Volume 14 Number 10 1986 Nucleic Acids Research

Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*

Leslie S.Leutwiler, Elliot M.Meyerowitz+ and Elaine M.Tobin*

Department of Biology, University of California, Los Angeles, CA 90024 and +Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

Received 14 February 1986; Accepted 21 April 1986

ABSTRACT

The genome of Arabidopsis thaliana is exceedingly small, in part because it lacks the large middle repetitive DNA component characteristic of other plants. In this paper we have characterized a member of the low copy DNA component: the gene family for the light-harvesting chlorophyll a/b-protein. This gene family is unusual in that it contains far fewer members than the 7-16 coding sequences for this protein found in other plants. We used cross-hybridization with a Lemna gene encoding a light-harvesting chlorophyll a/b-protein to isolate 3 genes from Arabidopsis, all of which are clustered on an 11-kb genomic clone. Southern blot analysis suggests that there is a fourth related gene in Arabidopsis. Sequence analysis of the three genes demonstrates that a) within the translated region the nucleic acid sequence homology is 96%, b) the deduced amino acid sequence of the mature proteins is identical for the three genes, and c) two of the genes have a high degree of sequence homology in both their 5' and 3' immediate flanking regions. The genes have regulatory sequences typical of eukaryotic genes upstream of the translation start sites. However, not all of these genes are equally expressed in plants grown under normal light-dark conditions.

INTRODUCTION

The major protein component of chloroplast thylakoid membranes is the light-harvesting chlorophyll a/b-protein complex (LHC II) (1) which is mainly associated with photosystem II. The chlorophyll binding proteins (LHCP's) of this complex are the products of nuclear genes and are synthesized in the cytoplasm as precursor polypeptides (30-32 kD). Following posttranslational transport into the chloroplast these polypeptides are cleaved to their mature form; bind noncovalently chlorophyll a, chlorophyll b and carotenoid molecules; and become embedded within the thylakoid membrane where they function to harvest light energy for photosynthesis. Neither the sequence of the above events following entry into the chloroplast nor the number and arrangement of the polypeptides within the complex is known. However, when the green LHC II complex is isolated on nondenaturing gels and then subjected to electrophoresis on denaturing gels, 2-5 polypeptide bands can be resolved within the region of 24-28 kD. The genes which code for the major polypeptide of this

complex are present in all plant species examined as a multigene family. For instance, petunia contains at least 16 (2), pea at least 8 (3), Lemna gibba 10-12 (4) and wheat at least 7 (5) coding sequences for the LHCP. Because multiple polypeptide bands can be resolved from the complex and because the multiple gene members can be organized into separate subfamilies coding for electrophoretically distinct precursor polypeptides (2,6), it is possible that these separate gene families might produce functionally distinct polypeptides (cf. 7). It is also possible that post-translational events have a role in producing the electrophoretically variant forms found in the thylakoids (cf. 8).

The crucifer Arabidopsis thaliana is an unusual higher plant in that it contains an extremely small genome (9-12). The small size is due mainly to the absence of the large middle repetitive DNA component (10) and the reduced size of the single copy component (11) relative to other higher plant genomes. In this paper we characterize the gene family for the light-harvesting chlorophyll a/b-binding protein (LHCP). In contrast to the situation in other plants, Arabidopsis apparently contains far fewer coding sequences for the LHCP. We have isolated and sequenced three LHCP genes and find that the amino acid sequence deduced from the DNA sequence is identical for all three mature polypeptides. This finding raises the possibility that if the variant polypeptides encoded by the different subfamilies of other plants have distinct functions, they may not be essential.

MATERIALS AND METHODS

Plants and DNA isolation

Arabidopsis thaliana strain Columbia was grown and the DNA isolated as previously reported (10).

Construction and screening of a genomic library

An Arabidopsis genomic library was constructed in the lambda phage vector Sep 6 (13) according to the procedure described in (14). The in vitro constructed phage library, which contained 120,000 recombinant phage or 20 genome equivalents, was amplified according to (15). Recombinant phage were propagated on E. coli K802 and purified by standard methods (15,16). Five genomic equivalents of the amplified phage library were screened by the method of (17) with a genomic subclone of Lemna gibba, pAB19/H5c, which contains a LHCP coding sequence (10). Phage DNA was extracted by the rapid formamide method "A" of (18).

Restriction, gel electrophoresis and filter binding of DNA

DNA was digested with various restriction endonucleases and subjected to electrophoresis in 1X TBE (0.09M Trizma-base, 3mM Na EDTA, and 0.09M boric acid) on 0.5-2% horizontal agarose gels for restriction endonuclease mapping and on 0.6% agarose gels for genomic digests. DNA was transferred from agarose gels to nitrocellulose and hybridized as described by (14). Filters were washed at room temperature four times as described in (14).

DNA labeling and hybridization

DNA was radioactively labeled by nick translation as described by (18). Hybridizations were carried out according to the method of Pruitt and Meyerowitz (12).

Bal-31 deletion cloning and dideoxy sequencing

The 1.65-kb Eco RI fragment containing a LHCP coding sequence (AB165) was subcloned into pUC8 (pAB165). Fifteen micrograms were linearized with SalI, digested with nuclease Bal-31 as described in (7), and the deleted insert was cleaved from the deleted vector by digestion with EcoRI. The deleted fragments were ligated into M13mp8 RF DNA (30 ng vector and 100 ng insert) and transfected into E. coli strain JM103; and single-stranded phage DNA was prepared from the resulting colorless plaques for insert identification, sizing and sequencing.

The single-stranded DNA was sequenced by the chain-termination method of Sanger et al. (19) with [^{35}S]dATP (800-1500 Ci/mmol; New England Nuclear Corp.). The complete sequence of the noncoding strand and a partial sequence of the coding strand of AB165 was obtained from an overlapping set of sequentially deleted Bal-31 fragments that were subcloned into M13mp8 or M13mp9. Most of the remaining sequence of the coding strand of AB165 was obtained from various restriction fragments subcloned in mp9 and mp10. The sequence for a single strand of AB180 and AB140 was obtained also by subcloning various restriction fragments into mp8, mp9 or mp10. Electrophoresis was on 8% acrylamide gels which were transferred to Whatman 3MM chromatography paper, dried and exposed to X-Omat AR X-ray film (Kodak) for 7 hr at -70°C.

Chemical sequencing

The sequence of several regions was verified by the Maxam and Gilbert (20) chemical sequencing method.

S1 nuclease analysis

In order to map the 5' termini of the LHCP genes an end-labeled probe was constructed by digesting 15 ug of pAB180 and labeling at the BamHI site (158 nucleotides downstream from the translation start site) using T4 polynucleo-

tide kinase (BRL) and the procedure of (20). The labeled DNA was digested with EcoRI, followed by electrophoresis in 1% low gelling temperature agarose (Seakem) to separate the 1.1-kb EcoRI-BamHI fragment. The fragment was cut out of the gel, the agarose melted at 65°C, phenol extracted three times and the DNA ethanol precipitated.

RNA isolation

Total RNA was extracted from 15g of whole plants by grinding with 7.5g glass beads in LN_2 in a mortar and pestle and allowing the powder to thaw in the presence of 30 ml "Kirby Reagent" (21). The solution was homogenized with 2 10-sec bursts of the Polytron and phenol extracted 3 times. The aqueous fractions were combined and ethanol precipitated. The pellet was resuspended in 5 ml of dH_2O; 5 ml of 4M LiCl was added, and the suspension was placed at 4°C overnight to selectively precipitate the RNA. Following centrifugation at 7000 rpm for 20 min in a Beckman SS-34 rotor, the pellet was resuspended in 5 ml 2M LiCl, vortexed and centrifuged again. The pellet was washed in 70% ethanol, air dried and resuspended in 100 µl H_2O.

Hybridization and S1 nuclease digestion

Total RNA (25 µg) from light-grown Arabidopsis plants was added to 50 ng digested probe and ethanol precipitated. The DNA-RNA pellet was resuspended in 30 µl hybridization buffer according to the procedure of (13). The DNA was denatured at 85°C 10 min then transferred to a 56°C water bath and incubated for 3 hr. Either 100 U/ml or 1000 U/ml S1 nuclease (BRL) in a total volume of 0.3 ml ice cold nuclease S1 buffer (13) was added to each tube, mixed and incubated at 36°C 0.5 hr. The reaction was terminated, the RNA-DNA hybrid precipitated with isopropanol, resuspended in loading buffer, denatured and submitted to electrophoresis on an 8% acrylamide gel. The duplex from the 5' untranslated region was run alongside the Maxam and Gilbert sequencing reactions for the same end-labeled fragment.

RESULTS

Isolation of a genomic clone containing 3 coding sequences for the LHCP

Total Arabidopsis DNA was partially digested with EcoRI and cloned into the lambda vector Sep 6 (13). Five genome equivalents of phage were screened with pAB19/H5c, a genomic subclone from Lemna gibba containing the coding region for a LHCP (7). Nine hybridizing phage were selected and their DNA restricted. All contained inserts of the same 11-kb DNA fragment with 6 EcoRI sites. The inserts were in both orientations relative to the lambda vector and thus derived from at least two independent isolates. One phage contained

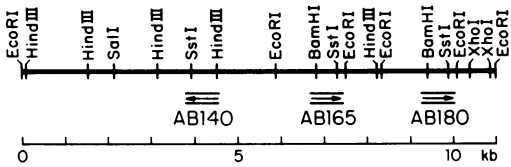

Figure 1. Restriction map of λbAt1005, an 11-kb genomic clone containing a cluster of three coding sequences for the LHCP. The leftmost sequence is largely contained on a 1.4-kb HindIII fragment (AB140); the remaining two coding sequences are located on 1.65-kb and 1.80-kb EcoRI fragments (AB165 and AB180). The arrows indicate the direction of transcription.

an additional 4-kb EcoRI fragment. Figure 1A shows the restriction endonuclease cleavage site map of the 11-kb DNA fragment (λbAt1005) which contains a cluster of three LHCP coding sequences. The restriction maps of the two rightmost sequences (on 1.65-kb and 1.80-kb EcoRI fragments) are identical (AB165 and AB180 respectively); the leftmost sequence (on a 6.0-kb EcoRI fragment) differs both in restriction map and transcription orientation.

The gene family for the major LHCP consists of three members

The three coding sequences were subcloned: the 1.65-kb and 1.80-kb EcoRI fragments contained the two sequences with identical restriction maps (pAB165 and pAB180); a 1.4-kb HindIII fragment contained most of the third coding sequence (pAB140). Since other plants in which the major LHCP has been characterized contain large multigene families encoding this protein, it was of interest to determine the number of possible genes in a plant which contains a genome as small as Arabidopsis. Total DNA was digested to completion with EcoRI or HindIII, subjected to gel electrophoresis, blotted to nitrocellulose and probed with nick-translated pAB165 DNA. Figure 2 shows that the three homologous bands in each digest are those predicted by the restriction map of λbAt1005. Therefore, it appears there are only three closely related coding sequences for the LHCP. In addition there seems to be a fourth more distantly related coding sequence, as under the hybridization conditions used, the HindIII digest of Fig. 2B shows a faint band approximately 2.0-kb in length.

DNA sequence studies show the three genes are highly homologous

The sequence of the noncoding strand of AB165 was obtained by digesting pAB165 with Bal-31 exonuclease to yield an overlapping set of sequentially deleted fragments from the 3' end which were subcloned into M13mp8 and sequenced by the dideoxy chain termination method of (19). The sequence of AB140, AB180 and the coding strand of AB165 were obtained by subcloning various restriction fragments into the M13 phage vectors mp8, mp9 or mp10.

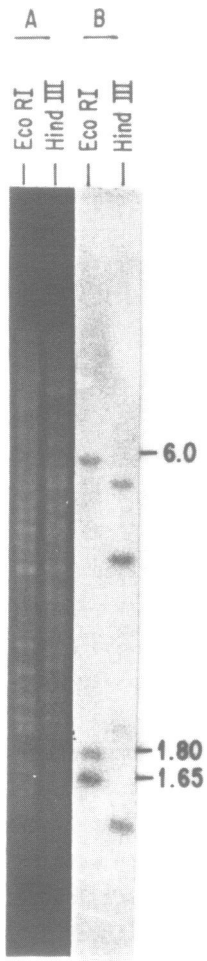

Figure 2. Southern blot analysis of total Arabidopsis DNA. Arabidopsis DNA was digested to completion with the restriction enzymes EcoRI (A) or HindIII (B), fractionated on a 0.6% agarose gel, transferred to nitrocellulose, and hybridized with nick-translated pAB165. The sizes of the three EcoRI bands containing each coding sequence are given in kb.

The sequence of several regions was confirmed by the chemical sequencing method of (20). The complete nucleotide sequence along with the deduced amino acid sequence is presented in Fig. 3. All three genes contain open reading frames starting with methionine and containing 267 amino acids up to the translation termination codon, TGA. At the amino acid level they show 100% homology in the mature coding sequence; the only difference is the substitution of an asparagine for lysine in the transit sequence of AB140. At the nucleotide level the homology is 96% for all three genes within the translated region. Additionally AB165 and AB180 show extreme homology in the noncoding region: the 50-bp immediately downstream from the translation stop site to the EcoRI site show 90% homology; the region up through the presumptive TATA

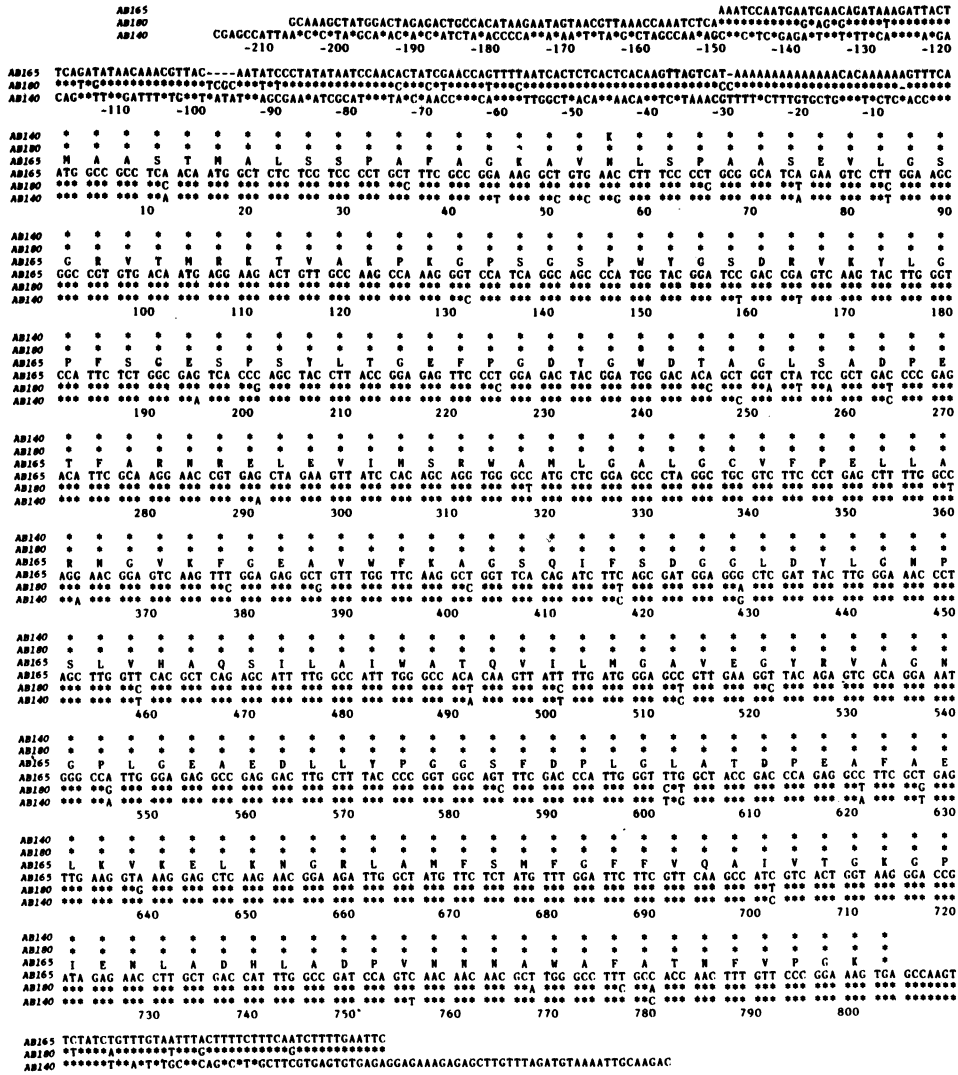

Figure 3. Nucleotide and deduced amino acid sequence of three Arabidopsis genes coding for the LHCP. The complete nucleotide sequence for AB165 is shown. Below this sequence are shown the nucleotide sequences for AB180 and AB140. The single-letter code for each amino acid of AB165 is given directly above the nucleotide sequence for AB165. Above the AB165 amino acid sequence are the amino acid sequences for AB180 and AB140. Asterisks denote identity.

box is 91% homologous; and that up through the presumptive CAAT box is 86% homologous. The homology continues upstream from the CAAT box. While AB140 is equally homologous to AB165 and AB180 in the translated region, its

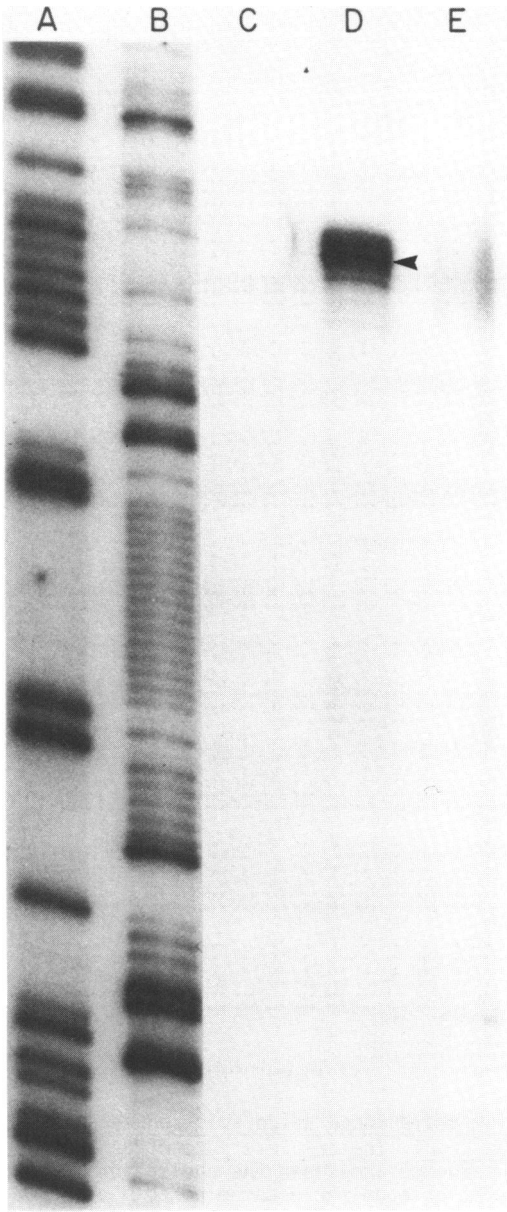

Figure 4. S1 nuclease analysis of AB180. A 1.1-kb EcoRI-BamHI fragment (50 ng) from AB180 was end labeled at the BamHI site, hybridized for 3 hr to 25 µg total Arabidopsis RNA (lanes D and E) or 25 g yeast tRNA (lane C). The DNA-RNA hybrids were digested with S1 nuclease -- either 100 U/ml (lanes C and D) or 1000 U/ml (lane E). The Maxam and Gilbert G (lane A) and C+T (lane B)

reactions were carried out with the same 5'-end-labeled fragment. Electrophoresis was with 8% acrylamide sequencing gels. The arrow denotes the major band of 210 bp.

sequence diverges almost immediately upstream from the translation start site as well as about 20-bp downstream from the translation stop site.

Expression of the LHCP genes

S1 nuclease protection experiments were carried out to determine if all three genes are expressed. Since AB140 differs from AB165 and AB180 upstream from the translation start site, a S1 nuclease protection experiment was devised using a 5' untranslated region probe. A 1.1-kb EcoRI-BamHI fragment derived from the 1.80-kb EcoRI fragment was end labeled at the BamHI site, denatured, hybridized to total RNA, and the hybrid treated with S1 nuclease. Figure 4 shows several predominant protected bands of which the major one is 210 bp. Since the probe included 158 bp of translated sequence the 5' untranslated region is 52 bp. The expression of AB165 and AB180 cannot be distinguished, since their 5' untranslated regions in this area show greater than 94% homology. However, AB140 can be distinguished from the other two, as the sequence diverges sharply 3 bp upstream from the translation start site. The 161 bp protected band, which would be expected if AB140 is expressed, is not evident.

DISCUSSION

The multigene family for the LHCP in *Arabidopsis* is much smaller than in other plants. It consists of only three strongly cross-hybridizing members and possibly a fourth less related gene as compared with the 7-16 members found in other plants (2-5); in addition the three *Arabidopsis* genes are found clustered in a region of about 6 kb. Other genes, including a seed storage protein (R.E. Pruitt, personal communication), the alcohol dehydrogenase (22), and the SSU of RuBisCo (J. Leemans and M. Timko, personal communication) which are also single copy genes or members of small clustered gene families in *Arabidopsis*, exist in larger gene families in other plants. Therefore, the small size of the *Arabidopsis* genome relative to the genome size in other angiosperms may be partly due to a relative reduction in the size of multigene families. This smaller size of gene families should simplify the study of the expression and regulation of individual genes.

The data presented here show that the three closely related genes produce identical mature proteins for the LHCP of the thylakoid membranes in *Arabidopsis*. These three genes show such strong homology -- 99.9% amino acid

sequence and 96% nucleic acid sequence conservation in their coding regions and 83% nucleic acid sequence conservation in their immediate 3' noncoding regions -- that they cannot be divided into the distinct gene subfamilies which specify unique polypeptides for the precursors of the LHCP as has been seen in petunia (2). Two of these genes are so closely related in nucleic acid sequence of their noncoding regions -- 86% homology in the 5' region up through the CAAT box and 90% homology in the immediate 3' region following the translation stop site -- that it is difficult to distinguish them by conventional methods of S1 analysis.

Analysis of flanking sequences

Upstream of the transcription start site all three genes contain sequences characteristic of many eukaryotic promoters. A TATA sequence, TATATAAT, is located 79 nucleotides upstream of the translation start site in AB165 and AB180. The sequence TATTATATATA occurs 92 bp upstream of the translation start site in AB140. The sequence CCAAT occurs 51, 55, and 51 nucleotides 5' of the putative TATA sequences in AB165, AB180 and AB140 respectively. An unusual feature of the 5' upstream region of AB165 and AB180 is the poly A region of 23 and 22 nucleotides which is interrupted by only two cytosines. This poly A tract, which is located seven nucleotides upstream of the translation start site, is absent in the AB140 gene.

Based on S1 analysis the transcription start site for AB180 is 52 nucleotides upstream from the translation start site and 27 nucleotides downstream of the TATA box near the beginning of the repeated sequence TCACTC. This repeated sequence is also present in the same location in AB165. Five petunia genomic clones contain sequences about 50 nucleotides upstream of the translation start site which are similar to the consensus transcription start sequence, TCAT (23). While this sequence is not present in the 5' upstream region of AB140, the sequence ACCAC is present 65 nucleotides upstream of the translation start site and 23 nucleotides downstream of the TATA sequence. This sequence also occurs 61 nucleotides 5' of the translation start site in a Lemna LHCP genomic clone (8); and a pea and a wheat LHCP gene have transcription start sites 68 and 70 nucleotides 5' of the translation start site at the sequence ACCAT (5).

In Arabidopsis, in addition to the major band from S1 digestion, faint bands can be seen between the AB180 transcription start and translation start sites. It seems likely that these bands are caused by the misalignment of AB165 RNA with the AB180 probe, since the only differences in nucleotide composition between these two genes in this region occur on either side of the

poly A region: AB165 contains a substitution and adjacent deletion of nucleotides immediately upstream and an addition immediately downstream of the poly A tract.

If AB140 RNA were present, a second band from S1 nuclease digestion would be expected three nucleotides 5' to the translation start site where the sequence diverges from that of AB165 and AB180. Since this band is not evident, it appears that in light-grown plants either or both AB165 and AB180 are heavily expressed whereas AB140 is either transcribed at very low levels or not at all. That it may be expressed at low level is indicated by the work of G.An (personal communication). He finds that the Arabidopsis AB180 promoter is 5 to 10 times more active than the AB165 promoter, which in turn is more active than the AB140 promoter in producing chloramphenicol transacetylase in transformed tobacco suspension culture cells.

In contrast to the 5' untranslated region, in which the sequence of AB140 diverges from AB165 and AB180 almost immediately upstream of the translation start site, the region immediately 3' of the protein-coding sequence shows strong homology for about 20 nucleotides. Another feature of the 3' region of the three genes in the 50 bp downstream of the translation stop site is the high percentage of thymine: 52%, 54% and 48% respectively in AB165, AB180 and AB140. The sequence of AB165 and AB180 further downstream is unknown since the subclones were contained on EcoRI fragments which contain only 50-bp 3' of the translated region. However, the 3' region for AB140 shows the sequence TTGTTT 64 bases downstream of the translation stop site. This sequence has been recognized as a conserved sequence in the 3' untranslated regions of four petunia cDNA clones (2) and a wheat genomic clone (5) for the LHCP.

Analysis of amino acid composition

The predicted amino acid sequence for the precursor polypeptide is 267 residues with a molecular weight of 28,195. We do not know the exact processing site; however, observations by (24) combined with studies on the sequence of genomic clones from pea (25), petunia (23), Lemna (7,8) and wheat (5) indicate that either methionine (amino acid 35) or arginine (amino acid 36) occurs at the N-terminus of the mature LHCP. If this is also true for Arabidopsis, then a 34 or 35 amino acid transit sequence would be cleaved to yield a mature protein of 233 or 232 residues with a molecular weight of 24,966 or 24,704 respectively.

In some plant species a second LHCP precursor (pLHCP) with a lower M_r has been detected. This putative pLHCP shows immunological relatedness to the higher M_r pLHCP at the protein level but no homology at the nucleotide

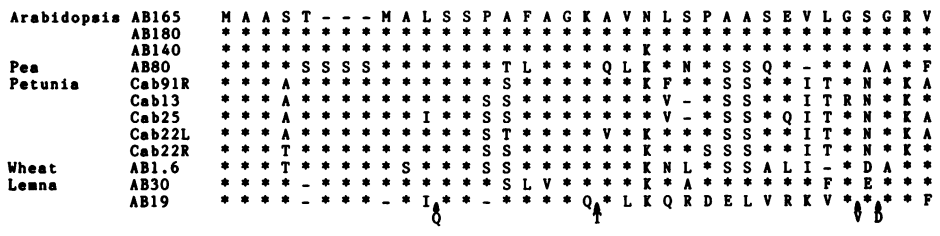

Figure 5. Comparison of the amino acid sequence for 3 Arabidopsis, 1 pea (24), 5 petunia (22), 1 wheat (5) and 2 Lemna (7,8) transit sequences for the LHCP. Asterisks denote identity with the Arabidopsis AB165 gene.

level. For instance, in pea LHC II contains at least two immunologically related polypeptides of 28,000 and 27,000 daltons (polypeptide 15 and polypeptide 16) which are synthesized in cell-free translation systems as larger forms of 33,000 and 32,000 daltons (6,26). Whereas pea poly A+ RNA translated in vitro yields the the two distinct precursor polypeptides of LHC II, the pea cDNA clone, pAB96 (27), and pea genomic clone, AB80 (25) selectively hybridize mRNAs which yield only the larger precursor which has been proposed to give rise to the larger polypeptide (polypeptide 15; 27). In petunia (2) poly A+ RNA translation products include at least two major and one minor pLHCPs which are immunoprecipitated by antisera raised against either the major LHCP of maize or the LHC II of spinach. Although at least three different precursor polypeptides can be distinguished following hybrid release translation of the gene-specific RNAs, none of these clones hybridizes to mRNA that yields the lowest molecular weight LHCP precursor in in vitro translation. It is possible that the faint 2.0 kb HindIII band seen in Fig. 2B could encode a more distantly related LHCP which might be analogous to the lower molecular weight precursor polypeptides of pea and petunia.

Analysis of transit sequence

The transit sequences for several nuclear encoded chloroplast proteins have been published. In all published sequences the transit peptide shows greater nucleotide and amino acid sequence divergence than that which specifies the mature protein. This is also true for the LHCP of Arabidopsis: the average nucleotide divergence within the transit peptide coding regions of the three LHCP genes is 6% and within the mature peptide coding regions is 3.6%. The predicted amino acid sequence shows a single difference between the AB140 transit peptide and the identical AB165 and AB180 transit peptides. The mature proteins coded by the three Arabidopsis genes are identical. Whereas previously nobody had noted any amino acid homology between the transit

peptides of the LHCP and the SSU of RuBisCo within or between species, the homology between transit sequences for the LHCP of different species is striking. Fig. 5 compares the amino acid sequence of the transit peptide from 12 LHCP genes. The homology in the first two-thirds of the sequence is particularly noteworthy. This homology between species for the LHCP transit peptide may be involved in determining the fate of this protein following translocation into the chloroplast. For instance, it is conceivable that these sequences are necessary for such events as the processing of the precursor to the mature protein, incorporation of the protein into the thylakoid membrane, binding of the pigment molecules, and/or assembly into the LHC II complex.

It should be noted that Karlin-Neumann and Tobin (28) have identified three major blocks of homology between all transit peptides so far sequenced (LHCP, SSU and ferredoxin). The blocks of homology, which are at the beginning, middle and end of the transit peptide, are separated by the longer lengths of unshared sequence which show homology within the individual proteins. They suggest that the three blocks of homology are necessary for the targeting and uptake of these precursors by the chloroplast. The information for the processing of the precursor peptide to the mature form may also be encoded within these homology blocks.

ACKNOWLEDGEMENTS

LSL had a National Institutes of Health postdoctoral fellowship. This work was supported by grants from NIH 23167 (EMT), USDA 83-CRCR-1-1319 (EMT) and NSF PCM-8408504 (EMM).

*To whom correspondence should be addressed

REFERENCES

1. Thornber, J.P. (1975) Ann. Rev. Plant Physiol. 67, 127-158.
2. Dunsmuir, P., Smith, S.M. and Bedbrook, J. (1983) J. Mol. Appl. Genet. 2, 285-300.
3. Coruzzi, G., Broglie R., Cashmore, A. and Chua, N.-H. (1983) J. Biol. Chem. 258, 1399-1402.
4. Tobin, E.M., Wimpee, C.F., Silverthorne, J., Stiekema, W.J., Neumann, G.A. and Thornber, J.P. (1984) in Biosynthesis of the Photosynthetic Apparatus, J.P. Thornber, L.A. Staehelin, R. Hallick, pp. 325-334. Alan Liss, New York.
5. Lamppa, G.K., Morelli, G. and Chua, N.-H. (1985) Mol. Cell. Biol. 5, 1370-1378.
6. Schmidt, G.W., Bartlett, S.G., Grossman, A.R., Cashmore, A.R. and Chua, N.-H. (1981) J. Cell Biol. 91, 468-478.

7. Karlin-Neumann, G.A., Kohorn, B.D., Thornber, J.P. and Tobin, E.M. (1985) J. Mol. Appl. Genet. 3, 45-61.
8. Kohorn, B.D., Harel, E., Chitnis, P.R., Thornber, J.P. and Tobin, E.M. (1986) J. Cell Biol. 102, 972-981.
9. Bennett, M.D. and Smith, J.B. (1976) Philos. Trans. R. Soc. London Ser. B 274, 227-274.
10. Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. (1984) Mol. Gen. Genet. 194, 15-23.
11. Meyerowitz, E.M. and Pruitt, R.E. (1985) Science 229, 1214-1218.
12. Pruitt, R.E. and Meyerowitz, E.M. (1986) J. Mol. Biol. In press.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. Meyerowitz, E.M. and Martin, C.H. (1984) J. Mol. Evol. 20, 251-264.
15. Maniatis, T., Hardison, R.C., Lacy, E., Lauer J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
16. Meyerowitz, E.M. and Hogness, D.S. (1982) Cell 28, 165-176.
17. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
18. Davis, R.W., Botstein, D. and Roth, J.R. (1980) in Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
20. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
21. Loenig, V. (1969) Biochem. J. 113, 131-138.
22. Chang, C. and Meyerowitz, E.M. (1986) Proc. Natl. Acad. Sci. USA 83, 1408-1412.
23. Dunsmuir, P. (1985) Nucl. Acids Res. 13, 2503-2518.
24. Mullet, J.E. (1983) J. Biol. Chem. 258:9941-9948.
25. Cashmore, A.R. (1984) Proc. Natl. Acad. Sci. USA 81, 2960-2964.
26. Schmidt, G.W., Bartlett, S., Grossman, A.R., Cashmore, A.R. and Chua, N.-H. (1980) in Genome Organization and Expression in Plants. C.J. Leaver Ed., pp. 337-351. Plenum Publishing Co., New York.
27. Broglie, R., Bellemare, G., Bartlett, S.G., Chua, N.-H. and Cashmore, A.R. (1981) Proc. Natl. Acad. Sci. USA 78, 7304-7308.
28. Karlin-Neumann, G.A. and Tobin, E.M. (1986) EMBO J. 5, 9-13.