Biochimica et Biophysica Acta, 1019 (1990) 110–114
Elsevier

BBABIO 43260

The rapidly phosphorylated 25 kDa polypeptide of the light-harvesting complex of Photosystem II is encoded by the Type 2 *cab-II* genes

Stefan Jansson, Eva Selstam and Petter Gustafsson

Department of Plant Physiology, University of Umeå, Umeå (Sweden)

(Received 11 January 1990)
(Revised manuscript received 24 April 1990)

Key words: Peripheral LHC II; Type 2 *cab-II* gene; Acylation; Gene product; Light harvesting complex; Photosystem II

The main light-harvesting complex of Photosystem II (LHC II) in higher plants consists of two sub-populations. The 'inner' pool consists only of a 27 kDa polypeptide, whereas in the 'outer' pool both the 27 kDa and a 25 kDa polypeptide are found. We purified the 25 and the 27 kDa LHC II polypeptides from Scots pine and 25 kDa LHC II polypeptide from spinach. Protein sequencing after cleavage with endoproteinase Lys-C showed that the 25 kDa polypeptide is encoded by the Type 2 *cab-II* genes and the 27 kDa polypeptide by the Type 1 *cab-II* genes. A fatty acid was not covalently attached to the peptides assembled into the pigment-protein complex. Our results show that the different polypeptides seen on a gel are different gene products, and not the result of different processing.

Introduction

The light-harvesting chlorophyll *a/b*-binding proteins of Photosystem II are a family of related pigment-protein complexes which are dominant in the thylakoid membranes of higher plants. Despite the fact that their apoproteins and their corresponding genes have been extensively studied (for reviews see Refs. 1–3) it is still not clear how many physiologically different polypeptides constitute the complex. By non-denaturating electrophoresis a number of minor chlorophyll *a/b*-binding protein complexes (CP29, CP27, CP26 and CP24) can be separated. The protein complex commonly referred to as LHC II is the most abundant and can be further resolved into at least two subpopulations, the 'inner' and the 'outer' pool [4,5]. The 'inner' tightly bound pool contains a 27 kDa polypeptide, whereas the 'outer' or 'peripheral' pool consists of a mixture of 27 kDa and 25 kDa polypeptides [4,6]. The 25 kDa polypeptide is phosphorylated by a thylakoid-bound LHC-II kinase at a much higher initial rate then the 27 kDa polypeptide. As a consequence of the phosphorylation the 'outer' pool of the LHC II detaches from Photosystem II and migrates from the grana stacks into the stroma exposed thylakoid regions [7,8]. A role in long-term light acclimation [9] and a correlation with the so-called α-centers [10] has also been proposed for the 25-kDa-enriched 'peripheral' LHC II pool.

The nuclear genes encoding the LHC II polypeptides, referred to as *cab* or *cab-II* genes, are present in a multi-gene family and have been sequenced from a number of plants, for review see Refs. 2. In angiosperms [2] and gymnosperms [11] the genes are present in two types, Type 1 and Type 2 whose amino acid sequences are 85–90% homologous. We have previously defined amino acid residues at 14 positions in the mature polypeptides where Type 1 and Type 2 gene products could be distinguished from each other (Fig 1, and Ref. 11). However, the relation between the different genes and the different polypeptides has not been established. The proteins are synthesized as larger precursors, pLHC II, which are imported into the chloroplasts, processed to their mature size and inserted into the thylakoids. The heterogeneity of the LHC II polypeptides has been thought to be a result of differences within the multigene family, but in addition, the production of multiple bands on SDS-PAGE in some in vitro import experiments [12–14] has led many authors to believe that different processing has occurred. Since labelling experiments with palmitylic acid have shown that LHC-II could be palmitoylated [15] some authors have suggested that different acylation could also account for some heterogeneity [1,4].

Correspondence: S. Jansson, Department of Plant Physiology, University of Umeå, 901 87 Umeå, Sweden.

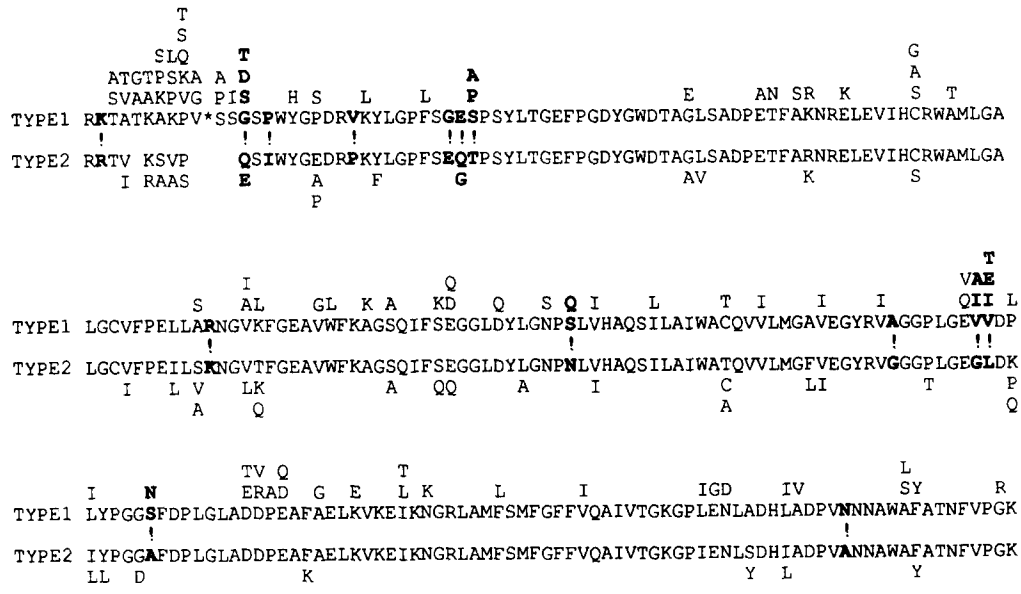

Fig. 1. Comparison of deduced amino acid sequences from published Type 1 and Type 2 *cab-II* genes from angiosperms and gymnosperms [2,11,23]. Bold letters and exclamation marks (!) indicate Type 1/Type 2-specific amino acids. * indicates a highly variable position.

The identification of type-specific amino acids enables us to determine whether the 25 and 27 kDa polypeptides are encoded by different types of *cab-II* genes or are the result of different post-translational events. We did this by comparing protein sequences of the 25 and 27 kDa LHC II polypeptides with deduced amino acid sequences from the Type 1 and Type 2 *cab-II* genes. In addition we also examined whether any

Fig. 2. Predicted protein sequences of Scots pine 25 kDa (P25), spinach 25 kDa (S25) and Scots pine 27 kDa (P27) LHC II polypeptides (see Material and Methods). Arrows correspond to the first 20 amino acids in predicted peptide fragments produced after endoproteinase Lys-C-digestion. Fragments smaller then 3 kDa are omitted. Bold letters indicate Type-1/Type-2-specific amino acids [11].

fatty acids, which could alter the relative electrophoretic mobility of the polypeptides, were covalently bound to the LHC II polypeptides.

Material and Methods

Prediction of protein sequences

The pine 25 kDa LHC II (P25) sequence was predicted from published sequences [11,23]. The spinach 25 kDa LHC II (S25) sequence was predicted from the consensus angiosperm Type 2 *cab-II* protein sequence [2,11] as no spinach Type 2 *cab-II* gene sequence has been published. Prediction of the pine 27 kDa LHC II (P27) protein sequence was made from the deduced amino-acid sequence of the two most highly expressed Scots pine Type 1 *cab-II* genes, *cab-II/IA* and *cab-II/IB* [11]. Proteolytic fragments produced after cleavage of the polypeptides P25, S25 and P27 with endoproteinase Lys-C are shown in Fig 2.

Purification of LHC II polypeptides

A thylakoid protein fraction enriched in LHC II was isolated from needles of 3–4-year-old greenhouse grown Scots pine (*Pinus sylvestris* L.) Needles were homogenized in 50 mM K_2HPO_4 (pH 6.5), 10 mM KCl, 0.3 M sucrose and 20% poly(ethylene glycol) (PEG) 4000 and squeezed through cheesecloth and nylon mesh (20 μm) Thylakoids were pelleted at $3000 \times g$, resuspended and pelleted three times in the same buffer without PEG. The pellet was homogenized in 5 mM EDTA, pelleted and resuspended in water. Sucrose gradient centrifugation was performed essentially according to [16] with a Triton/Chl ratio of 1 : 1. After centrifugation, the proteins were acetone-precipitated.

Spinach LHC II was purified as in Ref. 16 by sucrose gradient centrifugation of Triton-solubilized thylakoids.

The ratio between 25 and 27 kDa LHC II polypeptides was obtained by Coomassie-staining of SDS-polyacrylamide gels.

25 and 27 kDa polypeptides of LHC II were purified by SDS-polyacrylamide electrophoresis (SDS-PAGE). The Scots pine thylakoid protein fraction or spinach LHC II were subjected to SDS-PAGE on a linear 15% gel. Proteins were electro-blotted onto an Immobilon-membrane and stained with Amido black. The 25 and the 27 kDa polypeptide bands were excised and the proteins were recovered in 2% SDS, 1% Triton X-100 and acetone precipitated [17].

Cleavage of proteins and purification of fragments

The 25 and 27 kDa polypeptides from Scots pine and the 25 kDa polypeptide from spinach were digested with endoproteinase Lys C (Boehringer-Mannheim) in 50 mM Tris-HCl (pH 8.5), 0.1% SDS for 44 h at 37°C. To purify the proteolytic fragments we tried separation by gel electrophoresis and subsequent blotting onto Immobilon [18]. However, the LHC II fragments were unable to bind to the membrane although standard proteins of the same size did bind. Instead we precipitated the largest fragments with 1/10 vol. 1 M KCl, 200 mM $MgCl_2$ and 4 vol. of acetone overnight at $-20\,^\circ$C. After centrifugation, the pellet was dissolved in 2% SDS, 1% β-mercaptoethanol and 5% glycerol and analysed by SDS-PAGE on a 20% gel.

Peptide sequencing

Amino acid sequence analysis were made on an Applied Biosystems (Foster City, CA) 477A Pulsed Liquid Phase sequencer with an online PTH 120A Analyzer for 20 cycles. Before initiating Edman degradation of the peptides, excessive SDS was removed from the glass filter by a special procedure described in Ref. 19. The initial yield calculated from a sequenced standard protein, β-lactoglobulin, was 47% and repetitive yield about 97%. Due to some contaminating compound, Gly and Arg residues were unreadable in the first five cycles.

Fatty acid analysis

Covalently bound fatty acids were analysed according to [20] with some modifications. 2 mg of purified spinach LHC II was precipitated overnight at $-20\,^\circ$C in 80% acetone. The precipitate was centrifuged at $10\,000 \times g$ for 30 min and the pellet extracted with chloroform/methanol, 2 : 1, 1 : 1, 1 : 2 (by volume) and then with butanol-1/0.02 M phosphate buffer (pH 7.5) (1/1, by volume). The extractions were made at room temperature for 20–24 h. SDS-PAGE was performed on the delipated protein samples to ensure that the extraction did not alter the stoichiometry between the 25 and 27 kDa polypeptides. The lipid-free LHC II proteins were subjected to acid hydrolysis with 1.2 M HCl in dry methanol for 20 h at 80°C . Methylated and hydrolysed fatty acids were extracted from the acidic hydrolysate with petroleum ether (bp 40–60 °C). The extracted acids were then treated with 5% H_2SO_4 in dry methanol at 70 °C for 2 h to methylate residual hydrolysed free fatty acids.

Methylated fatty acids were analysed by gas-liquid chromatography on a packed 10% SP 2330 column [21]. Methyl pentadecanoate (23.4 nmol) was used as an internal standard.

After acidic hydrolysis of covalently bound fatty acids, the protein content was determined after hydrolysis of the LHC II to amino acids in 6 M HCl for 24 h at 110 °C according to Ref. 22.

Results

Purification of proteins and peptide fragments

The ratio between the 27 kDa and 25 kDa LHC II polypeptides in the thylakoids from Scots pine was

1.5–2, whereas the ratio in the spinach LHC II was roughly 4.

25 and 27 kDa LHC II polypeptides from Scots pine (hereafter called P25 and P27, respectively) and spinach 25 kDa LHC II polypeptide (S25) were purified to homogeneity as judged by SDS-PAGE. The polypeptides were cleaved with endoproteinase Lys-C and the proteolytic fragments were aceton precipitated. Under the conditions described, only the two largest fragments of each digestion were precipitated.

Identification of the 25 kDa polypeptide as the Type 2 cab-II gene product

We sequenced the precipitated fragments from P25, S25 and P27 and compared the sequences obtained with the sequences predicted from the DNA sequence (Fig 3).

The observed sequences from the proteolytic fragments of the different polypeptides corresponded to the sequences predicted from the deduced amino acid sequences. The sequence from the fragments P25: *Lys24* and S25: *Lys24* showed that the 25 kDa polypeptides had Type-2-specific amino acids (Glu-Gly/Gln-Thr) at positions 31–33. The absence of P27: *Lys24* sequence corroborates the predicted P27 sequence, where Lys24 is substituted with Leu. The P25 and S25 also had a Type 2-specific Asn residue at position 119, where the P27 polypeptide had a mixture between the diagnostic amino acids Ser and Gln as predicted. The expected mixtures Asn/Ser at position 117 and Val/Ile at position 121 were also obtained from the P27: *Lys101* sequence.

Thus, we conclude that the 25 kDa LHC II polypeptide is encoded by the Type 2 *cab-II* genes and the 27 kDa polypeptide by the Type 1 *cab-II* genes.

We also obtained a sequence from the P27: *Lys62* fragment, but as no type-specific amino acids are present in this region no additional information concerning the correlation between the genes and the proteins was derived.

One discrepancy between the predicted and observed sequences could be pointed out. Ser-108 was undetectable in all three polypeptides, which is probably a consequence of a modification of that amino acid residue.

Fatty acid analysis

The analysis of fatty acids covalently bound to LHC II showed that 44 nmol protein contained 1.9 nmol fatty acids (palmitic, palmitoleic, oleic and linoleic acid) corresponding to 0.043 mol fatty acid/mol protein. These small amounts of fatty acids are probably residual membrane lipids and not fatty acids covalently bound to the protein. Thus, we conclude that no fatty acid which could alter gel mobility is covalently bound either to the 27 or the 25 kDa LHC II polypeptides.

Discussion

In this paper we are able to correlate *cab-II* gene heterogeneity with LHC II protein heterogeneity, and we show that the rapidly phosphorylated 25 kDa polypeptide of the 'peripheral' LHC II pool is encoded by the Type 2 *cab-II* genes. The data presented in this paper show that specific LHC II polypeptides in vivo are products of specific *cab-II* genes, and that the variation in mobility of one *cab-II* gene product that has been demonstrated in some systems is probably an artifact. In fact, in most recent publications it is shown that in vitro-uptake of one gene product produces only one band on a gel (for example, Ref. 24). This is also the case when a Scots pine *cab-II* gene product is taken up by pea chloroplasts (Robinsson, C. and Jansson, S., unpublished data).

Spinach and Scots pine are very distant in evolution: more then 350 million years according to classical systematic botany [25]. The fact that we show that the 25 kDa polypeptide is encoded by the Type 2 genes in both spinach and Scots pine implies that this is the case for all angiosperms and gymnosperms, and that the type-specific (27/25-kDa-specific) amino acids have been conserved through evolution for at least 350 million years. Thus, the 27/25-kDa-specific amino acids must give the polypeptides different properties which are physiologically very important. The difference in phosphorylation rate and the localisation of the 25 kDa polypeptide in the 'peripheral' pool of LHC II is well-documented, but we do not exclude the possibility that other differences might be important.

(a)

S25:*Lys24*	Predicted	YLGPFSEQTPSYLTGEFPGD
	Observed	YLXPFSEQTPSYLTGEFPGD
P25:*Lys24*	Predicted	YLGPFSEGTPSYLTGEFPGD
	Observed	YLXPFXEGTPXYXTGXXXXX

(b)

S25:*Lys101*	Predicted	AGSQIFSEGGLDYLGNPNLV
	Observed	AXSQIFXEGGLDYLGNPNLV
P25:*Lys101*	Predicted	AGAQIFSEGGLDYLGNPNLI
	Observed	AXAQIFXEXGLDXLGNPNLI
		N S V
P27:*Lys101*	Predicted	AGAQIFSEGGLDYLGSPQLI
	Observed	AXAQIFXEGGLDYLGSPQLI
		N S V

Fig. 3. Comparison of predicted and observed amino acid sequences resulting from cleavage after (a) Lys-24 and (b) Lys-101. X denotes unidentified amino acid. Type-1/Type-2-specific amino acids are boxed.

A LHC II sequence from the moss *Physcomitrella patens* has recently been published [26]. The sequence of the moss protein shows characteristics of both the 25 and 27 kDa protein sequences of angiosperms and gymnosperms. Of the 14 type-specific amino acids, 7 are Type 1-specific, 6 are Type 2-specific and one is neither in the moss LHC II sequence. Thus it is not clear whether that polypeptide is functionally equivalent to the 25 or 27 kDa LHC II polypeptides.

It is notable that the ratios between the amounts of the 27 and 25 kDa LHC II polypeptides are different between Scots pine and spinach, the LHC II of Scots pine is enriched in the 25 kDa polypeptide. This is also the case when Scots pine is grown under physiological conditions when the total LHC pool (measured as Chl *a/b* ratio) is small (Jansson, S., unpublished results). Thus Scots pine has, in contrast to spinach and pea [9], a high proportion of 'peripheral' LHC II in a situation when the antenna size is small.

It has been suggested from N-terminal sequence data that a minor, smaller LHC II component is encoded by the Type 2 *cab-II* genes, but processed at a different site [27]. However, it has recently been shown that the N-terminal residue in vivo of the gene products of the Type 1 and Type 2 *cab-II* genes is an arginine residue, Arg^1, in an acetylated form [28], and thus, the proteins are processed at the same site. Moreover, the protein sequence from the minor, smaller LHC II component presented in Ref. 27 does not correspond to either Type 1 or Type 2 sequences. At the six Type-specific positions in the sequence presented, only two correspond to the Type 2 genes, one to the Type 1 genes, and three to neither of the types. As the N-terminus of both gene products are blocked in vivo, we think that the sequence presented originates from another smaller LHC II polypeptide (encoded by a 'Type 3' *cab-II* gene) comigrating in their gel system.

The molecular weight for the 25 and 27 kDa LHC II polypeptides predicted from the DNA sequence is very similar, in the case of the pine polypeptides 24.79 kDa and 25.04 kDa. We do not know why a minute difference in molecular weight could render such a different mobility in some (but not all) SDS gel systems. No fatty acid is covalently attached to the polypeptides assembled in the chlorophyll/protein complex. However, transient attachment of a palmitic acid during membrane insertion/processing/assembly of the LHC II proteins could be one reason for the heterogeneity sometimes found in in vitro uptake experiments.

Here we have shown that the two major polypeptides of LHC II are encoded by different types of gene. However, much is still unknown about the minor protein components of the LHC II, which have been objects of some recent studies (for example, Ref. 5). We think it is reasonable to assume that each of these is encoded by a unique type of gene. and we also think that it is a necessity for the future to combine DNA and protein sequencing to avoid confusion in the field.

Acknowledgements

We thank Dr Michael Spangfort for fruitful discussions and for preparation of spinach LHC II, and Dr Per-Ingvar Ohlsson, University of Umeå, for skilful performance of the sequence analysis. This work was supported by grants from Swedish Council for Forestry and Agricultural Research and Swedish Cancer Society (2515-B8-02XB).

References

1 Peter, G. and Thornber, J.P. (1988) in: Photosynthetic Light-Harvesting Systems (Scheer and Schneider, eds), pp. 175–186, Walter de Gruyter, Berlin New York.
2 Chitnis, P.R. and Thornber, J.P. (1988) Photosyn. Res. 16, 41–63.
3 Green, B.R. (1988) Photosyn. Res. 15, 3–32.
4 Spangfort, M. and Andersson, B. (1989) Biochim. Biophys. Acta 977, 163–170.
5 Bassi, R., Rigoni, F., Barbato, R. and Giacometti, G.M. (1988) Biochim. Biophys. Acta 936, 29–38.
6 Larsson, U.K., Sundby, C. and Andersson, B. (1987) Biochim. Biophys. Acta 894, 59–68.
7 Larsson, U.K. and Andersson, B. (1985) Biochim. Biophys. Acta 809, 396–402.
8 Islam, K. (1987) Biochim. Biophys. Acta 893, 333–341.
9 Larsson, U.K., Anderson, J.M. and Andersson, B. (1987) Biochim. Biophys. Acta 894, 69–75.
10 Mäenpää, P. and Andersson, B. (1989) Z. Naturforsch. 44C, 403–406.
11 Jansson S. and Gustafsson P. (1990) Plant. Mol. Biol. 14, 287–296.
12 Kohorn, B.D., Harel, E., Chitnis, P.R., Thornber, J.P. and Tobin, E.M. (1986) J. Cell Biol. 102, 972–981.
13 Kohorn, B.D. and Tobin, E.M. (1986) Plant Physiol. 82, 1172–1174.
14 Lamppa, G.K. and Abad, M.S. (1987) J. Cell Biol. 105, 2641–2648.
15 Mattoo, A.K. and Edelmann, M. (1987) Proc. Natl. Acad. Sci. USA 84, 1497–1501.
16 Ryrie, I.J., Andersson, J.M. and Goodchild, D.J. (1980) Eur. J. Biochem 107, 345–354.
17 Szewczyk, B. and Summers, D.F. (1988) Anal. Biochem 168, 48–53.
18 Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
19 Wilson, K.J. and Yuan, P.M. (1989) in Protein Sequencing A Practical Approach. pp. 1–41, IRL Press, Oxford.
20 Zdebska, E., Antoniewics, J. and Koscielak, J. (1989) Arch. Biochem. Biophys. 273, 223–229.
21 Selstam, E. and Öquist, G. (1985) Plant Sci. 42, 41–48.
22 Moore, J. (1968) J. Biol. Chem. 243, 6281–6283.
23 Yamamoto, N., Matsuoka, M., Kano-Murakimi, Y., Tanaka, Y. and Ohashi, Y. (1988) Nucleic Acids Res. 16, 11829.
24 Pichersky, E., Hoffman, N.E., Malik, V.S., Bernatzky, R., Tanksley, S.D., Szabo, L. and Cashmore, A.R. (1987) Plant Mol Biol 9, 109–120.
25 Frohne, D. and Jensen, U. (1973) Systematik des Pflantzenreich, Gustav Fisher, Stuttgart.
26 Long, Z., Wang, S.Y. and Nelson, N. (1989) Gene 76, 299–312.
27 Webber, A.N. and Gray, J.C. (1989) FEBS Lett. 249, 79–82.
28 Michel, H., Buvinger, W.E., Hunt, D.F. and Bennet, J. (1989) Physiol Plant 76 (3, Part 2), A21 (Abstr.).