Volume 293, number 1,2, 183–187 FEBS 10414 November 1991

Correlation of apoproteins with the genes of the major chlorophyll *a/b* binding protein of photosystem II in *Arabidopsis thaliana*

Confirmation for the presence of a third member of the LHC IIb gene family

Daryl T. Morishige and J. Philip Thornber

Department of Biology, University of California, Los Angeles, CA 90024-1606, USA

Received 3 August 1991; revised version received 1 October 1991

The major light-harvesting complex in higher plants is LHC IIb. The LHC IIb of *Arabidopsis thaliana* contains 2 pigment-binding apoproteins of 28 and 25 kDa. To determine the relationship between them and the LHC IIb gene family members, each protein was purified to homogeneity, subjected to direct protein sequencing, and the sequences compared with those deduced from LHC IIb genes in this organism. The 28 kDa protein is the product of Type I LHC IIb genes. The 25 kDa LHC IIb component is distinctly different from the 28 kDa LHC IIb protein, and is more closely related to the type III LHC IIb gene product of barley. Type III gene products lack the first 9–11 residues found in proteins of the Type I and II genes, a region that contains a phosphorylatable threonine residue. The lack of the N-terminal residues explains why this LHC IIb apoprotein has never been seen to be phosphorylated, and partly or wholly why it is smaller. The implication of the missing N-terminus on uptake of LHC II precursor proteins into the plastid and of the relative organization of the LHC IIb subunits in the PS II antenna is discussed.

Light-harvesting pigment-protein complex; Protein sequencing; *cab* gene; Photosynthesis; Membrane protein; *Arabidopsis thaliana*

1. INTRODUCTION

The major light-harvesting pigment-protein of higher plant chloroplasts is LHC IIb, whose chlorophyll–carotenoid binding apoproteins are associated with up to 43% of the total chlorophyll in normal plants [1]. In vivo the pigmented LHC IIb complex occurs as a trimer and/or as a higher order oligomer bound to CC II [1,2]. The oligomeric LHC IIb of approximately 72 kDa contains as many as 5 protein subunits of 29–25 kDa in some species, presumably all of which are apoproteins [1,3]. While the LHC IIb apoproteins are biochemically similar, their origin remains equivocal. The multiple proteins could be derived from a single precursor polypeptide which is processed to different sizes upon import into the chloroplast as observed in vitro [4–7] or to different LHC IIb genes being expressed [8]. It also remains possible that multiple LHC IIb gene products are each processed to different sizes upon import into the chloroplast.

In diverse plant species the genes coding for the LHC IIb apoproteins can be grouped into small nuclear gene families from approximately 3–16 members [9,10]. The *A. thaliana* genome contains the smallest (3–4) number of genes yet observed [10]. While their deduced protein sequences are very similar (see e.g. [11]), the LHC IIb genes themselves have been divided into two classes: Type I genes which contain no introns and Type II genes which contain one small intron [12; see also 4,13]. Similarities among the deduced LHC IIb Type I and Type II protein sequences are greater within either group than between the 2 groups. Although different species contain both Type I and II genes, the functional significance of the different apoproteins encoded by the different genes is unknown.

In barley the LHC IIb complex apparently contains at least 3 apoproteins of 28, 27 and 25 kDa, that are immunologically related, yet distinct proteins [1,14]. The protein sequence of the first 59 residues of the 25 kDa apoprotein has been determined [14]. The sequence is 11 or 9 residues shorter at its N-terminus than that of the predicted mature proteins of the Type I or II LHC IIb genes, respectively; its N-terminal sequence contains 3 regions of relatively high divergence from the other 2 types. The differences indicate that the gene coding for the 25 kDa protein is wholly distinct from either the Type I or II LHC IIb genes and we termed it a Type III LHC IIb gene [14]. Prior to this, Webber and Gray [15] had reported that the N-terminal sequence of the wheat 25 kDa LHC IIb protein corresponded to the sequence of a typical Type II LHC IIb gene, but a lack of simila-

Abbreviations: LHC, light-harvesting complex; CC, core complex; PS, photosystem; CNBr, cyanogen bromide; PVDF, polyvinylidene fluoride

Correspondence address: D.T. Morishige, Department of Biology, University of California, Los Angeles, CA 90024-1606, USA. Fax: (1) (213) 206 4386.

rity with the characteristic Type II residues would indicate that this is not the case. Correlation of the other barley LHC IIb subunits with their genes is more difficult since the LHC IIb gene family in barley appears to have 8–10 gene family members (Morishige and Thornber, unpublished).

While a large number of LHC IIb genes have been identified in several plants, it is not presently clear how the expression of the LHC IIb gene family members correlates to the LHC IIb proteins observed in vivo in any plant species. *A. thaliana* appears to be the best plant species to make the initial investigations. In *A. thaliana* 3 of the possible 4 gene family members have been sequenced and they code for polypeptides of the same sequence except for 1 amino acid difference in the transit sequence of 1 of the 3 genes [10]. Each of the genes is apparently expressed at the transcriptional level [16], but it is not known how their expression is correlated to the LHC IIb proteins observed in vivo. Although the 3 genes appear to code for an identical Type I mature protein, the oligomeric LHC IIb can be resolved by SDS-PAGE into at least 2 different apoproteins, as in many other plant species. Since the *A. thaliana* LHC IIb gene family is rather small and apparently contains 3 Type I but no Type II LHC IIb genes [10], this plant species should be particularly useful for determining the sequences of the multiple LHC IIb apoproteins found in vivo in order to determine their realtionship to their genes. To this end, we have purified the 25 and 28 kDa apoproteins of LHC IIb from *A. thaliana* to homogeneity and subjected each to direct amino acid sequencing. Our results indicate that the 25 kDa protein has homology to the Type III LHC IIb gene product of barley whose gene(s) has yet to be isolated, whereas partial sequences of the 28 kDa subunit indicate that this apoprotein is derived from the LHC IIb Type I gene sequences previously determined for *A. thaliana* [10].

2. MATERIALS AND METHODS

Arabidopsis thaliana var. Columbia was germinated and grown in a vermiculite-soil mixture in a greenhouse under natural light conditions for 5 weeks. Whole thylakoids were isolated [17] and PS II-enriched membrane preparations, lacking the oxygen-evolving complex, were prepared [18]. Detergent solubilization of thylakoids and fractionation of the extracts by non-denaturing polyacrylamide gel electrophoresis (PAGE) [19] yielded the oligomeric LHC IIb fraction, from which the LHC IIb apoproteins were obtained by fully-denaturing SDS-PAGE (see [18] for details).

Internal peptide fragments were obtained by cleavage of the 28 or 25 kDa LHC IIb apoproteins with CNBr. The isolated proteins were diluted in 98% formic acid so that the final concentration of formic acid was 70%. β-Mercaptoethanol was added to a final concentration of 0.01% (v/v). CNBr (2–3 small crystals; 5–10 mg, Pierce Chemical Co.) was added to the reaction mixture (typically 500 μl) and cleavage was allowed to proceed overnight in the dark at room temperature. The reaction mixture was dried in a Speed-vac (Savant), washed with distilled water and re-dried.

The resulting CNBr-derived peptides were resuspended in 62 mM Tris-HCl, pH 6.8, 1% (v/v) β-mercaptoethanol, 1% (w/v) SDS, 1% (v/v) glycerol and heated for 5 min at 80°C. Peptide fragmetns were separated on a 15% polyacrylamide gel with modifications as described in Morishige and Thornber [18] using a Mighty Small II electrophoresis apparatur (Hoeffer, San Francisco) and then transferred to polyvinylidene fluoride membranes (Immobilon-P; Millipore Corp.) in 10 mM 3-[cyclohexylamino]-1-propanesulfonic acid (CAPS) buffer, pH 11.0, containing 10% methanol [20] in a Mighty Small Transphor unit (Hoeffer). After lightly staining the PVDF membrane with Coomassie blue and destaining, the detected peptide bands were excised and subjected to N-terminal sequencing using an Applied Biosystems 470A gas-phase sequencer.

All electrophoresis chemicals were ultra-pure or electrophoresis grade. Chemicals and detergents used for isolation of the apoproteins were reagent grade. Silver staining of gels was carried out using a Gel-Code silver staining kit (Pierce).

3. RESULTS AND DISCUSSION

We have isolated from *A. thaliana* the individual apoproteins of the major light-harvesting complex LHC IIb associated with photosystem II. The oligomeric LHC IIb pigmented complex was obtained by non-denaturing PAGE of either detergent solubilized thylaloids or PS II-enriched thylakoid extracts. Separation of the individual apoproteins of the LHC IIb oligomer was achieved by fully-denaturing SDS-PAGE. Regardless of its initial source (either whole thylakoids (Fig. 1A; lane 1) or PS II-enriched membrane preparations (Fig. 1A; lane 2)), the purified LHC IIb had an identical apoprotein composition (Fig. 1A; lane 3); no major protein contaminants were apparent. The oligomeric LHC IIb appeared to contain 2 apoprotein subunits of approximately 28 and 25 kDa (Fig. 1A; lane 3) each of which was purified to homogeneity (Fig. 1A; lanes 4 and 5).

In order to clearly understand the relationship of the LHC IIb gene family members and the individual LHC IIb polypeptides encoded by them, partial amino acid sequences of the 28 and 25 kDa polypeptide subunits were obtained. Initial attempts to sequence either of the LHC IIb subunits from their native amino-termini were unsuccessful. This is in direct contrast to the 25 kDa LHC IIb subunit from barley which had an unmodified N-terminus [14]; however, the remaining LHC IIb apoproteins of barley also have blocked N-termini. The N-terminal modification is most likely due to an acetyl group being added post-translationally to the protein, as observed for LHC IIb proteins in other species [21].

Because of the modified N-termini, the LHC IIb proteins were cleaved with CNBr, producing smaller peptides that were ultimately sequenced. The cleavage products were separated by SDS-PAGE and the resulting peptide maps showed that the Met residues of the 28 and 25 kDa subunits were distributed differently (Fig. 1B; compare lanes 2 and 4). The 28 kDa LHC IIb yielded major peptide fragments at 23, 15.5, 9.5, 6.5 and 5.5 kDa (Fig. 1B; lane 2). Cleavage of the 25 kDa protein resulted mainly in peptide fragments of 13.5, 7.5 and 5.5 kDa (Fig. 1B; lane 4). Thus, the difference in the distribution of Met residues suggests that the 2 LHC

IIb proteins are the products of substantially different genes. In contrast, it would be expected that CNBr treatment of the 28 and 25 kDa subunits would produce very similar peptide maps if they corresponded to the Type I and II LHC IIb gene products, because of the spatial conservation of the Met residues in the Type I and II sequences (see e.g. [11]). The CNBr-derived peptide fragments separated by SDS-PAGE were transferred to PVDF membranes. Three distinct bands from each of the cleavage reactions were apparent after staining with Coomassie blue (data not shown) corresponding to fragments of 28, 15.5 and 9.5 kDa for the 28 kDa subunit and 25, 13.5 and 7.5 kDa for the 25 kDa subunit on the silver-stained gel (Fig. 1B).

Sequencing of an internal fragment from the 28 kDa subunit yielded an amino acid sequence that was essentially identical to the 3 derived from the published gene sequences of *A. thaliana* LHC IIb [10]. Approximately 150 pmol of the 15.5 kDa peptide fragment of the 28 kDa LHC IIb protein was sequenced (initial yield 16%; repetitive yield 94%) and a sequence of 32 residues was obtained. This sequence starts at residue 74 of the published translated gene sequence and is identical to it (Fig. 2A). An additional minor sequence of 22 residues was obtained at the same time which could be aligned with the derived protein sequence beginning at residue 136 of the Type I gene (Fig. 2B). In both cases the 2 internal sequences obtained from the 28 kDa fragments are directly preceded by Met residues in the sequence predicted from the gene, as would be expected. Thus, the 28 kDa LHC IIb apoprotein is the apparent product of the Type I gene(s) from *A. thaliana* [10].

Despite the great similarity of the Type I and II gene products (see e.g. [11]), a few amino acid residues throughout the deduced protein sequences can be used to distinguish a Type I from a Type II LHC IIb protein. For example, an Arg residue occurs at position 87 in all Type I mature proteins, while a Lys is found at the equivalent position in all the Type II proteins (see e.g. [11]). Similarly, an Ala residue occurs at position 144 in the Type I sequence and a Gly in the Type II sequence. The partial amino acid sequences of the 28 kDa subunit (Fig. 2A and 2B) indicate that at these positions only the characteristic Type I residues (Arg and Ala at residues 87 and 144, respectively) are present (Fig. 2). Although the relative amounts of Type II residues at these positions could have been below detection during sequencing, these results would indicate that the only protein represented in the major 28 kDa protein band is that derived from a Type I LHC IIb gene and that a Type II gene does not occur or is not expressed in *A. thaliana*. The presence of a Type I, but not Type II, gene product correlates with the individual LHC IIb genes isolated from *A. thaliana*, which encode the same mature Type I proteins [10]. Only Type I LHC IIb genes were isolated using a Type II gene (*ab19* from *Lemna gibba* [13]) as the probe. The fourth putative LHC IIb gene, corresponding to a weakly hybridizing band on genomic DNA blots [10], has not yet been isolated and could be that of the 25 kDa LHC IIb component (see below).

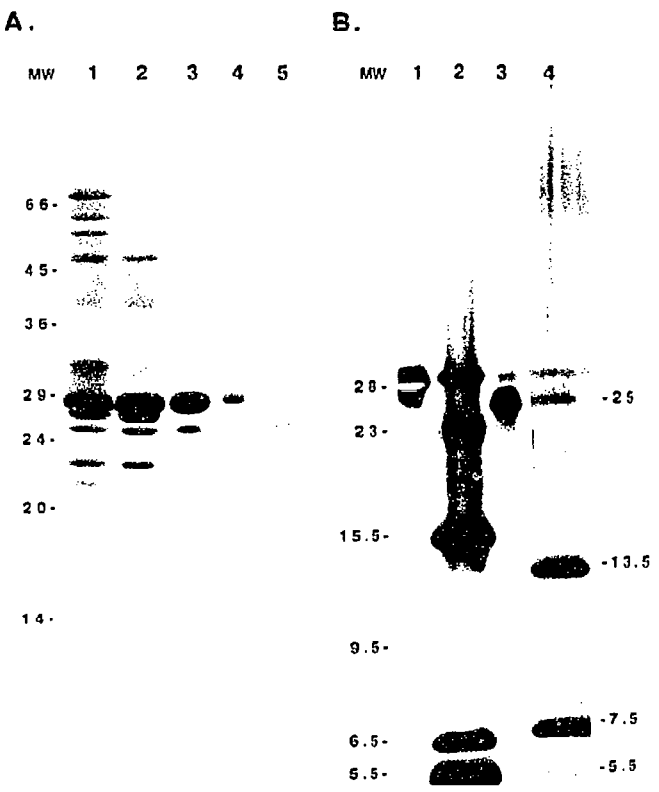

Fig. 1. SDS-PAGE analysis of *A. thaliana* LHC IIb proteins. (A) Coomassie blue stained gel of: (1) whole thylakoids; (2) PS II-enriched membrane fraction; (3) LHC IIb oligomer isolated from non-denaturing gels; (4) purified 28 kDa LHC IIb apoprotein; (5) purified 25 kDa LHC IIb apoprotein. (B) Silver stained gel of CNBr cleavage products of the 28 kDa (lane 2) and 25 kDa (lane 4) LHC IIb apoproteins. Untreated 28 and 25 kDa proteins are in lanes 1 and 3, respectively.

Approximately 100 pmol of the 13.5 kDa peptide fragment from the 25 kDa LHC IIb protein was sequenced (initial yield 37%; repetitive yield 88%) and its first 24 amino acid residues were obtained (Fig. 2A). This partial sequence gave a reasonable but far from identical match to the only known amino acid sequences (residues 74–96) predicted from the Type I gene sequences for LHC IIb in *A. thaliana* [10] and to the major sequence of the 15.5 kDa peptide fragment from the 28 kDa LHC IIb subunit; optimal alignment is observed at only 11 of 24 (46%) residues (Fig. 2A). In accordance with the putative model for the folding of LHC IIb in the membrane [13], residues 74–97 fall partly within the first membrane-spanning region and are partly exposed to the thylakoid lumen. The region around the first trans- membrane span, including the Met corresponding to residue 73 in the derived *A. thaliana* LHC IIb sequence, is highly conserved among all LHC proteins [22]. This Met residue appears to be highly susceptible to cleavage with CNBr, since protein sequences originating at this same Met residue have now been

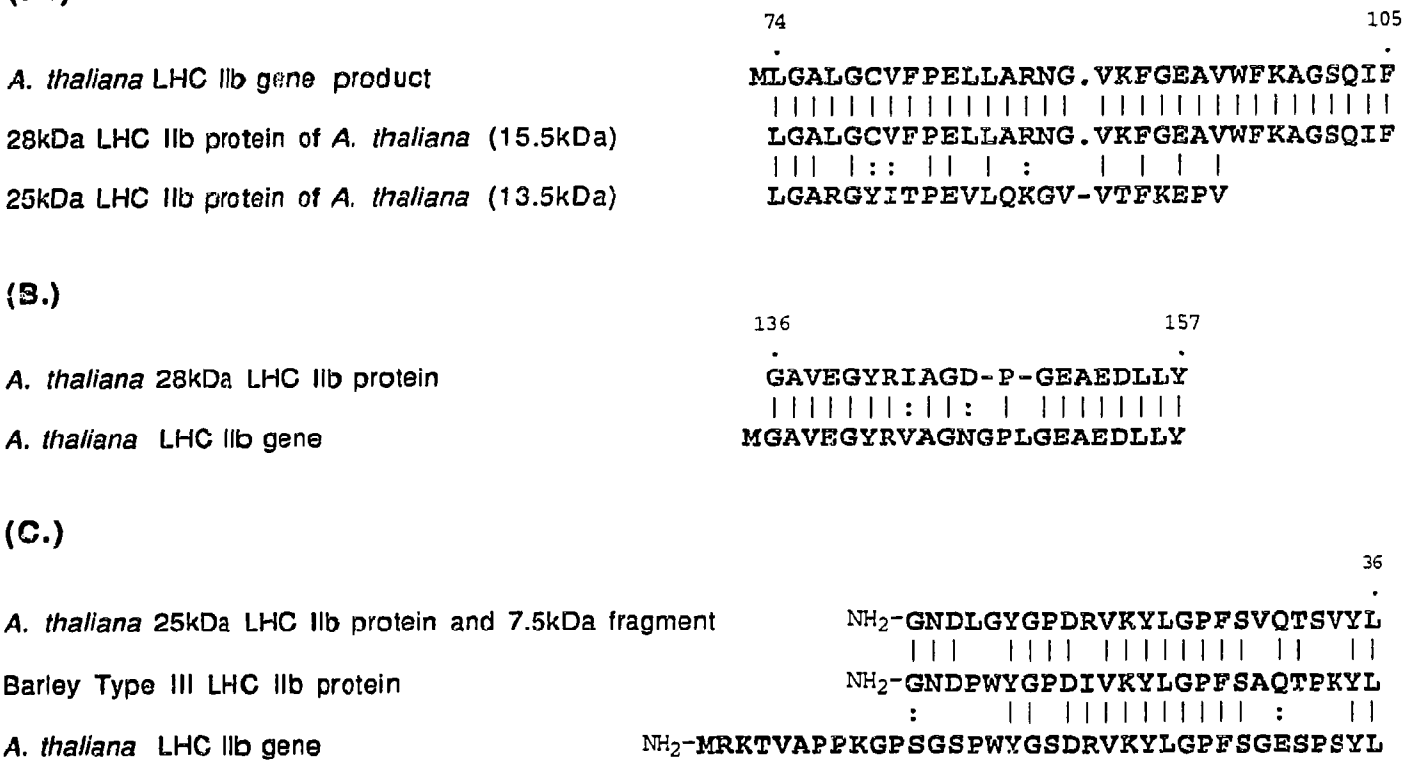

Fig. 2. Direct protein sequences from LHC IIb apoproteins and CNBr-derived peptide fragments. (A) Amino acid sequences from the 15.5 and 13.5 kDa CNBr-derived peptide fragments from the 28 and 25 kDa LHC IIb apoproteins, respectively, and compared to the derived protein sequence from the *A. thaliana* genes [10]. (B) Minor protein sequence from the CNBr-derived 15.5 kDa peptide of the 28 kDa LHC IIb protein and alignment with a region from the *A. thaliana* gene (residues 136–157). (C) N-terminal amino acid sequence from the 25 kDa LHC IIb protein and 7.5 kDa peptide fragment from the *A. thaliana* compared to the Type III 25 kDa protein from barley [14] and the deduced sequence of the mature protein from the *A. thaliana* gene. (|) indicates homology, (:) indicates a conserved amino acid difference with the uppermost sequence in each panel; (-) defines an undetermined amino acid residue; (·) within a sequence indicates a gap to produce an optimal alignment. Numbering refers to the position from the putative cleavage site for the mature protein sequences derived from the LHC IIb genes.

obtained from the 28 and 25 kDa LHC IIb proteins and from 2 other different LHC II proteins treated with CNBr (Morishige and Thornber, in press). Because of the high degree of conservation in this region, protein sequence data from this region are useful in identifying and comparing the different LHC proteins.

Sequencing 150 pmol (initial yield 33%; repetitive yield 91%) of the 7.5 kDa fragment of the 25 kDa protein resulted in a sequence with obvious homology to the N-terminal sequence of the 25 kDa LHC IIb apoprotein (Type III) of barley [14], but with less homology to the *A. thaliana* Type I gene product (Fig. 2C). Thus, after CNBr treatment it would appear that the native N-terminus of the 25 kDa subunit was unblocked. Therefore, we tested whether the 25 kDa apoprotein obtained after CNBr cleavage could also be sequenced. Approximately 90 pmol of the 25 kDa protein was sequenced (initial yield 29%; repetitive yield 91%) and a sequence of the first 25 N-terminal residues was obtained exhibiting complete similarity to the 7.5 kDa CNBr fragment, partial similarity at 19 of 25 residues (76%) to the barley Type III LHC IIb sequence [14], but similarity at only 14 of 25 (56%) residues to those in the derived sequence from the 3 Type I LHC IIb *A. thaliana* genes (Fig. 2C). It is not known how the N-terminus became unblocked during treatment with CNBr.

Alignment of the partial sequence of the 25 kDa LHC IIb protein with the protein sequences derived from the *A. thaliana* LHC IIb genes (Fig. 2C) indicates that the 25 kDa component is 11 residues shorter at its amino terminus. Although it is not known whether the sequences also vary at other positions along their entire lengths, the difference in length at the N-terminus possibly explains the difference in apparent mol. wts observed for the 28 and 25 kDa proteins. Within the first 11 residues of the Type I and Type II LHC IIbs is a Thr (residue 3), shown to be involved in the reversible phosphorylation of this apoprotein [24]. Absence of this Thr residue in the 25 kDa subunit explains why this LHC IIb apoprotein is not phosphorylated in vitro [1]. It is also interesting to note that deletion of the first 18 residues of the mature wheat ab-1.6 LHC IIb gene, a Type I LHC IIb gene, allows binding of the precursor protein to the envelope, but abolishes import [25], indicating that this region of the protein is necessary for import

into the chloroplast. Since this region in the 25 kDa protein is either missing or does not display any similarity until residue 17 of the Type I protein, one could speculate that only a very short stretch of amino acids around residue 17 or a higher order structure within the mature sequence, which was altered by the 18 residue deletion in the import experiments with the wheat ab-1.6 construct, is required for proper import across the chloroplast envelope.

In vivo the LHC IIb oligomer is an organization of 2 or more similar yet distinct pigment-proteins. The functional and structural roles of the individual LHC IIbs in the PS II holocomplex have yet to be resolved. Electron diffraction analysis of two-dimensional crystals of LHC IIb indicates that LHC IIb trimeric units are formed via monomer interactions, which most likely occur in the region of the polypeptides just preceding the first transmembrane helix of the LHC IIb proteins [2]. If correct, absence of the first 11 residues in the Type III protein could affect the assembly and arrangement of the 25 kDa subunit with the other subunits around the CC II and with the other LHC IIbs. The different sized LHC IIb monomers would appear to associate in varying ratios, thus creating essentially different trimeric LHC IIb units. This is presumably at least partly correct since the 28 kDa protein occurs in much greater abundance than the other LHC IIb subunit(s) (Fig. 1; see also [1]). Multimeric LHC II subcomplexes have been isolated from barley containing LHC IIa and LHC IId, 2 minor LHC II components which are similar, yet distinct from the LHC IIb subunits (reviewed in [26]), together with the 28 and 25 kDa LHC IIb subunits in a 2:1 ratio [1]. This subcomplex is predicted to bind closer to the core complex than another LHC IIb subcomplex containing only the 28 and 27 kDa LHC IIb subunits [1]. Similarly, in maize the smallest LHC IIb component of 26 kDa, presumably equivalent to the 25 kDa protein in barley and *A. thaliana*, is a component of the tightly bound LHC II, that is not phosphorylated and does not migrate from PS II upon state transition [27]. The importance of the 25 kDa subunit over the other LHC IIb subunits is also demonstrated in the *Chlorina* chl *b*-less barley mutant where increased levels of the 25 kDa protein are observed with a virtual absence of the 28 and 27 kDa subunits [1,28]. These data would indicate that the 25 kDa LHC IIb lies closer to CC II than the other LHC IIbs and likely serves a structural role, linking the remaining LHC IIbs to the rest of CC II (see e.g. [1]). We are currently examining the precise role the 25 kDa LHC IIb plays in the function and structure of the light-harvesting antenna complex of PS II.

Acknowledgements: We gratefully acknowledge Dr. Audree Fowler for carrying out the protein sequencing and Dr. Gary F. Peter for many helpful discussions and critical reading of the manuscript. This work was supported by a grant from the National Science Foundation (DMB 87-16230) to J.P.T. Protein sequencing performed at the UCLA Protein Microsequencing Facility was aided by a BRS Shared Instrumentation Grant 1 S10RR05554-01 from the National Institutes of Health.

REFERENCES

[1] Peter, G.P. and Thornber, J.P. (1991) J. Biol. Chem. 266, (in press).
[2] Kuhlbrandt, W. and Wang, D.N. (1991) Nature 350, 130–134.
[3] Bassi, R., Hoyer-Hansen, G., Barbato, R., Giacometti, G.M. and Simpson, D.J. (1987) J. Biol. Chem. 262, 13333–13341.
[4] Kohorn, B., Harel, E., Chitnis, P.R., Thornber, J.P. and Tobin, E.M. (1986) J. Cell Biol. 102, 972–981.
[5] Kohorn, B. and Tobin, E.M. (1986) Plant Physiol. 82, 1172–1174.
[6] Cline, K. (1988) Plant Physiol. 86, 1120–1126.
[7] Lamppa, G. and Abad, M. (1987) J. Cell Biol. 105, 2641–2648.
[8] Jansson, S., Selstam, E. and Gustafsson, P. (1990) Biochim. Biophys. Acta 1019, 110–114.
[9] Dunsmuir, P., Smith, S.M. and Bedbrook, J. (1983) J. Mol. Appl. Genet. 2, 285–300.
[10] Leutwiler, L.S., Meyerowitz, E.M. and Tobin, E.M. (1986) Nucleic Acids Res. 14, 4051–4064.
[11] Chitnis, P.R. and Thornber, J.P. (1988) Photosyn. Res. 16, 41–63.
[12] Stayton, M.M., Black, M., Bedbrook, J. and Dunsmuir, P. (1986) Nucleic Acids Res. 14, 9781–9796.
[13] Karlin-Neumann, G.A., Kohorn, B.D., Thornber, J.P. and Tobin, E.M. (1985) J. Mol. Appl. Genet. 3, 45–61.
[14] Morishige, D.T. and Thornber, J.P. (1990) in: Current Research in Photosynthesis, vol II (M. Baltscheffsky, ed.) pp. 261–264, Kluwer, Dordrecht.
[15] Webber, A.N. and Gray, J.C. (1989) FEBS Lett. 249, 79–82.
[16] Karlin-Neumann, G.A., Sun, L. and Tobin, E.M. (1988) Plant Physiol. 88, 1323–1331.
[17] Camm, E.L. and Green, B.R. (1982) Arch. Biochem. Biophys. 214, 563–572.
[18] Morishige, D.T. and Thornber, J.P. (1990) FEBS Lett. 264, 239–242.
[19] Dunahay, T.G. and Staehelin, L.A. (1986) Plant Physiol. 80, 429–434.
[20] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[21] Michel, H.P., Buvinger, W.E. and Bennett, J. (1990) in: Current Research in Photosynthesis, vol II (M. Baltscheffsky, ed) pp. 747–753, Kluwer, Dordrecht.
[22] Green, B.R., Pichersky, E. and Kloppstech, K. (1991) Trends Biochem. Sci. 16, 181–186.
[23] Morishige, D.T. and Thornber, J.P. (1992) Plant Physiol., (in press).
[24] Mullet, J.E. (1983) J. Biol. Chem. 258, 9941–9948.
[25] Clark, S.E., Oblong, J.E. and Lamppa, G.K. (1990) Plant Cell 2, 173–184.
[26] Thornber, J.P., Morishige, D.T., Anandan, S. and Peter, G.P. (1991) in: The Chlorophylls (Scheer, H. ed.) pp. 549–585, CRC UniScience Series, CRC Press, Boca Raton, Florida.
[27] Bassi, R., Rigoni, F., Barbato, R. and Giacometti, G.M. (1988) Biochim. Biophys. Acta 936, 29–38.
[28] White, M.J. and Green, B.R. (1987) Eur. J. Biochem. 165, 531–535.