Universiteit
Leiden
The Netherlands

Structure and expression of a light-inducible shoot-specific rice gene

Pater, B.S. de; Hensgens, L.A.M.; Schilperoort, R.A.

Citation

Pater, B. S. de, Hensgens, L. A. M., & Schilperoort, R. A. (1990). Structure and expression of a light-inducible shoot-specific rice gene. *Plant Molecular Biology*, *15*(3), 399-406. doi:10.1007/BF00019157

Version:	Publisher's Version
License:	Licensed under Article 25fa Copyright Act/Law (Amendment Taverne)
Downloaded from:	https://hdl.handle.net/1887/3753569

Note: To cite this publication please use the final published version (if applicable).

Plant Molecular Biology **15**: 399–406, 1990.
© 1990 *Kluwer Academic Publishers. Printed in Belgium.*

Structure and expression of a light-inducible shoot-specific rice gene

Sylvia de Pater,[1]* Lambert A. M. Hensgens and Rob A. Schilperoort
MOLBAS Research Group, Department of Plant Molecular Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, Netherlands; [1] *Current address: RUL/TNO Centre for Phytotechnology, Botanical Laboratory, Nonnensteeg 3, 2311 VJ Leiden, Netherlands (*author for correspondence)*

Received 6 March 1990; accepted in revised form 30 May 1990

Key words: light-inducible, rice, shoot-specific

Abstract

By differential screening of a cDNA library of two-week-old rice seedlings cDNA clones were obtained, corresponding to shoot-specific mRNAs. By sequence analysis two of these clones were found to be *rbcS* cDNA clones. The mRNA corresponding to a third cDNA clone (COS5) displayed an expression pattern similar to the expression pattern of *rbcS* genes. The mRNA (800 bases) was light-inducible and encoded by a single-copy gene. The genomic clone (GOS5) was isolated and the intron/exon structure was determined by comparing the nucleotide sequences of the mRNA and the genomic clone. The gene contains two introns. Transcription start sites were determined by S1-nuclease mapping and primer extension. The start site obtained by both methods is located 87 bp upstream of the translation start site and 23 bp downstream of TATA box-like sequence. In the 5′ non-coding region motifs can be found that are homologous to sequences in promoters that are light- or UV-inducible or confer leaf-specific expression. The open reading frame present in GOS5 codes for a protein (15 kDa) that contains a putative chloroplast transit peptide and does not show any significant homology to protein sequences in the NBRF protein database.

Introduction

To increase the current knowledge about the molecular biology of rice and to obtain tools for biotechnical engineering of this important crop, we have isolated and analysed a rice gene that is highly expressed in a shoot-specific manner. The use of tissue-specific promoters for genetic engineering of plants offers possibilities to make specific tissues or organs resistant to pathogens or herbicides or to alter the protein composition of consumable parts. Tissue-specific promoters have already been isolated from several, mostly dicotyledonous, plant species and these might be used for genetic engineering of rice. However, it has been observed in several cases that the activity of promoters in heterologous hosts is lower than in the plants from which they were isolated [2, 9, 10, 16, 17]. For biotechnical engineering of rice such problems might be circumvented by using promoters from rice itself.

An important class of plant genes, displaying a shoot-specific expression pattern, encodes chloroplast proteins involved in the conversion of light energy into chemical energy and in the fixation of carbon dioxide in carbohydrates. Most of these proteins are nuclear-encoded [4] and contain a transit peptide, targeting them into chloroplasts

[23]. Well studied members of this group of proteins are the small subunit of ribulose-1,5-bisphosphate carboxylase, chloropyll *a*/*b* binding protein, ferredoxin and plastocyanin.

Here we describe the structure and expression of a single-copy gene (GOS5) from rice (Indica variety IR36), which displays a shoot-specific and light-inducible expression pattern. The encoded protein contains a putative transit peptide possibly involved in transport into chloroplasts. We found no sequence homology to other known plant proteins.

Materials and methods

Plant growth

Rice (*Oryza sativa*, var. *indica* IR36 or var. *japonica* T309) was grown in a growth chamber at 28 °C with dark/light intervals of 14/10 h, at a relative humidity of 73%, in sterile sand with Luwasa nutrients for two weeks (seedlings) or three months (flowering plants).

RNA and DNA isolation and blot hybridization

Total RNA was isolated by grinding the tissue in liquid N_2, phenol-chloroform extraction followed by LiCl precipitation [21]. DNA was recovered from the supernatant by adding 0.6 volume isopropanol and further purified by CsCl-EtBr centrifugation. Poly(A) RNA was isolated using oligo(dT)cellulose columns [21]. RNA and DNA blot hybridizations were performed as described [15].

cDNA and genomic cloning

A cDNA plasmid library was constructed and amplified as described [15] using 12 μg poly(A) RNA from two-week-old IR36 rice seedlings. The library consisted of 7.3×10^4 independent transformants. Replica filters (Hybond-N, Amersham) containing 500 recombinant colonies were differentially screened for root- and shoot-specific sequences using 0.12 μg ^{32}P-labelled single-stranded cDNA (10^8 cpm/μg) synthesized on total RNA from roots or shoots. Preparation, hybridization and washing of the filters was done as described [15].

A non-amplified genomic library in Lambda Fix (Stratagene) was produced. DNA was isolated from green parts of two-week-old IR36 seedlings, partially digested with *Sau* 3AI and fractionated on agarose gels. 15–20 kb fragments were cloned in the *Xho* I sites of Lambda Fix after partially filling in of both the *Sau* 3AI and *Xho* I sticky ends with dGTP, dATP and dCTP, dTTP respectively. This procedure prevents (1) ligation of more than one *Sau* 3AI fragments in the Lambda vector and (2) ligation of the stuffer fragment. The library was screened for sequences hybridizing to cDNA clone COS5. Hybridization and washing conditions were as for DNA blot hybridization [15]. Isolation of Lambda DNA was done according to the plate lysate method [13].

Nucleic acid sequencing

The cDNA insert of COS5 was cloned in M13tg130 [11] and sequenced by the dideoxy chain-termination method [18].

The 4 kb *Sac* I and 3 kb *Bam* HI fragments of Lambda clone 5 were subcloned in Bluescript vector SKM13+ (Stratagene). For both orientations series of overlapping exonuclease III deletions were produced (Stratagene, Bluescript exo/mung DNA sequencing system) and appropriate clones were sequenced. Parts of the coding region were subcloned in M13tg130 or M13tg131 and sequenced. The sequence shown in Fig. 4 was determined on both DNA strands.

The 5′ end of the mRNA was sequenced by the dideoxy chain-termination method essentially as described [5] using 10 μg poly(A) RNA and the 17-mer 5′-GAGGGACGCCATTGCTG-3′.

Nucleotide sequence data were collected, assembled and analysed with a VAX computer fitted with the Genetics Computer Group Sequence Analysis Software Package [3].

S1-nuclease mapping

The end-labelled *Bgl* II/*Sma* I fragment containing the 5′ part of the coding region (Fig. 4) was used for S1-mapping essentially performed as described [1].

Results

Isolation of cDNA clones corresponding to shoot-specific and light-inducible mRNAs

A cDNA library made on poly(A) RNA from two-week-old seedlings of the rice variety IR36 was differentially screened for root- and shoot-specific sequences. cDNA clones that hybridized strongly with a single-stranded cDNA probe made on shoot RNA, but not with a single-stranded cDNA probe made on root RNA, were further characterized by cross-hybridization, RNA blot hybridization and sequence analysis. Besides two ribulose-1,5-bisphosphate carboxylase (*rbcS*) cDNA clones, another cDNA clone (COS5) was obtained that hybridized strongly with shoot RNA but not with root RNA of seedlings, as shown in Fig. 1A. When seedlings were grown in the dark no hybridization was detectable with RNA from shoots (Fig. 1A, lane 2). The estimated size of the mRNA that hybridized to the cDNA clone COS5 is 800 bases.

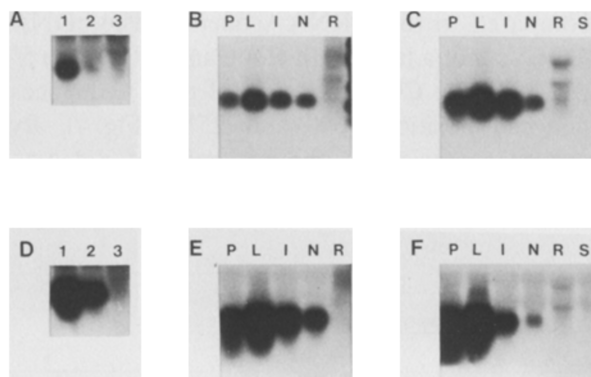

Fig. 1. *rbcS* and COS5 expression pattern. Gel blots containing RNAs isolated from green shoot tissue (1), etiolated shoot tissue (2) and roots (3) from two-week-old IR36 seedlings or RNAs isolated from panicles (P), leaves (L), internodes (I), nodes (N) and roots (R) from three-month-old plants of the variety IR36 (panels B and E) or T309 (panels C and F) and RNA isolated from a T309 cell suspension (lanes S) were hybridized with ^{32}P-labelled COS5 cDNA (panels A, B and C) or with ^{32}P-labelled rice *rbcS* cDNA (panels D, E and F). All lanes contain 20 μg of total RNA.

Figure 1B shows the expression pattern in flowering plants. RNA was isolated from panicles (P), leaves (L), internodes (I), nodes (N) and roots (R). The expression was highest in leaves (Fig. 1B, lane L), less in light green parts (Fig. 1B, lanes P, I, N) and not detectable in roots (Fig. 1B, lane R). The faint bands visible in the root lanes are background, since they are present after hybridization with different non-homologous cDNAs (*rbcS*, COS5 and others (not shown)). The reason for this background is that it is relatively difficult to isolate RNA from rice roots, resulting in preparations containing a high amount of impurities. These preparations were shown to give good specific signals on RNA blots probed with cDNAs corresponding to root-specific or constitutive mRNAs.

Although the expression pattern was roughly similar to the rice *rbcS* expression pattern, as shown in Fig. 1D and 1E, the expression level was 5–10 times lower (compare Fig. 1A and 1B with Fig. 1D and 1E) and COS5 did not cross-hybridize to different *rbcS* cDNA clones (results not shown) nor was there sequence homology. Furthermore, the *rbcS* cDNA clones hybridized to a mRNA of approximately 900 bases. Figures 1C and 1F show that the hybridization patterns of COS5 cDNA and *rbcS* cDNA with RNAs isolated from another rice variety (*japonica*, T309) were similar. The expression in a (white) cell suspension of T309 was below our detection level (Fig. 1C and 1F, lanes S).

Copy numbers of genes corresponding to rbcS *cDNAs and COS5 cDNA*

DNA blots containing IR36 genomic DNA were hybridized with *rbcS* cDNA clones or COS5 cDNA. The DNA was digested with *Eco* RI,

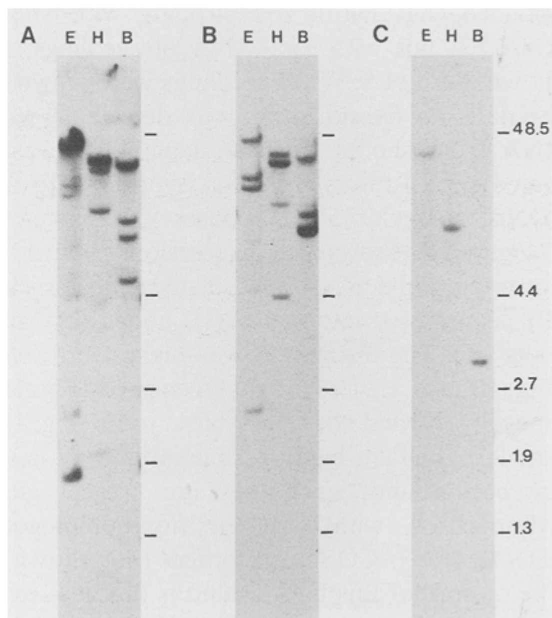

Fig. 2. rbcS and COS5 gene copy number. Gel blots containing rice DNA digested with *Eco* RI (E), *Bam* HI (B) or *Hin*d III (H) were hybridized with two different ^{32}P-labelled rice *rbcS* cDNA clones (A and B) or with the ^{32}P-labelled COS5 cDNA clone (C). All lanes contain 10 μg DNA. The size of marker fragments are indicated in kb.

Bam HI or *Hin*d III, which have no recognition sites in the cDNA inserts. The results are shown in Fig. 2. At least four bands are visible after hybridization with the *rbcS* cDNA clones (Fig. 2A and 2B). It is very likely that these cDNAs correspond to a gene family, unless the genes contain introns with recognition sites for *Eco* RI, *Bam* HI and *Hin*d III. Hybridization with cDNA COS5 resulted in one band in each lane (Fig. 2C). This indicates that the corresponding gene is present in one copy per genome unless (1) the hybridizing fragments contain more than one gene or (2) the haploid rice genome contains several identical copies of this gene (as far as the restriction sites used are concerned). The first possibility seems unlikely, since 2.5 kb of the 3.0 kb *Bam* HI fragment has been sequenced, and contains only one COS5 homologous region (see below). The second possibility cannot be completely ruled out, but the intensities of the hybridizing bands are similar to the signals obtained with other single-copy genes (results not shown).

In order to isolate an active promoter, we have isolated the gene encoding COS5 mRNA, as individual members of gene families (*rbcS*) might be inactive or expressed at a low level.

Isolation of the gene GOS5 encoding COS5 mRNA

A genomic library of IR36 DNA was constructed in Lambda Fix. By screening with COS5, seven genomic clones were obtained. The internal *Sac* I and *Bam* HI fragments that hybridized to COS5 had identical sizes in all seven clones (results not shown) indicating that the clones were different *Sau* 3A partials from the same gene. The *Sac* I and *Bam* HI fragments were subcloned. The restriction maps are shown in Fig. 3.

Structure of the gene

The nucleotide sequence of the GOS5 transcribed region including 1700 base pairs (bp) upstream and 200 bp downstream of the protein coding region, as shown in Fig. 4, was determined and compared to the mRNA sequence. The cDNA COS5 is not a full-length clone and misses the 5′ part (Fig. 3). COS5 consists of exon sequences between position 1698 and 2516 (Fig. 4). By primer extension in the presence of dideoxy-

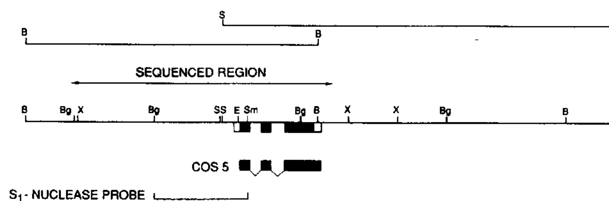

Fig. 3. Structure of GOS5. A restriction map of the overlapping *Bam* HI and *Sac* I fragments hybridizing to the COS5 cDNA clone is shown. B = *Bam* HI; X = *Xba* I; Bg = *Bgl* II; S = *Sac* I; E = *Eco* RI; Sm = *Sma* I. Both strands of the indicated region were sequenced. The exons are shown by bars: the top row of black bars represent the protein coding region, the bottom bars represent sequences present in the cDNA clone COS5. The *Bgl* II–*Sma* I fragment used for the S1-nuclease mapping of Fig. 5 is shown.

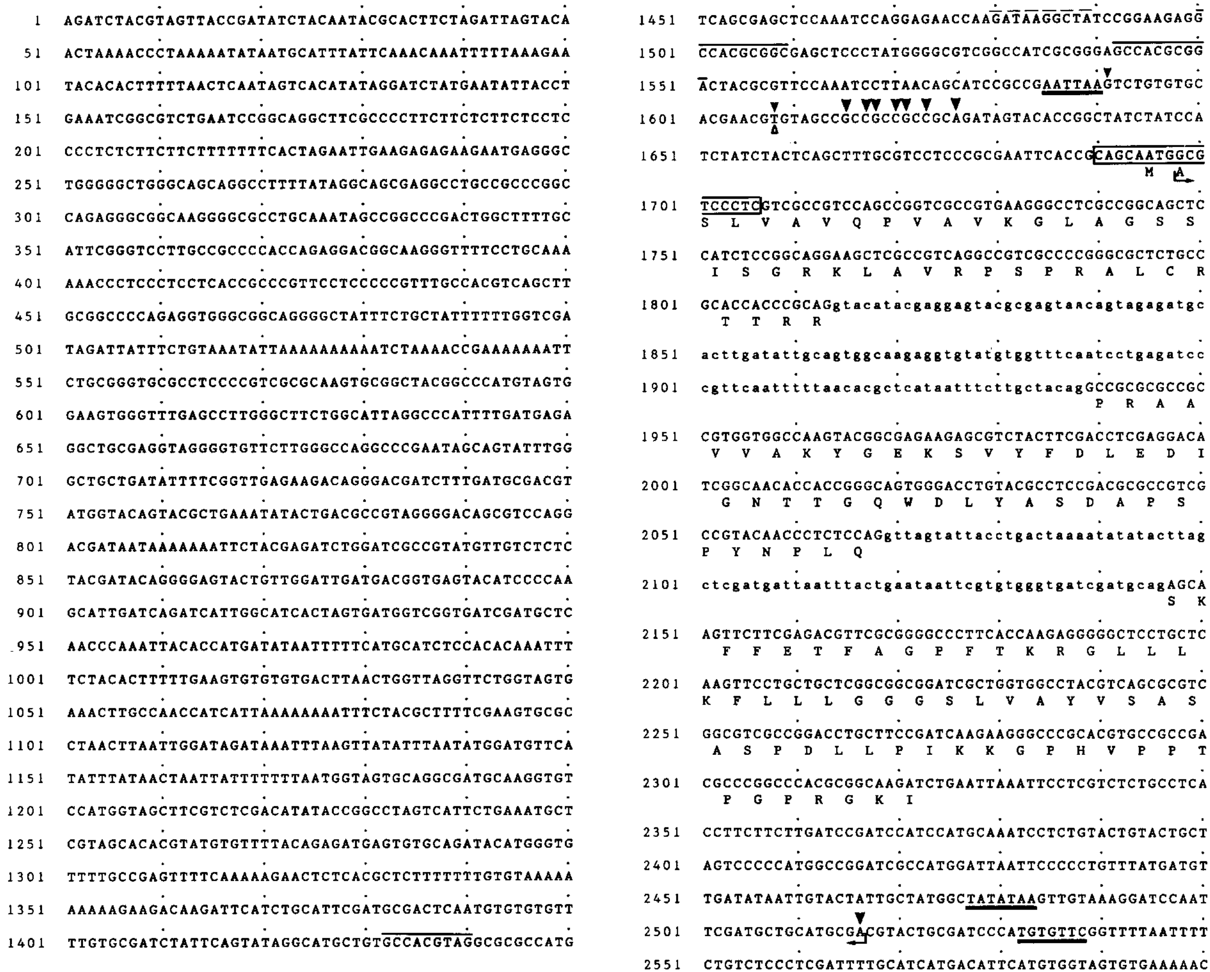

Fig. 4. Nucleotide sequence and deduced amino acid sequence of GOS5. The borders of the cDNA clone COS5 are shown by arrows. The poly(A) addition site and the 5′ ends obtained by primer extension are indicated with solid arrowheads. The transcription start site determined by S1-mapping is indicated by an open arrowhead. Intron sequences are shown in lower case letters. A putative TATA box and signals involved in processing of the 3′ end of the mRNA are underlined. Sequences in the 5′ flanking region with homology to the G box motif and as-2 binding site (dotted) are overlined. The sequence homologous to the 17-mer used for primer extension is boxed.

nucleotides using a 17-mer oligonucleotide hybridizing around the methionine start codon (Fig. 4) the nucleotide sequence of the 5′ part of the mRNA was determined, as shown in Fig. 5. The sequence of the mRNA and the corresponding sequence in the genomic clone GOS5 are identical, which is in agreement with the assumption that GOS5 is a single-copy gene.

The gene contains two introns (Fig. 3 and 4) of 124 and 78 bp. In both introns typical GT/AG donor-acceptor sites are present. The sequences surrounding these junctions (CAG * GTAcaTACG ...... aaTTTcTTGCTACAG * and CAG * GTtAGTATT ...... ggGTgATcGATGCAG *) are 75% homologous with the consensus sequence for splice sites of monocots (C/A/G)AG * GTA(T/A)G(T/C)(T/A)(T/C)N ... ... (T/C)-(T/C)(T/G)(T/A)(T/C)(T/A)(T/C)(T/G)NNTN-CAG * [7].

Protein coding region

Inspection of the composite sequence of the exons reveals an open reading frame of 426 bp, located between the first AUG codon at position 1695 and the UGA codon at position 2324 (Fig. 4). The context of the translation start site (gcAgCAATGGCg) resembles the plant consensus sequence (TAAACAATGGCT) rather well [8]. The open reading frame potentially encodes a protein of 142 amino acids (M_r 15052). The N-terminal part has some characteristics of chloroplast transit peptides [22]. The amino acid sequence begins with methionine-alanine and among the first 10 residues no arginine, lysine, asparagine, glutamine and glycine are found and only one proline. Thus, the protein might be transported into chloroplasts. In searching the NBRF protein database (release June 1989) with the UWGCG word search program no significant homology with any protein was found.

5′ and 3′ flanking regions

The transcription start site was determined by S1-nuclease mapping and primer extension. The results are presented in Fig. 5. S1-nuclease mapping using the *Sma* I-*Bgl* II fragment, as shown in Fig. 3, resulted in a 180 bp protected fragment (Fig. 5A), indicating that transcription starts around position 1608. The start sites determined by primer extension (Fig. 5B) are located near the start site determined by S1-nuclease mapping (Fig. 4). The majority of the 5′ ends (position 1615–1626) fall within a very GC-rich region and might be premature stops of reverse transcription, as they were not detected by S1 mapping. Further upstream, some minor start sites are present. One of these is located at position 1608 and coincides with the position found with S1 mapping. The most upstream 5′ end obtained by primer extension was also not detected by S1 mapping. Altogether it is not clear which bands represent 5′ ends of the mRNA. Position 1608 is most likely the transcription start site although it does not fit with the consensus sequence for plant transcription start sites (CÅTCA) [8].

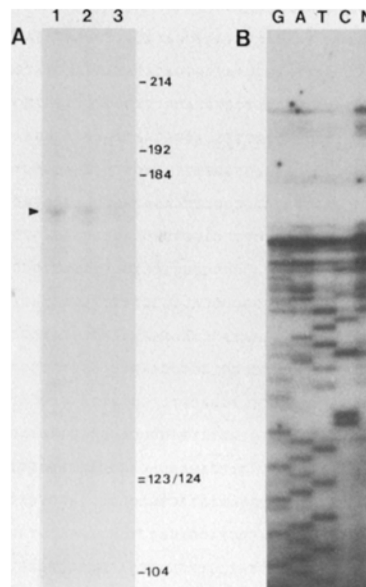

Fig. 5. Determination of the GOS5 transcription start sites. (A) S1-nuclease mapping of 20 μg of total RNA using the probe indicated in Fig. 3 and three different concentrations of S1-nuclease; 100 (1), 300 (2), and 1000 (3) units/ml. The protected fragments are indicated with an arrowhead and the size of marker fragments is in bases. (B) Primer extension on 10 μg of poly(A) RNA in the absence (N) or presence of dideoxy NTPs (G, A, T and C) using the 17-mer indicated in Fig. 4.

Twenty-three bp upstream of the putative transcription start site (1608) an AT-rich stretch (aATtaA) is present with some homology to the TATA box consensus sequence (TATATA) [8]. The region upstream of the TATA box (1431–1584) contains three motifs which have homology to a sequence (G box, TCCACGTGGC) present in several *rbcS* [6] and chalcone synthase (*chs*) genes [19, 20]. This sequence was found to bind nuclear protein factors (named GBF, CG-1) from different plant species. Mutation of this motif in the *chs* gene of parsley abolished UV light-induced expression [19]. The minimal binding site in the experiments with the *rbcS* genes seems to be 12 bp (TGACACGTGGCA) [6] whereas the minimal binding site with the *chs* gene of *A. majus*, where a five-copy situation was tested, was found to be the hexameric sequence CACGTG [20]. Thus it seems that at least the CACGTG motif is required

for binding. The mismatch present in the most upstream motif in GOS5 (CACGTA) resulted in loss of binding activity [20]. It is unknown what the influence is of the mismatch present in the other two motifs (CACGCG).

In addition, a sequence can be found (GATAAGGcTA) with homology to the ASF-2 binding site [12] (Fig. 4). This sequence is able to confer leaf-specific expression when fused to the − 90 derivative of the 35S promoter of CaMV and it is present in promoters of several light-responsive genes [12].

The cDNA clone COS5 contains a poly(A) tail of 8 bp. A sequence (tATAtAA) with homology to the poly(A) addition consensus (AATAAA) is present 40 bp upstream of the poly(A) tail. In the GOS5 gene 17 bp downstream of the poly(A) additional site a sequence (TGTGTTCg) is present with homology to the consensus sequence (YGTGTTYY) of which the significance in animal gene expression has been demonstrated [14].

Discussion

The rice gene GOS5, which displays a shoot-specific and light-inducible expression pattern, was isolated. The intron/exon structure of the gene was determined by comparing the mRNA sequence and the genomic sequence. GOS5 is a split gene composed of three exons.

The 5′ non-coding region contains motifs homologous to a sequence present in the 5′ non-coding region of several *rbcS* [6] and *chs* [19, 20] genes, which are, like the GOS5 gene, light-inducible. In the *chs* gene of parsley a nuclear protein factor binds to this sequence after irradiation. Replacement of this sequence led to complete loss of light-responsiveness [19].

The as-2-like sequence in the GOS5 promoter might also be (in part) responsible for the tissue-specific expression of GOS5. The factor (ASF-2) that binds to this sequence could not be detected in root extracts [12].

The level of GOS5 transcripts is high in dark green tissue, lower in light green tissue and absent in roots, white cell suspension or etiolated shoots. Thus the mRNA level seems to be correlated with the presence of chloroplasts. The N-terminus of the GOS5 protein shows some features of chloroplast transit peptides, indicating that it might be transported into chloroplasts. We did not find any sequence homology between the GOS5 amino acid sequence and other light-inducible plant proteins. The majority (at least 75%) of the chloroplast proteins is nuclear-encoded and genes of only a few of these have been isolated and characterized yet.

The steady-state mRNA level of GOS5 is 5–10 fold less than the *rbcS* steady-state mRNA level. Since GOS5 is a single-copy gene and *rbcS* is a gene family of 4–5 members, the GOS5 promoter activity is only a few fold less, with this restriction that the mRNAs of both genes are equally stable. Thus the GOS5 promoter seems to be very active.

By isolation of the single-copy gene GOS5 we assume to have obtained a promoter which can be used to express useful qualities in rice in a shoot-specific and light-dependent manner.

Acknowledgements

We thank Johan Memelink for comments on the manuscript and Peter Hock and the Biology Photographic Service for preparation of the figures. This work was financed by the Rockefeller Foundation.

References

1. Berk AJ, Sharp PA: Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12: 721–732 (1977).
2. Chen Z-L, Schuler MA, Beachy RN: Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA 83: 8560–8564 (1986).
3. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs of the VAX. Nucl Acids Res 12: 387–395 (1984).
4. Ellis RJ: Chloroplast proteins: synthesis, transport, and assembly. Ann Rev Plant Physiol 32: 111–137 (1981).
5. Geliebter J: Dideoxynucleotide sequencing of RNA and uncloned cDNA. Focus 9.1: 5–9 (1987).
6. Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik

PA, Cashmore AR: An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci USA 85: 7089–7093 (1988).

7. Hanley BA, Schuler MA: Plant intron sequences: evidence for distinct groups of introns. Nucl Acids Res 16: 7159–7176 (1988).
8. Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl Acids Res 15: 6643–6653 (1987).
9. Keith B, Chua N-H: Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J 5: 2419–2425 (1986).
10. Keller JM, Shanklin J, Vierstra RD, Hershey HP: Expression of a functional monocotyledonous phytochrome in transgenic tobacco. EMBO J 8: 1005–1012 (1989).
11. Kieny MP, Lathe R, Lecocq JP: New versatile cloning and sequencing vectors based on bacteriophage M13. Gene 26: 91–99 (1983).
12. Lam E, Chua N-H: ASF-2: A factor that binds to the Cauliflower Mosaic Virus 35S promoter and a conserved GATA motif in *Cab* promoters. Plant Cell 1: 1147–1156 (1989).
13. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY., (1982).
14. McLauchlan J, Gaffney D, Whitton JL, Clements JB: The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucl Acids Res 13: 1347–1368 (1985).
15. Memelink J, Hoge JHC, Schilperoort RA: Cytokinin stress changes the developmental regulation of several defence-related genes in tobacco. EMBO J 6: 3579–3583 (1987).
16. Rochester DE, Winer JA, Shah DM: The structure and expression of maize genes encoding the major heat shock protein, hsp70. EMBO J 5: 451–458 (1986).
17. Schernthaner JP, Matzke MA, Matzke AJM: Endosperm-specific activity of a zein gene promoter in transgenic tobacco plants. EMBO J 7: 1249–1255 (1988).
18. Sanger F, Nicklen S, Coulsen AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
19. Schulze-Lefert P, Dangl JL, Becker-André M, Hahlbrock K, Schulz W: Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. EMBO J 8: 651–656 (1989).
20. Staiger D, Kaulen H, Schell J: A CACGTG motif of the *Antirrhinum majus* chalcone synthase promoter is recognized by an evolutionarily conserved nuclear protein. Proc Natl Acad Sci USA 86: 6930–6934 (1989).
21. van Slogteren GMS, Hoge JHC, Hooykaas PJJ, Schilperoort RA: Clonal analysis of heterogeneous crown gall tumor tissues induced by wild-type and shooter mutant strains of *Agrobacteruim tumefaciens* – expression of T-DNA genes. Plant Mol Biol 2: 321–333 (1983).
22. von Heijne G, Steppuhn J, Hermann RG: Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem 180: 535–545 (1989).
23. Weisbeek P, Hageman J, Smeekens S, de Boer D, Cremer F: Chloroplast-specific import and routing of proteins. In: von Wettstein D, Chua N-H (eds) Plant Molecular Biology. Proceedings of NATO advanced study institute on plant molecular biology, pp. 77–91. Plenum Press, New York) (1987).