Proteomic Analysis of the Photosystem I Light-Harvesting Antenna in Tomato (*Lycopersicon esculentum*)†

Stefanie Storf,‡ Einar J. Stauber,§ Michael Hippler,*,|| and Volkmar H. R. Schmid*,‡

Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, Institut für Allgemeine Botanik, Johannes Gutenberg-Universität Mainz, Müllerweg 6, 55099 Mainz, Germany, and Lehrstuhl für Pflanzenphysiologie, Friedrich Schiller-Universität Jena, Dornburger Strasse 159, 07743 Jena, Germany

Received January 25, 2004; Revised Manuscript Received May 17, 2004

ABSTRACT: Until now, more genes of the light-harvesting antenna of higher-plant photosystem I (PSI) than proteins have been described. To improve our understanding of the composition of light-harvesting complex I (LHCI) of tomato (*Lycopersicon esculentum*), we combined one- and two-dimensional (1-D and 2-D, respectively) gel electrophoresis with immunoblotting and tandem mass spectrometry (MS/MS). Separation of PSI with high-resolution 1-D gels allowed separation of five bands attributed to proteins of LHCI. Immunoblotting with monospecific antibodies and MS/MS analysis enabled the correct assignment of the four prominent bands to light-harvesting proteins Lhca1−4. The fifth band was recognized by only the Lhca1 antibody. Immunodetection as well as mass spectrometric analysis revealed that these protein bands contain not only the eponymous protein but also other Lhca proteins, indicating a heterogeneous protein composition of Lhca bands. Additionally, highly sensitive MS/MS allowed detection of a second Lhca4 isoform and of Lhca5. These proteins had not been described before on the protein level in higher plants. Two-dimensional gel electrophoresis revealed an even more diverse composition of individual Lhca proteins than was apparent from 1-D gels. For each of the four prominent Lhca proteins, four to five isoforms with different isoelectric points could be identified. In the case of Lhca1, Lhca4, and Lhca3, additional isoforms with slightly differing molecular masses were identified. Thus, we were able to detect four to ten isoforms of each individual Lhca protein in PSI. Reasons for the origin of Lhca heterogeneity are discussed. The observed variety of Lhca proteins and their isoforms is of particular interest in the context of the recently published crystal structure of photosystem I from pea, which showed the presence of only four Lhca proteins per photosystem I. These findings indicate that several populations of photosystem I that differ in their Lhca composition may exist.

The two photosystems (PS)[1] of the photosynthetic apparatus of higher plants carry out the reactions that are responsible for the conversion of solar radiation into chemically fixed energy. In both PS, a central core complex is surrounded by a light-harvesting antenna, which is composed of several light-harvesting complexes (LHCs). The main function of LHCs is collection of solar radiation and transfer of exitation energy to the central core, thereby enlarging the absorption cross section of the PS. This in turn leads to a concomitant increase in photosynthetic performance. In addition, the LHCs can also function in attenuation of the excitation energy transmitted to the core complexes of the PS under high-light conditions. This can be achieved either by translocation of a subpopulation of LHCII from PSII to PSI (*1*) or by dissipation of excess excitation energy via zeaxanthin formed in the xanthophyll cycle (*2*, *3*).

The light-harvesting antenna of PSII is composed of six different apoproteins arranged in four LHCs: LHCII (Lhcb1−3), CP29 (Lhcb4), CP26 (Lhcb5), and CP24 (Lhcb6) (*4*, *5*). Until now, most biochemical analyses of PSI in various plants have identifed four different LHCI proteins (Lhca1−4) with molecular masses determined to be between 20 and 25 kDa by fully denaturing gel electrophoresis (*6*). By gentle separation of mildly solubilized PSI, two different subfractions of LHCI, named LHCI-680 and LHCI-730 (*7*), could be identified. LHCI-730 is composed of Lhca1 and Lhca4 (*7*, *8*) and exists in a heterodimeric state (*9*, *10*). The nature of the oligomerization behavior of Lhca2 and Lhca3 that form LHCI-680 is not clear yet. It has been reported that they also form dimers (*11*). Considering that Lhca2 comigrates (*12*, *13*) and coelutes (*14*) with the LHCI-730 proteins Lhca1 and Lhca4 and that a chemical cross-linking study supports a dimeric state for Lhca2 and Lhca3 (*9*), a homodimeric state of the LHCI-680 proteins is the most probable scenario. However, until now, reconstitution experiments with Lhca2 and Lhca3 have failed to produce dimers of Lhca2 and/or Lhca3 (*15*, *16*). For a long period, it was

† This project was supported by Grant "Nachwuchsgruppe Pflanzenphysiologie" from the federal state of Thüringen (to M.H.) and Grants Schm 1203/2-3 and 1203/2-4 (to V.H.R.S.) from the Deutsche Forschungsgemeinschaft.

* To whom correspondence should be addressed. M.H.: phone, 001-215-898-4974; fax, 001-215-898-8780; e-mail, mhippler@sas.upenn.edu. V.H.R.S.: phone, 0049-6131-3924203; fax, 0049-6131-3923787; e-mail, vschmid@mail.uni-mainz.de.

|| University of Pennsylvania.

‡ Johannes Gutenberg-Universität Mainz.

§ Friedrich Schiller-Universität Jena.

[1] Abbreviations: 1-D and 2-D, one- and two-dimensional protein separation, respectively; Lhca, apoprotein of a light-harvesting complex of photosystem I; *lhca*, gene encoding Lhca; LHC, light-harvesting complex; MS/MS, tandem mass spectrometry; PS, photosystem.

10.1021/bi0498196 CCC: $27.50
Published on Web 06/26/2004

assumed that six to eight Lhca proteins (two copies of each protein) surround the core of PSI (*17*, *18*). However, the recently published 4.4 Å crystal structure of PSI from pea revealed the presence of only four Lhca proteins per PSI (*19*).

Because of the pioneering work that has been done to identify genes encoding LHCI proteins in tomato (*Lycopersicon esculentum*), it can be considered a model plant for studying LHCI protein composition (*20*). Moreover, more than 155 000 EST entries available through The Institute For Genomic Research (TIGR) website (http://www.tigr.org/tdb/tgi/lgi) provide an extensive resource of nucleotide sequence information for investigations of tomato LHC proteins. Furthermore, the adequacy of ESTs for the investigation of LHCI genes and proteins was demonstrated by previous studies in which the complete nucleotide sequence for six *lhca* genes in *Arabidopsis thaliana* (*21*) and nine *Chlamydomonas reinhardtii* Lhca proteins (*22*) was fully documented by EST information. The first two *lhca* genes (*cab6a* and *cab6b*) which were identified in tomato encode two Lhca1 proteins that have a molecular mass of ~ 22 kDa (*23*, *24*). In the years following this discovery, genes which encode Lhca2 (~23 kDa; cab7; *25*), Lhca3 (~25 kDa; cab8; *26*), and Lhca4 (~22 kDa; cab11 and cab12; *27*) were identified in tomato. In *A. thaliana*, two new genes (*lhca5* and *lhca6*) that encode additional putative LHCI proteins have been identified by database analysis (*21*). However, the corresponding proteins have not yet been detected in higher plants. The Lhca5 gene product is the most similar to Lhca4 when the sequences of putative transmembrane helices are compared. In contrast, the Lhca6 gene product is highly homologous to Lhca2 and can therefore be presumably considered to be an additional Lhca2 isoform. The failure to detect these gene products on the protein level in higher plants until now may have been a consequence of their low expression level (*21*). The inability to detect proteins encoded by *cab6b* (*24*) and *cab12* (*27*) from tomato for a long period may also be explained this way. Only recently, refined analytical procedures have allowed further dissection of the PSI light-harvesting system of tomato. By combining HPLC separation of solubilized PSI with electrospray ionization mass spectrometric (ESI-MS) analyses of eluted proteins, Zolla et al. (*28*) were able to detect gene products of *cab6a* and *cab6b*. In this study, additional Lhca2 proteins could be detected. Another investigation of the tomato LHCI antenna composition using SDS−PAGE with high resolution in the LHCI protein region revealed the presence of a fifth band in the region of LHCI protein bands. The relevant protein(s) copurifies with LHCI-730 (*15*). Due to (i) the availability of monospecific antibodies raised against short sequences of individual Lhca proteins, (ii) the improvement in 2-D gel electrophoresis of transmembrane proteins (*29*), and (iii) proteomic approaches facilitated by the combination of extensive sequence information with tandem mass spectrometry (MS/MS), further dissection of the LHCI antenna of tomato appears to be feasible.

In this study, we applied all of these analytical tools to the LHCI antenna of tomato for the first time. In our first approach, we used high-resolution 1-D gel electrophoresis in combination with immunoblotting using monospecific antibodies against Lhca1−4. Further characterization of the resolved protein bands was achieved by proteolytic digestion of excised bands followed by MS/MS. Subsequently, PSI was separated by 2-D gel electrophoresis with isoelectric focusing (IEF) in the first dimension and SDS−PAGE in the second dimension. Characterization by immunoblotting and MS/MS measurements allowed resolution of various isoforms of Lhca proteins and revealed for the first time the very diverse composition of the PSI light-harvesting system in higher plants.

EXPERIMENTAL PROCEDURES

Isolation of Thylakoids and Photosystem I. Thylakoids were isolated according to the methods described in ref *30* from tomato plants (*L. esculentum*) that were grown in a greenhouse under a 16 h−8 h light−dark regime and at a light intensity of ~120 μmol of photons m^{-2} s^{-1} supplied by fluorescent tubes (L58W/11-860 LumiluxPlusEco, Osram, München, Germany). The PSI holocomplex and LHCI-730 were isolated according to the methods described in ref *15* and stored at −70 °C in 30% (w/v) sucrose, 10 mM Tricine/NaOH (pH 7.8), and 1 mM EDTA/NaOH (pH 7.8) at a chlorophyll (Chl) concentration of 1.5−2 mg/mL. The Chl concentration was determined using acetone as a solvent and the equations of ref *31*.

1-D Gel Electrophoresis. Polyacrylamide gradient gels were used for improved separation of LHCI proteins. The resolving gel contained an acrylamide gradient of 14 to 20% (w/v) and 0.375 M Tris-HCl (pH 8.8). Stacking gels had a uniform acrylamide concentration of 4.5% and contained 0.13 M Tris-HCl (pH 6.8). PSI samples were mixed with 2 volumes of denaturing solution [5% (w/v) LDS, 180 mM DTT, 70 mM Tris-HCl (pH 8.4), and 0.666 M sucrose] followed by incubation at 100 °C for 1 min. Aliquots equivalent to 15 μg of Chl were applied per lane. Electrophoresis was conducted with 2× concentrated Laemmli running buffer [50 mM Tris, 384 mM glycine, and 0.1% (w/v) SDS (*32*)] supplemented with 1 mM EDTA/NaOH (pH 7.8) in a Maxigel electrophoresis unit (Biometra, Göttingen, Germany). Gels were run overnight for at least 18 h at 15 mA per gel. Afterward, proteins were either stained with Coomassie Brilliant Blue or transferred onto nitrocellulose. Analysis of protein phosphorylation was carried out with the Pro-Q diamond phosphoprotein gel stain (Molecular Probes, Leiden, The Netherlands) according to the guidelines of the manufacturer. Stained gels and immunodecorated blots were documented with an Epson GT 7000 flat bed scanner.

2-D Gel Electrophoresis. For IEF separation in the first dimension, proteins were precipitated to remove pigments and lipids according to the methods described in ref *33*. Usually, an aliquot of PSI equivalent to 20 μg of Chl was diluted with distilled water to a final volume of 100 μL. For detection of Lhca2 in 2-D spots by MS/MS, larger aliquots equivalent to 100 μg of Chl had to be used. Protein precipitation, lyophilization, and solubilization were carried out as described in ref *29*. The IEF was performed in an IPGphor apparatus using Immobiline DryStrips (Amersham Pharmacia Biotech, Freiburg, Germany) with a linear pH range of 3−10 or 4−7 following the procedure described in ref *29*. Following IEF, the cysteine residues of the proteins were reduced and carbamidomethylated by incubation in two different equilibration buffers as described in ref *29*.

Afterward, the strips were washed briefly in water, loaded onto the second-dimension SDS gel, and covered with 0.5%

agarose. The second-dimension SDS gels contained 13% polyacrylamide, 0.4% piperazine diacrylamide, 0.357 M Tris-HCl (pH 8.8), and 0.025% (w/v) $Na_2S_2O_3$ and were run in a Bio-Rad (München, Germany) multicell apparatus at 8 °C and 30 mA per gel until the bromophenol blue front reached the end of the gel. Gels were either stained with coomassie or silver or blotted onto nitrocellulose.

Immunoblot Analysis. Western Blot analysis was conducted as described in ref *29* with the following modifications. After electrotransfer, the entire band/spot profile of the membrane was visualized with Ponceau-S which allowed later alignment of immunostained bands with the entire band pattern. The primary antibodies used for detection of Lhca1, Lhca3, and Lhca4 were purchased from AgriSera (Vännäs, Sweden). The Lhca2 antibody was a kind gift of L. A. Staehelin (University of Colorado, Boulder, CO) and is described in ref *34*. The antibodies were used at dilutions of 1:2000 (Lhca1), 1:300 (Lhca2), 1:500 (Lhca3), and 1:1000 (Lhca4). Anti-rabbit IgG conjugates with either horseradish peroxidase or alkaline phosphatase were used as the secondary antibody at a dilution of 1:10000 or 1:2500, respectively. Bound antibodies were detected either by enhanced chemiluminescence with the ECL kit of Amersham or by nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate according to the guidelines of the suppliers.

Liquid Chromatography−Mass Spectrometry. Samples for tandem mass spectrometry were prepared by tryptic in-gel hydrolysis of proteins in excised bands/spots essentially as described in ref *35*, with the exception of 2-D spots C−E (from the gel shown in Figure 4), for which the concentration of the trypsin solution was increased to 0.125 μg/μL. In addition, these spots were kept for 1 h on ice with the trypsin solution and were gently agitated using a rotary table shaker. The remainder of the procedure was unchanged. Prior to tryptic digestion, silver-stained protein spots were first destained by incubating the gel piece in a 30 mM $K_3Fe(CN)_6$/100 mM $Na_2S_2O_3$ (1:1, v/v) mixture for 8 min. Double-distilled water and HPLC-grade solvents were used throughout the procedure. The lyophilized samples were stored at −80 °C until analysis was carried out. For MS/MS analysis, the dried samples were diluted in 10 μL of buffer A [0.1% (v/v) formic acid in 5% (v/v) acetonitrile and 95% (v/v) water] and centrifuged for 5 min at 12000*g*. An aliquot of the supernatant (9 μL) was transferred into an autosampling vial. Analyte sampling and chromatography as well as production and acquisition of MS/MS data were performed on-line using fully automated instrumentation as described in ref *22*.

Bioinformatic Data Analysis. Computational analyses of MS/MS data were performed with the Finnigan Sequest/Turbo Sequest software (revision 2.0, ThermoQuest, San Jose, CA) using the parameters described in ref *22*. The databases that were used were as follows: one genomic database containing all tomato DNA sequences found in the NCBI (nr)-nucleotide database, one database containing all tomato LHC protein sequences present in the NCBI (nr)-protein database, and one small EST data bank containing tomato ESTs found by searching the NCBI EST database with the Lhca5 cDNA sequence reported for *A. thaliana* (*21*; GenBank accession number AF134121). The databases were also processed to detect the presence of acetylation (+42 Da for peptide N-termini or lysine residues) and carbamylation (+43 Da for N-termini or arginine or lysine residues) from the MS/MS data from the 2-D gel spots. Cysteine modification by carbamidomethylation (+57 Da) was also attributed. To search for N-terminally processed protein forms, Sequest analyses were performed by choosing the option "no enzyme" for the *in-silico* digestion of a database containing all tomato LHC deduced protein sequences.

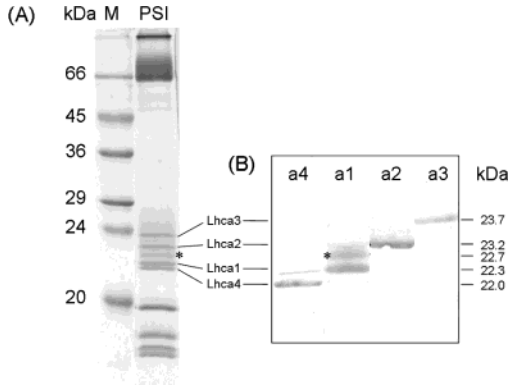

FIGURE 1: Protein composition of the PSI holocomplex of tomato. In panel A, a fully denaturing gel stained with Coomassie Blue is shown. Lhca1−4 bands and a fifth protein band in the LHCI mass range are marked. M represents the molecular mass standard. In panel B, immunoblots of proteins of the PSI holocomplex separated by SDS−PAGE as in panel A are shown. Blots were probed with antibodies against LHCI apoproteins Lhca1 (a1), Lhca2 (a2), Lhca3 (a3), and Lhca4 (a4).

BLAST searches were carried out against the NCBI databases (both genomic and EST) with the net-based blast program available at http://www.ncbi.nlm.nih.gov:80/BLAST using the default values. Alignments of peptide sequences derived from MS/MS data with complete protein sequences and alignment of the amino acid sequences of different Lhca proteins with each other were performed with the net-based align program at the GENESTREAM network server IGH (Montpellier, France, http://www2.igh.cnrs.fr/bin/align-guess.cgi) using the default values. Figure 3 was made on basis of these alignments using the GeneDoc program available at http://www.psc.edu/biomed/genedoc. Translation of cDNA and EST sequences into amino acid sequences and determination of theoretical molecular masses and isoelectric points of the Lhca proteins were carried out with the corresponding net-based tools at http://www.expasy.org/tools/.

RESULTS

Recently, a fifth protein band in the molecular mass range of LHCI proteins was resolved by SDS−PAGE for both the PSI holocomplex and LHCI-730 (*15*). By using long gels with an acrylamide gradient from 14 to 20% in the resolving gel and 2× Laemmli electrophoresis buffer, we could further improve the separation of protein bands in the LHCI region. As shown in Figure 1A, this procedure allowed clear resolution of four intensely stained bands with apparent molecular masses of ∼22.0, ∼22.3, ∼23.2, and ∼23.7 kDa. On the basis of their molecular mass, they were tentatively assigned to Lhca4 (22.0 kDa), Lhca1 (22.3 kDa), Lhca2 (23.2 kDa), and Lhca3 (23.7 kDa). In addition, a less intensely stained fifth band with a molecular mass of ∼22.7 kDa was clearly resolved (marked with an asterisk in Figure 1).

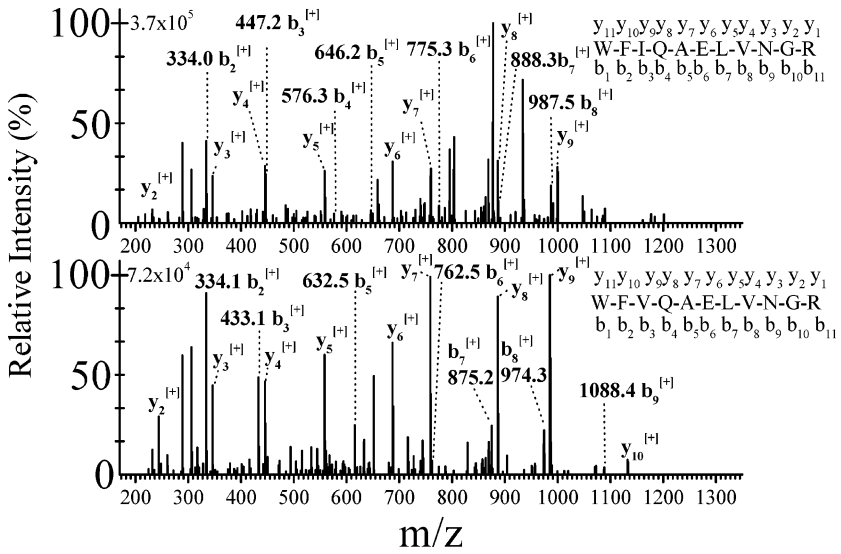

FIGURE 2: MS/MS spectra of tryptic peptide ions at *m*/*z* 666.7 ([2H$^+$]) (top graph) and *m*/*z* 660.2 ([2H$^+$]) (bottom graph) that are characteristic of the two different Lhca4 proteins (cab11 and cab12). The presence of isoleucine (cab11) rather than valine (cab12) at the third position of the peptide is reflected in a mass difference of 13.1 Da between ions of the b-series from the two spectra, beginning with the b_3 ion. Ions of the b-series (containing the N-terminus of the peptide) and y-series (containing the C-terminus of the peptide) are labeled. The amino acid sequences retrieved by the Sequest software are printed beside the corresponding spectra.

To ensure assignment of the protein bands to LHCI apoproteins already described for tomato (Lhca1−4), immunoblot analyses were performed. Proteins of the PSI holocomplex were separated by 1-D gel electrophoresis as described above, transferred onto nitrocellulose membranes, and probed with antibodies raised against short peptide sequences of the individual Lhca proteins. The monospecificity of the antibodies was tested with individual Lhca proteins obtained by overexpression of tomato Lhca genes in *Escherichia coli* (data not shown). Immunodetection using antibodies directed against Lhca2 and Lhca3 confirmed the previous assignment based on the band pattern on coomassie-stained gels. The Lhca3 antibody reacted with the band at 23.7 kDa and the Lhca2 antibody with the band resolved at 23.2 kDa (Figure 1B). In contrast, Lhca1 and Lhca4 antibodies recognized three and two distinct bands, respectively. The Lhca4 antibody cross reacted strongly with the band with an M_r of 22.0 kDa. Additionally, the Lhca4 antibody recognized the lower part of the band at 22.3 kDa, which was the main target of the Lhca1 antibody. The Lhca1 antibody also reacted strongly with the fifth band, which had not been identified until now (asterisk in Figure 1), and a weak reaction was observed with the 23.2 kDa band. These results suggest a heterogenic protein composition of the "Lhca1" and "Lhca2" band.

To examine further the protein composition of these bands, PSI and LHCI-730 were separated in high-resolution gels as described above. Following coomassie staining, bands were excised and digested with trypsin, and the tryptic peptides were subjected to liquid chromatography and MS/MS. Table 1 summarizes the results of these analyses. It is of note that more than one protein could be detected in all five LHCI bands.

In the 22.0 kDa band that was recognized by only the Lhca4 antibody, peptides specific for products of two different Lhca4 genes (*cab11* and *cab12*; *25*) were detected (Figure 2). Thus, the specific amino acid sequence information from highly sensitive MS/MS measurements enabled detection of the *cab12* gene product for the first time on the protein level. As a consequence of the very high level of amino acid sequence identity (91%) of the products of *cab11* and *cab12*, several additional peptides were found in the 22.0 kDa band that are common to both of these Lhca4 isoforms. In addition to these Lhca4 fragments, one peptide specific for Lhca1 was found.

In the 22.3 kDa band that strongly reacted with the Lhca1 antibody, several peptides specific for gene products of *cab6a* (*23*) and *cab6b* (*24*) were found. Also, single peptides specific for proteins derived from the two Lhca4 genes described above (*cab11* and *cab12*) were detected, which confirms immunoblot analyses showing comigration of Lhca4 with Lhca1 in the 22.3 kDa band. Two tryptic peptides (WYVQAELVHAR and DLPVWYEAGATK) identified from the MS/MS data matched the Lhca5 sequence from *A. thaliana*, but did not match any tomato protein retrieved from the NCBI databases. To determine if the MS/MS data could be used to identify tomato homologues of Lhca5 or Lhca6 from *A. thaliana*, BLAST searches were carried out using the tomato EST database (Cornell University, Ithaca, NY) and the cDNA sequences of Lhca5 and Lhca6 from *A. thaliana* (GenBank accession numbers AF134121 and U03395, respectively; *21*). Two EST clones (GenBank accession numbers AW041324 and AI778223) that share a very high level of nucleotide sequence identity with large sections of the *A. thaliana lhca*5 sequence were found. These two partially overlapping EST sequences were joined together, and the resulting nucleotide sequence was translated. Alignment of the obtained amino acid sequence with the Lhca5 sequence of *A. thaliana* demonstrates a very high level of sequence identity (Figure 3). The level of identity for the precursor protein is 68.8% and increases to 77.9% for the region that corresponds to the mature protein of *A. thaliana* as suggested by ref *21*. The two tryptic peptides that were identified within the Lhca5 amino acid sequence from *A. thaliana* are also present in the tomato protein sequence (Figure 3).

The additional band at 22.7 kDa resolved by high-resolution electrophoresis of PSI and LHCI-730 also contains

Table 1: Mass Spectrometric Identification of Proteins in the Bands Resolved in the Molecular Mass Range of LHCI Proteins following Separation of PSI[1] or LHCI-730[2] in Denaturing 1-D SDS Gels[a]

band[b]	protein[c]	peptide[d]	position[e]	ref[e]	*z*	Δ*M*	X_{corr}
Lhca4[1]	Lhca1 (*cab6a*/*6b*)	KYPGGAFDPLGYSK	168−180	P12360	2	−1.0	3.81
	Lhca4 (*cab11*/*12*)	WQDIKNPGSVNQDPIFK		S14305/6	3	0.4	4.66
		NPGSVNQDPIFK		S14305/6	2	−0.5	4.13
		NPGSVNQDPIFK		S14305/6	1	0.0	3.06
	Lhca4 (*cab11*)	WFIQAELVNGR	90−100	S14305	2	−0.7	4.07
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.2	2.33
Lhca1[1]	Lhca1 (*cab6a*/*6b*)	WAMLAVPGIIVPEALGLGNWVK	94−115	P12360	3	0.4	5.95
		PSYLDGSAPGDFGFDPLGLGEVPANLER	56−83	P12360	2	0.4	4.96
		YPGGAFDPLGYSKDPAKFEELK	168−189	P12360	3	0.4	4.41
		YPGGAFDPLGYSK	168−180	P12360	2	−0.1	4.30
		KYPGGAFDPLGYSK	167−180	P12360	3	−0.4	3.59
		DPAKFEELK	181−189	P12360	1	0.0	1.72
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.4	4.15
		WYDAGK		S14305/6	1	0.4	1.85
	Lhca4 (*cab11*)	WFIQAELVNGR	90−100	S14305	2	0.0	3.46
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.0	4.50
	Lhca5	WYVQAELVHAR	90−101		2	0.8	3.52
		WYVQAELVHAR	90−101		3	−0.2	3.45
		DLPVWYEAGATK	123−134		2	−0.8	3.04
new ∗[1]	Lhca1 (*cab6a*/*6b*)	YPGGAFDPLGYSK	168−180	P12360	2	0.2	4.06
		KYPGGAFDPLGYSK	167−180	P12360	2	0.0	3.37
		WAMLAVPGIIVPEALGLGNWVK	94−115	P12360	2	−0.4	3.51
		DPAKFEELK	181−189	P12360	1	0.6	2.06
		FEELK	185−189	P12360	1	0.2	1.56
	Lhca3 (*cab8*)	WLAYGEVINGR	95−105	P27522	2	0.5	3.43
		GLGGSGDPAYPGGPLFNPLGFGK	191−213	P27522	2	0.6	3.37
	Lhca4 (*cab11*/*12*)	WQDIKNPGSVNQDPIFK		S14305/6	2	0.3	3.85
		NPGSVNQDPIFK		S14305/6	2	−0.9	3.25
		NPGSVNQDPIFK		S14305/6	1	0.9	2.37
		WYDAGK		S14305/6	1	0.0	1.75
	Lhca4 (*cab11*)	WFIQAELVNGR	90−100	S14305	2	0.1	3.98
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.3	4.43
	Lhca5	DLPVWYEAGATK	123−134		2	0.5	3.98
new ∗[2]	Lhca1 (*cab6a*/*6b*)	WAM[#]LAVPGIIVPEALGLGNWVK[f]	94−115	P12360	3	0.4	4.91
		WAMLAVPGIIVPEALGLGNWVK	94−115	P12360	2	0.2	4.12
		KYPGGAFDPLGYSK	167−180	P12360	3	0.2	3.90
		YPGGAFDPLGYSK	168−180	P12360	2	0.0	3.54
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S13405/6	2	0.1	3.81
		WQDIKNPGSVNQDPIFK		S13405/6	3	0.4	4.37
		WYDAGK		S13405/6	1	0.1	2.03
	Lhca4 (*cab11*)	WFIQAELVNGR	90−100	S13405	2	−0.7	3.59
	Lhca4 (*cab12*)	WFVQAEVNGR	89−99	S13406	2	0.2	4.14
	Lhca5	DLPVWYEAGATKFNF	123−137		2	0.1	3.16
Lhca2[1]	Lhca1 (*cab6a*/*6b*)	KYPGGAFDPLGYSK	167−180	P12360	3	−0.7	3.66
		KYPGGAFDPLGYSK	167−180	P12360	2	0.8	2.76
	Lhca2 (*cab7*)	WAMLGAAGIFIPELLTK	113−129	P10708	2	0.8	3.94
		LTGTDVGYPGGLWFDPLGWGSGSPAK	189−214	P10708	2	−0.1	3.89
		WADIIK	169−174	P10708	1	0.2	1.51
	Lhca3 (*cab8*)	FAMLGAAGAIAPEILGK	106−122	P27522	2	0.9	3.26
	Lhca4 (*cab11*)	WFIQAELVNGR	90−100	S13405	2	0.7	3.57
	Lhca5	DLPVWYEAGATK	123−134		2	−0.4	3.65
Lhca3[1]	Lhca1 (*cab6a*/*6b*)	YPGGAFDPLGYSK	168−180	P12360	2	0.0	4.71
	Lhca2 (*cab7*)	LTGTDVGYPGGLWFDPLGWGSGSPAK	189−214	P10708	2	0.4	4.07
	Lhca3 (*cab8*)	FAMLGAAGAIAPEILGK	106−122	P27522	3	−0.8	4.33
		FAMLGAAGAIAPEILGK	106−122	P27522	2	−0.1	2.35
		GLGGSGDPAYPGGPLFNPLGFGKDEK	191−216	P27522	3	0.5	4.27
		GLGGSGDPAYPGGPLFNPLGFGK	191−213	P27522	2	0.0	3.76
		WLAYGEVINGR	95−105	P27522	2	1.1	3.35
		FQDWAKPGSMGK	171−182	P27522	2	0.1	2.93
	CP22	SALGLSEGGPLFGFTK	188−203	P54773	2	−0.6	4.63
		FVDDPTPPTGLEK	165−177	P54773	2	0.0	3.87

[a] The charge state of the measured ion (*z*), the deviation of the experimentally determined mass from the theoretical average mass of the peptide (Δ*M* in daltons; ± means that the calculated mass is larger or smaller than the measured mass), and the cross-correlation factor calculated by the Sequest algorithm (X_{corr}) are listed. [b] Labeling of the bands according to immunoblot assignment as in Figure 1B. [c] Proteins detected by MS/MS in proteolytically digested gel bands. The names of proteins and genes (in parentheses) are given. [d] Peptides found by a database search done with MS/MS data. [e] Position of the identified peptide in the protein sequence retrieved from the given data bank. The GenBank accession numbers are given. For peptides that match both Lhca4 isoforms (*cab11*/*12*), no peptide position is given as it differs in both isoforms. Positions of identified Lhca5 peptides are given for the precursor sequence shown in Figure 3. [f] Oxidation of methionine indicated.

peptides characteristic for several Lhca proteins. One of these is a peptide specific for Lhca5. In addition, numerous peptides of Lhca1 and the two isoforms of Lhca4 are present in this band. The occurrence of Lhca1 was expected from

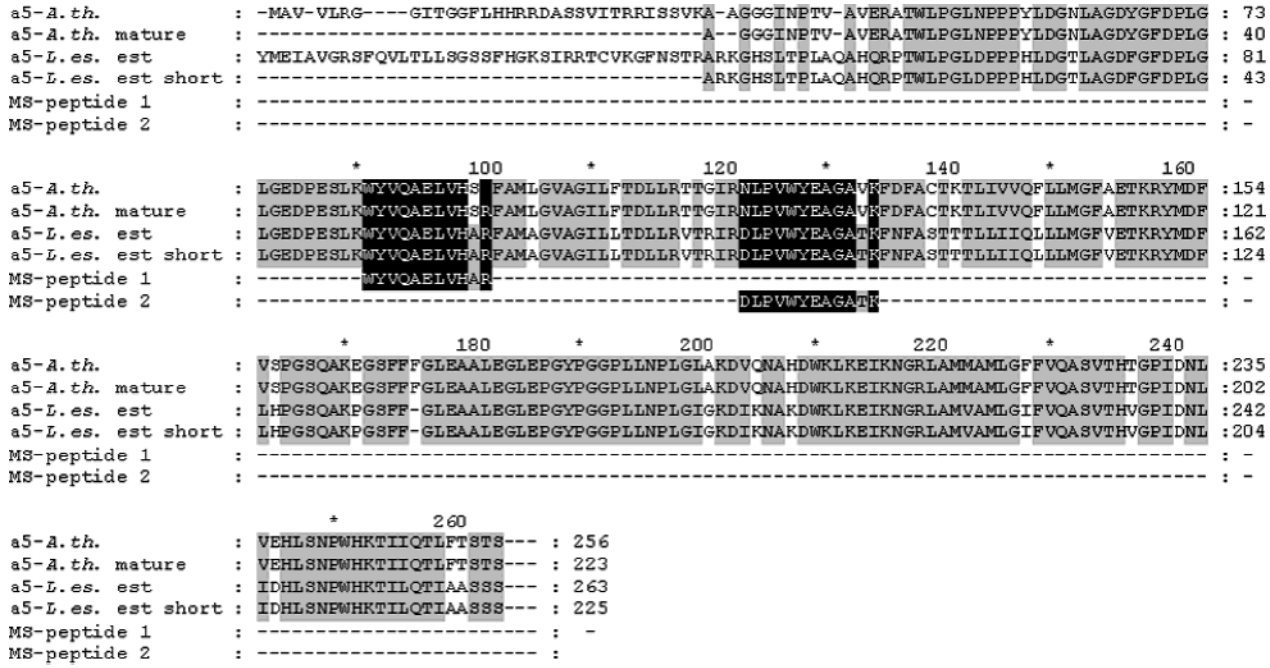

FIGURE 3: Alignment of the amino acid sequence of Lhca5 of *A. thaliana* (*21*; GenBank accession number AF134121), the reconstructed amino acid sequence of Lhca5 of *L. esculentum*, and Lhca5 peptides identified by MS/MS. The protein sequence of the Lhca5 protein of tomato was obtained by overlapping the EST clones (GenBank accession numbers AW041324 and AI778223) (see the text for further details).

the immunoblot analyses with the Lhca1 antibody. Finally, two Lhca3 peptides were detected also. These peptides were observed for only the band from the PSI holocomplex, but not for that from LHCI-730. This finding indicates that these two Lhca3 peptides are contaminations caused by smearing of Lhca3, which is found mainly in the 23.7 kDa band, as was demonstrated by immunodetection.

Surprisingly, fragments of all Lhca proteins could be detected in the 23.2 kDa band, which was specifically recognized by the Lhca2 antibody. However, only for Lhca2 were three different peptides found indicating its enrichment in this band in comparison with other bands. Additionally, single peptides of Lhca1, Lhca3, the *cab11* gene product, and Lhca5 could be detected.

The Lhca1 and Lhca2 proteins each could be identified by one peptide in the uppermost band at 23.7 kDa. In contrast, five different peptides characteristic of Lhca3 were found. This indicates the dominance of Lhca3 in this band in concurrence with the immunoblots. Interestingly, two additional peptides that matched CP22 of photosystem II were identified. Since no LHCII proteins could be detected, this may indicate weak contamination of the PSI preparation with PSII cores. In summary, highly sensitive MS/MS demonstrated that bands resolved by 1-D gel electrophoresis usually contain more than only one protein and thus confirms the results obtained with immunoblots.

To learn more about the heterogeneity of LHCI proteins, we used 2-D gel electrophoresis as described in ref *29* to separate PSI. Spots from 2-D gels were identified again by either immunoblot analyses or MS/MS. The top panel of Figure 4A shows a complete 2-D gel of PSI proteins stained with silver. Enrichment of neutral to slightly acidic proteins of the PSI core in the top right part and of LHCI proteins at more acidic pH in the middle section of the gel is clearly visible. In the bottom panel of Figure 4A, the typical resolution of the LHCI region is shown in more detail. Depending on the staining intensity, the individual horizontal spot rows consist of four to five spots for each Lhca protein. The assignment of the spots in Figure 4A is based on immunoblots like those shown in Figure 4B. All spots in silver-stained gels assigned to Lhca3, Lhca2, and Lhca1 on the basis of apparent molecular mass were recognized by antibodies raised against the respective protein. By contrast, due to the weak reaction of the Lhca4 antibody with Lhca4, only the Lhca4 spots that were stained most intensely by silver were detected. It is conspicuous that each Lhca protein forms one or two spot row(s) with proteins of the same molecular mass, but with different isoelectric points, indicative of protein isoforms with different isoelectric points. The row recognized by the Lhca2 antibody has the most acidic isoelectric points. The ones detected by the Lhca3 antibody exhibit the least acidic pI, and those of Lhca1 and Lhca4 are intermediate. This corresponds well with the theoretical pI values calculated for the different Lhca proteins. In addition to the heterogeneity regarding isoelectric points, Lhca3, Lhca1, and Lhca4 show a heterogeneity in molecular mass as is visible from detection of two spot rows by immunoblotting of these proteins (Figure 4B).

To gain specific amino acid sequence information from the proteins comprising individual spots resolved by 2-D gel electrophoresis, prominent spots were excised from gels, digested with trypsin, and analyzed by MS/MS. Table 2 summarizes the data from these analyses, and the assignment of the Lhca proteins to different spots is depicted in Figure 4C. In the spots recognized by the Lhca3 antibody, which have the same isoelectric point (pI ~6) but migrate at different molecular mass, only peptides corresponding to Lhca3 (labeled A and B in Figure 4C) were identified. By contrast, with the exception of spot C, all other spots that

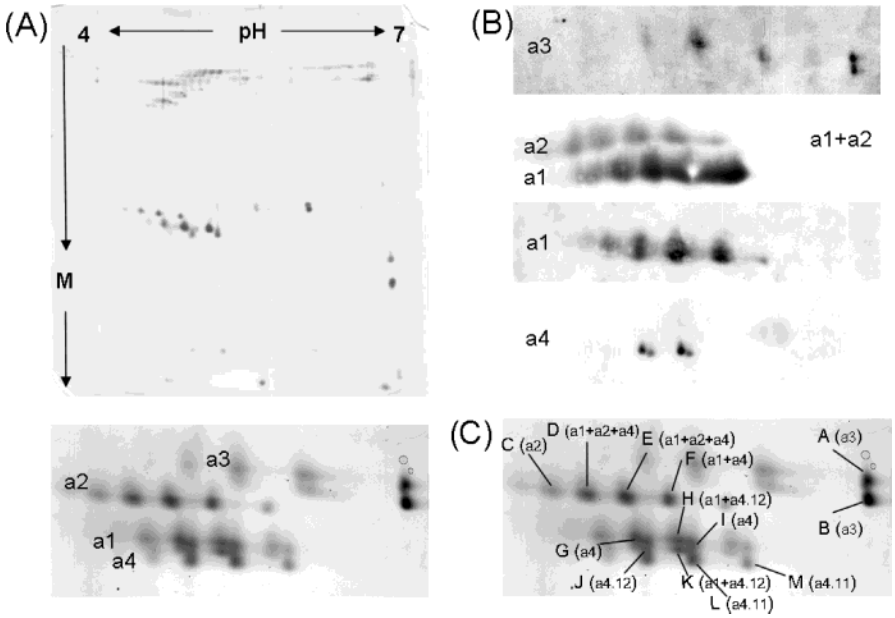

FIGURE 4: Protein maps of the PSI holocomplex obtained by 2-D gel electrophoresis using isoelectric focusing in the first and fully denaturing SDS−PAGE in the second dimension. In the top portion of panel A, the whole 2-D map stained with silver is shown. The bottom panel depicts the LHCI region at a higher magnification. The assignment of the different spot rows to the LHCI proteins was carried out according to the immunoblots shown in panel B. Depicted are immunoblots, which were developed with antibodies raised against the four different LHCI apoproteins (Lhca1−4) as indicated. Identification of Lhca1, Lhca3, and Lhca4 was achieved via incubation with a single antibody (a1, a3, and a4), whereas immunodetection of Lhca2 occurred simultaneously with that of Lhca1 to illustrate the relative position of the two spot rows. M is the molecular mass; a1−a4 are Lhca1−4, respectively. (C) Protein map of the molecular mass region of LHCI with Lhca spots identified by mass spectrometry as indicated: a1, Lhca1; a3, Lhca3; a4.11, Lhca4 derived from *cab11*; a4.12, Lhca4 derived from *cab12*; a4, Lhca4 derived from *cab11* or *cab12*.

were analyzed contained at least one of the Lhca4 isoforms. Spots L and M with higher pI contained the *cab11* gene product, whereas in the spots H, J, and K with a lower pI the *cab12* gene product was found. In spots D−G and I, peptides characteristic of Lhca4 were found that are common to both *cab11* and *cab12*. In spots D−F, H, and K, the Lhca1 protein (*cab6a/cab6b*) was also detected. Peptides identifying Lhca2 could be detected in spots C−E only after the amount of protein loaded on the gels and the trypsin concentration used for digestion were increased (see Experimental Procedures).

One possible explanation for the different electrophoretic forms of each Lhca protein could be that the proteins are phosphorylated to different degrees as suggested from a study with barley and antibodies that recognize phosphorylated proteins (*36*). To examine phosphorylation of Lhca proteins in tomato, proteins of PSI and thylakoids were separated in a 1-D gel. Subsequent detection of phosphorylated proteins revealed strong staining of phosphoproteins of photosystem II (e.g., Lhcb2, D1, and CP43), but not of Lhca proteins (data not shown). This shows that there is no significant phosphorylation of Lhca proteins, which could explain the electrophoretic behavior observed by 2-D electrophoresis.

DISCUSSION

Using high-resolution 1-D and 2-D gel electrophoresis in combination with immunoblotting and tandem mass spectrometry, our data yield a highly diverse protein composition of the PSI light-harvesting antenna of tomato. The data revealing this great diversity of Lhca proteins are summarized in Figure 5, which shows a 2-D map of the LHCI mass region. The assignment of individual spots to individual Lhca proteins is based on molecular mass determination, immunoblotting, and MS/MS analysis. Highly sensitive MS/MS additionally allowed detection of Lhca5 and of two different isoforms of Lhca4 for the first time on the protein level in a higher plant.

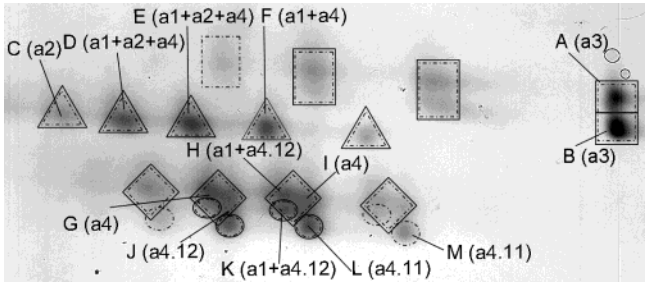

FIGURE 5: 2-D map summarizing the Lhca proteins of photosystem I. Identification of Lhca proteins is based on molecular mass determination of silver-stained gels (-•-), immunoblotting (—), and mass spectrometry (indicated with letters) of Lhca3 (rectangles), Lhca2 (triangles), Lhca1 (squares), and Lhca4 (ellipsoids).

Heterogeneity of the PSI Light-Harvesting Antenna. High-resolution SDS−PAGE resolved five bands in the molecular mass range of light-harvesting proteins (Figure 1A). Four of these bands are stained strongly by coomassie and are assigned to the well-known Lhca1−4 proteins. In addition, this gel system allowed resolution of a fifth band, which until now has been described for only barley (*13*) and tomato (*15*). From 1-D gels and immunoblotting with monospecific antibodies, it is obvious that the four bands attributed to Lhca1−4 do not contain only the eponymous proteins. The Lhca4 antibody reacted with the Lhca4 band and additionally with the band usually assigned to Lhca1. The reactivity of the Lhca1 antibody with the Lhca2 band indicated a heterogeneous protein composition of this band as well. This heterogeneity in molecular mass observed for Lhca4 and

Table 2: Protein Composition of Spots Resolved by 2-D Gels of PSI As Determined by MS/MS[a]

2-D spot[b]	protein[c]	peptide[d]	position[e]	ref[e]	*z*	Δ*M*	X_{corr}
A	Lhca3 (*cab8*)	WLAYGEVINGR	95−105	P27522	2	−0.5	2.94
		FAMLGAAGAIAPEILGK	106−122	P27522	2	−0.9	2.67
B	Lhca3 (*cab8*)	GLGGSGDPAYPGGPLFNPLGFGK	191−213	P27522	2	1.3	4.21
		WLAYGEVINGR	95−105	P27522	2	0.9	3.70
		FAMLGAAGAIAPEILGK	106−122	P27522	2	1.3	2.98
		FQDWAKPGSMGK	171−182	P27522	2	0.9	3.25
C	Lhca2 (*cab7*)	PGC#VNTDPIFPNNK[f]	175−188	P10708	2	0.1	2.52
D	Lhca2 (*cab7*)	PGC#VNTDPIFPNNK[f]	175−188	P10708	2	−1.2	2.62
	Lhca1 (*cab6a*/*6b*)	KYPGGAFDPLGYSK	167−180	P12360	2	−0.4	2.49
		YKESELIHC#R[f]	84−93	P12360	2	−0.2	2.35
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.4	2.85
E	Lhca1 (*cab6a*/*6b*)	YPGGAFDPLGYSK	168−180	P12360	2	0.9	4.63
		KYPGGAFDPLGYSK	167−180	P12360	2	1.1	3.93
	Lhca2	WADIIKPGC#VNTDPIFPNNK[f]	169−188	P10708	3	2.2	5.05
		WAM*LGAAGIFIPELLTK[g]	113−129	P10708	2	1.3	4.20
		W§AMLGAAGIFIPELLTK[h]	113−129	P10708	2	1.3	4.19
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.8	3.68
F	Lhca1 (*cab6a*/*6b*)	KYPGGAFDPLGYSK	167−180	P12360	2	0.0	3.20
		YPGGAFDPLGYSK	168−180	P12360	2	−1.2	3.06
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	−1.2	3.40
G	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	−0.9	3.61
H	Lhca1 (*cab6a*/*6b*)	YPGGAFDPLGYSK	168−180	P12360	2	−0.5	3.74
		KYPGGAFDPLGYSK	167−180	P12360	3	1.2	4.20
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.7	3.69
		WQDIKNPGSVNQDPIFK	157−173	S14305/6	3	1.2	3.83
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.0	3.94
I	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.0	2.79
J	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.2	3.98
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.1	3.69
K	Lhca1 (*cab6a*/*6b*)	YPGGAFDPLGYSK	168−180	P12360	2	0.1	2.68
	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.3	3.81
		WQDIKNPGSVNQDPIFK		S14305/6	3	1.2	4.80
	Lhca4 (*cab12*)	WFVQAELVNGR	89−99	S14306	2	0.8	3.98
L	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.8	4.03
		WQDIKNPGSVNQDPIFK		S14305/6	3	1.2	5.37
	Lhca4 (*cab11*)	C#GYPGGIFNPLNFAPTEEAK[f]	182−201	S14305	2	0.9	4.55
M	Lhca4 (*cab11*/*12*)	NPGSVNQDPIFK		S14305/6	2	0.8	3.74
		WQDIKNPGSVNQDPIFK		S14305/6	3	1.4	3.60
	Lhca4 (*cab11*)	C#GYPGGIFNPLNFAPTEEAK[f]	182−201	S14305	2	1.4	4.90
		WFIQAELVNGR	90−100	S14305	2	0.8	4.70

[a] Spots were excised from gels, proteolytically hydrolyzed, and analyzed by tandem mass spectrometry. For further details, see footnote a of Table 1. [b] The labeling of the spots is according to Figure 4C. [c] Proteins identified in the 2-D spots by MS/MS. [d] Peptides found by matching the MS data with tryptic peptides derived from proteins in tomato databases. [e] Position of the identified peptide within the protein sequence retrieved from the given data bank. The GenBank accession numbers are given. For peptides that match both Lhca4 isoforms (*cab11*/*12*), no peptide position is given as it differs in both isoforms. [f] Carbamidomethylation of cysteine indicated (+57 Da). [g] Oxidation of methionine indicated (+16 Da). [h] Keto, amino, or hydroxy modification of tryptophan indicated (+16 Da).

Lhca1 also could be detected by 2-D electrophoresis and subsequent immunodetection. Two-dimensional protein maps indicate that Lhca3 also exists in two isoforms with differing molecular masses. Due to the absence of a stacking gel in the second-dimension gel, the separation distance is longer than in 1-D gels. Therefore, separation of two differently migrating Lhca3 isoforms solely in 2-D and not in 1-D gels may be the result of a better resolution at this molecular mass in the second-dimension gel. Only Lhca2 did not show a heterogeneous behavior regarding electrophoretic mobility in 1-D or 2-D gels as judged from immunoblot analyses. MS/MS analyses, which are more sensitive, confirmed these results (Tables 1 and 2). The immunoblots and MS/MS data demonstrate that the light-harvesting antenna of PSI is very heterogeneous, and that the assignment of the four bands obtained by standard 1-D gel electrophoresis to individual proteins is an oversimplification. Therefore, quantification of individual Lhca proteins that relies on the staining intensity of these individual bands can be erroneous. In addition, our MS/MS analyses of trypsin-digested spots demonstrate the superior resolution of 2-D gels versus 1-D gels. In the majority of analyzed spots of 2-D gels, only one Lhca protein could be detected. From this observation, it can be concluded that minor constituents comigrating with the dominating protein in a band in 1-D gels are separated in 2-D gels as a consequence of the different pIs and escape detection in 2-D gels.

Isoforms with different masses have already been reported for various proteins of PSII (*37*, *38*) and the PSI core complex (*39*). Additionally, isoforms of individual Lhca proteins were observed by using different technical approaches. Using a combination of weakly and fully denaturing SDS gel electrophoresis in the first and second dimension, respectively, two Lhca3 proteins with slightly different electrophoretic mobility were observed in spinach PSI (*12*). Additionally, the same authors sometimes observed an additional band in 1-D gels with an electrophoretic migration similar to that of Lhca2. By HPLC separation coupled with ESI-MS it was possible to identify Lhca2 isoforms in PSI of tomato (*28*). In addition, two Lhca1 isoforms encoded by

cab6a (*23*) and *cab6b* (*24*) could be detected with this approach. We were not able to detect peptides specific for products of *cab6a* or *cab6b* or for Lhca2 isoforms. This failure can be explained by the different methodical approaches used in this study and in ref *28*. In contrast to the latter study in which entire proteins were analyzed, we employed tryptic protein fragments for detection and identification of proteins. The two Lhca1 isoforms differ by only three amino acids. Therefore, the most probable explanation is that proteolytic fragments specific for single isoforms have not been detected. For Lhca2, it is conspicuous that MS analyses of bands of 1-D gels resulted in the detection of only a few peptides (Table 1). It is also surprising that in 2-D gels only in spots C−E were Lhca2 peptides found by MS (Table 2), although spot F clearly contains Lhca2 as could be demonstrated by immunoblotting (Figure 4B). This could be due to poor extraction of Lhca2 tryptic peptides from a polyacrylamide matrix and/or to poor electrospray ionization efficiencies of these peptides. Such an interpretation is consistent with the mass spectrometric analysis of tryptic peptides of the Lhca2 1-D band, which resulted in the identification of only three peptides, whereas for the 1-D bands of Lhca1, Lhca3, and Lhca4 (*cab11* and *cab12*), six, five, and four peptides, respectively, were recognized (Table 1).

Due to the use of the 2-D gel electrophoresis procedure recently developed for the separation of membrane proteins (*29*), we could also demonstrate several isoelectric points for individual Lhca proteins (Figure 4). This technique allowed resolution of four to five isoforms with different pI values for Lhca1−4. The 2-D protein maps show that Lhca3 is the least acidic Lhca protein whereas Lhca2 is the most acidic one, which is consistent with the theoretical pI values of 5.95 (Lhca3), 5.22 (Lhca1; *cab6a* and *cab6b*), and 4.99 (Lhca2) as calculated from the amino acid sequence. The resolving power of this 2-D gel electrophoresis, in combination with the specific amino acid sequence tags generated by MS/MS, allowed the resolution and identification of the two Lhca4 isoforms derived from *cab11* and *cab12* genes (*27*). The product of *cab12* possessing a theoretical pI of 5.15 could be clearly separated from the *cab11* product with a pI of 5.34 (Figure 4C and Table 2), and in this way, its first detection on the protein level was facilitated.

Detection of four to five isoforms resolved at different isoelectric points for each of the prominent Lhca proteins raises the question about the origin of these charge differences. One possible explanation are posttranslational modifications. Various proteins of PSII could be isolated in a phosphorylated form (*38*, *40*). However, with regard to LHCI proteins, it is not clear whether phosphorylation occurs. ^{32}P labeling of plant material followed by gel electrophoresis and autoradiography did not prove phosphorylation of LHCI proteins under conditions in which the various PSII proteins are phosphorylated (*41*, *42*). These results are in contrast to an immunoblot study performed with anti-phosphothreonine and anti-phosphoserine antibodies (*36*). Both antibodies cross reacted with all Lhca proteins that were tested (Lhca1−4), indicating phosphorylation at various sites in all proteins. Such extensive phosphorylation at different sites of one protein could explain a heterogeneity in pI as was observed in this study. However, in addition to shifts in pI, phosphorylation may also cause shifts in electrophoretic mobility as was observed for phosphorylated PSII proteins, e.g., D1 and CP29 (*43*, *44*). To test whether the Lhca proteins of tomato are phosphorylated to an extent that would explain the spot pattern obtained by 2-D gels, we subjected PSI proteins to 1-D electrophoresis and phosphoprotein staining. However, we could not detect significant phosphorylation, which would support the idea that phosphorylation causes the heterogeneity of individual Lhca proteins. Another explanation for the heterogeneity in molecular mass observed for Lhca4, Lhca1, and Lhca3 may result from different processing of individual Lhca precursor proteins. Processing of LHC proteins at different sites has been reported in higher plants (*39*, *45*, *46*) and the green alga *C. reinhardtii* (*22*). Because the size differences observed in these investigations are comparable to those detected by us, differential processing may also be germane to the distinct electrophoretic behavior found for Lhca proteins of tomato. In addition to posttranslational modifications such as phosphorylation, acetylation or modification of the peptide amino terminus or the side chains of lysine or arginine by isocyanic acid (carbamylation) could be an explanation for the different Lhca isoforms. We specifically searched the MS/MS data of all Lhca proteins for protein carbamylation and acetylation. We also aimed to identify differentially N-terminally processed Lhca proteins by searching the MS/MS data for such processing events (see Experimental Procedures). However, no carbamylation or acetylation was detected, nor was it possible to determine the identity of any differentially N-terminally processed proteins. However, the existence of such modifications cannot be ruled out by the applied MS/MS procedure since only between 17% (*cab12*) and 36% (Lhca1) of the entire proteins were covered in the 1-D gel bands and even less in the spots of 2-D gels (Tables 1 and 2). Additionally, no peptides were detected that derive from the N-terminus, which is especially predestined to differential processing and protein modifications. Therefore, we are left with the speculation that the distinct Lhca isoforms are caused by posttranslational modifications, protein processing, or even charge heterogeneity that occur in the protein segments, which have not been detected. Further proteomic analyses using other proteases for in-gel digestion may be suitable for clarifying the origin of the distinct electrophoretic behavior. Additionally, plants grown under different environmental conditions (e.g., high light, low temperature, and nutrient deficiency) hold promise for providing insight into the physiological significance of these isoforms.

Detection of Lhca5 on the Protein Level in a Higher Plant. Our approach allowed not only identification of Lhca4 isoforms (gene products of *cab11* and *cab12*) but also detection of Lhca5, a protein that was not yet identified on the protein level in higher plants. The *lhca*5 gene has a low expression level and was first described for *A. thaliana* (*21*). Very recently, a detailed MS/MS analysis indicated its presence in the green alga *C. reinhardtii* (*22*). In the study presented here, combining electrophoresis and MS/MS with bioinformatic analyses, we were able to demonstrate the presence of Lhca5 on the protein level in a higher plant (Table 1). A contig (TC116242) resulting in a protein with the same amino acid sequence as in Figure 3 that lacks the final five amino acids (AASSS) can be found at The Institute for Genomic Research (TIGR) tomato gene index (http://www.tigr.org/tdb/tgi/lgi). This confirms the sequence of the reconstructed Lhca5.

As a consequence of the presence of Lhca5 in bands with apparent molecular masses of 22.3, 22.7, and 23.2 kDa, it appears that this protein also occurs in different isoforms. This corresponds to the observation in *C. reinhardtii* in which two different isoforms of Lhca5 have been found (*22*). On the basis of the molecular masses, one can speculate about the processing site at which the protein is cleaved during maturation. For *A. thaliana*, it was suggested that the mature protein starts with the sequence AAGGGIN..., which corresponds to a protein of 24.1 kDa (*21*). By using the net-based ChloroP software (at http://www.cbs.dtu.dk/services/ChloroP/), we obtained the same processing site for tomato Lhca5. Thus, if one assumes that the mature tomato Lhca5 starts with the sequence ARKGHS... (Figure 3), a protein of 24.5 kDa is obtained, which does not correspond to the observed molecular masses. Recent analyses demonstrated that the predictions of the ChloroP software are only reliable for soluble proteins and not for membrane proteins (*47*), which may also apply for Lhca5. Therefore, we calculated the molecular masses for shortened versions of Lhca5. Given the same N-terminus for tomato as suggested for *A. thaliana* (*21*), removal of 12, 16, and 19 amino acids from the N-terminus of tomato Lhca5 would result in proteins of 23.2, 22.7, and 22.3 kDa, respectively, which would fit to the molecular masses obtained by 1-D SDS−PAGE. Therefore, we assume that mature Lhca5 of tomato starts with one of the following three amino acid sequences: LPGLDP (minus 19 amino acids), PTWLPG (minus 16 amino acids), and AHQRPT (minus 12 amino acids). Due to the presence of Lhca5 in LHCI-730, it may interact with either Lhca1 or Lhca4. It will be interesting to perform dimerization experiments as described in ref *48* with Lhca5 and other Lhca proteins to elucidate the dimer formation capability of this protein.

Significance of LHCI Heterogeneity for PSI Composition. Depending on the staining intensity and resolution of 1-D and 2-D gels, we were able to resolve five different Lhca proteins occurring in several isoforms. This variety of Lhca proteins found in the present work and by Zolla and coworkers (*28*) for Lhca2 is surprising considering that according to the recently published crystal structure of PSI from pea (*19*) and single-particle analyses of PSI from spinach (*18*) isolated PSI complexes of higher plants contain between four and eight Lhca proteins per PSI core complex. The detected variety indicates the existence of different populations of PSI in higher plants, which differ in their Lhca protein composition. In fact, using phase partitioning, the existence of two different PSI populations (PSI_α and PSI_β) was demonstrated (*49*). Immunoblot analyses showed that PSI_α located in the inner parts of grana thylakoids possesses a smaller Lhca antenna than PSI_β localized in stroma thylakoids (*50*). In addition, it was shown recently that high levels of light resulted in a selective reduction of Lhca4 and Lhca1 (*51*). Finally, it can be inferred from analyses of mutants of barley and *A. thaliana* that PSI complexes with a varying complement of antenna proteins exist (*52*−*54*). All these examples show that PSI with varying Lhca composition occurs depending on the location in the thylakoid membrane, physiological state, and genotype. Our results imply that even under normal growth conditions several PSI populations differing in their Lhca protein composition are present. Therefore, it will be interesting to analyze plants grown under different environmental conditions with regard to up- or downregulation of individual Lhca proteins and their isoforms. This will shed light on the physiological significance of Lhca heterogeneity and the existence of PSI with different Lhca compositions.

NOTE ADDED AFTER ASAP POSTING

After being posted on the Web on 06/26/04, this paper was altered (subscript Greek characters in last paragraph of Discussion). The correct version was posted 06/29/04.

ACKNOWLEDGMENT

We thank Prof. L. A. Staehelin (Department of Molecular, Cellular and Developmental Biology, University of Colorado) for the kind gift of Lhca2 antibody and Christine Markert (Lehrstuhl für Pflanzenphysiologie, Friedrich Schiller-Universität Jena) for help with MS measurements.

REFERENCES

1. Haldrup, A., Jensen, P. E., Lunde, C., and Scheller, H. V. (2001) Balance of power: a view of the mechanism of photosynthetic state transitions, *Trends Plant Sci. 6*, 301−305.
2. Demmig-Adams, B. (1990) Carotenoid and photoprotection in plants: A role for the xanthophyll zeaxanthin, *Biochim. Biophys. Acta 1020*, 1−24.
3. Yamamoto, H., and Bassi, R. (1996) Carotenoids: Localization and Function, in *Oxygenic Photosynthesis: The Light Reactions* (Ort, D. R., and Yocum, C. F., Eds.) pp 507−521, Kluwer Academic Publishers, Dordrecht, The Netherlands.
4. Peter, G. F., and Thornber, J. P. (1991*)* Biochemical composition and organization of higher plant photosystem II light-harvesting pigment−proteins, *J. Biol. Chem. 266*, 16745−16754.
5. Bassi, R., Pineau, B., Dainese, P., and Marquardt, J. (1993) Carotenoid-binding proteins of photosystem II, *Eur. J. Biochem. 21*, 297−303.
6. Haworth, P., Watson, J. L., and Arntzen, C. J. (1983) The detection, isolation, and characterization of a light-harvesting complex which is specifically associated with photosystem I, *Biochim. Biophys. Acta 724*, 151−158.
7. Lam, E., Ortiz, W., and Malkin, R. (1984) Chlorophyll a/b proteins of photosystem I, *FEBS Lett. 168*, 10−14.
8. Bassi, R., Machold, O., and Simpson, D. (1985) Chlorophyll-proteins of two photosystem I preparations from maize, *Carlsberg Res. Commun. 50*, 145−162.
9. Jansson, S., Andersen, B., and Scheller, H. V. (1996) Nearest-neighbor analysis of higher-plant photosystem I holocomplex, *Plant Physiol. 112*, 409−420.
10. Schmid, V. H. R., Cammarata, K. V., Bruns, B. U., and Schmidt, G. W. (1997) In vitro reconstitution of the photosystem I light harvesting complex LHCI-730: heterodimerization is required for antenna pigment organization, *Proc. Natl. Acad. Sci. U.S.A. 94*, 7667−7672.
11. Croce, R., Morosinotto, T., Castelletti, S., Breton, J., and Bassi, R. (2002) The Lhca antenna complexes of higher plant photosystem I, *Biochim. Biophys. Acta 1556*, 29−40.
12. Ikeuchi, M., Hirano, A., and Inoue, Y. (1991) Correspondence of apoproteins of light-harvesting chlorophyll a/b complexes associated with photosystem I to cab genes: evidence for a novel type IV apoprotein, *Plant Cell Physiol. 32*, 103−112.
13. Knoetzel, J., Svendsen, I., and Simpson, D. J. (1992) Identification of the photosystem-I antenna polypeptides in barley: isolation of 3 pigment-binding antenna complexes, *Eur. J. Biochem. 206*, 209−215.
14. Tjus, S. E., Roobol-Boza, M., Palsson, L. O., and Andersson, B. (1995) Rapid isolation of photosystem I chlorophyll-binding proteins by anion exchange perfusion chromatography, *Photosynth. Res. 45*, 41−49.
15. Schmid, V. H. R., Potthast, S., Wiener, M., Bergauer, V., Paulsen, H., and Storf, S. (2002) Pigment Binding of Photosystem I Light-harvesting Proteins, *J. Biol. Chem. 277*, 37307−37314.

16. Castelletti, S., Morosinotto, T., Robert, B., Caffarri, S., Bassi, R., and Croce, R. (2003) Recombinant Lhca2 and Lhca3 Subunits of the Photosystem I Antenna System, *Biochemistry 42*, 4226−4234.
17. Boekema, E. J., Wynn, R. M., and Malkin, R. (1990) The structure of spinach photosystem I studied by electron microscopy, *Biochim. Biophys. Acta 1017*, 49−56.
18. Boekema, E. J., Jensen, P. E., Schlodder, E., van Breemen, J. F. L., van Roon, H., Scheller, H. V., and Dekker, J. P. (2001) Green plant photosystem I binds light-harvesting complex I on one side of the complex, *Biochemistry 40*, 1029−1036.
19. Ben-Shem, A., Frolow, F., and Nelson, N. (2003) Crystal structure of plant photosystem I, *Nature 426*, 630−635.
20. Pichersky, E., and Jansson, S. (1996) The light-harvesting chlorophyll a/b-binding polypeptides and their genes in angiosperm and gymnosperm species, in *Oxygenic Photosynthesis: The Light Reactions* (Ort, D. R., and Yocum, C. F., Eds.) pp 507−521, Kluwer Academic Publishers, Dordrecht, The Netherlands.
21. Jansson, S. (1999) A guide to the Lhc genes and their relatives in *Arabidopsis*, *Trends Plant Sci. 4*, 236−240.
22. Stauber, E. J., Fink, A., Markert, C., Kruse, O., Johanningmeier, U., and Hippler, M. (2003) Proteomics of *Chlamydomonas reinhardtii* light-harvesting proteins, *Eukaryotic Cell 2*, 978−994.
23. Hoffman, N. E., Pichersky, E., Malik, V. S., Castresana, C., Ko, K., Darr, S. C., and Cashmore, A. R. (1987) The nucleotide sequence of a tomato cDNA clone encoding a photosystem I protein with homology to photosystem II chlorophyll a/b-binding polypeptides, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8844−8848.
24. Pichersky, E., Hoffman, N. E., Bernatzky, R., Piechulla, B., Tanksley, S. D., and Cashmore, A. R. (1987) Molecular characterization and genetic mapping of DNA sequences encoding type I chlorophyll *a*/*b*-binding polypeptide of photosystem I in *Lycopersicon esculentum* (tomato), *Plant Mol. Biol. 9*, 205−216.
25. Pichersky, E., Tanksley, S. D., Piechulla, B., Stayton, M. M., and Dunsmuir, P. (1988) Nucleotide sequence and chromosomal location of cab-7, the tomato gene encoding the type II chlorophyll a/b-binding polypeptide of photosystem I, *Plant Mol. Biol. 11*, 69−71.
26. Pichersky, E., Brock, T. G., Nguyen, D., Hoffman, N. E., Piechulla, B., Tanksley, S. D., and Green, B. R. (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.
27. Schwartz, E., Shen, D., Aebersold, R., McGrath, J. M., Pichersky, E., and Green, B. R. (1991) Nucleotide sequence and chromosomal location of cab11 and cab12, the genes for the fourth polypeptide of the photosystem I light-harvesting antenna (LHCI), *FEBS Lett. 280*, 229−234.
28. Zolla, L., Rinalducci, S., Timperio, A. M., and Huber, C. G. (2002) Proteomics of Light-Harvesting Proteins in Different Plant Species. Analysis and Comparison by Liquid Chromatography-Electrospray Ionization Mass Spectrometry. Photosystem I, *Plant Physiol. 130*, 1938−1950.
29. Hippler, M., Klein, J., Fink, A., Allinger, T., and Hoerth, P. (2001) Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from *Chlamydomonas reinhardtii*, *Plant J. 28*, 595−606.
30. Paulsen, H., and Schmid, V. H. R. (2002) Analysis and Reconstitution of Chlorophyll-Proteins, in *Heme, Chlorophyll and Bilins: Methods and Protocols* (Smith, A. G., and Witty, M., Eds.) pp 235−253, Humana Press, Totowa, NJ.
31. Porra, R. J., Thompson, W. A., and Kriedemann, P. E. (1989) Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy, *Biochim. Biophys. Acta 975*, 384−394.
32. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680−685.
33. Wessel, D., and Fluegge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, *Anal. Biochem. 138*, 141−143.
34. Sigrist, M., and Staehelin, L. A. (1994) Appearance of type 1, 2, and 3 light-harvesting complex II and light-harvesting complex I proteins during light-induced greening of barley (*Hordeum vulgare*) etioplasts, *Plant Physiol. 104*, 135−145.
35. Mortz, E., Vorm, O., Mann, M., and Roepstorff, P. (1994) Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search, *Biol. Mass Spectrom. 23*, 249−261.
36. Knoetzel, J., Meyer, D. U., and Grimme, L. H. (1998) Phosphorylated Photosystem I antenna proteins in Barley, in *Photosynthesis: from Light to Biosphere* (Mathis, P., Ed.) Vol. 1, pp 131−134, Kluwer Academic Publishers, Dordrecht, The Netherlands.
37. Huber, C. G., Timperio, A.-M., and Zolla, L. (2001) Isoforms of Photosystem II Antenna Proteins in Different Plant Species Revealed by Liquid Chromatography-Electrospray Ionization Mass Spectrometry, *J. Biol. Chem. 276*, 45755−45761.
38. Gómez, S. M., Nishio, J. N., Faull, K. F., and Whitelegge, J. P. (2002) The Chloroplast Grana Proteome Defined by Intact Mass Measurements from Liquid Chromatography Mass Spectrometry, *Mol. Cell. Proteomics 1*, 46−59.
39. Obokata, J., Mikami, K., Hayashida, N., Nakamura, M., and Sugiura, M. (1993*)* Molecular Heterogeneity of Photosystem I, *Plant Physiol. 102*, 1259−1267.
40. Bennett, J. (1991) Protein phosphorylation in green plant chloroplasts, *Annu. Rev. Plant Physiol. Plant Mol. Biol. 42*, 281−311.
41. Dunahay, T. G., Schuster, G., and Staehelin, L. A. (1987) Phosphorylation of spinach chlorophyll-protein complexes, *FEBS Lett. 215*, 25−30.
42. Bassi, R., Giacometti, G. M., and Simpson, D. J. (1988) Changes in the organization of stroma membranes induced by in vivo state 1−state 2 transition, *Biochim. Biophys. Acta 935*, 152−165.
43. Elich, T. D., Edelman, M., and Mattoo, A. K. (1992) Identification, characterization, and resolution of the in vivo phosphorylated form of the photosystem II reaction center protein, *J. Biol. Chem. 267*, 3523−3529.
44. Bergantino, E., Sandona, D., Cugini, D., and Bassi, R. (1998) The photosystem II subunit CP29 can be phosphorylated in both C3 and C4 plants as suggested by sequence analysis, *Plant Mol. Biol. 36*, 11−22.
45. Clark, S. E., Abad, M. S., and Lamppa, G. K. (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.
46. Pichersky, E., Hoffman, N. E., Malik, V. S., Bernatzky, R., Szabo, L., and Cashmore, A. R. (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.
47. Gómez, S. M., Bil', K. Y., Aguilera, R., Nishio, J. N., Faull, K. F., and Whitelegge, J. P. (2003) Transit peptide cleavage sites of integral thylakoid membrane proteins, *Mol. Cell. Proteomics 2*, 1068−1085.
48. Schmid, V. H. R., Paulsen, H., and Rupprecht, J. (2002) Identification of N- and C-terminal amino acids of Lhca1 and Lhca4 required for formation of the heterodimeric peripheral photosystem I antenna LHCI-730, *Biochemistry 41*, 9126−9131.
49. Svensson, P., Andreasson, E., and Albertsson, P.-A. (1991) Heterogeneity among photosystem I, *Biochim. Biophys. Acta 1060*, 45−50.
50. Jansson, S., Stefansson, H., Nyström, U., Gustafsson, P., and Albertsson, P.-A. (1997) Antenna protein composition of PS I and PS II in thylakoid subdomains, *Biochim. Biophys. Acta 1320*, 297−309.
51. Bailey, S., Walters, R. G., Jansson, S., and Horton, P. (2001) Acclimation of *Arabidopsis thaliana* to the light environment: the existence of separate low light and high light responses, *Planta 213*, 794−801.
52. Bossmann, B., Knoetzel, J., and Jansson, S. (1997) Screening of chlorina mutants of barley (*Hordeum vulgare* L.) with antibodies against light-harvesting proteins of PSI and PSII: absence of specific antenna proteins, *Photosynth. Res. 52*, 127−136.
53. Ganeteg, U., Strand, A., Gustafsson, P., and Jansson, S. (2001) The Properties of the Chlorophyll *a*/*b*-Binding Proteins Lhca2 and Lhca3 Studied in Vivo Using Antisense Inhibition, *Plant Physiol. 127*, 150−158.
54. Zhang, H., Goodman, H. M., and Jansson, S. (1997) Antisense inhibition of the Photosystem I Antenna Protein Lhca4 in *Arabidopsis thaliana*, *Plant Physiol. 11*, 1525−1531.

BI0498196