Proteomic characterization and three-dimensional electron microscopy study of PSII-LHCII supercomplexes from higher plants.

Pagliano, C; Nield, J; Marsano, F; Pape, T; Barera, S; Saracco, G; Barber, J

doi:10.1016/j.bbabio.2013.11.004

Title

Proteomic characterization and three-dimensional electron microscopy study of PSII-LHCII supercomplexes from higher plants

Author names and affiliations

Cristina Pagliano[a*], Jon Nield[b], Francesco Marsano[c], Tillmann Pape[d], Simone Barera[c], Guido Saracco[a], James Barber[a,d*]

[a]Applied Science and Technology Department - BioSolar Lab, Politecnico di Torino, Viale T. Michel 5, 15121 Alessandria, Italy

[b]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom

[c]Department of Science and Technological Innovation - University of Piemonte Orientale, Viale T. Michel 11, 15121 Alessandria, Italy

[d]Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, United Kingdom

Corresponding authors*

Cristina Pagliano E-mail: cristina.pagliano@polito.it; Tel. No. +39 131 229301; Fax No. +39 131 229344; Politecnico di Torino, Applied Science and Technology Department - BioSolar Lab, Viale T. Michel 5, 15121 Alessandria, Italy

James Barber E-mail: j.barber@imperial.ac.uk; Tel. No. +44 208 747 1165; Fax. No. +44 207 594 5267; Imperial College London, Department of Life Sciences, Faculty of Natural Sciences, London SW7 2AZ, United Kingdom

E-mail addresses of other authors:

Jon Nield j.nield@qmul.ac.uk

Francesco Marsano francesco.marsano@mfn.unipmn.it

Tillmann Pape t.pape@imperial.ac.uk

Simone Barera simone.barera@mfn.unipmn.it.

Guido Saracco guido.saracco@polito.it

Abstract

In higher plants a variable number of peripheral LHCII trimers can strongly (S), moderately (M) or loosely (L) associate with the dimeric PSII core (C_2) complex via monomeric Lhcb proteins to form PSII-LHCII supercomplexes with different structural organizations.
By solubilizing isolated stacked pea thylakoid membranes either with the α or β isomeric forms of the detergent n-dodecyl-D-maltoside, followed by sucrose density ultracentrifugation, we previously showed that PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2, respectively, can be isolated [S. Barera et al., Phil. Trans. R Soc. B 67 (2012) 3389–3399]. Here we analyzed their protein composition by applying extensive bottom-up and top-down mass spectrometry on the two forms of the isolated supercomplexes. In this way, we revealed the presence of the antenna proteins Lhcb3 and Lhcb6 and of the extrinsic polypeptides PsbP, PsbQ and PsbR exclusively in the $C_2S_2M_2$ supercomplex. Other proteins of the PSII core complex, common to the $C_2S_2M_2$ and C_2S_2 supercomplexes, including the low molecular mass subunits, were also detected and characterized.
To complement the proteomic study with structural information, we performed negative stain transmission electron microscopy and single particle analysis on the PSII-LHCII supercomplexes isolated from pea thylakoid membranes solubilized with n-dodecyl-α-D-maltoside. We observed the $C_2S_2M_2$ supercomplex in its intact form as the largest PSII complex in our preparations. Its dataset was further analyzed *in silico*, together with that of the second largest identified sub-population, corresponding to its C_2S_2 subcomplex. In this way, we calculated 3D electron density maps for the $C_2S_2M_2$ and C_2S_2 supercomplexes, approaching respectively 30 and 28 Å resolution, extended by molecular modelling towards the atomic level.

Keywords: thylakoids; PSII-LHCII supercomplex; proteomics; transmission electron microscopy; single particle analysis; structure

Abbreviations: BN-PAGE, blue native polyacrylamide gel electrophoresis; Chl, chlorophyll; 1D/2D SDS-PAGE, mono-dimensional/bi-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis; 3D, three-dimensional; α-DM, n-dodecyl-α-D-maltoside; β-DM, n-dodecyl-β-D-maltoside; FEG, field emission gun; FSC, Fourier shell correlation; LC-ESI-MS/MS, liquid chromatography electrospray ionization tandem mass spectrometry; LHC, light harvesting complex; LMM, low molecular mass; MALDI-TOF/TOF, matrix-assisted laser desorption/ionization-time of flight/time of flight; MS, mass spectrometry; OEC, oxygen evolving complex; PS, photosystem; PTM, post translational modification; RC, reaction centre; TEM, transmission electron microscopy

1. Introduction

Photosystem II (PSII) is one of the key protein complexes of the light reactions of photosynthesis, carrying out the conversion of solar energy into electrochemical potential energy required to drive the water splitting reaction which it catalyzes, together with the production of reducing equivalents needed for driving CO_2 fixation. In plants and green algae, the PSII core complex has associated with it membrane-bound light-harvesting antenna complexes (LHCII), to form large macromolecular complexes called PSII-LHCII supercomplexes. The LHCII complexes, functioning as peripheral solar energy collectors, absorb most of the sunlight subsequently directed to the photochemical reaction centre (RC) of PSII.

In plants and green algae the PSII core complex is mainly embedded in the stacked regions of the thylakoid membranes where it is organized as a dimer, each monomer consisting of several proteins including: 1) D1 and D2, making up the photochemical RC; 2) CP47 and CP43, acting as inner antenna proteins; 3) several low molecular mass subunits (LMM subunits, <10 kDa), accounting for more than half of the entire complex and playing a role in stabilizing the binding of cofactors to the PSII core; and 4) the extrinsic polypeptides PsbO, PsbP, PsbQ and PsbR, forming the oxygen evolving complex (OEC) on the lumenal side of the membrane (for a recent review see [1]). Up to now the highest resolution structure available for the plant PSII core complex has been obtained by electron crystallography [2,3], which led to the assignment of the major subunits and location of their transmembrane helices. Moreover, crystal structures have been determined for the isolated extrinsic polypeptides PsbP [4] and PsbQ [5,6].

The most abundant PSII-associated LHCII complex, called “major”, consists of homo- or hetero-trimers of Lhcb1, Lhcb2 and Lhcb3 polypeptides, usually occurring in a ratio of about 8:3:1 [7–9], whose high-resolution structures have been solved by X-ray crystallography [10,11]. According to these studies, all LHCIIs have three membrane-spanning regions connected by both stromal and lumenally-exposed loops and bind a total of 14 chlorophyll (Chl) molecules (8 Chl *a* and 6 Chl *b*) plus 4 carotenoid molecules. In addition, there are three “minor ” LHCII antenna polypeptides, termed Lhcb4 (CP29), Lhcb5 (CP26) and Lhcb6 (CP24), which usually occur in monomeric form. So far, among the minor LHCII antenna proteins, the three-dimensional (3D) structure is available at high resolution only for Lhcb4 [12], revealing three transmembrane α-helices with 13 Chls binding sites (8 assigned as Chl *a* sites, 4 as Chl *b* sites and 1 putative mixed site occupied by both Chl *a* and Chl *b*) and 3 carotenoid binding sites.

A variable number of LHCII can associate with the dimeric PSII core complex to form different types of PSII-LHCII supercomplexes, named according to their composition [13]. The dimeric PSII core complex (C_2) strongly binds two copies of the monomeric Lhcb4 and Lhcb5 and two LHCII trimers (S-trimer) in order to form the C_2S_2 supercomplex [14], which can be regarded as a basic building block of PSII *in vivo*. Larger PSII-LHCII supercomplexes, containing two extra copies of the monomeric Lhcb6 with two additional LHCII trimers (M-trimer) moderately bound to the dimeric PSII core complex via Lhcb4 and Lhcb6, are known as $C_2S_2M_2$ and have been found to represent the basic organization of the PSII in *Arabidopsis thaliana* thylakoid membranes [13,15]. Occasionally even larger supercomplexes have been observed in isolated spinach thylakoids fragments, with one or two additional LHCII trimers (L-trimer) even more loosely bound to the dimeric PSII core complex via Lhcb6, and are known as $C_2S_2M_2L_{1-2}$ [16].

Note that the classification of LHCII trimers within the PSII-LHCII supercomplexes in strongly (S), moderately (M) or loosely (L) bound to the PSII dimeric core complex is based on susceptibility to solubilization by

detergent. Thus the typology and composition of the isolated supercomplexes reflect the mildness of the detergent(s) used and the overall conditions of solubilization. By solubilizing isolated stacked pea thylakoid membranes either with the α or β isomeric forms of the detergent n-dodecyl-D-maltoside (DM), followed by sucrose density ultracentrifugation, we isolated PSII-LHCII supercomplexes with different molecular masses, shown to be respectively of type $C_2S_2M_2$ and C_2S_2, demonstrating the milder detergent action of α-DM with respect to β-DM [17].

In order to gain insights into the primary and tertiary structure of the isolated $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes, we applied extensive multiple approaches of mass spectrometry (MS), combining bottom-up and top-down methods. Bottom-up MS techniques involve approaches where the intact protein is enzymatically cleaved to peptides before measurements via tandem MS; top-down MS targets intact proteins rather than peptides for analysis, with the aim to define the protein primary structure by providing highly accurate structural assignment of MS/MS fragments. In this way, we obtained a detailed overview of the proteins in the isolated PSII-LHCII supercomplexes of different organization, revealing the presence of the antenna proteins Lhcb3 and Lhcb6 and of the extrinsic polypeptides PsbP, PsbQ and PsbR exclusively in the $C_2S_2M_2$ supercomplex. Other proteins of the PSII core complex, common to the $C_2S_2M_2$ and C_2S_2 supercomplexes, including the LMM subunits, were also detected and characterized. Conversely, the LHCII-like PsbS protein was not detected in either the $C_2S_2M_2$ or C_2S_2 supercomplex.

To date, the only 3D structure available of a PSII-LHCII supercomplex has been obtained at 17 Å resolution by cryo-transmission electron microscopy (cryo-TEM) and single particle analysis of C_2S_2 isolated particles containing only one LHCII trimer (S-trimer) per RC core and lacking the minor antenna Lhcb6 [18–20]. For the supercomplex of type $C_2S_2M_2$ only 2D projection maps obtained by TEM analysis of negatively stained single particles derived either from fully or partially solubilized thylakoids are available [15–17,21]. In this paper we show 3D electron density maps, derived from negatively stained samples, for the $C_2S_2M_2$ supercomplex as well as for its C_2S_2 subcomplex from pea (*Pisum sativum*), with resolutions respectively of 30 and 28 Å, subsequently extended by molecular modelling towards atomic level.

2. Material and methods

2.1. PSII-LHCII supercomplexes isolation

Stacked thylakoid membranes were isolated from pea plants according to [22]. By solubilizing thylakoid membranes either with α-DM or β-DM, followed by sucrose density gradient ultracentrifugation, PSII-LHCII supercomplexes of different size were isolated, attributable to the $C_2S_2M_2$ and C_2S_2 organization, respectively, as described in our previous paper [17]. Sucrose bands, containing PSII-LHCII supercomplexes, were carefully removed using a syringe and, if necessary, concentrated by membrane filtration with Amicon Ultra 100 kDa cut-off devices (Millipore) and then stored at -80°C. The Chl concentration was determined spectrophotometrically after extraction in 80% (v/v) acetone according to [23].

2.2. Gel electrophoresis and western blotting

PSII-LHCII supercomplexes were analyzed in native conditions by using the blue-native polyacrylamide gel electrophoresis (BN-PAGE) system according to [24], with a 3-12% acrylamide separating gel and a 4% acrylamide stacking gel. Prior to loading, samples were supplemented with a one-sixteenth volume of the loading buffer (750 mM ε- amino caproic acid, 5% (w/v) Coomassie G250) and incubated for 10 min on ice. After centrifugation at 21,000x*g* for 10 min, the supernatants were loaded onto the 20 cm gradient gel and run for 7 h at a constant voltage of 70 V, using as anode buffer a solution made of 50 mM Bis-Tris-HCl pH 7.0 and as cathode buffer a solution made of 50 mM Tricine, 15 mM Bis-Tris-HCl pH 7.0, and 0.02% (w/v) Coomassie G250. After two-thirds of the run, the cathode buffer containing Coomassie G250 was replaced by a buffer with the same composition but devoid of Coomassie G250 and run overnight at a constant voltage of 60 V. For molecular mass markers, a mixture of lyophilized standard proteins (Amersham, high molecular weight calibration kit (code 17-0445-01), GE Healthcare) was used. For bi-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (2D SDS-PAGE), bands corresponding to $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes resolved on BN-PAGE were cut out and equilibrated in a buffer made of 66 mM Na_2CO_3, 2% (w/v) SDS and 0.66% (v/v) 2-mercaptoethanol at 25°C for 30 min and subjected to 15% acrylamide SDS-PAGE containing 6 M urea using Laemmli's system [25].
Mono-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (1D SDS-PAGE) was performed on a linear gradient gel (18-22% acrylamide) containing 6 M urea using Kashino's system [26], to improve the resolution of LMM subunits.
The proteins separated in 1D or 2D SDS-PAGEs were either stained by 0.25% (w/v) Coomassie R250 for 1 h in a solution made of 50% (v/v) methanol and 10% (v/v) acetic acid, and destained by a solution made of 25% (v/v) methanol and 7.5% (v/v) acetic acid, or transferred onto nitro-cellulose membrane and immunodetected with a specific antiserum against the PsbS polypeptide, by using the alkaline phosphatase conjugate method, with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium as chromogenic substrates (Sigma-Aldrich).

2.3. Mass spectrometry

For liquid nano chromatography electrospray ionization tandem mass spectrometry (nanoLC-ESI-MS/MS) analysis, spots from the 2D SDS-PAGE and bands from the 1D SDS-PAGE were cut out and proteins were digested in-gel with trypsin (Roche), as described in Hellmann et al. [27]. NanoLC-ESI-MS/MS data from each protein sample were obtained by using a Q-star XL (AB SCIEX) as previously described [28]. Mascot.dll v 1.4804.0.22 (Matrix Science/AB SCIEX) was used to generate Mascot (.mgf) files with peak lists from the Analyst QS 2.0 (.wiff) files; the default parameters were used (http://www.matrixscience.com). The principal parameter settings for the Mascot search were as follows: UniProtKB/Swiss-Prot (http://www.expasy.ch) release 2013_02. Common variable modifications such as methionine oxidation and cysteine carbamidomethylation were considered. A tolerance of 60 ppm and 0.3 Da, respectively for precursor ion mass and fragment masses, was allowed for identification.
For matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) and matrix-assisted laser desorption/ionization-time of flight/time of flight (MALDI-TOF/TOF) MS analyses, the isolated PSII-LHCII supercomplexes were initially dialyzed for 18 h against 5% (v/v) acetic acid, using a 12-14 kDa cut-off membrane (Spectra/Por, SpectrumLabs), and further concentrated to 1/10 of the initial volume by membrane

filtration with Amicon Ultra 100 kDa cut-off devices (Millipore). 1 µl of each concentrated sample was mixed with 9 µl of saturated matrix (sinapic acid, Laser Biolabs) solution which consists of 60% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid. After drying droplets of sample onto a target plate, MALDI-TOF and MALDI-TOF/TOF MS analyses were performed using respectively the mass spectrometers Voyager-DE PRO MALDI-TOF (AB SCIEX) and MALDI-TOF/TOF^{TM} 5800 System (AB SCIEX).
The MALDI-TOF mass spectrometer was operated in linear mode at 25 keV accelerating voltage, grid 96.5%, guide wire 0.05% and 800 ns ion extraction delay; the nitrogen laser working at 337 nm and 3 Hz. Two hundred laser shots were accumulated per spectrum over a *m/z* range of 3,500–10,000. Internal calibration was performed on the samples premixed with Calibration mixture 2 of the Peptide Mass Standards Kit for Calibration of AB SCIEX MALDI-TOF Instruments.
MALDI TOF-TOF spectra were acquired using the AB SCIEX TOF/TOF^{TM} 5800 system operated with positive ionization either in linear mode, to determine the average molecular weight, or in reflector mode, to analyze the fragments. An internal calibration was performed on the samples premixed with polyethylene glycol (PEG4000). MS/MS was carried out on the top precursors. Between twenty thousand and two hundred thousand shots were accumulated to get the best S/N, with laser frequency of 1 kHz, acceleration voltage of 2 keV and using air as collision gas. The MS/MS spectra, obtained from the main proteins peaks in MS, were analyzed in Mascot Distiller (ver. 2.3.2.0) by the *de novo* sequencing function coupled with MS-Blast (http://dove.embl.de/Blast2/msblast.html) search at EMBL http://www.embl.de/ [29], using default parameter values.

2.4. Transmission electron microscopy and 3D single particle image analysis

PSII-LHCII supercomplexes isolated with α-DM were negatively stained with 2% uranyl acetate, as described previously [17], and imaged at a calibrated magnification of 50,000x and an acceleration voltage of 200 keV with a Philips CM200 transmission electron microscope equipped with a field emission gun (FEG) at the Electron Microscopy Centre, Imperial College London, UK. A total of 600 CCD images were recorded under low dose conditions (electron dose of approximately 20 electrons/Å^2) on a 4,096 x 4,096 15 µm/pixel slow-scan CCD camera TemCam-F415MP (TVIPS, Germany), leading to a final pixel size corresponding to 1.76 Å at the specimen level. Imaging conditions were optimised, in terms of defocus and astigmatism, to ensure the first minima of the power spectrum to be within a range of 10 to 25 Å, equivalent to an underfocus range of 0.5 to 1.5 µm. Particles were floated out into boxes using EMAN2 [30]. All subsequent image processing was performed within the IMAGIC-V software (Imagic Science, Berlin, Germany) environment [31], at a sampling frequency of 7.04 Å per pixel, until the final reference-free alignment [32] iteration reverted back to 1.76 Å per pixel. A dataset of 15,563 negatively stained single particle images were obtained by picking all discernible single particles present. Several sub-populations of particles, differing in size and shape, were identified. The two largest sub-populations, corresponding to the $C_2S_2M_2$ and the C_2S_2 PSII-LHCII supercomplexes, were in turn analysed as separate datasets, with the reference free alignment giving the initial class averages necessary for multi-reference alignment. Relative orientations were determined for the class averages by the angular reconstitution technique [33] and initial 3D reconstructions gained from implementation of the exact back projection technique [34]. Reprojections were taken from each 3D model and used to identify additional atypical views and further refine the class averages within each sub-

population dataset. Through iterative refinement the data converged to give the best 3D reconstructions shown. Resolution was determined by calculating the Fourier shell correlation (FSC) at the 3σ criterion between two independent 3D reconstructions [35]. Relevant crystallographic co-ordinate atom data (PDB identifiers: 3ARC, 2BHW, 3PL9) were modelled into molecular maps derived from the sub-populations using PyMOL (The PyMOL Molecular Graphics System, Version 1.1r1, Schrödinger, LLC) and UCSF Chimera [36] modelling software. Surface rendered views were calculated at a threshold of 2.5 σ.

3. Results and discussion

3.1. Different protein composition of PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2

PSII-LHCII supercomplexes isolated from pea thylakoid membranes solubilized with α- or β-DM, and shown previously to be of type $C_2S_2M_2$ and C_2S_2 respectively [17], were extensively subjected to in-depth proteomic analyses in order to detect specific proteins (*i.e.*, peripheral antenna proteins, extrinsic polypeptides, and LMM subunits) that may be related to the presence of the additional LHCII M-trimers associated with the C_2S_2 basic supercomplex of PSII.

To investigate the association of the LHCII proteins with the dimeric PSII core complex, in the two differently isolated PSII-LHCII supercomplexes, BN-PAGE followed by 2D SDS-PAGE, coupled with nanoLC-ESI-MS/MS analysis, were performed. In Fig. 1A the BN-PAGE profile of the PSII-LHCII supercomplexes isolated from pea thylakoid membranes solubilized with α-DM (lane 1) and β-DM (lane 2) shows, in the former case, a predominant green band, attributable to the $C_2S_2M_2$ supercomplex, with a higher molecular weight with respect to that of the band observed in the latter, and attributable to the C_2S_2 supercomplex. The difference in mass (~300 kDa) between these two bands is indicative of the retention, by the $C_2S_2M_2$ supercomplex, of two additional LHCII M-trimers. The two green bands corresponding to the $C_2S_2M_2$ and C_2S_2 supercomplexes were cut out from the native gel and subsequently separated by denaturing 2D SDS-PAGE, whose profile, after Coomassie staining, shows the two corresponding maps of spots derived from their membrane polypeptide components (Fig. 1B). After in-gel trypsin digestion of all the spots, nanoLC-ESI-MS/MS analysis of the digested peptides revealed the identity of the RC core subunits CP47, CP43, D2 and D1, and of the six Lhcb antenna proteins (Lhcb1-6) (Table S1), allowing their positioning on the 2D SDS-PAGE maps to be identified (Fig. 1B). From these analyses, it was found that the antenna proteins Lhcb3 and Lhcb6 are present only in the $C_2S_2M_2$ supercomplex, in line with findings suggesting Lhcb3 as an exclusive subunit of the LHCII M-trimer present in the $C_2S_2M_{1\text{-}2}$ supercomplexes [16] and Lhcb6 functioning as a linker for this trimer to the C_2S_2 supercomplex [15,17].

In lane 1 of Fig. 1A, however, two additional faint bands were detected below the predominant one: the one with higher molecular weight corresponds to C_2S_2 supercomplexes and the second, with lower molecular weight (~150 kDa), to free trimers of LHCII. Despite care taken during the PSII-LHCII isolation process, the presence of these complexes in the preparation obtained with α-DM indicates a degree of instability of the isolated $C_2S_2M_2$ supercomplex and the easy detachment of LHCII M-trimers from the $C_2S_2M_2$ supercomplex.

In order to determine the exact protein composition of the $C_2S_2M_2$ and C_2S_2 supercomplexes, also in terms of extrinsic polypeptides and LMM subunits, we performed a separation of their proteins in 1D SDS-PAGE according to the system of Kashino and co-authors [26], shown in Fig. 2A. This electrophoretic system, in

combination with nanoLC-ESI-MS/MS analysis (as for proteins separated through 2D SDS-PAGE), allowed the detection and identification of the extrinsic polypeptides PsbO, PsbP, PsbQ and PsbR (Table S1, Fig. 2A). The level of PsbO appeared to be less sensitive, compared to the other OEC subunits, to perturbation during the solubilization with both forms of DM. The relative stability of PsbO in both isolated PSII-LHCII supercomplexes is due to the interaction of its N-terminal region with several PSII RC subunits, including CP47, CP43, D1 and D2, as shown in the cyanobacterial crystal structure [37–39]. Moreover, numerous cross-linking studies have indicated that the PsbO protein may be cross-linked to CP47 in higher plants [40–42]. The milder action of α-DM as the solubilizing agent also facilitated the retention of the extrinsic subunits PsbP, PsbQ and PsbR in the $C_2S_2M_2$ supercomplex. These polypeptides were not present in the C_2S_2 supercomplex, indicating that they are not absolutely required to stabilize the binding of LHCII S-trimers to the PSII RC core. On the other hand, their presence in the $C_2S_2M_2$ supercomplex suggests their possible involvement, as individual subunit or in cooperation, in the binding of the LHCII M-trimers and/or in the overall organization and stabilization of this very large macromolecular PSII complex. This is in agreement with recent findings by Allahverdiyeva et al. [43] attesting to the important role played by the PsbQ and PsbR subunits in PSII-LHCII supercomplex macro-organization. Of note is the fact that the isolated $C_2S_2M_2$ supercomplex also retains the PsbP extrinsic subunit, which was not found in a similar preparation obtained from the thylakoid membranes of *Arabidopsis thaliana* solubilized with α-DM by Caffarri et al. [15].

The Lhcb-like PsbS protein seems to play a role in the distribution of light to the PSII RC by regulating non-photochemical quenching [44]. It has been argued that it does so by controlling PSII-LHCII supercomplex macro-organization [45]. Our MS analyses did not reveal the presence of this protein in either isolated $C_2S_2M_2$ or C_2S_2 supercomplexes, in agreement with previous reports [15,46]. As this protein is believed to be present in non-stoichiometric amounts compared with other PSII proteins, we undertook a sensitive immunological analysis using an antibody with strong reactivity to PsbS in pea thylakoid membranes. This analysis did not detect the PsbS protein in any of the isolated supercomplexes (Fig. 2B), and showed that the protein co-migrated with the free LHCII band in the sucrose density gradient step adopted in the isolation of supercomplexes (data not shown). This finding contrasts with that of Caffarri et al. [15], who concluded that PsbS co-migrated with the $C_2S_2M_2$ supercomplex in their sucrose density gradient.

3.2. Common content in LMM subunits in PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2

The high resolution of Kashino's electrophoretic system [26], especially appropriate for the low molecular masses, allowed for the separation of the LMM subunits present in the $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes. After in-gel trypsin digestion of all bands with masses <10 kDa, nanoLC-ESI-MS/MS analysis of the digested peptides revealed the presence of PsbE and PsbH (Table S1, Fig. 2A) in both supercomplexes, but failed to detect PSII components with lower molecular weights. This was likely due to the higher hydrophobicity of these transmembrane proteins, almost completely embedded in the membrane, due to their short length, which lowers the accessibility of trypsin enzyme and the number of tryptic cleavage sites present in their sequences. On the contrary, by applying MALDI-TOF and MALDI-TOF/TOF MS directly to the isolated $C_2S_2M_2$ and C_2S_2 supercomplexes, it was possible to identify most of the expected LMM subunits present in the two samples: PsbX, PsbTc, PsbJ, PsbI, PsbK, PsbL, PsbF, PsbW, and PsbE (Fig. 3, Table 1).

It is worth noting that MALDI-TOF measurements performed on three independent preparations of $C_2S_2M_2$ and C_2S_2 supercomplexes were highly reproducible, and that in both types of samples corresponding peaks were observed at coincident *m/z* values, given a mass tolerance of 50 ppm (the MALDI TOF technique is accurate to a 100 to 50 ppm average error, achieved by internal calibration). The overlapping of MALDI-TOF mass spectra for the $C_2S_2M_2$ and C_2S_2 samples in the range of *m/z* below 10,000 confirmed the common composition in LMM subunits of the two isolated supercomplexes (Fig. 3). In this range of *m/z*, several peaks were detected, most of which have been identified as follows: 3,986.2 ± 0.2 as PsbX, 4,064.1 ± 0.2 as PsbTc, 4,161.9 ± 0.2 as PsbJ, 4,213.4 ± 0.2 as PsbI, 4,289.1 ± 0.2 as PsbK, 4,355.3 ± 0.2 as PsbL, 4,399.7 ± 0.2 as PsbF, 5,932.1 ± 0.3 as PsbW, 9,265.3 ± 0.5 as PsbE (Fig. 3, Table 1).

The identification of the proteins PsbX, PsbTc, PsbJ, PsbK and PsbF was assigned by MS/MS analysis on the corresponding selected precursor main peaks combined with *de novo* sequencing and homology searching (see Table S2, and Figs. S1-S5). Among these five assignments:

1) the observed *m/z* values for PsbK and PsbF are in good agreement with *m/z* values measured on these isolated proteins from pea, spinach and barley [47–49], and also with expected molecular masses calculated from the corresponding genomic sequences from pea, taking into account, in the case of PsbF, the Met1 removal and Thr2 acetylation suggested as post translational modifications (PTMs) by Sharma et al. [47];

2) the observed *m/z* value obtained for PsbTc is respectively higher than that measured for this protein isolated from spinach by Zheleva et al. [48] and lower than that experimentally found for its homolog from barley by Plöscher et al. [49]. This can be explained by the different length of the protein sequence among the three plants: 35 amino acid residues in pea, 33 in spinach and 38 in barley (accession numbers in the UniProtKB database: Q8HS25, P61840 and P69669, respectively). Moreover, the mass difference between the observed value and that expected from the calculated corresponding genomic sequence from pea, could indicate a formyl-methylation as PTM, as found for this protein in spinach by Zheleva et al. [48];

3) in the literature there are measured values of *m/z* for PsbX and PsbJ proteins purified only from barley for comparison [49]. The observed value of *m/z* for PsbJ is higher than that measured for its homolog in barley. Despite the same length of the protein sequence in pea and barley, the different *m/z* value observed in the two plants can be explained by the presence of five amino acid substitutions (Asn3 *vs* Asp3, Ile12 *vs* Leu12, Val20 *vs* Pro20, Ile25 *vs* Val25, Leu27 *vs* Val27) in the sequence from pea with respect to barley (accession numbers in the UniProtKB database: P13555 and P20175, respectively). The *m/z* value observed for PsbJ closely matches the expected mass calculated from the corresponding genomic sequence, with a slight difference possibly linked to a N-acetylation suggested as PTM by Plöscher et al. [49] in barley. In the case of PsbX, the experimental *m/z*, that is slightly lower than that measured for its homolog in barley, accounts for less than a half of the expected mass calculated from its genomic sequence and the sequence obtained by *de novo* sequencing for this protein covers the C-term portion of the amino acid sequence present in the UniProtKB database (Table 1 and Table S2).

The putative assignment of PsbI, PsbL, PsbW and PsbE was done by comparing the experimental *m/z* measured by MALDI-TOF either with results obtained by ESI MS/MS and/or MALDI-TOF MS and N-terminal amino acid sequencing present in the literature for these LMM subunits purified from higher plants [47–49], or with protein masses calculated from the corresponding nucleotide sequences from pea (Table 1). For PsbI, PsbW and PsbE both correlations were good; in case of PsbL, the measured *m/z* closely matched that obtained for this subunit in spinach and barley [48,49], and the difference observed between the measured

and the expected molecular mass calculated from its genomic sequence from pea, can be likely due to the Met1 removal, a PTM found in its corresponding analog in spinach and barley [48,49].
Despite the reproducible signal strength of peaks at *m/z* 4,028.6 ± 0.2 and 7,744.9 ± 0.4, it was not possible to assign a specific identity to these masses by MS/MS. However, the measured *m/z* value at 7,744.9 can be tentatively assigned to PsbH, due to the identification of this subunit by nanoLC-ESI-MS/MS analysis in both PSII-LHCII supercomplexes (Table S1, Fig. 2A), and the good correlation between the measured weight and the expected molecular mass calculated from its genomic sequence in *P. sativum* (accession number in the UniProtKB database Q9XQR3, calculated average mass after Met1 removal 7,726.95).

3.3. 3D reconstructions of the PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2 revealed by TEM single particle analysis and angular reconstitution

Negative stain TEM from the sucrose density gradient fraction containing PSII-LHCII supercomplexes obtained by solubilizing pea thylakoid membranes with α-DM provided for a single particle dataset of 15,563 images, as described previously [17]. In the current study, this dataset was re-subjected to more intensive computer-based purification analyses, so that more rigorously defined sub-populations of particles might be identified prior to the application of the 3D reconstruction technique of angular reconstitution [34]. In so doing, sub-populations of 4,760 and 1,868 particles were identified relating to the largest complexes, in terms of surface area with two-fold symmetry. These were attributed to the $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes respectively. Following *de novo* reference-free alignments, the relative angular orientations of the particles observed within each sub-population were strongly biased towards top and side views; however, subsequent iterative refinements were able to identify a small amount of slightly tilted views which aided in the calculation of the final 3D electron density maps.
In Fig. 4A the 3D electron density map of the $C_2S_2M_2$ supercomplex is represented in green mesh as viewed from the top lumenal side and the C_2S_2 3D electron density map has been incorporated within, surface-rendered in light blue. The $C_2S_2M_2$ map has maximum dimensions of 375 Å length by 210 Å width by 105 Å height, with a two-fold axis and an approximate resolution of 30 Å. The C_2S_2 map is also shown with two-fold imposed symmetry, having dimensions of 340 Å length by 200 Å width by 105 Å height with a resolution approaching 28 Å. To interpret these 3D electron density maps, we compared their internal density distribution with surface-rendered X-ray structures (see Fig. 4B) of the PSII dimeric core of cyanobacteria at 1.9 Å [39], the LHCII trimeric complex of pea at 2.5 Å [11] and the monomeric Lhcb4 of spinach at 2.8 Å [12]. The latter was extrapolated to represent densities we attributed previously to the Lhcb5 and Lhcb6 subunits [17], whose X-ray structures have yet to be solved. Modelling the co-ordinates within the two molecular envelopes was done by visual inspection using the internal densities (not shown) as a guide, starting centrally with the C_2S_2 model of Nield and Barber [20], treated as a rigid whole, and extending out to include two Lhcb6 subunits and two LHCII M-trimers. The entire modelling environment encompassed a depth of 130 Å and this thickness is shown in full in Fig. 4A-B. The cyanobacterial PSII dimeric core X-ray co-ordinates (excluding those of PsbV and PsbU), when shown surface-rendered (Fig. 4B), are able to emphasise the key differences between the cyanobacterial lumenal surface and our molecular envelopes from pea. However, the prevalence of top and side views, coupled with the availability of only a few tilted views, due to the negative stain methodology employed, was found in this particular work to suppress the

density expected for the lumenally-exposed polypeptides of the OEC and therefore limit our maps' interpretability in this regard. Reducing the slabbed area to 65 Å in Fig. 4C-D enabled the visualisation of the entire membrane domain, the interactions between the LHCII antennae proteins with specific subunits of the PSII core complex and the relative overall positioning of the additional LHCII M-trimers and the Lhcb6 subunits present in the $C_2S_2M_2$ supercomplex (the latter two in 1:1 stoichiometry with the PSII monomeric core which they bind). In Fig. 4D the C_2S_2 map was removed from the modelling environment and, noting an approximate 10 Å wide boundary for the detergent shell (yellow line) encompassing the TEM-derived $C_2S_2M_2$ green mesh, the major domains of the largest PSII-LHCII supercomplex we observed in α-DM solubilized pea thylakoids can be interpreted more readily.

4. Conclusions

By means of bottom-up and top-down MS we have conducted an in-depth characterization of the polypeptide composition of PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2 isolated from pea thylakoid membranes by one-step treatment with α- and β-DM detergents, respectively, as described previously [17]. Their protein composition, spanning from RC and LMM intrinsic subunits to antennae proteins and extrinsic polypeptides, was revealed. In addition to a common composition in the main PSII RC proteins, the $C_2S_2M_2$ and C_2S_2 supercomplexes showed an identical set of LMM subunits. Thus, we conclude that none of the LMM subunits are specifically required for the binding of the additional LHCII M-trimers to the basic C_2S_2 unit. In contrast to the LMM subunits, the two isolated supercomplexes revealed basic differences in their Lhcb antennae polypeptides: Lhcb1, Lhcb2, Lhcb4 and Lhcb5 were found in both the $C_2S_2M_2$ and C_2S_2 supercomplexes, whereas Lhcb3 and Lhcb6 were present only in the largest supercomplex, suggesting that Lhcb3 acts as an exclusive subunit of the LHCII M-trimer and Lhcb6 is functioning as a linker for this LHCII trimer to the C_2S_2. The Lhcb-like PsbS protein was not found to be associated with the isolated supercomplexes, indicating that this subunit does not influence the interaction between the PSII core and the outer Lhcb antenna system. Due to the high hydrophobicity of this protein, its absence from the isolated supercomplexes and abundance in the free LHCII trimers (LHCII band in the sucrose density gradient adopted in the isolation of supercomplexes), suggests that its location is in the peripheral boundary of the PSII-LHCII supercomplexes or in the LHCII-enriched domains of the thylakoid membranes. The proteomic data indicate that, in addition to the PsbO subunit, which is stably bound to the PSII RC core in both types of supercomplexes, the $C_2S_2M_2$ supercomplex retains the PsbP, PsbQ and PsbR subunits. These polypeptides were not present in the C_2S_2 supercomplex, indicating a dispensable role in the stable binding of LHCII S-trimers to the PSII dimeric core, while suggesting their possible involvement in the overall macromolecular organization and/or stabilization of the larger $C_2S_2M_2$ PSII-LHCII supercomplex.

A pseudo-atomic 3D structural model of the spinach C_2S_2 supercomplex, based upon a cryo-TEM molecular envelope calculated at 17 Å resolution, was previously reported by Nield and Barber [20]. However, for the larger $C_2S_2M_2$ supercomplex, only 2D projection maps obtained by TEM of negatively stained single particles have been published to date [15–17,21].

Here we report the first 3D structural model of an isolated $C_2S_2M_2$ supercomplex obtained by single particle analysis of negatively stained samples as imaged by TEM. The resolution of the model was estimated to be approximately 30 Å according to the 3σ FSC criterion, being restricted partly by the relatively low size of the

dataset used for the analysis and the lack of sufficient random orientations of the complex on the TEM grid. Both of these limitations reflect the use of samples prepared for TEM by negative staining and can, in principle, be overcome by using unstained vitrified samples and associated cryo-TEM techniques, as shown for the C_2S_2 supercomplex [18,19]. Nonetheless, this is the first time that such a large PSII-LHCII supercomplex has been shown in 3D from solubilized thylakoids in pea (*P. sativum*). The 3D reconstructions presented here once more confirm the central positioning of the C_2S_2 supercomplex within the larger volume of the $C_2S_2M_2$ molecular envelope and the relative positioning of various major subunits. The density for the OEC proteins was not fully resolved, hence making it difficult at this stage to suggest the location of the PsbP, PsbQ and PsbR subunits, albeit these polypeptides are present within the $C_2S_2M_2$ particles analyzed. An improved dataset, ideally derived from cryo-TEM, will be required to resolve and assign the OEC proteins.

Appendix A. Supplementary data

Short legends for supporting information

Table S1

List of PSII RC core subunits, extrinsic polypeptides, LMM subunits and LHCII proteins identified by nanoLC-ESI-MS/MS present in the isolated $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes as shown in Figs. 1-2. The table reports: sequences of peptides obtained by nanoLC-ESI-MS/MS (third column) with their corresponding precursor ion mass (second column); for each identified protein (first column), the calculated molecular mass (Mr, fourth column), the accession number and the database in which the protein was found (fifth column), and the percentage of residue identities with *Pisum sativum*, when available, or the homolog *Arabidopsis thaliana* (sixth column). Underlined amino acid residues (third column) indicate modifications such as methionine or asparagine oxidation (M, N) and cysteine carbamidomethylation (C).

Table S2

Assignments of LMM subunits in the isolated $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes based on MS/MS spectra (shown in Figs. S1-S5), *de novo* sequencing and homology searching.

Fig. S1. *De novo* sequencing results obtained through Mascot Distiller on the MS/MS spectrum corresponding to the experimental precursor at *m/z* 3,982.2.

Fig. S2. *De novo* sequencing results obtained through Mascot Distiller on the MS/MS spectrum corresponding to the experimental precursor at *m/z* 4,059.2.

Fig. S3. *De novo* sequencing results obtained through Mascot Distiller on the MS/MS spectrum corresponding to the experimental precursor at *m/z* 4,156.3.

Fig. S4. *De novo* sequencing results obtained through Mascot Distiller on the MS/MS spectrum corresponding to the experimental precursor at *m/z* 4,283.3.

Fig. S5. *De novo* sequencing results obtained through Mascot Distiller on the MS/MS spectrum corresponding to the experimental precursor at *m/z* 4,393.4.

Acknowledgments

This work was supported by funds from the European Commission (ArtipHyction project (303435) FCH-JU-2011-1; Eco^{2}CO$_2$ project (309701) FP7-NMP2012). We thank Dr Waidelich Dietmar (AB SCIEX- Darmstadt, Germany) for kindly performing MALDI-TOF/TOF analyses.

References

[1] C. Pagliano, G. Saracco, J. Barber, Structural, functional and auxiliary proteins of photosystem II, Photosynth. Res. (2013) doi:10.1007/s11120-013-9803-8.

[2] K.-H. Rhee, E.P. Morris, J. Barber, W. Kühlbrandt, Three-dimensional structure of the plant photosystem II reaction centre at 8 Å resolution, Nature 396 (1998) 283–286.

[3] B. Hankamer, E.P. Morris, J. Nield, C. Gerle, J. Barber, Three-dimensional structure of the photosystem II core dimer of higher plants determined by electron microscopy, J. Struct. Biol. 135 (2001) 262–269.

[4] K. Ifuku, T. Nakatsu, H. Kato, F. Sato, Crystal structure of the PsbP protein of photosystem II from *Nicotiana tabacum*, EMBO Rep. 5 (2004) 362–367.

[5] V. Calderone, M. Trabucco, A. Vujicić, R. Battistutta, G.M. Giacometti, F. Andreucci, R. Barbato, G. Zanotti, Crystal structure of the PsbQ protein of photosystem II from higher plants, EMBO Rep. 4 (2003) 900–905.

[6] M. Balsera, J.B. Arellano, J.L. Revuelta, J. De Las Rivas, J.A. Hermoso, The 1.49 Å resolution crystal structure of PsbQ from photosystem II of *Spinacia oleracea* reveals a PPII structure in the N-terminal region, J. Mol. Biol. 350 (2005) 1051–1060.

[7] S. Jansson, The light-harvesting chlorophyll a/b binding proteins, Biochim. Biophys. Acta 1184 (1994) 1–19.

[8] G. Jackowski, K. Kacprzak, S. Jansson, Identification of Lhcb1/Lhcb2/Lhcb3 heterotrimers of the main light-harvesting chlorophyll a/b-protein complex of photosystem II (LHCII), Biochim. Biophys. Acta 1504 (2001) 340–345.

[9] S. Caffarri, R. Croce, L. Cattivelli, R. Bassi, A Look within LHCII: differential analysis of the Lhcb1−3 complexes building the major trimeric antenna complex of higher-plant photosynthesis, Biochemistry 43 (2004) 9467–9476.

[10] Z.F. Liu, H.C. Yan, K.B. Wang, T.Y. Kuang, J.P. Zhang, L.L. Gui, X.M. An, W.R. Chang, Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution, Nature 428 (2004) 287–292.

[11] J. Standfuss, A.C.T. van Scheltinga, M. Lamborghini, W. Kühlbrandt, Mechanisms of photoprotection and nonphotochemical quenching in pea light harvesting complex at 2.5 Å resolution, EMBO J. 24 (2005) 918–928.

[12] X. Pan, M. Li, T. Wan, L. Wang, C. Jia, Z. Hou, X. Zhao, J. Zhang, W. Chang, Structural insights into energy regulation of light-harvesting complex CP29 from spinach, Nat. Struct. Mol. Biol. 18 (2011) 309–316.

[13] J.P. Dekker, E.J. Boekema, Supramolecular organization of thylakoid membrane proteins in green plants, Biochim. Biophys. Acta 1706 (2005) 12–39.

[14] E.J. Boekema, B. Hankamer, D. Bald, J. Kruip, J. Nield, A.F. Boonstra, J. Barber, M. Rogner, Supramolecular structure of the photosystem II complex from green plants and cyanobacteria, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 175–179.

[15] S. Caffarri, R. Kouřil, S. Kereïche, E.J. Boekema, R. Croce, Functional architecture of higher plant photosystem II supercomplexes, EMBO J. 28 (2009) 3052–3063.

[16] E.J. Boekema, H. van Roon, F. Calkoen, R. Bassi, J.P. Dekker, Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes, Biochemistry 38 (1999) 2233–2239.

[17] S. Barera, C. Pagliano, T. Pape, G. Saracco, J. Barber, Characterization of PSII-LHCII supercomplexes isolated from pea thylakoid membrane by one-step treatment with α- and β-dodecyl-D-maltoside, Phil. Trans. R Soc. B 67 (2012) 3389–3399.

[18] J. Nield, E.V. Orlova, E.P. Morris, B. Gowen, M. van Heel, J. Barber, 3D map of the plant photosystem two supercomplex obtained by cryoelectron microscopy and single particle analysis, Nat. Struct. Biol. 7 (2000) 44–47.

[19] J. Nield, M. Balsera, J. De Las Rivas, J. Barber, Three-dimensional electron cryo-microscopy study of the extrinsic domains of the oxygen-evolving complex of spinach: assignment of the PsbO protein, J. Biol. Chem. 277 (2002) 15006–15012.

[20] J. Nield, J. Barber, Refinement of the structural model for the photosystem II supercomplex of higher plants, Biochim. Biophys. Acta 1757 (2006) 353–361.

[21] A.E. Yakushevska, P.E. Jensen, W. Keegstra, H. van Roon, H.V. Scheller, E.J. Boekema, J.P. Dekker, Supermolecular organization of photosystem II and its associated light-harvesting antenna in *Arabidopsis thaliana*, Eur. J. Biochem. 268 (2001) 6020–6028.

[22] C. Pagliano, S. Barera, F. Chimirri, G. Saracco, J. Barber, Comparison of the α and β isomeric forms of the detergent n-dodecyl-D-maltoside for solubilizing photosynthetic complexes from pea thylakoid membranes, Biochim. Biophys. Acta 1817 (2012) 1506–1515.

[23] D.J. Arnon, Copper enzymes in isolated chloroplasts, polyphenoloxidase in *Beta vulgaris*, Plant Physiol. 24 (1949) 1–14.

[24] H. Schagger, G. von Jagow, Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form, Anal. Biochem. 199 (1991) 223–231.

[25] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.

[26] Y. Kashino, H. Koike, K. Satoh, An improved sodium dodecyl sulfate-polyacrylamide gel electrophoresis system for the analysis of membrane protein complexes, Electrophoresis 22 (2001) 1004–1007.

[27] U. Hellmann, C. Wernstedt, J. Gonez, C.H. Heldin, Improvement of an “In-Gel” digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing, Anal. Biochem. 224 (1995) 451–455.

[28] C. Pagliano, F. Chimirri, G. Saracco, F. Marsano, J. Barber, One-step isolation and biochemical characterization of a highly active plant PSII monomeric core, Photosynth. Res. 108 (2011) 33–46.

[29] A. Shevchenko, S. Sunyaev, A. Loboba, A. Shevchenko, P. Bork, W. Ens, K.G. Standing, Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of flight mass spectrometry and BLAST homology searching, Anal. Chem. 73 (2001) 1917–1926.

[30] G. Tang, L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang, I. Rees, S.J. Ludtke, EMAN2: an extensible image processing suite for electron microscopy, J. Struct. Biol. 157 (2007) 38–46.

[31] M. van Heel, G. Harauz, E. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system, J. Struct. Biol. 116 (1996) 17–24.

[32] P. Dube, P. Tavares, R. Lurz, M. van Heel, The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry, EMBO J. 12 (1993) 1303–1309.

[33] M. Van Heel, Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction, Ultramicroscopy 21 (1987) 111–123.

[34] M. Radermacher, Three-dimensional reconstruction of single particles from random and nonrandom tilt series, J. Electron. Microsc. Tech. 9 (1988) 359–394.

[35] M. van Heel, M. Schatz, Fourier shell correlation threshold criteria, J. Struct. Biol. 151 (2005) 250–262.

[36] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[37] K.N. Ferreira, T.M. Iverson, K. Maghlaoui, J. Barber, S. Iwata, Architecture of the photosynthetic oxygen-evolving center, Science 303 (2004) 1831–1838.

[38] J. De Las Rivas, J. Barber, Analysis of the structure of the PsbO protein and its implications, Photosynth. Res. 81 (2004) 329–343.

[39] Y. Umena, K. Kawakami, J.R. Shen, N. Kamiya, Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å, Nature 473 (2011) 55–60.

[40] T.M. Bricker, W.R. Odom, C.B. Queirolo, Close association of the 33-kDa extrinsic protein with the apoprotein of Cpa1 in photosystem II, FEBS Lett. 231 (1988) 111–117.

[41] I. Enami, T. Miyaoka, Y. Mochizuki, J.R. Shen, K. Satoh, S. Katoh, Nearest neighbor relationships among constituent proteins of oxygen-evolving photosystem II membranes – binding and function of the extrinsic 33 kDa protein, Biochim. Biophys. Acta 973 (1989) 35–40.

[42] A. Seidler, A.W. Rutherford, H. Michel, On the role of the N-terminus of the extrinsic 33 kDa protein of photosystem II, Plant Mol. Biol. 31 (1996) 183–188.

[43] Y. Allahverdiyeva, M. Suorsa, F. Rossi, A. Pavesi, M.M. Kater, A. Antonacci, L. Tadini, M. Pribil, A. Schneider, G. Wanner, D. Leister, E.M. Aro, R. Barbato, P. Pesaresi, Arabidopsis plants lacking PsbQ and PsbR subunits of the oxygen-evolving complex show altered PSII super-complex organization and short-term adaptive mechanisms, Plant J. 75 (2013) 671–684.

[44] X.P. Li, O. Björkman, C. Shih, A.R. Grossman, M. Rosenquist, S. Jansson, K.K. Niyogi, A pigment-binding protein essential for regulation of photosynthetic light harvesting, Nature 403 (2000) 391–395.

[45] S. Kereïche, A.Z. Kiss, R. Kouřil, E.J. Boekema, P. Horton, The PsbS protein controls the macro-organisation of photosystem II complexes in the grana membranes of higher plant chloroplasts, FEBS Lett. 584 (2010) 759–764.

[46] J. Nield, C. Funk, J. Barber, Supermolecular structure of photosystem II and location of the PsbS protein, Phil. Trans. R Soc. B 355 (2000)1337–1344.

[47] J. Sharma, M. Panico, J. Barber, H.R. Morris, Characterization of the low molecular weight photosystem II reaction center subunits and their light-induced modifications by mass spectrometry, J. Biol. Chem. 272 (1997) 3935–3943.

[48] D. Zheleva, J. Sharma, M. Panico, H.R. Morris, J. Barber, Isolation and characterization of monomeric and dimeric CP47-reaction center photosystem II complexes, J. Biol. Chem. 273 (1998) 16122–16127.

[49] M. Plöscher, B. Granvogl, M. Zoryan, V. Reisinger, L.A. Eichacker, Mass spectrometric characterization of membrane integral low molecular weight proteins from photosystem II in barley etioplasts, Proteomics 9 (2009) 625–635.

Tables and figures captions

Table 1

Proposed identification of measured *m/z* peaks in PSII-LHCII supercomplexes of type $C_2S_2M_2$ and C_2S_2 isolated from pea thylakoid membranes. The table reports the reproducible *m/z* values measured by MALDI-TOF (first column) on three independent preparations of $C_2S_2M_2$ and C_2S_2 supercomplexes (similar values for both types of samples were obtained. See text for details and Fig. 3), assigned by *de novo* sequencing and homology searching (second column) on MS/MS data (see Table S2 and Figs. S1-S5), or putatively assigned (third column) according to matches with referenced values (fourth column). For each identified subunit, expected mass values of unprocessed precursors and processed proteins (sequences from UniProtKB-Swiss Prot), including annotated PTMs, are reported (fifth column), along with referenced masses, when available (fourth column).

Fig. 1. BN/2D SDS-PAGE profiles of PSII-LHCII supercomplexes isolated from pea thylakoid membranes. A. BN-PAGE of PSII-LHCII supercomplexes isolated from pea thylakoid membranes solubilized with α-DM (lane 1) and β-DM (lane 2) (6 µg Chl per lane). Protein marker (Native high molecular weight, GE Healthcare) positions indicated on the right. B. 2D SDS-PAGE separation of $C_2S_2M_2$ and C_2S_2 supercomplexes, after Coomassie staining. Protein marker (Precision plus, Bio-Rad) positions indicated on the right.

Fig. 2. 1D SDS-PAGE and western blot analysis of PSII-LHCII supercomplexes isolated from pea thylakoid membranes. A. Profiles of protein composition of $C_2S_2M_2$ and C_2S_2 supercomplexes (10 µg Chl per lane) resolved by 1D SDS-PAGE according to [26]. Protein standards (Precision plus, Bio-Rad) are loaded on lane M. B. Western blot analysis using the antibody against PsbS. Pea thylakoid membranes (Thyl) loaded as control (10 µg Chl on each lane).

Fig. 3. MALDI-TOF mass spectra of PSII-LHCII supercomplexes of type $C_2S_2M_2$ (green line) and C_2S_2 (red line) isolated from pea thylakoid membranes. A. Peaks with *m/z* values between 3,500–5,000. B. Peaks with *m/z* values between 5,000–10,000. Denoted above the peaks, names of proteins in bold refer to LMM subunits identified by MS/MS analysis and *de novo* sequencing (see Table S2), those in italics refer to putative assignments based on good correlation between observed *m/z* and referenced values for masses measured on isolated LMM subunits from pea, spinach and barley [47-49].

Fig. 4. Top lumenal views of 3D reconstructions of the $C_2S_2M_2$ and C_2S_2 PSII-LHCII supercomplexes isolated from pea thylakoid membranes, derived from TEM and single particle analysis, with modelled high-resolution X-ray structures of the PSII dimeric core from cyanobacteria [39] (PDB: 3ARC; subunits D1, D2, CP47, CP43 and PsbO are in yellow, orange, red, sandy brown and purple, other subunits in grey, respectively; PsbU and PsbV have been omitted from the PDB file), the LHCII trimer [11] (PDB: 2BHW; in blue) and Lhcb4 [12] (PDB: 3PL9; in pale green), the latter two from higher plants. A. Top lumenal view of the $C_2S_2M_2$ 3D electron density map (green mesh), with the C_2S_2 3D electron density map, inset, surface-rendered in light blue. Maximum dimensions (in plane) of the 3D maps, inclusive of the detergent shell, are 375 Å (length) x 210 Å (width) for $C_2S_2M_2$ and 340 Å (length) x 200 Å (width) for C_2S_2 supercomplexes. Scale bar for all panels represents 5 nm. B. As per panel A, with the cyanobacterial PSII dimeric core present, highlighting lumenal surface differences, together with LHCII trimer and monomeric Lhcb atomic co-ordinates shown as surface-rendered spheres (coloured as described above). C. The modelling environment, cut away by 65 Å, to reveal its lower half, also 65 Å thick, thus emphasising the position of the X-ray co-ordinates (surface-rendered and coloured as described above). D. The $C_2S_2M_2$ 3D cat away map (as in C) on its own with modelled subunits labelled (surface-rendered and coloured as described above), its membrane domain also shown as a 65 Å thick slab from the lumenal top view. The Lhcb5 and Lhcb6 atomic co-ordinates, whose structures have not yet been solved, are assumed to be similar to that of Lhcb4. A delineating boundary (yellow line) represents the α-DM detergent shell, approximately 10 Å within the outer edge of the $C_2S_2M_2$ three-dimensional mesh.

Table 1

Observed *m/z*	*De novo* sequencing assignment	Putative assignment	Referenced *m/z*, organism	Calculated average molecular mass of unprocessed precursor/accession UniProtKB (*P. sativum* or *S. oleracea*)/annotated PTMs (calculated average molecular mass of processed precursor)
3986.2 ± 0.2	**PsbX**		4141.32 [49], *H. vulgare*	8683.1/Q8VYY1 (*P. sativum*)
4064.1 ± 0.2	**PsbTc**		3849.6 [48], *S. oleracea*; 4422.05 [49], *H. vulgare*	4032.9/Q8HS25 (*P. sativum*)
4161.9 ± 0.2	**PsbJ**		4011.15 [49], *H. vulgare*	4115.9/P13555 (*P. sativum*)
4213.4 ± 0.2		**PsbI**	4195.5 [48], *S. oleracea*; 4209.5 [47], *P. sativum;* 4193.28 [49], *H. vulgare*	4182.0/D5MAJ9 (*P. sativum*)/N-formyl Met (4210.0)
4289.1 ± 0.2	**PsbK**		4292.1 [48], *S. oleracea;* 4282.40 [49], *H. vulgare*	6910.4/D5MAJ8 (*P. sativum*)/N-term 1-24 removed (4285.2)
4355.3 ± 0.2		**PsbL**	4365.5 [48], *S. oleracea;* 4363.17 [49], *H. vulgare*	4497.2/P60147 (*P. sativum*)
4399.7 ± 0.2	**PsbF**		4394.6 [47], *P. sativum*; 4409.1 [48], *S. oleracea;* 4406.40 [49], *H. vulgare*	4424.3/P62096 (*P. sativum*)
5932.1 ± 0.3		**PsbW**	5927.4 [48], *S. oleracea;* 5889.00 [49], *H. vulgare*	14177.2/Q41387 (*S. oleracea*)/N-term 1-83 removed (5927.7)
9265.3 ± 0.5		**PsbE**	9283.6 [47], *P. sativum*; 9255.1 [48], *S. oleracea;* 9307.83 [49], *H. vulgare*	9414.6/P13554 (*P. sativum*)/Met1 removed (9283.4)

Figure 1
Click here to download high resolution image

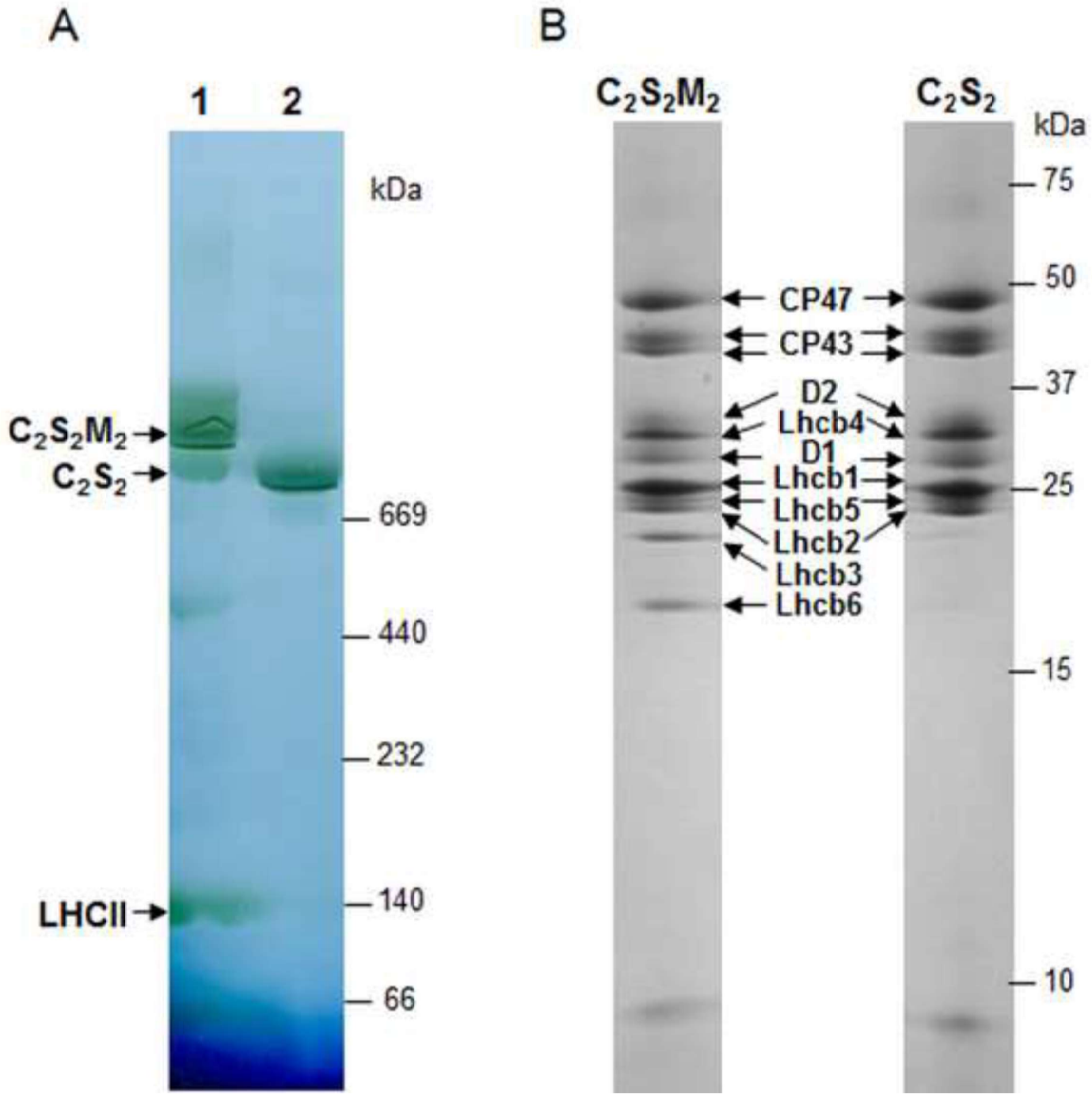


Figure 2

Click here to download high resolution image

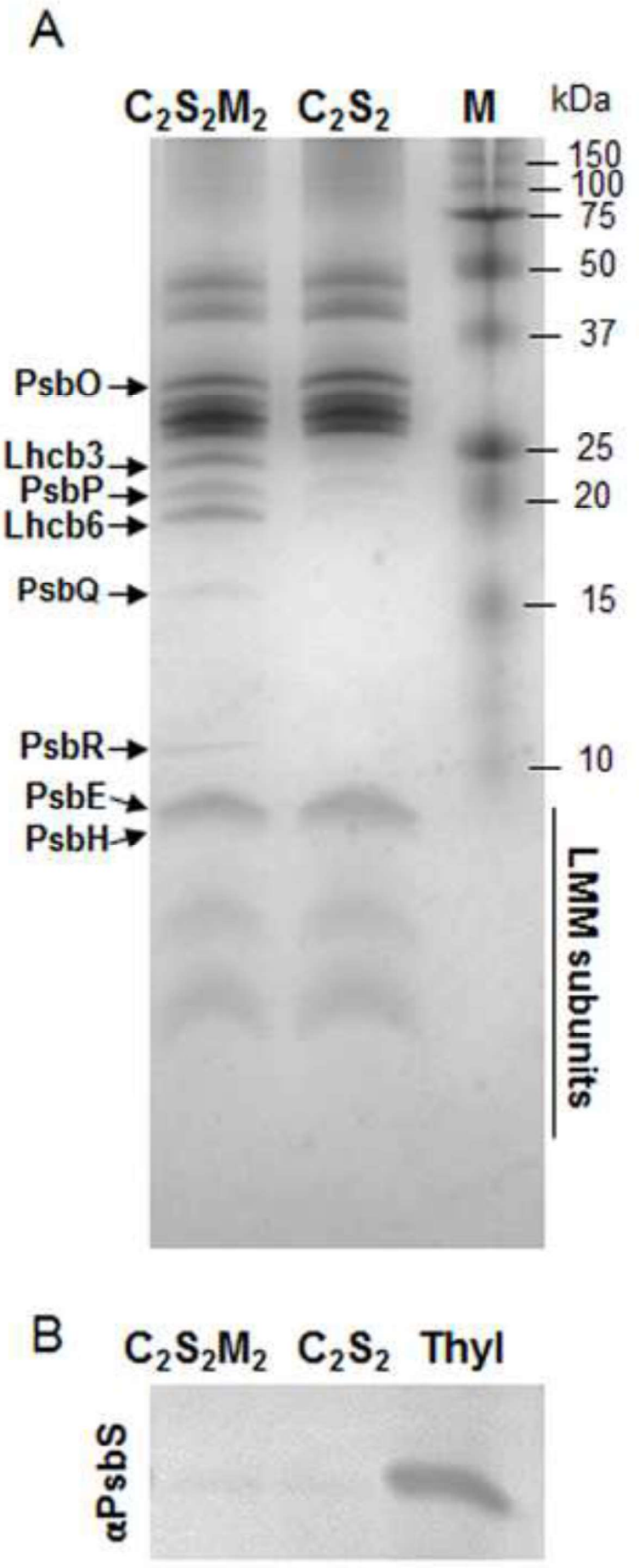


Figure 3

Click here to download high resolution image

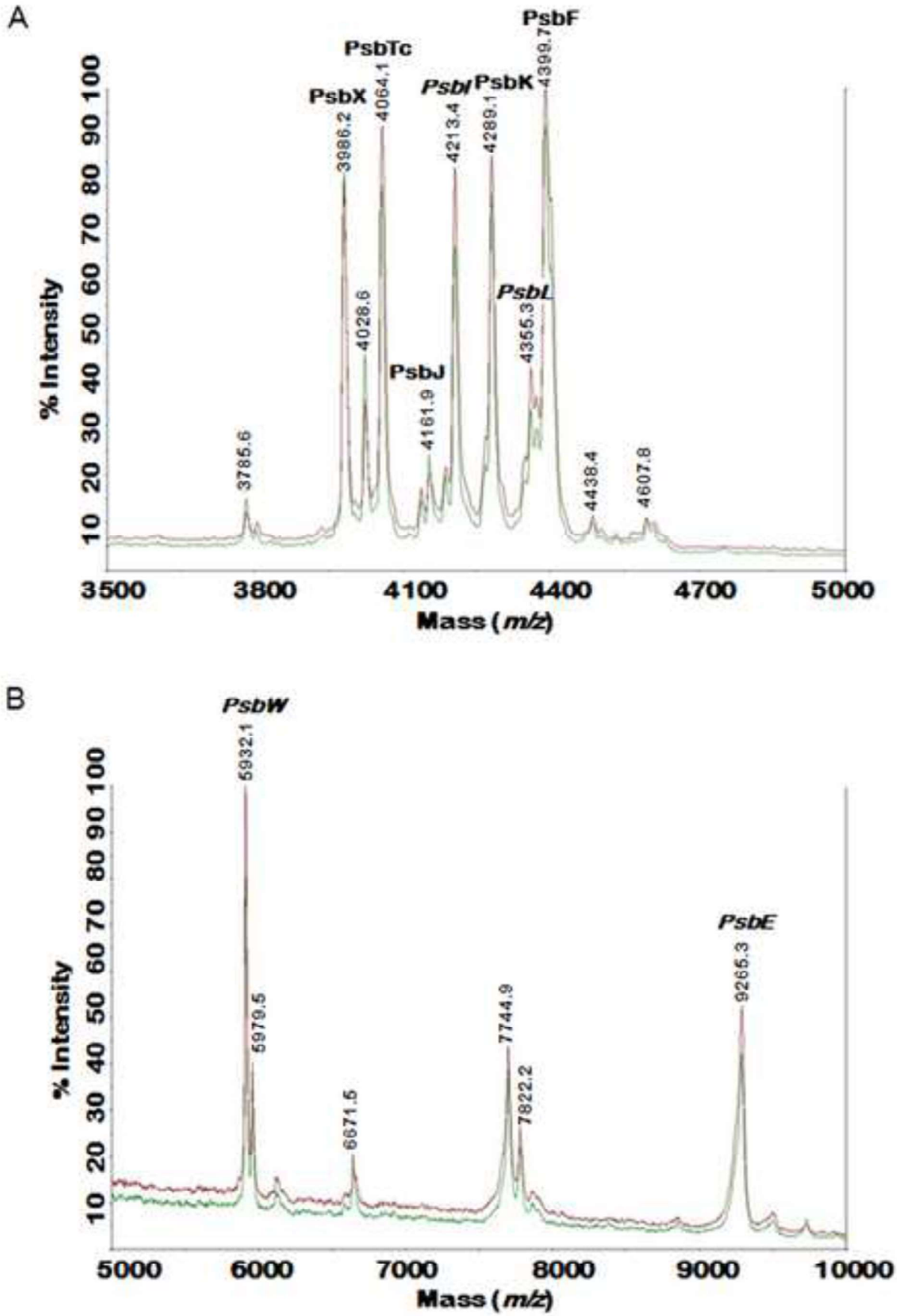
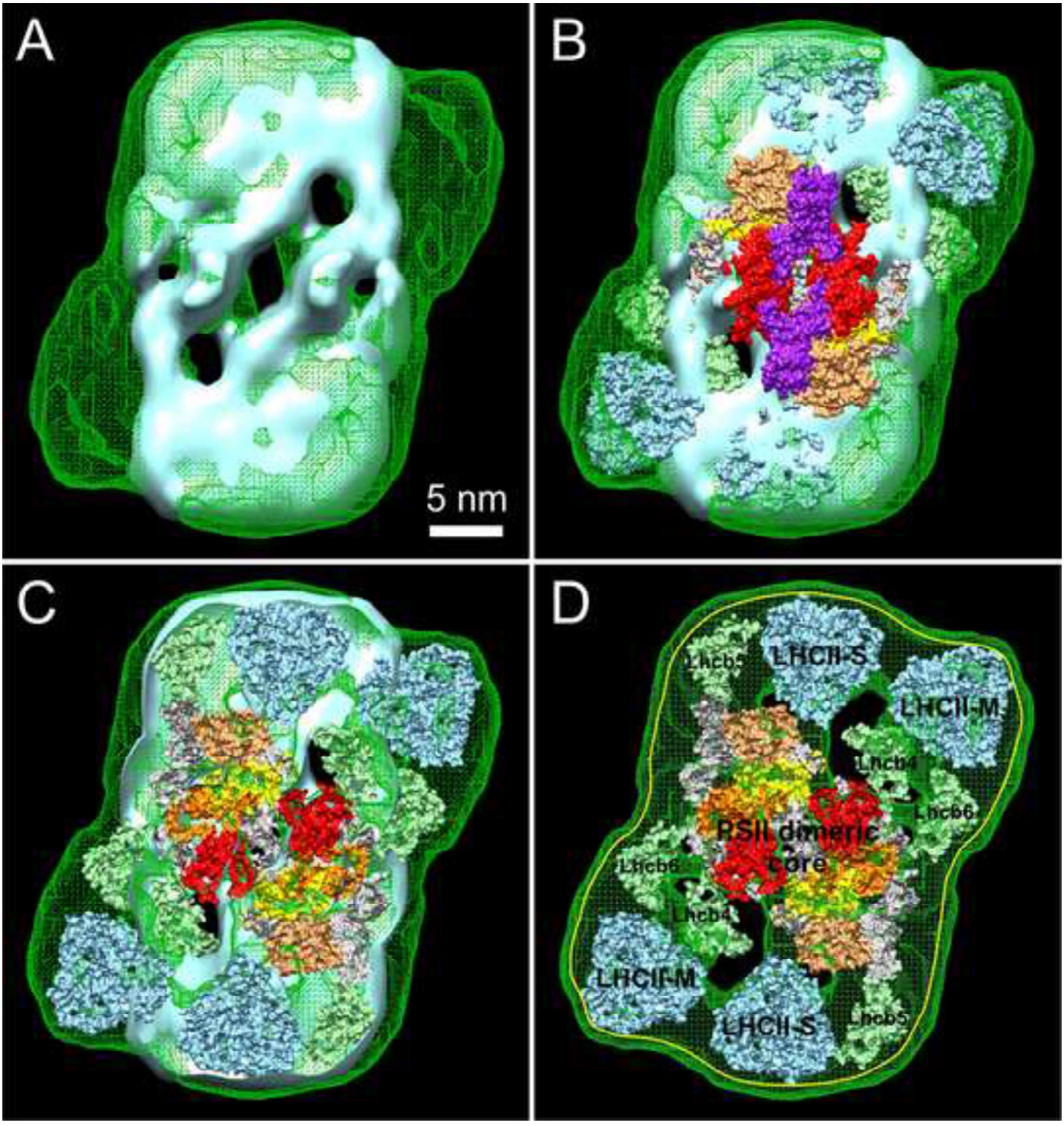


Figure 4
Click here to download high resolution image



Supplementary Material Table S1
Click here to download Supplementary Material (for online publication): Table S1.xls

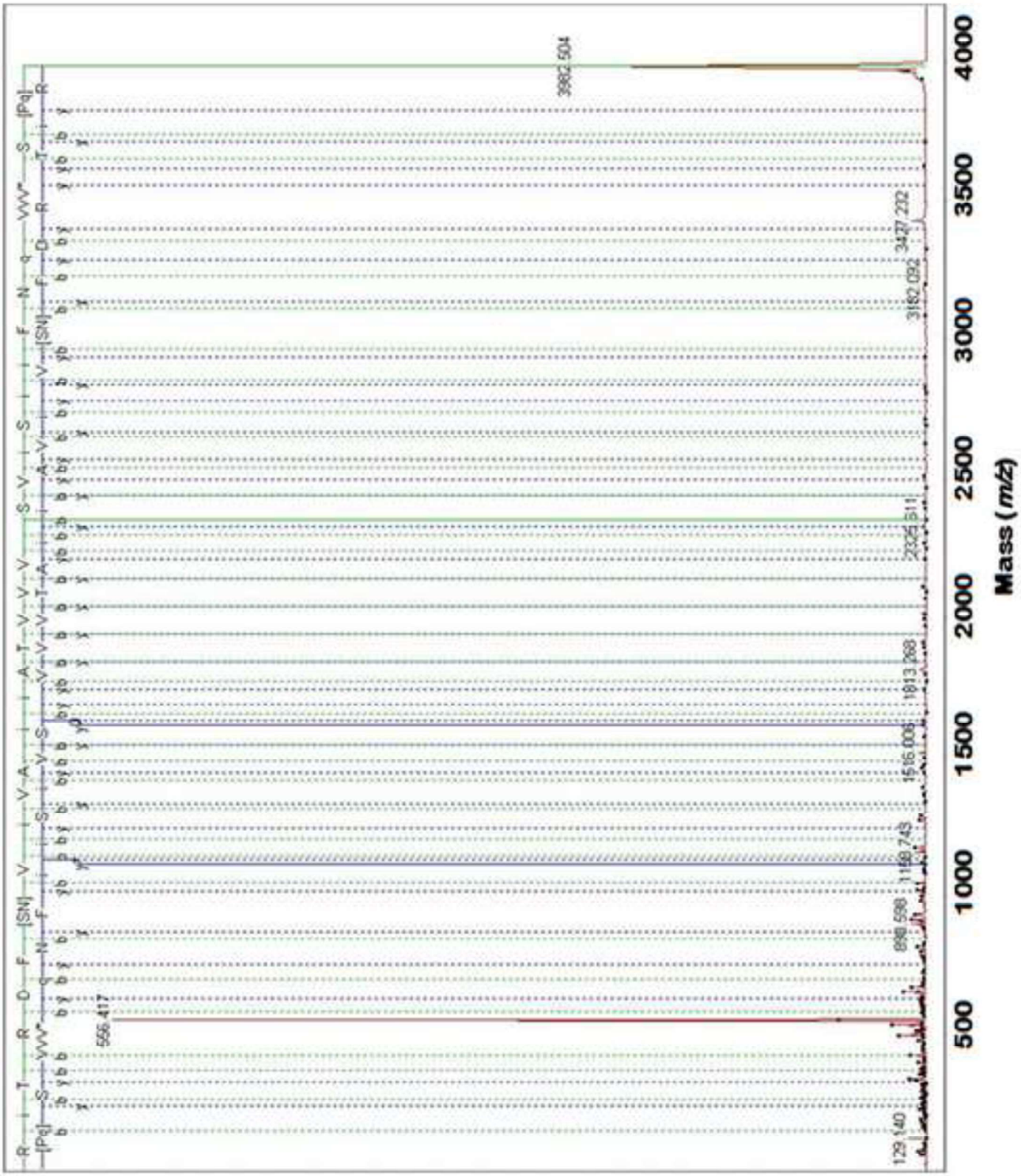
Protein	Precursor ion mass *m/z*	Sequence	Mr (kDa)	AC number (gi NCBI) reference organism	% Identity with *P. sativum or A. thaliana* (sequence coverage)
CP47	1922.8587 1759.8741 1484.7299 2186.0215 1289.6615	YQWDQGYFQQEIYR VGGGLVENQSLSEAWSK LAFYDYIGNNPAK AGSMDNGDGIAVGWLGHPIFR AQLGEIFELDR	54208	gi|27446507 (*Pisum sativum*)	100% gi|27446507 *P. sativum* (15%)
CP43	1708.7594 1890.9952 1818.8822	DQETTGFAWWAGNAR APWLEPLRGPNGLDLSR GIDRDFEPVLSMTPLN	51955	gi|295136994 (*Pisum sativum*)	100% gi|295136994 *P. sativum* (10%)
D2	2044.9160 1226.5931 1546.6827 1040.5978 2636.2693	AFNPTQAEETYSMVTANR AYDFVSQEIR AAEDPEFETFYTK NILLNEGIR AWMATQDQPHENLIFPEEVLPR	35386	gi|27435890 (*Pisum sativum*)	100% gi|27435890 *P. sativum* (10%)
D1	1458.7255 1313.7092 1110.5458	LIFQYASFNNSR VINTWADIINR NAHNFPLDLA	38962	gi|131252 (*Pisum sativum*)	100% gi|131252 *P. sativum* (9%)
PsbO	1235.6510 2440.1329 1421.7588 2269.1002 2293.1226 963.5793 2283.1535 2433.1408 3359.7348 1292.6401	RLTFDEIQSK GTGTANQCPTIDGGVDSFSFKPGK KLCLEPTSFTVK LTYTLDEIEGPFEVSADGSVK FEEKDGIDYAAVTVQLPGGER VPFLFTIK QLVASGKPDSFSGEFLVPSYR GASTGYDNAVALPAGGRGDEEELGK ITLSVTQTKPETGEVIGVFESIQPSDTDLGAK IQGVWYAQLES	34872	gi|131384 (*Pisum sativum*)	100% gi|131384 *P. sativum* (59%)
PsbP	1604.6995 2373.1336 1571.7355 3396.6474 1255.6561 1839.9327 1386.6667	TNTDYLPYNGDGFK YEDNFDATSNVSVLVQTTDKK SITDYGSPEEFLSK QAFFGQTDSEGGFDTNAVAVANILESSAPVIGGK QYYNISVLTR TADGDEGGKHQLITATVK KFVEDTASSFSVA	28030	gi|131390 (*Pisum sativum*)	100% gi|131390 *P. sativum* (48%)
PsbQ	1990.0120 1110.6761 1655.8770 1258.6458 1470.7467 1530.8194	VGGPPPLSGGLPGTLNSDEAR DLKLPLKER FFIQPLAPTEAAAR AWPFLQNDLR LFQDISNLDHAAK YYAIAVSTLNDVLSK	25265	gi|31096349 (*Pisum sativum*)	100% gi|31096349 *P. sativum* (35%)
PsbR	2537.2068 2608.2486	IKTDTPYGTGGGMDLPNGLDASGRK GVYQFVDKYGANVDGYSPIYEPK	14170	gi|33694227 (*Trifolium pratense*)	71% gi|15219268 *A. thaliana* (29%)
PsbE	1121.6081 953.5658 1484.6783	SFADIITSIR QGIPLITGR FDSLEQLDEFSR	9381	gi|293338576 (*Pisum sativum*)	100% gi|293338576 *P. sativum* (41%)
PsbH	1732.9247	TVVGDLLKPLNSEYGK	7824	gi|295137033 (*Pisum sativum*)	100% gi|295137033 *P. sativum* (23%)
Lhcb1	1377.6313 3944.7581 1251.6683 1934.9747 982.4913 4548.2999 3525.7555 3123.5050	VASSGSPWYGPDRVK YLGPFSGESPSYLTGEFPGDYGWDTAGLSADPETFSK NRELEVIHSR WAMLGALGCVFPELLSR FGEAVWFK AGSQIFSEGGLDYLGNPSLVHAQSILAIWATQVILMGAVEGYR IAGGPLGEVVDPLYPGGSFDPLGLADDPEAFAELK GPLENLADHLADPVNNNAWSYATNFVPGK	28635	gi|115788 (*Pisum sativum*)	100% gi|115788 *P. sativum* (72%)
Lhcb2	1601.7838 4053.8585 1950.9583 982.4913 3455.6773	SAPESIWYGPDRPK YLGPFSEQIPSYLTGEFPGDYGWDTAGLSADPETFAR WAMLGALGCTFPELLEK FGEAVWFK VGGGPLGEGLDPLYPGGAFDPLGLADDPDSFAELK	28866	gi|115797 (*Pisum sativum*)	100% gi|115797 *P. sativum* (43%)
Lhcb3	1247.6299 3926.7839 1901.9743 1293.6758	DLWYGPDRVK YLGPFSAQTPSYLTGEFPGDYGWDTAGLSADPEAFAK WAMLGALGCITPEVLQK VDFKEPVWFK	28710	gi|20671 (*Pisum sativum*)	100% gi|20671 *P. sativum* (24%)
Lhcb4	1768.8784 1315.6455 1418.7082 985.5808	STPFQPYTEVFGLQR FRECELIHGR FFDPLGLAADPEK ATLQLAEIK	31290	gi|346987811 (*Dimocarpus longan*)	80% gi|38502901 *A. thaliana* (16%)
Lhcb5	1157.6445 1497.6711	IFLPDGLLDR YGANCGPEAVWFK	30138	gi|15235029 (*Arabidopsis thaliana*)	100% gi|15235029 *A. thaliana* (8%)
Lhcb6	1854.8384	TAENFNNSTGLQGYPGGK	23518	gi|168009690 (*Physcomitrella patens*)	82% gi|4741960 *A. thaliana* (7%)

Supplementary Material Table S2

Click here to download Supplementary Material (for online publication): Table S2.xlsx

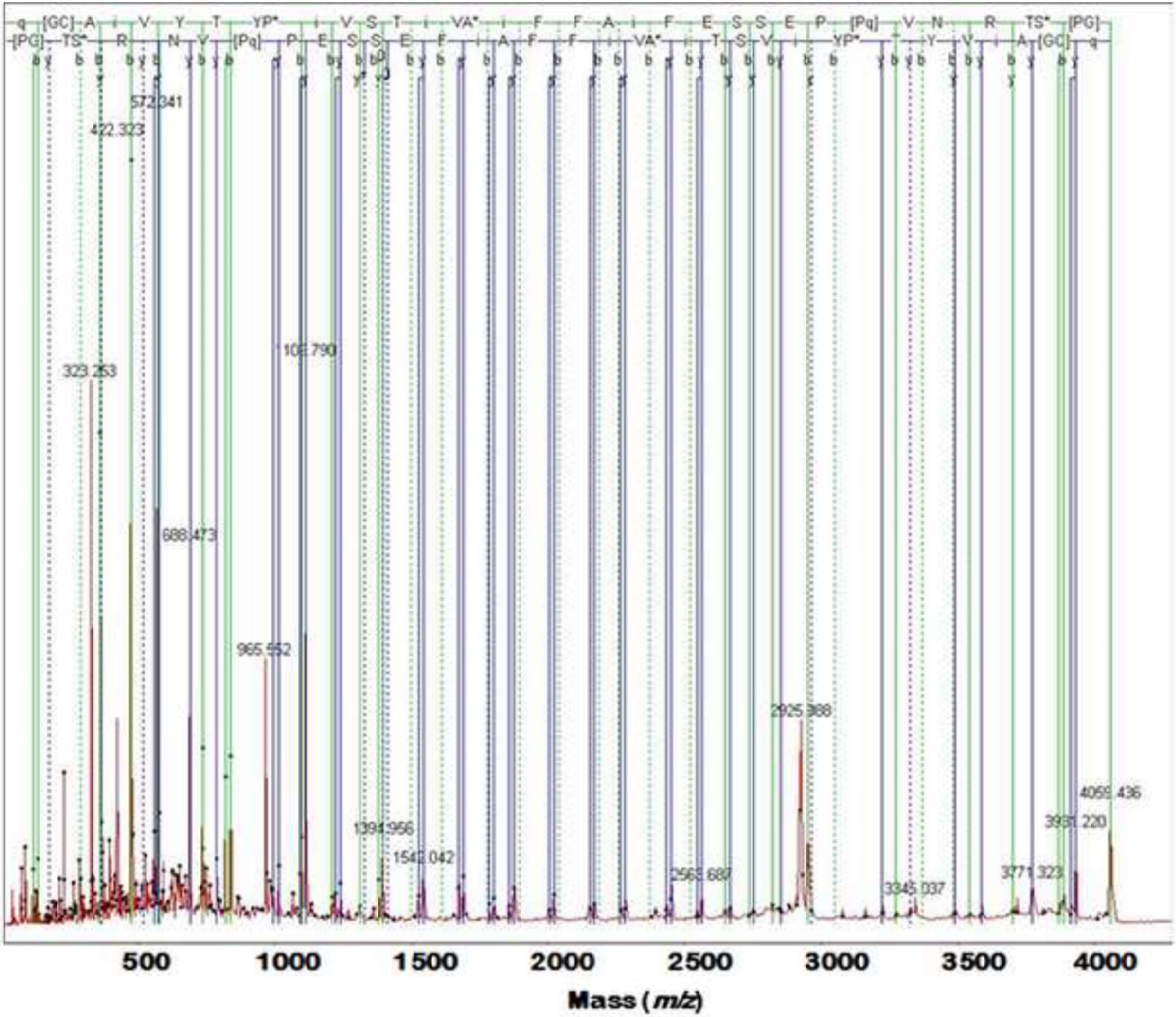
Experimental precursor *m/z*	Protein	Identified peptide sequence(s) from *de novo* sequencing	UniProtKB/Tremble	MS Blast results
3982.2	**PsbX**	R i T G R D F [SN] V G i V A G i i A T V V V G G S V i S i i F N q [VVV|VTP|PPC|PEA|IIA|HGC|SPi|LAi] S [Pq]	Q8VYY1	Total Score = 185 (97.4 bits) Identities = 25/32 (78%), Positives = 29/32 (90%) Query: 284 SXXXZNFLLSLVSGGVVVTALLGAVLGVSNFD 315 S +NFLLS+VSGGVVVTA+LGAV+GVSNFD Sbjct: 48 SPSLKNFLLSIVSGGVVVTAILGAVIGVSNFD 79
4059.2	**PsbTc**	q [GC] A i V Y T [YP|ME|Fi] i V S T i [VA|Gi] i F F A i F E S S E P [Pq] V N G R [TS|MG] [PG]	Q9GF83	Total Score: 170 Score = 60 (30.2 bits) Identities = 7/8 (87%), Positives = 8/8 (100%) Query: 22 GIIFFAIF 29 GI+FFAIF Sbjct: 15 GILFFAIF 22 Score = 44 (22.6 bits) Identities = 4/7 (57%), Positives = 7/7 (100%) Query: 13 FIIVSTI 19 F++VST+ Sbjct: 8 FLLVSTL 14 Score = 37 (19.3 bits) Identities = 4/6 (66%), Positives = 6/6 (100%) Query: 4 AIVYTY 9 A+VYT+ Sbjct: 3 ALVYTF 8 Score = 29 (15.5 bits) Identities = 4/5 (80%), Positives = 4/5 (80%) Query: 33 EPPQV 37 EPP V Sbjct: 25 EPPRV 29
4156.3	**PsbJ**	i P S V i i G [SM|FA|DC] A q W i i G E A A G i V V i G i i [AA] [HA] [YW] A [PP|HG] G T T R R	P13555	Total Score = 73 (39.3 bits) Identities = 8/15 (53%), Positives = 13/15 (86%) Query: 915 WLLGEAAGLVVLGLL 929 W++G AG+VV+GL+ Sbjct: 11 WIIGTVAGIVVIGLI 25
4283.3	**PsbK**	[Pi] i T F [CC] W V F A i i F F i i P i V P M F D P [VN|RG] P E F A Y A V [SG] [CC]	Q5SCX1	Total Score: 171 Score = 99 (51.4 bits) Identities = 12/15 (80%), Positives = 14/15 (93%) Query: 1227 MPVLPLLFFLLAFVW 1241 MP++PL FFLLAFVW Sbjct: 32 MPIIPLFFFLLAFVW 46 Score = 37 (20.3 bits) Identities = 5/7 (71%), Positives = 6/7 (85%) Query: 1180 VFALLFF 1186 VF L+FF Sbjct: 9 VFHLIFF 15 Score = 35 (19.3 bits) Identities = 5/5 (100%), Positives = 5/5 (100%) Query: 1124 AYAVS 1128 AYAVS Sbjct: 21 AYAVS 25
4393.4	**PsbF**	R i H V A [TN|Sq] E T V S G i i [RN] A H V A i G H V A i W R V [PN] q W C [SS|CA] R G M G	Q6EYQ4	Total Score = 83 (44.3 bits). Identities = 14/27 (51%), Positives = 17/27 (62%) Query: 3908 VRWLAVHGLAVHARNLLGSVTESZAVH 3934 VRWLAVHGLAV LGS+ ++ Sbjct: 12 VRWLAVHGLAVPTVSFLGSISAXXXIQ 38

Supplementary Material Figure S1
Click here to download high resolution image



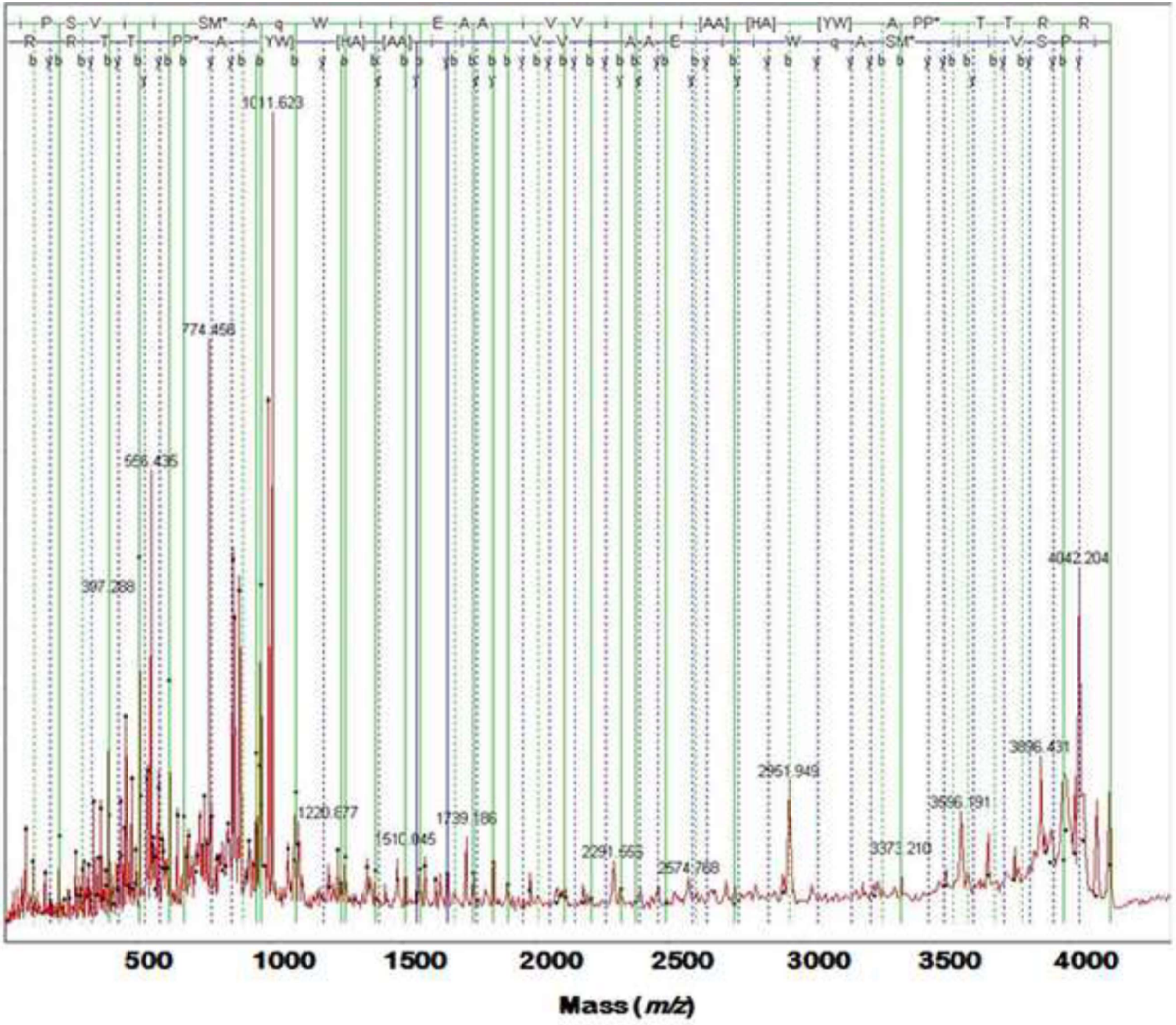
Supplementary Material Figure S2

Click here to download high resolution image



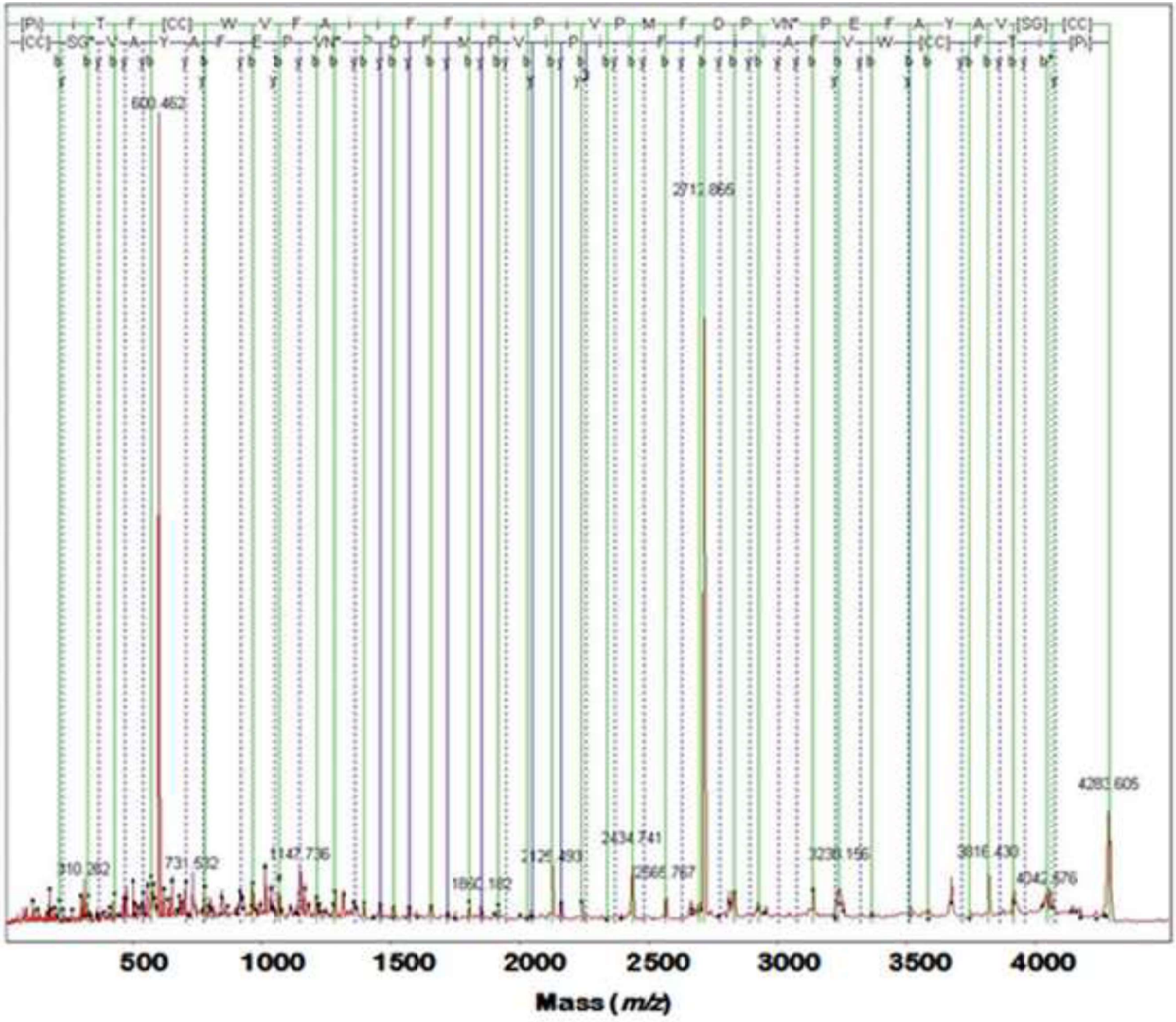
Supplementary Material Figure S3

Click here to download high resolution image



Supplementary Material Figure S4

Click here to download high resolution image



Supplementary Material Figure S5
Click here to download high resolution image