The Plant Cell, Vol. 9, 491–507, April 1997 © 1997 American Society of Plant Physiologists

# A Myb-Related Transcription Factor Is Involved in the Phytochrome Regulation of an Arabidopsis *Lhcb* Gene

**Zhi-Yong Wang, David Kenigsbuch,[1] Lin Sun,[2] Eitan Harel,[3] May S. Ong, and Elaine M. Tobin[4]**

Department of Molecular, Cellular, and Developmental Biology, University of California, Los Angeles, California 90095-1606

**We have isolated the gene for a protein designated CCA1. This protein can bind to a region of the promoter of an Arabidopsis light-harvesting chlorophyll *a/b* protein gene, *Lhcb1*3*, which is necessary for its regulation by phytochrome. The CCA1 protein interacted with two imperfect repeats in the *Lhcb1*3* promoter, $AA^A/_CAATCT$, a sequence that is conserved in *Lhcb* genes. A region near the N terminus of CCA1, which has some homology to the repeated sequence found in the DNA binding domain of Myb proteins, is required for binding to the *Lhcb1*3* promoter. Lines of transgenic Arabidopsis plants expressing antisense RNA for CCA1 showed reduced phytochrome induction of the endogenous *Lhcb1*3* gene, whereas expression of another phytochrome-regulated gene, *rbcS-1A*, which encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, was not affected. Thus, the CCA1 protein acts as a specific activator of *Lhcb1*3* transcription in response to brief red illumination. The expression of CCA1 RNA was itself transiently increased when etiolated seedlings were transferred to light. We conclude that the CCA1 protein is a key element in the functioning of the phytochrome signal transduction pathway leading to increased transcription of this *Lhcb* gene in Arabidopsis.**

## INTRODUCTION

Plants use several different photoreceptors to respond to changes in environmental conditions. Among the best studied of these photoreceptors are the phytochromes, comprising a family of closely related chromoproteins that are involved in regulating a diverse array of developmental processes in higher plants, including seed germination, photomorphogenesis, and flowering (Kendrick and Kronenberg, 1994). Phytochromes can exist in either of two photochemically interconvertible forms: Pr (absorption maximum in the red region of the spectrum) and Pfr (absorption maximum in the far-red region of the spectrum). A classical diagnostic hallmark of a phytochrome response is that it can be elicited by brief red illumination (R), and the effect of R can be substantially reversed by far-red illumination (FR) given immediately after an R treatment. Although the primary mechanism of action of any phytochrome is not known, activation of phytochrome signal transduction by light has been shown to affect transcription of a number of different genes. Different phytochrome family members can have redundant functions, and at least three phytochrome types can regulate the expression of genes encoding light-harvesting chlorophyll *a/b* proteins associated with photosystem II (*Lhcb* genes) (Furuya and Schäfer, 1996).

A range of experimental approaches has been taken to understand the ways in which phytochrome acts and to identify components involved in transducing the original light signal. There is biochemical and pharmacological evidence that G proteins and a $Ca^{2+}$/calmodulin system are involved in phytochrome signal transduction mechanisms (reviewed in Deng, 1994; Barnes et al., 1995; Quail et al., 1995). Proteins affecting processes that can be regulated by phytochrome have been identified by the mutation of genes (designated *DEETIOLATED* [*DET*]/*CONSTITUTIVE PHOTOMORPHOGENIC* [*COP*]/*FUSCA* [*FUS*]) that have pleiotropic effects, including repression of photomorphogenesis. Several of these genes have now been cloned and characterized (reviewed in Chory et al., 1996; von Arnim and Deng, 1996). The proteins identified by this strategy include those that are involved in a diverse array of responses and not necessarily directly involved in mediating the action of phytochrome, but there is evidence that the *DET1* and *COP1* genes may play a specific role as repressors in photomorphogenesis (Chory et al., 1996; Wei and Deng, 1996).

Another approach to understanding the phytochrome signal transduction chain is to identify promoter sequences and the cognate factors mediating the regulation of transcription. There have been many studies of promoter regions responsible for differences in gene expression between light-grown and dark-treated plants, but fewer studies have focused specifically on phytochrome regulation (Tobin and Kehoe, 1994; Terzaghi and Cashmore, 1995). Studies that have defined

[1] Current address: Life Sciences Department, Bar Ilan University, Ramat Gan, Israel.
[2] Current address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.
[3] Current address: Department of Plant Sciences, Hebrew University, Jerusalem 91904, Israel.
[4] To whom correspondence should be addressed. E-mail etobin @ucla.edu; fax 310-206-4386.

promoter regions involved in conferring phytochrome responsiveness have been performed with *rbcS* (encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase), *Lhcb*, and *PHYA* (encoding an apoprotein of phytochrome) genes (Nagy et al., 1987; Gilmartin and Chua, 1990; Bruce et al., 1991; Sun et al., 1993; Anderson et al., 1994; Kehoe et al., 1994; Anderson and Kay, 1995; Kenigsbuch and Tobin, 1995; Degenhardt and Tobin, 1996). Although motifs that are necessary for phytochrome responsiveness have been defined for some of these promoters, no single consensus motif has been found for such regulation. In several instances, it has been shown that multiple sequence elements are needed for responsiveness (Kehoe et al., 1994; Degenhardt and Tobin, 1996; Puente et al., 1996), and recently, two different combinations of elements have been used to confer phytochrome regulation to a reporter gene (Puente et al., 1996). The hypothesis that multiple sequence elements are involved is also supported by extensive analysis of promoters of light-responsive genes (Argüello-Astorga and Herrrera-Estrella, 1996).

The *Lhcb1*3* gene of Arabidopsis (Jansson et al., 1992), originally designated *cab140* (for chlorophyll *a/b* binding protein gene 140; Leutwiler et al., 1986), has been found to be regulated by phytochrome action in etiolated seedlings, with brief R resulting in a large increase in the level of its mRNA (Karlin-Neumann et al., 1988). We have identified a DNA binding activity, designated CA-1, that interacts with the promoter of this gene (Sun et al., 1993). The region to which CA-1 binds in the Arabidopsis *Lhcb1*3* promoter is necessary for both high-level expression and normal phytochrome regulation of this gene (Kenigsbuch and Tobin, 1995). It is located immediately downstream of an ACGT core sequence, which occurs in many different plant promoters and binds transcription factors of the basic leucine zipper type (Menkens et al., 1995); however, the ACGT sequence is not itself critical for binding of CA-1 (Sun et al., 1993).

DNA binding proteins that can interact with phytochrome-regulated promoters have been found in a number of different species (reviewed in Tobin and Kehoe, 1994; Teakle and Kay, 1995; Terzaghi and Cashmore, 1995). Several of these interact with sequences shown to be important for phytochrome responsiveness. However, none of them thus far has been shown to have a functional role in vivo specific for phytochrome action. In this study, we report the identification and characterization of a gene for such a transcription factor, which interacts with the promoter of the *Lhcb1*3* gene of Arabidopsis.

## RESULTS

### Isolation and Sequence Characterization of CCA1 cDNA and Genomic Clones

An Arabidopsis λ cDNA expression library was screened with the radiolabeled A2 fragment of the *Lhcb1*3* promoter. This fragment was used previously to identify CA-1 activity in plant extracts. To isolate clones encoding proteins that interact specifically with the CA-1 binding site, positive plaques were also screened with a mutant probe (m1) for which the CA-1 binding activity in plant extracts has little affinity (Sun et al., 1993). Two cDNA clones (21 and 24) were found to produce proteins that showed specific binding to the wild-type but not to the mutant probe. Sequence analysis showed that the two clones overlap by 470 nucleotides and are partial cDNAs derived from the same RNA. Clone 24 included a polyadenylated tail of 15 bases and therefore encompassed the entire 3′ region of the mRNA. The 5′ end of the mRNA was determined by primer extension analysis (Figure 1A), and this result demonstrated that clone 21 included the complete 5′ region of the transcript. A full-length cDNA clone, designated clone 25, was constructed by joining the 5′ and 3′ fragments of clones 21 and 24, respectively, at the unique PstI site in the overlapping region.

Screening of an Arabidopsis genomic library with a cDNA fragment allowed the isolation of the corresponding gene, which we designated *CCA1*. The sequence of this gene is shown in Figure 1B, along with the predicted amino acid sequence for the CCA1 protein. The gene sequence includes seven introns, 237 nucleotides of 5′ untranslated sequence, and 193 nucleotides of 3′ untranslated sequence. The 1824-nucleotide open reading frame (ORF) encodes a protein of 608 amino acids with a calculated molecular weight of 66,970 and an isoelectric point of 5.6. An ORF of 24 nucleotides is present in the 5′ untranslated region of the transcript and is in phase with the main ORF. In several cases, such ORFs have been shown to be involved in translational regulation of gene expression (Lohmer et al., 1993; Hinnebusch, 1994) and have also been found in other plant transcription factor genes (Singh et al., 1990; Ruberti et al., 1991; Carabelli et al., 1993; Lohmer et al., 1993).

### *CCA1* Is a Single-Copy Gene Encoding a Novel Myb-Related Protein

The predicted amino acid sequence of the CCA1 protein has a basic region at the N terminus (K-13 to K-107). Within this region is a sequence similar to the repeat sequence highly conserved in Myb-related proteins. Figure 1C shows the predicted amino acid sequence of CCA1 from amino acid residue 24 to 75 compared with the repeat sequences of various Myb proteins from animals, plants, and yeast. Within this sequence, there is a limited amino acid identity (16 of 52; 31%) and substantial similarity (29 of 52; 56%) when compared with the third repeat of human c-Myb. The sequence identity includes two of the three conserved tryptophans present in most Myb proteins. The conserved residues also include seven of the 11 residues that are known to be important for forming the hydrophobic core and maintaining the three-dimensional structure of the Myb repeat, which forms a helix-turn-helix structure (Ogata et al., 1992). However, the amino acid residues of human Myb that

contact the bases of DNA are not conserved in CCA1 (N-183 in hMyb; S in CCA1; K-182 versus R; N-186 versus Q; N-179 versus V) (Ogata et al., 1994). In contrast to most other Myb proteins that have been characterized, this region is not repeated in the CCA1 protein. No other significant homology to any protein in the database was found.

We tested whether *CCA1* is a member of a gene family in Arabidopsis by using genomic DNA gel blot analysis. The results of low-stringency hybridization of Arabidopsis DNA with a *CCA1* probe are shown in Figure 1D. There was a single band of hybridization in the lanes that were digested with EcoRI, SstI, and PstI, which have no cleavage site in the probe fragment. There were two bands in the lane that was digested with HindIII, which has a cleavage site in the probe 188 bp from one end. An identical pattern was seen when the blot was hybridized with the same probe under the high-stringency conditions (see Methods; data not shown). The DNA gel blots were also hybridized under both low- and high-stringency conditions with a probe consisting of nucleotides 267 to 949 of the CCA1 cDNA, which includes the region of similarity to the Myb repeat. This probe gave no evidence for any additional closely related sequences (data not shown). We conclude that although the *CCA1* gene includes a small region with amino acid sequence homology to the Myb repeat, there are no genes that are closely related to *CCA1* in the Arabidopsis genome.

### The Myb-Related Region of the CCA1 Protein Is Required for DNA Binding

The fact that the two cDNA clones isolated in the initial filter-binding screening each contained the sequence similar to the Myb repeat suggested that this region is necessary for DNA binding. We tested this possibility and further characterized the CCA1 protein by expressing the polypeptides encoded by various fragments of the CCA1 cDNA. The CCA1 cDNA clones were fused to the coding sequence for glutathione *S*-transferase (GST) and used to produce the polypeptides in *Escherichia coli.* These constructs are diagrammed in Figure 2A. The polypeptides produced were tested for their ability to bind to the A2 fragment of the *Lhcb1*3* promoter by electrophoretic gel mobility shift assays (EMSAs). The polypeptides corresponding to the pXCA-23 and pXCA-23 constructs were produced as isopropyl β-D-thiogalactopyranoside (IPTG)–inducible GST fusion proteins and were also tested as purified proteins after cleavage from GST. Those corresponding to the cDNA clones CCA1-21 and CCA1-25 (pXCA-21 and pXCA-25) contained stop codons in the 5′ untranslated region of the cDNA and thus were not produced as fusion proteins. Figure 2B shows an EMSA using either *E. coli* extracts (lanes 1 to 10, 15, and 16) or purified proteins before (lanes 11 and 13) and after (lanes 12 and 14) cleavage from GST. DNA binding activities induced by IPTG were observed for proteins produced from constructs pXCA-21, pXCA-24, and pXCA-25, but binding activity could not be detected for the protein produced from construct pXCA-23, which lacked the N-terminal 82 amino acids. In conjunction with the finding that the N-terminal 11 amino acids are not necessary for binding (construct pXCA-24), these experiments demonstrate that the sequence containing amino acid residues 11 to 82 of CCA1, which includes the region with similarity to the Myb DNA binding domain, is essential for the DNA binding activity of the CCA1 protein.

### Binding Characteristics of CCA1 and CA-1

To compare the binding characteristics of the CCA1 protein and CA-1 activity from the plants, we performed footprint analyses and binding competition experiments using the A2 fragment of the *Lhcb1*3* promoter as a probe. The results of 1,10-phenanthroline–copper footprinting are shown in Figure 3A. At left is the EMSA that was performed to resolve the free probe and the DNA–protein complexes. DNA was recovered from each band after treatment of the gel with phenanthroline–copper and resolved on the sequencing gel shown at right. With increasing amounts of the CCA1 protein purified from *E. coli*, two complexes (B1 and B2) of different mobilities could be observed. The nucleotides protected from cleavage in each of the complexes can be seen on the sequencing gel on the right in Figure 3A. In complex B1, the −92 to −105 region was protected, and in complex B2, regions from −92 to −105 and from −111 to −122 were protected. This result suggests that the two complexes of different mobilities are a result of the presence of two separate binding sites on this fragment and that the −92 to −105 region is the higher affinity binding site for CCA1. A nearly perfect repeated sequence of $AAA^{A}/_{C}$ AATCTA occurs in each of these footprinted regions.

The CA-1 protein–DNA complex obtained with the plant cell extract (Figure 3A, lane 4) showed protected nucleotides in the region from −94 to −105, and a second experiment confirmed these boundaries, demonstrating that CA-1 (from plant extracts) and CCA1 (from the clone expressed in *E. coli*) can bind to the same region of the *Lhcb1*3* promoter.

Figures 3B and 3C show the results of methylation interference and depurination interference experiments performed with the CCA1 protein. Figure 3B shows the interfering nucleotides on sequencing gels, and their positions on the A2 fragment of the promoter are shown in Figure 3C, along with the results of footprinting experiments. Interference with the protein–DNA binding by the modification of a base residue is manifested by increased intensity in the lane with the free DNA fraction and reduced intensity in the lane of protein-bound DNA compared with the lane of control DNA that was not incubated with protein. Both methylation and depurination interference assays identified the same nucleotides and showed that nucleotides within both nearly perfect repeats ($AA^{A}/_{C}$AATCT) interact directly with the CCA1 protein.

Figure 3C also summarizes the results of the phenanthroline–copper footprinting and shows the extent of the DNase I


gcagtggttcacttacaagaacctggtcttcaaaccagacaggttaacccaattctctctttaactctgtgtttggttgcatgtaatactgagaatggaagactcaaatttctcgaggaaat -891
tgtttgttatctgtttcagggaggctttgtttgagaaggtcaagagcacatacaaagacatattagggagcagctgaatcaaaggagaagaagaagaagagcgccttttgaggccat -771
tcatgaattggaatgaaggatatcaaaagaatctaacacaaaggccacgtccttccttcaatctttccttcttgtaactaaataattttcatcctttctctctctgtctctggtcttt -651
tttagctcaaagtatcatccatttatgtcaaagtgttgtaaatttcctcaagactatatatgagatgttttgtttcatttttccaaaatttcaaacttttgtccccatttagtcttctaccct -531
tcatgcatggttagcttagcttaatgctgaactgttgaataacgatatgggccttatgctaaaagaacaaaacctttatgggtctaaaaaaaaataagcccaatatataaactatggcccaaa -411
taagtttaggtccattagagtgtgagaatagcgcgtgtagtgaacccgcacgagaatgcgcgttcgattgttggtgaagtagtcgtctagattcccgggtccactgatgtttctagtgtat -291
cagacacgtgtcgacaaactggtgggagagattaacgatcttaagtaggtcccactagatgaattgacctttttaacctttcaggtagtcccggaactcgtggcct -171
agaatacaaagaaggttgtgaacaagttgatgttaagatggacaagaatgtaacttgaacaaaagctgaatcatctcttcagccactagtatgttgacatatggcagtttcttttgtagc -51
ctcgaaataaataaatttaaaaagtttgaggttaaagataatttatagtggcTGAGATTTCTCCATTTCCGTAGCTTCTGGTCTCTTTTCTTTGTTTCATTGATCAAAAGCAAATCACTTCT 70
TCTTCTTCTTCTCGATTTCTTACTGTTTTCTTATCCAACGAAATCTGGAATTAAAAATGGAATCTTTATCGAATCCAAGCTGATTTTGTTTCTTTCATTGAATCATCTCTCTAAAGg 190
tacttaagattgatttattgtcatggtctttcttattgtttgatgaataacttgacttgattgttttttgttttgtggattagTGGAATTTTGTAAAGAGAAGATCTGAAGTTGTGTAGA 310
GGAGCTTAGTGATGGAGACAAATTCGTCTGGAGAAGATCTGGTTATTAAGgtaaattaactaaatttttaggggggaagatgattgttttaggtgtcaaagattgagaatttttaatgaaact 430
1 M E T N S S G E D L V I K
tgatatagACTCGGAAGCCATATACGATAACAAAGCAACGTGAAAGGTGGACTGAGGAAGAACATAATAGATTCATTGAAGCTTTGAGGCTTTATGGTAGAGCATGGCAGAAGATTGAAG 550
14 T R K P Y T I T K Q R E R W T E E E H N R F I E A L R L Y G R A W Q K I E
gttgatttttatttccctttatatgtcttatttttgtgtttgcagaggtttgtcttcaaactgatttgctttttttcatttggacagAACATGTAGCAACAAAAACTGCTGTCCAGATA 670
51 E H V A T K T A V Q I
AGAAGTCACGCTCAGAAATTTTTCTCCAAGgtaaaatcggttaattttgaaatgatgttctcatcttcattggcttaatgcttaagacttattgaaagccaggcaagttttctgcttctt 790
62 R S H A Q K F F S K
ttgcttcttagtcaggagatagatagattacgtttttagagtttagtaatgagcaataagtcttaaaatagttggagaaatgacgagatgtaatcgttttcttttgtttatgcctatatc 910
ttgttaatccacaaacatgtacatagattcttcagaagaatgttagtttctttagattcttcagataaacttgtgtcttcttaccgattctgaggtagtggcaaaagtgggctgagtgct 1030
agaaatttttgaatgttccttgtgatgaagccatagaggtaaaccatttttgattttccagttctgtcatttaaacttgttaggtgtcattagatttttgtttgtttacgtttgtttagag 1150
ggtaacaaaactactctcatctctctcagGTAGAGAAAGAGGCTGAAGCTAAAGGTGTAGCTATGGGTCAAGCGCTAGACATAGCTATTCCTCCTCCACGGCCTAAGCGTAAACCAAACA 1270
72 V E K E A E A K G V A M G Q A L D I A I P P P R P K R K P N
ATCCTTATCCTCGAAAGACGGGAAGTGGAACGATCCTTATGTCAAAAACGGGTGTGAATGATGGAAAAGAGTCCCTTGGATCAGAAAAAGTGTCGCATCCTGAGgtgattttcatggtca 1390
102 N P Y P R K T G S G T I L M S K T G V N D G K E S L G S E K V S H P E
tatggcatcttttgcagtgtgtcacattgctcctcatgttattaatacagattgtgtgcttcgtttatagATGGCCAATGAAGATCGACAACAATCAAAGCCTGAAGAGAAAACTCTGC 1510
137 M A N E D R Q Q S K P E E K T L
AGGAAGACAACTGTTCAGATTGTTTCACTCATCAGTATCTCTCTGCTGCATCCTCCATGAATAAAAGTTGTATAGAGACATCAAACGCAAGCACTTTCCGCGAGTTCTTGCCTTCACGGG 1630
153 Q E D N C S D C F T H Q Y L S A A S S M N K S C I E T S N A S T F R E F L P S R
AAGAGgtaaaaaacaatctttcattgctatttgaggttttaagacgattagtacttttcatgaaactaaaaccgtgggggaataacagGGAAGTCAGAATAACAGGGTAAGAAAGGAGTC 1750
193 E E G S Q N N R V R K E S
AAACTCAGATTTGAATGCAAAATCTCTGGAAAACGGTAATGAGCAAGGACCTCAGACTTATCCGATGCATATCCCTGTGCTAGTGCCATTGGGGAGCTCAATAACAAGTTCTCTATCACA 1870
206 N S D L N A K S L E N G N E Q G P Q T Y P M H I P V L V P L G S S I T S S L S H
TCCTCCTTCAGAGCCAGATAGTCATCCCCACACAGTTGCAGGAGATTATCAGTCGTTTCCTAATCATATAATGTCAACCCTTTTACAAACACCGGCTCTTTATACTGCCGCAACTTTCGC 1990
246 P P S E P D S H P H T V A G D Y Q S F P N H I M S T L L Q T P A L Y T A A T F A
CTCATCATTTTGGCCTCCCGATTCTAGTGGTGGCTCACCTGTTCCAGGGAACTCACCTCCGAATCTGGCTGCCATGGCCGCAGCCACTGTTGCAGCTGCTAGTGCTTGGTGGGCTGCCAA 2110
286 S S F W P P D S S G G S P V P G N S P P N L A A M A A A T V A A A S A W W A A N
TGGATTATTACCTTTATGTGCTCCTCTTAGTTCAGGTGGTTTCACTAGTCATCCTCCATCTACTTTTGGACCATCATGTGATGTAGAGTACACAAAAGCAAGCACTTTACAACATGGTTC 2230
326 G L L P L C A P L S S G G F T S H P P S T F G P S C D V E Y T K A S T L Q H G S
TGTGCAGAGCCGAGAGCAAGAACACTCCGAGGCATCAAAGGCTCAGAGGATGTTGAAAATAAGAGTAAACCAGTTTGTCATGAGCAGCCTTCTGCAACACCTGA 2350
366 V Q S R E Q E H S E A S K A R S S L D S E D V E N K S K P V C H E Q P S A T P E
GAGTGATGCAAAGGGTTCAGATGGAGCAGGAGACAGAAAACAAGTTGACCGGTCCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCAGAAAGGCAAGA 2470
406 S D A K G S D G A G D R K Q V D R S S C G S N T P S S S D D V E A D A S E R Q E
GGATGGCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAACCTCAAACTTCAGAGTCCAATGCACGCCGCAGTAGAATCAGCTCCAATATAACCGATCCATGGAAGTCTGT 2590
446 D G T N G E V K E T N E D T N K P Q T S E S N A R R S R I S S N I T D P W K S V
GTCTGACGAGgtacttacttggactaaagatcaacttccttattccaaatcatttttctcatataaaatattgtacattcggGGTCGAATTGCCTTCCAAGCTCTCTTCTCCAGGTATGAGGTAT 2710
486 S D E G R I A F Q A L F S R E V
TGCCGCAAAGTTTTACATATCGAGAAGAACACAGAGAGGAAGAACAACAACAACAAGAACAAAGATATCCAATGGCACTTGATCTTAACTTCACAGCTCAGTTAACACCAGTTGATGATC 2830
502 L P Q S F T Y R E E H R E E E Q Q Q Q E Q R Y P M A L D L N F T A Q L T P V D D
AAGAGGAGAAGAGAAACACAGGATTTCTTGGAATCGGATTAGATGCTTCAAAGCTAATGAGTAGAGGAAGAACAGGTTTTAAACCATACAAAAGATGTTCCATGGAAGCCAAAGAAAGTA 2950
542 Q E E K R N T G F L G I G L D A S K L M S R G R T G F K P Y K R C S M E A K E S
GAATCCTCAACAACAATCCTATCATTCATGTGGAACAGAAAGATCCCAAACGGATGCGGTTGGAAACTCAAGCTTCCACATGAGACTCTATTTTCATCTGATCTGTTGTTTGTACTCTGT 3070
582 R I L N N N P I I H V E Q K D P R M R L E T Q A S T *
TTTTAAGTTTTCAAGACCACTGCTACATTTTCTTTTCTTTTGAGGCCTTTGTATTTGTTTCCTTGTCCATAGTCTTCCTGTAACATTTGACTCTGTATTATTCAACAAATCATAAACTG 3190
TTTAATCTTTTTTTTTTCCAAcctggaaagaacttcactcaaggggctcttgttcttgatatatgcaaacgacagagttccaaaacgtaatcttagcccatccatcaccctaagttgtct 3310
cataactcataagtaagcacaaaa 3334

**Figure 1.** Sequence and Genomic Organization of the *CCA1* Gene.

footprint previously determined for CA-1 activity (Sun et al., 1993). We used unlabeled competitor DNAs in the EMSA to compare binding specificities of the CCA1 protein produced in *E. coli* and the CA-1 activity from the plant extracts. The wild-type and mutant promoter fragments used as competitors are shown at the bottom of Figure 3C. A representative result of such experiments is shown in Figure 3D. The binding of the *E. coli*–produced CCA1 protein to the probe was efficiently competed for either by a fragment of the A2 probe that contained the repeated sequence or by a promoter fragment (WT2) of another closely related *Lhcb* gene (*Lhcb1*1*, originally called AB165; Leutwiler et al., 1986) that contains one copy of this sequence (AAAAATCT). The m3 fragment, which had altered nucleotides in the downstream repeat region, was a less effective competitor than was the wild type; m1, m2, and m4 fragments, which had alterations in both repeats, showed the least competition.

When plant extracts were used, all of the fragments showed some degree of competition, which is likely in part to be the result of low amounts of the CA-1 protein in plant extracts. The results are not directly comparable to those with the purified CCA1 protein because the absolute amounts of the specific binding proteins are not known. Nonetheless, it can be seen that the m2 fragment served as a better competitor for CA-1 than did the m1 fragment, whereas the opposite was found with CCA1. Even more striking are the contrasting results with the m4 competitor DNA. This fragment, in which the C residues of both TCT motifs in the two repeats were changed, was even more effective than was the wild type in competing for the CA-1 activity, whereas it was not a particularly good competitor for CCA1. Thus, although both activities interact with the AAAAATCT sequence, there are differences in the importance of individual nucleotides in this sequence for the binding of CA-1 and CCA1.

### The Product of the CCA1 cDNA Is Nuclear Localized

We used a transient expression assay in onion epidermal cells to test whether the product of the *CCA1* gene was localized to nuclei, as might be expected for a transcription factor. The *uidA* gene, which encodes β-glucuronidase (GUS), was fused in frame to the coding sequence of *CCA1* so that GUS activity could be used to localize the compartmentation of the CCA1 protein. This transient assay should result in the expression of GUS activity in individual cells into which the DNA is effectively introduced. Figures 4A and 4B show that when a cauliflower mosaic virus 35S::*uidA* construct (pMF::GUS) was used in this assay, GUS activity was detected throughout the cytoplasm. When a construct (pMF::B::GUS) with the *opaque2* gene, which encodes a well-characterized transcription factor from maize, was fused to the *uidA* gene and used, GUS activity was detected specifically in nuclei (Figures 4C and 4D). Similarly, specific nuclear localization was found for the CCA1–GUS (p35S–CCA1–GUS) fusion protein, as shown in Figures 4E to 4H. These results show that the CCA1 protein is targeted to nuclei, and they are consistent with the proposed function of the CCA1 protein as a transcription factor.

In view of results showing that light and dark treatments can affect the nuclear localization of another protein involved

**Figure 1.** (continued).

**(A)** Transcription start site of the *CCA1* gene. The products of the primer extension reaction (lane P) were resolved on a 9% sequencing gel, along with the sequencing reactions (lanes T, G, C, and A). The arrow indicates the largest extended product detected. The sequence around the junction of the cDNA and the SalI adapter is shown.

**(B)** Nucleotide and deduced amino acid sequences of *CCA1*. The nucleotide sequence of the cDNA is in uppercase letters, and those of the promoter, introns, and 3′ flanking sequences are in lowercase letters. The deduced amino acid sequence is shown in single-letter code below the corresponding codons, and the Myb-related region is in boldface letters. The overlapping region of cDNA clones 21 and 24 is underlined. A small ORF in the 5′ untranslated region is shown in boldface italics. The numbers at left indicate the positions of the first amino acid on each line, and those at right show the positions of the last nucleotides relative to the transcription start (boldface T). The GenBank accession numbers for the cDNA and genomic sequences are U28422 and U79156, respectively.

**(C)** Sequence similarity between CCA1 and Myb proteins. The deduced amino acid sequence of a region of CCA1 (amino acids 24 to 75) is compared with one of the repeat sequences of the DNA binding domains of the Myb proteins from potato (St1, potato MybSt1; Baranowskij et al., 1994), human (HMyb, human c-Myb; Majello et al., 1986), chicken (CMyb, chicken c-Myb; Gerondakis and Bishop, 1986), Drosophila (DMyb, Drosophila c-Myb; Peters et al., 1987), maize (ZmC1, C1 protein of *Zea mays*; Paz-Ares et al., 1987), yeast (YBAS1, yeast BAS1 protein; Tice-Baldwin et al., 1989), and Arabidopsis (AtGl1, Arabidopsis Gl1 protein; Oppenheimer et al., 1991). The amino acid residues identical to CCA1 are boxed in black; those similar to CCA1 are boxed in gray. Asterisks show the positions of the conserved tryptophan residues. Dashes indicate gaps introduced to maximize alignment. The number at the right of each sequence indicates the position of the last amino acid in the corresponding protein.

**(D)** *CCA1* is a single-copy gene in the Arabidopsis genome. Shown is a genomic DNA gel blot analysis of the *CCA1* gene in Arabidopsis under low-stringency conditions with a 1.3-kb fragment of the 3′ region of *CCA1* cDNA (nucleotides 950 to 2254 of the cDNA). Each lane contains 5 μg of Arabidopsis genomic DNA digested with the restriction enzyme indicated. The positions and sizes of molecular markers are shown at left.

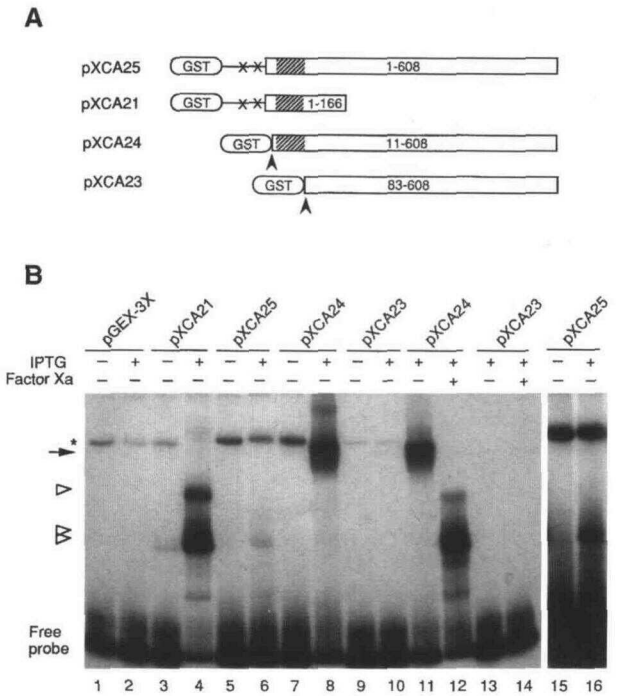


**Figure 2.** The Myb-Related Region of the CCA1 Protein Is Necessary for Its DNA Binding Activity.

**(A)** Diagrams of the constructs used for expression of CCA1 polypeptides in *E. coli.* Ovals represent the *GST* coding sequence, and bars represent the *CCA1* coding sequence; numbers indicate the amino acid residues encoded, and the Myb-related region is indicated by hatched boxes. Crosses (X) mark the positions of stop codons in the 5′ untranslated region of *CCA1* cDNA (solid line). Arrowheads show the positions of factor Xa cleavage.

**(B)** EMSA for DNA binding activities of polypeptides produced from the cDNA constructs shown above each pair of lanes. Each reaction included 0.3 ng of $^{32}$P-labeled A2 DNA probe and 1 μg of poly(dI-dC). Lanes 1 to 10, 15, and 16 each contain 1 μg of protein from extracts of *E. coli* induced (+) or not induced (−) with IPTG. Lanes 11 and 13 each contain 50 ng of purified GST–CCA1 fusion proteins (factor Xa [−]), and lanes 12 and 14 each contain 50 ng of purified CCA1 polypeptides released from the GST–CCA1 fusion proteins by factor Xa cleavage (Factor Xa [+]). The asterisk marks a nonspecific DNA binding activity. The arrow and the triangles denote the positions of the major protein–DNA complexes formed by the GST–CCA1 fusion protein and the nonfusion CCA1 polypeptides, respectively. Lanes 15 and 16 are a longer exposure of lanes 5 and 6.

in photomorphogenesis (COP1; von Arnim and Deng, 1994) and of a G-box (ACGT core motif) binding factor (Harter et al., 1994), we tested whether the partitioning of CCA1 was altered if the transient assay was performed in the dark. Unlike COP1, which was also tested in the onion transient assay system and was nuclear localized only in the dark, no difference in the nuclear localization of CCA1 was seen between the light and dark samples (data not shown).

### Expression of *CCA1* in an Antisense Orientation Affects the Phytochrome Response of the Endogenous *Lhcb1*3* Gene

We have previously shown that the promoter region to which the CA-1 activity binds is essential for phytochrome regulation of the *Lhcb1*3* gene (Kenigsbuch and Tobin, 1995). Therefore, if the product of the *CCA1* gene interacts with this promoter in vivo, it might be expected to affect the phytochrome induction of *Lhcb1*3* expression. We addressed this possibility by transforming Arabidopsis with portions of the *CCA1* gene in an antisense orientation driven by the constitutive cauliflower mosaic virus 35S promoter (Figure 5A). For each independent transgenic line, $T_2$ seeds homozygous for the T-DNA insertion were selected by analysis of the segregation of kanamycin resistance, and seedlings from these homozygous seeds were tested for the phytochrome responsiveness of both the endogenous *Lhcb1*3* gene and another phytochrome-regulated gene, the *rbcS-1A* gene. Figure 5B shows that in five of these seven lines, the level of *Lhcb1*3* mRNA after the red treatment was substantially lower than that of the wild type. However, no substantial effect of the antisense construct was seen for *rbcS-1A* gene expression in the same lines. The mRNA levels for *Lhcb1*3* and *rbcS-1A* were normalized to a ubiquitin RNA (*ubq3*; Brusslan and Tobin, 1992), and the relative expression levels of these two genes for all the lines and treatments are shown below the autoradiogram. The fold-increase of *Lhcb1*3* RNA in response to R was reduced in lines 4, 14, 17, 21, and 34, ranging from 37 to 53% of that of the wild type, but the induction of the *rbcS-1A* RNA was not comparably affected. Also, the antisense construct did not affect the mRNA levels in plants that had been given no light treatment or had been given FR after the R treatment. No visible phenotype was apparent in the antisense lines.

We also used the $T_3$ generation of four of the lines to test whether the reduction in *Lhcb1*3* RNA correlated with a reduction in levels of *CCA1* RNA. In this generation, there was a smaller effect of the antisense construct. It is not unusual for antisense effects to be lost or diminished in subsequent generations. For example, Chamnongpol et al. (1996) found that seven of eight lines expressing an antisense construct for a catalase gene lost the catalase suppression phenotype in their progeny. The R induction of *Lhcb1*3* RNA ranged from 68 to 86% of the wild type in the $T_3$ seedlings of these four lines, and the levels of *CCA1* RNA were 68 to 75% that of the wild-type plants. Although the effect of the antisense constructs was substantially reduced in this generation, the reduction of *CCA1* RNA was accompanied by a similar decrease in the induction of *Lhcb1*3* RNA by R. Our results demonstrate that the CCA1 protein can have a specific effect on the phytochrome induction of expression of the endogenous *Lhcb1*3* gene in vivo, and they strongly support the hypothesis that this protein is a part of the normal transduction pathway for the phytochrome response of this gene.

### Expression of the *CCA1* Gene Is Transiently Induced by Light Treatment of Etiolated Seedlings

We tested whether the expression of the *CCA1* gene might be regulated by light by examining the level of its RNA after transferring etiolated seedlings to continuous white light. The results of this experiment are shown in Figure 6. Two bands corresponding to *CCA1* RNA are detected at 2.9 and 2.4 kb on RNA gel blots. The lower band corresponds to the size of the full-length cDNA; the upper band may represent an incompletely spliced form of the RNA or an alternate transcript from this gene. The level of *CCA1* mRNA shows a large increase within the first hour of the light treatment and then declines over the next half day. This increase peaks earlier than does the increase in *Lhcb* RNA and is consistent with the hypothesis that CCA1 normally is involved in the induction of *Lhcb* gene expression.

### The AATCT Motif Is Highly Conserved in Promoters of *Lhc* Genes

Genomic sequences of 61 *Lhc* (apoproteins of light-harvesting complexes) genes were examined, and 55 of these were found to contain an AATCT motif in the region from −350 to −1 relative to translation start or −250 to −1 relative to transcription start. Of these, 35 occur just upstream of a CCAAT box. In some cases, the sequence is found on the complementary strand. A summary of these data, showing nucleotide identities (compared with the *Lhcb1*3* promoter), is shown in Table 1. Additional adenines are frequently found upstream of the AATCT motif. Analysis of another family of phytochrome-regulated genes, the *rbcS* genes, did not reveal any striking conservation of the AATCT sequence.

## DISCUSSION

The isolation of the Arabidopsis *CCA1* gene has allowed us to further our understanding of an essential step in the phytochrome regulation of *Lhcb* gene expression. It is of particular importance that phytochrome regulation of the endogenous *Lhcb1*3* gene was reduced in lines transformed with a *CCA1* antisense construct because it demonstrates that the CCA1 protein not only interacts with the promoter of the *Lhcb1*3* gene in vitro but that it also functions in the phytochrome regulation of the expression of this gene in vivo. The fact that the binding site is conserved in *Lhcb* genes supports the possibility that proteins related to CCA1 play a role in other species as well.

The predicted amino acid sequence of CCA1 includes a region near its N terminus with some sequence identity to a repeated region of Myb proteins. The Myb repeats are known to be involved in DNA binding (Ogata et al., 1994). By synthesizing truncated versions of the CCA1 protein in *E. coli* and testing them for the ability to bind to the *Lhcb1*3* promoter, we have been able to demonstrate that the Myb-related region of CCA1 is also required for binding to this promoter, although this region is not repeated in CCA1.

A large number of Myb proteins have been found in different plant species, and they comprise a gene family. For example, in Arabidopsis, at least seven *myb* homologs have been characterized to date (Oppenheimer et al., 1991; Shinozaki et al., 1992; Urao et al., 1993; Li and Parish, 1995; Li et al., 1996; Quaedvlieg et al., 1996); sequences of 16 additional members of this family from Arabidopsis are provided in the GenBank database, and an additional 22 *myb*-related sequences are present in the Arabidopsis expressed sequence tag database. However, the CCA1 protein cannot unequivocally be considered a member of the Arabidopsis *myb* gene family for several reasons. First, it has only a limited degree of amino acid sequence identity to any other Myb protein, and although this region of similarity includes two of the three conserved tryptophan residues of the Myb repeat, the region is not repeated in the CCA1 protein. One other Myb-like protein, MybSt1, which has a single region with some identity to the Myb repeat, has been reported in potato (Baranowskij et al., 1994); however, the CCA1 protein has no greater similarity to this protein than to other Myb proteins, and no homology could be detected between other regions of CCA1 and MybSt1. Second, *CCA1* probes can only detect a single gene on a DNA gel blot even at low stringency and using a probe including the *myb*-related sequence. In addition, the DNA region to which it binds, and which is conserved in a number of plant *Lhcb* genes, does not include the plant consensus sequence $CC^{T}/_{A}ACC$ for recognition by the maize Myb P protein (Grotewold et al., 1994), the recognition sequence for other plant Myb proteins (e.g., see Solano et al., 1995), or consensus sequences determined for Myb proteins from other kingdoms (TAACTG) (Tanikawa et al., 1993). We conclude, therefore, that *CCA1* is a novel *myb*-related gene rather than a member of the *myb* gene family in Arabidopsis.

In addition to the Myb repeat region of CCA1, there are other regions of potential interest. A basic region was found at the C terminus (K-559 to T-608, with 13 basic residues). There is an acidic region in the C-terminal half (D-352 to E-544; 42 acidic residues of 194) that is also serine and threonine rich (40 of 194). Within this region, the acidic residues are clustered in three stretches of amino acids (E-370 to E-389, seven of 20; D-434 to D-458, 11 of 25; and E-510 to E-544, 11 of 35) that are predicted to form α-helices (Chou and Fasman, 1978). This region might function as an acidic transcription activation domain for the CCA1 protein. These three acidic stretches also contain serine and threonine residues that match potential phosphorylation sites for casein kinase II and protein kinase C. Phosphorylation of these residues would increase the negative charge density in these regions. There are also a number of additional possible phosphorylation sites throughout the protein, which is consistent with our previous finding that the CA-1 activity can

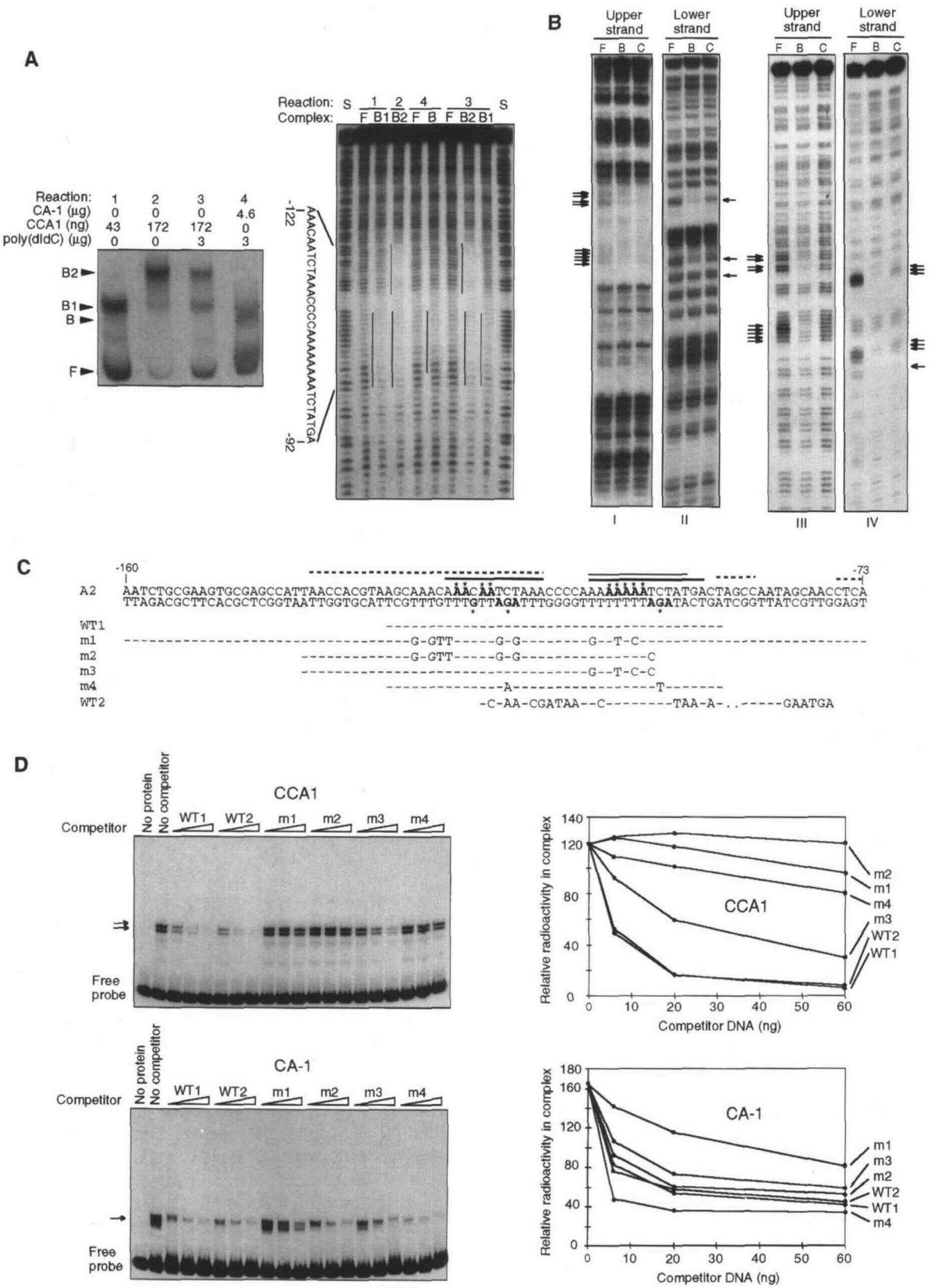


**Figure 3.** DNA Binding Specificity of the CCA1 Protein.

be phosphorylated (Sun et al., 1993). The *CCA1* clone has been mapped to a position near the bottom of chromosome 2 next to marker Athb-7 at 94.8 centimorgans on a yeast artificial chromosome-anchored physical map (Zachgo et al., 1996).

We have also demonstrated that the CCA1 protein is transported to nuclei in vivo, and this is consistent with its proposed function as a transcription factor. The N-terminal basic region contains a proline/basic rich sequence (P-92 to K-107) that has two clusters of basic amino acids. These features are characteristic of a bipartite nuclear localization signal (Raikhel, 1992).

The CCA1 protein interacted with two closely spaced binding sites with nearly perfect 10-bp repeated sequences ($AAA^A/_CAATCTA$) in the *Lhcb1*3* promoter. Our previous results (Sun et al., 1993) and the results of the phenanthroline–copper footprinting (Figure 3A) showed that the CA-1 activity could protect the same sequence as CCA1. The predicted molecular weight of the encoded protein (66,970) is approximately the same as the apparent molecular weight of the partially purified CA-1 activity (Sun et al., 1993). There are, however, some differences in the relative importance of specific nucleotides for the binding of the two activities. The binding of CCA1 was more affected by alteration in the TCT sequence than by alterations in the AAAAA, whereas the opposite was observed with the plant extract activity (cf. m3 and m4 in Figure 3D). It is possible that the differences observed are the result of differences in modifications of the protein in *E. coli* and plants or that the CA-1 activity in the plant extracts might be associated with another protein that might alter its binding characteristics. It is also possible that CA-1 and CCA1 are the products of two genes, or the result of alternative splicing, in which case they may compete for the same binding sites. Further studies are required to address definitively the relationship between CCA1 and CA-1.

We have previously shown that the region of the *Lhcb1*3* promoter to which the CA-1 protein binds is essential both for a high level of expression and for its phytochrome responsiveness (Kenigsbuch and Tobin, 1995). A mutation was made in a 1.15-kb segment of this promoter and consisted of the same nine nucleotide changes as in the m1 construct used in the competition experiments (Figure 3C); the activity of mutant and wild-type promoter and reporter constructs was examined in transgenic Arabidopsis lines. The expression of the mutant construct was much lower than that of the wild-type construct and showed no responsiveness to phytochrome. Based on these data, we suggested that CA-1 might act as an activator of transcription of the *Lhcb1*3* gene. The m1 mutation also affects the binding of the CCA1 protein. Therefore, by analyzing the effects of expressing the CCA1 cDNA in an antisense orientation in transgenic Arabidopsis plants, we tested whether the protein encoded by *CCA1* could affect expression or phytochrome regulation of transcription in vivo. In five of seven such transgenic lines, a substantial decrease in the phytochrome-induced increase in *Lhcb1*3* RNA was observed. This result demonstrates that CCA1 can affect the phytochrome regulation of *Lhcb1*3* in vivo. The fact that none of the lines we recovered showed strong suppression of *CCA1* RNA suggests the possibility that complete loss of function might be highly deleterious. Our results strongly support the

**Figure 3.** (continued).

**(A)** Phenanthroline–copper footprints of the affinity-purified CCA1 protein expressed in *E. coli* and of the partially purified plant CA-1 protein. Shown at left are the results of an EMSA, with the A2 fragment and with the amounts of proteins and poly(dI-dC) shown above each reaction. This gel was treated with phenanthroline–copper to cleave the DNA, and the DNA of each band was recovered and separated on a sequencing gel, shown at right. The labeling of reactions and complexes is the same as for the EMSA; lanes S are the G+A chemical sequencing reactions. The lines between the lanes highlight the regions protected by the proteins. The sequence of the protected region is also shown. B, CA-1 protein–DNA complex; B1 and B2, CCA1 protein–DNA complexes; F, free probe.

**(B)** DNA modification interference with CCA1 binding. Partially methylated (gels I and II) or depurinated (gels III and IV) A2 probes, which were labeled at the 3′ end of either the coding strand (I and III) or the noncoding strand (II and IV) of the DNA, were incubated with the CCA1 protein. The free DNA (F), protein-bound DNA (B), and DNA not incubated with the protein (C) were cleaved by piperidine and separated on an 8% polyacrylamide–urea sequencing gel. Arrows indicate the positions at which modification of DNA interferes with the CCA1 binding.

**(C)** Summary of binding site interactions and protection by CCA1 and CA-1, and sequences of DNA fragments used in the competition EMSA. A2 is the sequence of the A2 fragment of the *Lhcb1*3* promoter. Thick lines and the thin line indicate regions protected by CCA1 and CA-1, respectively, in the phenanthroline–copper footprint assay **(A)**. Dashed lines indicate regions protected by CA-1 in the DNase I footprint assay (Sun et al., 1993). Asterisks designate nucleotides that interfere with CCA1–DNA binding when methylated; boldface letters indicate nucleotides that interfere with the CCA1–DNA binding when depurinated **(B)**. WT1, WT2, and m1 to m4 are sequences of wild-type and mutated DNA fragments used as competitors. Dashes represent nucleotides identical to those of the A2 fragment. WT2 is the −117 to −78 region relative to transcription start of the *Lhcb1*1* gene. Dots denote gaps introduced to optimize the alignment of the conserved sequence elements.

**(D)** Competition EMSA analysis of the CCA1 protein and CA-1 activity, using wild-type and mutant promoter fragments. All lanes contain 0.3 ng of the $^{32}$P-labeled A2 probe. Triangles above the lanes denote increasing amounts (6, 20, and 60 ng) of competitor DNA. Except for lanes labeled "No protein," each lane contains 4.4 ng of purified CCA1 protein expressed in *E. coli* (top) or 0.46 μg of protein in plant extracts enriched for CA-1 (bottom). Arrows indicate the protein–DNA complexes. Autoradiograms and quantification of the binding are shown. The sequences of the competitors are given in **(C)**.

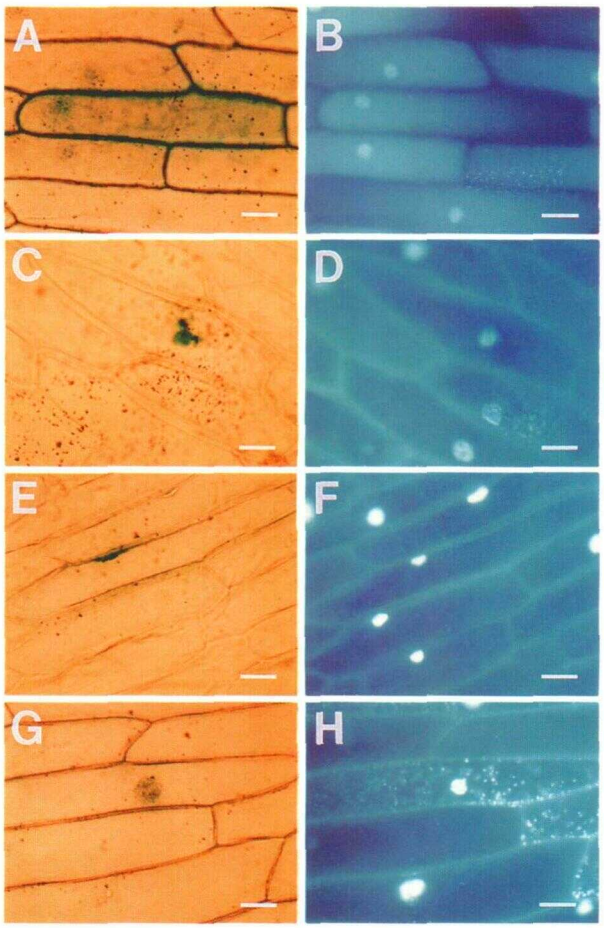


**Figure 4.** The CCA1–GUS Fusion Protein Is Localized to Nuclei in Onion Epidermal Cells.

Tissues were analyzed both for GUS activity by colorimetric X-gluc staining (**[A]**, **[C]**, **[E]**, and **[G]**) and for nuclei by 4′-6-diamidino-2-phenylindole (DAPI) fluorescence staining (**[B]**, **[D]**, **[F]**, and **[H]**). Shown is microscopy of onion cells after staining.

**(A)** and **(B)** The same onion cells transformed with pMF::GUS as a negative control.

**(C)** and **(D)** The same onion cells transformed with pMF::B::GUS as a positive control (Varagona et al., 1992).

**(E)** to **(H)** Two sets of cells transformed with p35S–CCA1–GUS.

Bars = 20 μm.

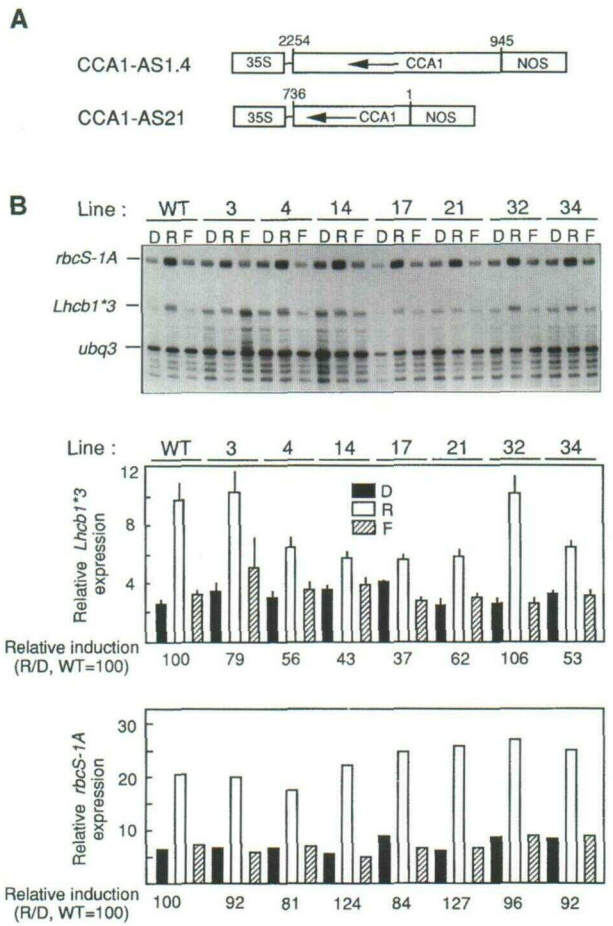


**Figure 5.** Phytochrome Regulation of *Lhcb1*3* and *rbcS-1A* Genes in Transgenic Plants Expressing Antisense *CCA1* RNA.

**(A)** Diagrams of constructs for expression of antisense *CCA1* in transgenic plants. The 35S indicates the 35S promoter of cauliflower mosaic virus. NOS is the transcription termination sequence of the nopaline synthase gene. Arrows show the sense direction of the *CCA1* gene. Numbers show the nucleotide positions of the cDNA fragments.

**(B)** RNase protection assays for *Lhcb1*3* and *rbcS-1A* RNA. Five-day-old dark-grown seedlings were given no light treatment (D), 2 min of R, or 2 min of R followed by 10 min of FR at 4 hr before harvesting (F). Line 21 was transformed with construct CCA1–AS21. All of the other lines were transformed with CCA1–AS1.4. The *ubq3* gene was used as an internal control. Quantification of the *Lhcb1*3* and *rbcS-1A* RNA levels, normalized to the *ubq3* RNA, is shown below the gel. The averages of three independent experiments are shown with standard errors for the *Lhcb1*3* expression. The averages of two independent experiments are shown for *rbcS-1A* gene expression. Relative induction is the ratio of the R sample to the D sample as a percentage of this ratio for the wild-type line, which is set to 100 for each RNA. WT, wild type.

idea that CCA1 normally interacts with this promoter in vivo and acts as an activator of the phytochrome response of the *Lhcb1*3* gene.

The expression level of *Lhcb1*3* was not affected in dark-grown seedlings of any of the antisense lines. This latter observation suggests the possibility that the large reduction of expression seen in transgenic plants of the m1 mutant *Lhcb1*3* promoter and reporter gene construct (Kenigsbuch and Tobin, 1995) may have been caused by an effect on the binding of a transcription factor in addition to CCA1. Be-

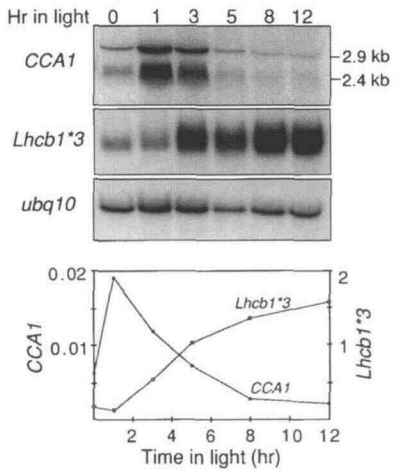


**Figure 6.** *CCA1* RNA Is Induced by Light in Etiolated Seedlings before the *Lhcb1*3* RNA.

Arabidopsis seedlings were grown in the dark for 6 days, transferred to continuous light, and harvested at the specified time after the transfer. RNA (12.5 μg for the 5-hr sample and 15 μg for all other samples) was analyzed on gel blots by hybridization with RNA probes of *Lhcb1*3*, *ubq10*, and *CCA1* genes. The three gels represent different exposure times. The approximate length of each band of *CCA1* RNA is shown at right. Quantitation of the amounts of signal from *CCA1* and *Lhcb1*3* RNAs relative to the *ubq10* RNA is also shown.

cause there is a G-box (ACGT-containing) sequence adjacent to the binding site for CA-1, it is possible that the mutation also affected the binding of a member of the G-box transcription factor family (Menkens et al., 1995) either directly, because of the DNA sequence changes, or indirectly, by a mechanism that might involve an interaction with the CCA1 protein. It has been frequently observed in other systems that Myb proteins act in combination with other transcription factors (e.g., Tice-Baldwin et al., 1989; Goff et al., 1992; Burk et al., 1993; Larkin et al., 1994). The presence of a CCAAT box downstream of the conserved CCA1 binding sequence (Table 1) also raises the possibility that CCA1 could interact with a CCAAT-box binding protein. The CCAAT motif occurs in many eukaryotic promoters, and it has been proposed that factors binding to this sequence can facilitate recruitment of transcription factors binding upstream of this sequence (Wright et al., 1994).

The proposed role in transcriptional activation of the protein(s) binding to the AAA$^{A}/_{C}$AATCTA region contrasts with our earlier proposal that CA-1 might act to repress transcription in dark-grown seedlings (Sun et al., 1993). This idea was based on the lack of CA-1 binding activity in seedlings of a *det1* mutant, in which *Lhcb* RNAs accumulate in dark-grown seedlings (Chory et al., 1989; Chory and Peto, 1990). However, we have since found that CA-1 activity is present in *det1-6*, a null allele of *det1* (Pepper et al., 1994). It is possible that the *det1-1* allele, used in our earlier work (Sun et al., 1993), actually makes a truncated protein (Pepper et al., 1994) whose presence affects the DNA binding activity of CA-1. CA-1 activity is also present in the *cop1-4* (Deng and Quail, 1992) mutant (Z.-Y. Wang and E.M. Tobin, unpublished data).

It is also of interest that the expression of another phytochrome-regulated gene, the *rbcS-1A* gene, was not affected in the antisense lines. It is striking that the phytochrome regulation of one gene can be altered without affecting the regulation of another. It has been suggested that there are several signal transduction pathways leading to phytochrome-activated gene expression (reviewed in Barnes et al., 1995), and a number of different observations have indicated that at least the final steps of the phytochrome signal transduction pathway differ between *Lhcb* and *rbcS* genes (reviewed in Tobin and Kehoe, 1994). The results with the antisense transgenic plants and the sequence conservation of the CCA1 binding sequence in *Lhcb* genes, but not in *rbcS* genes, suggest that CCA1 is specifically involved in regulating the transcription of *Lhcb* genes.

The conservation of the AAAAATCT motif in *Lhc* genes of many species (Table 1) suggests the possible general involvement of CCA1-related proteins in the regulation of *Lhc* genes. An AAAAATCTA sequence is present in the closely related *Lhcb1*1* gene of Arabidopsis, and the results of competition assays suggested that both the CA-1 activity and the CCA1 protein can also interact with this sequence in the promoter of the *Lhcb1*1* gene. This sequence is present in a 38-bp region sufficient for circadian regulation of the *Lhcb1*1* gene and that has been found to bind several activities (CUF2/3) that differ from CCA1/CA-1 (Carré and Kay, 1995). Further studies are necessary to test whether the CCA1 protein might also be involved in the circadian regulation of this gene.

Transferring etiolated seedlings to white light resulted in a transient increase in *CCA1* RNA. An Antirrhinum *myb* gene has been reported to be expressed at a lower level in light-grown seedlings than in etiolated seedlings (Jackson et al., 1991), and some alleles of another *myb* gene, the *C1* gene of maize, are expressed at a higher level in light- than in dark-grown seedlings (Scheffler et al., 1994; Kao et al., 1996). The finding that the level of *CCA1* RNA increases in response to light more rapidly than the increase in *Lhcb* RNA is consistent with the idea that the CCA1 protein plays an important role in the phytochrome signal transduction chain leading to increased expression of *Lhcb* genes. Although further studies are necessary to characterize the effects of light on the level and/or activity of the CCA1 protein, it seems likely that the phytochrome-induced increase in *Lhcb* expression is caused at least in part by an increase in *CCA1* RNA levels.

In summary, the *CCA1* gene that we have cloned encodes a Myb-related protein that binds to a region of the *Lhcb1*3*

**Table 1.** Conservation of AAAAATCT and CCAAT Sequences in the *Lhc* Promoters

| Species | Locus Name[a] | Sequence[b] | Position[c] |
|---|---|---|---|
| Arabidopsis | *ATHCAB1A* | aaAcAAACAATCTAaacc | −109Tx |
| | | ccAAAAAAAATCTAtgactagCCAATagca | −79Tx |
| | *AT5FCAB2* | aaAAcAAAAATCTtaaaatCCAATgaatga | −78Tx |
| | *AT5FCAB3* | aaAAccAAAATCTcaaaaatCCAATgagta | −95Tx |
| | *ATLHB1B1* | aaAccAAAAATCTgatgccagctcgaca | −207Tl |
| | | agAcAAAAAATCTA | −193Tl |
| | | agctcAAAAATgTcaacaaCCAATagaaat | −163Tl |
| | | gcAAAttAAATCTtcaagttttgCAATtca | −106Tl |
| | *ATLH1B2* | aatcccAAAATCTcataagCCAATcactaa | −134Tl |
| | *ATHCAB6P* | taatatacg*TAGATT*c*T*gc*TT*tt**ATCT**gaa | −116Tx |
| | *ATLHCB4* | tgacCCAATc*AGATT*g | −242Tl |
| | | tacctc**ATCT***AGATTTT*a*TT*gCCAATggac | −212Tl |
| *Brassica napus* | *BNPHIICHL* | aatcccAAcATCTAactcttatCAATatcc | −133Tl |
| Cotton | *COTIIABINA* | aacAtcAAAATCTtggaagctCCAATgaaa | −78Tx |
| | *COTIIABINB* | acAtAtAAAATCTtcacttCCATTttcata | −38Tl |
| | *GHCAB* | tagccttAAATCTgt | −135Tl |
| | | tgAAAgAAAATCTAacg-n12-catCCAAT | −98Tl |
| Soybean | *GMCAB3* | ccAAAcAAAATCTcaaacccCCAATgagat | −144Tl |
| *Lemma gibba* | *LGIAB19A* | atAcAgAAAATCTgccat-n12-aaCCAAT | −128Tx |
| Rice | *OSCABR1* | tcAccAAgAATCTtcgagaaacttataaac | −232Tl |
| | | cgccgAAAAATCTcggacaaacccgcggct | −200Tl |
| Petunia | *PECAB22L* | gctAAtggAATCTtgaaacCCAATgaaatt | −144Tl |
| | *PHCAB37* | caAggAAAAATCagaaggaaatCCAAATcc | −98Tx |
| Pea | *PSCABIIC* | atAtgAAgAATCTAtcccataagtggatca | −127Tx |
| | | tagAcAAAAATCTccaactcCCAATgaaat | −82Tx |
| | *PEACAB66* | gcAAAgAAAATCTccaactagCCATAgctt | −64Tx |
| | *PEACAB80* | gcAAAgAAAATCTccaactagCCATAgctt | −64Tx |
| Mustard | *SACAB* | aactAAAgAATCTAaaccaaaagctgtggc | −142Tx |
| | | atAtccAAAATCTcattagCCAATcaggaa | −80Tx |
| Spinach | *SOCAB1PRO* | tagAAAtAAATCTccaaaatgcatggctta | −140Tx |
| | | aacAccAAAATCTtcctaagCAAATaagaa | −91Tx |
| Tobacco | *TOBCABA* | catgtAtggATCTttgaaacCCAATgaaat | −143Tl |
| | *TOBCABB* | atAAAAAgtATCTAAAAAAATCTtgtcatg | −152Tx |
| | *TOBCABC* | gctAgtgAAATCTtgaaatCCAATgaaaat | −138Tl |
| | *NTCAB7* | ttAAAcAAAtTgTtATCTtCCAATgagaaa | −151Tl |
| Tomato | *LECAB1AUP* | acACAtgggATCTtgatacCCAATgagatt | −148Tl |
| | *LECAB9* | tccAcAttcATCTtccctgtggctCCAATa | −149Tl |
| | *LECAB11* | ttccAAAAAATCTtagt | −79Tx |
| | | gactcttca*AGATT*gaaccagagccaCAAT | −49Tx |
| | *LELHBC1g* | acAcAAAAcATCTccaaccaCAATttctac | −125Tl |
| | *TOMCAB* | tatagCCAATAAcAATCTAaattc*AGATTT* | −56Tx |
| | *TOMCBPA* | acAcAtgggATCTtgatacCCAATgagatc | −123Tl |
| | *TOMCBPB* | tttgagt*TAGATTTTTT*-n13-tcaCCAAT | −145Tl |
| | *TOMCABIN* | tggtccgAAATCTAtccaccagagatcatt | −159Tl |
| | *TOMCBBIN* | tggtcAgAAATCTAtccac*TAGATTT*ca*T*c | −138Tl |
| | *ZMCAB1* | ggtgAgAAtATCTggcgactggcggagacc | −149Tx |
| | *ZMCABM7* | tccaagcatc*AGATTT*a*TTTT*tattttctg | −83Tx |

[a] From GenBank.

[b] Nucleotides identical to those of *Lhcb1*3* (*ATHCAB1A*) at or near the AATCT sequence are in boldface uppercase letters (complementary nucleotides are italicized). Lowercase letters indicate nonidentical nucleotides. CCAAT boxes are in uppercase letters.

[c] Positions of the last nucleotides relative to translational start (Tl) or transcriptional start (Tx) of the gene.

promoter that is essential for phytochrome regulation of this gene. The encoded protein can affect the phytochrome regulation of this promoter in vivo but does not affect an *rbcS* gene that also can be regulated by phytochrome. The expression of the *CCA1* gene itself can be regulated by light. Our data show that CCA1 is a part of the phytochrome signal transduction chain leading to increased transcription of the *Lhcb1*3* gene in Arabidopsis.

## METHODS

### Plant Materials and Growth Conditions

*Arabidopsis thaliana* ecotype Columbia was used in all experiments, except ecotype Nossen (No-O) was used for transformation with the antisense *CCA1* constructs. The medium used for plant growth (MS2S medium) contained 1 × Murashige and Skoog salts (Gibco BRL), 0.05% Mes, pH 5.7, 0.8% Phytagar (Gibco BRL), and 2% sucrose. Light-grown plants were maintained at 24°C in a growth chamber with light intensity of 150 $\mu E\ m^{-2}\ sec^{-1}$. Growth and light treatments of etiolated seedlings for phytochrome experiments were as described previously (Brusslan and Tobin, 1992). White onions used for nuclear localization experiments were purchased from a local supermarket.

### Cloning and Sequencing of CCA1 cDNA and Genomic Clones

Poly(A) RNA was isolated from leaves of Arabidopsis grown for 3 weeks on soil in continuous white light. A directional cDNA expression library was constructed in λgt22A by using the SuperScript Lambda system (Bethesda Research Laboratory instruction manual). The library was screened essentially as described by Singh et al. (1989), except that NEB buffer (25 mM Hepes-NaOH, pH 7.2, 40 mM KCl, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 10% glycerol; Sun et al., 1993) was used as the binding buffer, and the washing solution was supplemented with 0.25% nonfat dry milk and 0.1% Triton X-100. Approximately 640,000 unamplified recombinant phage plaques were screened in the first round by using double-stranded A2 DNA probe (A2; Figure 3D; Sun et al., 1993). The positive plaques from the initial screening were rescreened using both the A2 probe and a mutant probe (m1 probe; Figure 3C) that binds poorly to the CA-1 activity (Sun et al., 1993). Two phage clones (clones 21 and 24) that bound only to the A2 and not to the m1 probe were isolated as individual plaques. The cDNA inserts were subcloned into the SalI and NotI restriction sites of pGEM-11Zf(−) (Promega). To reconstruct a full-length cDNA (clone 25), the 5′ end of the cDNA clone 24 was removed as a SalI-PstI fragment and replaced with that of clone 21. Sequencing of the region spanning the PstI junction of clone 25 confirmed the reconstitution of wild-type sequence.

A genomic clone corresponding to the CCA1 cDNA was isolated by screening a genomic library of Arabidopsis ecotype Columbia in λGEM11 (Promega), using the SstI-NotI fragment of CCA1 cDNA clone 24 (corresponding to nucleotides 950 to 2254 of the full-length cDNA).

The sequences of the CCA1 cDNA and overlapping fragments of the genomic clone were determined by the dideoxy chain termination method using a Sequenase kit (United States Biochemical) and double-stranded plasmid DNA. Both strands of the cDNA and genomic DNA were completely sequenced.

### Sequence Analysis and Database Searching

The protein and DNA sequences were analyzed using the MacVector software (IBI, New Haven, CT) and the Genetics Computer Group (Madison, WI) software package. The GenBank database was searched with the amino acid sequence of CCA1 by using the BLAST (Altschul et al., 1990) and FASTA programs (Pearson and Lipman, 1988) on the National Center for Biotechnology Information (NCBI) on-line service. Sequence alignment was assembled manually, based on the results of database searches.

Genomic sequences of light-harvesting complex apoprotein (*Lhc*) genes and small subunit of ribulose bisphosphate carboxylase/oxygenase (*rbcS*) genes were retrieved from the GenBank database by using text search on the NCBI worldwide web site. The presence of AATCT sequences in the promoter regions of the genes was detected using the FASTA program and further analyzed visually.

### DNA and RNA Gel Blot Analyses

Genomic DNA isolation and DNA gel blotting were performed as described by Brusslan et al. (1993). Membranes were hybridized with $^{32}P$-labeled CCA1 cDNA fragments under high-stringency conditions (final washes were at 65°C in 0.1% SSC [1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate] and 0.1% SDS) and then stripped and reprobed under low-stringency conditions (hybridization at 32°C in buffer containing 50% formamide, 0.25 M $NaHPO_4$, pH 7.2, 0.75 M NaCl, 7% [w/v] SDS, and 1 mM EDTA, and final washes at 45°C in 2 × SSC and 0.1% SDS).

Total RNA was extracted from Arabidopsis seedlings as described by Brusslan and Tobin (1992). Total RNA was separated on a 1% agarose gel containing formaldehyde and blotted onto Zeta-Probe membrane (Bio-Rad, Richmond, CA), following the manufacturer's instructions. RNA probes were synthesized by in vitro transcription, using linearized plasmid DNA. *CCA1* RNA probe was synthesized from CCA1 clone 24. To make the *ubq10* RNA probe, a fragment of the 3′ untranslated region of the *ubq10* gene (Callis et al., 1995) was amplified by polymerase chain reaction (PCR), using the primers 5′-CTGTTATGCTTAAGAAGTTCAATGT-3′ and 5′-CCACCCTCGAGTAGAACACTTATTCAT-3′. The amplified fragment was digested with HindIII and XhoI and cloned into pGEM-11Zf(−). This plasmid DNA was digested with HindIII and used as a template for synthesis of the *ubq10* RNA probe. The *Lhcb1*3* RNA probe was made as described by Brusslan and Tobin (1992). The membrane blot was hybridized overnight with the RNA probes, in buffer containing 50% formamide, 0.3 M NaCl, 0.05 M $NaHPO_4$, pH 6.5, 1 mM EDTA, 1% SDS, 0.1% Ficoll (type 400), 0.1% polyvinylpyrrolidone, 0.1% BSA, 0.5 mg/mL yeast tRNA, and 0.5 mg/mL herring sperm DNA. Hybridization of *Lhcb1*3*, *ubq10*, and *CCA1* probes was performed at 55, 52, and 58°C, respectively. Final washes were performed at 65°C in 0.1 × SSC and 0.1% SDS. After hybridization with *Lhcb1*3* and *ubq10* probes, the blot was stripped by boiling in 0.1 × SSC and 0.1% SDS, and then hybridized with the *CCA1* probe. The blots were imaged and quantified using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The measurement of the signal for each probe was adjusted for the uridine content of the probe and the exposure time, and the *Lhcb1*3* and *CCA1* signals were normalized to the *ubq10* signal.

### Primer Extension

The primer extension experiment was performed as described by Ausubel et al. (1987). The sequence of the oligonucleotide primer corresponded to nucleotides +42 to +22 of CCA1 cDNA clone 21. Fifty femtomoles of $^{32}$P-labeled primer was annealed to 45 μg of total RNA. The primer was extended with 9.5 units of avian myeloblastosis virus reverse transcriptase (Promega) at 37°C for 1 hr. The dideoxy sequencing reactions were performed using the same $^{32}$P-labeled oligonucleotide primer and *CCA1* clone 21 plasmid DNA.

### Protein Expression in *Escherichia coli* and Purification of Glutathione *S*-Transferase–CCA1 Proteins

The constructs diagrammed in Figure 2A were made by cloning the CCA1 cDNA fragments into pGEX-3X (Pharmacia) by using PCR-aided cloning with the following 5′ primers: 5′-GGCCGGGATCCAA-TTCGTCGACCCACGCG-3′ for pXCA-21, pXCA-24, and pXCA-25; and 5′-TAAAGGGATCCATATGGGTCAAGCGCTAG-3′ for pXCA-23. A 3′ primer (5′-ATAGAATTCTCGAGCTTATGCATGCGG-3′) was used for pXCA-21, pXCA-24, pXCA-25, and pXCA-23. The appropriate plasmid DNA (0.5 μg) was amplified for 10 cycles, and the PCR products were digested with EcoRI and BamHI. Cloning of the cDNAs of clones 21, 24, and 25 and the 483- to 2254-nucleotide region into pGEX-3X yielded pXCA-21, pXCA-24, pXCA-25, and pXCA-23, respectively. Sequencing of the junction region between the glutathione *S*-transferase (*GST*) gene and cDNA confirmed the construction of a translational fusion in pXCA-24 and pXCA-23.

The plasmid constructs were transformed into *E. coli* BL21(DE3). Protein expression, purification of GST–CCA1 fusion proteins by using glutathione agarose, and purification of CCA1 polypeptides by cleavage of the matrix-bound GST fusion protein with factor Xa were performed following the procedure of Ausubel et al. (1987). Protein concentrations were determined by the Bradford assay (Bio-Rad), using BSA as standard.

### Electrophoretic Mobility Shift Assays and Competition

Partial purification of Arabidopsis CA-1 protein, A2 probe labeling, and the electrophoretic mobility shift assays (EMSAs) were performed as described by Sun et al. (1993). Competitor DNA fragments were prepared by annealing synthetic oligonucleotides. Competition EMSA experiments were performed by adding partially purified plant CA-1 protein or affinity-purified CCA1 polypeptide expressed from pXCA-24 in *E. coli* to the DNA binding reaction mixture containing A2 probe and specified amounts of competitor DNA fragments. The dried gels were imaged and quantified using a PhosphorImager.

### Phenanthroline–Copper Footprinting

The A2 fragment was labeled with phosphorus-32 at the 3′ end of the sense strand by end filling (Sun et al., 1993). Footprint experiments were performed as described by Kuwabara and Sigman (1987). The EMSA reactions were scaled up fivefold; $10^6$ cpm of probe and specified amounts of protein and poly(dI-dC) were used in each reaction. After electrophoresis, the gel was treated with phenanthroline–copper and then exposed wet to x-ray film for 40 min. The bands representing free DNA and protein–DNA complexes were excised from the gel. DNA was eluted from the gel slices, recovered by ethanol precipitation, and loaded on an 8% polyacrylamide–urea sequencing gel. The G+A chemical cleavage sequencing reaction was performed as described by Maxam and Gilbert (1980).

### Methylation Interference and Depurination Interference Assays

Partial methylation and depurination of the A2 DNA probe were performed following the procedure of DNA chemical sequencing (Maxam and Gilbert, 1980). Five nanograms ($10^5$ cpm) of modified DNA probe was incubated with 0.8 μg of affinity-purified CCA1 protein in 50 μL of NEB buffer containing 5 μg of poly(dI-dC) and 10 μg of BSA. The protein-bound DNA and free DNA were separated by filtering the mixture through a nitrocellulose membrane (Ausubel et al., 1987). Free and bound fractions of DNA were recovered and cleaved with piperidine following the DNA chemical sequencing procedure. An aliquot of probes not incubated with protein was also cleaved with piperidine as a control. Equal amounts of radioactivity from each sample were used on an 8% polyacrylamide–urea sequencing gel.

### Nuclear Localization Assays

An XbaI site and a BamHI site were introduced into *CCA1* by PCR amplification of cDNA clone 25, using the 5′ primer (5′-GAAGT-TGTCTAGAGGAGCTAAGTG-3′) and 3′ primer (5′-ATGTGGATCCT-TGAGTTTCCAACCGC-3′) (mismatches are underlined). The resulting PCR product was digested with XbaI and BamHI and inserted in pBI221 (Clontech, Palo Alto, CA), yielding p35S–CCA1–GUS. This construct contains *CCA1* coding sequence as a 1828-bp XbaI-BamHI fragment inserted between the cauliflower mosaic virus 35S promoter and the *uidA* gene. pMF::GUS and pMF::B::GUS were obtained from N. Raikhel (Michigan State University, East Lansing, MI); construction of these plasmids is described in Varagona et al. (1992). Onion epidermal peels were transformed by biolistic transformation and analyzed for GUS activity, and nuclei localization was as described in Varagona et al. (1992). Histochemical staining was visualized using a Zeiss Axiophot microscope and photographed using Kodak Ektachrome (Elite Series) ASA 400 film.

### Plant Transformation

The SstI-NotI fragments of *CCA1* clones 21 and 24 in pGEM11Zf(−) were cloned into pBluescript KS− (Stratagene, La Jolla, CA) at the corresponding sites. The resulting plasmids were digested with BamHI and SstI, and the cDNA fragments were cloned into the BamHI and SstI sites of binary vector pBI121 (Clontech), replacing the *uidA* gene coding sequence. The binary vectors were transformed into *Agrobacterium tumefaciens* A2260. Arabidopsis ecotype No-O plants were transformed with the above constructs, using the Agrobacterium-mediated root transformation procedure described by Valvekens et al. (1988).

### RNase Protection Assays

The RNA probe synthesis and RNase protection assays were performed as described by Brusslan and Tobin (1992).

## ACKNOWLEDGMENTS

This research was supported by National Institutes of Health (NIH) Grant No. GM23167. M.S.O. was supported in part by an NIH Minority Predoctoral Fellowship (No. F31 GM18046). We thank Drs. Rita Varagona and Natasha Raikhel (Michigan State University) for the pMF::GUS and pMF::B::GUS constructs used in the nuclear localization experiments and Drs. John Mulligan and Ron Davis (Stanford University, Stanford, CA) for the Arabidopsis genomic library. We are grateful to Drs. Rachel Green, Karam Singh, Shoji Sugano, and Sharlene Weatherwax for comments on the manuscript.

Received December 9, 1996; accepted February 17, 1997.

## REFERENCES

**Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

**Anderson, S.L., and Kay, S.A.** (1995). Functional dissection of circadian clock- and phytochrome-regulated transcription of the *Arabidopsis CAB2* gene. Proc. Natl. Acad. Sci. USA **92,** 1500–1504.

**Anderson, S.L., Teakle, G.R., Martino-Catt, S.J., and Kay, S.A.** (1994). Circadian clock- and phytochrome-regulated transcription is conferred by a 78 bp *cis*-acting domain of the *Arabidopsis CAB2* promoter. Plant J. **6,** 457–470.

**Argüello-Astorga, G.R., and Herrera-Estrella, L.R.** (1996). Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. Plant Physiol. **112,** 1151–1166.

**Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K.,** eds (1987). Current Protocols in Molecular Biology. (New York: Greene Publishing Associates and Wiley-Interscience).

**Baranowskij, N., Frohberg, C., Prat, S., and Willmitzer, L.** (1994). A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J. **13,** 5383–5392.

**Barnes, S.A., Quaggio, R.B., and Chua, N.-H.** (1995). Phytochrome signal-transduction: Characterization of pathways and isolation of mutants. Philos. Trans. R. Soc. Lond. B **350,** 67–74.

**Bruce, W.B., Deng, X.-W., and Quail, P.H.** (1991). A negatively acting DNA sequence element mediates phytochrome-directed repression of *phyA* gene transcription. EMBO J. **10,** 3015–3024.

**Brusslan, J.A., and Tobin, E.M.** (1992). Light-independent developmental regulation of *cab* gene expression in *Arabidopsis thaliana* seedlings. Proc. Natl. Acad. Sci. USA **89,** 7791–7795.

**Brusslan, J.A., Karlin-Neumann, G.A., Huang, L., and Tobin, E.M.** (1993). An Arabidopsis mutant with a reduced level of *cab140* RNA is a result of cosuppression. Plant Cell **5,** 667–677.

**Burk, O., Mink, S., Ringwald, M., and Klempnauer, K.-H.** (1993). Synergistic activation of the chicken *mim*-1 gene by v-*myb* and C/EBP transcription factors. EMBO J. **12,** 2027–2038.

**Callis, J., Carpenter, T., Sun, C.W., and Vierstra, R.D.** (1995). Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins in *Arabidopsis thaliana* ecotype Columbia. Genetics **139,** 921–939.

**Carabelli, M., Sessa, G., Baima, S., Morelli, G., and Ruberti, I.** (1993). The *Arabidopsis Athb-2* and *-4* genes are strongly induced by far-red–rich light. Plant J. **4,** 469–479.

**Carré, I.A., and Kay S.A.** (1995). Multiple DNA-protein complexes at a circadian-regulated promoter element. Plant Cell **7,** 2039–2051.

**Chamnongpol, S., Willekens, H., Langebartels, C., Van Montagu, M., Inzé, D., and Van Camp, W.** (1996). Transgenic tobacco with a reduced catalase activity develops necrotic lesions and induces pathogenesis-related expression under high light. Plant J. **10,** 491–503.

**Chory, J., and Peto, C.A.** (1990). Mutations in the *DET1* gene affect cell-type–specific expression of light-regulated genes and chloroplast development in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **87,** 8776–8780.

**Chory, J., Peto, C., Feinbaum, R., Pratt, L., and Ausubel, F.** (1989). Arabidopsis thaliana mutant that develops as a light-grown plant in the absence of light. Cell **58,** 991–999.

**Chory, J., Chatterjee, M., Cook, R.K., Elich, T., Fankhauser, C., Li, J., Nagpal, P., Neff, M., Pepper, A., Poole, D., Reed, J., and Vitart, V.** (1996). From seed germination to flowering, light controls plant development via the pigment phytochrome. Proc. Natl. Acad. Sci. USA **93,** 12066–12071.

**Chou, P.Y., and Fasman, G.D.** (1978). Empirical prediction of protein conformations. Annu. Rev. Biochem. **47,** 251–276.

**Degenhardt, J., and Tobin, E.M.** (1996). A DNA binding activity for one of two closely defined phytochrome regulatory elements in an *Lhcb* promoter is more abundant in etiolated than in green plants. Plant Cell **8,** 31–41.

**Deng, X.-W.** (1994). Fresh view of light signal transduction in plants. Cell **76,** 423–426.

**Deng, X.-W., and Quail, P.H.** (1992). Genetic and phenotypic characterization of *cop*1 mutants of *Arabidopsis thaliana*. Plant J. **2,** 83–95.

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

**Furuya, M., and Schäfer, E.** (1996). Photoperception and signalling of induction reactions by different phytochromes. Trends Plant Sci. **9,** 301–307.

**Gerondakis, S., and Bishop, J.M.** (1986). Structure of the protein encoded by the chicken proto-oncogene c-*myb*. Mol. Cell. Biol. **6,** 3677–3684.

**Gilmartin, P.M., and Chua, N.-H.** (1990). Localization of a phytochrome-responsive element within the upstream region of pea *rbcS-3A*. Mol. Cell. Biol. **10,** 5565–5568.

**Goff, S.A., Cone, K.C., and Chandler, V.L.** (1992). Functional analysis of the transcriptional activator encoded by the maize *B* gene: Evidence for a direct functional interaction between two classes of regulatory proteins. Genes Dev. **6,** 864–875.

**Grotewold, E., Drummond, B.J., Bowen, B., and Peterson, T.** (1994). The *myb*-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell **76,** 543–553.

**Harter, K., Kircher, S., Frohnmeyer, H., Krenz, M., Nagy, F., and Schäfer, E.** (1994). Light-regulated modification and nuclear

*translocation of cytosolic G-box binding factors in parsley.* Plant Cell **6,** 545–559.

**Hinnebusch, A.G.** (1994). Translational control of GCN4—An in vivo barometer of initiation-factor activity. Trends Biochem. Sci. **19,** 409–414.

**Jackson, D., Culianez-Macia, F., Prescott, A.G., Roberts, K., and Martin, C.** (1991). Expression patterns of *myb* genes from *Antirrhinum* flowers. Plant Cell **3,** 115–125.

**Jansson, S., Pichersky, E., Bassi, R., Green, B.R., Ikeuchi, M., Melis, A., Simpson, D.J., Spangfort, M., Staehelin, L.A., and Thornber, J.P.** (1992). A nomenclature for the genes encoding the chlorophyll *a/b*-binding proteins of higher plants. Plant Mol. Biol. Rep. **10,** 242–253.

**Kao, C.-Y., Cocciolone, S.M., Vasil, I.K., and McCarty, D.R.** (1996). Localization and interaction of the *cis*-acting elements for abscisic acid, VIVIPAROUS1, and light activation of the *C1* gene of maize. Plant Cell **8,** 1171–1179.

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

**Kehoe, D.M., Degenhardt, J., Winicov, I., and Tobin, E.M.** (1994). Two 10-bp regions are critical for phytochrome regulation of a *Lemna gibba Lhcb* gene promoter. Plant Cell **6,** 1123–1134.

**Kendrick, R.E., and Kronenberg, G.H.M.,** eds (1994). Photomorphogenesis in Plants, 2nd ed. (Dordrecht, The Netherlands: Kluwer Academic Publishers).

**Kenigsbuch, D., and Tobin, E.M.** (1995). A region of the *Arabidopsis Lhcb1*3* promoter that binds to CA-1 activity is essential for high expression and phytochrome regulation. Plant Physiol. **108,** 1023–1027.

**Kuwabara, M.D., and Sigman, D.S.** (1987). Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: *Escherichia coli* RNA polymerase–*lac* promoter complexes. Biochemistry **26,** 7234–7238.

**Larkin, J.C., Oppenheimer, D.G., Lloyd, A.M., Paparozzi, E.T., and Marks, M.D.** (1994). Roles of the *GLABROUS1* and *TRANSPARENT TESTA GLABRA* genes in Arabidopsis trichome development. Plant Cell **6,** 1065–1076.

**Leutwiler, L.S., Meyerowitz, E.M., and Tobin, E.M.** (1986). Structure and expression of three light-harvesting chlorophyll *a/b*-binding protein genes of *Arabidopsis thaliana.* Nucleic Acids Res. **14,** 4051–4064.

**Li, S.-F., and Parish, R.W.** (1995). Isolation of two novel *myb*-like genes from *Arabidopsis* and studies on the DNA-binding properties of their products. Plant J. **8,** 963–972.

**Li, S.-F., Santini, J.M., Nicolaou, O., and Parish, R.W.** (1996). A novel *myb*-related gene from *Arabidopsis thaliana.* FEBS Lett. **379,** 117–121.

**Lohmer, S., Maddaloni, M., Motto, M., Salamini, F., and Thompson, R.D.** (1993). Translation of the mRNA of the maize transcriptional activator *Opaque-2* is inhibited by upstream open reading frames present in the leader sequence. Plant Cell **5,** 65–73.

**Majello, B., Kenyon, L.C., and Dalla-Favera, R.** (1986). Human *c-myb* protooncogene: Nucleotide sequence of cDNA and organization of the genomic locus. Proc. Natl. Acad. Sci. USA **83,** 9636–9640.

**Maxam, A.M., and Gilbert, W.** (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65,** 499–580.

**Menkens, A.E., Schindler, U., and Cashmore, A.R.** (1995). The G-box—A ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. Trends Biochem. Sci. **20,** 506–510.

**Nagy, F., Boutry, M., Hsu, M.-Y., Wong, M., and Chua, N.-H.** (1987). The 5′-proximal region of the wheat *Cab-1* gene contains a 268-bp enhancer-like sequence for phytochrome response. EMBO J. **6,** 2537–2542.

**Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y.** (1992). Solution structure of a DNA-binding unit of Myb: A helix-turn-helix–related motif with conserved tryptophans forming a hydrophobic core. Proc. Natl. Acad. Sci. USA **89,** 6428–6432.

**Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y.** (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell **79,** 639–648.

**Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D.** (1991). A *myb* gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell **67,** 483–493.

**Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A., and Saedler, H.** (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. **6,** 3553–3558.

**Pearson, W.R., and Lipman, D.J.** (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85,** 2444–2448.

**Pepper, A., Delaney, T., Washburn, T., Poole, D., and Chory, J.** (1994). *DET1*, a negative regulator of light-mediated development and gene expression in Arabidopsis, encodes a novel nuclear-localized protein. Cell **78,** 109–116.

**Peters, C.W.B., Sippel, A.E., Vingron, M., and Klempnauer, K.-H.** (1987). *Drosophila* and vertebrate *myb* proteins share two conserved regions, one of which functions as a DNA-binding domain. EMBO J. **6,** 3085–3090.

**Puente, P., Wei, N., and Deng, X.-W.** (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. EMBO J. **15,** 3732–3743.

**Quaedvlieg, N., Dockx, J., Keultjes, G., Kock, P., Wilmering, J., Weisbeek, P., and Smeekens, S.** (1996). Identification of a light-regulated *MYB* gene from an *Arabidopsis* transcription factor gene collection. Plant Mol. Biol. **32,** 987–993.

**Quail, P.H., Boylan, M.T., Parks, B.M., Short, T.W., Xu, Y., and Wagner, D.** (1995). Phytochromes: Photosensory perception and signal transduction. Science **268,** 675–680.

**Raikhel, N.V.** (1992). Nuclear targeting in plants. Plant Physiol. **100,** 1627–1632.

**Ruberti, I., Sessa, G., Lucchetti, S., and Morelli, G.** (1991). A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. EMBO J. **10,** 1787–1791.

**Scheffler, B., Franken, P., Schütt, E., Schrell, A., Saedler, H., and Wienand, U.** (1994). *Molecular analysis of C1 alleles in Zea mays* defines regions involved in the expression of this regulatory gene. Mol. Gen. Genet. **242,** 40–48.

**Shinozaki, K., Yamaguchi-Shinozaki, K., Urao, T., and Koizumi, M.** (1992). Nucleotide sequence of a gene from *Arabidopsis thaliana* encoding a *myb* homologue. Plant Mol. Biol. **19,** 493–499.

**Singh, H., Clerc, R.G., and LeBowitz, J.H.** (1989). Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. BioTechniques **7,** 252–261.

**Singh, K., Dennis, E.S., Ellis, J.G., Llewellyn, D.J., Tokuhisa, J.G., Wahleithner, J.A., and Peacock, W.J.** (1990). OCSBF-1, a maize Ocs enhancer binding factor: Isolation and expression during development. Plant Cell **2,** 891–903.

**Solano, R., Nieto, C., Avila, J., Canas, L., Diaz, I., and Paz-Ares, J.** (1995). Dual DNA binding specificity of a petal epidermis–specific MYB transcription factor (MYB.Ph3) from *Petunia hybrida.* EMBO J. **18,** 1773–1784.

**Sun, L., Doxsee, R.A., Harel, E., and Tobin, E.M.** (1993). CA-1, a novel phosphoprotein, interacts with the promoter of the *cab*140 gene in Arabidopsis and is undetectable in *det*1 mutant seedlings. Plant Cell **5,** 109–121.

**Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A.** (1993). Recognition of specific DNA sequences by the *c-myb* protooncogene product: Role of three repeat units in the DNA-binding domain. Proc. Natl. Acad. Sci. USA **90,** 9320–9324.

**Teakle, G.R., and Kay, S.A.** (1995). The GATA-binding protein CGF-1 is closely related to GT-1. Plant Mol. Biol. **29,** 1253–1266.

**Terzaghi, W.B., and Cashmore, A.R.** (1995). Light-regulated transcription. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46,** 445–474.

**Tice-Baldwin, K., Fink, G.F., and Arndt, K.T.** (1989). BAS1 has a Myb motif and activates *HIS4* transcription only in combination with BAS2. Science **246,** 931–935.

**Tobin, E.M., and Kehoe, D.M.** (1994). Phytochrome regulated gene expression. Semin. Cell Biol. **5,** 335–346.

**Urao, T., Yamaguchi-Shinozaki, K., Urao, S., and Shinozaki, K.** (1993). An Arabidopsis *myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5,** 1529–1539.

**Valvekens, D., Van Montagu, M., and Van Lijsebettens, M.** (1988). *Agrobacterium tumefaciens*–mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc. Natl. Acad. Sci. USA **85,** 5536–5540.

**Varagona, M.J., Schmidt, R.J., and Raikhel, N.V.** (1992). Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein Opaque-2. Plant Cell **4,** 1213–1227.

**von Arnim, A.G., and Deng, X.-W.** (1994). Light inactivation of Arabidopsis photomorphogenic repressor COP1 involves a cell-specific regulation of its nucleocytoplasmic partitioning. Cell **79,** 1035–1045.

**von Arnim, A., and Deng, X.-W.** (1996). Light control of seedling development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47,** 215–243.

**Wei, N., and Deng, X.-W.** (1996). The role of the *COP/DET/FUS* genes in light control of Arabidopsis seedling development. Plant Physiol. **112,** 871–878.

**Wright, K.L., Vilen, B.J., Itoh-Lindstrom, Y., Moore, T.L., Li, G., Criscitiello, M., Cogswell, P., Clarke, J.B., and Ting, J.P.-Y.** (1994). CCAAT box binding protein NF-κ facilitates in vivo recruitment of upstream DNA binding transcription factors. EMBO J. **13,** 4042–4053.

**Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M., and Goodman, H.M.** (1996). A physical map of chromosome 2 of *Arabidopsis thaliana.* Genome Res. **6,** 19–25.