Volume 280, number 2, 229–234 FEBS 09498 March 1991

ADONIS 001457939100624O6

Nucleotide sequence and chromosomal location of Cab11 and Cab12, the genes for the fourth polypeptide of the photosystem I light-harvesting antenna (LHCI)

Egbert Schwartz[1], Dingren Shen[2], Ruedi Aebersold[3], J. Mitchell McGrath[1], Eran Pichersky[1] and Beverley R. Green[2]

[1]*Biology Department, University of Michigan, Ann Arbor, MI48109, USA and* [2]*Botany Dept. and* [3]*Biomedical Research Centre and Dept. of Biochemistry, University of B.C., Vancouver, B.C., Canada*

(Received 19 December 1990)

Tryptic peptide sequences from the 22 kDa polypeptide of tomato LHCI were used to construct a probe for gene cloning. The two genes cloned, *cab*11 and *cab*12, encode proteins of 251 and 250 residues that are 88% identical in overall amino acid sequence and 93% identical in the deduced mature protein. Each gene is present in a single copy per haploid genome; *cab*11 on chromosome 3 and *cab*12 on chromosome 6, and each has 2 introns located in similar positions to introns in other members of the Chl *a/b*-binding (CAB) protein gene family. Comparison of the amino acid sequences of LHCI, LHCII, CP29 and CP24 polypeptides confirms that all CABs share two regions of very high similarity which include the first and third transmembrane helices and the stroma-exposed sequences preceding them. However, near the N-terminus and between the conserved regions, the LHCI polypeptides have sequence motifs which appear to be PSI-specific.

Photosystem I light-harvesting complex; Chlorophyll *a/b*-(CAB) protein; Gene, *cab*11; Gene, *cab*12; *Lycopersicon*

1. INTRODUCTION

The chlorophyll *a/b*-binding (CAB) light harvesting antennas of higher plants comprise a diverse group of chlorophyll-protein complexes, one associated with Photosystem I (LHCI) and three with Photosystem II (LHCII, CP29 and CP24) [1–3]. Their polypeptides are immunologically related [4–6], and comparison of sequences obtained to date by gene cloning indicate that they are part of a large gene family [7,8].

Genomic and cDNA clones encoding three of the four LHCI polypeptides have been isolated and characterized in tomato [9–12]. We report here the isolation and characterization of cDNA clones for 2 tomato genes, *cab*11 and *cab*12, which encode the fourth LHCI protein (Type IV). Comparison of the deduced amino acid sequences for all four of the LHCI polypeptides shows that while they are structurally related to the CAB polypeptides of PSII in having the two highly conserved regions characteristic of all Chl *a/b*-binding polypeptides, outside of these regions they have some unique sequence motifs which may be involved in their association with Photosystem I.

Correspondence address: B.R. Green, Botany Department, University of B.C., B.C., Canada

Abbreviations: Chl, chlorophyll; PSI, Photosystem I; LHCI, Chl *a/b* light-harvesting antenna of PSI; CPI, PSI reaction centre complex; CAB, Chl *a/b* binding

2. METHODS

2.1 *Isolation and protein sequencing of tomato LHCI polypeptides*

Lycopersicon esculentum var. Best of All, was grown in a mixture of soil and vermiculite under natural illumination in a greenhouse. PSI particles (PSI-200) were made according to [13] with a Triton/*Chl* ratio of 6.25. LHCI was prepared by solubilizing PSI-200 in 1% dodecyl maltoside, 130 mM NaCl, 10 mM Tricine, pH 8.0, and fractionating on a 0.2–1.0 M sucrose gradient [14].

Polypeptides were separated by electrophoresis for 22 h at 17 mA on 22 cm long 14% polyacrylamide gels containing 0.8 M Tris, pH 8.8, and 0.1% SDS, at 4° [15] and electrotransferred onto nitrocellulose membranes for tryptic hydrolysis or immunostaining, or onto polyvinylidene difluoride (Immobilon P) membranes for N-terminal sequencing. LHCI was identified by Western blotting using antibodies raised against barley CP1a (CP1+LHCI) [5]. Subunit II (psaD gene product) was identified using antisera donated by Drs R. Malkin and N. Nelson.

Protein bands on nitrocellulose membranes were cut out, digested on the support with trypsin [16] and separated by narrow-bore reversed phase HPLC on a Waters peptide analyzer equipped with a Vydac C-4 column. Individual peptides were collected manually and sequenced using standard pulsed-liquid-phase or solid-phase sequencing procedures [17]. For N-terminal sequencing, Coomassie-stained bands [18] were excised and sequenced in an Applied Biosystems Model 477 protein sequenator.

2.2 *Gene cloning and sequencing*

Oligonucleotides were prepared using sequences derived from 2 of the tryptic peptides: for PGSVNQDPIF, the sense 20-mer 5′-GT(A/T/G/C)AA(C/T)CA(G/A)GA(C/T)CC(A/T/G/C)AT(A/T-

/C)TT; and for GYPGXIFNPLNF, the antisense 23-mer 5'-GC(A/G)AA(A/G)TT(A/G/C/T)A(A/G)A(A/G/C/T)CG(A/G)TT-(A/G)AA(A/T/G)AT . Plasmid DNA (0.1 μg) from a tomato leaf cDNA library [10] was used as the source for DNA amplification with *Taq* DNA polymerase (35 cycles at 92°, 37°, and 72° for 2, 3, and 2 min respectively). After amplification, total amplified DNA was passed through a Sephadex G-50 column, then labelled with ^{32}P using the random-primer method. The labelled DNA was used as a probe to screen tomato genomic and leaf cDNA libraries in the lambda-phage vectors Charon 35 and Charon 16, respectively [10,19]. Probe labelling, library screening and the isolation and characterization of clones was done as previously described [10,20]. DNA was sequenced by chemical [21] and enzymatic [22] methods.

3. RESULTS

The resolution of tomato PSI-200 and LHCI polypeptides on a long separating gel is shown in Fig. 1. The LHCI polypeptides were clearly separated from the Subunit II (psaD) polypeptide, which was identified by Western blotting using two different antisera (data not shown). The LHCI polypeptides were immunostained (lane a) with antisera raised against the barley complex CPIa (CPI + LHCI) [5]. The lowest LHCI polypeptide (22 kDa) was the most strongly stained band in all preparations. A similar predominance of one polypep-

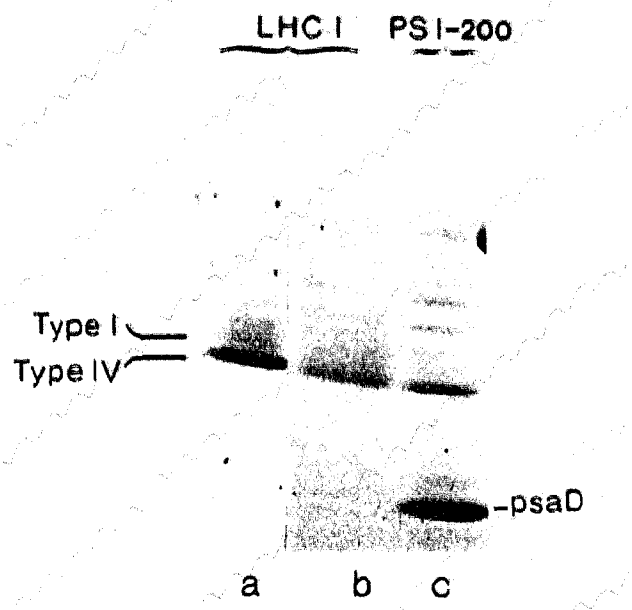

Fig. 1. Polypeptides of tomato LHCI (a,b) and PSI-200 (c). Lane a, immunodecorated with anti-CPIa, which recognizes the LHCI polypeptides [5]; lanes b, c, Coomassie stained. The psaD polypeptide was identified by immunoblotting with two independent antisera (data not shown).

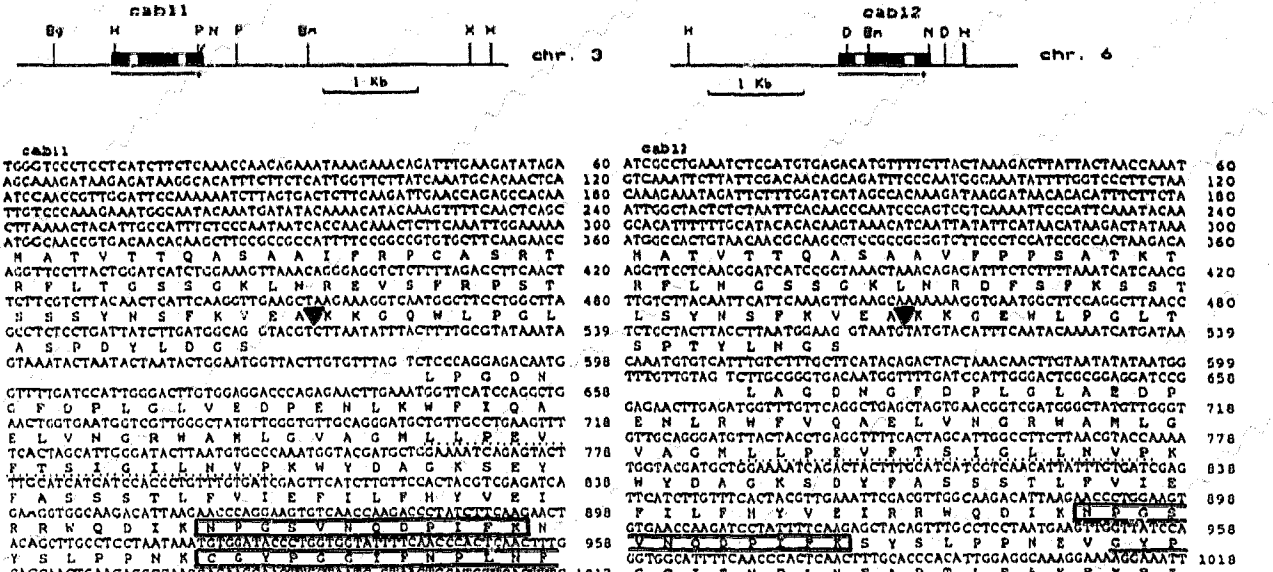

Fig. 2. Structure of *cab*11 and *cab*12. (Above) Restriction maps. Arrows indicate direction of transcription. Cleavage sites: Bg, *Bgl* II; Bm, *Bam* HI; D, *Dra*I; H, *Hind*III; N, *Nco*I; P, *Pst*I; and X, *Xba*I. (Below) Nucleotide and deduced amino acid sequences. Intron positions were determined by comparison with cDNA clones. Exact matches to 22 kDa tryptic peptides XPGSVNQDPIFK, XGYPGXIFNPLNFA and XXXXNLLQXLSDP are boxed (X denotes amino acid not unambiguously identified); similar but not identical sequences are underlined in *cab*12. Arrow-head denotes the most likely transit peptide cleavage site [28]. The sequence corresponding to the DCCD-labeled pea 20 kDa polypeptide of [32] is underlined with dots.

tide has been observed in LHCI from some other higher plants [23,24] and the green alga *Chlamydomonas rheinhardtii* [25]. The antibody did not detect the minor polypeptides of 17 and 11 kDa reported in maize and barley LHCI [26,27].

The amino acid sequences of three tryptic peptides from the 22 kDa LHCI band (Type IV, Fig. 1) were unique but clearly related to sequences in the Type II and Type III tomato LHCI proteins [11,12]. Two of them appeared to belong to adjacent regions by comparison with other CAB sequences and were therefore used to prepare synthetic oligonucleotides for amplification of the intervening gene segment by the polymerase chain reaction. The sequence of one tryptic peptide from the band labelled Type I (Fig. 1), (K)GIFPN, uniquely identified it as the C-terminus of *cab*6a, the first LHCI gene isolated [9].

The amplified DNA segment was used as a probe to isolate two different cDNA clones and the corresponding genomic clones (Fig. 2). We designated the two genes *cab*11 and *cab*12; genes *cab*1 through *cab*10 encode other CAB polypeptides of PSI and PSII [7,8,19]. The gene *cab*11 encodes a protein of 251 amino acid residues, and *cab*12 encodes a very similar polypeptide of 250 residues (88% overall sequence identity).

All 3 tryptic peptide sequences derived from the 22 kDa LHCI band were identical to sequences encoded by *cab*11 (boxed), and two were identical to *cab*12-encoded sequences. Although the sequence XGYPGXIFNPLNFA matched a *cab*12 sequence it should not have been produced by tryptic cleavage of a *cab*12-encoded protein because the preceeding Lys has been replaced by Glu. The third sequence, NLLQXLSDP, differs in one residue (L instead of I) from the corresponding *cab*12 peptide. This suggests that the *cab*11 gene codes for most of the 22 kDa LHCI polypeptide produced by tomato leaves. In our tomato cDNA library the frequency of clones for *cab*11 is approximately 100 times that of *cab*12; work is in progress to determine whether *cab*12 mRNA is also rare in vivo.

Comparison of the sequences of the cDNA and genomic clones revealed that both *cab*11 and *cab*12 contain two introns in identical positions (Fig. 2); the first intron is in the same position as an intron found in the Type II and Type III LHCI CAB genes and in the CP24 CAB genes [11,12,19]; the second intron is in the same position as one found in the Type I LHCI CAB genes [9]. Since *cab*11 and *cab*12 have the same number and position of introns, as well as 88% sequence identity, we consider them both to be Type IV LHCI.

DNA from *cab*11 and *cab*12 was used to probe Southern blots of tomato (*Lycopersicon esculentum*) and its close relative, *L. pennellii*, to determine if additional genes encoding the Type IV LHCI CAB protein existed in the genomes of these plants (Fig. 3). With one exception, only two hybridizing fragments were observed in all restriction digests probed with *cab*12 under

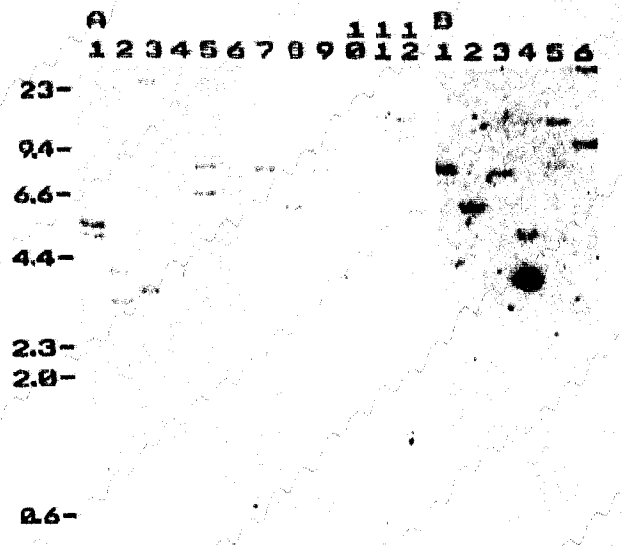

Fig. 3. Southern blots of *L. pennellii* and *L. esculentum* genomic DNA. (A) Probed with the 0.5 kb *Nco*I-*Bam*HI fragment from the *cab*12 cDNA clone. Odd-numbered lanes, *L. pennellii* DNA; even-number lanes, *L. esculentum* DNA; digested with *Hin*dIII (lanes1,2); *Bgl*II (lanes 3,4), *Xba*I (lanes 5,6); *Dra*I (lanes 7,8); *Eco*RV (lanes 9,10); and *Eco*RI (lanes 11,12). Hybridization conditions were 6 X SSC, 65°, washing conditions were 2 X SSC, 65°, so that both Type IV genes were detected with the *cab*12 probe. (B) Higher stringency hybridization with the 0.7 kb *Hin*dIII-*Pst* I fragment from the *cab*11 cDNA clone. Lanes 1-6 are replicas of lanes 7-12 in (A). Hybridization conditions as in A but washed with 0.1 X SSC, 65°. Numbers at left indicate position and size (in kb) of marker fragments.

low-stringency conditions (Fig. 3A, lanes 1-12), where the probe hybridizes with both *cab*11 and *cab*12 genes (the second *Dra* I bands of about 1 kb are faint and may not be visible in the printed figure). The three hybridizing fragments in the *Eco*RI digest of *L. pennellii* (Fig. 3A, lane 11) are likely the results of an *Eco*RI site occurring inside one of these two genes. The fragments containing *cab*11 were identified under high stringency conditions (Fig. 3B, lanes 1-6). In the case of *Hin*dIII, comparison of the *L. esculentum* restriction map (Fig. 2A) and the Southern blot (Fig. 3A, lane 2) showed that *cab*11 is on the 4.0 kb fragment and *cab*12 on the 3.0 kb fragment. The cloned *Hin*dIII fragments were fine-mapped and Southern blotted to rule out any tightly-linked duplications; no additional CAB genes were found (data not shown). We therefore conclude that *L. esculentum* and *L. pennellii* both contain the two genes encoding the Type IV CAB polypeptide of LHCI.

Due to the presence of polymorphisms in fragment sizes between the two species of *Lycopersicon* (Fig. 3), we were able to map the two loci in segregating F2's of the interspecific cross [20]. Results indicate that *cab*11 maps to chromosome 3, approximately 4 cM from the marker TG242, and the *cab*12 gene maps to chromosomes 6, approximately 7 cM from marker TG275 (data not shown).

4. DISCUSSION

LHCI can be subdivided into LHCI-730 and LHCI-680, with fluorescence emission maxima at 730 and 680 nm respectively [23,28]. Ikeuchi et al.[29] have reported partial peptide sequences from all four of the mature LHCI polypeptides of spinach and pea, including a spinach 20 kDa polypeptide of LHCI-730 with an N-terminal sequence that closely matches that of tomato *cab*11 and *cab*12-encoded polypeptides, starting with Lys51 and Lys52 (Fig. 2, arrowheads). In tomato, most of this protein appeared to be N-terminally blocked, but a low level of clearly identifiable amino acids were detected in cycles 4(Gln), 6(Leu), 9(Leu) and 10(Ala), suggesting that the transit peptide of the tomato Type IV precursor polypeptide is also removed at this point, giving a mature polypeptide of 200 residues. The fact that there is Gln rather than Glu at the 4th position and Ala rather than Thr at the 10th position is consistent with the tryptic peptide results which indicate that the dominant polypeptide in the tomato 22 kDa band originates from *cab*11 rather than *cab*12. By comparison with their peptide sequences from spinach and pea, Ikeuchi et al.[29] also confirmed that the tomato *cab*6A,B (LHCI Type I) genes encode the other LHCI-730 CAB polypeptide, and that the

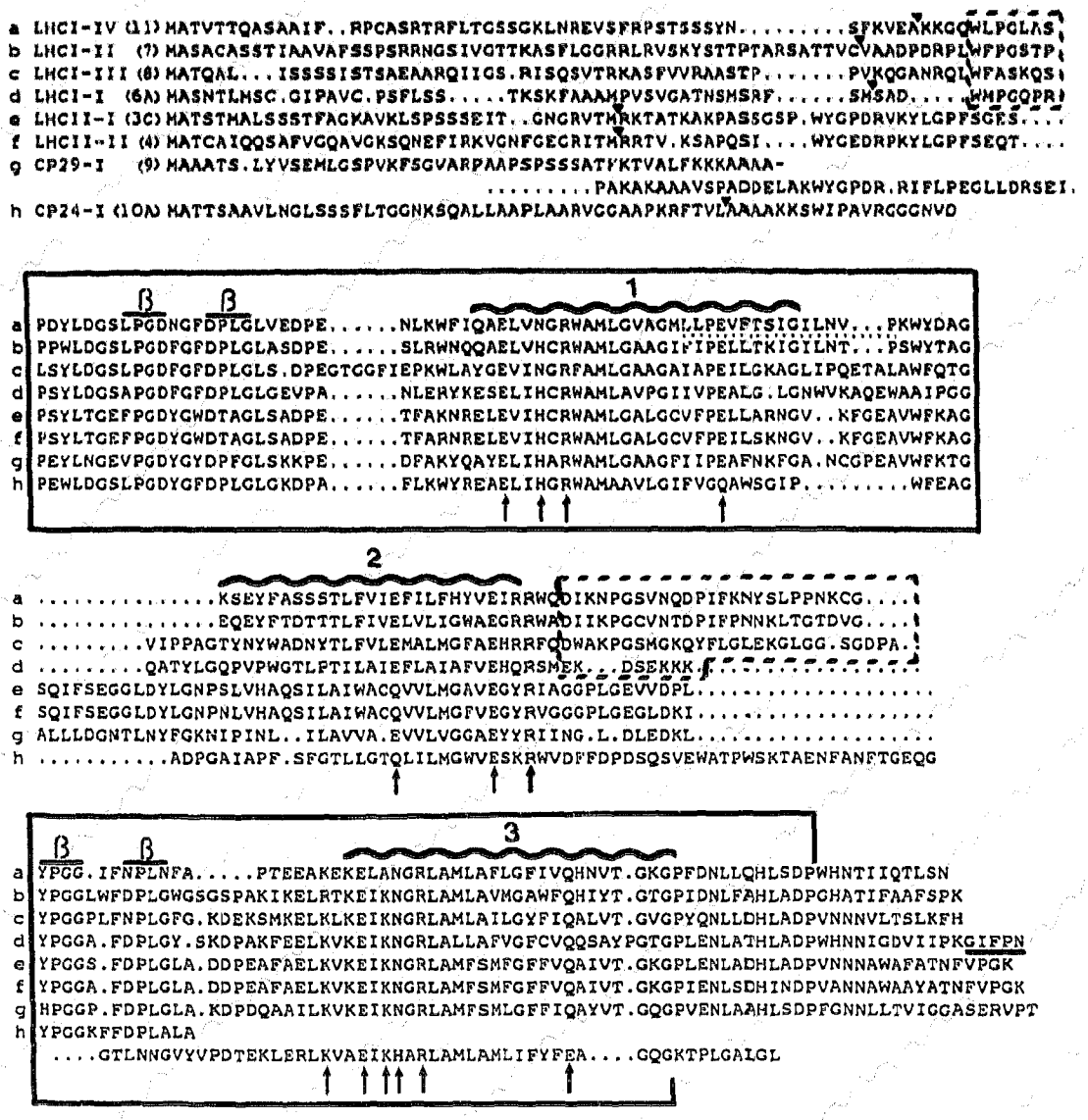

Fig. 4. Comparison of tomato CAB protein sequences. Thick wavy lines, predicted trans-membrane helices; betas, predicted beta turns. The two regions of high sequence similarity are enclosed in solid boxes. Sequence corresponding to pea 20 kDa DCCD-labeled polypeptide [32] is underlined with dots. C-terminal tryptic peptide identifying *cab*6A polypeptide is underlined. Dashed boxes are proposed LHCI-specific motifs. Arrows point to highly conserved residues discussed in text. (a) LHCI Type IV(*cab*11); (b) LHCI TypeII(*cab*7); (c) LHCI TypeIII(*cab*8), (d) LHCI TypeI(*cab*6A); (e) LHCII TypeI(*cab*3C); (f) LHCII TypeII(*cab*4); (g) CP29 TypeI(*cab*9); and (h) CP24 Type I(*cab*10A).

tomato *cab*7 (LHCI Type II) and *cab*8 (LHCI Type III) genes encode the two polypeptides of tomato LHCI-680.

Fig. 4 shows the deduced amino acid sequences of the four types of LHCI genes lined up with the sequences for Type I and II LHCII (*cab*3 and *cab*4), Type I CP29 (*cab*9) and the CP24 polypeptide (*cab*10A). All the proteins have three hydrophobic regions long enough to be trans-membrane helices (heavy wavy lines). There is a high degree of conservation (35% identity plus conservative substitutions) in the two homology regions diagnostic of CAB proteins (boxed). These regions include the first and third trans-membrane helices and the highly polar sequences N-terminal to them which are enriched in turn-promoting amino acids (Asp, Gly, Pro). According to the current model for LHCII topology [30], the N-termini of helices 1 and 3 are exposed on the stroma side of the thylakoid membrane. Within the highly conserved regions, all CABs should have the same folding pattern, with the core 3D conformations within 1.2–1.5 Å (root mean square deviation) of each other [31]. Direct evidence for the 22 kDa polypeptide having the same relative orientation comes from DCCD-labelling of pea CAB polypeptides [32]. A DCCD-labeled CNBr fragment of a pea 20 kDa polypeptide had the sequence LXPEVFTSIGIINVP which matches the corresponding *cab*11 and *cab*12 sequences except for an Ile/Leu substitution (Figs 2 and 4). Since DCCD was used under conditions where it reacts with acidic residues in a hydrophobic environment [32], this supports the location of a Glu within the first membrane-spanning helix.

Arrows in Fig. 4 mark the positions of conserved His, Asn and Gln residues that could be ligands to the Chl Mg^{+2} atoms, as well as several conserved Glu, Gln and Arg residues within trans-membrane helices. There are also a number of highly conserved Leu and aromatic residues. The conserved polar residues could be involved in H-bonding to polar groups on the porphyrin rings [33]. Note that there are a number of charged residues within the predicted trans-membrane helices; this is comparable to the situation in bacteriorhodopsin where the trans-membrane helices have a number of ionizable side-chains that are not part of the proton pore [34].

Differences between LHCI and the PSII CAB polypeptides can be seen in Helix 2 and the sequences flanking it. The PSI LHCI sequences have very short connectors between the first conserved region and Helix 2. With the exception of *cab*6A, the LHCI's have longer connectors between Helix 2 and the second conserved region. In all four LHCI polypeptides, this connector has more positively-charged residues than in the PSII polypeptides. The CP24 polypeptide also has a long connector, but it has 6 negative charges and no detectable similarity with the other CAB polypeptides. In addition to this region, predicted to be exposed on the stromal side of the membrane, there is a second motif near the N-terminus which appears to be LHCI-specific (Fig. 4). These motif sequences could be involved in specific binding to the PSI core or other PSI proteins, or could be required for targetting the protein to PSI during assembly of the holocomplex. In addition, Helix 2 appears to have some periodicity in conserved residues: this could indicate that one side of it is making specific contacts with other proteins.

Acknowledgements: Our work was supported by U.S. Dept of Agriculture (EP), Natural Sciences and Engineering Research Council of Canada (BRG) and Medical Research Council of Canada (RA). We should like to thank Drs N. Nelson and R. Malkin for their generous gifts of antibodies, Drs J. Golbeck and I. Damm for helpful discussions about PSI, Drs P. Jahns, W. Junge and M. Ikeuchi for communication of unpublished results, and Mr Ted Bures for expert technical help in protein sequencing.

REFERENCES

[1] Green, B.R. (1988) Photosyn. Res. 15, 3–32.
[2] Thornber, J.P. (1986) in: Encyclopedia of Plant Physiology 19 (New Series) (Staehelin, L.A. and Arntzen, C.J., eds) Springer-Verlag, Berlin, pp. 95–142.
[3] Chitnis, P.R. and Thornber, J.P. (1988) Photosyn. Res. 16, 41–63.
[4] Evans, P.K. and Anderson, J.M. (1986) FEBS Lett. 199, 227–233.
[5] White, M.J. and Green, B.R. (1987) Eur. J. Biochem. 163, 545–551.
[6] Darr, S.C., Somerville, S.C. and Arntzen C.J. (1986) J. Cell Biol. 103, 733–740.
[7] Pichersky, E. and Green, B.R. (1990) Curr. Res. in Photosyn. 3, 553–556.
[8] Green, B.R., Pichersky, E. and Kloppstech, K (1991) Trends Biochem. (in press).
[9] Hoffman, N.E., Pichersky, E., Malik, V.S., Castresana, C., Ko, K., Darr, S.C. and Cashmore, A.R. (1987) Proc. Natl. Acad. Sci. USA 84, 8844–8848.
[10] Pichersky, E., Hoffman, N.E., Malik, V.S., Bernatzky, R., Tanksley, S.D., Szabo, L. and Cashmore, A.R. (1987) Plant Mol. Biol. 9, 205–216.
[11] Pichersky, E., Tanksley, S.D., Piechulla, B., Stayton, M.M. and Dunsmuir, P. (1988) Plant Mol. Biol. 11, 69–71.
[12] Pichersky, E., Brock, T.G., Nguyen, D., Hoffman, N.E., Piechulla, B., Tanksley, S.D. and Green, B.R. (1989) Plant Mol. Biol. 12, 257–270.
[13] Mullet, J.E., Burke, J.J. and Arntzen, C.J. (1980) Plant Physiol. 65, 814–822.
[14] Haworth, P., Watson, J.L. and Arntzen, C.J. (1983) Biochim. Biophys. Acta 724, 151–158.
[15] Camm, E.L. and Green, B.R. (1989) Biochim. Biophys. Acta 974, 180–184.
[16] Aebersold, R., Leavitt, J., Hood L.E. and Kent S.B.H. (1987) Proc. Natl. Acad. Sci. USA 184, 6970–6974.
[17] Aebersold, R., Pipes, G.D., Wettenhall, R.E.H., Nilco, H. and Hood, L.E. (1990) 187, 56–65.
[18] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[19] Schwartz, E. and Pichersky, E. (1990) Plant Mol. Biol. 15, 157–160.
[20] Pichersky, E., Bernatsky, R., Tanksley, S.D., Breidenbach, R.B., Kausch, A.P. and Cashmore, A.R. (1985) Gene 40, 247–258.
[21] Maxam, A.M. and Gilbert, W. (1980) Methods. Enzymol. 65, 499–560.
[22] Sanger, R., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[23] Anderson, J.M. (1984) Photobiochem. Photobiophys. 8, 221–228.
[24] Nechushtai, R., Peterson, C.C., Peter, G.F. and Thornber, J.P (1987) Eur. J.Biochem. 164, 345–350.
[25] Herrin, D.L., Plumley, F.G., Ikeuchi, M., Michaels, A.S. and Schmidt, G.W. (1987) Arch. Biochem. Biophys. 254, 397–408.
[26] Vainstein, A., Peterson, C.C. and Thornber, J.P. (1989) J.Biol.Chem. 264, 4058–4063.
[27] Anandan, S., Vainstein, A. and Thornber, J.P. (1989) FEBS Lett. 256, 150–154.
[28] Lam, E., Ortiz, W. and Malkin, R. (1984) FEBS Lett. 168, 10–14.
[29] Ikeuchi, M., Hirano, A. and Inoue, Y. (1991) Plant and Cell Physiol. (in press).
[30] Karlin-Neumann, G.A., Kohorn, B.D., Thornber, J.P. and Tobin, E.M. (1985) J. Mol. Appl. Genet. 3, 45–61.
[31] Chothia, C. and Lesk, A.M. (1986) EMBO J. 5, 823–826.
[32] Jahns, P. and Junge, W. (1990) Eur. J. Biochem. (in press).
[33] Loach, P.A., Michalski, T. and Parkes-Loach, P.S. (1990) Curr. Res. Photosyn. 2, 69–72.
[34] Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. and Downing, K.H.(1990) J. Mol. Biol. 213, 899–929.