Mol Gen Genet (1991) 230:413–422

MGG

Determination of steady-state mRNA levels of individual chlorophyll a/b binding protein genes of the tomato *cab* gene family

Birgit Piechulla[1], **Jan-Wolfhard Kellmann**[1], **Eran Pichersky**[2], **Egbert Schwartz**[2], **and Hans-Heinrich Förster**[3]

[1] Institut für Biochemie der Pflanze, Untere Karspüle 2, 3400 Göttingen, FRG
[2] University of Michigan, Department of Biology, Ann Arbor, MI 48109, USA
[3] Max-Planck Institut für experimentelle Medizin, Hermann Rein Str. 3, 3400 Göttingen, FRG

Received July 26, 1991

Summary. The steady-state levels of mRNA produced by 14 genes encoding members of the tomtato chlorophyll a/b binding protein family were quantified. All genes were found to be expressed in leaf tissue, but the mRNAs accumulated to significantly different levels. The transcripts of *cab* 1A, *cab* 1B, *cab* 3A and *cab* 3B, encoding the Type I LHC proteins of photosystem II, are abundant, while low levels were measured for mRNAs encoding the Type II LHC II and the LHC I proteins. Sequences from the 5′ upstream regions (−400 to translational start) of some *cab* genes were determined in this study, and a total of 16 tomato *cab* gene promoters for which sequences are now available were analyzed. Significant sequence conservation was found for those genes which are tandemly linked on the chromosome. However, the level of sequence conservation is different for the different *cab* subfamilies, e.g. 85% similarity between *cab* 1A and *cab* 1D vs. 45% sequence similarity between *cab* 3A and *cab* 3C upstream sequences. Characteristic GATA repeats with a conserved spacing were found in 5′ upstream sequences of *cab* 1A–D, *cab* 3A–C, *cab* 11 and *cab* 12. The consensus sequence CCTTATCAT, which is believed to mediate light responsiveness, was found at different locations in the upstream sequences of *cab* 6B, *cab* 7, *cab* 8, *cab* 9, *cab* 10A, *cab* 10B and *cab* 11. In 11 out of 15 genes the transcription initiation site was found to center on the triplet TCA.

Key words: Chlorophyll a/b binding proteins – Tomato – Gene family – mRNA accumulation – Promoter analysis

Introduction

The light energy used in the process of photosynthesis in the chloroplasts of plants is first captured in the macromolecular structures known as Light Harvesting Complexes (LHCs). These complexes are found in the thylakoid membranes in close association with the 'core' complexes of PS I and PS II reaction centers. LHC I is associated with PS I, while LHC II and two other minor light harvesting complexes, known as CP24 and CP29, are associated with PS II (Green et al. 1991). The chlorophyll molecules in these LHCs are bound to proteins known as chlorophyll a/b binding (CAB) polypeptides. Examination of the nucleotide sequences of the *cab* genes and the predicted amino acids of the encoded polypeptides revealed extensive sequence similarities, indicating that the CAB polypeptides of the various LHCs are structurally and evolutionary related to each other (Pichersky and Green 1990; Green et al. 1991).

The *cab* genes have been classified into ten different types based on coding sequence similarities/divergences and intron positions (Green et al. 1991; Jansson and Gustafsson 1991). The large number of genes encoding structurally and functionally related proteins raises several questions regarding the mechanism(s) which regulate the expression of these genes. Many previous reports have dealt with the expression pattern of a single type of *cab* genes, the one encoding the LHC II Type I CAB polypeptide (Pichersky et al. 1985; Piechulla and Gruissem 1987; Castresana et al. 1987), and a few other reports have presented the expression characteristics of one other type of *cab* genes (Stayton et al. 1986; Pichersky et al. 1987b; Pichersky et al. 1989). These reports have indicated that the *cab* genes examined were under the control of exogenous and endogenous stimuli (organ- and tissue-specificity, circadian rhythmicity, developmental control, light control) (Piechulla and Gruissem 1987; Kuhlemeier et al. 1987; Kellmann et al. 1990).

We have isolated and characterized *cab* genes of eight different types from the diploid dicot species *Lycopersicon esculentum* (tomato). This collection of genes constitutes the largest set of *cab* genes available from any plant species and represents an almost complete set of the *cab* genes in the plant genome. The availability of these

---

Offprint requests to: B. Piechulla

genes and their sequences has allowed us to investigate the specific mode of expression of each type of *cab* gene and the expression characteristics of particular individual genes, and to examine whether the entire set of *cab* genes is coordinately expressed.

Materials and methods

Determination of steady-state mRNA levels. RNA was isolated from leaves of 50-day-old tomato plants (*Lycopersicon esculentum* Mill. VFNT LA 1221; grown in the greenhouse at the University of Göttingen), harvested at 1:30 PM on July 25, 1990 (sunrise 4:35 AM, sunset 8:21 PM), according to the method described elsewhere (Kellmann et al. 1990). To determine the steady-state mRNA levels corresponding to products of individual *cab* genes, specific oligonucleotides were used for primer extension analysis. The oligonucleotides were labeled at the 5′ end and specific activity was determined by Cerenkov counting by spotting aliquots on Nylon membranes. A total of 40 µg of isolated RNA was combined with 0.2 pmol oligonucleotide, coprecipitated, resuspended in 10 µl of annealing buffer, and incubated for 5 min at 80° C. Annealing conditions were optimized by variation of the KCl concentrations and the hybridization temperatures. The conditions were 30° C, 1000 mM KCl for the *cab* 1A, *cab* 1B, *cab* 1C, *cab* 1D, *cab* 3A, *cab* 3C, and *cab* 8 primers; 40° C, 500 mM KCl for the *cab* 9 primer; 40° C, 750 mM KCl for the *cab* 3B, *cab* 6, and *cab* 7 primers, and 40° C, 1000 mM KCl for the *cab* 4 and *cab* 5 primers. The MMLV reverse transcriptase (Gibco-BRL, Eggenstein, Germany) was used for the synthesis of the primer-extended ssDNA fragments, which were analyzed on 8–10% polyacrylamide/urea sequencing gels. Relative levels of the individual *cab* mRNAs were determined by cutting out the respective DNA band, subjecting it to Cerenkov counting, taking the specific radioactivity of each oligonucleotide solution into account.

Table 1. Sequence similarity between 5′ upstream regions and coding regions of the tomato *cab* genes

Genes compared[a]	Coding sequence	5′ upstream[b] sequence	GATA repeat[b]	TATA –mRNA start[b]
cab 1A–*cab* 1B	95	76(−226)[c]	79	79
cab 1A–*cab* 1C	95	84	82	95
cab 1A–*cab* 1D	93	85	87	98
cab 1B–*cab* 1C	97	81	82	79
cab 1B–*cab* 1D	98	77	79	76
cab 1C–*cab* 1D	98	76	77	93
cab 3A–*cab* 3B	98	46(−221)[c]	47	42
cab 3A–*cab* 3C	88	45	56	49
cab 3B–*cab* 3C	88	64	63	59
cab 6A–*cab* 6B	99			
cab 10A–*cab* 10B	93	71(−307)[c]		68
cab 11–*cab* 12		81(140)[c]	70	44

[a] Degrees of sequence similarity are expressed as percentages
[b] The upstream regions compared are shown in Figs. 5, 6 and 3A
[c] Numbers in parentheses indicate the lengths or position of sequences used for comparison

Sequence comparison. Sequence comparison was performed using the sequence alignment method of Needleman and Wunsch (1970) implemented in the UWGCG sequence analysis software package (Devereux et al. 1984) and visual inspection. Calculations of similarities (Table 1) are based on the alignment presented in Figs. 3A, 5 and 6; deletions were set as zero.

Determination of transcription initiation sites. Determination of transcription start sites was performed by the primer extension technique as described elsewhere (Kellmann et al. 1990) and by the S1 nuclease method (Sambrook et al. 1989). In some cases cDNA sequences were available which are compatible with the results of the two methods mentioned.

Results

Determination of mRNA levels of tomato cab *genes*

For exact quantification of the inidividual steady-state mRNA levels of the *cab* genes by the primer extension method, several precautions had to be taken. First, specific oligonucleotide primers were used which give rise to only one ssDNA fragment on extension (Fig. 1A). To achieve this, the annealing and primer extension conditions were optimized for each RNA/oligonucleotide combination. Based on the known positions of the primers (*cab* 1A, *cab* 1B, *cab* 1C, *cab* 1D, *cab* 3A, *cab* 3B, *cab* 3C, *cab* 7 oligonucleotides were complementary to sequences 5′ upstream of the translational start point, while for *cab* 4, *cab* 5, *cab* 6B, *cab* 8 and *cab* 9 the oligonucleotides were situated within 100 nucleotides downstream of the initiating ATG codon, within the sequence encoding the transit peptide), the lengths of the resulting primer-extended fragments were examined and verified either by S1 nuclease analysis or by comparison with a full-length cDNA clone sequence. Primers complementary to different positions in the DNA of a given *cab* gene should give rise to different ssDNA fragments but the signal intensity should remain the same. However, two pairs of oligonucleotides, one pair for *cab* 1C and one for *cab* 3A, revealed DNA fragments of different intensities. The determination of the expression levels (Fig. 1B) was based on the strongest signal; we hypothesize that in the case of the weaker signal, secondary or tertiary structure of the mRNA may have prevented complete binding of the primer.

Using this methodology, we determined the steady-state mRNA levels for 14 genes (Fig. 1B). This analysis revealed that all genes included in this survey are expressed in tomato leaves, however the levels were significantly different. The most abundant transcripts corresponded to the *cab* 3B and *cab* 1B genes, accounting for respectively 28.4% and 20% of the total *cab* mRNAs. This level of expression is followed by *cab* 1A

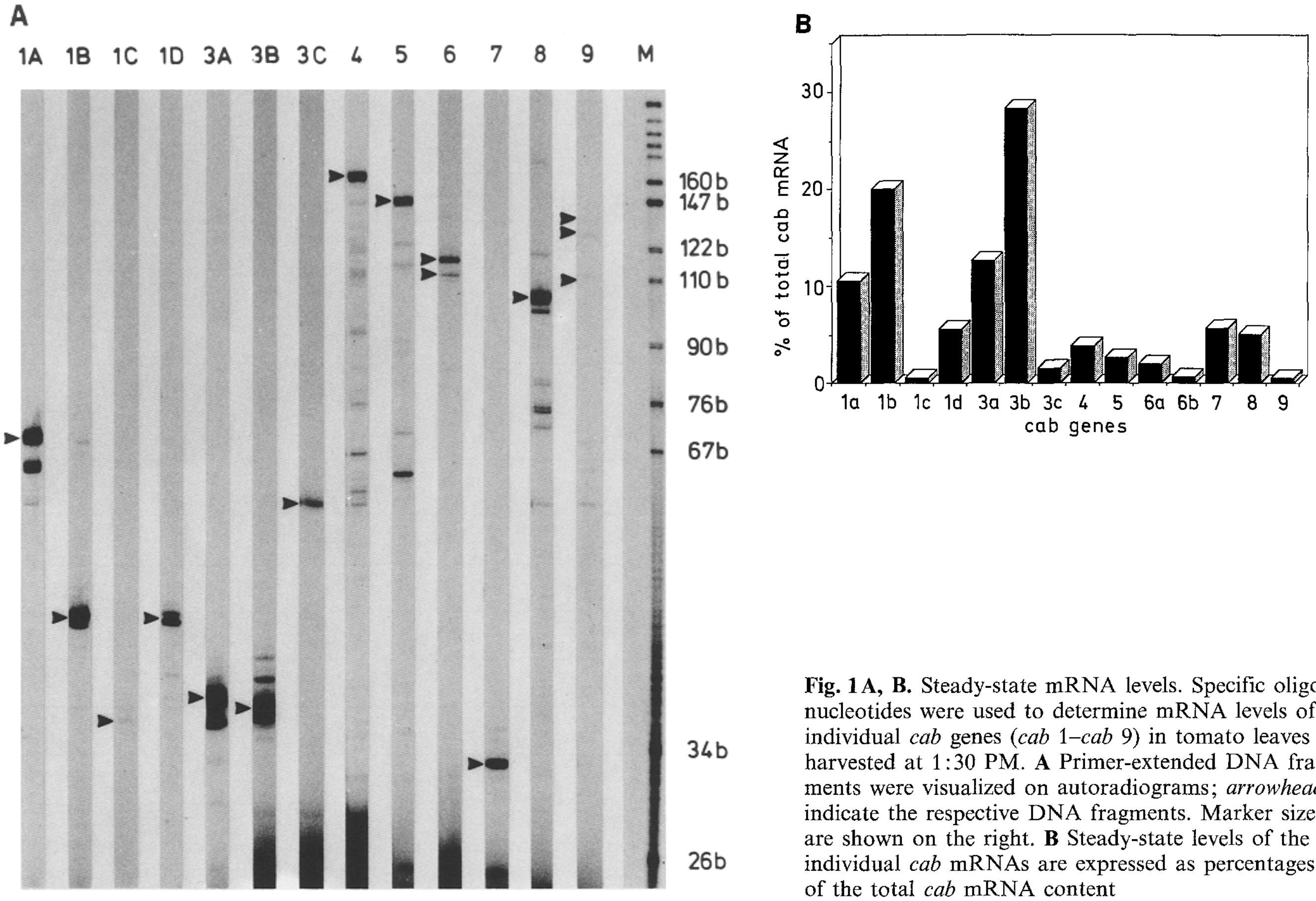


Fig. 1A, B. Steady-state mRNA levels. Specific oligonucleotides were used to determine mRNA levels of individual *cab* genes (*cab* 1–*cab* 9) in tomato leaves harvested at 1:30 PM. **A** Primer-extended DNA fragments were visualized on autoradiograms; *arrowheads* indicate the respective DNA fragments. Marker sizes are shown on the right. **B** Steady-state levels of the individual *cab* mRNAs are expressed as percentages of the total *cab* mRNA content

(10.6%) and *cab* 3A (12.7%). The mRNAs of each of the residual *cab* genes analyzed accumulate to approximately 5% or less.

Transcription initiation sites

The primer extension method was also used to identify the transcription start site for each member of the tomato *cab* gene family. In the case of *cab* 1A, *cab* 1B, *cab* 1C, *cab* 3B, and *cab* 3C the transcription initiation site was also verified by S1 nuclease digestion experiments, while sequences of cDNA clones confirm the correct start point for *cab* 4, *cab* 6B, *cab* 7, *cab* 8, and *cab* 11. The transcription start sites of almost all tomato *cab* genes examined were localized 22 to 28 nucleotides downstream of the putative TATA box. Exceptions to the rule are *cab* 6B and *cab* 11, where this region is 37 or 19 nucleotides long, respectively. The distances between the transcriptional and translational start sites vary between 37 and 80 nucleotides (Fig. 2).

A comparison of the 5′ ends of the different *cab* mRNAs is presented in Fig. 3A. In seven cases the first nucleotide of the *cab* mRNA was determined to be T, in four cases a C and in three cases an A. In 11 instances, the sequence centered around the 5′ ends is TCA. The quantitative distribution of the nucleotides around the putative 5′ end of the different tomato *cab* mRNAs (Fig. 3B) strongly suggests that the first three nucleotides are indeed TCA, downstream of the TCA triplet, there is less evidence that particular nucleotides are favoured in particular positions.

To determine whether this common transcription initiation site of the tomato *cab* gene family is also present in *cab* gene sequences of other plant species, we surveyed all *cab* sequences presently compiled in the EMBL database and have indicated the transcription start sites based on published data (Fig. 3C). We also compared the transcription initiaton sites for members of large gene families from species other than tomato. A TCA motif located at or close to the transcription initiation site was detected for the *Arabidopsis thaliana cab* 2 and *cab* 3 genes (Leutwiler et al. 1986; Karlin-Neumann et al. 1988; Mitra et al. 1989), *Glycine max cab* 5 (Walling et al. 1988; Demmin et al. 1989), *Nicotiana plumbaginifolia cab* E (Castresana et al. 1987; Castresana et al. 1988), *Zea mays cab* 1 (Sullivan et al. 1989), *Pisum sativum cab* 805 and *cab* 80 (Cashmore 1984; Simpson et al. 1985), and *Petunia hybrida cab* 22R gene (Dunsmuir 1985; Stayton et al. 1986; Gidoni et al. 1989). In cases where the transcription start sites were not published (*Hordeum vulgare cab* 2, Chitnis et al. 1988; *Lemna gibba cab* 19A and *cab* 30, Karlin-Neumann et al. 1985; Kohorn et al. 1986; *Oryza sativa cab* R1 and *cab* R2, Luan and Bogorad 1989; *Physcomitrella patens cab*, Long et al. 1989; *Triticum aestivum cab* 1, Lamppa et al. 1985)

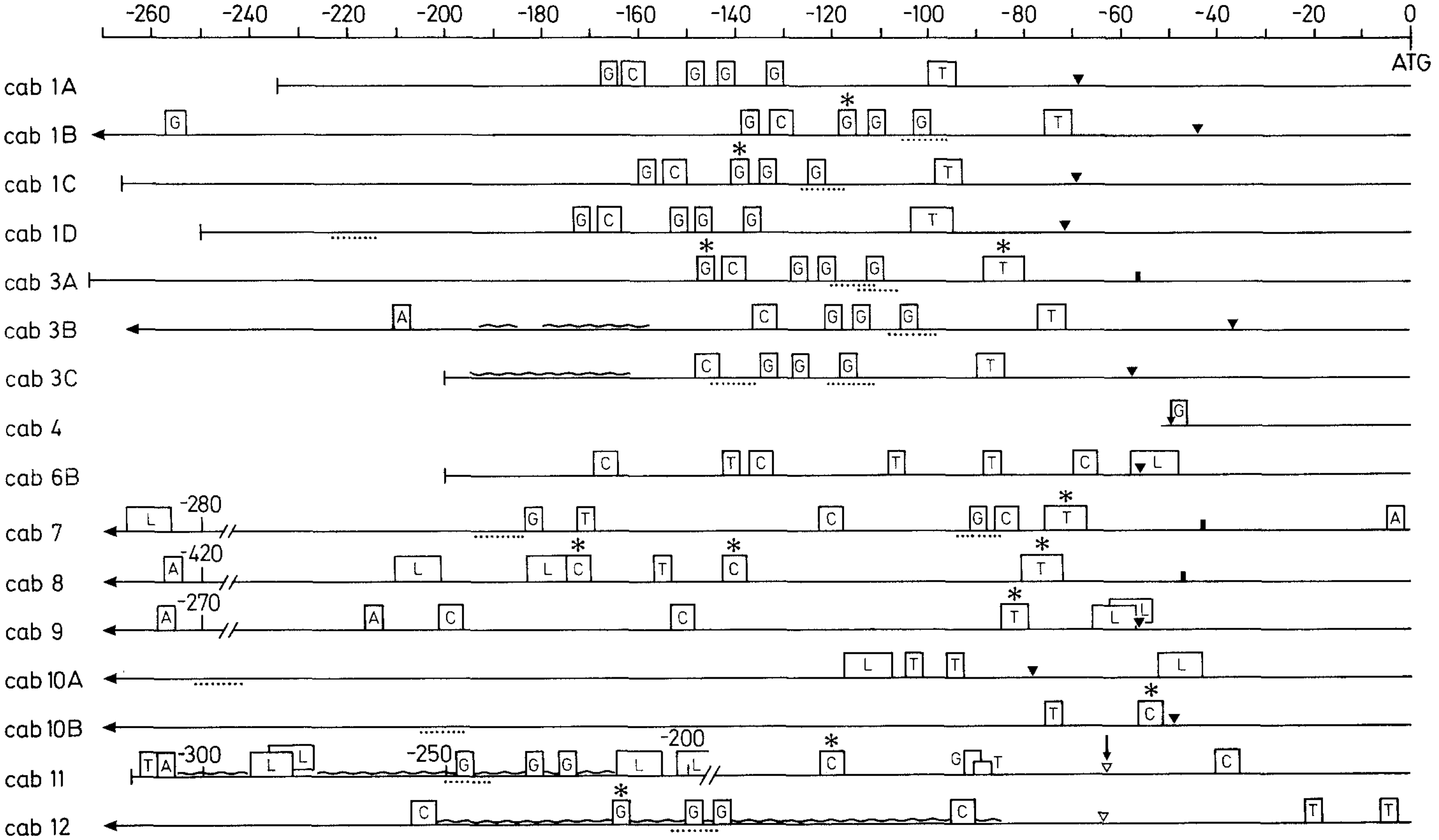


Fig. 2. Schematic representation of sequence motifs and repeats in 5′ upstream sequences of the tomato *cab* genes. Sequences upstream of the translational start site (ATG) of each *cab* gene are represented as lines. T represents the TATA box; C, the CCAAT box; G, the GATA motif; A, the ACGT core of the G box (Weisshaar et al. 1991); L, the CCTTATCAT light-responsive element; the *dotted line* indicates similarity to the ATGATAAGA sequence. Mutated sequences are denoted by *asterisks*; transcription start sites beginning with a TCA sequence are indicated by the *arrowheads*; transcription start sites that lack TCA sequences are denoted by *black bars*. The *open arrowheads* indicate putative TCA transcription start sites; the *arrow* indicates the first nucleotide of the corresponding cDNA sequence. The *wavy line* indicates regions of similarity between either *cab* 3B and *cab* 3C or *cab* 11 and *cab* 12

the nucleotide sequence centered around 25 to 30 nucleotides downstream of the putative TATA box was examined and is given in Fig. 3C.

5′ upstream sequences of tomato cab *genes*

We have determined the nucleotide sequence in the promoter region of most of the genes under investigation for which these sequences were not previously available (Fig. 4). These include the seven genes in the two clusters of genes in loci *cab* 1 and *cab* 3. Short upstream sequences were previously reported for most of these genes (Pichersky et al. 1985), but reexamination of these regions revealed errors; thus, the promoter sequences published here for *cab* 1 and *cab* 3 genes replace the previously reported ones when they differ. Figure 4 also shows the promoter sequences of *cab* 1B and *cab* 3C (Pichersky et al. 1985), *cab* 4 (Pichersky et al. 1987a), *cab* 6B (Pichersky et al. 1987b), *cab* 7 (Pichersky et al. 1988), *cab* 8 (Pichersky et al. 1989), *cab* 9 (Pichersky et al. 1991), *cab* 10A and *cab* 10B (Schwartz and Pichersky 1990), *cab* 11 and *cab* 12 (Schwartz et al. 1991).

Best fit alignment of 5′ upstream sequences

Visual inspection of the aligned 5′ upstream sequences in Fig. 4 does not reveal substantial sequence similarities between the *cab* genes. However, using the sequence comparison program for a best fit alignment of *cab* 1A, *cab* 1B, *cab* 1C and *cab* 1D or *cab* 3A, *cab* 3B and *cab* 3C or *cab* 10A and *cab* 10B significant similarities are revealed (Fig. 5). The sequence identity within the *cab* 1 subfamily is 76% to 85% (Table 1). In the case of the *cab* 3 subfamily, the similarity between *cab* 3A and *cab* 3B was calculated to be 46%, while *cab* 3B and *cab* 3C are even more similar (64%). Analysis of the nucleotide sequences of *cab* 10A and *cab* 10B revealed 71% similarity. These data demonstrate that i) the promoter regions of the genes which are in close proximity on the same chromosome of the tomato genome are more similar to each other than to promoters of genes unlinked to them; and ii) for each gene class, the sequences of the promoter regions are less similar than the sequences encoding the transit and mature proteins (Table 1).

A computer analysis of the promoter sequences of

A

```
                                              ●
cab 1a        TATATA..TGGTGAATTAATTCCCTTGTAACTTCATC.TCATCACA
                                              ●
cab 1b        TATATA..TTCTCAA.....CCCCAACTAACTTCATCTTCATCACC
                                              ●
cab 1c        TATATA..TGCTGAA.TAATTCCCTTGTAACTTCAAC.TCATCACA
                                              ●
cab 1d      TATATATA..TGGTGAATTAATT.CCTTGTAACTTCATC.TCATTACA
                                               ●
cab 3a      TATAAATA..GTGTTAT.TAATCACAAAATGAAA.CATAACAACAACC
                                              ●
cab 3b        TATATA..CACTTCGGTGACTCAAGCCTCAAATCATCTCTTCTTTT
                                               ●
cab 3c        TATATA..CAGTTAGTGCAAAGCTCATGAAACTCAAGCTTCAAAAC
                                                ●
cab 6b          TATACATTCCACAATTACACCAATTTTTTTTCATTCTCATATCC
                                              ●
cab 7     TAATAAAATA..CCACAA.AATCTCATTGTCCTTGGT.ATCTCTCATAAT
                                                ●
cab 8        TAATTTA...TTCTTTGTGGAGCTAAGT..GTTT.ATATTATTCTTC
                                               ●
cab 9          AAATA..CCATCCTCAATT.CTCACTTCTCATCATCAACTCGACC
                                               ●
cab 10a        TATA..AACTTAT   ATCTCACTACTTCATTCATAGAAGAGACA
                                              ●
cab 10b        TATA..ACCAGGA   ATCTCACTTCAATATTCAAAAAACAGAAA
                                                ▽
cab 11          TATA..CAAAACATACAAAGTT.......TTCAACTCAGCCTTA
cab 12             CCCAATCCCAGTCGTCAAAATTCCCATTCAAATACAAGCAC
```

B

	5'end of mRNA									
nucleotide	1	2	3	4	5	6	7	8	9	10
T	13	–	–	10	1	5	7	3	3	6
A	–	–	14	4	4	3	3	5	6	4
C	–	12	–	–	8	2	4	6	3	4
G	–	–	–	–	1	1	–	–	2	–
deletion	1	2	–	–	–	3	–	–	–	–

C

```
                                                  ●●●● ○
Arabidopsis          cab1     TATATATA    15 ATACCAAACCACCCA   64 ATG
thaliana                                             ○
                                                  ●●●●
                     cab2     TATATAAT    18 TTTTAATCACTCTCA   48 ATG
                                                     ○
                                                  ●●●●
                     cab3     TATATAAT    18 TTTCAATCACTCTCA   48 ATG
Glycine max          cab1     AATAA*      14 TTTCAGCCGCATAGTT  27 ATG
                     cab2     TATATA      24 ACCCCTCTCTTCACTT  42 ATG
                     cab3     TATAAATA    20 AAGTCACTCACCACT   52 ATG
                                                  ●
                     cab4     TATATATA    18 AACTACAAGCAACAT   67 ATG
                                                  ●
                     cab5     TATATATA    14 TATGAATCAACAATT   55 ATG
Hordeum vulgare      cab2     TATTA       20 TTACTAGCATGGTTA   25 ATG
Lemna gibba          cab19A   TATTAA      20 CTCTCTCTATCTCCTC  55 ATG
                                                   ●
                     cabAB30  TATTAAA     22 TATCCCTACACCACTC  58 ATG
Nicotiana            cabC     TATATAAATA  15 CTACATCACCACAGCT  56 ATG
plumbaginifolia                                    ●
                     cabE     TATATA      19 GAAACTCAAGCCTCA   49 ATG
                     cabF     TAAATA      16 TACATGACCACAGCT   52 ATG
Oryza sativa         cabR1    TATTTAAA    21 CCAACTCACACTCGCT  43 ATG
                     cabR2    TATTAATA    20 TGACACAGCACGCACA  58 ATG
                                                  ○
Petunia      sp.     cab13    TTATAA      20 TTGTTAGTAGCTGCGTT 35 ATG
                                               ○  ●  ●  ●
             sp.     cab22L   TATATATA    19 CTAACAGCACCACAG   58 ATG
                                                  ○
                                                  ●
             sp.     cab22R   TAAATAAA    19 AACTCATCAACTCTT   48 ATG
                                               ○
             sp.     cab25    TATTTATATA  15 GCCAATAAAACTCAAG  65 ATG
                                                         ○
             hybrida cab37    TATTTTT     19 CCAAGCAATATAGCAGA 75 ATG
                                             ○
             sp.     cab91R   TATATATA    20 AATTCAAGCAACAAAT  56 ATG
Physcomitrella       cab      TTATAAA     20 GACGGCATCGAGGCAG  72 ATG
patens                                             ○
Pisum sativum        cab80    TATAAATA*   20 AAAATCACCATTGATA  58 ATG
                                                   ●●●
                     cab805   ATTATAAA    21 CAAAATCACCATTGA   84 ATG
                                                  ●●●●
Triticum aestivum    cab1     TTTAAATA    17 CTCTTAAACCATCT    63 ATG
                                                  ●●●
Zea mays             cab1     TATTTA      21 ACACTCCACCAGCGG   43 ATG
```

Fig. 3A–C. Transcription initiation sites **A** 5′ upstream sequences beginning 30 to 40 nucleotides downstream of the TATA box of 16 *cab* genes are presented. Definitive transcription start points are indicated by *dots*. The *open triangle* indicates the first nucleotide of a cDNA clone sequence. The nucleotide sequence TCA surrounding the transcription initiation site is indicated in *boldface*. **B** Nucleotide composition of the 5′ ends of the *cab* mRNAs, starting with the putative conserved sequence TCA. The *cab* 12 sequence (Fig. 3A) was not included in this calculation. **C** Compilation of all presently available 5′ upstream sequences of *cab* genes of other plant species. The regions between the TATA box and the translational start point are indicated and the sequences surrounding the (putative) transcriptional start sites are presented. The *closed circle* indicates the transcriptional start site based on the original publication, the *open circle* indicates the transcriptional start site presented in Joshi (1987), and the nucleotide sequence TCA at the transcriptional start site is printed in *boldface*

```
            -397
cab 1b      CTTGACCAGT AAACTCTAAA ACCAGAGAGA ATAGATGCAT CAAAAAAAAA AGACTAATGG ACCTACATTG AATGAGCTAG CTGTAGAAAC AGATGGTATC
cab 3b                                                                                                         CTTTCTTATT
cab 7       CATCGATGAA TAACCAAATG AAGAGTTGAA GAATGAAGCA TGAAAAAAAC AAGTGAATAA GGAAAAATAA ATATTTACTT TGTTAACAAA AAACAACAAA
cab 8       ACTTGTAGTT ATCACAAGAA ACACAAATTT CAAATTCAGA ATCCGTCACG AGAAAAAACT ATATACCCTA GTAAATCTAT GAGAAGGCAA GGTGGCAATA
cab 9                                                                                       TAAAGAA AGTATTGTGG CATTGCAAAT
cab 10b                                                             AACCTCACT TTCGTTTTTA GCGCAATGTT GATTACAAAC TTCTTTTTTT
cab 11                                                                                        ATATA TCTACGTGGC AAATTTTTGG
cab 12                                                                                                          TTTATTATC

            -297
cab 1a                                                                        CCCTTAAGTA ATAAACATCA TGCAGATTGG AGATTGCCAA
cab 1b      TGTTATCCTG CTGGTAGAGG TGTCATAATG TAAGCTCACG AGATAGAGTG AAATCCTTGT AACCATACAT AGGAGGACAA CATTGGTTGG GTCCCCCTCG
cab 1c                                   GTA GCCAATTAAA GGTGGACAAC ATTAGTTGGG TCCCCACTGT AAACATCCTG TAAAATTGTT GGACTAGAGA
cab 1d                                                     GGGTC CCTTCAGTAA ACATCTTAAA AAATTAGAAG GATTAGAGAT TGTCAATTTG
cab 3a                         T CTAGACTAAA GAATCTTACT TGACAGCATA GAGGACGAAG TTTGCAATAT TCATAGCCAC ATATTTGTTG GACCCCATTA GTAAAAATAT
cab 3b      TTGTCTCATT GTCAATTTGC CACACATAAA AGTGGTCCTA TTTCAATGTC AGCCAAATAA ATTTAATTGC AATAATCAAC AATCATACGT GGCGACTATT
cab 3c                                                                                                                CTGCA
cab 6b                                                                                                    ATAT CCTCAAAAAT
cab 7       TTCATTATGA TCTCATAAAA GATTCGAAAA AAATAAATAT ACACAAATTT TACCGAATCG AATCGAGAAG TACTTAATTC CAACCACTAT TAGAATGGGG
cab 8       CCAAACAAGA GTATCTAAGA TTTTCATTTG TGACTATAGG AATATAATAT CTCTTATCTG ATTTAATGAA TCCACATGTT CACTTCTCAT TTGTCCACAA
cab 9       TGCAATGTGG ATAGGGTCAC GTTATCCAAG TCTGGTCCAG GAGTTTTGCG ATGCATGCCA TGTCAGCACT GTCATCTGCC ACGTCACAAA CTCCTTCCAA
cab 10a         TAGAGT ACAGAAATAA CAACTAAGAC AGAGAATCAA AACTAAAGGA GAAGGGAGTG TCCACGGGTA TGGTACAAAT AAGGACAGAG ATGTAACTCA
cab 10b     CTCTATTTTC ACTAAGATGT CATTTTCTTA GGAACCCAAT ATCTTCTTAT TAGAGAGAGG AGAATTACAG AGAGGTGTCC ACGGGTATGG TTATGAATAG
cab 11      GTCCCTCCTC ATCTTCTCAA ACCAACAGAA ATAAAGAAAC AGATTTGAAG ATATAGAAGC AAAGATAAGA GATAAGGCAC ATTTCTTCTC ATTGGTTCTT
cab 12      GCCTGAAATC TCCATGTGAG ACATGTTTTC TTACTAAAGA CTTATTACTA ACCAAATGTC AAATTCTTAT TCGACAACAG CAGATTTCCC AATGGCAAAT

            -197
cab 1a      GTGCATTAAT CGCTACACAT GGGATCTTGA TACCCAATGA GATTATAGAT ATAGATATCA CTAGATAATT ACGGCTCTTT CCCTCTTTCT TAATCCCTAT
cab 1b      TAAAATCATG AAGAAGTTGA TGGATTATAG ATTGCCAAGT GTGCTACACA TGGGATCTTG ATACCCAATG AGATCATACA TATAGATATC ACTTGATAAG
cab 1c      TTGCCAAGTA GCATTACTTG CTGTATATGG GATCTTGATA CCCAATGAGA TCATAAATAT AGATATCACT AGATAAGGAC TCTTTCCCTC TTAATCCCTA
cab 1d      CATTCATTGC TACACTTAGG ATCTTGATAT CCAATGAGAT CATAGATATA GATATCATTA GATAATTTGG ACTCTTTCCC TCTTAATTAC TCCCTATATA
cab 3a      CTTGAAAATA TTGATGGGTT ATAAAATGCC AAGTGCCTAA TAAAATCTTG AAAACCAATG AAATTGTAGA TAGAGATATG ATAAGATAAG AACCAACCAT
cab 3b      GATTGGATGT AAGAGAGAAA ATCTTCATTG GGTTAGATTT TTTAACAAGT ATCTAGTGAT GTTTAATCCC ACCAATGAAA AAAGAGATAT AGATATTCAT
cab 3c      TAAGAGGAAA CAACATTTGA GTTAGATTTT TTAAAAAATT GTCATTCACC AATGAAAAAG CAGATAATGA TATTCTAAGA TAAGGATTTT GGGCCTGTTG
cab 6b      ATCCACTTCA TCTTCCAGGT GGACCAATCA CAAAACAGAA ACCAGTCTCT ATAGCCAATA ACAATCTAAA TTCAGATTAG CCTATATCTC AAACTCCCTC
cab 7       ACATGATGAG TGTGATAGAG GGGGTATAAG AACCACAATA TTGGGTTGTG GTTGCCACAT GGCAATTTAA GTAGCCAATC ACATATTGAC TCTTCTATCC
cab 8       GATCACAACT TTATCTTCAA TATTCACAAC TTGTTATATC AACCACAACA ATTTCTATTC TTTTCACTCA GTCCCACAAA ATACTTTGTT CCCTTATTTG
cab 9       TCAAAACGCC GCATCAGATT CCACATTCAT CTTCCCTGTG GCTCCAATAA CAGAGCGACA CTTGTCACGC TCTCATCCAC CAGAAAAAGC TTTAAGCTAG
cab 10a     ACACCTTATT GGTCCGAAAT CTATCCACCA GAGATCATTT GCAGATTTCA TTTATCCTAC TTGGCTCCTT ACAAGGTCCC CTTTTGATCT TATAAACTTA
cab 10b     GGGACATAGA GATGGAGCTC AACAACTCAT TGGTCAGAAA TCTATCCACT AGATTTCATC TGCAGATTTC GTTTATCCTA CTTGGCTACT TACTATTTGC
cab 11      ATCAAATGCA CAACTCAATC CAACCGTTGG ATTCCAAAAA ATCTTAGTGA CTCTTCAAGA TTGAACCAGA GCCACAATTG TCCCAAAGAA ATGGCAATAC
cab 12      ATTTTGGTCC CTTCTAACAA AGAAATAGAT TCTTTGGATC ATAGCCACAA AGATAAGGAT AACACACATT TCTTCTAATT GGCTACTCTC TAATTCACAA

            - 97                                                                                                    +1
cab 1a      ATATGGTGAA TTAATTCCCT TGTAACTTCA TCTCATCACA GCCTTCAACA ATATTTAATA CCATAAAATA CTCAACACTT TTCTCTTAAT ATAAATCATG
cab 1b      ATGATTCTCT CTCTTTTCTC CTATATATTC TCAACCCCAA CTAACTTCAT CTTCATCACC CATCAAACAC TTAATTCTTC TCTTAAAATA AACACAAATG
cab 1c      TATATGCTGA ATAATTCCCT TGTAACTTCA ACTCATCACA GCAAACTTCA AAAAGTTTAC CATCAAACAC TTACATTTTC TCTTGATATA AACACAAATG
cab 1d      TATGGTGAAT TAATTCCTTG TAACTTCATC TCATTACAGC CAACTTCAAC AATATCTCAT ACCATCAAAC ACTTACATTT CTCTTGATAT AAACACCATG
cab 3a      ATTCCCTCTT ATAAATAGTG TTATTAATCA CAAAATGAAA CATAACAACA ACCATCGAAA ACACAATTCA TTTCTTTTTA TTTATTAAAA TTAAACCATG
cab 3b      GGATAAGGGT ATTGGGCTTG TGGAGTCATT TATATACACT TCGGTGACTC AAGCCTCAAA TCATCTCTTC TTTTTTTGTA CATTCTAAGA GTTCATAATG
cab 3c      AGTAATTTAT ATACAGTTAG TGCAAAGCTC ATGAAACTCA AGCTTCAAAA CAACTTTTCT TTTTGTACAT TCAAGAGTTT CTCATTCTAC TTCTATAATG
cab 4                                                G ATATCATCAG AAAGAAACAA AAAAGCTTAA GCAAATTAAA AAAAAAAATA AAAAAAAATG
cab 6b      CTATACATTC CACAATTACA CCAATTTTTT TTCATTCTCA TATCCAAACT TTTTTTTGTA CATCTTTTAA ATACCAAAAA AAAAAGAGGA AGAAGATATG
cab 7       ATCAAGATAA GCCAATTCTC ATAATAAAAT ACCACAAAAT CTCATTGTCC TTGGTATCTC TCATAATCAC AAACACAAGA GTGAAGAACG TGCCGACATG
cab 8       CCACCTTTTG TATTTAATTT ATTCTTTGTG GAGCTAAGTG TTTATATTAT TCTTCTTCTC AAAAAAACAA AAACAAACAA AAAAGAGAAG AGAAATTATG
cab 9       TATCTCCACT CCAAATACCA TCCTCAATTC TGACTTCTCA TCATCAACTC GACCTCAATT TTTTTTACCT CTTGCCAGCG ACACCGTTTA GCTACAAATG
cab 10a     TATCTCACTA CTTCATTCAT AGAAGAGACA CAAGAAACAA CCATCCATAT CTTTGCATAT TACTATCATT TCCATTCTTA AACATATCAA AACAAACATG
cab 10b     CCTCTTTTGA TCTCATGACC ATATAACCAG GAATCTCACT TCAATATTCA AAAAACAGAA ACTTCGTATT TGCAGTTTAC CACTATCCAA AATAAACATG
cab 11      AAATGATATA CAAAACATAC AAAGTTTTCA ACTCAGCCTT AAAACTACAT TGCCATTTCT CCCAATAATC ACCAACAAAC TCTTCAAATT GGAAAAAATG
cab 12      CCCAATCCCA GTCGTCAAAA TTCCCATTCA AATACAAGCA CATTTTTTGC ATACACACAA GTAAACATCA ATTATATTCA TAACATTAGA CTATAAAATG
```

Fig. 4. Sequence compilation. All presently known 5′ upstream sequences of tomato *cab* genes were aligned relative to the translational initiation site. The A of the ATG was set as +1

the single *cab* genes which are localized on either separate loci on the same chromosome such as *cab* 7 and *cab* 8, or on different chromosomes such as *cab* 9, *cab* 11 and *cab* 12, does not reveal significant similarities (data not shown). Unfortunately no sequences upstream of the transcription start site are available for the single genes *cab* 4 and *cab* 5. The comparison of *cab* 11 and *cab* 12 promoter sequences revealed an interesting feature: although the two genes are located on different chromosomes, chromosomes 3 and 6 respectively, 81% sequence similarity was identified between positions −175 to −311 of *cab* 11 and positons −87 to −206 of *cab* 12 (Figs. 2 and 5).

Search for known elements of light-regulated plant gene promoters

We have searched the *cab* gene promoters for several sequence motifs previously identified in several light-re-

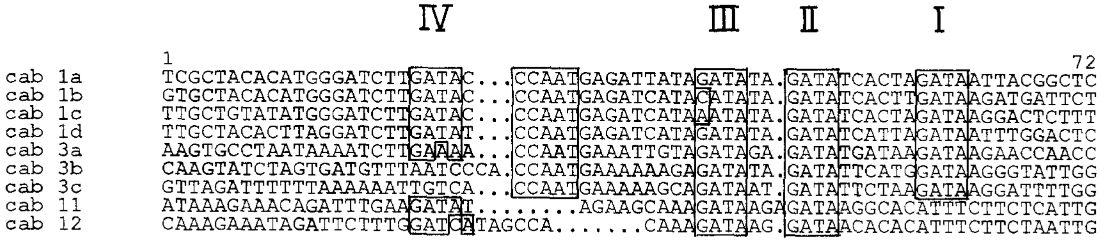
```
cab 1a                                          CCCtta aGTAA..... ..tAAacATc atgca..... .GATTgGAGA TTGCCAA.GT .GCATTaaT
cab 1b                                                    At..... ..gAAgaAgt tgAtg..... .GATTAtAGA TTGCCAA.GT .G.......
cab 1c      gtagccaatt aaaggtggac aacattagtt gggtCCCcac tGTAA..... ..acAtccTg taAaattgtt gGAcTAGAGA TTGCCAA.GT aGCATTacT
cab 1d                                       gggtCCCttc aGTAAacatc ttaAAaaATt agAag..... .GATTAGAGA TTGtCAAttT .GCATTcaT

cab 1a      cGCTACACATG GGATCTTGAT ACCCAATGAG ATtATAgATA TAGATATCAC TAGATAAtta cggctcTtTC cCTCTtTCTT AaTCCC..TA TATATGgTG
cab 1b      TGCTACACATG GGATCTTGAT ACCCAATGAG ATCATAcATA TAGATATCAC TtGATAA... ....gaTgat tCTCTcTCTT ttctCC..TA TATATtcTc
cab 1c      TGCTgtAtATG GGATCTTGAT ACCCAATGAG ATCATAaATA TAGATATCAC TAGATAAgga c.......TC ttTCccTCTT AaTCCC..TA TATATGcTG
cab 1d      TGCTACACtTa GGATCTTGAT AtCCAATGAG ATCATAgATA TAGATATCAt TAGATAAttt ggactcTtTC cCTCTtaaTT AcTCCCtaTA TATATGgTG

cab 1a      AAtTAATTCCC TTGTAACTTC ATC.TCATCA CAGC...ctt CaaCAAtAtt TaaTACCATa AAAtACTcAA cacTTtTCTc ttaATAtA.A A...tcATG
cab 1b      AA.....cCCC aacTAACTTC ATCtTCATCA C......... .......... .....CCATC AAACACTtAA ttcTTcTCTt aaaATAaAcA c...aaATG
cab 1c      AA.TAATTCCC TTGTAACTTC AaC.TCATCA CAGC...aaa CttCAAaAag T.tTACCATC AAACACTtA. .caTTtTCTc ttgATAtA.A AcacaaATG
cab 1d      AAtTAATT.CC TTGTAACTTC ATC.TCATtA CAGCcaactt CaaCAAtAtc TcaTACCATC AAACACT.tA catTTcTCTt gatATAaA.c A...ccATG


cab 3a                                                                                                A aaatAtcttG AAAATATTgA
cab 3b                                                                                                A TGtAAgAGAG AAAATcTTcA
cab 3c                                                                                                c TGcAtaAGAG gAAAcAacat

cab 3a      .TGGGTTAtA aaaTgccAAg tGccTa.... .ATaaAaTcT tgaaaACCAA TGAAAttGtA GATAgAGATA TgaTAAGATA AGaA..ccaa cCaTaTTccc
cab 3b      TTGGGTTAGA TTTTTTAAcA AGtaTctagt gATgttTaaT Ccc..ACCAA TGAAAAAagA GATAtAGATA TTCatgGATA AGGgTaTTGG GCtTGTgGAG
cab 3c      TTGaGTTAGA TTTTTTAAAA A......... .....ATTgT CattcACCAA TGAAAAAGcA GATAatGATA TTCTAAGATA AGGATtTTGG GCcTGTTGAG

cab 3a      TCtTaTAaAT AgtGTTAtTa atcAcaAAAT G..AAA.cat AaCaaCAAcc atcgaaaaCa caaTtcAttT cTTTTTaTTT ...ATTAAAA TTaaAccATG
cab 3b      TCATTTATAT A.cacTtcgG tgActCAAgc ctcAAA.TCA tctcTt.... .......... .........c TTTTTTtgTa CattcTAAgA gTtcATAATG
cab 3c      TaATTTATAT AcaGTTAgTG caAAgCtcAT G..AAAcTCA AgCtTCAAaa caacttttCt tttTgtAcaT TcaagagTTT CtcATTctAc TTctATAATG


cab 10a                                                                                                            tagAGtAcag
cab 10b                                                                                                            cttAGgAacc

cab 10a     aAATAaCaaC TaAgacAGAG AatcaAaAcT aAagGAGAaG ggagTGTCCA CGGGTATGGT acaaAtaAGG ....AcAGAG ATGtAaCTCA ACAcCTtATT
cab 10b     cAATAtCttC TtA.ttAGAG AgaggAgAaT tAcaGAGAgG ....TGTCCA CGGGTATGGT tatgAatAGG ggacAtAGAG ATGgAgCTCA ACAaCTcATT

cab 10a     GGTCcGAAAT CTATCCACcA GAgaTCATtT GCAGATTTCa TTTATCCTAC TTGGCTcCTT ACaA..gGtC CcCTTTTGAT CT........ TATAAaCtta
cab 10b     GGTCaGAAAT CTATCCACtA GAttTCATcT GCAGATTTCg TTTATCCTAC TTGGCTaCTT ACtAtttGcC CtCTTTTGAT CTcatgacca TATAAcCagg

cab 10a     tATCTCACTa CttcATTCAt AgAAgAGAcA Caagaaacaa ccatccatat ctttgcatat tactatcATT TcCAtTcTta aACatatCAA AAcAAACATG
cab 10b     aATCTCACTt CaatATTCAa AaAAcAGAaA Cttc...... .......... .......... .....gtATT TgCAgTtTac cACtatcCAA AAtAAACATG


cab 11      TGGCAAATtT TTgGGTCCCT cCTcatcttc tcaaaccaac agaAAtAAAG AAAcAGATT. TgaaGAT.AT AGaagCAAAG ATAAGaGATA AggCACATTT
cab 12      TGGCAAATaT TTtGGTCCCT tCT....... .......... ...AAcAAAG AAAtAGATTc TttgGATcAT AGccaCAAAG ATAAG.GATA AcaCACATTT

cab 11      CTTCTcATTG GtTctTaTCa AATgCACAAC tCAATCCaA - (-174 upstream of ATG)
cab 12      CTTCTaATTG GcTacTcTCt AATtCACAAC cCAATCCcA - (- 87 upstream of ATG)
```

Fig. 5. Sequence alignment. 5′ upstream sequences of the subfamilies *cab* 1, *cab* 3, *cab* 10 and *cab* 11/12 were aligned by visual inspection and using a computer program. Deletions were included to maximize degrees of identity. The putative TATA and CCAAT boxes, the GATA repeats and the transcription initiation sites are indicated. The alignment of the subfamilies *cab* 1, *cab* 3, and *cab* 10 were arranged to start with the translational start point. Sequences upstream of −174 of *cab* 11 and upstream of −87 of *cab* 12 were also aligned. Bases are given in *upper-case letters* where 3 out of 4, 2 out of 3 or 2 nucleotides were identical

gulated plant gene promoters (Figs. 2 and 6). The eukaryotic RNA polymerase II-specific binding sites CCAAT and TATA were found in almost all *cab* genes (Fig. 2). The TATA sequences of all *cab* genes, except *cab* 6B, are localized approximately 25 nucleotides upstream of the transcription start site. The distance from the putative TATA box to the putative CCAAT box is 50–59 nucleotides and highly conserved within the gene clusters *cab* 1 and *cab* 3. The putative boxes are separated by 42, 58 and 64 nucleotides in the cases of *cab* 7, *cab* 8 and *cab* 9, respectively. No CCAAT box was identified in the 5′ upstream sequences of *cab* 10A and *cab* 10B (only CAAT in the case of *cab* 10A).

The so-called GATA box was identified in several *cab* genes from different plant species (Castresana et al. 1987; Gidoni et al. 1989). Four of the GATA sequences were detected in 5′ upstream sequences of the *cab* 1 gene cluster, three in the *cab* 3 gene cluster and in *cab* 11 and *cab* 12 (Figs. 2 and 6). In the case of the *cab* 1 subfamily one GATA motif was found upstream (IV), while three (I–III) were located downstream of the putative CCAAT box. The *cab* 3B and *cab* 3C upstream sequences lack GATA box IV, and box I is not present in *cab* 11 and *cab* 12 upstream sequences. Using the GATA motif with the surrounding sequences as a basis for calculations of sequence similarities (Fig. 6), the sequences of the *cab* 1 gene cluster are 76% to 86% similar to each other, while the *cab* 3 gene cluster are 43% to 64%, and *cab* 11 and *cab* 12 are 70% identical (Table 1).

The consensus sequence CCTTATCAT was predicted to be a feature of all light-regulated genes (Grob and Stüber 1987). It should be noted that this sequence contains within it the sequence complementary to GATA. A computer search using the motifs CCTTATCAT and ATGATAAGG, allowing 2 mismatches, uncovered 29



```
cab 1a   TCGCTACACATGGGATCTTGATAC...CCAATGAGATTATAGATATA.GATATCACTAGATAATTACGGCTC
cab 1b   GTGCTACACATGGGATCTTGATAC...CCAATGAGATCATACATATA.GATATCACTTGATAAGATGATTCT
cab 1c   TTGCTGTATATGGGATCTTGATAC...CCAATGAGATCATAAATATA.GATATCACTAGATAAGGACTCTTT
cab 1d   TTGCTACACTTAGGATCTTGATAT...CCAATGAGATCATAGATATA.GATATCATTAGATAATTTGGACTC
cab 3a   AAGTGCCTAATAAAATCTTGAAAA...CCAATGAAATTGTAGATAGA.GATATGATAAGATAAGAACCAACC
cab 3b   CAAGTATCTAGTGATGTTTAATCCCA.CCAATGAAAAAAGAGATATA.GATATTCATGGATAAGGGTATTGG
cab 3c   GTTAGATTTTTTAAAAAATTGTCA...CCAATGAAAAAGCAGATAAT.GATATTCTAAGATAAGGATTTTGG
cab 11   ATAAAGAAACAGATTTGAAGATAT.......AGAAGCAAAGATAAGAGATAAGGCACATTTCTTCTCATTG
cab 12   CAAAGAAATAGATTCTTTGGATCATAGCCA.......CAAAGATAAG.GATAACACACATTTCTTCTAATTG
```

Fig. 6. GATA repeats. The GATA motif and surrounding sequences of *cab* 1, *cab* 3, *cab* 11 and *cab* 12 were aligned. The putative CCAAT box and the GATA repeats are indicated. GATA boxes are numbered as proposed by Gidoni et al. (1989)

positions of sequence identity at different locations in the 5′ upstream sequences of the tomato *cab* genes (Fig. 2). The latter matrix, containing a GATA sequence, was found 16 times and the majority of these is present in GATA repeat motifs. Sequences homologous to CCTTATCAT were detected in the 5′ upstream sequences of *cab* 6B, *cab* 7, *cab* 8, *cab* 9, *cab* 10A, and *cab* 11. Our analysis shows that the 5′ upstream sequences of the tomato *cab* genes either contain the GATA motif or the light-responsive element CCTTATCAT, consistent with the observation that one sequence is the complement of the other. However, exceptions to this rule are encountered in *cab* 6B, where both elements are not found, and in the *cab* 11 promoter, where both motifs are present.

Discussion

Expression levels and patterns

The expression of *cab* genes encoding the Type I polypeptide of LHC II has been extensively investigated in several species with respect to light-dependence, developmental control and/or organ specificity. It was found that the Type I LHC II *cab* mRNAs in all plants investigated so far accumulate in green tissue and after illumination (with the exception of gymnosperms, where *cab* mRNAs accumulation is observed in the dark [Jansson and Gustafsson 1990; Alosi et al. 1990]). Furthermore, the steady-state level of mRNA is highest in leaf tissue, lower in other green organs, and below the limits of detection in roots. In this report, we have extended the investigaton of the *cab* gene family to examine the expression characteristics of six additional genes belonging to five other subfamilies of the *cab* gene family (*cab* 4, *cab* 5: PS II, Type II; *cab* 6B: PS I, Type I; *cab* 7: PS I, Type II; *cab* 8: PS I, Type III; *cab* 9: CP29), as well as seven members of the PS II Type I subfamily (*cab* 1A, *cab* 1B, *cab* 1C, *cab* 1D, *cab* 3A, *cab* 3B, *cab* 3C). Our results indicate that all these genes are expressed in leaf tissue, although the levels of expression vary significantly within and among the different subfamilies. For example, in the PS II Type I subfamily, *cab* 1A, *cab* 1B, *cab* 3A, *cab* 3B are highly expressed, whereas the mRNAs of *cab* 1C, *cab* 1D, and *cab* 3C are more weakly expressed. In general, steady-state levels of mRNAs of *cab* genes encoding PS I and CP29 proteins were 2- to 15-fold lower than those of the highly expressed PS II *cab* genes.

The fact that all the *cab* genes are expressed in leaf tissue strongly suggests that all types of CAB polypeptides are required for efficient light harvesting. Consistent with this hypothesis is the finding by Ikeuchi et al. (1991) of all the PS I CAB proteins in both pea and spinach thylakoid membranes. The reason for the presence of multiple copies of genes encoding the same type of CAB polypeptide is less clear. For example, why should tomato have seven or more genes encoding the PS II Type I CAB polypeptide, and why are all of these genes not expressed at the same level?

Transcription initiation site

Our determination of transcription initiation sites revealed that most such sites included the sequence TCA (Fig. 3A). In some other plant species, the same sequence was found to be present at or close to the transcription start point of *cab* genes (Fig. 3C). While our results lead us to propose the tomato *cab* mRNAs start with a T, Joshi (1987) indicated the A of the sequence TCACC as the first nucleotide of *cab* genes from *Pisum sativum, Petunia hybrida* and *Arabidopsis thaliana*. A generalization about the significance of the transcription initiation site of a gene in the *cab* gene family is presently not possible, since the determination of the transcription initiation site by both primer extension and S1 nuclease techniques is somewhat uncertain (±1 nucleotide), and also because in many published *cab* sequences of various plant species information about the 5′ nucleotide of the mRNA is lacking. However, it has been noted that the nucleotides CA often appear at the transcription initiation site of other plant genes as well, such as cereal storage proteins, dicot storage proteins, leghemoglobin and nodulins, enzymes, actin and lectins (Joshi 1987), and the sequence CAA was frequently found in the cases of ribulose-1,5-bisphosphate carboxylase small subunit genes (Morelli et al. 1985).

Comparisons of promoter sequences of different cab *genes*

A search for characteristic sequence motifs in the 5′ upstream regions of each tomato *cab* gene, extending up to 400 nucleotides upstream of the translation start point was carried out. A summary of sequence motifs found is presented in Fig. 2. In tomato the GATA repeat motif was found in upstream sequences of the *cab* genes in the *cab* 1 and *cab* 3 loci, and in the promoters of *cab* 7, *cab* 11 and *cab* 12. Four GATA repeats were found in the 5′ upstream sequences of the *cab* 1 subfamily and in *cab* 3A; *cab* 3B, *cab* 3C, *cab* 11 and *cab* 12 contain three repeats, *cab* 7 promoter has two boxes, and in the *cab* 6B, *cab* 8 and *cab* 9 upstream regions no GATA motif is present. Three GATA repeats have been described for each of several other photoregulated genes from various plant species (Castresana et al. 1987; Grob and Stüber 1987; Gidoni et al. 1989), and two were identified in the promoter region of the 35S gene of cauliflower mosaic virus (CaMV) (Lam and Chua 1989). A binding factor, GA-1, was found to bind at the GATA motif of the *cab* E gene of tobacco (Schindler and Cashmore 1990) and the ASF-2 factor interacts with the GATA-containing *as-2* site of the CaMV 35S promoter (Lam and Chua 1989).

The lack of the GATA repeats in the 5′ upstream sequences of some of the tomato *cab* genes correlates with the presence of a CCTTATCAT sequence (L-motif, Grob and Stüber 1987). This sequence was proposed to be located directly upstream of the TATA box in all known light-responsive genes (Grob and Stüber 1987). A computer search using this sequence as a matrix identified this sequence at various positions in the tomato *cab* 1D, *cab* 6B, *cab* 7, *cab* 8, *cab* 9, *cab* 10A, *cab* 10B, and *cab* 11 5′ upstream sequences (Fig. 2); however, conservation of spacing or location was not apparent. It should be noted that within the L-motif the nucleotide sequence GATA is present on the complementary strand. This analysis supports the idea that either GATA repeats and/or L-motifs are present in the 5′ upstream sequences of light-dependent, phytochrome-regulated plant genes, including the tomato *cab* genes.

The analysis of the 5′ upstream sequences of the tomato *cab* gene family also indicated that the promoters of genes which are tandemly linked share extensive sequence similarity. This observation suggests that during the process of gene duplication not only coding sequences but also significant stretches of flanking regions are duplicated. Alternatively, gene conversion events must have involved flanking as well as coding regions. Additionally, this analysis demonstrates clearly that the overall sequence similarity of the 5′ upstream sequences as well as the absence/presence of specific motifs and repeats does not necessarily imply similar expression levels. However, it should be noted that the L-motif, or the GATA motif are present in the 5′ upstream sequences of all light-responsive tomato *cab* genes. At present, we cannot exclude the possibility that additional sequences relevant for control of the expression levels are localized further upstream of position −400.

The study of the expression of individual members of this gene family in tomato has clearly indicated that all the *cab* genes are coordinately expressed. However, the exact mechanisms by which such coordinated expression is achieved are not yet understood. Our sequence comparisons indicate that neither overall sequence similarity nor presence or absence of specific nucleotide motifs is strongly correlated with the levels and pattern of gene expression.

Acknowledgements. This work was supported by a grant of the DFG to B.P. (Pi 153/2-4), a fellowship of the Graduiertenförderung of the University of Göttingen to J.-W.K., and a NATO grant for collaborative research to E.P. and B.P. (04377/88).

References

Alosi CM, Neale DB, Kinlaw CS (1990) Expression of *cab* genes in Douglas fir is not strongly regulated by light. Plant Physiol 93:829–832

Cashmore AR (1984) Structure and expression of a pea nuclear gene encoding a chlorophyll a/b-binding polypeptide. Proc Natl Acad Sci USA 81:2960–2964

Castresana C, Staneloni R, Malik VS, Cashmore AR (1987) Molecular characterization of two clusters of genes encoding the Type I CAB polypeptides of PS II in *Nicotiana plumbaginifolia.* Plant Mol Biol 10:117–126

Castresana C, Garcia-Luque I, Alonso E, Malik VS, Cashmore AR (1988) Both positive and negative regulatory elements mediate expression of a photoregulated *cab* gene from *Nicotiana plumbaginifolia.* EMBO J 7:1929–1936

Chitnis PR, Morishige DT, Nechusthtai R, Thornber JP (1988) Assembly of the barley light-harvesting chlorophyll a/b proteins in barley etiochloroplasts involves processing of the precursor on thylakoids. Plant Mol Biol 11:95–107

Demmin DS, Stockinger EJ, Chang YC, Walling LL (1989) Phylogenetic relationships between the chlorophyll a/b binding (*cab*) multigene family: an intra- and interspecies study. J Mol Evol 29:266–279

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dunsmuir P (1985) The petunia chlorophyll a/b binding protein genes: a comparison of *cab* genes from different gene families. Nucleic Acids Res 13:2503–2518

Gidoni D, Brosio P, Bond-Nutter D, Bedbrock J, Dunsmuir P (1989) Novel cis-acting elements in *Petunia cab* gene promoters. Mol Gen Genet 215:337–344

Green BR, Pichersky E, Kloppstech K (1991) Chlorophyll a/b binding proteins: an extended family. Trends Biol Sci 16:180–186

Grob U, Stüber K (1987) Discrimination of phytochrome-dependent, light-inducible from non-light-inducible plant genes. Prediction of a common light-responsive element (LRE) in phytochrome-dependent, light-inducible plant genes. Nucleic Acids Res 15:9957–9973

Ikeuchi M, Hirano A, Inoue Y (1991) Correspondence of apoproteins of light-harvesting chlorophyll a/b complexes associated with photosystem I to *cab* genes: Evidence for a novel Type IV apoprotein. Plant Cell Physiol 32:103–112

Jansson S, Gustafsson P (1990) Type I and Type II genes for the chlorophyll a/b-binding protein in the gymnosperm *Pinus sylvestris* (Scots pine): cDNA cloning and sequence analysis. Plant Mol Biol 14:287–296

Jansson S, Gustafsson P (1991) Evolutionary conservation of the chlorophyll a/b binding proteins: cDNAs encoding Type I, II and III LHC I polypeptides from the gymnosperm Scots pine. Mol Gen Genet, in press

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6653

Karlin-Neumann GA, Kohorn BD, Thornber JP, Tobin EM (1985) A chlorophyll a/b-protein encoded by a gene containing an intron with characteristics of a transposable element. J Mol Appl Genet 3:45–61

Karlin-Neumann GA, Sun L, Tobin EM (1988) Expression of light-harvesting chlorophyll a/b protein genes is phytochrome regulated in etiolated *Arabidopsis thaliana* seedlings. Plant Physiol 88:1323–1331

Kellmann JW, Pichersky E, Piechulla B (1990) Analysis of the diurnal expression patterns of the tomato chlorophyll a/b binding protein genes. Influence of light and characterization of the gene family. Photochem Photobiol 52:35–41

Kohorn BD, Harel E, Chitnis PR, Thornber JP, Tobin EM (1986) Functional and mutational analysis of the light-harvesting chlorophyll a/b protein of thylakoid membranes. J Cell Biol 102:972–981

Kuhlemeier C, Green PJ, Chua NH (1987) Regulation of gene expression in higher plants. Annu Rev Plant Physiol 38:221–257

Lam E, Chua NH (1989) 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

Lamppa G, Morelli G, Chua NH (1985) Structure and developmental regulation of a wheat gene encoding the major chlorophyll a/b-binding polypeptide. Mol Cell Biol 5:1370–1378

Leutwiler LS, Meyerowitz EM, Tobin EM (1986) Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*. Nucleic Acids Res 14:4051–4063

Long Z, Wang SY, Nelson N (1989) Cloning and nucleotide sequence analysis of genes coding for the major chlorophyll-binding protein of the moss *Physcomitrella patens* and the halotolerant alga *Dunaliella salina*. Gene 76:299–312

Luan S, Bogorad L (1989) Nucleotide sequences of two genes encoding the light harvesting chlorophyll a/b binding protein of rice. Nucleic Acids Res 17:2357–2358

Mitra A, Choi HK, An G (1989) Structural and functional analyses of *Arabidopsis thaliana* chlorophyll a/b-binding protein (cab) promoters. Plant Mol Biol 12:169–179

Morelli G, Nagy F, Fraley RT, Rogers SG, Chua NH (1985) A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose-1,5-bisphosphate carboxylase small subunit of pea. Nature 315:200–204

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Pichersky E, Bernatzky R, Tanksley SD, Breidenbach RB, Kausch AR, Cashmore AR (1985) Molecular characterization and genetic mapping of two clusters of genes encoding chlorophyll a/b-binding proteins in *Lycopersicon esculentum* (tomato). Gene 40:247–258

Pichersky E, Hoffman NE, Malik VS, Bernatzky R, Tanksley SD, Szabo L, Cashmore AR (1987a) The tomato *cab*-4 and *cab*-5 genes encode a second type of CAB polypeptide localized in photosystem II. Plant Mol Biol 9:109–120

Pichersky E, Hoffman NE, Bernatzky R, Piechulla B, Tanksley SD, Cashmore AR (1987b) Molecular characterization and genetic mapping of DNA sequences encoding the Type I chlorophyll a/b-binding polypeptides of photosystem I in *Lycopersicon esculentum* (tomato). Plant Mol Biol 9:205–216

Pichersky E, Tanksley SD, Piechulla B, Stayton MM, Dunsmuir P (1988) Nucleotide sequence and chromosomal location of *cab*-7, the tomato gene encoding the Type II chlorophyll a/b-binding polypeptide of photosystem I. Plant Mol Biol 11:69–71

Pichersky E, Brock TG, Nguyen D, Hoffman NE, Piechulla B, Tanksley SD, Green BR (1989) A new member of the CAB gene family: structure, expression and chromosomal location of *cab*-8, the tomato gene encoding the Type III chlorophyll a/b-binding polypeptide of photosystem I. Plant Mol Biol 12:257–270

Pichersky E, Green BR (1990) The extended family of chlorophyll a/b binding proteins of PS I and PS II. In: Baltscheffsky M (ed), Current research in photosynthesis, vol 3, Kluewer Academic Publishers, The Netherlands, pp 553–556

Pichersky E, Subramaniam R, White MJ, Reid J, Aebersold R, Green BR (1991) Chlorophyll a and b binding (CAB) polypeptides of CP 29, the internal chlorophyll a and b complex of PS II: characterization of the tomato gene encoding the 26 kDa (Type I) polypeptide, and evidence for a second CP 29 polypeptide. Mol Gen Genet 227:277–284

Piechulla B, Gruissem W (1987) Diurnal mRNA fluctuations of nuclear and plastid genes in developing tomato fruits. EMBO J 6:3593–3599

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York, pp 5.78–5.79

Schindler U, Cashmore AR (1990) Photoregulated gene expression may involve ubiquitous DNA-binding proteins. EMBO J 9:3415–3427

Schwartz E, Pichersky E (1990) Sequence of two tomato genes encoding chlorophyll a/b-binding proteins of CP 24, a PS II antenna component. Plant Mol Biol 15:157–160

Schwartz E, Shen D, Aebersold R, McGrath MJ, Pichersky E, Green BR (1991) Nucleotide sequence and chromosomal location of *cab* 11 and *cab* 12, the genes for the fourth polypeptide of the photosystem I light-harvesting antenna (LHC I). FEBS Lett 280:229–234

Simpson J, Timko MP, Cashmore AR, Schell J, Van Montagu M, Herrera-Estrella L (1985) Light-inducible and tissue-specific expression of a chimeric gene under control of the 5′flanking sequence of a pea chlorophyll a/b-binding protein gene. EMBO J 4:2723–2729

Stayton MM, Black M, Bedbrock J, Dunsmuir P (1986) A novel chlorophyll a/b binding (*cab*) protein gene from *Petunia* which encodes the lower molecular weight CAB precursor protein. Nucleic Acids Res 14:9781–9796

Sullivan TD, Christensen AH, Quail PH (1989) Isolation and characterization of a maize chlorophyll a/b binding protein gene that produces high levels of mRNA in the dark. Mol Gen Genet 215:431–440

Walling LL, Chang CY, Demmin DS, Holzer FM (1988) Isolation, characterization and evolutionary relatedness of three members from the soybean multigene family encoding chlorophyll a/b binding proteins. Nucleic Acids Res 16:10477–10493

Weisshaar B, Block A, Armstrong GA, Herrmann A, Schulze-Lefert P, Hahlbrock K (1991) Regulatory elements required for light-mediated expression of the *Petroselinum crispum* CHS gene. In: Jenkins and Schuch (eds) Molecular Biology of Plant Development, SEB Seminar Series, Great Britain, in press

Communicated by R. Devoret