Volume 305, number 1, 18–22 FEBS 11193 June 1992

Identification of the polypeptides of the major light-harvesting complex of photosystem II (LHCII) with their genes in tomato*

Beverley R. Green[a], Dingren Shen[a], Ruedi Aebersold[b] and Eran Pichersky[c]

[a]*Botany Department, University of British Columbia, BC, Canada* [b]*Biomedical Research Centre and Department of Biochemistry, University of British Columbia, Vancouver, BC, Canada* and [c]*Biology Department, University of Michigan, Ann Arbor, MI, USA*

Received 17 March 1992; revised version received 7 May 1992

Using an improved SDS-PAGE system, the polypeptides of the major chlorophyll *a/b* light-harvesting complex of PSII (LHCII) from tomato leaves were resolved into five polypeptide bands. All the polypeptides were matched with the genes encoding them by comparing amino acid sequences of tryptic peptides with gene sequences. The two major LHCII bands (usually comigrating as a '27 kDa' polypeptide) were encoded by *cab*1 and *cab*3 (Type I LHCII) genes. A third strong band of about 25 kDa was encoded by *cab*4 (Type II) genes. Polypeptides from two minor bands of 23–24 kDa were not N-terminally blocked; their N-terminal sequences showed they were Type III LHCII proteins. One complete cDNA clone and several incomplete clones for Type III polypeptides were sequenced. Combined with the peptide sequences, the results indicate that there are at least four different Type III genes in tomato, encoding four almost identical polypeptides. Thus, all the LHCII CAB polypeptides have been identified, and each type of LHCII polypeptide is encoded by distinct gene or genes in tomato.

Chlorophyll *a/b*-(CAB)protein; gene, Type III LHCII; Light-harvesting antenna; *Lycopersicon*

1. INTRODUCTION

LHCII is the major chlorophyll (Chl) *a/b* light-harvesting complex of green plants, accounting for up to 50% of the total Chl in the thylakoid membrane [1]. Although this Chl–protein complex has been extensively studied since its discovery more than 25 years ago, there is still some question about the number of polypeptides it contains, with estimates ranging from 2 to 6 [1,2]. Some of the confusion results from different definitions of LHCII which in turn are the result of different isolation procedures [2]. Our definition of LHCII is operational: based on the fact that LHCII can be precipitated by divalent cations even in the presence of detergents [3] and that one of its oligomeric forms can be isolated as a Chl–protein complex (CPII*) on mildly-denaturing SDS-PAGE [4,5]. Thus it does not include the Chl *a/b* complexes CP29 and CP24 which are also associated with PSII [6]. Gene sequences of three distinct types of LHCII genes have been reported [7–9], but it has not been clear which polypeptide corresponded to which gene product. It has also been suggested that some of the multiple bands observed on SDS-PAGE could be due to alternative processing of a single precursor [10,11]

On most gel systems, LHCII has two major polypeptides of about 27 and 25 kDa. The latter is enriched in the 'mobile' LHCII which migrates to the stroma lamellae in response to changes in illumination or temperature (reviewed in [12]). In addition, there is often a third minor band in LHCII [5,13]. Using very long gels containing 4 M urea, we have been able to resolve tomato LHCII polypeptides that differ by less than 1 kDa in molecular weight, and correlate each one with the respective gene type by tryptic peptide sequencing. In the process, we have found that there are at least four Type III genes in tomato, giving rise to two separable polypeptides of 23–24 kDa. A preliminary report on the first Type III gene sequence from tomato has been published [7].

2. MATERIALS AND METHODS

Chloroplasts were isolated from greenhouse-grown tomato (*Lycopersicon esculentum* var. Best of All). Thylakoids were washed several times with 10 mM Tricine-NaOH, 1 mM EDTA, pH 8.0 containing the protease inhibitors phenylmethylsulfonyl fluoride (1 mM), *p*-aminobenzamidine (6 mM) and aminocaproic acid (40 mM). LHCII was isolated according to [3] as modified by [14]; oxygen-evolving PSII reaction centre cores ('G&Ys') according to [6].

Samples were denatured by heating to 80°C in 2% SDS, 65 mM Tris-HCl, pH 6.5, 50 mM dithiothreitol, 20% glycerol and the polypeptides separated by electrophoresis on 14% polyacrylamide gels containing 0.8 M Tris-HCl, pH 8.8 and 4 M urea with a 2 cm long stacking gel. Gels were run for 21–24 h at 4°C, until the buffer front had moved about 30 cm. Polypeptides were either stained with

*Dedicated to Professor O. Machold on the occasion of his retirement.

Abbreviations: Chl, chlorophyll; CAB, chlorophyll *a/b*-binding; LHCII, major light-harvesting complex of Photosystem II

Correspondence address: B.R. Green, Botany Department, University of British Columbia, Vancouver, BC, Canada.

Published by Elsevier Science Publishers B.V.

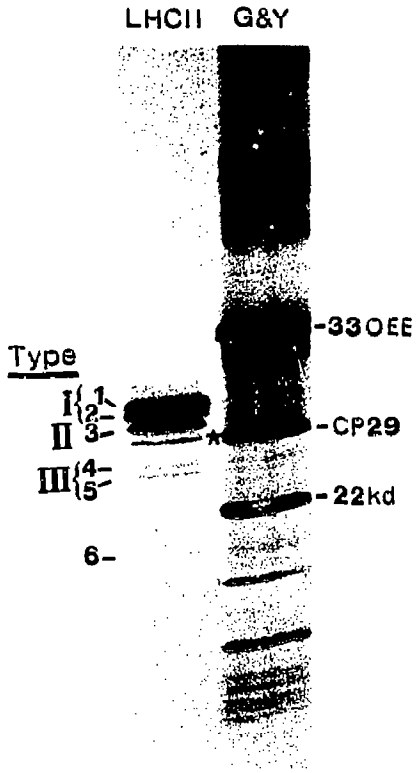

Fig. 1. Tomato LHCII polypeptides (left lane) resolved on 14% acrylamide-4 M urea gels and Coomassie-stained. Type I, II, III: gene types encoding the bands. Asterisk (*) marks a small amount of CP29 contamination (the two polypeptides of tomato CP29 run together on urea-containing gels). G&Y, Oxygen-evolving PSII preparation (right lane) showing positions of CP29, 33 kDa OEE (oxygen-evolving enhancer) and intrinsic 22 kDa polypeptides of Photosystem II.

Coomassie blue or electro-transferred to nitrocellulose or polyvinylidene difluoride (Immobilon P) membranes prior to amino acid sequencing as in [15]. Cloning and nucleotide sequencing of Type III genes were carried out as in [7].

3. RESULTS

3.1. *Identification of major polypeptides with their genes*

A sample of tomato LHCII cation, precipitated from detergent-solubilized PSII membranes is shown in Fig. 1. For comparison purposes, a sample of PSII reaction centre cores depleted of LHCII ('G&Y') was run in the next lane. The identity of the minor CP29 contaminant in LHCII (asterisk) was confirmed by immunoblotting and peptide sequencing [6,15]. Tryptic peptide sequences obtained from the other six bands are given in Fig. 2.

Two tryptic peptides were obtained from the combined Bands 1 and 2. One matched the deduced amino acid sequence from Type I but not Type II or III genes (Fig. 2). The Type I polypeptides are encoded by two gene clusters (*cab*1 and *cab*3) that are located on two different chromosomes [16] and encode polypeptides which differ in only 8 positions out of 232 in the mature

Band 1 and 2	
Tryptic Peptide:	?GESPSYLTGEFPGDYGWDT
Type I gene (*cab*1B/3C	*SGESPSYLTGEFPGDYGWDT*
Type II gene (*cab*4,5)	*SEQTPSYLTGEFPGDYGWDT*
Type III gene (*cab*13)	*SAQTPSYLNGEFPGDYGWDT*
Tryptic Peptide:	FGEAVWFK
Type I or II gene	FGEAVWFK
Type III gene (*cab*13)	FKEPVWFK
Band 3	
Tryptic Peptide	GPIENLSDHINDPVANNA
Type II gene (*cab*4)	GPIENLSDHINDPVANNA
Type II gene (*cab*5)	GPIENLSDHIADPVANNA
Type I gene (*cab*1B,3C)	GPLENLADHLADPVNNA
Type III gene (*cab*13)	GPLENLLDHLDNPVANNA
Band 4	
N-terminal protein sequence	XNDLXYGPDXVXYL
Tryptic Peptide	NDLXYGPD
Type III gene (*cab*13)	SNDLWYGPDRVKYL
Tryptic Peptide	SAQTPSYL (40%)
	SAQTPNFL (60%)
Type III gene (*cab*13)	SAQTPSYL
Type I gene (*cab*1B,3C)	SGESPSYL
Type II gene (*cab*4,5)	SEQTPSYL
Tryptic Peptide	INGLPGVGEGNDLYPGGQYFDPL
Type III gene (*cab*13)	INGLPGVGEGNDLYPGGQYFDPL
Type I gene (*cab*1B,3C)	IAGGPLGEVVDPLYPGGS.FDPL
Type II gene (*cab*4,5)	VGGGPLGEGLDKIYPGGA.FDPL
Tryptic Peptide	FVPGA
Type III gene (*cab*13)	KFVPGA
Type I or II gene	NFVPGK
Band 5	
N-terminal protein sequence	XNDLWYGPDXVXYL
Type III gene (*cab*13)	SNDLWYGPDRVKYL
Tryptic Peptide	FVPGS
Type III gene (*cab*13)	KFVPGA
Type I or II gene	NFVPGK
Tryptic Peptide	YLGPFSAQTP
Tryptic Peptide	.....SAQTPSYLTGE
Type III gene (*cab*13)	YLGPFSAQTPSYLNGE
Type I gene (*cab*1B,3C)	YLGPFSGESPSYLTGE
Type II gene (*cab*4,5)	YLGPFSEQTPSYLTGE
Band 6 (CP24)	
Tryptic Peptide	SXIPAV
CP24 gene (*cab*10A)	KSWIPAV
Type I gene (*cab*1B,3C)	SPWYGPD
Type II gene (*cab*4,5)	SIWYGED
Type III gene (*cab*13)	DLWYGPD

Fig. 2. Comparison of tryptic and N-terminal peptide sequences of tomato LHCII polypeptides with sequences deduced from CAB genes. Bands 1 and 2 are the Type I polypeptides (see Fig. 1); Band 3, the Type 2 polypeptides; Bands 4 and 5, the Type III polypeptides; Band 6, the CP24 polypeptide. Boldface: identical amino acids; italic, non-identical; X or ?, not unambiguously determined. A complete alignment of all tomato CAB polypeptides except those encoded by Type III genes is given in [22].

protein. Assuming that their precursors are cleaved at the same position, the mature *cab*3 polypeptides are two amino acids longer and have one more positively-charged amino acid than *cab*1 polypeptides. Since most of the variant amino acids are located near the mature N-terminus, which is blocked, we did not attempt to determine which of Bands 1 and 2 corresponded to which gene. However, this tryptic peptide sequence is not found in any other type of CAB gene. A second tryptic peptide with the sequence FGEAVWFK is found in both Type I and Type II, but not in Type III gene sequences.

A

		Peptide near N-terminus	Peptide at C-terminus	Gene
BAND 4:	40%	SAQTPSYL(N)	FVPGA	*cab*13
	60%	SAQTPNFL		*cab*14
BAND 5		SAQTPSYLTGE	FVPGS	*cab*15, *cab*16

B

```
          V  K  Y  L  G  P  F  S  A  Q  T  P  S  Y  L  T  G  E  F  P  G  D
cab15:    GTCAAGTACTTGGGACCATTTTCTGCTCAAACTCCTTCATACTTGACTGGAGAATTCCCTGGTGAT
cab16:     T              C                                    G           C
```

Fig. 3. The four LHCII Type III genes of tomato. (A) Correspondence of the distinctive tryptic peptide sequences from polypeptide Bands 4 and 5 with the completely sequenced gene (*cab*13), partially-sequenced PCR clones (*cab*15, *cab*16) and gene whose existence was deduced from protein sequence (*cab*14). In Band 4, (N) is known only from the DNA sequence (see text). (B) Nucleotide sequence determined for PCR products corresponding to *cab*15 and *cab*16 and deduced amino acid sequence.

Band 3 is uniquely identified as the product of *cab*4, one of the two Type II CAB genes in tomato [17]. Since the two Type II genes, *cab*4 and *cab*5 are almost identical, their products would be expected to co-migrate. Our data suggest that *cab*4 is more highly expressed than *cab*5 in tomato leaves. It also supports the earlier conclusion [17] that in tomato the Type II polypeptide(s) have an apparent molecular weight lower than Type I polypeptides on SDS-polyacrylamide gels.

Band 6 was identified as the product of *cab*10A/B which encode the polypeptide of CP24 [18]. This amino acid sequence is not found in any other CAB gene and there is no cleavage site for trypsin at the homologous site in any of the LHCII polypeptide sequences.

3.2. *Type III Polypeptides*

Both Bands 4 and 5 had unblocked N-termini (Fig. 2), and the sequences obtained from these termini were very similar to each other and to those reported for the unblocked minor CAB polypeptides of wheat [19], barley [20], *Arabidopsis* [21] and corn (R. Bassi, pers. comm.). This type of CAB polypeptide has been designated LHCII Type III [20]. Despite the high level of similarity of the N-terminal sequences from Bands 4 and 5 of tomato LHCII, distinctive differences were found in a region close to the N-terminus (Figs. 2 and 3A). Two variants of a sequence motif, SAQTPSYL and SAQTPNFL (in 40% and 60% proportions respectively) were found in Band 4. These data show that Band 4 contains two different polypeptides encoded by at least two distinct genes. Although one of the sequences from Band 4 appears to be identical to the corresponding sequence of Band 5 (SAQTPSYLTGE), the C-terminal sequences obtained from proteins in each of these two bands were different (Figs. 2 and 3A), suggesting that the Type III protein in Band 5 was encoded by yet another gene(s).

Comparison of the peptide sequences obtained from Band 4 and Band 5 proteins with the predicted sequence of the protein encoded by *cab*13, a tomato gene recently isolated and characterized [7], indicates that this gene encodes a polypeptide found in Band 4. The predicted sequence is in complete agreement with the band 4 C-terminal sequence, a 23-residue internal sequence, the SAQTPSYL sequence which comprises 40% of the signal in that tryptic peptide, and the N-terminal sequence. This means that the SAQTPNFL sequence also found in Band 4 must have come from another protein, encoded by an as yet uncloned gene which we have designated *cab*14. It is possible the N-terminal tryptic peptide (which would be lacking a Ser if it came from the protein encoded by *cab*13) was encoded by *cab*14, but it is also possible that the first amino acid was post-translationally cleaved.

Using total DNA from a cDNA library of tomato leaf tissue as a template, we have performed a Polymerase Chain Reaction (PCR) [15] using appropriate oligonucleotides, to amplify DNA sequences encoding the N-terminal portion of LHCII Type III polypeptides. Two PCR products were cloned into plasmids and their nucleotide sequences determined (Fig. 3B). The two clones differ in their nucleotide sequences but they specify the same amino acid sequence. This sequence includes the peptide sequence SAQTPSYLTGE unique to Band 5. Note that the corresponding sequence in Band 4 must have N instead of T in the ninth position. Thus, the two PCR clones are derived from two additional LHCII Type III genes which we have designated *cab*15 and *cab*16, and these genes most likely encode the Band 5 proteins. Within the cloned and sequenced region

(which extends 12 base pairs upstream and 30 base pairs downstream of the sequence given in Fig. 3B) the nucleotide sequences of the *cab*15 and *cab*16 genes differ from the *cab*13 sequence at additional sites besides the N/T codon, but these differences do not lead to any change in amino acid sequence (comparison not shown).

4. DISCUSSION

In this paper, we have related the members of the CAB gene family encoding the LHCII polypeptides to their respective polypeptides. In other published work, we have identified most of the other members of this extended family of proteins in tomato. (Table I) [7,15–18,22–25]. Our data indicate that each of the separable polypeptides is a different type. Most of the resolvable polypeptides are encoded by one or two genes, with the exception of the closely related Type I polypeptides. (Table I)

Tomato is unusual in having two separable Type III polypeptides, each of which appears to be encoded by two different genes. Spinach, barley and *Brassica napus* had only one Type III Band on our gel system (data not shown). In all plants so far investigated, the Type III polypeptide is a part of CPII*, the oligomeric form of LHCII isolated on mildly denaturing SDS-PAGE, although the amount relative to Types I and II polypeptides depends on the detergent concentration used in the initial solubilization [13,26].

We still have much to learn about the organization of the different components of the light-harvesting complex associated with PSII. We do not even know if all its polypeptides bind the same ratio of Chl *a* to *b*. An appreciable amount of Type III LHCII polypeptide was found by immunoblotting thylakoid membranes of the Chl *b*-less barley mutant chlorina f2 [27] and in intermittent-light-grown barley which has very low levels of Chl *b* [28], both of which have very reduced amounts of Types I and II polypeptides. This suggests that Type III polypeptides may bind less Chl *b* than Types I and II, or for some other reason are not as susceptible to turnover in the absence of Chl *b*. Morissey et al. [29] found that the Type III polypeptide was assembled early in thylakoid development in soybeans raised under conditions inhibiting the development of the full light-harvesting antenna, and that it was maintained at a fairly constant level per PSII unit. A barley Type III gene is expressed in dark-grown seedlings, in contrast to other CAB genes [30]. These observations suggest that the Type III polypeptide(s) may play a special role in the development of the full light-harvesting apparatus or may act as a linker to join one or more units containing the major LHCII polypeptides to the PSII core.

REFERENCES

[1] 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.
[2] Green, B.R. (1988) Photosyn. Res. 15, 3–32.
[3] Burke, J.J., Ditto, C.L. and Arntzen, C.J. (1978) Arch. Biochem. Biophys. 187, 252–263.
[4] Hiller, R.G., Genge, S. and Pilger, D. (1974) Plant Sci. Lett. 2, 239–242.
[5] Machold, O. (1981) Biochem. Physiol. Pflanzen. 176, 805–827.
[6] Camm, E.L. and Green, B.R. (1989) Biochim. Biophys. Acta 974, 180–184.
[7] Schwarz, E., Siasys, R., Aebersold, R., McGrath, J.M., Green, B.R. and Pichersky, E. (1991) Plant Mol. Biol. 17, 923–925.
[8] Jansson, S. and Gustafsson, P. (1991) Mol. Gen. Genet. 229, 67–76.
[9] Chitnis, P.R. and Thornber, J.P. (1988) Photosyn. Res. 16, 41–63.
[10] Kohorn, B.D. and Tobin, E.M. (1986) Plant Physiol. 82, 1172–1174.
[11] Clark, S.E., Abad, M.S. and Lamppa, G.K. (1989) J. Biol. Chem. 264, 17544–17550.
[12] Anderson, J.M. and Andersson, B. (1988) Trends Biochem. Sci. 13, 351–355.
[13] Machold, O. (1986) Carlsberg Res. Commun. 51, 227–238.
[14] Ryrie, I.J., Anderson, J.M. and Goodchild, D.J. (1980) Eur. J. Biochem. 107, 345–354.
[15] Pichersky, E., Subramaniam, R., White, M.J., Reid, J., Aebersold, R. and Green, B.R. (1991) Mol. Gen. Genet. 227, 277–284.
[16] Pichersky, E., Bernatzky, R., Tanksley, S.D., Breidenbach, R.B., Kausch, A.P., Cashmore, A.R. (1985) Gene 40, 247–258.
[17] Pichersky, E., Hoffman, N.E., Malik, V.S., Bernatzky, R.,

Table 1

Tomato chlorophyll *a/b* proteins and their genes

Complex	Role/location	Chl *a/b* ratio	Polypeptides	Gene types	No. of gene copies	No. of introns	Ref.
LCHII	major antenna PSII	1.2	2 major	Type I	8	0	16
			1	Type II	2	1	17
			2	Type III	4+	2	7
CP29	core antenna PSII	4–5	1	Type I	1	5	15
			1	Type II	(n.c.)	–	–
CP24	minor PSII antenna	<1	1	–	2	1	18
LHCI	PSI antenna	3–5	4	Type I	2	3	23
				Type II	1	4	24
				Type III	1	2	25
				Type IV	2	2	22

n.c. = not cloned yet.

Tanksley, S.D., Szabo, L. and Cashmore, A.R. (1987) Plant Mol. Biol. 9, 109–120.
[18] Schwartz, E. and Pichersky, E. (1990) Plant Mol. Biol. 15, 157–160.
[19] Webber, A.N. and Gray, J.C. (1989) FEBS Lett. 249, 79–82.
[20] Morishige, D.T. and Thornber, J.P. (1990) Curr. Res. Photosyn. 2, 261–264.
[21] Morishige, D.T. and Thornber, J.P. (1991) FEBS Lett. 293, 183–187.
[22] Schwartz, E., Shen, D., Aebersold, R., McGrath, J.M., Pichersky, E. and Green, B.R. (1991) FEBS Lett. 280, 229–234.
[23] 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.
[24] Pichersky, E., Tanksley, S.D., Piechulla,B., Stayton, M.M. and Dunsmuir, P. (1988) Plant Mol. Biol. 11, 69–71.
[25] 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.
[26] Machold, O. (1991) J. Plant Physiol. 138, 678–684.
[27] White, M.J. and Green, B.R. (1987) Eur. J. Biochem. 165, 531–535.
[28] White, M.J. and Green, B.R. (1988) Photosyn. Res. 15, 195–203.
[29] Morrisey, P.J., Glick, R.E. and Melis, A. (1989) Plant Cell Physiol. 30, 335–344.
[30] Brandt, J., Nielsen, V.S., Thordal-Christensen, H., Simpson, D.J. and Okkels, J.S. (1992) Plant Mol. Biol. (in press).