg) isolated from leaves of an ST-LS1/470-transformed tobacco plant was probed for expression of the resident gene homologous to ST-LS1 and ST-LS1/470 as described in Fig. 5. Lane 2 (upper part), RNA isolated from the plant kept in the dark for 10 days; lane 1 (upper part), RNA isolated from leaves of the same plant after being put back into light for 2 days. The lower part of lanes 1 and 2 shows the same blot after hybridizing with a nopaline synthase probe

3, 4) and subsequently put back into the light for 2 days (Fig. 5b, lanes 1, 2). The expression of ST-LS1/470 is clearly light dependent in the same manner as that of the endogenous gene. The specificity of the light induction for both of these genes is demonstrated by the fact that the expression of the nopaline synthase gene is not influenced by light as measured by RNA blot (cf. Fig. 5b, lower part).

*High-level, light-dependent expression of ST-LS1/470 is also observed in tobacco, a heterologous plant.* In order to analyse whether the expression of the modified ST-LS1/470 gene is influenced by the host plant, experiments similar to those described above were performed using transgenic tobacco shoots. Using the coding sequence of ST-LS1 as a probe, a single RNA of about 700 nucleotides was found in leaves of untransformed tobacco (Fig. 6a, lane 1), indicating that a gene homologous to ST-LS1 is also expressed in tobacco.

RNA isolated from leaves of transgenic tobacco shoots contains two species with molecular weights of 700 and 1,200 nucleotides hybridizing to a probe for the coding region of ST-LS1 (Fig. 6a, lanes 2–6). The larger RNA species again hybridizes specifically to the 470 bp fragment (data not shown) indicating that the transferred ST-LS1/470 gene is active in transgenic tobacco shoots. Figure 6a, lanes 2–6, shows the hybridization obtained with five independently derived transgenic tobacco shoots using the coding sequence of ST-LS1 (Fig. 6a, upper part) or of the nopaline synthase gene (Fig. 6a, lower part) as the probe. Two conclusions can be drawn from these data: (1) The hybridiza-



tion intensity obtained with the RNA made from the ST-LS1/470 gene in transgenic tobacco shoots can be as high as the intensity obtained with the RNA from the resident gene (cf. Fig. 6a, lanes 3 and 5). (2) The relative levels of expression of the transferred and the resident gene vary considerably (Fig. 6a, lanes 2–6) and are not linked to the expression of the nopaline synthase gene. As was the case in potato, expression of ST-LS1/470 in tobacco is light dependent (Fig. 6b) and is induced simultaneously with the equivalent resident gene. In contrast no influence of light was observed for the expression of the nopaline synthase gene.

## Discussion

ST-LS1, a single-copy gene from potato, which displays a leaf/stem-specific and light-inducible pattern of expression, was isolated from a genomic library and its physical structure was determined by comparison of the sequences of different cDNAs with that of a genomic clone. The transcriptional start site was determined by RNA protection experiments. ST-LS1 is a split gene composed of five exons. The open reading frame could code for a protein of 138 amino acids. The 5' upstream region of the gene contains motifs homologous to the TATA and CAT box consensus sequences. In addition four sequence elements resembling enhancer core structures were found. Analysis of different cDNAs showed that two different polyadenylation sites are used.

A tagged derivative of the gene was reintroduced into both a heterologous (tobacco) and a homologous (potato) plant using *Agrobacterium* Ti-plasmid based vectors. The type of tagging chosen, i.e. the insertion of a foreign piece of DNA into an exon, allows a direct comparison between the expression of the newly introduced gene and that of the equivalent resident gene in the same plant, thus avoiding difficulties arising from differences in the physiological state of different explants. Furthermore this method is not biased by any preconceived notion of the location of regulatory sequences. The results show that the transferred segment of genomic DNA contains all the information necessary for a high level of expression as well as for light inducibility. The introduction of the 470 bp fragment into the last exon resulted in the production of a chimaeric RNA of about 1,200 nucleotides. Since the size of the chimaeric RNA is in agreement with the expected increase of 470 nucleotides we conclude that the chimaeric gene must have been correctly spliced and processed. That the expression of the newly introduced gene was actually driven by its own promoter was indicated by the fact that a deletion of the promoter region and of part of the first exon resulted in a complete disappearance of this chimaeric RNA (cf. Fig. 4, lane 3).

One of the most important results is the high level of expression of the reintroduced gene as compared to the resident gene. Thus in three out of eight independently transformed potato shoots, and in three out of six cases analysed of independent transgenic tobacco shoots, the amount of RNA transcribed from the added gene was about as high as that from the resident gene (cf. Figs. 4–6). In experiments reported previously (see Introduction) chimaeric genes were reintroduced in relatively distant species (e.g. from pea into tobacco or petunia, Broglie et al. 1984; Herrera-Estrella et al. 1984; Morelli et al. 1985; or from

bean into tobacco, Sengupta-Gopalan et al. 1985). In these cases the level of expression of the introduced gene was lower when compared to that of the equivalent endogenous gene. One instance of introduction of a gene into its homologous plant, using a chlorophyll a/b gene from petunia, has been reported (Jones et al. 1985). It was found that in some transgenic plants the level of expression of the added gene reached nearly half the level of the resident gene. However, the proportion of transgenic plants expressing the introduced gene at a high level relative to those expressing at a low level was 1 out of 20.

The following explanations could account for our results. First of all these constructions differ from other constructions in that the modification of the gene was kept as small as possible, thereby leaving intact all other possible regulatory regions which might be located in introns (cf. Gillies et al. 1983). These regions might be important for transcriptional as well as for post-transcriptional processes. Since we analysed steady-state RNA we cannot differentiate between these possibilities.

The presence of sequences other than the 5' upstream signals however cannot be the only explanation for the observed high level of expression. Thus when a completely unmodified ssRubsico gene from pea was transferred to petunia, a 50-fold drop in the level of expression was observed (Broglie et al. 1984). This experiment therefore excludes the possibility that the presence of the intact gene as such is sufficient to get high level expression and points to the possible role played by the host in determining the level of expression of introduced genes. However the data reported in this paper do not show any significant difference in the level of expression of the same gene introduced in either an homologous (potato) or a heterologous (tobacco) host. No significant influence of the host was also reported for the expression of a chimaeric chlorophyll a/b gene from petunia after transfer into petunia and tobacco (Jones et al. 1985). Thus it seems that at least two factors are important, i.e. maximum conservation of the gene structure as well as evolutionary relatedness between the plant from which the gene originates and the plant serving as a host for the transferred gene.

Due to the fact that the expression of the new gene can be directly compared to the expression of the resident gene on the very same RNA blot, physiologically determined differences in the level of expression in different independent transgenic plants are normalized. Thus the observed variation in the relative level of expression of the added versus the resident gene must be due to other factors. Since all transformants analysed contain two to five copies of the added gene as estimated by Southern blot analysis/reconstruction experiments (data not shown), the variation in levels of expression cannot be due simply to different copy numbers. Therefore the more likely explanation for the observed variation is that the random integration of foreign DNA will bring the chimaeric genes into different chromosomal environments which can influence their expression. The existence of a position effect is well known from studies in *Drosophila* (Spradling and Rubin 1983). However the observation that the relative expression of a second linked and cotransferred gene (nopaline synthase) does not always parallel the variation of the ST-LS1/470 gene, although both genes are separated by only 6 kb, argues against a position effect as the sole factor responsible for the observed variation.

As outlined in the Introduction, ST-LS1 was isolated using a strategy which should yield (preferentially) single-copy genes which are under strict developmental control. In addition to being under developmental control this gene is also light inducible. As shown in Figs. 5 and 6 the transferred segment of DNA carrying ST-LS1 also carries all the necessary signals conferring light inducibility on its expression. This trait has previously been observed to be associated with 5' upstream regions for several ssRubisco genes (Herrera-Estrella et al. 1984; Morelli et al. 1985) as well as for the chalcone synthase gene (Kaulen et al. 1986) and a gene encoding chlorophyll *a/b* binding protein from pea (Simpson et al. 1985). The fact that for ST-LS1 light inducibility is again controlled by sequences close to the structural gene makes it likely that this might be a general phenomenon of light inducibility. As a final point in our analysis we compared the level of expression of the chimaeric gene in a homologous (potato) versus a heterologous background (tobacco). The data indicate that in these two related plants there is no significant difference in either the quantitative or the qualitative level of expression.

*Acknowledgements.* Our special thanks go to Astrid Blau for her very careful plant tissue culture work. We thank Jutta Freitag for editing the manuscript and M. Kalda for photographic work. This work was supported by a grant from the BMFT of the German Federal Republic (Molekulare Zell- und Gentechnologie 035/2).

## References

- Beachy RN, Chen ZL, Horsch RB, Rogers SG, Hoffmann NJ, Fraley RT (1985) Accumulation and assembly of soybean  $\beta$ -conglycinin in seeds of transformed petunia plants. *EMBO J* 4:3047–3053
- Bevan M, Flavell RB, Chilton MD (1983) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* 304:184–187
- Brogliè R, Coruzzi G, Fraley RT, Rogers SG, Horsch RB, Niedermeyer JG, Fink CL, Flick JS, Chua NH (1984) Light-regulated expression of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene in transformed plant cells. *Science* 224:838–843
- De Block M, Herrera-Estrella L, Van Montagu M, Schell J, Zambryski P (1984) Expression of foreign genes in regenerated plants and in their progeny. *EMBO J* 3:1681–1684
- Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H, Bedbrook J (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo. *Nucleic Acids Res* 14:2229–2240
- Eckes P, Schell J, Willmitzer L (1985) Organ-specific expression of three leaf/stem specific cDNA's from potato is regulated by light and correlated with chloroplast development. *Mol Gen Genet* 199:216–224
- Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC (1983) Expression of bacterial genes in plant cells. *Proc Natl Acad Sci USA* 80:4803–4807
- Frischauf AM, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* 170:827–842
- Gielen J, De Beuckeleer M, Seurinck J, Deboeck F, De Greve H, Lemmers M, Van Montagu M, Schell J (1984) The complete nucleotide sequence of the TL-DNA of the *Agrobacterium tumefaciens* plasmid pTiAch5. *EMBO J* 3:835–846
- Gillies SD, Morrison SL, Oi YT, Tonegawa S (1983) A tissue specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717–728
- Gorman CM, Moffat CF, Howard BA (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol Cell Biol* 2:1044–1051
- Herrera-Estrella L, Depicker A, Van Montagu M, Schell J (1983a) Expression of chimaeric genes transferred into plant cells using a Ti-plasmid derived vector. *Nature* 303:209–213
- Herrera-Estrella L, De Block M, Messens E, Hernalsteens JP, Van Montagu M, Schell J (1983b) Chimaeric genes as dominant selectable markers in plant cells. *EMBO J* 2:987–995
- Herrera-Estrella L, Van den Broeck G, Maenhaut R, Van Montagu M, Schell J, Timko M, Cashmore A (1984) Light-inducible and chloroplast-associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a Ti-plasmid vector. *Nature* 310:115–120
- Horsch RB, Fry FE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. *Science* 227:1229–1231
- Jones J, Dunsmuir P, Bedbrook J (1985) High level expression of introduced chimaeric genes in regenerated transformed plants. *EMBO J* 4:2411–2418
- Joos H, Inzew D, Caplan A, Sormann M, Van Montagu M, Schell J (1983) Genetic analysis of T-DNA transcripts in nopaline crown galls. *Cell* 32:1057–1067
- Kaulen H, Schell J, Kreuzaler F (1986) Light-induced expression of the chimeric chalcone synthase NPTII gene in tobacco cells. *EMBO J* 5:1–8
- Lamppa G, Nagy F, Chua NH (1985) Light-regulated and organ-specific expression of a wheat Cab gene in transgenic tobacco. *Nature* 316:750–752
- Maniatis T, Fritsch E, Sambrook J (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Press, New York, USA
- Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage sp6 promoter. *Nucleic Acids Res* 12:7035–7056
- Messing J, Geraghty D, Heidecker G, Hu NT, Kridl J, Rubenstein J (1983) Plant gene structure. In: Kosuge T, Meredith CP, Hollaender A (eds) Genetic engineering of plants – an agricultural perspective. Plenum Press, NY, pp 211–227
- Morelli G, Nagy F, Fraley RT, Rogers RG, Chua NH (1985) A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose-1,5-bisphosphate carboxylase small subunits of pea. *Nature* 315:200–204
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol Plant* 15:473–497
- Odell JT, Nagy F, Chua NH (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35 S promoter. *Nature* 313:810–812
- Rosahl S, Eckes P, Schell J, Willmitzer L (1986a) Organ-specific gene expression in potato: isolation and characterization of tuber-specific cDNA sequences. *Mol Gen Genet* 202:368–373
- Rosahl S, Schmidt R, Schell J, Willmitzer L (1986b) Isolation and characterization of a gene from *Solanum tuberosum* encoding patatin, the major storage protein of potato tubers. *Mol Gen Genet* 203:214–220
- Sanchez-Serrano J, Schmidt R, Schell J, Willmitzer L (1986) Nucleotide sequence of proteinase inhibitor II encoding cDNA of potato and its mode of expression. *Mol Gen Genet* 203:15–20
- Schoeffl F, Baumann G (1985) Thermo-induced transcripts of a soybean heat shock gene after transfer into sunflower using a Ti-plasmid vector. *EMBO J* 4:1119–1124
- Sengupta-Gopalan C, Reichert NA, Barker RF, Hall TC, Kemp JD (1985) Developmentally regulated expression of the bean phaseolin gene in tobacco seed. *Proc Natl Acad Sci USA* 82:3320–3324
- Shaw CH, Carter GH, Watson MD, Shaw CH (1984) A functional map of the nopaline synthase promoter. *Nucleic Acids Res* 12:7831–7846



- Simpson J, Timko MP, Cashmore AR, Schell J, Van Montagu M, Herrera-Estrella L (1985) Light-inducible and tissue-specific expression of a chimaeric gene under control of the 5'-flanking sequence of a pea chlorophyll a/b binding protein gene. *EMBO J* 4:2723-2729
- Spradling AC, Rubin GM (1983) The effect of chromosome position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* 34:47-57
- Van Haute E, Joos H, Maes M, Warren G, Van Montagu M, Schell J (1983) Intergenic transfer and exchange recombination of restriction fragments cloned in pBR322: a novel strategy for the reversed genetics of the Ti-plasmid of *Agrobacterium tumefaciens*. *EMBO J* 2:411-418
- Weiber H, Koenig M, Gruss P (1983) Multiple point mutations affect the SV40 enhancer. *Science* 219:626-631
- Werr W, Frommer WB, Maas C, Starlinger P (1985) Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. *EMBO J* 4:1373-1380
- Willmitzer L, Dhaese P, Schreier P, Schmalenbach W, Van Montagu M, Schell J (1983) Size, location and polarity of T-DNA encoded transcripts in nopaline crown gall tumors, common transcripts in octopine and nopaline tumors. *Cell* 32:1045-1056
- Zambryski P, Joos H, Genetello C, Leemans J, Van Montagu M, Schell J (1983) Ti-plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J* 2:2143-2150

Communicated by P. Starlinger

Received March 28, 1986