Mol Gen Genet (1991) 227:277–284
002689259100174D

MGG

Chlorophyll *a/b* binding (CAB) polypeptides of CP29, the internal chlorophyll *a/b* complex of PSII: characterization of the tomato gene encoding the 26 kDa (type I) polypeptide, and evidence for a second CP29 polypeptide

Eran Pichersky [1], Rajagopal Subramaniam [2], Michael J. White [3], James Reid [1], Ruedi Aebersold [4], and Beverley R. Green [2]

[1] Biology Department, University of Michigan, Ann Arbor, MI 48109, USA
[2] Botany Department, University of British Columbia, Vancouver, B.C. V6T 2B1, Canada
[3] Botany Department, N. Carolina State University, Raleigh, N.C. USA
[4] Biomedical Research Centre and Biochemistry Dept., University of British Columbia, Vancouver, B.C. V6T 1W5, Canada

Summary. CP29, the core chlorophyll *a/b* (CAB) antenna complex of Photosystem II (PSII), has two nuclear-encoded polypeptides of approximately 26 and 28 kDa in tomato (*Lycopersicon esculentum*). Cab9, the gene for the Type I (26 kDa) CP29 polypeptide was cloned by immunoscreening a tomato leaf cDNA library. Its identity was confirmed by sequencing tryptic peptides from the mature protein. Cab9 is a single-copy gene with five introns, the highest number found in a CAB protein. In vitro transcription-translation gave a 31 kDa precursor which was cleaved to about 26 kDa after import into isolated tomato chloroplasts. The Cab9 polypeptide has the two highly conserved regions common to all CAB polypeptides, which define the members of this extended gene family. Outside of the conserved regions, it is only slightly more closely related to other PSII CABs than to PSI CABs. Sequence analysis of tryptic peptides from the Type II (28 kDa) CP29 polypeptide showed that it is also a member of the CAB family and is very similar or identical to the CP29 polypeptide previously isolated from spinach. All members of the CAB family have absolutely conserved His, Gln and Asn residues which could ligate the Mg atoms of the chlorophylls, and a number of conserved Asp, Glu, Lys and Arg residues which could form H-bonds to the polar groups on the porphyrin rings. The two conserved regions comprise the first and third predicted trans-membrane helices and the stroma-exposed segments preceding them.

Key words: Chlorophyll *a/b* antenna polypeptide – CAB gene family – CP29 – Cab9

Introduction

The chlorophyll (Chl) *a/b*-protein complexes of higher plants and green algae act as light-harvesting antennas for both Photosystem I (PSI) and Photosystem II (PSII). There are three clearly distinguishable Chl *a/b* complexes in PSII: LHCII, CP29 and CP24 (Fig. 1). The CP29 complex is of particular interest because it remains with the active PSII core when LHCII and CP24 are removed with octylglucoside, and appears to be an internal or "core" antenna (Camm and Green 1989; Ghanotakis et al. 1987a; Staehelin 1986). Most CP29 preparations contain two polypeptides in the 26–30 kDa range (White and Green 1987a; Camm and Green 1989); prior to the experiments reported here it was not known whether

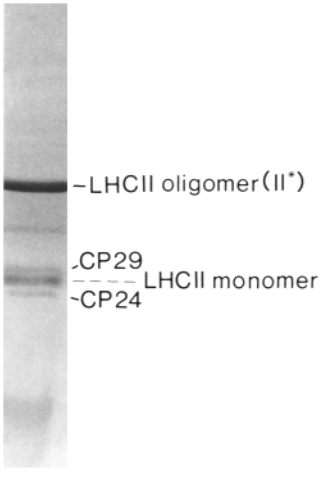

Fig. 1. Tomato Photosystem II (PSII) chlorophyll (Chl)-protein complexes (unstained). PSII membranes were solubilized with 300 mM octylglucoside in 30% (w/v) glycerol, 2 mM TRIS-maleate, pH 8 and separated on a 10% polyacrylamide gel containing 1.32 M TRIS-HCl, pH 9.8, 0.05% SDS, in the dark at 4° C. The LHCII oligomer is also called CPII*

Offprint requests to: B.R. Green

Abbreviations: Chl, chlorophyll; CP29, core Chl *a/b* antenna of Photosystem II; LHCII, major Chl *a/b* antenna of Photosystem II; LHCl, Chl *a/b* antenna of Photosystem I; PS, Photosystem; CPII*, oligomeric form of LHCII isolated by SDS-PAGE; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; TMH, trans-membrane (membrane-spanning) helix

they were encoded by different genes or were derived from a single translational product. These proteins are not phosphorylated (Bassi et al. 1987; Dunahay et al. 1987) and they remain associated with the reaction centre under conditions where some of the LHCII migrates to the stromal regions of the thylakoid (Dunahay and Staehelin 1987). CP29 is distinguished from LHCII by its higher Chl *a*/*b* ratio (4–5 vs 1.2), the proteolytic fragmentation pattern of its polypeptides and (in some species) their higher apparent molecular weights (Green 1988). However, its polypeptides are immunologically related to each other and to the other Chl *a*/*b* polypeptides (White and Green 1987a; Camm and Green 1989), suggesting that they share some sequence similarity with them.

In this paper, we show that tomato CP29 has two different polypeptides of about 26 and 28 kDa, and report the isolation and sequencing of the gene for the 26 kDa polypeptide. Comparison of the deduced amino acid sequence with those of other tomato Chl *a*/*b* or CAB polypeptides reveals that they form an "extended family" of genes. We discuss their structural and evolutionary relationships.

Materials and methods

Antibodies and immunoscreening. Polyclonal antibodies were raised to highly purified barley CP29 (White and Green 1987a). Antiserum at a dilution of 1/25 was pre-adsorbed with wild-type lambda lysate on nitrocellulose filters, then further treated with tomato LHCII made by Triton solubilization and cation precipitation (Ryrie et al. 1980). This step was necessary to remove antibodies to the epitopes shared by all Chl *a*/*b* proteins (Evans and Anderson 1986; White and Green 1987a, b). The purified antiserum was tested against gel-purified CP29 and LHCII and the procedure repeated until the antiserum no longer reacted with LHCII proteins. A tomato leaf cDNA expression library in Charon 16 (Hoffman et al. 1987) was screened with the antibodies according to the method of Young and Davis (1983) except that fish skin gelatin (Norland Products, New Brunswick, N.J.) was used as the blocking agent and the alkaline phosphatase-coupled second antibody was detected using Fast Red and Naphthol AS-MX phosphate (White and Green 1987a, b).

DNA sequence analysis. The tomato leaf cDNA library and procedures were as described in Pichersky et al. (1985, 1987) and references therein. Nucleotide sequence determination was by the chemical method (Maxam and Gilbert 1980). The genomic clone was isolated from the tomato EMBL3 library described in Sugita et al. (1987).

Protein isolation and sequencing. Chloroplasts were isolated from greenhouse-grown tomato (*Lycopersicon esculentum* var. Manitoba), PSII membranes prepared, and CP29 and CPII* (LHCII oligomer) purified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using low SDS concentrations at 4° C in the dark ("green gels") as described (Camm and Green 1989). A Triton X-100/Chl ratio of 10/1 rather than 20/1 was used for preparing PSII membranes. For protein sequencing, oxygen-evolving PSII reaction centre cores were prepared from PSII membranes (Ghanotakis et al. 1987a; Camm and Green 1989) and washed with 3 M NaSCN, 50 mM $CaCl_2$, 0.01% Triton X-100, 10 mM MES, pH 6.5 to remove extrinsic proteins (Henrysson et al. 1989; Ljungberg et al. 1986). After solubilization with 2% SDS in 65 mM TRIS-HCl, pH 6.5, 50 mM dithiothreitol, 20% ethylene glycol, the polypeptides were separated by electrophoresis for 24 h at 4° C on a 12%–22% polyacrylamide gradient containing 0.05% SDS and 1.32 M TRIS, pH 8.8. Proteins were electrotransferred to nitrocellulose and stained with Ponceau S or Amido Black to locate the CP29 polypeptides. The protein bands corresponding to the CP29 polypeptides were cut out and the protein was digested on the support with trypsin as described (Aebersold et al. 1987). The resultant peptides were separated by narrow-bore reversed phase HPLC on a Waters Peptide Analyzer equipped with a Vydoc C-4 column. Individual peptides were collected manually and sequenced in an Applied Biosystems Model 477 or a Milligen/Biosearch Model 6600 protein sequenator.

In vitro translation and import. The full-length insert from clone CP29-2 was excised with *Sal*I and cloned into pGEM4. Purified plasmid DNA was linearized with *Hin*dIII and transcribed using SP6 polymerase (Pichersky et al. 1987). The transcript was translated in a wheat-germ extract (Amersham) using [^{35}S]methionine, and imported into intact tomato chloroplasts according to Pichersky et al. (1987, 1989). Thermolysin-treated chloroplasts were solubilized in 2% SDS, and the polypeptides separated on 7%–15% polyacrylamide gradient gels and fluorographed.

Results

Tomato CP29 polypeptides and antiserum purification

We have previously shown that spinach CP29 isolated by successive rounds of electrophoresis on green gels contains two polypeptides, and that these two polypeptides are part of the oxygen-evolving PSII core particles of Ghanotakis et al. (1987a), which are the smallest O_2-evolving PSII units that retain normal electron acceptor preferences (Ghanotakis et al. 1987b). Tomato CP29 also has two polypeptides, of approximately 26 and 28 kDa, one of which is preferentially removed from the core particles by cation-exchange chromatography (Green and Camm 1990).

In preparation for immunoscreening a cDNA library, antiserum raised to barley CP29 was purified by extensive pre-adsorption against tomato LHCII, in order to remove antibodies directed towards epitopes shared with the other Chl *a*/*b* polypeptides (White and Green 1987a,

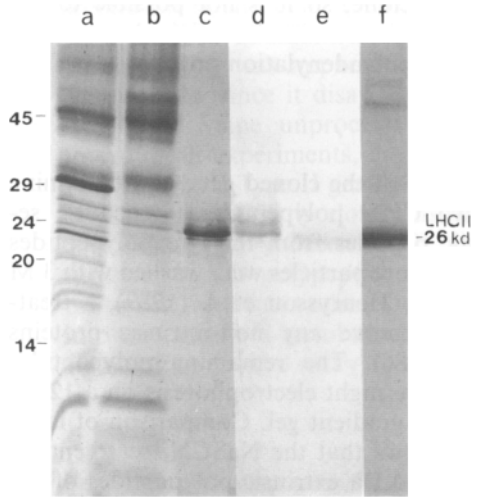

Fig. 2. Polypeptides of oxygen-evolving PSII core particles and LHCII. Lane a, PSII core particles; b, d, PSII core particles washed with 3 M NaSCN, 50 mM $CaCl_2$, 10 mM MES, pH 6.5, 0.01% Triton X-100; c, LHCII prepared by Triton solubilization and cation precipitation; e, f, LHCII oligomer (CPII*) from a "green gel". Lanes a–c and f, Coomassie stained; lanes d and e, immunoblotted with purified anti-CP29 antiserum. Samples were denatured by heating in 2% SDS and separated on a 12%–22% polyacrylamide gel, 1.32 M TRIS-HCl, pH 8.8

b; Camm and Green 1989). Figure 2 shows that the purified antiserum recognized the two polypeptides of tomato CP29 but had almost no reaction with the polypeptides of the oligomeric form of tomato LHCII isolated electrophoretically (CPII*). In both tomato and spinach (Camm and Green 1989) the lower CP29 polypeptide has approximately the same electrophoretic mobility as the major LHCII polypeptides, in contrast to barley where both the CP29 polypeptides have molecular weights that are several kilodaltons greater (White and Green 1987a, b). Tomato CP29 purified on gels as the green complex has both 26 and 28 kDa polypeptides, both of which react with the purified anti-CP29 antiserum (data not shown). Note that LHCII prepared by Triton solubilization and cation precipitation (Ryrie et al. 1980) sometimes contains a small amount of CP29 (Fig. 2c) while the electrophoretically purified oligomer does not (Fig. 2f).

Cloning and sequencing of Cab9

The purified antiserum was used to screen a tomato leaf cDNA expression library by standards methods (Young and Davis 1983). The single clone recovered, CP29-2,

```
TAAAGAAAGTATTGTGGCATTGCAAATTGCAATGTGGATAGGGTCACGTTATCCAAGTCTGGTCCAGGAGTTTTGCGATGCATGCCATGTCAGCACTGTC  100
ATCTGCCACGTCACAAACTCCTTCCAATCAAAACGCCGCATCAGATTCCACATTCATCTTCCCTGTGGCTCCAATAACAGAGCGACACTTGTCACGCTCT  200
CATCCACCAGAAAAAGCTTTAAGCTAGTATCTCCACTCCAAATACCATCCTCAATTCTCACTTCTCATCATCAACTCGACCTCAATTTTTTTTACCTCTT  300
GCCAGCGACACCGTTTAGCTACAA ATG GCT GCA GCT ACC TCT CTA TAC GTC TCC GAG ATG CTC GGA AGT CCG GTC AAG TTT  381
                          M  A  A  A  T  S  L  Y  V  S  E  M  L  G  S  P  V  K  F
AGC GGT GTT GCA AGG CCT GCT GCT CCA TCT CCA TCC TCC TCT GCT ACC TTC AAA ACT GTT GCC CTT TTC AAA AAG   456
 S  G  V  A  R  P  A  A  P  S  P  S  S  S  A  T  F  K  T  V  A  L  F  K  K
AAG GCG GCT GCA GCT CCT GCC AAA GCT AAG GCT GCC GCT GTC TCC CCT GCT GAC GAT GAG CTC GCC AAG TGG TAC G 532
 K  A  A  A  A  P  A  K  A  K  A  A  A  V  S  P  A  D  D  E  L  A  K  W  Y
GTGAGTAACATTACTACCTCCTTCCCTTTTATGAGTTTGATTGGACTCGAAATTTAATGCAGTAGTCGTATTAATTGATAACAAGATGAGTACTACATTT  632
TCTCACTTTATAAGAGAAGCTATTTAATTTCGCACAAAGTTGTATTAAATTTAAGTTCCCTCTAGCTGTACTGTAATTAATGTCAAACAGAACGGAAGAG  732
TAGAGTTCTCCTAAGTCCACGCAGCCTCTGTTTAACATCTTCCATTTAGTTTTTATATGCTCTTAAATTTTGCGAATATTTTCAG GA CCT GAC AGA  828
                                                                                        G  P  D  R
AGA ATT TTC CTG CCC GAG GGT CTA TTG GAC AGA TCA GAA ATC CCC GAG TAC TTG AAC GGA GAA GTT CCC GGA GA G  903
 R  I  F  L  P  E  G  L  L  D  R  S  E  I  P  E  Y  L  N  G  E  V  P  G  D
TAAGCTTCGATCTACTTGCTGCATCTCGTAATTTTGTTGATAAATTATATTACTTAGCATATACAAATAGAGAAATTGTCCATATATATAAGAATTTTCT 1003
TGTACTTGCTGCAG C TAC GGT TAC GAT CCT TTT GGT TTG AGC AAG AAG CCA GAA GAC TTC GCC AA GTAAGTGTAGTACTAG 1084
               Y  G  Y  D  P  F  G  L  S  K  K  P  E  D  F  A  K
TGACACATTAATTAATTTTACAGCTCAGTCAGCCAACTTTGTTCTTGTTGATTGTGGAGATGGATCTAAGTTTAAAATTGTTCGATCTGCTGCAG A TAT 1183
                                                                                                    Y
CAG GCA TAT GAG CTG ATC CAC GCC AGA TGG GCT ATG CTC GGT GCT GCT GGG TTT ATT ATC CCA GAA GCC TTC AAC  1258
 Q  A  Y  E  L  I  H  A  R  W  A  M  L  G  A  A  G  F  I  I  P  E  A  F  N
AAA TTT GGT GCC AAC TGT GGT CCT GAA GCT GTT TGG TTC AAG GTATATATATATACACACATCACCACTAATTGCAATTTGTTTAA 1344
 K  F  G  A  N  C  G  P  E  A  V  W  F  K
ATGTGAGGGGAAATCCTGATTATGAAATTTCGTCTCATGTGTAG ACT GGA GCT TTA CTT CTA GAC GGA AAC ACA TTG AAT TAC TTC 1430
                                               T  G  A  L  L  L  D  G  N  T  L  N  Y  F
GGA AAG AAC ATT CCC ATC AAC CTT ATT CTT GCT GTC GTC GCA GAG GTT GTT CTT GTT GGT GGT GCT GAG TAC TAC  1505
 G  K  N  I  P  I  N  L  I  L  A  V  V  A  E  V  V  L  V  G  G  A  E  Y  Y
AGA ATT ATT AAT GGC TTG GTGAGCTCCCCTTTTCCTTCACTACAACCTTCATCATAATATTATGTGCATTGACAGTTCTTGTATTTATGTTAAG 1599
 R  I  I  N  G  L
AAATCTACTTAATATGTATAATTAAATTTATTCAAAACTCAGTAATTTGCCTCGATCATAAAAACTGACAGTTGATTAGATTGTGGATTTAACGAATTGT 1699
GTCGGGATGCAG GAT TTG GAG GAC AAG CTA CAC CCA GGT GGT CCA TTT GAC CCA CTA GGC CTA GCA AAG GAC CCA GAC 1777
             D  L  E  D  K  L  H  P  G  G  P  F  D  P  L  G  L  A  K  D  P  D
CAG GCA GCA ATC TTG AAG GTA AAA GAA ATC AAG AAT GGA AGA CTA GCC ATG TTT TCT ATG CTT GGT TTC TTC ATC  1852
 Q  A  A  I  L  K  V  K  E  I  K  N  G  R  L  A  M  F  S  M  L  G  F  F  I
CAA GCC TAT GTG ACC GGA CAG GGT CCT GTT GAG AAT TTG GCT GCC CAT TTG AGT GAC CCC TTT GGC AAC AAC TTG  1927
 Q  A  Y  V  T  G  Q  G  P  V  E  N  L  A  A  H  L  S  D  P  F  G  N  N  L
CTC ACC GTC ATC GGT GGT GCT TCT GAA AGA GTC CCA ACC CTG TAATTTTCTTTTTGGCCGGCGGAACGTACAATTCTGCAATACAT 2013
 L  T  V  I  G  G  A  S  E  R  V  P  T  L
TTCAGAGCTGTATGTGAAAAACCAGCTTGTAGTTTATGAAGATAAACTGTACTTATTCTCATACAGCTACTACTCTGGCCACATTCGTCCTTCATTCCAA 2113
TTTATGCAGTAACATTATTTTTTTATGAGTATAACTTATATATATCGTCATTATTTCTAAAGAAAAACACGTAATCATATTCACAATGTGAGGATACAAT 2213
ATTTGTCTATACTTGAAAAAACTTAATTTACCTGTAGTGAAGAATTTTTATTTGCCACACAAACGTTAGATGTATTTAAGACTACAAATTTTAATATATT 2313
AAAAATACAAGCTT                                                                                        2327
```

Fig. 3. Nucleotide and deduced amino acid sequence of the Cab9 gene. Nucleotide sequences referred to in the text are *underlined*. *Boxed* amino acid sequences were confirmed by peptide sequencing. The *bottom box* contains two contiguous sequences, IINGLDLEDK and LHPGGPFDPLGLAK

produced a fusion protein of about 70 kDa, which was recognized by the purified antiserum (data not shown). It was not recognized by a monoclonal antibody that is thought to be specific for the N-terminal end of one of the CP29 apoproteins (Darr et al. 1986; Camm and Green 1989). A labelled 0.4 kb *Sst*I – *Xba*I DNA fragment from the 5′ end of this clone was then used as a probe to screen an EMBL3 tomato genomic library. Several clones were isolated that, upon restriction digest analysis, proved to carry the same region from the tomato nuclear DNA as the cDNA clone. The probe was also used to isolate additional clones from the tomato cDNA library. Sequence analysis showed that all cDNA clones had the same nucleotide sequence (within their overlapping regions), suggesting that tomato has a single expressed nuclear gene encoding this CP29 protein. Southern blots also indicated the presence of a single nuclear gene, located on chromosome 6 by restriction fragment length polymorphism (RFLP) mapping (E. Pichersky and S. Tanksley, unpublished). This gene has been designated Cab9. Tomato genes designated Cab1 to Cab8 encode other CAB polypeptides of PSI and PSII (Pichersky and Green 1990).

In the two longest of the new cDNA clones, CP29-21 and CP29-22, the first nucleotide was the A at position 270 (Fig. 3). The original cDNA clone isolated, CP29-2, is shorter and its first nucleotide is at position 517. Comparison of the nucleotide sequence of CP29-21 and CP29-22 with the genomic clone revealed that this gene contains five introns, the largest number of introns in any of the CAB genes so far characterized. Only the first two have counterparts in other CAB genes (Pichersky et al. 1989). Furthermore, the comparison showed that the cDNA clone CP29-2 was made from an mRNA from which intron II had not yet been spliced out. The last nucleotides of CP29-21 and CP29-22 prior to the poly(A) tract are at positions 2137 and 2140, respectively; the last nucleotide of the CP29-2 clone is at position 2154. Note that all three positions are succeeded by an A in the genomic sequence and by a poly(A) in the cDNA clone, so it is not possible to tell whether this A was present in the original transcript or added later in the polyadenylation process.

The protein sequence

To verify the identity of the cloned gene and determine which of the two CP29 polypeptides it encodes, sequences of tryptic peptides from the two polypeptides were obtained. PSII core particles were washed with 3 M NaSCN according to Henrysson et al. (1989), a treatment reported to remove any non-intrinsic proteins (Ljungberg et al. 1986). The remaining polypeptides were separated by overnight electrophoresis on a 12%–22% polyacrylamide gradient gel. Comparison of lanes a and b in Fig. 2 shows that the NaSCN treatment removes the 23 and 33 kDa extrinsic polypeptides of the water-splitting apparatus and diminishes the amounts of several others. However, it does not remove either of the two CP29 polypeptides, although a small amount of the 26 kDa species is detached if there is excess residual detergent (B.R. Green and T. Peng, unpublished results).

After electro-transfer of the separated polypeptides to a nitrocellulose membrane, the two CP29 polypeptide bands were cut out, digested with trypsin and the resulting peptide cleavage fragments were separated by HPLC and sequenced. Four separate tryptic fragments from the lower molecular weight (26 kDa) polypeptide matched perfectly with the deduced amino acid sequence of the cloned gene (boxed in Fig. 3), and did not match any of the previously sequenced tomato CAB genes (see below). We can therefore conclude that the gene we have isolated, Cab9, encodes the 26 kDa CP29 polypeptide.

Four tryptic fragments from the higher molecular weight (28 kDa) polypeptide were also sequenced (Fig. 4). Two of them were similar but not identical to sequences in the 26 kDa gene, and a third was similar to a peptide from spinach CP29 isolated by HPLC (Henrysson et al. 1989). The fourth fragment from the 28 kDa

```
                          87
(a)  Tomato CP29-26kDa    IPEYLNGEVPGDYGYDPFGLSKKPE      (gene)
                           * ** *   **** *****
     Tomato CP29-28kDa    ?P?YLDGSLVGDYGFDPFGLG?PA       (peptide)
                            S'

                         206
(b)  Tomato CP29-26kDa     LHPGG.PFDPLGLAKDPDQ           (gene)
                           * ***  ******* **
     Tomato CP29-28kDa     LYPGGSFFDPLGLAADPPK           (peptide)
                           A               E

(c)  Spinach CP29          NLAGDIIGTRTESADVKSTSLQPYSEVFG      (peptide)
                                          *  ********
     Tomato CP29-28 kDa                   TPFQPYSEVFGLQR     (peptide)

                         60
(d)  Tomato Cab7(LHC I)    ADPDRPLWFPGST     (gene)
                             *****  **
     Tomato CP29-28 kDa    QIPDRPL?YPGAK     (peptide)
```

Fig. 4. Tryptic peptide sequences from the 28 kDa CP29 polypeptide, aligned with closest matching sequence from CP29-26 kDa or other CAB polypeptide. ? denotes uncertain identification; *slash* between vertical amino acids, alternative assignment; *, identical residues; ·, gap introduced for optimal alignment. Amino acid numbering refers to the amino acid sequence of the unprocessed precursor as deduced from gene sequencing (see Fig. 5)

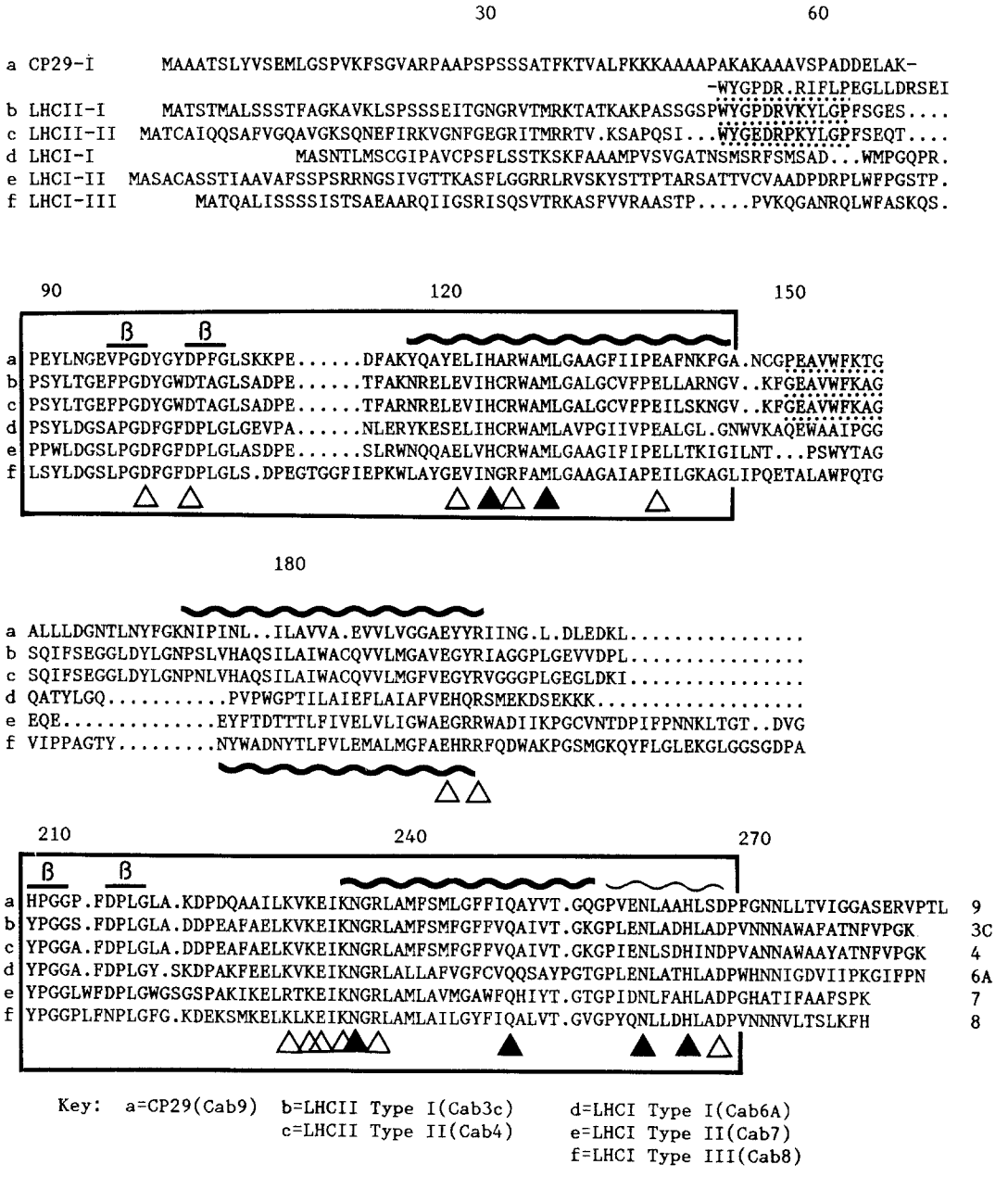

Fig. 5. Comparison of tomato Chl *a/b*-binding polypeptides. The two regions of high homology shared by polypeptides of both photosystems are *boxed* (*dark lines*). Additional homology between the 26 kDa polypeptide of CP29 (Cab9) and LHCII (Cab3c; Cab4) sequences is indicated by *dotted underlining*; there is some additional homology in the second trans-membrane helix (TMH). Beta-turns predicted according to Wilmot and Thornton (1988) are indicated by a *short horizontal bar* and the letter *beta*; membrane-spanning helices by *wavy lines* (Green 1990). *Arrowheads* indicate possible Chl ligating residues conserved in all tomato sequences. *Open arrowheads* indicate polar residues conserved in all six sequences

band was more similar to a region near the N-terminus of Cab7, one of the LHCI polypeptides (Fig. 5). Note that the first two tryptic peptides (a and b) are from the first and second conserved regions shared by all Chl *a/b* polypeptides, clearly showing that the 28 kDa polypeptide is also a member of the CAB family and not an unrelated polypeptide.

In vitro transcription, translation and import

Examination of the complete sequence of the 26 kDa CP29 protein reveals no obvious cleavage site for removal of the transit peptide. The first 85 amino acids of the sequence are almost completely unrelated to any other CAB sequences with the exception of the hexapeptide WYGPDR (Fig. 5, underlined). In vitro transcription of the pGEM construct, followed by translation in the wheat-germ system gave a 31 kDa precursor (Fig. 6c), which was taken up by intact tomato chloroplasts and cleaved to about 26 kDa (Fig. 6a), suggesting that approximately 50 amino acids were removed from the N-terminal end of the precursor. Unfortunately, the N-terminus of the CP29 26 kDa protein is blocked, so we were unable to determine the exact cleavage site. A

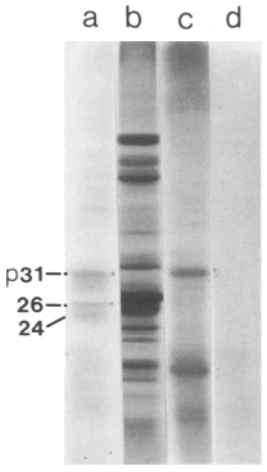

Fig. 6a–d. In vitro translation and import of Cab9 polypeptide. Lanes **a**, **c**, **d** autoradiograms; **b** Coomassie stained. **a**, **b** Product taken up and processed by intact tomato chloroplasts. **c** Translation mixture with Cab9 mRNA; **d** translation mixture without added RNA

second processing product of about 24 kDa was found in all experiments. It is not clear if this results from the action of residual protease or from cleavage at a second site (Clark et al. 1989). An additional translation product of 14–17 kDa was detected in most translation mixtures, but was apparently not taken up by intact chloroplasts, since it disappeared after the thermolysin treatment. Some unprocessed 31 kDa precursor was found in all experiments, as has been reported for the Lemna LHCII Type II gene product (Kohorn and Tobin 1986). Since CP29 is closely associated with the PSII core (Camm and Green 1989), there is perhaps little turnover of this polypeptide and consequently less opportunity for extra copies to be correctly incorporated into the thylakoid membrane, especially within the time frame of an in vitro assay.

Discussion

The complete protein sequence of the tomato 26 kDa CP29 polypeptide deduced from the DNA sequence establishes it as a distinct member of the CAB gene family (Fig. 5, line a). It has the two regions of conserved sequence shared by all Chl *a*/*b* polypeptides whose complete sequences have been determined to date (boxed). However, outside these regions the CP29 and LHCII sequences (Fig. 5, lines b and c) are only slightly more related to each other than they are to the three PSI CABs (Fig. 5, lines d–f). The first two tryptic peptides, which confirm the identity of polypeptide and gene product, are from the first conserved region but have amino acids unique to Cab9; the third is from a much less conserved region following the second transmembrane helix (TMH) and is contiguous with the fourth polypeptide which starts the second conserved region but has a unique His (H) where all other tomato CAB genes have Tyr (Y).

Henrysson et al. (1989) have isolated a highly purified Chl-protein from spinach PSII membranes by cation exchange HPLC chromatography, and identified it as CP29 by its comigration with the CP29 band from PSII membranes on SDS-PAGE under mild conditions (green gel), and by the similarity of their absorption spectra. However, the HPLC-isolated spinach CP29 had only one polypeptide, and a partial protein sequence obtained from it was very similar to one of the tryptic peptide sequences we obtained from the tomato 28 kDa polypeptide (Fig. 4c). This is consistent with our previous results indicating that cation exchange chromatography of dodecyl maltoside-solubilized PSII core particles results in the preferential removal of the 28 kDa polypeptide, which is eluted from the column with a NaCl gradient (Green and Camm 1990). The 26 kDa polypeptide remains with the other polypeptides of the PSII core and elutes in the void volume.

In keeping with the nomenclature of LHCI and LHCII genes (Hoffman et al. 1987; Pichersky et al. 1985, 1987, 1989), we propose to call the protein whose gene we have cloned "CP29 Type I" and the 28 kDa protein "CP29 Type II" (Pichersky and Green 1990). The 28 kDa CP29 protein is structurally related to the other members of the CAB family; two of its tryptic peptide sequences (a and b) are clearly related to sequences beginning the first and second conserved regions in all the other CAB proteins. However, the 28 kDa polypeptide must be encoded by an independent gene, since all of its tryptic peptide sequences are unique. Therefore, the two polypeptides of CP29 are the products of two different genes, which are themselves different from any other CAB gene.

We have not yet succeeded in isolating clone(s) encoding the 28 kDa protein by screening a cDNA library with probes derived from the tryptic peptide sequences. One possible explanation is that the mRNA, and consequently cDNA clones, encoding this protein are rare. The leaf steady-state level of mRNA for the Cab9 gene, encoding the 26 kDa CP29 polypeptide, is lower than that of any other CAB gene, including LHCI genes (J. Kellman, E. Pichersky and B. Piechulla, in preparation). The rarity of these mRNAs may be correlated with the stability of the proteins even in the absence of Chl *b* (White and Green 1987b), consistent with their role as core antennas (Camm and Green 1989; Green and Camm 1990; Greene et al. 1988) and with the low degree of incorporation into isolated chloroplasts (Fig. 6).

Secondary structure analysis (Green 1990) of the Cab9-encoded protein sequence predicts that it has three TMHs. This folding pattern is common to all Chl *a*/*b* and Chl *a*/*c* proteins, as well as to a related group of polypeptides that have not been shown to bind Chl but may be involved in the development of the photosynthetic apparatus (Grimm et al. 1989; Kolanus et al. 1987). The regions of high similarity with the other CABs comprise the first and third TMHs (Fig. 5, wavy line) and a markedly polar segment preceding each of them. These segments are rich in Pro and Gly as well as charged amino acids and are predicted to have high probabilities for Type I and II beta-turns, calculated according to the method of Wilmot and Thornton (1988). They are probably both exposed on the stromal side of the thylakoid membrane, since the N-terminus of the major LHCII polypeptide is on the stromal side and the C-terminus on the lumenal side (Mullet 1983; Burgi et al. 1987).

In the two highly conserved regions, the 26 kDa CP29 sequence is 58% and 62% identical to the Type I LHCII sequence. It is only slightly less related to the three LHCI polypeptide sequences. Overall similarity is considerably less, because of the divergence of the non-conserved regions. In the two regions of high homology, 40% of the amino acids are identical in all six sequences of Fig. 5, implying conservation of essential protein-protein or protein-pigment contacts. Chothia and Lesk (1986) compared the X-ray crystallographic structures of members of several protein families and found that the folding pattern was well conserved even when the amino acid identity was considerably lower than in the CAB proteins. Based on their formula, the conserved regions of all CAB polypeptides sequenced to date should have very similar conformations, with a root mean square deviation of about 1 Å. This means that when the three-dimensional structure of LHCII is solved (Kühlbrandt 1990), it will be possible to model the structures of the other CAB polypeptides with reasonable confidence.

Six conserved residues which could be ligands to the Chl Mg^{2+} are indicated by arrowheads in Fig. 5. His is a Chl ligand in the bacterial reaction centre (Michel et al. 1986); Asn and Gln can replace His in some bacterial mutants (Coleman and Youvan 1990). However, it has been estimated that LHCII contains about 15 Chl *a* and *b* per polypeptide chain (Kühlbrandt 1990). Highly conserved polar residues (marked with open arrowheads in Fig. 5), including three Glu and three Arg in the TMHs and several Asp, Glu, and Lys in the stroma-exposed regions, could be involved in H-bonding to the porphyrin rings. Studies on reconstitution of bacterial light-harvesting complexes emphasize the importance of polar interactions between the protein and the polar substituents of the prophyrin rings (Loach et al. 1990).

The complete sequence of the 26 kDa CP29 protein deduced from its gene, Cab9, firmly establishes it as a member of the extended family of CAB proteins, which we define here as any protein exhibiting the two blocks of homology identified in Fig. 5 and likely to have three TMHs, one in each of the homology boxes and a third in between. The obvious relatedness of the Chl *a*/*b* polypeptides to each other suggests that they arose as a result of successive gene duplications. Furthermore, there is some internal homology between the two conserved regions (Hoffman et al. 1987) suggesting that the ancestral gene could have arisen from the duplication of a smaller gene with one TMH and a polar N-terminus, followed by divergence of the intervening sequences. We should note here that some previous reports refer to the genes encoding the major polypeptide(s) of LHCII as "the" CAB proteins. However, with the discovery of homologous proteins in other Chl *a*/*b* antenna systems, such as LHCI, CP24 and CP29 such usage is no longer valid. It is also important to note that since many of the CAB proteins (as defined here) are so divergent from each other, the DNA sequences encoding them are even more divergent and they do not usually hybridize with each other (but see Pichersky et al. 1989). Thus, it is not possible to estimate the number of CAB genes in a given organism using a single type of CAB gene as a probe.

Acknowledgements. This work was initiated under a contract with the National Science Council of Canada (BRG). We thank the Natural Sciences and Engineering Research Council of Canada (BRG), Medical Research Council of Canada (RA), and U.S. Department of Agriculture (EP) for financial support, and Ted Peng, Dingren Shen and Edwards Bures for excellent technical assistance in the later stages of this work.

References

Aebersold R, Leavitt J, Hood LE, Kent SBH (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after *in situ* protease digestion on nitrocellulose. Proc Natl Acad Sci USA 184:6970–6974

Bassi R, Hoyer-Hansen G, Barbato R, Giacometti GM, Simpson DJ (1987) Chlorophyll-proteins of the photosystem II antenna system. J Biol Chem 262:13333–13341

Burgi R, Suter F, Zuber H (1987) Arrangement of the light-harvesting chlorophyll *a*/*b* protein complex in the thylakoid membrane. Biochim Biophys Acta 890:346–351

Camm EL, Green BR (1989) The chlorophyll *ab* complex, CP29, is associated with the Photosystem II reaction centre core. Biochim Biophys Acta 974:180–184

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826

Clark SE, Abad MS, Lamppa GK (1989) Mutations at the transit peptide-mature protein junction separate two cleavage events during chloroplast import of the chlorophyll a/b-binding protein. J Biol Chem 264:17544–17550

Coleman WJ, Youvan DC (1990) Spectroscopic analysis of genetically modified photosynthetic reaction centers. Annu Rev Biophys Chem 19:333–367

Darr SC, Somerville SC, Arntzen CJ (1986) Monoclonal antibodies to the light-harvesting chlorophyll a/b protein complex of photosystem II. J Cell Biol 103:733–740

Dunahay TG, Staehelin LA (1987) Immunolocalization of the chl *a*/*b*-light harvesting complex and CP29 under conditions favoring phosphorylation and dephosphorylation of thylakoid membranes. In: Biggins J (ed) Progress in photosynthesis research, vol 2. Martinus Nijhoff Publisher, Boston, pp 701–704

Dunahay TG, Schuster G, Staehelin LA (1987) Phosphorylation of spinach chlorophyll-protein complexes. CPII*, but not CP29, CP27, or CP24, is phosphorylated in vitro. FEBS Lett 215:25–30

Evans PK, Anderson JM (1986) The chlorophyll *a*/*b*-proteins of PSI and PSII are immunologically related. FEBS Lett 199:227–233

Ghanotakis DF, Demetriou DM, Yocum CF (1987a) Isolation and characterization of an oxygen-evolving Photosystem II reaction center core preparation and a 28 kDa Chl *a* binding protein. Biochim Biophys Acta 891:15–21

Ghanotakis DG, Waggoner CM, Bowlby NR, Demetriou DM, Babcock GT, Yocum CF (1987b) Comparative structural and catalytic properties of oxygen-evolving photosystem II preparations. Photosyn Res 14:191–199

Green BR (1988) The chlorophyll-protein complexes of higher plant photosynthetic membranes, or Just what green band is that? Photosyn Res 15:3–32

Green BR (1990) Structure prediction methods for membrane proteins: comparison with the x-ray structure of *R. viridis* photosynthetic reaction centre. In: Villafranca JJ (ed) Current research in protein chemistry: Techniques, structure, and function, Academic Press, San Diego, pp 395–404

Green BR, Camm EL (1990) Relationship of Chl *a*/*b*-binding and related polypeptides in PSII core particles. Curr Res Photosyn 1:659–662

Greene BA, Allred DR, Morishige DT, Staehelin LA (1988) Hierarchical response of light harvesting chlorophyll-proteins in a light-sensitive chlorophyll *b*-deficient mutant of maize. Plant Physiol 87:357–364

Grimm B, Kruse E, Kloppstech K (1989) Transiently expressed early light-inducible thylakoid proteins share transmembrane domains with light-harvesting chlorophyll binding proteins. Plant Mol Biol 13:583–593

Henrysson T, Schroeder WP, Spangfort M, Akerlund H-E (1989) Isolation and characterization of the chlorophyll *a/b* protein complex CP29 from spinach. Biochim Biophys Acta 977:301–308

Hoffman NE, Pichersky E, Malik VS, Castresana C, Ko K, Darr SC, Cashmore AR (1987) A cDNA clone encoding a photosystem I protein with homology to photosystem II chlorophyll *a/b*-binding polypeptides. Proc Natl Acad Sci USA 84:8844–8848

Kohorn BD, Tobin EM (1986) Chloroplast import of light-harvesting chlorophyll *a/b*-proteins with different amino termini and transit peptides. Plant Physiol 82:1172–1174

Kolanus W, Scharnhorst C, Kuhne U, Herzfeld F (1987) The structure and light-dependent transient expression of a nuclear-encoded chloroplast protein gene from pea (*Pisum sativum* L.) Mol Gen Genet 209:234–239

Kühlbrandt W (1990) Structure of the light-harvesting chlorophyll *a/b* protein complex by high-resolution electron crystallography. Curr Res Photosyn 2:217–222

Ljungberg U, Akerlund H-E, Andersson B (1986) Isolation and characterization of the 10-kDa and 22-kDa polypeptides of higher plant photosystem 2. Eur J Biochem 158:477–482

Loach PA, Michalski T, Parkes-Loach PS (1990) Probing the bacteriochlorophyll binding site requirements of the core light-harvesting complex of photosynthetic bacteria using BChl analogs. Curr Res Photosyn 2:69–72

Maxam AM, Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical cleavage. Methods Enzymol 65:499–560

Michel H, Epp O, Deisenhofer J (1986) Pigment-protein interactions in the photosynthetic reaction centre from *Rhodopseudomonas viridis*. EMBO J 5:2445–2451

Mullet JE (1983) The amino acid sequence of the polypeptide segment which regulates membrane adhesion (grana stacking) in chloroplasts. J Biol Chem 258:9941–9948

Pichersky E, Green BR (1990) The extended family of chlorophyll *a/b*-binding proteins of PSI and PSII. Curr Res Photosyn 3:553–556

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

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

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

Ryrie IJ, Anderson JM, Goodchild DJ (1980) The role of the light-harvesting chlorophyll *a/b*-protein complex in chloroplast membrane stacking. Eur J Biochem 107:345–354

Staehelin LA (1986) Chloroplast structure and supramolecular organization of photosynthetic membranes. In: Staehelin LA, Arntzen CJ (eds) Encyclopedia of plant physiology, new series, vol 19. Springer, Berlin, pp 1–84

Sugita M, Manzara T, Pichersky E, Cashmore AR, Gruissem W (1987) Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Mol Gen Genet 209:247–256

White MJ, Green BR (1987a) Antibodies to the photosystem I chlorophyll *a+b* antenna cross-react with polypeptides of CP29 and LHCII. Eur J Biochem 163:545–551

White MJ, Green BR (1987b) Polypeptides belonging to each of the three major chlorophyll *a+b* protein complexes are present in a chlorophyll *b*-less barley mutant. Eur J Biochem 165:531–535

Wilmot CM, Thornton JM (1988) Analysis and prediction of the different types of beta-turn in proteins. J Mol Biol 203:221–232

Young RA, Davis RW (1983) Efficient isolation of genes by using antibody probes. Proc Natl Acad Sci USA 80:1194–1198

Communicated by R.G. Herrmann