Structural insights into PsbS and its interactions with the PSII-LHCII supercomplex under active and inactive nonphotochemical states in *Spinacia oleracea*

November 2022, Utrecht University (Utrecht, Netherlands)

Author: Isabel de Moya Clark

Supervisors: Pascal Albanese (daily) and Richard Scheltema

Abstract

Nonphotochemical quenching (NPQ) is a photoprotective mechanism activated under strong light conditions to safely dissipate excess light energy into heat. This mitigates the production of radical oxygen species (ROS) and the photooxidative damage of thylakoidal proteins. The PsbS protein is a key component of NPQ in higher order plant photosynthesis. Higher plants' thylakoid membranes are characterized for being enriched in Photosystem II (PSII) and light-harvesting complex II (LHCII) which together form the PSII-LHCII supercomplex (PSII-LHCII sc) and form stacked structures in the grana. Under excess light conditions, the thylakoid lumen acidifies activating PsbS which mediates light-regulated interactions in the PSII-LHCII sc leading to its macroorganization and causing structural changes in the grana stacks. The light-dependent conformational changes of PsbS and its interactions with PSII and LHCII components are still largely unknow. Additionally, it has been shown that altering NPQ could have significant implications in crop engineering to increase plant productivity. This makes understanding the dynamic behavior of PsbS and the photosynthetic structures it interacts with of great interest. In this project, we aim to extract structural details of the PsbS conformation and its interactions with the PSII-LHCII sc under NPQ-inactive (dark) and NPQ-active (light) states. To achieve this, we make use of crosslinking mass spectrometry (XL-MS) structural proteomics approaches and molecular modelling. We also explore the incorporation of structural evolutionary predictions, via AlphaFold 2 and Evolutionary couplings, to complement the experimental data. Our results show that we find dimeric PsbS under both dark and light conditions with different conformations in its N-termini. Moreover, we use the XL-MS data to build a structural model which shows that PsbS, Lhcb6 and the LHCII moderately bound trimers disassociate together from the PSII-LHCII sc under NPQ conditions.

Key words: Photosynthesis, nonphotochemical quenching (NPQ), PsbS, photoprotection, Photosystem II (PSII), Light harvesting complex II (LHCII), cross-linking mass spectrometry (XL-MS), structural modelling, Lhcb1, Lhcb2, Lhcb3, Lhcb6, Evolutionary couplings, AlphaFold 2 and *Spinacia oleracea*.

Plain Layman summary

Solar energy is one of the most abundant renewable energy sources available in our planet. Photosynthesis arose early on in Earth's Life History as a way to convert solar energy into chemical energy usable by living organisms. The earliest forms of photosynthesis were anoxygenic (does not involve oxygen). However, around 2.4 billion years ago, oxygenic photosynthesis (involves oxygen) emerged. This led to a rise in atmospheric oxygen and resulted in a surge of novel ways to produce biological molecules and the appearance of new forms of life. There is a wide range of oxygenic photosynthetic organisms going from cyanobacteria (unicellular, aquatic photosynthetic organisms) to vascular plants, and these are all adapted to different environmental conditions. Early photosynthetic organisms, from which land plants evolved, survived under low light conditions in which a more diluted light energy reached the photosynthetic components. The transition to higher light conditions caused the parts of the organisms performing photosynthesis to receive more light than they could process. Moreover, the excess light energy accumulated in the photosynthetic components generated adverse effects, called photooxidative damage, harming these photosynthetic components. To counter this, oxygenic organisms such as cyanobacteria, algae and plants have evolved to develop protective mechanisms against excess light collectively called nonphotochemical quenching (NPQ). This is a very fast process by which excess light is dissipated into heat diminishing the photooxidative damage. PsbS is a key protein involved in the NPQ process and it is present in both algae and plants. Nevertheless, the details on the structure of this protein and the details on the dynamics of how this protein interacts with the photosynthetic components is yet not fully understood. Furthermore, it has been shown that altering NPQ could have significant implications on crop engineering to increase plant productivity. This makes understanding the dynamic behavior of PsbS and the photosynthetic structures it interacts with of great interest. In this project, we aim to gain further insight in the structural details of the structure of PsbS and its interactions with the photosynthetic proteins using the combination of both novel experimental and computational approaches.

Index

1. INTRODUCTION 4

1.1. Oxygenic photosynthesis and the excess light dilemma 4

1.2. Nonphotochemical Quenching (NPQ) and the qE component 5

1.3. The structure of the PSII-LHCII supercomplex 6

1.4. Analyzing interactions in highly dynamic and complex structures 7

1.5. Aim of the project 7

2. METHODS 8

2.1. Thylakoid purification 8

2.2. Optimization of crosslinking conditions 8

2.3. NPQ induction and crosslinking of thylakoid membranes 9

2.4. SDS-PAGE and lpBN-PAGE gel preparation 9

2.5. Trypsin in gel digestion 9

2.6. Desalting and fractionation 10

2.7. LC–MS/MS 10

2.8. RAW data Analysis 10

2.9. Structural modelling and crosslinking mapping 11

2.10. EvoMAS bundle development 11

2.11. Evolutionary predictions 12

3. RESULTS 12

3.1. PsbS has at least 3 N-termini proteoforms in S. oleracea 12

3.2. PsbS appears in its dimeric state both dark and NPQ conditions. 13

3.3. PsbS has varying N-termini conformations under dark and light (NPQ) conditions 13

3.4. PsbS and Lhcb6 interact with the LHCII trimers under NPQ conditions. 14

3.5. Evolutionary predictions of the interactions of PsbS with Lhcb6 16

4. DISCUSSION 17

5. CONCLUSIONS 19

6. REFERENCES 19

1. INTRODUCTION

1.1. Oxygenic photosynthesis and the excess light dilemma

Both algae and plants, as oxygenic photoautotrophic organisms, need light to be able to carry out photosynthesis for their survival. In higher order plants, photosynthesis takes place in organelles called chloroplasts (Figure 1). Chloroplasts have a unique architecture consisting of an outer double membrane (envelope) which encloses single membrane bound compartments (thylakoids). These thylakoids can form stacked structures called grana. Moreover, grana structures can be connected with each other through unstacked thylakoid domains (lamella). The volume enclosed inside the thylakoid membranes is called the lumen and the volume enclosed in the chloroplasts but outside the thylakoidal structures is called stroma. The photosynthetic components (protein structures and pigments) involved in the light-dependent reactions of oxygenic photosynthesis can be found in the thylakoid membrane system. (Koochak *et al.,* 2018; Bassi *et al.,* 2021).

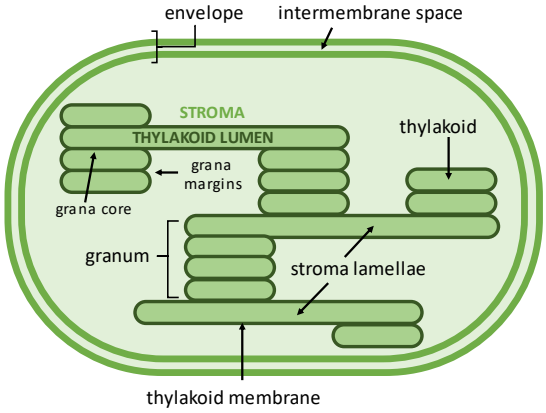


Figure 1. Architecture of chloroplasts in higher order plants. Chloroplasts are organelles that consist of a double outer membrane (envelope) that encloses single membrane structures called thylakoids. Thylakoids can form stacked structures called grana and these can be connected through unstacked domains called lamellae. The volume inside the thylakoids is called lumen and the volume between the thylakoid membranes and the inner membrane of the envelope is called stroma. The figure is based on Figure 2 of Bassi *et al.* (2021).

Briefly, we will go over the key functional units and streps involved in the light-dependent reactions of oxygenic photosynthesis. Firstly, the photosynthetic chain consists of several subunits being the most relevant: photosystem II (PSII), cytochrome b6f, photosystem I (PSI) and ATP synthase (which appear in the chain in this order). Additionally, PSII and PSI can interact with light harvesting complexes (LHCII and LCHI respectively) which contain pigment (chlorophylls and carotenoids) binding proteins. The LHC structures form arrays which are referred to as antenna systems. Under normal non-saturating light conditions, photons reach the chlorophyll and other pigments in the antenna systems, exciting them in the process. Then the antenna systems transfer the excitation energy to the reaction centers of the photosystems PSII and PSI which contain special pairs of chlorophylls (P680 and P700 respectively) that, in turn, also get excited. The chlorophylls then return to their resting states through photoquenching mechanisms releasing and transferring electrons to the photosynthetic chain. This triggers a series of reactions, including proton (H^+) pumping into the lumen, which generates an electrochemical gradient that is utilized to generate adenosine triphosphate (ATP). This ATP production will provide the energy needed to fix inorganic carbon from CO_2, by the RUBISCO enzyme, in slower dark (light-independent) photosynthetic reactions (Wolfgang *et al.*, 2019; Bassi *et al.*, 2021). Under excess light conditions, the excitation pressure in the antenna increases, the photosynthetic chain becomes saturated, and the thylakoid lumen acidifies (through an increase in proton pumping). The build-up of excess excitation energy reduces the efficiency of photosynthesis and generates adverse effects such as the photooxidative damage of the proteins in the photosynthetic chain (Björkman *et al.,* 1995).

Excessive amounts light during photosynthesis can cause photooxidative damage through the reduction of O_2 to superoxide anion (O_2^-), hydroxyl radical (OH·) and reactive oxygen species (ROS) as well as the release of singlet oxygen ($^1O_2$*). Most of the ROS are generated in the PSII reaction center, PSI and LHCII (Nelson *et al.,* 2005). These byproducts can lead to photoinhibition decreasing photosynthetic productivity (Tjus *et al*., 1998; Tjus *et al*., 2001). In extreme cases, excess light conditions, can even cause pigment bleaching and death. Therefore, protective mechanisms have been developed, by algae and plants, to prevent the detrimental effects of excess light (Muller *et al.,* 2011). The light received by plants varies seasonally and daily and therefore different photoprotective processes have been developed for each case. One of them is called acclimation, which takes several days and by which plants can change the architecture of chloroplasts to regulate the amount of light energy that reaches the photosynthetic chain. This mechanism is useful for plants to adapt to seasonal changes, but it is too slow to deal with quick fluctuations in photon fluxes such as those caused by diurnal changes. There is a collection of mechanisms that tackle the photooxidative damage caused by rapid shifts in light availability which are collectively called nonphotochemical quenching (NPQ) (Bassi *et al.,* 2021).

1.2. Nonphotochemical Quenching (NPQ) and the qE component

To mitigate the negative effects of photooxidative damage, caused by fast fluctuations in light intensity, a collection of energy dissipation mechanisms take place, and they are referred to as NPQ (Müller *et al*., 2011). The strongest NPQ component is qE and it safely transforms excess light energy into heat. This process is rapidly forming in high light (0.5-2 min) and rapidly reversible in low light (Bassi *et al.*, 2021). Two synergistic molecular mechanisms are crucial in activating and sustaining qE: one involves a specific photoprotective carotenoid pigment, zeaxanthin (Muller *et al*., 2001), and the other involves a conformational change in the thylakoid membrane and is activated by the LHC-like protein, PsbS (Demmig-Adams *et al.,* 1990; Li *et al.,* 2000). However, the actual site of energy quenching is still unknown. The mainstream theory posits that in plants quenching occurs in trimeric LHCII subunits, which aggregate to form quenched clusters (Horton *et al*., 1996; Johnson *et al*., 2011). Recent observations have shown that plants deprived of LHCII trimers retain ~40% of their NPQ capacity (Nicol *et al.* 2019) suggesting that a quenching site is also located in the PSII core. Alternative models place the quenching site in the monomeric antennae Lhcb4, Lhcb5 and Lhcb6 (Walters *et al.,* 1994). Ultimately, these intensive, decades-long research efforts suggested that the site of quenching is not comprised of a single mechanism

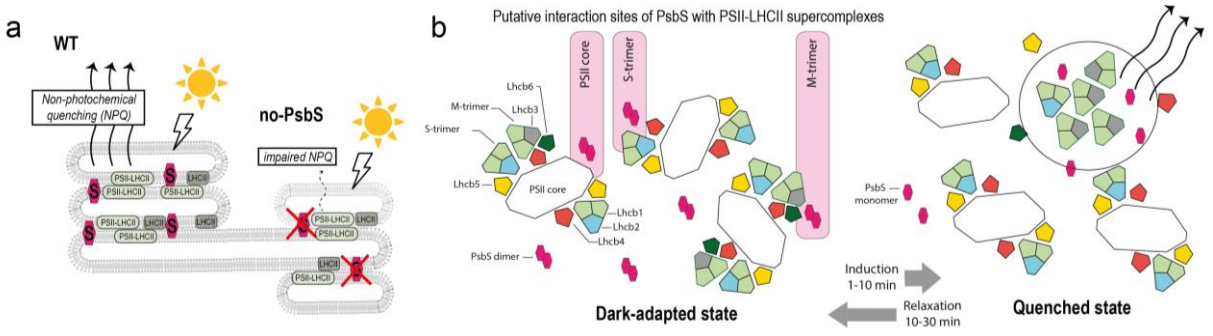


Figure 2. Overview of PsbS-related processes and non-photochemical quenching. **A)** Plants lacking PsbS are severely impaired in their capacity to cope with high irradiances. Multiple putative binding sites of PsbS to PSII-LHCIIsc (the components of which are schematically shown in **B)** have been so far indirectly assessed in mosses (Gerotto *et al.,* 2015) and plants (Correa-Galvis *et al.,* 2016). **B)** The role of PsbS in the structural remodeling of the PSII-LHCII sc, and consequently of the thylakoid membranes, is still matter of debate, but can roughly be summarized as depicted.

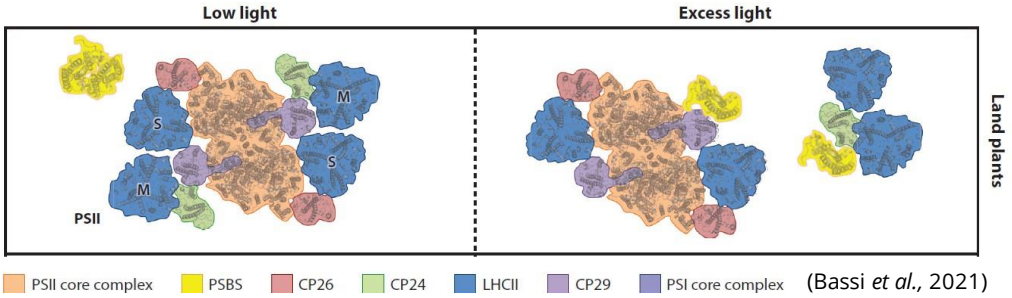


Figure 3. Proposed model for the macrostructural reorganization of PSII-LHCII sc induced by PsbS under excess light conditions in higher order plants. Under low light conditions PsbS remains as a dimer and does not interact with the PSII-LHCII sc. Under excess light conditions, PsbS monomerizes and interacts with PSII-LHCII sc leading to the dissociation of the moderate (M) LHCII trimers and the protein Lhcb6. This figure and model are extracted from Bassi *et al*. (2021).

(Holzwarth *et al.,* 2009). Nonetheless, consensus assigns a pivotal role to PsbS which, under excess light conditions, acts as a pH sensor that can detect the acidification of the lumen and mediate light-dependent reactions. Also, plants lacking PsbS are severely impaired in their capacity to cope with high irradiances (Figure 2a). While probably not the actual quenching site itself, as it does not bind pigment molecules (Bonente *et al.,* 2008; Funk *et al.,* 1995), PsbS is thought to promote a structural rearrangement of PSII-LHCII supercomplex (PSII-LHCII sc) that triggers energy dissipation (Johnson *et al.,* 2011; Kereïche *et al.,* 2010), possibly through undocking of some specific antenna modules (Figure 2b) (Betterle *et al.,* 2009). A proposed model for this reorganization is shown in Figure 3 (Bassi *et al.,* 2021). PsbS-triggered remodeling was also suggested to affect the stacking of thylakoids, with PsbS indirectly involved as a promoter of PSII-LHCII sc dis-assembly and re-assembly (Kiss *et al.,* 2008) that implies a dynamically regulated structural interaction. The PSII-LHCII sc, on which extensive research has been done, has a highly complex structure that needs to be understood in other to analyze its interactions with PsbS.

1.3. The structure of the PSII-LHCII supercomplex

In this project, we are specifically interested in unveiling structural information on the PsbS structure and its interactions with the PSII-LHCII sc. To better understand the results, we will describe the structure of this supercomplex (Figure 4). It consists of a PSII dimeric core (C_2), two strongly bound LHCII trimers (S_2) and two moderately bound trimers (M_2). The predominantly occurring PSII–LHCII sc are: $C_2S_2M_2$, C_2S_2M and C_2S_2 (Caffari *et al.,* 2009). Their relative abundances, in the thylakoid membranes, depend on the light intensity (Albanese *et al.,* 2016; Kouřil *et al.,* 2013). These supercomplexes stack in the grana (Figure 2a) forming a sandwich with exposed flexible N-termini in the inter-stromal gap. (Albanese *et al*., 2020).

The structural organization of the PSII catalytic core has been fundamentally conserved throughout the evolution of photosynthetic organisms from cyanobacteria to higher plants (Pagliano *et al.,* 2013). PSII is involved in the water-splitting reactions, and it is characterized by its large reaction center D1 and D2 subunits. The antenna system has had a greater evolutionary

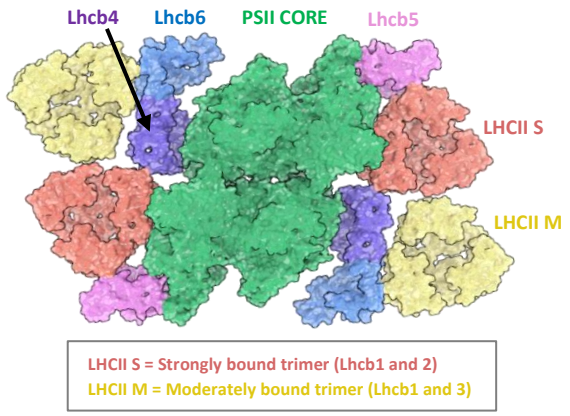


Figure 4. The structure of the PSII-LHCII supercomplex in *A. thaliana* (PDB: 7oui; Graca *et al.,* 2021)**.** In green is C_2, in red S_2 and in yellow M_2. The monomeric Lhcbs are shown in purple, pink and blue.

diversification. There are a wide range of species-specific LHCII isoforms encoded by multiple genes, but they still maintain a strictly conserved fold and structural organization (Koziol *et al.,* 2007; Ballottari *et al.,* 2012). The LHCII trimers are composed of the Lhcb1, Lhcb2 and Lhcb3 proteins being Lhcb2 found in the S-trimer, Lhcb3 in the M-trimer and Lhcb1 in both (Albanese *et al*., 2020; Caffari *et al.,* 2021). Furthermore, the binding of the trimers to the PSII core relies on three monomeric subunits: Lhcb4, Lhcb5 and Lhcb6. Lhcb4 acts as a linker for both trimers, Lhcb5 exclusively for the S-trimer and Lhcb6 exclusively for the M-trimer (Dekker *et al.,* 2005). It is proposed (model in Figure 3) that PsbS, under NPQ conditions, monomerizes and interacts with Lhcb6 and M-trimers. To unveil the interactions of PsbS, the complexity of the PSII-LHCII sc needs to be considered and therefore different approaches have been considered.

1.4. Analyzing interactions in highly dynamic and complex structures

The PSII-LHCII sc is highly complex due to its light-driven structural dynamics and its heterogeneous composition. For this project, we considered the combination of crosslinking mass spectrometry (XL-MS) with evolutionary predictions to be a suitable approach to gain a greater insight into the PSIILHCII sc and PsbS interactions. XL-MS is an experimental approach, that can be applied to proteomics, that uses small chemical crosslinkers to find peptide sequences close to each other in the 3D structure of proteins. This technique has demonstrated considerable potential in investigating protein-protein interactions in complex protein mixtures in an unbiased manner. The combination of XL-MS data together with other techniques such as cryo-EM has been successfully used to generate final protein complex models (Schweppe *et al.,* 2017; Albanese *et al.,* 2020). Nevertheless, XL-MS is limited by the accessibility of the crosslinker which in this case can be restricted by the thylakoid membrane. Therefore, complementary approaches are needed to extract information on the interactions between the protein regions embedded in the thylakoid membrane. Here we used evolutionary predictions. In the recent years, powerful tools to predict protein structures and interactions have been developed. One of the most famous is AlphaFold 2 (AF2; Hockenberry *et al.,* 2019; Evans *et al.,* 2021) which extracts evolutionary information from multiple sequence alignments (MSA) to predict the structure of proteins. It has been shown to predict monomeric structures to a nearly experimental resolution (monomer pipeline) and can also predict protein interactions (multimer pipeline). There is another tool called EVolutionary couplings (EV couplings) (Hockenberry *et al.,* 2019), which is also based on the extraction of evolutionary information from MSAs, that predicts coevolving amino acids within the same structure (intralinks) or between two different structures (interlinks). Both these tools will be incorporated to try and complement the experimental data.

1.5. Aim of the project

Structural models for the dynamics of PsbS and its interactions are urgently needed to gain mechanistic insights into the driving factors behind NPQ. Fundamentally, this knowledge could be used to optimize the light-use efficiency of crop plants that are constantly exposed to photoinhibitory conditions or sudden changes in light cues in the canopy (Raven *et al*., 2011; Ort *et al.,* 2015), hence increasing their resilience to the adverse effects of climate change. Therefore, the aim of this project will be to gain further insight into the details of the structure of PsbS and its interactions with the photosynthetic proteins in PSII-LHCII sc using a combination of both novel experimental (XL-MS) and computational (EV couplings and AF2) approaches (Figure 5). We will focus on higher order plant photosynthesis of the organism *Spinacia oleracea*. Structural information from *Arabidopsis thaliana* and *Pisum sativum* will also be incorporated given their evolutionary proximity and structural similarity to *S. oleracea* (Ballotari *et al.,* 2012).

2. METHODS

A schematic overview of the steps followed in the methodology can be found in Figure 4.

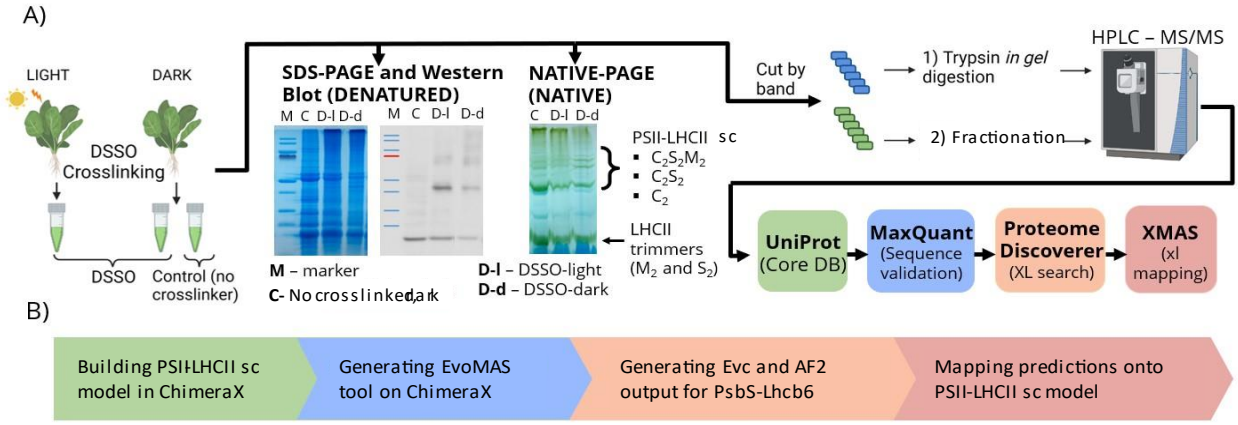


Figure 5. Experimental and computational approach used to unveil structural details of PsbS and its interactions with PSII-LHCII sc. A) Experimental approach. Thylakoid purification is followed by DSSO crosslinking and *in gel* separation of proteins under denatured (SDS-PAGE) and native conditions (lpBN-PAGE). Gel bands are cut, digested, fractionated, and run through HPLC-MS/MS. Then the raw data is analyzed using Thermo Proteome Discoverer (version 3.0.0.757) with XlinkX search nodes and the crosslinks are mapped onto 3D structure in ChimeraX using XMAS. **B) Computational approach.** EV coupling predictions are generated and mapped onto the structure using EvoMAS. AlphaFold 2 structural predictions are also generated and compared.

2.1. Thylakoid purification. The protocol for the thylakoid extraction was based on Caffarri *et al*. (2009). Thylakoid membranes were isolated from *Spinacia oleracea* leaves under cold (4ºC) and dark conditions and using buffers (see Supplementary Table 1) supplemented with divalent cations (Mg^{2+}) to mimic native chloroplast ionic conditions to preserve the stacked morphology of the grana membranes (Pagliano *et al.,* 2012; Schröppel-Meier *et al.,* 1988; Albanese *et al.,* 2020). Leaves were blended and homogenized for 3x6s at highest speed using buffer B1. Then, they were filtered through 4 layers of cheese cloth and centrifuged at 1400 x g for 10 min at 4ºC. The supernatant was discarded. The pellet was resuspended in buffer B2 and centrifuged at 4000 x g for 10 min at 4ºC. Next, the pellet was resuspended again in buffer B3 and centrifuged at 6000 x g for 10 min at 4ºC. Finally, the pellet containing the thylakoids was resuspended in buffer B4 to obtain the final purified thylakoids. The concentrations were of 1 mg/ml of chlorophyll (Chl) for NPQ induction samples and of 3 mg/ml Chl for the dark condition samples.

2.2. Optimization of crosslinking conditions. Crosslinking reaction conditions were optimized within the range of 0.25-4 mM for PhoX, 0.25-4 mM for disuccinimidyl sulfoxide (DSSO) and 2-40 mM for 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) on the dark-adapted purified thylakoids. No significant crosslinking was obtained with PhoX, so this crosslinker was not used. Furthermore, the EDC crosslinked samples were not included in this project due to temporal limitations. The optimal concentrations found for DSSO was 2mM (see Supplementary Figure 1). DSSO targets primary amines including lysine and amino termini of proteins and was considered as a "long-range" crosslinker which provides distance constraints of 0-35 Å. This distance constraint considers the overall protein flexibility (flexibility of the sidechains and the alpha-carbon backbone) (Leitner *et al.,* 2012; Albanese *et al.,* 2020).

2.3. NPQ induction and crosslinking of thylakoid membranes. The purified thylakoid samples were used to obtain both dark-adapted thylakoid samples and light-adapted thylakoid samples (NPQ is induced). For dark-adapted thylakoid samples, samples of 2µg total Chl (20 µg total protein) were crosslinked for 30 min on ice and in the dark with 2 mM DSSO. This was done for a final protein concentration of 1 mg/mL in a buffer of 0.1 M MES pH 6.0. A mildly acidic pH was used to preserve the macro-organization and functionality of the paired PSII-LHCII sc (Su *et al.,* 2017; Albanese *et al.,* 2017). The crosslinking reaction was quenched using 100 mM Tris buffer pH 8. For light-adapted thylakoid samples, the crosslinking quenching reaction was performed at the moment when NPQ occurs at its maximum. The protocol for NPQ induction was based on Bielczynski *et al*. (2022). For the NPQ induction, 50 µM methylviologen and 30 mM sodium ascorbate were added to 500 µl of freshly purified thylakoids (1 mg/ml Chl). Fluorescence measurements were performed using DualPAM-100. The intensities of the measuring light (ML), actinic light (AL) and saturating pulse (SP) were of 5, 531 and 2000 µmol of photons m^{-2} s^{-1}. The SP was 500 ms long. Samples were stirred during the NPQ induction and subjected to a SP to estimate the maximal fluorescence (F_M) from the dark-adapted sample, after a prior minimal fluorescence (F_0) from dark-adapted samples measurement. Right after the SP, AL was switched on, and over 10 min of illumination, each 2 min a SP was used to measure the F_M from light-adapted sample. The crosslinkers were added to the NPQ-induced sample and the sample was incubated at room temperature for 10 min in AL at 4 ºC. This was done without stirring to minimize induced conformational changes and unspecific aggregation. The reaction was stopped by adding 100 mM Tris buffer pH 8.

2.4. SDS-PAGE and lpBN-PAGE gel preparation. Aliquots of 20 µg of Chl (200 µg protein) of the DSSO crosslinked samples (dark-adapted thylakoids and NPQ-induced thylakoids samples) and a control (non-crosslinked dark-adapted thylakoids sample) were loaded onto SDS-PAGE and lpBN-PAGE gels (see Figures 7.a and 9.a). The protocol for SDS-PAGE was based on Schagger *et al.* (2006) and the protocol for lpBN-PAGE was based on Jarvi *et al.* (2011).

2.5. Trypsin in gel digestion. Each of the lanes, of both the SDS-PAGE and lpBN-PAGE gels, were cut into 7 slices (see Supplementary Figures 2 and 3). To remove compounds which could interfere with the LC-MS analysis, samples were washed with miliQ water and dehydrated for 15-min in 100% acetonitrile (ACN) (the volumes used were those large enough to cover the gels completely). Then, the ACN was removed, and proteins were reduced by incubating them in 6.5 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate (NH_4HCO_3) pH 8.5 for 1h at 60ºC under shaking conditions. The DTT was removed, and samples were dehydrated for 15-min in 100% ACN. The ACN was removed, and samples were alkylated by incubating them in 55 mM iodoacetamide (IAA) in 50 mM NH_4HCO_3 pH 8.5 in the dark for 30-min. The IAA was removed, and pieces were dehydrated ad rehydrated twice with 100% ACN and 50 mM NH_4HCO_3 pH 8.5 respectively, in intervals of 15-min, at room temperature. Samples were centrifuged at 15,000 × g for 15 min at 4 °C to remove the ACN completely before the digestion. Trypsin digestion was conducted by incubating samples at a 1:50 protein:protease ratio in 50 mM NH_4HCO_3 overnight at 37 °C. The supernatant was transferred to a new vial and dried down by vacuum centrifugation in a Savant SpeedVac Vacuum Concentrator (Thermo Scientific). Finally, the proteins were resuspended in the desired volume of 10% formic acid. The protocol was based on Tsaytler *et al.,* 2009.

2.6. Desalting and fractionation. Before the LC-MS analysis, samples needed to be processed further to remove potential compounds that could interfere with the LC-MS analysis. First, both crosslinked (DSS=-dark and DSSO-NPQ) and non-crosslinked (control) samples were desalted. For this, they were acidified to pH < 2 with 0.5% trifluoroacetic acid (TFA) and desalted using an Oasis HLB µElution 30 µm plate (Waters). Then the plate columns were washed twice with 150 µL of ACN and twice with 150 µL of 0.1% TFA. Acidified samples were loaded onto the column and then washed twice with 150 µl 0.1% TFA. Lastly, the peptides were eluted with a 35 µl 50% ACN / 0.1% TFA. The obtained eluates were evaporated by vacuum centrifugation in a Savant SpeedVac Vacuum Concentrator (Thermo Scientific). Next, only crosslinked samples were enriched by strong cation exchange (SCX) chromatography using the AssayMAP Bravo Platform with a protocol specifically optimized for the purpose of micro-fractionation of crosslinked peptides pairs. First, dried samples were resuspended in 100 µL 20% ACN/0.1% TFA using SCX cartridges in an AssayMAP Agilent Bravo robot (both Agilent Technologies). Then, fractions were retrieved using a series of consecutive elutions of 25 µL using acidifying buffers (20% ACN / 0.1% TFA, pH < 2) containing increasing concentrations of ammonium acetate. For the SDS-PAGE gel samples, three consecutive elutions with 250, 450 and 750 mM concentrations of ammonium bicarbonate were performed. For lpBN-PAGE, four consecutive elutions with 250, 400, 550, 750, 800 mM ammonium bicarbonate were performed (the 70 mM and 800 mM ammonium bicarbonate samples were combined). After the fractionation, fractions were desalted on an Oasis HLB Elution 30 µm plate as described previously. All fractions were evaporated again in a Savant SpeedVac Vacuum Concentrator and stored at -20°C prior to MS analysis. The protocol was based on the one of Lagerwaard *et al.,* 2022.

2.7. LC–MS/MS. For MS acquisition, dried samples were resuspended in 10% (v/v) formic acid in water accordingly so that the final injection volumes contained 200 ng of protein. Peptides were injected and separated on a 50 cm × 75 µm C18 analytical column, packed in-house (Poroshell 120 EC-C18/ 2.7 µm) connected to an Ultimate 3000 LC system and an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany). Peptide separation was achieved by a 55 min gradient for the controls and a 73 min gradient for the DSSO samples. For the controls, a standard acquisition method was used. For DSSO crosslinked samples, a variation of the Kolbowski *et al.,* 2022 acquisition method was used. Briefly, for the controls, a survey MS1 scan at 60,000 resolution and scan range of 375-1600 *m/z* was followed by a precursor charge selection of 2-6 for fragmentation using a stepped HCD collision energy mode (normalized energies of 28%). Full MS2 Orbitrap scans were acquired at a 15000 resolution. For the DSSO crosslinked samples, a survey MS1 scans were acquired at a 120000 resolution and scan range of 400-1450 *m/z*. This was followed by a precursor selection of a 3-6 charge state for fragmentation via stepped HCD collision energy mode with normalized energies of 19%, 25% and 31%. Then full MS2 Orbitrap scans were acquired at a 60000 resolution and a scan range of 200-2000 m/z.

2.8. RAW data Analysis. Raw data from both controls and crosslinked samples before fractionation were processed with MaxQuant (version 2.1.5.0.) using the UniProtKB repository for *S. oleracea* proteins (26-06-2022, 23924 sequences). A reduced database, which we will call the Core-DB (see Supplementary Figure 4), was generated using all possible *Arabidopsis thaliana* orthologue protein sequences in *S. oleracea* for PsbS and PSII-LHCII sc proteins. The presence of these proteins was validated with the MaxQuant results. The MaxQuant group-specific search settings were as follows: variable modifications—

oxidation (M), acetyl (protein N-termini), Phospho (STY); variable modifications—carbamidomethylation (C); label-free quantification (LFQ) — minimum ratio count of 2 and a classic normalization type; crosslinks— none. For the global-parameters: FDR—0.01; digestion— specific and Trypsin/P; match between runs— true. Moreover, no TD-MS was performed to get the full protein sequences due to temporal limitations. Therefore, an alternative approach was used to retrieve some additional information on the true length of the N-termini of the PsbS and LHCII proteins in the Core-DB. For this, a second MaxQuant run was performed using the N-termini semitriptic option and the Core-DB. The results of this run were used to generate a second database which we will call the Proteoform-DB and it includes both the Core-DB protein sequences and the N-termini variants for PsbS and LHCII proteins.

Next, crosslinked spectra were analyzed using Thermo Proteome Discoverer (version 3.0.0.757) software with incorporated XlinkX nodes, using DSSO specific settings. For peptide/protein identification Sequest HT was used. For the XlinkX detect, the acquisition strategies used was MS2. The NonCleavable (NC) option was not used due to temporal limitations. The XlinkX search settings for DSSO crosslinked samples were as follows: enzyme name—trypsin (full); maximum number of missed cleavages—3; minimum peptide length—5 amino acids. Precursor mass tolerance was set to 10 ppm and FTMS fragment mass to 20 ppm. Modifications allowed were carbamidomethylation of cysteines, as a fixed modification, and oxidation of methionines, protein N-terminal acetylation, and threonine phosphorylation as variable modifications. For the DSSO crosslinked samples the XlinkX was set to focus on K linkage. Crosslinks with a false discovery rate at 1% and score >39 were kept for further analysis. For both Sequest HT and XlinkX search, the reduced database called Proteoform-DB was used.

2.9. Structural modelling and crosslinking mapping. An available C_2S_2 high-resolution *S. oleracea* experimentally resolved structure (PDB: 3jcu; Wei *et al.,* 2016) was mapped onto the high-resolution $C_2S_2M_2$ *A. thaliana* structure (PDB: 7oui; Graca *et al.,* 2021) given that 3jcu is missing the structure of the LHCII M-trimers. Then, structural predictions, for each PSII-LHCII sc protein, were generated with AlphaFold version 2 (AF2; Hockenberry *et al.,* 2019; Evans *et al.,* 2021) using the monomer pipeline and the full amino acid sequences of the most abundant sequence variants in the Core-DB. AF2 predictions for C_2S_2 were mapped onto the *S. oleracea* structure and predictions for M_2 were mapped onto the *A. thaliana* structure. This was done given that the N-termini was not always resolved in the high-resolution experimental structures. N-termini were trimmed according to the XL-MS results, UniProt description and Target-P predictions (if the N-termini position was not found through the XL-MS data or was not described on UniProt). Additionally, a PsbS dimer model was predicted using the AF2 complex pipeline. Both these models were used to construct a PsbS-LHCII model. These models were generated using ChimeraX 1.4 (Pettersen *et al.,* 2020). Finally, circos-XL plots were used to visualize all crosslinks found and they were produced in R with "circlize" package (Gu *et al.,* 2014).

2.10. EvoMAS bundle development. Evolutionary couplings Mapping and AnalySis (EvoMAS; see Supplementary Figure 5) was implemented as a ChimeraX 1.4. bundle written in Python 3.9. It was packaged as a Python wheel according to the ChimeraX developer guide for Building and Distributing Bundles, which can be found at https://www.cgl.ucsf.edu/chimerax/docs/devel/writing_bundles.html. EvoMAS is the second version of the Xlink Mapping and AnalySis (XMAS) bundle developed by Lagerwaard

et al. (2022). It was developed as part of this project to be able to visualize and analyze both crosslinks and evolutionary links onto the 3D structure of proteins on ChimeraX.

Of the third-party Python packages pre-included in the ChimeraX environment, EvoMAS uses PyQt5 (for the graphical user interface (GUI)), NumPy (for fast processing of large datasets), Matplotlib (for generating plots) and Biopython (for sequence parsing). Additionally, Pandas (for handling tabular data format files), openpyxl (for importing Excel files), matplotlib-venn (for plotting Venn diagrams), seaborn (for plotting box and strip plots), and QtRangeSlider (to create slider widgets for defining a range of values) are employed. These external libraries are automatically installed in ChimeraX upon installation of EvoMAS when necessary.

EvoMAS code can be visualized and downloaded from https://bms-developer.science.uu.nl/isabel/chimerax_evomas. Instructions for manual installation are provided in the 'INSTALL.TXT' file included in the install package and a manual can be found under "MANUAL.PDF". To ensure support of all EvoMAS functionalities, ChimeraX version 1.4 or higher needs to be installed. Moreover, the XMAS bundle source code, including a user manual, is freely available at https://github.com/ScheltemaLab under an Apache License 2.0 software license.

2.11. Evolutionary predictions. To further analyze the PsbS interaction with PSII-LHCII sc, under both dark and NPQ conditions, evolutionary predictions were used to complement the crosslinking data. Both the software tools AF2 and EV couplings (Hockenberry *et al.,* 2019) were used. EV couplings was implemented to predict coevolving amino acids between PsbS and Lhcb6. These were then filtered to predictions with a probability greater or equal to 0.9 and were plotted onto the PSII-LHCII sc *S. oleracea* structural model using EvoMAS. Additionally, a set of predictions between PsbS and an increasing number of proteins belonging to the PSII-LHCII sc were generated using AF2 multimer.

3. RESULTS

3.1. PsbS has at least 3 N-termini proteoforms in S. oleracea. Due to temporal limitations, no TD-MS was performed to obtain the full protein sequences in our samples. Instead, a curated DB was generated using the *S. oleracea* sequences from UniprotKB. Given that PsbS is our protein of interest and that the PSII-LHCII sc thylakoid related proteins have a very variable N-termini (in length and PTMs), we analyzed the N-termini-specific crosslinks found for PsbS to find where the signal peptide ends and where its chain truly begins. Three PsbS proteoforms (PsbS#63, PsbS#68 and PsbS#70), with varying lengths in their N-termini, were found in *S. oleracea* (Figure 6). The UniProtKB annotation indicates that the PsbS chain for *S. oleracea* starts at position 70, but we found proteoforms beginning at positions 63 and 68. Furthermore, many of the PsbS crosslinks with other proteins (interlinks), which give us information on the interactions of PsbS, were found in this additional region. This information will be used to trim PsbS accordingly when making the molecular models.

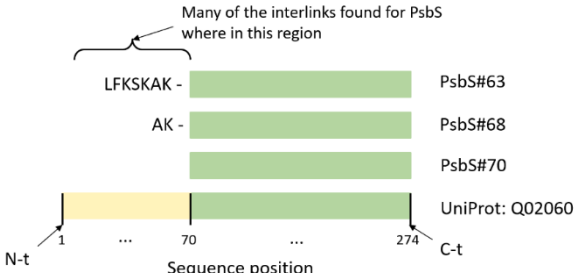


Figure 6. PsbS has at least 3 N-termini proteoforms in *S. oleracea*. Three proteoforms were found with varying lengths in the N-termini.

3.2. PsbS appears in its dimeric state both dark and NPQ conditions. One of the main aims of the project is to understand the conformation of PsbS throughout the dynamic process of NPQ. In Figures 7.a and 7.b. we can see a Coomassie stained SDS-PAGE gel and its western blot (using antibodies against PsbS) respectively. In the western blot, we find a band for both monomeric (22 KDa) and dimeric PsbS (44 KDa) for DSSO-crosslinked samples under dark and NPQ (to which will refer to as light conditions) conditions. Additionally, especially for the dark sample, we observe that PsbS is also found in bands > 44 KDa, but that the band pattern differs between both conditions. Therefore, we can say that PsbS appears as a dimer in *S. oleracea* under both dark and light conditions, but that the crosslinking pattern with other proteins (upper bands) varies between both conditions. Regarding the performance of the crosslinking reaction, we can observe in Figure 2.b that the reaction worked as there is no dimeric PsbS (44 KDa) found in the control lane. Moreover, in Figure 7.c, there is a histogram showing the total number of unique crosslinks for proteins in the CoreDB (PsbS and PSII-LHCII sc proteins for *S. oleracea*). A significant and similar number of crosslinks is found for both conditions (105 for DSSO-dark and 99 for DSSO-dark). Given that many PsbS and PSII-LHCII sc crosslinks have been found, we will proceed to further analyze them to see if we can get any additional details on the structure and dynamics of PsbS.

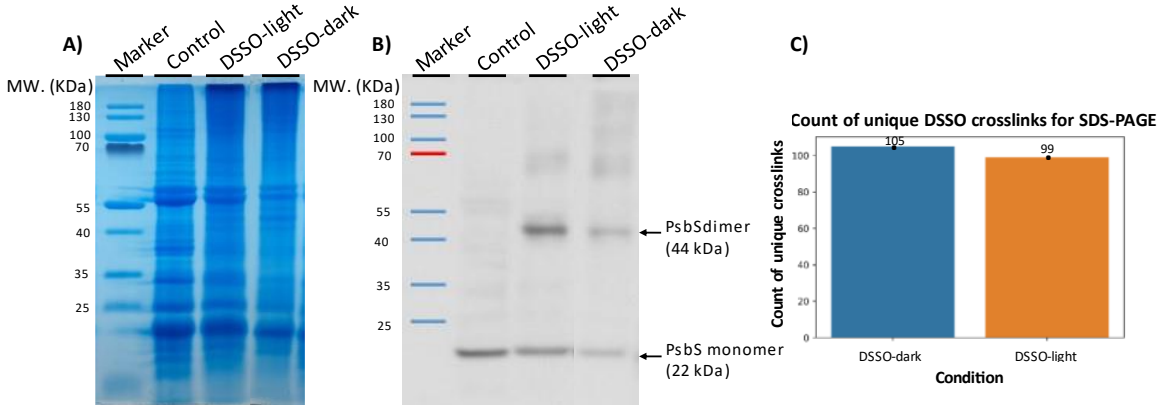


Figure 7. Dimeric PsbS appears under both light (NPQ) and dark conditions. A) SDS-PAGE gel (1.5 mm spacer; 4h 40 min; 80 V; Coomassie stained) for *S. oleracea* purified and DSSO crosslinked dark-adapted and NPQ-induced thylakoid samples (20 ug of Chl/sample). A non-crosslinked dark adapted thylakoid sample is also included as a control for crosslinking. **B) Western blot** of the SDS-PAGE gel using antibodies against *A. thaliana* PsbS (5s exposure time). Dimeric PsbS is found under both dark and light conditions. **C) Total count of unique** crosslinks for PsbS or PSII-LHCII sc related proteins in *S. oleracea*. A total 105 and 99 crosslinks were found for DSSO-dark and DSSO-light respectively.

3.3. PsbS has varying N-termini conformations under dark and light (NPQ) conditions. In this section, we will look into the crosslinks, specifically selflinks, to see if there is a difference in the conformation of the dimeric PsbS. Firstly, we will go through some key concepts (intralinks, interlinks and selflinks) with the purpose of making these results more understandable. In Figure 8.a, three types of crosslinks are plotted using the XMAS tool in ChimeraX: interlinks (links between two amino acids belonging to protein chains of two different proteins), intralinks (links between two amino acids within the same protein chain) and selflinks (links between two amino acids belonging to two different protein chains from the same protein). All the crosslinks from the Proteome Discoverer search classified as intralinks could in fact be either an intralinks or a selflinks. Furthermore, by looking at the crosslinked peptide sequences there are cases where you can verify if the crosslink is a selflink. These would be the cases in which the peptide sequences coincide or overlap (Figure 8.b). From now on, we will refer to the PsbS intralinks as potential selflinks (they could be both intralinks and selflinks) and we will search for the verifiable selflinks. In Figure 8.c we

can observe a histogram with the total potential and true unique selflinks found for PsbS. Two unique and verifiable selflinks are found under DSSO-dark conditions, but not under DSSO-light conditions. Furthermore, these selflinks were found in band 44 KDa and bands higher above this weight. This indicates that a PsbS dimeric formed was captured crosslinked to other proteins in the higher bands. Lastly, in Figure 8.d. we mapped all possible intralinks and selflinks onto the PsbS structure using XMAS. In green the verified selflinks are shown. Here, we can observe a difference in the possible potential selflinks for the N-termini of PsbS. For dark conditions, the N-termini seem to be closer together (green selflink joining the N-termini) and for light conditions, the N-termini seem to be closer to the thylakoid membrane. Now that we analyzed the denatured samples, we will try to capture more interactions under native conditions to find all possible PsbS crosslinks with the different conformations of the PSII-LHCII sc ($C_2S_2M_2$, C_2S_2M, C_2S_2 and loose M-trimers) which vary under dark and NPQ conditions.

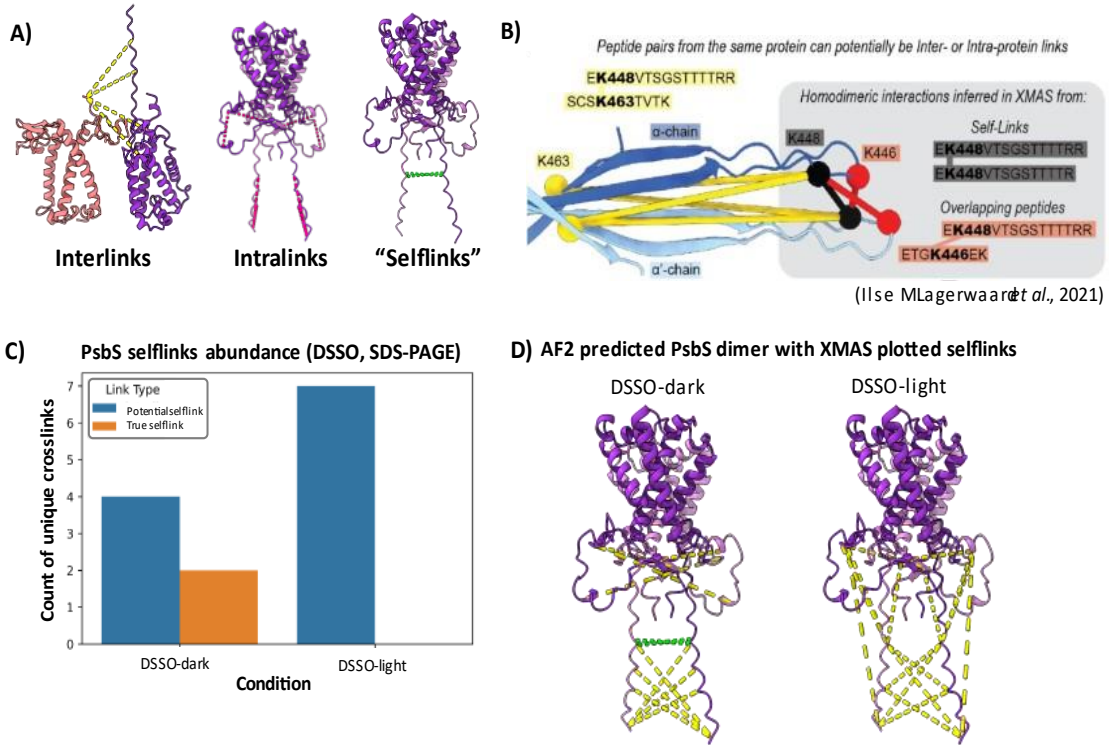


Figure 8. The N-termini of PsbS has a different conformation under dark and light (NPQ) conditions. **A)** Types of crosslinks: intralinks, interlinks or selflinks. **B)** Informative diagram to visualize how some selflinks can be verified**. C)** Histogram of potential and verified PsbS selflinks in dark vs light conditions. **D)** XMAS mapped intralinks and selflinks on AF2 predicted PsbS dimer.

3.4. PsbS and Lhcb6 interact with the LHCII trimers under NPQ conditions. We tried to capture PsbS interactions under native conditions with the different conformations of the dynamic PSII-LHCII sc ($C_2S_2M_2$, C_2S_2M, C_2S_2 and loose M-trimers) as the proportions of these complexes vary under dark and light conditions. In Figure 9.a. we can visualize the two lpBN-PAGE gels (native conditions) prepared to separate PSII-LHCII sc in its different conformations and capture its interactions with PsbS. In figure 9.b. there is a diagram to visualize the PSII-LHCII supercomplexes that we expect to separate. In the gels, we can locate the PSII-LHCII sc by their molecular weight and we can find a clear band for C_2 and another one for the LHCII trimers. Additionally, we can see that the DSSO-dark samples appear to be lighter. Lastly, in Figure 9.c., there is a histogram showing the total number of crosslinks found for DSSO-dark and DSSO-light conditions being these 11 and 81 respectively. Furthermore, of the 11 crosslinks found for DSSO-dark,

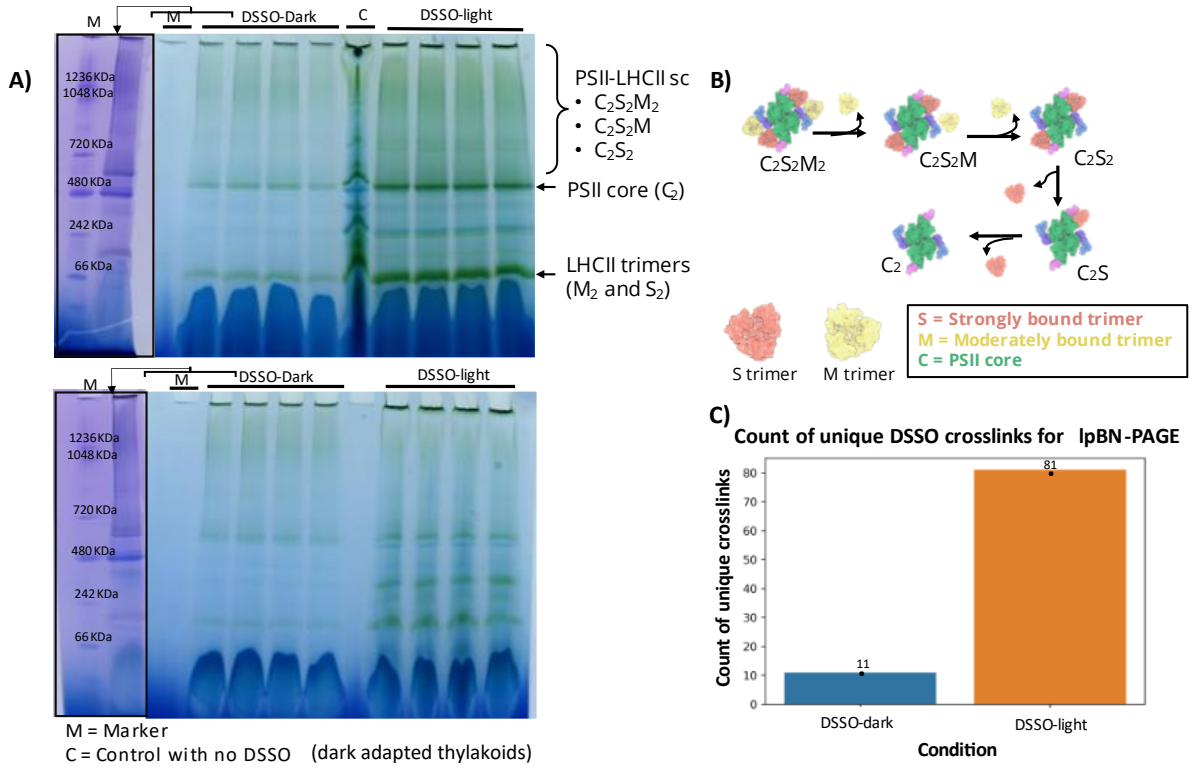


Figure 9. Separation of the PSII-LHCII supercomplexes and trimers of DSSO-crosslinked thylakoid samples under native conditions A) lpBN-PAGE gels. Each gel contains 4 replicates of each condition (DSSO-dark and DSSO-light) and a lane with the marker. The first gel also contains a lane with the control for crosslinking. Lastly to the left of each gel you can see the marker lane and the first lane stained with Coomassie. **B)** Figure illustrating the PSII-LHCII sc separation in the native gel. **C)** Count of total unique crosslinks found for each condition (DSSO-dark and DSSO-light).

none of them contained PsbS so nothing can be said for the interactions of PsbS in the dark. Therefore, we will only be able to analyze PsbS crosslinks under light conditions.

Under light conditions, we only find PsbS interlinks in the LHCII trimer band. Furthermore, most of these crosslinks are with the Lhcb isoforms presents in the trimers (Lhcb1 isoforms, Lhcb2, Lhcb3). In Figure 11.b. there is a circos plot with the PsbS interlinks found corresponding to the LHCII trimers. In Figure 11.c. there is a proposed model for the PsbS interaction with an LHCII trimer consisting of Lhcb2, Lhcb1.1 and Lhcb1.4 (the isoforms most crosslinked). Lastly, we can see that all the crosslinks found reside in the N-termini of the proteins.

In Figure 10 we can find two more circos plots that include all the crosslinks found for PsbS, Lhcb6 (monomeric antenna protein) and the LHCII trimer proteins (Lhcb1 isoforms, Lhcb2 and Lhcb3) in the lpBN-PAGE band corresponding to the LHCII trimers (left) and for the sum of all the bands (right). Here we see that we do not find crosslinks

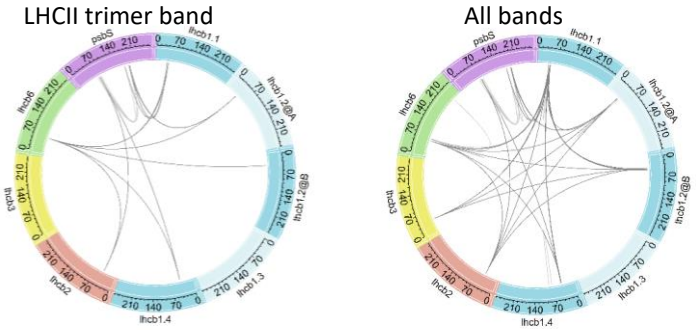


Figure 10. Crosslinks with the LHCII trimer proteins are found for both PsbS and Lhcb6. Circos plot containing all crosslinks found for the LHCII trimer proteins (Lhcb1 isoforms (blue), Lhcb2 (red), hcb3 (yellow)), the Lhcb6 (green) monomeric protein and PsbS (purple) in the trimer band and in the sum of all bands of the lpBN-PAGE gel under NPQ conditions.

between Lhcb6 and PsbS in the trimer band, but we do find crosslinks between Lhcb6 and the LHCII trimers. This suggest that both PsbS and LHCII leave with the LHII trimer in the light-induced dissociation during the NPQ process.

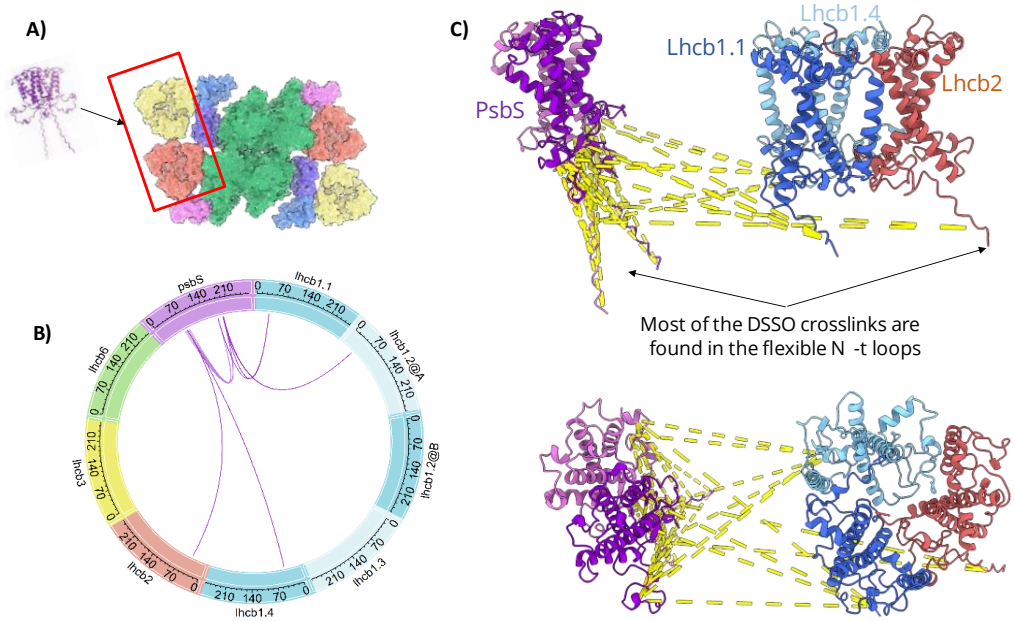


Figure 11. PsbS interacts with LHCII trimer proteins under light (NPQ) conditions. A) Diagram illustrating with which part of the PSII-LHCII sc PsbS interacts**. B)** Cricos plot of the crosslinks onto the full sequences of PsbS (purple), Lhcb1 isoforms (blue), Lhcb2 (red), Lhcb3 (yellow) and Lhcb6 (green)**. C)** PsbS crosslinks found mapped with XMAS onto trimer consisting of Lhcb1.1, Lhcb1.4 and Lhcb2 isoforms. All the crosslinks found are between the N-termini regions of the proteins.

3.5. Evolutionary predictions of the interactions of PsbS with Lhcb6. The EV couplings predictions of coevolving amino acids between PsbS and Lhcb6 (monomeric antenna protein with which PsbS is expected to interact under NPQ conditions) are shown in Figure 12.a. The coevolving amino acids are linked together with yellow bonds using EvoMAS. Only intralinks are predicted between Lhcb6 and PsbS meaning that no interacting amino acids between both proteins are found. Additionally, in the AF2 predictions (Figure 12.b), all the core proteins are predicted to nearly an experimental level (green and cyan) while PsbS and Lhcb6 are predicted to be far away from each other and not interacting with the core. The evolutionary predictions did not to work for this particular application.

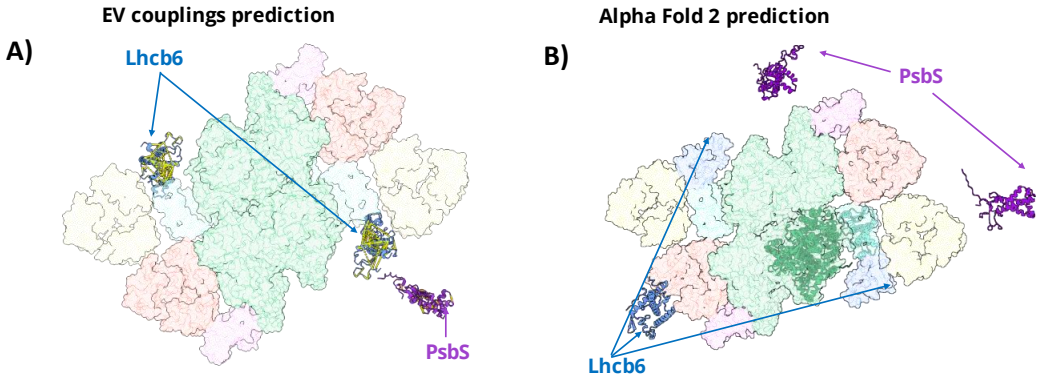


Figure 12. Evolutionary predictions of the interactions of PsbS with Lhcb6 and the PSII-LHCII sc were not successful. A) EV couplings coevolving amino acids prediction between PsbS and Lhcb6 shown in yellow as links using EvoMAS. **B)** AlphaFold 2 multimeric structure prediction of PsbS (purple), Lhcb6 (blue), Lhcb4 (cyan) and a series of proteins in the core including D1 (green). Both predictions are mapped onto the *A. thaliana* PSII-LHCII sc model (PDB: 7oui; Graca *et al.,* 2021)**.**

4. DISCUSSION

Nonphotochemical quenching is a rapid occurring process in which the PSII-LHCII sc undergoes remodeling of its supramolecular organization to dissipate excess light energy during photosynthesis. It is clear that PsbS is an essential protein for this process, but its behavior is yet not fully understood. Additionally, gaining more structural information on PsbS and its interactions is of importance as qE could be a key point of interest in crop engineering for the optimization of light-use efficiency in higher order plants (Glowacka *et al.,* 2018). Furthermore, a greater understanding of PsbS and NPQ in higher order plants could later be translated into a better comprehension of the functioning of qE in other oxygenic photosynthetic organisms such as algae and mosses (Alboresi *et al*., 2010; Li X-P *et al.*, 2000; Peers *et al.,* 2009). Therefore, structural models for the dynamics and interactions of PsbS are needed to gain mechanistic insights into the driving factors behind NPQ (specifically qE).

In this project, we explored the combination of a crosslinking proteomics mass spectrometry approach together with evolutionary predictions. Even though not all results were successful, key information has been able to be extracted about PsbS in *S. oleracea.* Firstly, regarding the structure of PsbS, we have found its dimeric form to be present under both under dark and light conditions. Nevertheless, the crosslinks obtained under each condition varied leading us to believe that the conformation of the dimer may vary. A pair of true selflinks were found under dark conditions that were not present under light conditions. Furthermore, the potential selflinks found under both conditions also differed and when mapping them onto the dimeric structure and visually it seems like the N-termini adopts a different conformation. A proposed model could be that under dark conditions, in its inactive state, the N-termini of the PsbS dimer are closer together (being able to obtain the true selflinks found). However, once NPQ begins, the change in pH in the lumen induces a conformational change of PsbS making the N-termini move towards the thylakoid membrane. Additionally, it is relevant to remember that, before NPQ conditions, the grana are stacked and the PSII-LHCII sc from adjacent thylakoid membranes are forming a sandwich structure stabilized through the interaction of the N-termini of the proteins in the macromolecular structures. Under NPQ conditions, the PSII-LHCII sc, forming sandwich structures in the grana, unstack and the conformations of N-termini shift (Albanese *et al.,* 2020). Going back to PsbS, it could be that the shift of its N-termini towards the thylakoid membrane could facilitate the beginning of its interaction with the pertinent PSII-LHCII sc proteins. Lastly, the group of Liguori *et al.* (2019) reached similar conclusions, regarding the dynamics of the N-termini of PsbS, through computational simulations of its behavior at pH = 8 (dark conditions) and pH= 5 (NPQ conditions).

Moving on to the interactions of PsbS, we were not able to conclude much for its behavior under dark conditions. Barely any crosslinks were acquired for the dark-adapted thylakoid sample under native conditions and no crosslinks were acquired specifically for PsbS. We believe there was a problem in the solubilization of the thylakoid membranes when preparing the sample. However, it would be of great interest to repeat this experiment for this condition. Nevertheless, several crosslinks were extracted for PsbS under NPQ conditions that match with the proposed model proposed in the review of Bassi *et al*. (2021, Fig. 5). The crosslinking data points towards PsbS interacting with LHCII under NPQ conditions. Also, as expected, no crosslinks were found of PsbS with the PSII core. Once proper results for dark conditions are acquired, it would be of interest to look at if there is a difference in to where the PsbS

crosslinks appear (upper PSII-LHCII sc bands vs lower LHCII trimer bands in the lpBN-PAGE gel) and if there is an increase is a significant PsbS-LHCII enrichment in crosslinks in the LHCII trimer band of the lpBN-PAGE gel under NPQ conditions. Moreover, we also analyzed crosslinks containing Lhcb6 as in the proposed model of Bassi *et al.* (2021) it is suggested that Lhcb6 migrates together with PsbS and LHCII. For Lhcb6, we get no direct crosslinks with PsbS, but we do with LHCII and they are mostly located in the LHCII trimer band of the lpBN-PAGE gel. All thing considered, our experimental data seems to go hand in hand with the proposed model mentioned in Figure 3.

Lastly, the computational approaches (evolutionary predictions with EV couplings and AF2) did not prove useful for finding the interactions of PsbS relevant to this project. The evolutionary predictions did not align with our experimental results and those acquired by other groups concerning the behaviors of PsbS and Lhcb6. We believe that this is due to the nature of our specific application and not so much the tools themselves. PsbS and Lhcb6 are both evolutionary quite recent (Ballottari*, et al.* 2012). Moreover, both EV couplings and AF2 are based in extracting structural information from evolutionary information. They both use multiple sequence alignments to extract evolutionary information and given that both proteins are quite recent this is probably why both applications are not working well. Nevertheless, the combination of crosslinking mass spectrometry proteomics approach with the co-evolution amino acid links could be quite useful in other applications where the proteins are not that recent (in terms of evolution).

Overall, this project has been developed to get some insights into the structure and dynamics of PsbS and to see if it is relevant to continue pursuing this research question. Interesting results have already been obtained and with some improvements in the method used the continuation of this project looks promising. Some steps to take in the future would be to use biological replicates and incorporate the use of other crosslinkers that provide different distance constraints such as EDC. As explained in the methods section, EDC results were not included in this report due to temporal limitations but could definitely give further insights. It would also be interesting to explore alternative methods of structure separation such as the usage of a sucrose gradient that has proven to be useful in other similar applications (Albanese *et al.,* 2020). Also, PsbS in *Spinacia oleracea* has lots of Lys in its N-termini making it difficult for the software to identify all crosslinks. So, it might be of interest exploring this with another model organism which has an N-termini in PsbS with less Lys. Moreover, there is a clear limitation with crosslinking for this application and it is the inaccessibility of the crosslinker to the portions of the protein inside the membrane. Consequently, other methods, such as Cryo-EM, that can be combined with XL-MS are an important aspect to explore. Lastly, in this project, again due to temporal limitations, TD-MS was not performed. If the project were to be continued and on the methodology is optimized, it would be very useful to perform TD-MS to acquire the full sequences of the proteins in the PSII-LHCII sc to be able to build more accurate 3D structural models. With the combination of new distance restraints (through the addition of new crosslinkers) and the acquisition of the full lengths of PsbS and the PSII-LhCII sc of the organism of choice (TD-MS) it will be possible to build an even more complete model. Furthermore, these new distance constraints and models could be used to explore alternative computational approaches such as protein docking given that evolutionary predictions were not successful.

In conclusion, it is important to remember that the NPQ process is not static, and that we must think in terms of dynamics when trying to understand the function and structure of the PsbS. In this project, some relevant insights of PsbS have been extracted. Nevertheless, if we want to even attempt to alter this process in the future for crop engineering, more steps should be taken to further understand the role of PsbS in NPQ.

5. CONCLUSIONS

The aim of this project was to gain a greater insight into the structure and conformation of PsbS and its dynamic interactions with the PSII-LHCII sc during NPQ. To achieve this, we analyzed crosslinked thylakoid samples under two different conditions: dark and light (NPQ-induced). From the results we find that (1) PsbS has, at least, 3 isoforms (#70, #68 and #63), in *S. oleracea,* with varying lengths in its amino termini. Additionally, (2) PsbS acts as a dimer under both dark and NPQ conditions. Moreover, the PsbS dimer has a variable N-termini conformation under dark and NPQ conditions. Under dark conditions the N-termini are closer together and under NPQ conditions the N-termini are further apart from each other and closer to the thylakoidal membrane. Furthermore, (3) PsbS and Lhcb6 bind to the LHCII trimers under light conditions. Also, no interactions are found with the PSII core and PsbS under NPQ conditions. Lastly, (4) the evolutionary predictions approach did not succeed for this particular application.

6. REFERENCES

1. Albanese, Pascal, et al. “How Paired PSII–LHCII Supercomplexes Mediate the Stacking of Plant Thylakoid Membranes Unveiled by Structural Mass-Spectrometry.” *Nature Communications*, vol. 11, no. 1, 2020, https://doi.org/10.1038/s41467-020-15184-1.
2. Albanese, Pascal, et al. “PEA PSII-LHCII Supercomplexes Form Pairs by Making Connections across the Stromal Gap.” *Scientific Reports*, vol. 7, no. 1, 2017, https://doi.org/10.1038/s41598-017-10700-8.
3. Albanese, Pascal, et al. “Dynamic Reorganization of Photosystem II Supercomplexes in Response to Variations in Light Intensities.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1857, no. 10, 2016, pp. 1651–1660., https://doi.org/10.1016/j.bbabio.2016.06.011.
4. Alboresi, Alessandro, et al. “*Physcomitrella Patens* Mutants Affected on Heat Dissipation Clarify the Evolution of Photoprotection Mechanisms upon Land Colonization.” *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, 2010, pp. 11128–11133., https://doi.org/10.1073/pnas.1002873107.
5. Ballottari, Matteo, et al. “Evolution and Functional Properties of Photosystem II Light Harvesting Complexes in Eukaryotes.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1817, no. 1, 2012, pp. 143–157., https://doi.org/10.1016/j.bbabio.2011.06.005.
6. Bassi, Roberto, and Luca Dall'Osto. “Dissipation of Light Energy Absorbed in Excess: The Molecular Mechanisms.” *Annual Review of Plant Biology*, vol. 72, no. 1, 2021, pp. 47–76., https://doi.org/10.1146/annurev-arplant-071720-015522.
7. Betterle, Nico, et al. “Light-Induced Dissociation of an Antenna Hetero-Oligomer Is Needed for Non-Photochemical Quenching Induction.” *Journal of Biological Chemistry*, vol. 284, no. 22, 2009, pp. 15255–15266., https://doi.org/10.1074/jbc.m808625200.
8. Björkman, O., and B. Demmig-Adams. “Regulation of Photosynthetic Light Energy Capture, Conversion, and Dissipation in Leaves of Higher Plants.” *Ecophysiology of Photosynthesis*, 1995, pp. 17–47., https://doi.org/10.1007/978-3-642-79354-7_2.
9. Bonente, Giulia, et al. “Interactions between the Photosystem II Subunit Psbs and Xanthophylls Studied in Vivo and in Vitro.” *Journal of Biological Chemistry*, vol. 283, no. 13, 2008, pp. 8434–8445., https://doi.org/10.1074/jbc.m708291200.
10. Caffarri, Stefano, et al. “Functional Architecture of Higher Plant Photosystem II Supercomplexes.” *The EMBO Journal*, vol. 28, no. 19, 2009, pp. 3052–3063., https://doi.org/10.1038/emboj.2009.232.
11. Correa-Galvis, Viviana, et al. “PSBs Interactions Involved in the Activation of Energy Dissipation in Arabidopsis.” *Nature Plants*, vol. 2, no. 2, 2016, https://doi.org/10.1038/nplants.2015.225.
12. Dekker, Jan P., and Egbert J. Boekema. “Supramolecular Organization of Thylakoid Membrane Proteins in Green Plants.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1706, no. 1-2, 2005, pp. 12–39., https://doi.org/10.1016/j.bbabio.2004.09.009.
13. Demmig-Adams, Barbara. “Carotenoids and Photoprotection in Plants: A Role for the Xanthophyll Zeaxanthin.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1020, no. 1, 1990, pp. 1–24., https://doi.org/10.1016/0005-2728(90)90088-l.
14. Evans, Richard, et al. “Protein Complex Prediction with Alphafold-Multimer.” 2021, https://doi.org/10.1101/2021.10.04.463034.

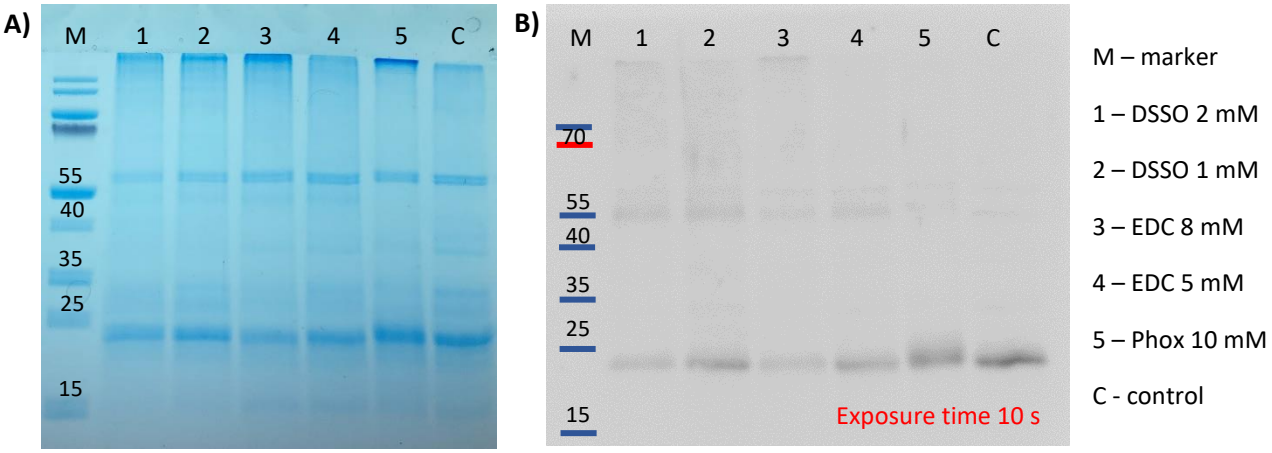
15. Funk, Christiane, et al. “The Nuclear-Encoded Chlorophyll-Binding Photosystem II-S Protein Is Stable in the Absence of Pigments.” *Journal of Biological Chemistry*, vol. 270, no. 50, 1995, pp. 30141–30147., https://doi.org/10.1074/jbc.270.50.30141.
16. Gerotto, Caterina, et al. “In Vivo Identification of Photosystem II Light Harvesting Complexes Interacting with Photosystem II Subunit S.” *Plant Physiology*, vol. 168, no. 4, 2015, pp. 1747–1761., https://doi.org/10.1104/pp.15.00361.
17. Głowacka, Katarzyna, et al. “Photosystem II Subunit s Overexpression Increases the Efficiency of Water Use in a Field-Grown Crop.” *Nature Communications*, vol. 9, no. 1, 2018, https://doi.org/10.1038/s41467-018-03231-x.
18. Goddard, Thomas D., et al. “UCSF Chimerax: Meeting Modern Challenges in Visualization and Analysis.” *Protein Science*, vol. 27, no. 1, 2017, pp. 14–25., https://doi.org/10.1002/pro.3235.
19. Graca, A.T., et al. “Structure of c2s2m2-Type Photosystem Supercomplex from Arabidopsis Thaliana (Digitonin-Extracted).” 2021, https://doi.org/10.2210/pdb7oui/pdb.
20. Gu, Zuguang, et al. “Circlize Implements and Enhances Circular Visualization in R.” *Bioinformatics*, vol. 30, no. 19, 2014, pp. 2811–2812., https://doi.org/10.1093/bioinformatics/btu393.
21. Hockenberry, Adam J., and Claus O. Wilke. “Evolutionary Couplings Detect Side-Chain Interactions.” *PeerJ*, vol. 7, 2019, https://doi.org/10.7717/peerj.7280.
22. Hohmann-Marriott, Martin F., and Robert E. Blankenship. “Evolution of Photosynthesis.” *Annual Review of Plant Biology*, vol. 62, no. 1, 2011, pp. 515–548., https://doi.org/10.1146/annurev-arplant-042110-103811.
23. Holzwarth, Alfred R., et al. “Identification of Two Quenching Sites Active in the Regulation of Photosynthetic Light-Harvesting Studied by Time-Resolved Fluorescence.” *Chemical Physics Letters*, vol. 483, no. 4-6, 2009, pp. 262–267., https://doi.org/10.1016/j.cplett.2009.10.085.
24. Horton, P., et al. “Regulation of Light Harvesting in Green Plants.” *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 47, no. 1, 1996, pp. 655–684., https://doi.org/10.1146/annurev.arplant.47.1.655.
25. Järvi, Sari, et al. “Optimized Native Gel Systems for Separation of Thylakoid Protein Complexes: Novel Super- and Mega-Complexes.” *Biochemical Journal*, vol. 439, no. 2, 2011, pp. 207–214., https://doi.org/10.1042/bj20102155.
26. Johnson, Matthew P., et al. “Photoprotective Energy Dissipation Involves the Reorganization of Photosystem II Light-Harvesting Complexes in the Grana Membranes of Spinach Chloroplasts.” *The Plant Cell*, vol. 23, no. 4, 2011, pp. 1468–1479., https://doi.org/10.1105/tpc.110.081646.
27. Jumper, John, et al. “Highly Accurate Protein Structure Prediction with Alphafold.” *Nature*, vol. 596, no. 7873, 2021, pp. 583–589., https://doi.org/10.1038/s41586-021-03819-2.
28. Kereïche, Sami, et al. “The Psbs Protein Controls the Macro-Organisation of Photosystem II Complexes in the Grana Membranes of Higher Plant Chloroplasts.” *FEBS Letters*, vol. 584, no. 4, 2009, pp. 759–764., https://doi.org/10.1016/j.febslet.2009.12.031.
29. Kereïche, Sami, et al. “The Psbs Protein Controls the Macro-Organisation of Photosystem II Complexes in the Grana Membranes of Higher Plant Chloroplasts.” *FEBS Letters*, vol. 584, no. 4, 2009, pp. 759–764., https://doi.org/10.1016/j.febslet.2009.12.031.
30. Kiss, Anett Z., et al. “The Psbs Protein Controls the Organization of the Photosystem II Antenna in Higher Plant Thylakoid Membranes.” *Journal of Biological Chemistry*, vol. 283, no. 7, 2008, pp. 3972–3978., https://doi.org/10.1074/jbc.m707410200.
31. Koochak, Haniyeh, et al. “The Structural and Functional Domains of Plant Thylakoid Membranes.” *The Plant Journal*, vol. 97, no. 3, 2018, pp. 412–429., https://doi.org/10.1111/tpj.14127.
32. Kolbowski, Lars, et al. “Improved Peptide Backbone Fragmentation Is the Primary Advantage of MS-Cleavable Crosslinkers.” *Analytical Chemistry*, vol. 94, no. 22, 2022, pp. 7779–7786., https://doi.org/10.1021/acs.analchem.1c05266.
33. Kouřil, Roman, et al. “High-Light vs. Low-Light: Effect of Light Acclimation on Photosystem II Composition and Organization in Arabidopsis Thaliana.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1827, no. 3, 2013, pp. 411–419., https://doi.org/10.1016/j.bbabio.2012.12.003.
34. Koziol, Adam G., et al. “Tracing the Evolution of the Light-Harvesting Antennae in Chlorophyll *a*/*b*-Containing Organisms.” *Plant Physiology*, vol. 143, no. 4, 2007, pp. 1802–1816., https://doi.org/10.1104/pp.106.092536.
35. Lagerwaard, Ilse M., et al. “Xlink Mapping and Analysis (XMAS) - Smooth Integrative Modeling in Chimerax.” 2022, https://doi.org/10.1101/2022.04.21.489026.
36. Leitner, Alexander, et al. “The Molecular Architecture of the Eukaryotic Chaperonin TRIC/CCT.” *Structure*, vol. 20, no. 5, 2012, pp. 814–825., https://doi.org/10.1016/j.str.2012.03.007.
37. Li, Xiao-Ping, et al. “A Pigment-Binding Protein Essential for Regulation of Photosynthetic Light Harvesting.” *Nature*, vol. 403, no. 6768, 2000, pp. 391–395., https://doi.org/10.1038/35000131.
38. Liguori, Nicoletta, et al. “Molecular Anatomy of Plant Photoprotective Switches: The Sensitivity of Psbs to the Environment, Residue by Residue.” *The Journal of Physical Chemistry Letters*, vol. 10, no. 8, 2019, pp. 1737–1742., https://doi.org/10.1021/acs.jpclett.9b00437.
39. Long, Stephen P., et al. “Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential.” *Cell*, vol. 161, no. 1, 2015, pp. 56–66., https://doi.org/10.1016/j.cell.2015.03.019.
40. Müller Patricia, et al. “Non-Photochemical Quenching. A Response to Excess Light Energy.” *Plant Physiology*, vol. 125, no. 4, 2001, pp. 1558–1566., https://doi.org/10.1104/pp.125.4.1558.
41. Nelson, Nathan, and Adam Ben-Shem. “The Complex Architecture of Oxygenic Photosynthesis.” *Nature Reviews Molecular Cell Biology*, vol. 5, no. 12, 2004, pp. 971–982., https://doi.org/10.1038/nrm1525.
42. Nicol, Lauren, et al. “Disentangling the Sites of Non-Photochemical Quenching in Vascular Plants.” *Nature Plants*, vol. 5, no. 11, 2019, pp. 1177–1183., https://doi.org/10.1038/s41477-019-0526-5.
43. Ort, Donald R., et al. “Redesigning Photosynthesis to Sustainably Meet Global Food and Bioenergy Demand.” *Proceedings of the National Academy of Sciences*, vol. 112, no. 28, 2015, pp. 8529–8536., https://doi.org/10.1073/pnas.1424031112.

44. Pagliano, Cristina, et al. “Comparison of the α and β Isomeric Forms of the Detergent N-Dodecyl-D-Maltoside for Solubilizing Photosynthetic Complexes from Pea Thylakoid Membranes.” *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, vol. 1817, no. 8, 2012, pp. 1506–1515., https://doi.org/10.1016/j.bbabio.2011.11.001.
45. Pagliano, Cristina, et al. “Structural, Functional and Auxiliary Proteins of Photosystem II.” *Photosynthesis Research*, vol. 116, no. 2-3, 2013, pp. 167–188., https://doi.org/10.1007/s11120-013-9803-8.
46. Peers, Graham, et al. “An Ancient Light-Harvesting Protein Is Critical for the Regulation of Algal Photosynthesis.” *Nature*, vol. 462, no. 7272, 2009, pp. 518–521., https://doi.org/10.1038/nature08587.
47. Pettersen, Eric F., et al. “UCSF Chimerax : Structure Visualization for Researchers, Educators, and Developers.” *Protein Science*, vol. 30, no. 1, 2020, pp. 70–82., https://doi.org/10.1002/pro.3943.
48. Raven, John A. “The Cost of Photoinhibition.” *Physiologia Plantarum*, vol. 142, no. 1, 2011, pp. 87–104., https://doi.org/10.1111/j.1399-3054.2011.01465.x.
49. Saccon, Francesco, et al. “Rapid Regulation of Photosynthetic Light Harvesting in the Absence of Minor Antenna and Reaction Centre Complexes.” *Journal of Experimental Botany*, vol. 71, no. 12, 2020, pp. 3626–3637., https://doi.org/10.1093/jxb/eraa126.
50. Schägger, Hermann. “Tricine–SDS-PAGE.” *Nature Protocols*, vol. 1, no. 1, 2006, pp. 16–22., https://doi.org/10.1038/nprot.2006.4.
51. Schröppel-Meier, Gabriele, and Werner M. Kaiser. “Ion Homeostasis in Chloroplasts under Salinity and Mineral Deficiency.” *Plant Physiology*, vol. 87, no. 4, 1988, pp. 828–832., https://doi.org/10.1104/pp.87.4.828.
52. Schweppe, Devin K., et al. “Mitochondrial Protein Interactome Elucidated by Chemical Cross-Linking Mass Spectrometry.” *Proceedings of the National Academy of Sciences*, vol. 114, no. 7, 2017, pp. 1732–1737., https://doi.org/10.1073/pnas.1617220114.
53. Su, Xiaodong, et al. “Structure and Assembly Mechanism of Plant $C_2S_2M_2$-Type PSII-LHCII Supercomplex.” *Science*, vol. 357, no. 6353, 2017, pp. 815–820., https://doi.org/10.1126/science.aan0327.
54. Erling Tjus, Staffan, et al. “Photosystem I Is an Early Target of Photoinhibition in Barley Illuminated at Chilling temperatures1.” *Plant Physiology*, vol. 116, no. 2, 1998, pp. 755–764., https://doi.org/10.1104/pp.116.2.755.
55. Tjus, Staffan Erling, et al. “Active Oxygen Produced during Selective Excitation of Photosystem I Is Damaging Not Only to Photosystem I, but Also to Photosystem II.” *Plant Physiology*, vol. 125, no. 4, 2001, pp. 2007–2015., https://doi.org/10.1104/pp.125.4.2007.
56. Tsaytler, Pavel A., et al. “Novel hsp90 Partners Discovered Using Complementary Proteomic Approaches.” *Cell Stress and Chaperones*, vol. 14, no. 6, 2009, pp. 629–638., https://doi.org/10.1007/s12192-009-0115-z.
57. Walters, Robin G., et al. “Higher Plant Light-Harvesting Complexes Lhciia and Lhciic Are Bound by Dicyclohexylcarbodiimide during Inhibition of Energy Dissipation.” *European Journal of Biochemistry*, vol. 226, no. 3, 1994, pp. 1063–1069., https://doi.org/10.1111/j.1432-1033.1994.01063.x.
58. Wei, X.P., et al. “Cryo-EM Structure of Spinach Psii-LHCII Supercomplex at 3.2 Angstrom Resolution.” 2016, https://doi.org/10.2210/pdb3jcu/pdb.

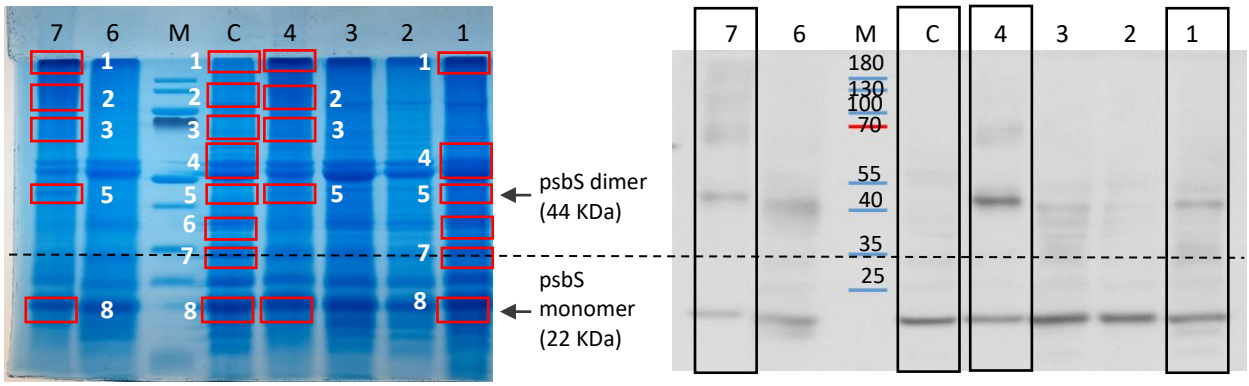
Supplementary Information

Supplementary Table 1. Compositions of buffers used for thylakoid purification.

Buffer Name	Component	Concentration (mM)
B1	Sorbitol	400
** Last two are protease inhibitors*	$MgCl_2$	5
and not strictly required.	tricine/KOH pH 7.8	20
	EDTA	5
	benzamidine	0.2
	ε-aminocaproic acid	1
B2	Sorbitol	150
	$MgCl_2$	5
	tricine/KOH pH 7.8	20
	EDTA	2.5
	benzamidine	0.2
	ε-aminocaproic acid	1
B3	$MgCl_2$	5
	NaCl	15
	Hepes/KOH pH 7.5	20
B4	Sorbitol	400
	$MgCl_2$	5
	NaCl	15
	Hepes/KOH pH 7.5	10

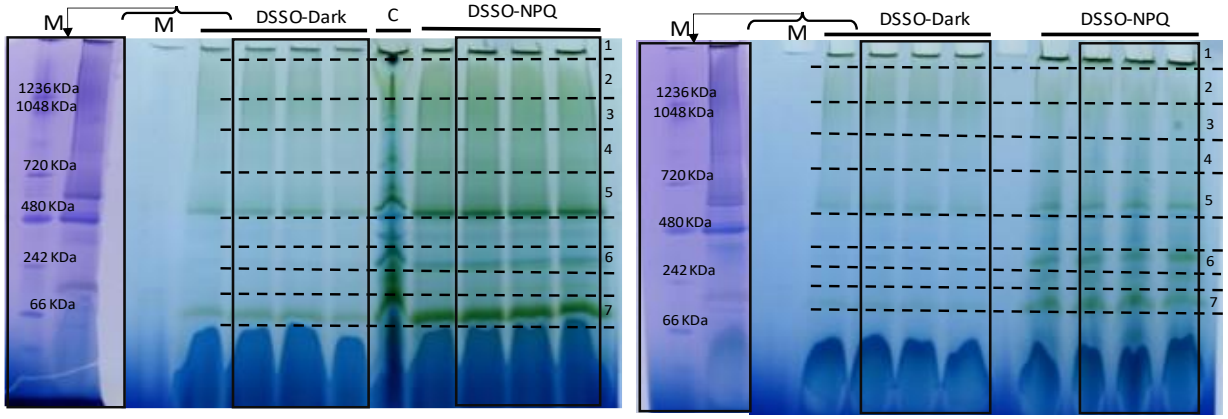


Supplementary Figure 1. Crosslinker concentration optimization. Dark-adapted thylakoid samples (c=2 ug total Chl) were crosslinked with different concentrations of PhoX, EDC and DSSO and 20 uL of crosslinked sample was loaded onto each well. **A) SDS-PAGE gel** with Coomassie staining. **B) Western Blot** using antibodies for *Arabidopsis thaliana* PsbS.



C – control
M – marker
1 – NPQ 8 mM EDC low pH
2 – X – wrong sample
3 – NPQ 8 mM neutral pH
4 – NPQ 2 mM DSSO
6 – dark 8 mM EDC low pH
7 – dark 2 mM DSSO

Supplementary Figure 2. SDS-PAGE gel band cutting. To the right we have the SDS-PAGE gel with the selected bands in red boxes. To the left we have the western blot for the SDS-PAGE gel using antibodies for Arabidopsis thaliana PsbS.



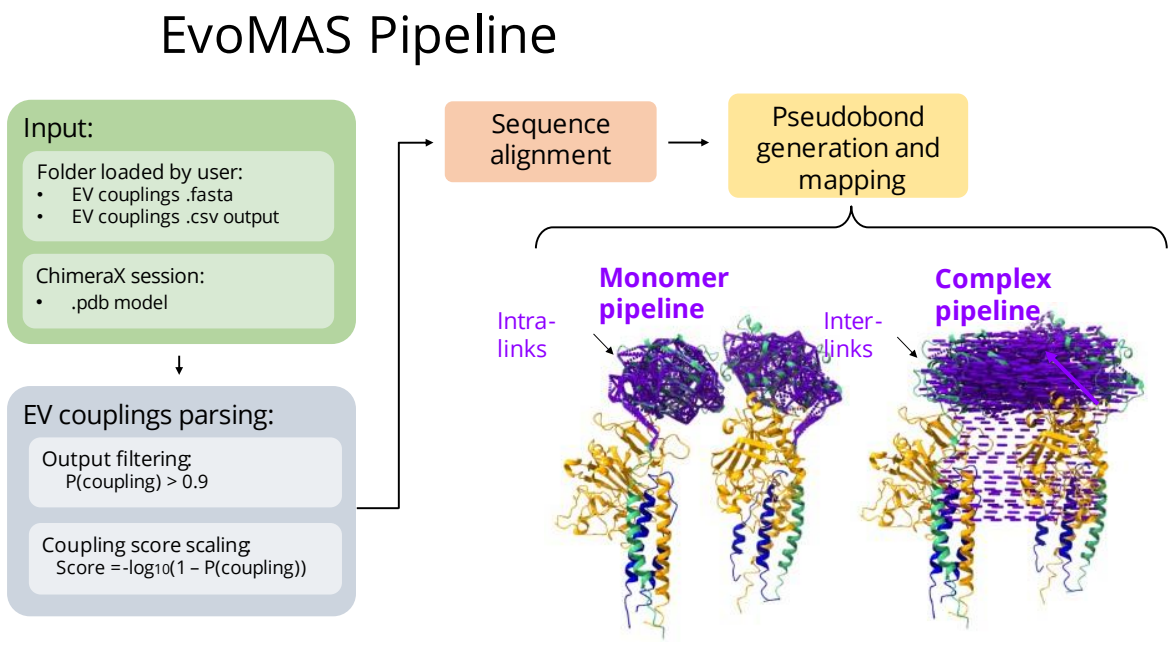
M = Marker
C = Control with no DSSO (dark-adapted thylakoids)

Supplementary Figure 3. lpBN-PAGE gel band cutting. Two native gels were prepared. Each containing 4 replicates of DSSO 2 mM and 4 replicates of DSSO 2 mM NPQ samples. The gel to the right also includes a lane with the control (no crosslinking, dark-adapted thylakoids). To the left of each lane, you can see the Coomassie stained gel for the marker and the first lane. The other lanes were not stained to avid interfering with the native conditions and were used for the future steps of the method. Lastly, 3 lanes for each sample (DSSO-dark and DSSO-NPQ) were selected from each band and the 6 lanes of each sample were combined by band for the trypsin digestion.

Spinacia oleracea Core Database

```
psii_lhcbii_prots = {
        'Q02060' : 'psbS',
        'A4F3R4' : 'lhcb1.1',
        'A4F3R3' : 'lhcb1.2@A', 'A0A0K9RBU9' : 'Lhcb1.2@B',
        'B3WFZ6' : 'lhcb1.3',
        'D2T2F7' : 'lhcb1.4',
        'A0A0K9QLM7' : 'lhcb2',
        'A0A0K9QIA4' : 'lhcb3',
        'A0A0K9RW58' : 'lhcb4@A',
        'A0A0K9QLX6' : 'lhcb4@B',
        'W6JKE4' : 'lhcb5@A',
        'A0A0K9QUQ7' : 'lhcb5@B',
        'P36494' : 'lhcb6',
        'P69560' : 'psbA/D1',
        'P04160' : 'psbB/CP47',
        'P06003' : 'psbC/CP43@A',
        'Q6EY72' : 'psbC/CP43@B',
        'P06005' : 'psbD/D2',
        'P69383' : 'psbE/cytb559A@A',
        'Q6EYP6' : 'psbE/cytb559A@B',
        'P60128' : 'psbF/cytb559B',
        'P05146' : 'psbH',
        'P62103' : 'psbI',
        'P12163' : 'psbK',
        'P60150' : 'ppsbL',
        'P62112' : 'psbM',
        'P12359' : 'PSBO',
        'P61840' : 'psbT',
        'Q41387' : ' psbW',
        'A0A0K9RTU3': 'psbX',
        'Q9M3M6' : 'psbZ',
        'A0A0K9RHP1' : 'psbTn'
}
```

Supplementary Figure 4. Spinacia oleracea PsbS and PSII-LHCII sc core database. As keys we have the UniprotKB accessions and as values we have the name we assigned them. In light grey are the proteins which could not be verified via MaxQuant.



Supplementary Figure 5. Ev couplings Mapping Analysis and Crosslinking (EvoMAS) pipeline.