The EMBO Journal vol.9 no.5 pp.1337–1346, 1990

Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*

Csaba Koncz[1,2], Reinhold Mayerhofer[1], Zsuzsanna Koncz-Kalman[1], Christiane Nawrath[1], Bernd Reiss[1], George P.Redei[3] and Jeff Schell[1]

[1]Max-Planck-Institut für Züchtungsforschung, D-5000 Köln 30, Carl-von-Linne-Weg 10, FRG, [2]Institute of Plant Physiology, Biological Research Center of Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521, Hungary and [3]Department of Agronomy, University of Missouri, Curtiss Hall, Columbia, MO 65211, USA

Communicated by J.Schell

A recessive *pale* mutation, designated as *cs*, was identified by transferred-DNA (T-DNA)-mediated insertional mutagenesis in *Arabidopsis thaliana*. The *pale* mutation, cosegregating with the hygromycin resistance marker of the T-DNA, was mapped to the position of the *ch*-42 (*chlorata*) locus on chromosome 4. Lack of genetic complementation between *cs* and *ch*-42 mutants indicated allelism. Plant boundaries of the T-DNA insert rescued from the *pale* mutant were used as probes for the isolation of genomic and full-length cDNA clones of the wild-type *cs* gene. Transformation of the *pale* mutant with T-DNA vectors carrying these clones resulted in a normal green phenotype, thus demonstrating positive complementation of the T-DNA induced mutation. DNA sequence comparison of the *cs* mutant and its wild-type allele revealed that the T-DNA insertion occurred 11 bp upstream of the stop codon. A fusion protein, seven amino acids longer than its wild-type counterpart of M_r 46 251, is therefore synthesized in the *pale* mutant. Transcript analysis during dark–light transition, *in vitro* protein transport assay, and the absence of DNA sequence homology between *cs* and known genes indicates that the light regulated expression of the *cs* gene results in the synthesis of a novel chloroplast protein.

Key words: Arabidopsis/*ch*-42 (*chlorata*) locus/genetic complementation/insertional mutagenesis/protein transport to chloroplast/T-DNA rescue/transformation

Introduction

Insertion elements are commonly used as genetic tools and molecular probes for mutagenesis and identification of genes in diverse organisms. In maize and snapdragon, endogenous transposable (controlling) elements were expoited to develop efficient gene tagging systems (Coen *et al.*, 1989; Gierl and Saedler, 1989). Since *Arabidopsis thaliana*, due to its small genome size and excellent genetics, has become a model for plant molecular biology (Redei, 1975; Meyerowitz, 1987; Chang *et al.*, 1988; Finkelstein *et al.*, 1988; Nam *et al.*, 1989), the possible application of gene tagging techniques to this plant was explored. Apart from a family of defective retrotransposons, the *Arabidopsis* genome does not possess active transposable elements (Voytas and Ausubel, 1988). Therefore alternative approaches were initiated using foreign transposons, such as Ac/Ds from maize (Van Sluys *et al.*, 1987; Masterson *et al.*, 1989) or the transferred DNA (T-DNA) of *Agrobacterium* Ti plasmids (Schell, 1987) as mutators.

The T-DNA is a unique insertion element that is integrated into the plant nuclear genome after transfer from agrobacteria (see Zambryski, 1988). To detect and isolate T-DNA insertions in plant genes, gene fusion techniques were developed (Andre *et al.*, 1986; Koncz *et al.*, 1987a). T-DNA constructs carrying a plant selectable marker gene, a bacterial replicon and a promoterless reporter gene linked to the right integration site of the T-DNA were used to induce and identify transcriptional or translational plant gene fusions and to rescue mutated plant genes in *Escherichia coli*. Similar high frequencies of T-DNA-induced gene fusions in diverse plant species indicated that T-DNA is preferentially integrated into potentially active chromosomal loci (Koncz *et al.*, 1989). Genetic evidence for a number of T-DNA-linked mutations in *Arabidopsis* has been reported recently (Feldmann *et al.*, 1989).

Here we describe the isolation via T-DNA tagging and the characterization of a nuclear gene, *ch*-42, that had been defined earlier by X-ray mutagenesis (Fischerova, 1975) and is located on *Arabidopsis* chromosome 4 (Koornneef *et al.*, 1983).

Results

Genetic analysis and mapping of a T-DNA-induced cs (pale) mutation

A total of 450 transgenic *A.thaliana* (var. Columbia) plants were regenerated and grown to maturity following transformation with T-DNA gene fusion vectors pPCV621 and pPCV6NFHyg (Redei *et al.*, 1988; Koncz *et al.*, 1989). Then, 100 T2 progenies of each plant were planted in soil and in hygromycin-containing medium to screen for induced mutations and to follow the inheritance of the T-DNA-encoded hygromycin resistance (Hyg^R) marker.

A pPCV6NFHyg T-DNA-transformed line, N6H-14, produced green and yellow–pale green offsprings in a ratio of 76:24 in soil and of 49:26 on selective medium. The 2000 T3 progenies obtained by selfing of four hygromycin resistant pale plants were all pale and Hyg^R, indicating that pale T2 progenies were homozygous for the mutation designated as *cs*.

In four independent crosses of homozygous *cs* mutants with wild-type plants, 24 063 green and 7918 pale F2 progenies were collected. A homogeneity test of the 3.04:1 segregation ratio ($\chi^2 = 5.228$; $P > 0.15$) confirmed the monogenic inheritance of the recessive *cs* mutation.

To explore chromosomal linkage of the mutation, homozygous pale plants were crossed as male with wild-

© Oxford University Press

type trisomic tester lines (Lee-Chen and Steinitz-Sears, 1967). A 9.9:1 segregation of 365 green and 36 pale plants was observed in combined disomic and trisomic F2 progenies of tester *concave* representing trisome 4. Crosses with other testers yielded ~3:1 segregation ratios. This showed that the *cs* mutation is located on chromosome 4 because a synteny between the mutation and the trisome results in a distorted 35:1 to 8:1 segregation ratio, depending on the transmission rate of disomic male and female gametes and on the frequency of crossing over between the gene and centromere.

The mutation was further mapped in repulsion and coupling to *bp* and *cer*-2 markers on chromosome 4 by identification of recombinant classes in F2 and F3 progenies derived from crosses of Landsberg 'erecta' W100 ms^+ (Koornneef *et al.*, 1986) and Columbia testers with pale plants as male and female. Statistical means of recombination frequencies did not differ significantly ($P = 0.35$) in repulsion and coupling phases. The mapping indicated that *cs* is located close to *cer*-2 within the *bp*−*cer*-2 region, but a significant difference between recombinant fractions observed in Landsberg and Columbia testers (Table I) biased the linkage analysis. This suggested that the experimentally determined linkage intensities may depend on the genetic background of the tester line. Results obtained with the Landsberg tester were therefore compared with the standard linkage map made in the same background (Koornneef *et al.*, 1983; Koornneef, 1987).

Test of allelism and resolution

A significant correlation between *cer*-2−*ch*-42 distance on the standard map and *cer*-2−*cs* distance determined in the W100 ms^+ tester was intriguing because of the similarity of *cs* and *ch*-42 mutant phenotypes. The *ch*-42 X-ray-induced mutation also yields yellow−pale green seedlings, but is homozygous lethal in soil. To test possible allelism, heterozygous *ch*-42 and homozygous *cs* plants were crossed and the F1 progenies were scored according to their colour phenotype on hygromycin selective medium (Figure 1). All 156 offspring tested were hygromycin resistant but split to 79 green and 77 pale. The 1:1 segregation ratio of green and pale F1 progenies confirmed ($P = 0.86$) the allelism and the absence of complementation between *cs* and *ch*-42 mutations, permitting *cs* to be assigned to position 39.4−4 of *ch*-42 on the standard map (Figure 1). Additivity and correlation of linkage intensities obtained for *bp*−*cs*−*cer*-2 distances in both Landsberg and Columbia testers suggested that the actual map distances on the standard map may be higher and need correction.

To test rigorously the separability of the *cs* mutation and the hygromycin resistance marker of the T-DNA, *cs* was crossed in repulsion to *bp* and *cer*-2. Recombinant classes homozygous for *cs* and either of the flanking markers were identified in F2 and tested in F3 for possible segregation of *cs* and Hyg^R traits. Among 7872 individuals representing 414 recombinant families and 621 independent chromosomes, no recombination was observed between *cs* and the T-DNA insert (Table II). The resolution of the mapping, computed either for the experimentally determined or for the standard *bp*−*cer*-2, *cer*-2−*ch*-42 and *ch*-42−*bp* distances, showed that recombinant categories were detected in the experiment with a maximal error of 0.079 map units, equal to ~11 kb (Meyerowitz, 1987), approaching the size of the T-DNA.

Table I. Genetic mapping of the *cs* (*ch*-42) mutation

Interval	W100 ms^+ tester		Columbia testers		
	N	*R*(%)	*N*	*R*(%)	*P*
bp−*cer*-2	4523	41.91 ± 0.73	1728	36.06 ± 1.16	0.997
bp−*cs*	2911	33.75 ± 1.16	2295	28.87 ± 0.95	0.990
cs−*cer*-2	2911	7.50 ± 0.49	2773	9.81 ± 0.56	0.964

	Standard	Map distances (cM) W100 ms^+ tester	Columbia tester
bp−*cer*-2	31.4	60.8	45.5
bp−*cs*/*ch*-42	24.1	50.0	32.9
cer-2−*cs*/*ch*-42	7.3	7.6	9.9

cs was crossed with *bp* and *cer*-2 testers in repulsion and coupling. Recombination frequencies (*R*%) were calculated from the F2 data by the maximum likelihood method and converted to map units (cM) using the Kosambi mapping function. Probability of differences (*P*) between data obtained in W100 ms^+ Landsberg 'erecta' and Columbia testers was computed on the basis of lack of overlaps between the means ± multiples of their standard error.

Table II. Probing the separability of the T-DNA encoded Hyg^R marker and the *cs* mutation by genetic recombination

Interval	Number of recombinant families	Number of F3 progenies tested	Number of Hyg^S-*cs* recombinants
bp−*cs*	352	6688	0
cs−*cer*-2	62	1184	0

	Standard	Resolution (cM) W100 ms^+	Columbia
bp−*cs*/*ch*-42	0.034	0.071	0.047
cer-2−*cs*/*ch*-42	0.058	0.061	0.079

cs was crossed in repulsion with *bp*−*cer*-2. +*cs*+/*bp*+*cer*-2 F1 plants were selfed. Non-parental marker arrangements were scored in F2. From 414 recombinants homozygous for *cs* and for either *bp* or *cer*-2, 27 201 F3 progenies were harvested. On average 20 F3 progenies of each recombinant were tested for hygromycin resistance by germination on selective medium. If *cs* and Hyg^R were separated loci and recombination occurred, four different marker constitutions are possible depending whether the T-DNA insert (Hyg^R) was located to the left or to the right of the *cs* locus. These are *bp* h *cs*+/*bp* h *cs*+ and *bp* H *cs*+/*bp* h *cs*+ or +*cs* h *cer*-2/+*cs* h *cer*-2 and +*cs* h *cer*-2/+*cs* H *cer*-2, where H and h stand for hygromycin resistance and sensitivity, respectively. Since none of the recombinant families were h/h or h/H, both of their chromosomes can be counted as non-recombinants regarding the *cs* and Hyg^R loci. Therefore, the resolution of the mapping was computed cM = map units/(2 × number of recombinant families); for the *bp*−*cs* interval, for example, 32.9/(2 × 352) = 0.047 cM. For comparison, the resolution was also computed using standard map distances and data obtained in the Landsberg 'erecta' tester.

T-DNA rescue and isolation of wild-type cs (ch-42) gene

Using probes derived from the left or right arms of pPCV6NFHyg T-DNA, a single *Hin*dIII fragment of 7.8 kb, and unique *Xba*I, *Bgl*II and *Eco*RI fragments of diverse lengths were detected by Southern DNA hybridization in the nuclear DNA of homozygous *cs* plants (Figure 2). Since there was no *Hin*dIII site within the T-DNA, this indicated that the *cs* mutant carried a single T-DNA insert of 6.4 kb within a 1.4 kb *Hin*dIII genomic DNA fragment.

Following *Hin*dIII digestion and self-ligation of *cs* nuclear

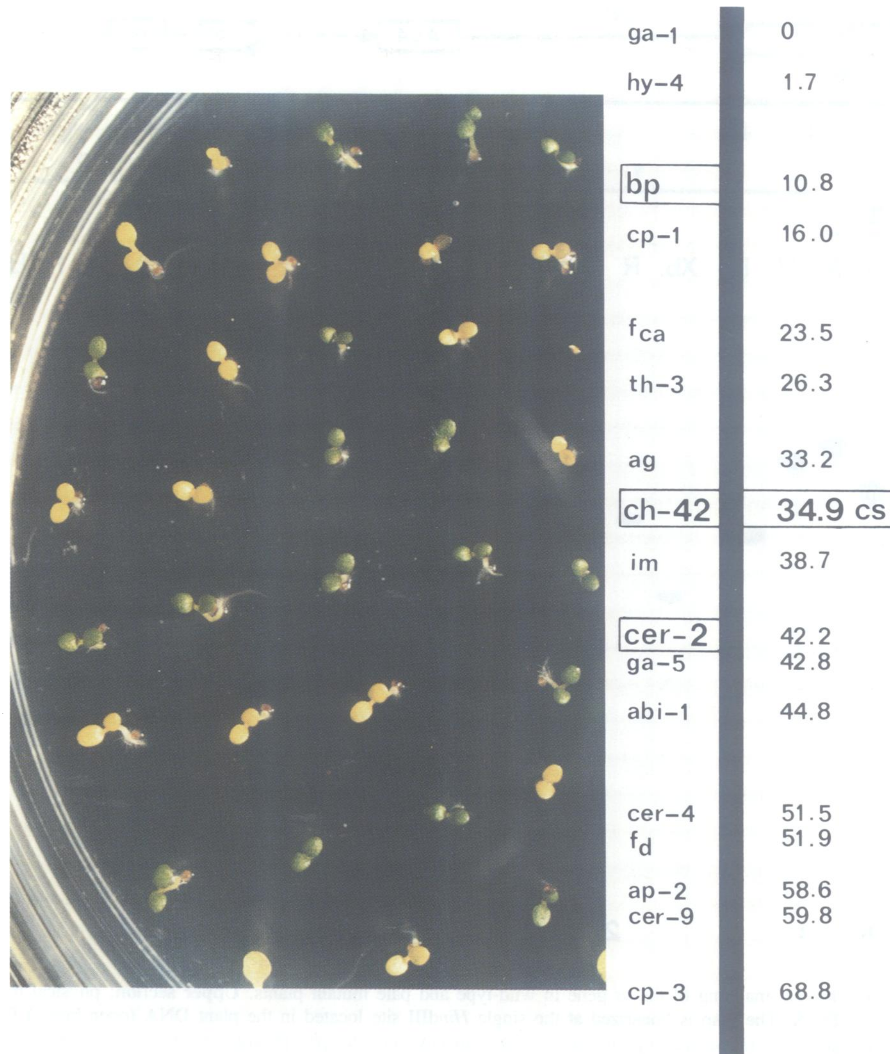

Fig. 1. The 1:1 segregation of green and pale F1 progenies derived from a cross between *ch*-42/+ and *cs*/*cs* plants (**left**). Allelism between *cs* and *ch*-42 allow us to assign *cs* to position 39.4 of *ch*-42 on the linkage map of chromosome 4 (**right**).

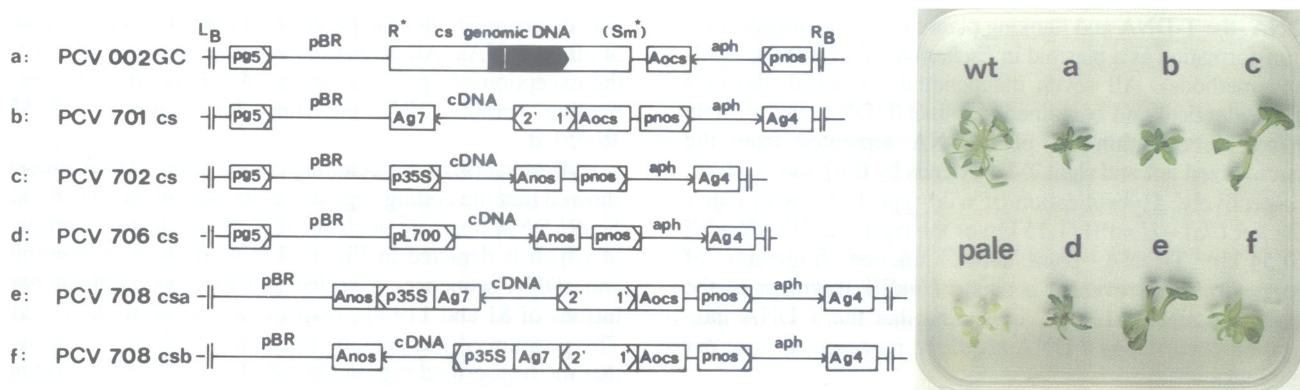

Fig. 5. Complementation of the *cs* mutant by transformation with wild-type genomic and cDNAs. **Left (a–f)**: schematic map of vectors carrying an *Eco*RI–*Sma*I fragment of clone GC7 in pPCV002 and *cs* cDNAs in pPCV701, 702, 706 and 708. Ag4, Ag7, A*osc* and A*nos* are polyadenylation signals of T-DNA genes 4, 7, octopine synthase and nopaline synthase, respectively. p35S, p*nos*, 1′2′, pg5 and pL700 are plant promoters derived from genes of the CaMV 35S RNA, nopaline synthase, mannopine synthase, T-DNA gene 5 and from a potato gene, ST-LS1. Other abbreviations are listed in Figure 3. **Right (a–f)**: Green phenotype of pale plants transformed by T-DNAs of vectors a–f in comparison with wild-type and non-transformed *cs* mutant plants as controls.

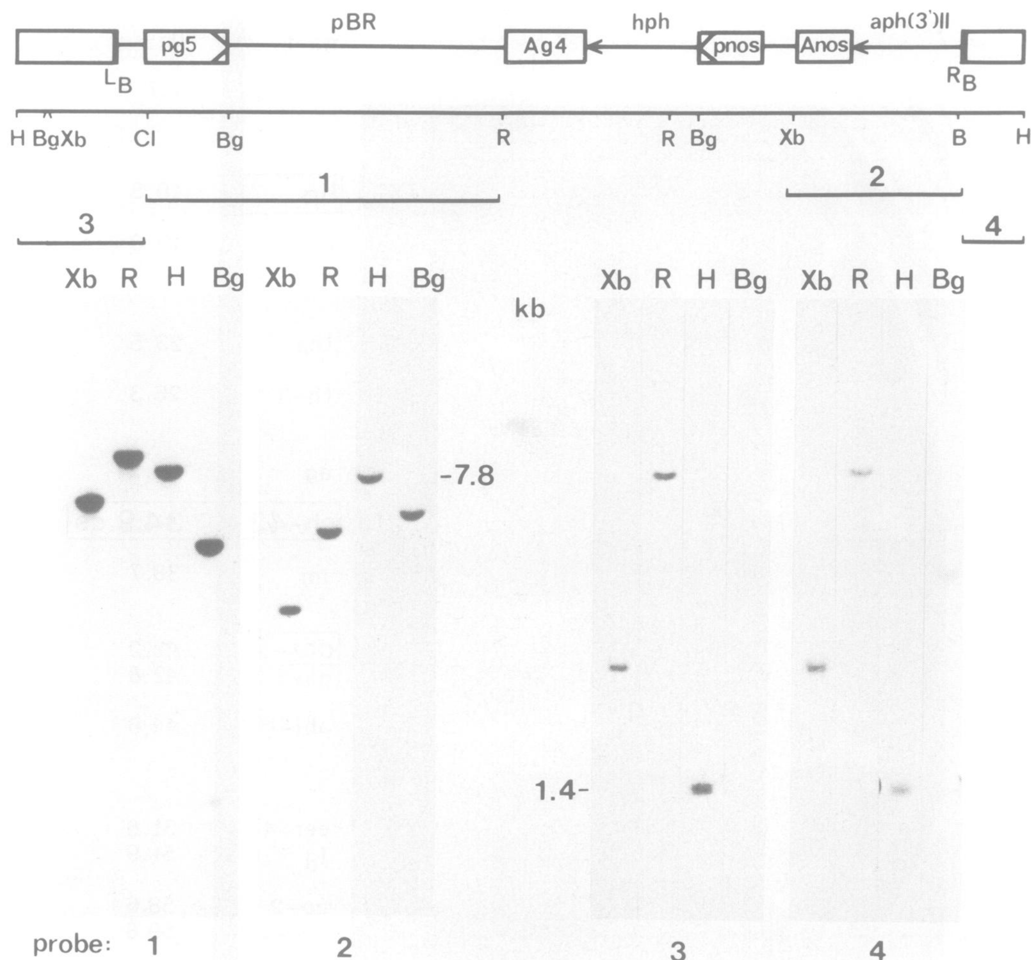

Fig. 2. T-DNA rescue and physical mapping of the *cs* gene in wild-type and pale mutant plants. **Upper section**: physical map of rescued plasmids containing pPCV6NFHyg T-DNA. The map is linearized at the single *Hin*dIII site located in the plant DNA (open boxes) flanking the T-DNA. *aph*(3')II and *hph* are aminoglycoside (kanamycin) phosphotransferase and hygromycin phosphotransfersae genes, respectively. pg5 and p*nos*: promoters of T-DNA gene 5 and nopaline synthase gene; Ag4 and A*nos*: polyadenylation signals of T-DNA gene 4 and *nos*; pBR: pBR322 replicon within the T-DNA; L_B and R_B: left and right boundaries of the T-DNA. Restriction endonuclease cleavage sites: Bg, *Bgl*II; Cl, *Cla*I; R, *Eco*RI; Xb, *Xba*I; B, *Bam*HI. **Lower section**: nuclear DNA prepared from homozygous *cs* mutant (1 and 2) and wild-type (3 and 4) plants was digested with *Xba*I, *Eco*RI, *Hin*dIII and *Bgl*II enzymes and hybridized to probes 1–4 after Southern blotting. Probe 1 and 2: *Cla*I–*Eco*RI and *Xba*I–*Bam*HI fragments representing the left and right arms of the T-DNA. Probes 3 and 4: *Hin*dIII–*Cla*I and *Bam*HI–*Hin*dIII left and right T-DNA–plant DNA junction fragments.

DNA, the T-DNA and flanking plant DNA was isolated after transformation as a plasmid in *Escherichia coli* (see Materials and methods). All seven independently rescued plasmids were identical and contained an intact T-DNA and a single *Hin*dIII site within the plant DNA separated from the circularized left and right T-DNA ends by 0.87 and 0.54 kb, respectively. Hybridization of wild-type DNA with either the left *Cla*I–*Hin*dIII (1.15 kb) or the right *Bam*HI–*Hin*dIII (0.54 kb) T-DNA–plant DNA junction fragments of rescued clones revealed a single *Hin*dIII fragment of the predicted size of 1.4 kb, indicating that the T-DNA integration occurred in a DNA sequence represented once per haploid genome.

Both junction fragments were used as probes to screen for homologous clones in λgt10 cDNA and EMBL 4 genomic DNA libraries made from wild-type plants of Columbia background. Six genomic and six cDNA clones were isolated and found to overlap to various degrees (Figure 3). Following *Eco*RI subcloning of cDNAs from λ into pUC vectors, the nucleotide sequence of the longest cDNA (cp 12) was determined and compared with that of 5' and 3' ends of other cDNAs. All cDNA clones were identical and, with the exception of cp52, carried an ATG codon of an open reading frame (ORF) encoding for a protein of M_r 46 251 d.

Hybridization of the genomic clones to the cDNA probes showed that the coding region was located within a 6.85 kb *Eco*RI DNA fragment of clone GC7, the nucleotide sequence of which is depicted in Figure 4. Comparison of genomic and cDNA sequences revealed that the gene contains two introns of 81 and 119 bp, respectively, close to its 5' end. The position of the poly(A) track in the cDNAs indicated that the transcribed region extends 122 bp 3' of the coding sequence.

T-DNA insertion in cs generated a 3' gene fusion

The T-DNA insert was mapped precisely by comparison of the nucleotide sequences of rescued T DNA–plant DNA junction fragments with that of wild-type genomic and cDNA clones. The DNA sequence analysis showed that in the *cs*

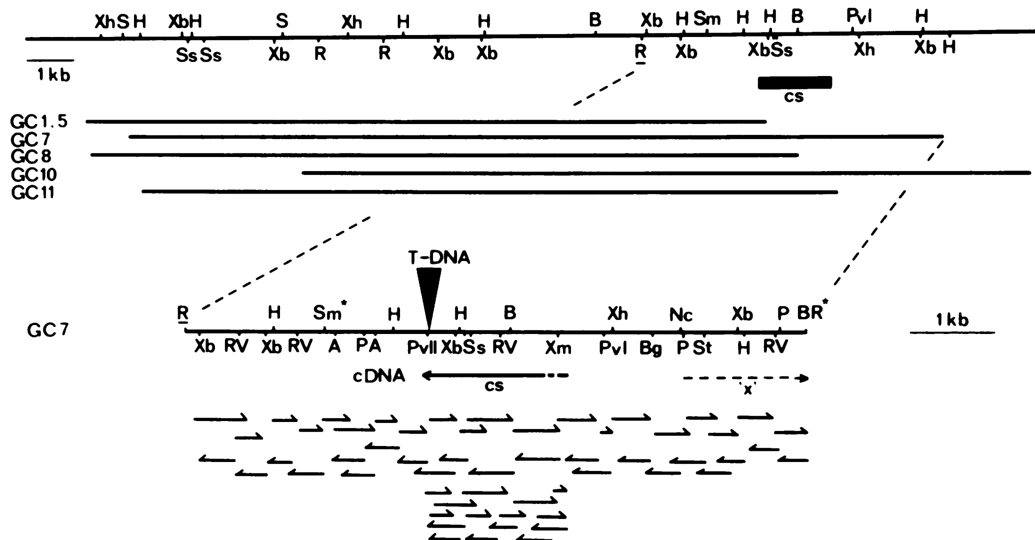

Fig. 3. Physical map of *cs* genomic and cDNA clones. GC1, 5, 7, 8, 10 and 11 are EMBL 4 genomic DNA clones carrying wild-type *cs* DNA sequences. Black bar shows the position of the *cs* gene determined by hybridization of genomic DNA clones with cDNA probes. A detailed map of a 6.85 kb *Eco*RI fragment of clone GC7 and the strategy used for sequencing of genomic and cDNAs are shown in the lower section and described in Materials and methods. The position of T-DNA insertion in the *cs* mutant is indicated by a black arrow. 'x' is ORF 'X' located 5' upstream of gene *cs*. Restriction endonuclease cleavage sites not listed in Figure 2 are: Xh, *Xho*I; S, *Sal*I; Ss, *Sst*I; Sm, *Sma*I; PvI, *Pvu*I; PvII, *Pvu*II; RV, *Eco*RV; A, *Acc*I; P, *Pst*I; Xm, *Xmn*I; Nc, *Nco*I; St, *Stu*I. The *Sma*I–*Eco*RI fragment subcloned in pPCV002 and used for complementation of the *cs* mutation is labeled by stars.

mutant the insertion of the T-DNA caused a deletion of 70 bp which removed 14 bp from the 3' end of the coding sequence. The fusion of the truncated gene to the end of the T-DNA altered the reading frame such that four codons for the C-terminal amino acids of the wild-type protein were replaced by 11 new codons (Figure 4). A DNA sequence of 16 bp located at the T-DNA–plant DNA junction showed homology to neither T-DNA nor the deleted target site. At the T-DNA boundaries 21 bp from the left and 3 bp from the right 25 bp border repeats were retained. The right T-DNA border and the linked promoterless aminoglycoside (kanamycin) phosphotransferase (*aph*(3')II) reporter gene faced the 3' end of the gene. This explains why *aph*(3')II transcripts could not be detected in any part of the *cs* mutant plants throughout their development (data not shown). Since no further divergence between mutant and wild-type DNA sequences was found, the data suggested that the T-DNA-induced 3' gene fusion caused the *cs* mutation.

Genetic complementation of the cs mutation

To demonstrate complementation of the mutant by the wild-type allele, homozygous pale plants were transformed with the isolated genomic and cDNA clones. To delimit the boundaries of the gene, all ORFs were examined in the genomic DNA sequence. A long ORF (ORF'X', Figure 4) was identified 5' upstream of the gene at a distance of 2.0 kb, while a short ORF occurred at ~1.0 kb 3' downstream. This suggested that the entire gene involving the promoter is located within a 5.5 kb *Sma*I–*Eco*RI DNA fragment of genomic clone GC7 (Figure 3). This fragment was therefore inserted in the plant gene vector pPCV002 (Koncz and Schell, 1986). To shorten the length of the poly(A) sequence of the cDNA, a 3' *Pst*I–*Bgl*II fragment of cDNA clone cp12 was replaced by that of clone cp 52. The modified cDNA was inserted into plant gene expression vectors pPCV701, 702, 706 and 708 between the promoter and polyadenylation signal sequences derived from the cauliflower mosaic virus (CaMV)35S RNA gene, and from either the T-DNA-encoded genes 4, 7, *ocs* (octopine synthase), *nos* (nopaline synthase) or *mas* (mannopine synthase), or from a light regulated potato gene, ST-LS1 (Figure 5). Following *Agrobacterium*-mediated transfer of the vector constructs into the pale mutant, 100 kanamycin resistant (Km^R) transformed calli from each experiment were regenerated to plants. Apart from two pPCV002GC, four pPCV702 *cs* and four pPCV706 *cs* transformants, all kanamycin resistant plants were normal green thus demonstrating functional complementation of the T-DNA induced *cs* mutation (Figure 5). F2 generations derived from green plant showed independent segregation of the two different T-DNAs since they produced Km^S pale and Km^R green progenies.

cs is a light-regulated gene and encodes a chloroplast protein

Screening of DNA and protein data banks with the nucleotide or derived amino acid sequences of *cs* clones did not reveal significant homology to known genes or proteins. The observation that the *ch*-42 mutant is lethal in soil but survives in sucrose-containing medium, however, suggested that the gene function is possibly involved in photosynthesis or chloroplast development. Comparison of the ultrastructure and pigment composition of chloroplasts from wild-type and *cs* mutant plants did not show obvious differences although the amount of some proteins (i.e. antenna complex; R.Herrmann, personal communication) was reduced in the thylakoid membranes of the *cs* mutant. To demonstrate that the pale and chlorata phenotypes resulted from mutations in a true photosynthetic gene and were not the result of some pleiotropic effect, the light regulation and chloroplast specificity of *cs* gene expression was explored.

Following exposure of dark-grown wild-type and pale mutant plants to normal light (375 lux), mRNA was purified

and hybridized to an internal *Bam*HI–*Xba*I fragment of cDNA probe cp 12 (Figure 6a). In comparison with low levels in the dark, the amount of transcript increased by 60- to 100-fold after 4 h of illumination. mRNA of similar length (1600 nt) was detected in both wild-type and mutant plants. This indicated that the T-DNA insertion did not abolish the production of polyadenylated transcript in the *cs* mutant. Putative polyadenylation signals (ATTAAA/T) were in fact found downstream of the stop codon in the T-DNA sequence of the 3′ gene fusion at positions −90, −136, −180 and −212.

To synthesize the wild-type transcript, a cDNA carrying a shortened 3′ end (see above) was inserted into SP6 promoter vectors. *In vitro* translation of the transcript resulted in a protein of 46 kd that showed specific binding to isolated chloroplasts in protein transport assays (Figure 6b). Washed or protease treated chloroplasts contained a protein of reduced size indicating that the 46 kd precursor is specifically processed during uptake into the chloroplast.

Discussion

The function of the *cs* (*ch*-42) gene remains to be determined. Preliminary data indicate that the Cs protein is located in the stroma fraction of lysed chloroplasts (B.Reiss, unpublished). Distribution of hydrophobic domains in the predicted protein sequence supports this notion. Apart from the N-terminal transit peptide domain, the only hydrophobic

```
   1 GGATCCAGTGGTAGCTTTTCACTCAAATCTTGTACCTTGGCAGTTTGGCTTGTACGAGTGCCTGGTGATATTTTGCCTGAGAGGGTTGTTAGAGAATGTCCAGCATCTGAGTTATACAGT
     AspLeuProLeuLysGluSerLeuAspGlnValLysAlaThrGlnSerThrArgThrGlyProSerIleLysGlySerLeuThrThrLeuSerHisGlyAlaAspSerAsnTyrLeuA

 121 GCTCCTTTAGTGTTATCCTGTATTTCTGCCTGAGTGTCTGTACCCATTTCTTCCTGTTGATGTTCTATCTTGTCTGAACATAAATGAGATGAGATGCTTGGTGAAGTCTGTTCTATCTTC
     laGlyLysThrAsnAspGlnIleGluAlaGlnThrAspThrGlyMetGluGluGlnGlnHisGluIleLysAspSerCysLeuHisSerSerIleSerProSerThrGlnGluIleLysG

 241 TCATAACTTGATTCCATGGTGGTAATCATTTTCTTGTTGGCATTTTCTGGATCTATAGGACAGTCTTCTTTTGTACCAGACCTGCAGAGTTCATGGATATCTGGACTGGTTTCTCTGTTT
     luTyrSerSerGluMetThrThrIleMetLysLysAsnAlaAsnGluProAspIleProCysAspGluLysThrGlySerArgCysLeuGluHisIleAspProSerThrGluArgAsnV

 361 ACTGCAATATGTGCAATCTCAGAAGCATGTAATGGCGAGGGTGCTATGGTCTCATTTGAAGGACATGGTTCATTCTTTTCTTCAACCTTTGAACATGATGTCTTCAATGGTGAAGTATCT
     alAlaIleHisAlaIleGluSerAlaHisLeuProSerProAlaIleThrGluAsnSerProCysProGluAsnLysGluGluValLysSerCysSerThrLysLeuProSerThrAspV

 481 ACGGATTGAGGTTGCACACACATTGTATGAGTTTGTGAGGCTTTTTTTCTTGAACCCTGAGGTGATTCCACCGCCTCAAAATCTCCCATGCTCTCTAATGCAGCTTGATTCGAGTTTTTG
     alSerGlnProGlnValCysMetThrHisThrGlnSerAlaLysLysArgSerGlyGlnProSerGluValAlaGluPheAspGlyMetSerGluLeuAlaAlaGlnAsnSerAsnLysS

 601 CTCTTATTAACAACCAGAGTATCTGTAGTCTTGTAATTTTGGCTCTCTTTTTCAGCTGAAGCTTTTGCCTGAACATCTGTGTTTGCCTTGGAGGGAACTTGTCTAGTTATTGGTAAGTCT
     erLysAsnValValLeuThrAspThrThrLysTyrAsnGlnSerGluLysGluAlaSerAlaLysAlaGlnValAspThrAsnAlaLysSerProValGlnArgThrIleProLeuAspL

 721 AGACCAAAGTCCAAATCGTCTGACTTATCACTCTTACTACTTGCACAAATTTCTTCACAGTCCATGCTCTTGGTCGTAGTCTTGCTTCCCTTCTTCGGGGGGCTTGAACCAAGATCAAAG
     euGlyPheAspLeuAspAspSerLysAspSerLysSerSerAlaCysIleGluGluCysAspMetSerLysThrThrThrLysSerGlyLysLysProProSerSerGlyLeuAspPheA

 841 TCGTCTAACCTGCTTGCAGAAAAACATAACATGGCGTTAAGCAATTTTAGATGACATAGATAATTCTAAAAGGAAATTGATTTACTGAAGGATTCATACGCATCAAAGTCGTAAGAGAAG
     spAspLeuArgSerAlaSerPheCysLeuMet          ORF "X"

 961 GCAAATGGATTCTTTTTCTGTTTTGAATTTCCACTAGGTTTATCATCAGAGGTCTCTTTCGTTTTTGTGGTTTTCTTGGCTGGGCTTGAGAAATCAAAGTCTGGCATGTCCATTTTAAAT
1081 GATGATCCAAAATCTCCATCCAGATCGAAACCCATATCCCTGAGAAAGAGACAAAGAAACAGAACTATATGAGACACTAGGCCTGTGCCAAGCAACAGAAACTTATAGTAGTAATCATGA
1201 TAAAGGAGGAAAAATTTCACGATTGAACTACATTTATTTAGGTGCAAACATGGTTTAAGAAACTTATCATATTGGTTACACTCACAACTTGAAGGCATTCTTCTTATTTTTTGCTGGTGA
1321 GCCAAAGCCAAAATCCATTGGATCATCATCTGTCATTGACTTGAAGTTCCATGTATCCTTTTCCATGTCTTTATCTGCAGCTAACCATGGAAATTAAAAGACAACGAACCTTTGAATCAC
1441 CCATTTTAACAATGGGAACTACATCTTATGGTTTCTATCTACGCCCAAAAAATTAAAAGCTGGCCAAAATTGAGAAATAGAACAATAAGGAGATAAAGAAGAATAAAAGATAGTGTGCAT
1561 GATCTTTTTGCCTATAGCATCAAGTTATTAAGGCAGTCTACTTGGGGATTCCTTTTCTTCAATGGAATTTTCGCCGCAAGCGGCTTAAAGACATTCCACTAGCAATGCCATGTTAGTTAT
1681 GAATCATATCACATGAACATTTGAACCATACCGAGCATTAGATCTTTGTCCTGAGTATCTGTAGAACTTCTTTTAGCAGGTGCAGACATGGCAATATCTGTTCAGTATCACTCTTGAAAA
1801 CCAACTGCAAAGAAAAAACTACCATCATAATCATTCTAGCACTCACTCACGAAAACACTTATCTGTTATGTTATGATAATCACAAGGCAACCAACAAAAAAATCATCCTACTACGTCACA
1921 AACTCAACTAGCATAGACAAATCATCAGTTCCATTAACACTCTTCTAAAACTATTTCTAAGGCGATATGACGCATAACGTACGTAGGTATGATTATCGGAGAGACTGAGAAACATAACAG
2041 ATCAAACAATGTGATTATCCGAGAACCAATCTAAACGAAGAAGAACAGTAGCAGGAATCAAATTATGAGATTCAACACCAAATTATAAAAAATAACCAGGCGGTTCGTCGATTTTAAGCA
2161 AGGAGACGCCACTCGAGACTAAATTTCAGGGACAACGAGCAGCACGAAACTAAGTTTAGGAGAAAGTGCATCTTAGCATTGATCGAACATTTCTTACTCAAATCTACGCATAAACGTCAC
2281 CTCTAACACAGAAATGTTCATCGATTATGATCAACCGATCGTCAATCGTCGAACCTTAGCAAACCGAAGCTAAAACAACGCCTGACAGTGAGATTCTACTCAATCGACGAGCAACGAGGG
2401 TAAATTCTTACCGATTGAATCGCTTTGCAGTTGTATATGTAGTAGGCGCTAACGATAAACGTTCAACGGCAACAAGACGACGACACCGGAGAGAAAATCGCCGACGGAACTCGAAGGGGG
2521 GAGATTTTGAAATTAGCTGGGCTCCGAATGATTTGACTTGGGCCTTTATATTCTTATGTTTGGTATATATATATAGAAGCTCGCCTTTTTTTTTGGTACGGCTATTTTATTTTCTATATC
2641 ATTTTTGATTTATAAATAAAAAAAAGGAGGAGATAGAGGAGGAAATCTATCTCCTTATCCTGACCACAGAAACAAACATAAGCCAATGGAAACAAAAAAATTCTAAGGTGAAAAAAGCCA

2761 ATCAGAAGCTGAGTAAACTCTAAGCTCACACAAAAATGGCGTCTCTTCTTGGAACATCTTCTTCTGCAATCTGGGCTTCTCCTTCACTCTCTTCTCCTTCCTCAAAACCTTCCTCCTCCC
   1         ------ cs cDNA ----------- MetAlaSerLeuLeuGlyThrSerSerSerAlaIleTrpAlaSerProSerLeuSerSerProSerSerLysProSerSerSerP

2881 CCATTTGCTTCAGGCCAGGTAAACCAGTCTCTCCAAACCTTAAAATCTTACTGTTTTTTCTCAATTCGAAATCATGGTGAATTGAATTTTTGTATACAGGAAAATTGTTTGGAAGCAAGT
  29 roIleCysPheArgProGly ----------------------------- INTRON I ----------------------------------- LysLeuPheGlySerLysL

3001 TAAATGCAGGAATCCAAATAAGGCCAAAGAAGAACAGGTCTCGTTACCATGTTTCGGTTATGAATGTAGCCACTGAAATCAACTCTACTGAACAAGTAAGCATTTTTTGATGATGTTCAC
  42 euAsnAlaGlyIleGlnIleArgProLysLysAsnArgSerArgTyrHisValSerValMetAsnValAlaThrGluIleAsnSerThrGluGln ----------------------

3121 TGAAAGTTTCAGACTTTGCTTGATTTTGAAGTTAGGAACATGATTGTTGTTGGTTTTTTATGTAATGGTGGTTTTGGGTTGTAGGTAGTAGGGAAGTTTGATTCAAAGAAGAGTGCGAGA
  74 ------------------------------- INTRON II -------------------------------------- ValValGlyLysPheAspSerLysLysSerAlaArg
```

```
3241  CCGGTTTATCCATTTGCAGCTATAGTAGGGCAAGATGAGATGAAGTTATGTCTTTTGTTGAATGTTATTGATCCAAAGATTGGTGGTGTTATGATTATGGGAGATAGAGGAACTGGAAAA
  86  ProValTyrProPheAlaAlaIleValGlyGlnAspGluMetLysLeuCysLeuLeuLeuAsnValIleAspProLysIleGlyGlyValMetIleMetGlyAspArgGlyThrGlyLys

3361  TCTACAACTGTTAGATCATTAGTTGATCTGTTACCTGAGATTAATGTAGTTGCAGGTGACCCGTATAACTCGGATCCGATAGATCCTGAGTTTATGGGTGTTGAAGTAAGAGAGAGAGTT
 126  SerThrThrValArgSerLeuValAspLeuLeuProGluIleAsnValValAlaGlyAspProTyrAsnSerAspProIleAspProGluPheMetGlyValGluValArgGluArgVal

3481  GAGAAAGGAGAGCAAGTTCCTGTTATTGCGACTAAGATTAATATGGTTGATCTTCCTTTGGGTGCAACAGAAGATAGAGTTTGTGGAACCATCGATATCGAAAAGGCTTTGACAGAAGGT
 166  GluLysGlyGluGlnValProValIleAlaThrLysIleAsnMetValAspLeuProLeuGlyAlaThrGluAspArgValCysGlyThrIleAspIleGluLysAlaLeuThrGluGly

3601  GTAAAAGCCTTTGAGCCTGGTTTGTTGGCTAAAGCTAATAGAGGGATTCTTTATGTTGATGAAGTTAATCTCTTGGATGATCATTTGGTTGATGTTCTTTTGGATTCAGCTGCTTCTGGT
 206  ValLysAlaPheGluProGlyLeuLeuAlaLysAlaAsnArgGlyIleLeuTyrValAspGluValAsnLeuLeuAspAspHisLeuValAspValLeuLeuAspSerAlaAlaSerGly

3721  TGGAATACGGTTGAGAGAGAAGGGATTTCGATTTCTCACCCGGCGAGGTTTATCTTGATCGGTTCAGGAAATCCGGAAGAAGGAGAGCTTAGGCCACAGCTTCTTGATCGGTTTGGTATG
 246  TrpAsnThrValGluArgGluGlyIleSerIleSerHisProAlaArgPheIleLeuIleGlySerGlyAsnProGluGluGlyGluLeuArgProGlnLeuLeuAspArgPheGlyMet

3841  CATGCACAAGTAGGGACGGTTAGAGATGCTGATTTACGGGTCAAGATTGTTGAAGAGAGAGCTCGTTTCGATAGTAACCCAAAGGATTTCCGTGACACTTACAAAACCGAGCAGGACAAG
 286  HisAlaGlnValGlyThrValArgAspAlaAspLeuArgValLysIleValGluGluArgAlaArgPheAspSerAsnProLysAspPheArgAspThrTyrLysThrGluGlnAspLys

3961  CTTCAAGACCAGATTTCAACTGCTAGGGCAAACCTTTCCTCGGTTCAGATTGATAGGGAATTGAAGGTGAAGATCTCTAGAGTTTGTTCAGAGCTCAATGTTGATGGGTTGAGAGGAGAC
 326  LeuGlnAspGlnIleSerThrAlaArgAlaAsnLeuSerSerValGlnIleAspArgGluLeuLysValLysIleSerArgValCysSerGluLeuAsnValAspGlyLeuArgGlyAsp

4081  ATAGTGACTAACAGAGCAGCAAAAGCACTTGCAGCTCTCAAAGGAAAAGATCGAGTAACTCCAGATGATGTTGCAACCGTTATCCCTAACTGCTTAAGGCACCGTCTGAGGAAAGATCCA
 366  IleValThrAsnArgAlaAlaLysAlaLeuAlaAlaLeuLysGlyLysAspArgValThrProAspAspValAlaThrValIleProAsnCysLeuArgHisArgLeuArgLysAspPro

                                                   GlyGluTyrAspAspSerGlyTyrIleGlnLeuEnd
                                                   GTGAGTATGATGATTCAGGATATATTCAATTGTAA-----  T-DNA left border fusion----
4201  CTGGAATCTATTGATTCAGGAGTTCTAGTTTCCGAGAAGTTCGCCGagattttcagctgaataggagattcgattctgtgatcatttttctttaactttttctactctgtaaacggTGAG
 406  LeuGluSerIleAspSerGlyValLeuValSerGluLysPheAlaGluIlePheSerEnd-----------------------------------------------------------

4321  AAGAACCAAAACTATAAACAGGCAAATCTGATTGAATATAGAGTCTATGCTTTATTCATTAAGTTGAGTTGTTCAAACTTCAAATCATAGACTCAATAGAACTCTGTTTATATTTGTGTA
      -------------------------------------------------------------   c DNA

4441  ACTACATACAAACATTCCTGCATTGATTATGGCTATAAAGCTATCAGAAAACGTTGTCAGGATCAACTGTTACTATATCTTGAAACGCAGGTGATAACTTATTGATTAGAAGGACATAGC

4561  GTTTTGCCGGTTCTGGTAAAAGTGTGGAGCAACTTCTGTCAGCCAAGAAGGATTGATGGTGACGACATTCCGCATGTATTGACGCTCAGTTGAGACAATGGATTGGTAAACCACCCATTT

4681  TGGATTCACCCTGAAGATTGTATAAAGAGTATATATACCAAATAAGCTTTCACGTTCAAAAGAAAAAATCTTCAACTTTTAGATGTTATTGAGGTTTTTGCTAAATCCATAGCAACAAGA

4801  TATACTACTAACCTGAATAATACAGATGACGGATGAATATAAACTTCTTCAGAGCCTCTAATGGTCTTGTATACTCCATTGCTGTGTGGCTAAAACATCGAAAAGAATTAAGATGATGAA

4921  AATCAGATTTCTTTATAGGGCTGAAAGGATGTGCAGAAAAAGAGAACTGAAATGGGTCACTTACTTCTAAACGGCATGCATTGGCAAAAAAACCTGCAGTCACAGCTTTTCTCACAGCCT

5041  GAAACAGTGTAACGTTAGAAACATGCAACCTTTTTTCAAATGAAGCAAAGTGCTACAGTTTCTTACCTCCATGTCTCCGTCACATGATTTCAAGGTGATGCCTAATCTCCGAGCAATCCG

5161  TTTGAGTTGATCTCTGATTTCAACTACTTTTTTCTGTGAAAGGAAACAAAAACAAATTAAATTGGAGAAGGGAAGATAAGGACTACTTGCCTGTTCTGTTATTATACGCACCATGGATTG

5281  GTAGTTGAGGAAATTCTTGTAACACCACTGTGTAGGCTTTTTCGACTCAAGAAACCCTTTGTATACATTCAGGAATGTGACATGGTCACCCTATATAGACAAGAACAATGATTTTACTCA

5401  GTCAAAAAAAGATTTTCAGAGATATAGTTAGCTCAAAGTTATATATCAATACATACCTCTGCGGCTGCAAATCTCAGTTTTGCTTCATCTTGTTCTTTCTGTACTCCCCGGGCAATGATC

5521  CAAACATACTGCATAGAAACATATTCAAGATTGCATCTGTTAAGTTTCAAATCCCATAGTTCAAGTCGAAAGCAAATAATCACAAATTTGCAATATGTTCGGTTGTTCCTTATGTAAATA

5641  AAGAGAAGAAACTTTAAGCAAAGGAAAAAGGGGTAAAGAGCTTTTACTTGGACAGAGAGAACTGCAGCGATTGTAATGATCTCATGTGAACAACCAAGCTCACTTGAAGCCAAAATCATT

5761  TTTGATATCATTGGGTCCTAGCAAAAAATGACGAGCATGAGACATGTAATATGCAGAACAAACTGCAAATGAACATTACATTTTAGATATATAATAGCTCATGGTCAGGAAGAAAAAGTA

5881  CCAGTGGAAGTTCTGCAACTTGGAATCCTGTTGGCGAAGTGAGTTTAGCATCATCATCGAGGATCTGAAGTGAATAAAGTACTTCAAGTGCTCGAATCATTGCTTCAGACGATGGAGGTG

6001  CAGGCCAATCGAAACCCAAGATATTATCTATGCCCAAAGCTTTTAGCTGAAAATACCAAAAAGATCAACAGAGTTTGAGAAATCTAGAGGCAAGGATTCTTCTTTACTGTCTTTTTGTTG

6121  CAAAATCCCATGGAACGAACAGATGTATTTCTCATACAGATATTGCTGATACATCATTGAGAGTAACTTTCTGTGTCGTTTCTTTATTCAGGGCGTTGTCTATGTGATTGATAGTGGTTT

6241  CTCGAAGCAGAAGTTCTATAACCCAGTAAGTCCATCTCCATTTCTTTTGTTTATATATTTTTAGTTCCATCTTTTGTAATATGTTGGACATTTCTAAAGCAACTTACATTTCTTACCCTC

6361  ACAGATTTCAGATATCGAAAGTGTTGAGGAGGCTCCAATATCAAAAGCATCTGCAAGGCAAAGGTCTGGTAGGGCTGGGCGAGTTCGGCCTGGGAAGTGTTACAGGTAACTCTCAATCCG

6481  TTTATTAAAGTACAGAATTTCGTCTGATATAGTTGTTTAGTGCTCATTAATCCTCTTCCTGATATTTTTATGGTAACTAGGCTTTACACCGAAGATTATTTTCTCAACCAAATGCCTGGA

6601  GAAGGCATACCGGAGATGCAGAGGTCGAATCTTGTTTCTACTGTAATACAGGTACTAGACTTGCTCAACTCTTGTCATTTAGTTTTTTTTTATAAGAACTCAAAAGTTGGGAGGCGTTAC

6721  GCAAAACCACACAAGTTTTTTCTTTTCACCTTTTCTATTAAGTAATTAATATCCAGATGAGAGGAAGCTCTTGCTTCTAGA   6801
```

Fig. 4. Nucleotide sequence of wild-type *cs* gene and cDNA. Nucleotide position 1 corresponds to the *Bam*HI site in the polylinker of EMBL 4 vector. The DNA sequence extends to the *Xba*I site that preceeds the *Eco*RI site underlined in Figure 3. The translation start of the *cs* coding region at position 2787 and the polyadenylation site at position 4367 are underlined. Dashed lines below the genomic DNA sequence indicate the ends of cDNA clone cp12. Amino acid sequences obtained by translation of *cs* cDNA and ORF 'X' genomic DNA are shown below the DNA sequence. The position of the T-DNA insertion (4246) and the sequence alteration in the *cs* mutant are depicted above the genomic DNA sequence. A DNA sequence of 16 bp at the T-DNA–plant DNA junction that showed homology to neither T-DNA nor the deleted target site (printed in lower case between positions 4246 and 4317) is underlined.

region is located within the central part of the protein between amino acid residues 120 and 240. Comparison of the transit peptide sequence to that of other chloroplast proteins, such as Cab or SSU of *Arabidopsis* (Leutwiler *et al.*, 1986; Krebbes *et al.*, 1988), did not reveal striking homologies although they all contained a similar MAS . . . SS/A N-terminal sequence. The viability of the *cs* mutant in soil and the polyadenylation of mutant transcript suggest that in the mutant a C-terminal fusion protein is synthesised, the last 11 terminal amino acids of which are encoded by T-DNA sequences. Further assays should determine whether the transport or the activity of the mutant protein is reduced by the altered C-terminal end. In contrast to the 'leaky' *cs* mutation, *ch*-42 is presumably a 'null' mutant resulting in lethality in the absence of sucrose. Since the *ch*-42 mutation can also be complemented by transformation with wild-type

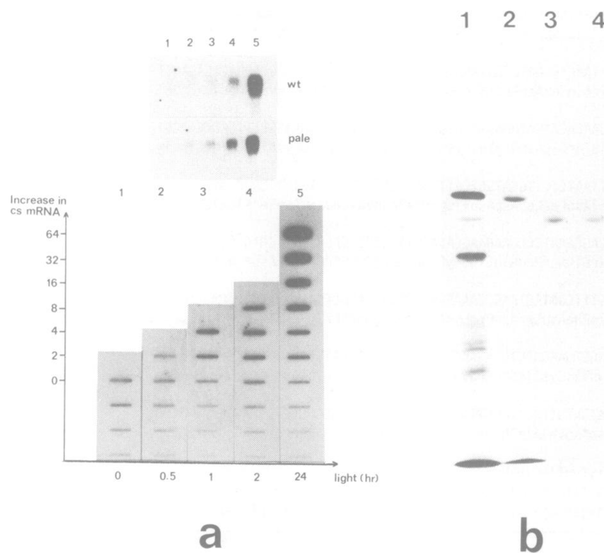

Fig. 6. (**a**) Northern and slot-blot analysis of *cs* transcripts in wild-type and pale mutant plants during the dark−light transition. Poly(A)$^+$ RNA (2.5 μg) prepared from dark grown plants (1) or from plants illuminated for 0.5 (2), 1 (3), 2 (4) and 24 (5) h was hybridized with an internal *Bam*HI−*Xba*I fragment of cDNA probe cp12. (Slot blots carry a series of 2-fold dilutions of RNAs starting at 1.25 μg in the upper lanes, followed by dilutions of 0.625, 0.312, 0.156 μg, etc.) Increase in mRNA concentration was measured by scintillation counting of ^{32}P signal in each slot as well as by quantitative scanning of autoradiographs. (**b**) *In vitro* binding and transport of the Cs protein to chloroplasts. *cs* RNA was synthesized by SP6 RNA polymerase from pSP64 and 65 vectors carrying a cDNA with a shortened 3′ end and translated in wheat germ extracts in the presence of [^{35}S]methionine. 1: labeled protein products of *in vitro* translation; 2: elution of 46 kd Cs precursor bound to intact chloroplasts; 3 and 4: processed Cs protein in lysates of washed and/or protease treated chloroplasts.

genomic and cDNA clones (unpublished), it may serve as a key to elucidate the function of this gene.

As is the case for many genes in *Arabidopsis* (Meyerowitz, 1987), the structure of the *cs* gene is simple and it occurs only once per haploid genome. Consistent with the light-regulated transcription, DNA sequence homology to box I, II (GT-1) and III regions of the pea *rbc*S-3A promoter (Kuhlemeier *et al.*, 1987) was detected at positions −205, −170 and −116, respectively, upstream of the initiation site of the *cs* transcript (C.Koncz, unpublished). DNA sequence data indicate that *cis*-regulatory elements of the *cs* gene are probably part of a dual promoter region which also directs the transcription of a putative gene, designated as ORF 'X', with opposite polarity.

The analysis described above has opened a window of about 0.048 centiMorgans (cM) from a total of 430 cM genetic length of the *Arabidopsis* genome. The resolution obtained by genetic mapping of the *cs* mutation approached this value. The conversion of linkage intensities to physical distances may not be completely correct due to observed variation of recombination frequencies in diverse genetic backgrounds. T-DNA insertions in the genome provide a useful tool to correlate distances between genetic, RFLP and physical maps. Using the T-DNA-encoded selectable marker, the insertions can be simply assigned to chromosomes by trisomic analysis and mapped further by crosses with available marker lines. A bacterial replicon within the T-DNA facilitates the isolation of plant DNA fragments joined to the inserts. These, in turn, can be used as probes for RFLP mapping or selection of the corresponding genomic DNA clones. Recently, we reported that T-DNA insertions in functional plant genes occur frequently (Koncz *et al.*, 1989). Therefore, with the help of a promoter-less reporter gene linked to the end of the T-DNA, a great number of diverse gene fusions can be mapped, identified and isolated. A screening of 450 out of 3000 available T-DNA-transformed *Arabidopsis* plants showed that only a portion (0.2−1.0%) of T-DNA-induced mutations cause visible alterations in the plant phenotype. One of these was the pale colour mutation, an allele of the *ch*-42 mutation on chromosome 4. This example and recent data on the identification of genes *agamous* (*ag*, E.Meyerowitz, personal communication), *apetala*-2 (*ap*-2, M.van Montagu, personal communication) on chromosome 4 and *glabrous*-1 (*gl*-1, Marks and Feldmann, 1989; Herman and Marks, 1989) on chromosome 3 of *Arabidopsis* illustrate that T-DNA tagging is an attractive alternative to transposable elements for the insertional mutagenesis and functional analysis of plant genes.

Materials and methods

Plant tissue culture and transformation

A.thaliana plants were maintained on modified MS medium (Murashige and Skoog, 1962) containing (g/l): 0.37 $MgSO_4 7H_2O$, 1.0 NH_4NO_3, 2.0 KNO_3, 0.17 KH_2PO_4, 0.435 $CaCl_2 \cdot 2H_2O$, 0.1 $CaH_4(PO_4)_2H_2O$ and 0.83 (mg/l) KI as a replacement for regular MS salts. Plants with elongated stems were dissected and incubated for 20 min with agrobacteria suspended at an OD_{550} of 0.5 in MSAR medium. Infected explants were incubated for 2 days on solid MSAR1 then subcultured for 7−10 days on MSAR1 medium containing (mg/l): 500 Claforan (Hoechst) and either 15 hygromycin (Boehringer) or 100 kanamycin (Sigma Co.). Calli, shoots and embryos were obtained on selective MASR2 medium. Regenerated plants were subcultured on MSAR3 and rooted on MSAR4 medium before planting in soil. Plant hormone concentrations were (mg/l): MSAR1 [2.0 indole-3-acetic acid (IAA), 0.5 N^6-(2-isopentyl)adenosine (9-iP), 0.2 benzylaminopurin (BAP) and 0.2 2,4-dichlorophenoxyacetic acid (2,4-D)], MSAR2 [2.0 9-iP, 0.05 1-naphthaleneacetic acid (NAA)], MSAR3 [1.5 9-iP, 0.05 NAA or 0.5 BAP, 0.05 NAA] and MSAR4 [1.0 indole-3-butyric acid (IBA), 0.05 NAA]. MSAR1−3 media were solidified by 0.2% Gelrite (Kelco Co.). MSAR4 contained one-fifth concentration of macroelements, 0.5% sucrose and 0.6% agar. All transgenic plants were maintained according to the German Guidelines for Recombinant DNA Research.

Genetic analysis and mapping

A total of 100 T2 progenies of each transgenic plant were grown in soil and on selective MSAR medium to screen for induced mutations and for 3:1 segregation of the T-DNA encoded hygromycin resistance marker. To determine the chromosomal linkage of the *cs* mutation, homozygous *cs* mutant plants were crossed as male with trisomic tester lines (Lee-Chen and Steinitz-Sears, 1967) and the ratio of green:pale plants was determined in a mixed trisomic and disomic population of F2 progenies (Redei, 1982). To map the *cs* mutation on chromosome 4, homozygous *cs* mutants were crossed as male or female with Landsberg 'erecta' *bp*−*cer*-2 (W100 *ms*$^+$, Koornneef *et al.*, 1986), and Columbia *bp*, *cer*-2, *bp*−*cer*-2 and *bp*−*ap*-2−*cer*-2 testers. F1 seeds derived from these crosses were individually germinated in soil. Relevant phenotypes were classified and counted in F2. *bp*−*cs* and *cs*−*cer*-2 recombinants were used for crosses in coupling. Recombinant fractions were determined by the maximum likelihood procedure (Bailey, 1961) and converted to map units (cM) using the Kosambi mapping function (Kosambi, 1944; Koornneef *et al.*, 1983). Calculation of homogeneity and standard deviation were as described (Redei, 1982). For probing the separability of the T-DNA-encoded Hyg^R marker and the *cs* mutation, *cs* was crossed with *bp*−*cer*-2 in repulsion and from each cross an average of 100 F2 progenies were germinated in soil. Green seedlings were discarded, while *bp*−*cs* and *cs*−*cer*-2 recombinant plants were harvested for seeds. On average 20 F3 progenies of each recombinant were tested for hygromycin resistance by germination on selective medium.

Vector constructions

Isolation of plasmid DNAs, ligation of synthetic oligonucleotide linkers and other DNA manipulations were performed according to Sambrook *et al.* (1989). Construction of pPCV002 (Koncz and Schell, 1986), pPCV701 (Koncz *et al.*, 1987b), pPCV702, 621 and 6NFHyg (Koncz *et al.*, 1989) was as reported previously. Replacing the CaMV 35S promoter fragment of pPCV702, a 1.6 kb *Eco*RI–*Bam*HI promoter fragment of potato gene ST-LS1 (Stockhaus *et al.*, 1987) was cloned in pPCV706. To obtain pPCV708, an *Eco*RI–*Hin*dIII cassette of CaMV 35S promoter–*nos* polyadenylation site was cloned from pPCV702 into the *Eco*RI site of pPCV701 following the conversion of the *Bam*HI site to a *Bgl*II site within the cassette and *Cla*I–*Bgl*II deletion of T-DNA gene 5 promoter from plasmid pPCV701.

T-DNA rescue and characterization of wild-type cs gene

Preparation of plant nuclear DNA and mapping of T-DNA insertions were as described (Koncz and Schell, 1986). To rescue the T-DNA insert, 10 μg of nuclear DNA isolated from homozygous pale plants was digested with *Hin*dIII, ligated and transformed into *E.coli* DH1 competent cells (Koncz *et al.*, 1989). *Cla*I–*Hin*dIII (1.15 kb) and *Bam*HI–*Hin*dIII (0.54 kb) T-DNA–plant DNA junction fragments of rescued clones were isolated and either used as hybridization probes or subcloned into pUC18 and 19 vectors (Yanisch-Perron *et al.*, 1985) to determine their nucleotide sequence. T7 DNA polymerase and double-stranded templates were used for DNA sequencing (Tabor and Richardson, 1987). Following preparation of *Mbo*I digested and size-fractionated nuclear DNA and poly(A)$^+$ RNA from wild-type *A.thaliana* (var. Columbia) plants, genomic and cDNA libraries were generated in EMBL 4 and λgt10 vectors as described (Sambrook *et al.*, 1989; Pharmacia cDNA Synthesis Kit). Six genomic and six cDNA clones were isolated by screening of the libraries with the T-DNA–plant DNA junction fragment probes. The established physical maps were confirmed by hybridization of wild-type DNA with fragments of genomic and cDNA clones. To determine the nucleotide sequence of the *cs* gene located on a 6.85 kb *Eco*RI fragment of EMBL 4 clone GC7, fragments listed below (referred to by the positions shown in Figure 4) were subcloned in pUC vectors and sequenced using the universal forwards and reverse primers: *Eco*RI/*Bam*HI–*Hin*dIII (0–660), *Bam*HI–*Xba*I (0–720), *Bam*HI–*Pst*I (0–325), *Pst*I–*Xba*I (325–720), *Eco*RI–*Eco*RV (0–340), *Eco*RV–*Xba*I (340–720), *Hin*dIII–*Xho*I (660–2180), *Hin*dIII–*Dra*I (660–1080), *Hin*dIII–*Stu*I (660–1165), *Hin*dIII–*Pst*I (660–1400), *Hin*dIII–*Bgl*II (660–1720), *Stu*I–*Xho*I (1165–2180), *Nco*I–*Xho*I (1405–2180), *Bgl*II–*Xho*I (1720–2180), *Xho*I–*Bam*HI (2180–3430), *Xho*I–*Xmn*I (2180–2820), *Pvu*I–*Xmn*I (2320–2820), *Pvu*I–*Bam*HI (2320–3430), *Xmn*I–*Bam*HI (2820–3430), *Bam*HI–*Xba*I (3430–4030), *Hin*dIII–*Sma*I (3960–5510), *Hin*dIII–*Pvu*II (3960–4260), *Hin*dIII–*Hin*dIII (3960–4720), *Hin*dIII–*Acc*I (3960–4870), *Hin*dIII–*Pst*I (3960–5020), *Pvu*II–*Hin*dIII (4260–4720), *Hin*dIII–*Pst*I (4720–5020), *Hin*dIII–*Sma*I (4720–5510), *Acc*I–*Acc*I (4870–5340), *Pst*I–*Sma*I (5020–5510), *Sma*I–*Xba*I (5510–6080), *Sma*I–*Eco*RV (5510–5765), *Eco*RV–*Xba*I (5765–6080), *Xba*I–*Xba*I (6080–6800), *Xba*I–*Eco*RV (6080–6370) and *Eco*RV–*Xba*I (6370–6800). cDNAs isolated as *Eco*RI fragments from λgt10 clones cp3, 5, 6, 12, 13 and 52 were inserted into pUC vectors for comparison of their 5′ and 3′ nucleotide sequences. The complete sequences of cDNAs cp12 and 13 were determined using subcloned fragments *Eco*RI–*Pvu*II (2770–3715), *Eco*RI–*Bam*HI (2770–3430), *Eco*RI–*Xba*I (2770–4035), *Xmn*I–*Eco*RI (2815–4380), *Xmn*I–*Xba*I (2815–4035), *Xmn*I–*Bam*HI (2815–3430), *Eco*RI–*Eco*RV (2770–3575), *Eco*RV–*Hin*dIII (3735–3960), *Pvu*II–*Pvu*II (3715–4260), *Hin*dIII–*Eco*RI (3960–4260), *Sst*I–*Sst*I (3910–4060), *Sst*I–*Eco*RI (4060–4380), *Pvu*II–*Eco*RI (4260–4380), *Bgl*II–*Eco*RI (4030–4380), *Xba*I–*Eco*RI (4035–4380) and *Bam*HI–*Eco*RI (3430–4380). (Numbers in brackets refer to positions in the genomic DNA sequence.) DNA sequences were analysed using a WISGEN program package (Deveraux *et al.*, 1984) adapted to VAX/VMS computer version V5.1–1. The nucleotide sequence data reported in this paper will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X51799/A.thaliana cs/ch-42 gene.

Complementation of the cs mutation

To complement the *cs* mutation by transformation with wild-type genomic and cDNAs, a 5.5 kb *Eco*RI–*Sma*I fragment of genomic clone GC7 carrying the entire gene was inserted into *Eco*RI–*Bam*HI sites of plasmid pCV002 DNA. To shorten the poly(A) track within the cDNA, the 3′ end of clone cp12 was replaced by a *Pst*I–*Bgl*II fragment of clone cp52 that carried only 17 A residues. The modified cDNA was inserted as an *Eco*RI fragment into the single *Bam*HI or *Bgl*II sites of expression vectors pPCV701, 702, 706 and 708. Conjugation of T-DNA vectors from *E.coli* donor strain S17-1 to *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) and *Agrobacterium*-mediated transformation of plant tissues were as described (Koncz and Schell, 1986; see above).

Transcript analysis and chloroplast protein transport assays

Wild-type and *cs* mutant plants grown on MSAR medium were placed into darkness for 5 days, then exposed to normal light of 375 lux. For purification of poly(A)$^+$ RNA, samples were collected after time periods of 0.5, 1, 2, 3, 4, 5, 6, 12 and 24 h following the start of illumination. Purification, electrophoresis, Northern and slot blotting, and filter hybridization of RNA were as described (Koncz and Schell, 1986; Sambrook *et al.*, 1989). To detect the *cs* transcript, an internal *Bam*HI–*Xba*I fragment of cDNA cp12 was used as hybridization probe.

Following subcloning of the shortened cDNA in plasmids pSP64 and 65 (Melton *et al.*, 1984), wild-type *cs* transcript was synthesized by SP6 RNA polymerase, capped and translated to radioactively labelled protein precursors in wheat germ lysate according to Reiss *et al.* (1987). *In vitro* protein transport assays with isolated chloroplasts were performed as described (Reiss *et al.*, 1989).

Acknowledgements

We thank Ms A.Radermacher and Ms A.Lossow for skillful technical assistance, and Gisela Kobert for typing the manuscript. This work was supported as part of a joint project between the MPI (Köln) and BRC (Szeged) by the Deutsche Forschungsgemeinschaft and Hungarian Academy of Sciences.

References

Andre,A., Colau,D., Schell,J., Van Montagu,M. and Hernalsteens,J.-P. (1986) *Mol. Gen. Genet.*, **204**, 512–518.

Bailey,N.T.J. (1961) *The Mathematical Theory of Genetic Linkage*. Clarendon Press, Oxford.

Chang,C., Bowman,J.L., DeJohn,A.W., Lander,E.S. and Meyerowitz,E.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6856–6860.

Coen,E.S., Robbins,T.P., Almeida,J., Hudson,A. and Carpenter,R. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. ASM, Washington, DC, pp. 413–436.

Deveraux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.

Feldmann,K.A., Marks,M.D., Christianson,M.L. and Quatrano,R.S. (1989) *Science*, **243**, 1351–1354.

Finkelstein,R., Estelle,M., Martinez-Zapater,J. and Somerville,C. (1988) In Verma,P.D.S. and Goldberg,R.B. (eds), *Plant Gene Research: Basic Knowledge and Application: Temporal and Spatial Regulation of Plant Genes*. Springer Verlag, Vienna, pp. 1–26.

Fischerova,H. (1975) *Biol. Plant. (Praha)*, **17**, 182–188.

Gierl,A. and Saedler,H. (1989) *Annu. Rev. Genet.*, **23**, 71–85.

Herman,P.L. and Marks,D.M. (1989) *Plant Cell*, **1**, 1051–1055.

Koncz,C. and Schell,J. (1986) *Mol. Gen. Genet.*, **204**, 383–396.

Koncz,C., Martini,M., Koncz-Kalman,Zs., Olsson,O., Radermacher,A., Szalay,A.A. and Schell,J. (1987a) In Bruening,G., Harada,J., Kosuge,T. and Hollaender,A. (eds), *Tailoring Genes for Crop Improvement*. Plenum Press, New York, pp. 197–209.

Koncz,C., Olsson,O., Langridge,W.H.R., Schell,J. and Szalay,A.A. (1987b) *Proc. Natl. Acad. Sci. USA*, **84**, 7943–7947.

Koncz,C., Martini,N., Mayerhofer,R., Koncz-Kalman,Zs., Körber,H., Redei,G.P. and Schell,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8467–8471.

Koornneef,M., Van Eden,J., Hanhart,C.J., Stam,P., Braaksma,F.J. and Fenstra,W.J. (1983) *J. Hered.*, **74**, 265–272.

Koornneef,M., Hanhart,C.J., Van Locnen Martinet,E.P. and Van der Veen,J.M. (1986) *Arabidopsis Inf. Serv.*, **23**, 47–50.

Koornneef,M. (1987) In O'Brien,S.J. (ed.), *Genetic Maps*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 4, pp. 742–745.

Kosambi,D.D. (1944) *Ann. Eugen.*, **12**, 172–175.

Krebbers,E., Seurinck,J., Herdies,L., Cashmore,A.R. and Timko,M.P. (1988) *Plant Mol. Biol.*, **11**, 745–759.

Kuhlemeier,C., Fluhr,R., Green,P.J. and Chua,N.-H. (1987) *Genes. Dev.*, **1**, 247–255.

Lee-Chen,S., and Steinitz-Sears,L.M. (1967) *Can. J. Genet. Cytol.*, **9**, 381–384.

Leutwiler,L.S., Meyerowitz,E.M. and Tobin,E.M. (1986) *Nucleic Acids Res.*, **14**, 4051–4076.

Marks,D.M. and Feldmann,K.A. (1989) *Plant Cell*, **1**, 1043–1050.

Masterson,R., Furtek,D.B., Grevelding,C. and Schell,J. (1989) *Mol. Gen. Genet.*, **219**, 461–466.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7046.
Meyerowitz,.E.M. (1987) *Annu. Rev. Genet.*, **21**, 93–111.
Murashige,T. and Skoog,F. (1962) *Physiol. Plant.*, **15**, 473–497.
Nam,H.G., Giraudat,J., DenBoer,B., Moonman,F., Loos,W.D.B., Hauge,B.M. and Goodman,H.M. (1989) *Plant Cell*, **1**, 699–706.
Redei,G.P. (1975) *Annu. Rev. Genet.*, **9**, 111–117.
Redei,G.P. (1982) *Genetics*. MacMillan Press, New York.
Redei,G.P., Koncz,C. and Schell,J. (1988) In Gustafson,J.P. and Appels,R. (eds), *Chromosome Structure and Function.* Plenum Press, New York, pp. 175–200.
Reiss,B., Wasmann,C.C. and Bohnert,H.J. (1987) *Mol. Gen. Genet.*, **209**, 116–121.
Reiss,B., Wasmann,C.C., Schell,J. and Bohnert,H.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 886–890.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schell,J. (1987) *Science*, **237**, 1176–1183.
Stockhaus,J., Eckes,P., Rocha-Sosa,M., Schell,J. and Willmitzer,L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7943–7947.
Tabor,S. and Richardson,C.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4767–4771.
Van Sluys,M.A., Tempe,J. and Fedoroff,N. (1987) *EMBO J.*, **6**, 3881–3890.
Voytas,D.F. and Ausubel,F.M. (1988) *Nature*, **336**, 242–244.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Zambryski,P. (1988) *Annu. Rev. Genet.*, **22**, 1–30.

Received on January 30, 1990