

**THE ASSEMBLY AND REPAIR OF
CYANOBACTERIAL PHOTOSYSTEM II**

by

Shengxi Shao

A thesis submitted to Imperial College London
for the Degree of Doctor of Philosophy

2017

Department of Life Sciences, Imperial College London
London SW7 2AZ

Statement of Originality

I hereby declare that this thesis, submitted in fulfilment of the requirements for the degree of Doctor of Philosophy of Imperial College London, represents my own work and has not been previously submitted to this or any other institute for any degree, diploma or other qualification.

Shengxi Shao

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Abstract

Photosystem II (PSII) is the multi-subunit membrane-anchored light-driven water:plastoquinone oxidoreductase located in the thylakoid membranes of chloroplasts and cyanobacteria. This macrocomplex harvests solar energy to power oxygenic photosynthesis, a reaction that renews atmospheric oxygen, supplies fixed carbon for the food chain and maintains the global carbon level. The D1 subunit is one of the two reaction centre proteins of PSII and provides amino-acid ligands to the Mn_4CaO_5 cluster involved in water oxidation. CP43, which plays a light-harvesting role in PSII, also provides a ligand to the cluster. In the structurally related photosystem I (PSI) complex, the equivalent to CP43 and D1 are fused and synthesised as a single subunit. In this work, a CP43-D1 fusion strain was successfully constructed in the cyanobacterium *Synechocystis* sp. PCC 6803 and shown to assemble a functional PSII complex capable of water oxidation. However, PSII activity *in vivo* was sensitive to high irradiances, and photoinhibition analyses showed impaired PSII repair in the fusion mutant and slower degradation of the CP43-D1 fusion protein. This work supports the hypothesis that CP43 and D1 are synthesised as separate subunits to allow prompt and efficient repair of PSII following damage by light. I have shown by phylogenetic analysis that PSII repair mediated by FtsH proteases might have also played an important role in the evolution of oxygenic photosynthesis. There are at least three orthologous groups of FtsH in the tree of life, with the FtsH involved in PSII repair diverging earlier than the radiation of FtsH from all other bacteria. Additionally, I carried out a structural conservation analysis on Psb29, an interacting partner of FtsH, to reveal potential interacting sites on this protein that are of interest for future studies for understanding the regulation of the PSII repair mechanism.

Publication

Excerpts of the present work have been submitted for publication:

(Chapter 4)

Shao S., Cardona T., Nixon P. J.: Divergence of Photosystem II-specific FtsH proteases at the dawn of oxygenic photosynthesis. *Photosynthetica*. (under review)

(Chapter 5)

Beckova M., Yu J., Krynická V., Kozlo A., Shao S., Konik P., Komenda J., Murray J. W., Nixon P. J.: Structure of Psb29/Thf1 and its association with the FtsH protease complex involved in photosystem II repair in cyanobacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* (in press)

Acknowledgements

I would like to thank my supervisor Prof. Peter Nixon and co-supervisor Dr. James Murray for supporting my PhD scholarship application, from where this journey began.

I sincerely thank Prof. Peter Nixon for his elaborate guidance, inspiration and constructive criticisms throughout the course of my study. From Peter I learn the beauty of science, I am deeply grateful for all that I have learned from you over the past years.

I particularly thank Dr. Jianfeng Yu, who has been helping me in both research and life in London since the first day I joined the Nixon group. I thank Dr. Tanai Cardona and Prof. Bill Rutherford for the enlightening discussions on the evolution of PSII. Conversations with Bill about his experiences, as a scientist, father and musician, will benefit me for a life time. I am also thankful for the support from former members of the Nixon group, Dr. Wojciech Bialek and Dr. Karim Maghlaoui. Also, I would like to express my gratitude to all my colleagues in the 7th-floor photosynthesis lab of the Sir Ernst Chain Building, Imperial College: all of you have made the past four years an invaluable experience in my life.

I sincerely thank Prof. Josef Komenda for inviting me to his laboratory, at Třeboň, Czech Republic, for two one-month visits. Moreover, I am grateful for the help from Dr. Jana Knoppová during my stay in the Czech Republic.

I thank the Department of Life Sciences of the Imperial College London for nominating me for the CSC/Imperial PhD scholarship, and I thank the Chinese Scholarship Council for awarding me this PhD scholarship to fund my research and life in London.

Special thanks go to Dr. Cheng-Yi Tang who inspired my curiosity on the secrets of life. It is your encouragement and inspiration that led me to the field of life sciences.

I am deeply indebted to my parents for their constant and loving support which always allowed me the luxury to let interest be my guide for the studies I pursued. I am sincerely grateful for the supports I received from my parents-in-law.

Finally, I thank my wife, Jing Yao, and my son, Qiyuan Shao, for coming into my life.

You are the reason I want to make the world better.

Contents

Title	1
Statement of Originality	2
Copyright declaration.....	2
Abstract	3
Publication	4
Acknowledgements.....	5
List of figures.....	11
List of tables.....	13
Abbreviations.....	14
Chapter 1. General introduction.....	16
1.1. Oxygenic photosynthesis and the global energy and food crisis.....	16
1.2. PSII: the engine of life.....	19
1.2.1. PSII initiates the energy flow of oxygenic photosynthesis.....	19
1.2.2. Architecture of cyanobacterial PSII	20
1.2.3. Cyanobacteria as model organisms for PSII studies.....	23
1.3. Photoinhibition and the maintenance of PSII	25
1.3.1. Mechanisms of PSII photoinhibition.....	25
1.3.2. The step-wise modular assembly model of PSII	27
1.3.3. The FtsH-mediated PSII repair model.....	29
1.4. Project scope.....	30
Chapter 2. Materials and methods	32
2.1. Biological materials.....	32
2.1.1. <i>E. coli</i> strain and growth condition.....	32
2.1.2. Oligonucleotide primers used in this work.....	32
2.1.3. Plasmids used in this work	35
2.1.4. Cyanobacteria strains and growth conditions.....	35
2.1.5. Estimation of cell concentration of liquid <i>E. coli</i> and <i>Synechocystis</i> sp. PCC 6803 cultures	38

2.2.	DNA and RNA techniques	39
2.2.1.	DNA transformation of cells	39
2.2.2.	Extraction and purification of nucleic acids	40
2.2.3.	Amplification and analyses of nucleic acids	41
2.3.	Protein biochemistry techniques.....	44
2.3.1.	Thylakoid membrane and PSII purification from <i>Synechocystis</i>	44
2.3.2.	Polyacrylamide gel electrophoresis (PAGE).....	46
2.3.3.	Western blotting and semi-quantification.....	47
2.3.4.	Sample preparation for mass spectrometry	48
2.4.	Physiological analyses.....	49
2.4.1.	Growth experiment under light stress.....	49
2.4.2.	Physiological measurements	50
2.5.	Bioinformatic techniques.....	52
2.5.2.	Evolutionary analyses of the FtsH protein family	53
2.5.3.	Conservation analyses of Psb29/THF1.....	55
Chapter 3. Repair and the evolution of photosystem II		57
3.1.	Introduction	57
3.1.1.	Structural similarity between two types of photosynthetic reaction centres ...	57
3.1.2.	The evolutionary controversy of two types of reaction centres.....	58
3.1.3.	A potential role of repair in the evolution of oxygenic photosynthesis.....	60
3.2.	A CP43-D1 fusion PSII is assembled and active	61
3.2.1.	Construction of the CP43-D1 fusion PSII	61
3.2.2.	Purification and composition analysis of the CP43-D1 fusion PSII.....	65
3.2.3.	Characteristics of the CP43-D1 fusion strain	70
3.2.4.	Characterisation of a parallel CP43-D1 fusion strain and its Δ FtsH2 derivative	75
3.3.	Repair of the fusion PSII is defective.....	79
3.3.1.	PSII repair is defective in the fusion strain.....	79
3.3.2.	The expression level of the CP43-D1 fusion gene	83
3.3.3.	The CP43-D1 fusion protein can be degraded under extreme highlight stress	84
3.4.	Split for survival: genomic study of fusion suppressors.....	85
3.4.1.	FuBH, a CP43-D1 split highlight suppressor	85
3.4.2.	Mock-evolution: mutagenesis-assisted CP43-D1 split suppressors	96
3.5.	Discussion	102
3.5.1.	Incorporation of CP47 and CP43 into PSII need not be sequential.....	102
3.5.2.	Efficient repair selects against the fusion of CP43 and D1	104
Chapter 4. An evolutionary view of the cyanobacterial FtsH proteases.....		106

4.1.	Introduction	106
4.1.1.	FtsH proteases and oxygenic photosynthesis	106
4.1.2.	Diverse function of FtsH proteases	106
4.1.3.	Common structure of FtsH complexes	108
4.2.	A phylogenetic survey of FtsH.....	111
4.2.1.	Early diversification of FtsH proteases	112
4.2.2.	Classification of cyanobacterial FtsH paralogs	118
4.2.3.	Multiplicity of FtsH in photosynthetic eukaryotes	122
4.3.	Structural characteristics of the cyanobacterial FtsH	125
4.3.1.	The structural conservation of FtsH in domain Bacteria	125
4.3.2.	Cyanobacterial characteristics near the structurally conserved regions of FtsH 128	
4.4.	Preliminary work on the function of FtsH4 in <i>Synechocystis</i>	131
4.5.	Discussion	133
4.5.1.	The overlooked evolution of FtsH.....	133
4.5.2.	Mutations might explain the diverse symmetries seen in FtsH crystal structures 134	
4.5.3.	The flexible linker and lid helix in the FtsH complex might interact.....	136
4.5.4.	A speculated action model of the FtsH protease complex.....	139
Chapter 5.	Towards the role of Psb29/THF1	143
5.1.	Introduction	143
5.2.	Preliminary characterization of a <i>psb29</i> null mutant of <i>Synechocystis</i>	145
5.2.1.	Accumulation of FtsH is affected in Δ Psb29	145
5.2.2.	Mixotrophic defect/D-glucose sensitivity of Δ Psb29.....	149
5.3.	Analyses of suppressor mutations in Δ Psb29 <i>Synechocystis</i>	150
5.3.1.	Phenotypes of Δ Psb29 suppressor mutants	150
5.3.2.	Genotype of Δ Psb29 suppressor mutants	151
5.3.3.	Δ Psb29-OCP: An OCP mutation in the sequenced Δ Psb29 strain	154
5.4.	Conservation of the Psb29/THF1 family and its implication	156
5.4.1.	Psb29/THF1 is closely related to oxygenic photosynthesis	156
5.4.2.	Structural conservation of the Psb29/THF1 family	158
5.5.	Discussion	161
5.5.1.	Psb29 likely provides site of diverse interaction	161
5.5.2.	Two conserved motifs of structural significance.....	162
5.5.3.	Psb29 and the accumulation of FtsH2	163
5.5.4.	Psb29 and D-glucose sensitivity.....	163
Chapter 6.	Conclusion and future works.....	165

References.....	168
Supplementary Table 1	189
Supplement 2. Access to the interactive phylogenetic trees in Figure 4-2	193
Supplement 3. Sequences of the 732 and 716 bp homologous region shown in Figure 3-18A	193

List of figures

Figure 1-1 World population and the primary energy supply.....	17
Figure 1-2 The “Z-scheme” of the electron transfer during the oxygenic photosynthesis.....	19
Figure 1-3 Organisation of the protein components of cyanobacterial PSII complex.	21
Figure 1-4 Light response curve for PSII activity.	25
Figure 1-5 An illustration of the step-wise modular assembly model of cyanobacterial PSII.....	27
Figure 1-6 Current model of PSII repair adapted from Nixon et al. (19).	30
Figure 3-1 Comparison of the arrangement of PSI and PSII transmembrane helices.....	58
Figure 3-2 Structural view of the CP43-D1 fusion design.	62
Figure 3-3 Plasmid map of the CP43-D1 fusion construct and the segregation of transformant.....	64
Figure 3-4 Western blotting of the CP43-D1 protein in thylakoid preparation.	65
Figure 3-5 Western blotting of the CP43-D1 PSII.	66
Figure 3-6 Components of the CP43-D1 fusion PSII analysed by mass spectrometry.....	68
Figure 3-7 Clear-native and 2D PAGE analyses of the CP43-D1 fusion PSII.	70
Figure 3-8 Growth assay and growth curve.....	72
Figure 3-9 Oxygen evolution activity and electron transfer property.....	74
Figure 3-10 Detection of CP43-D1 fusion PSII in CP43-D1/ Δ PsbA and CP43-D1/ Δ PsbA/ Δ FtsH2.....	76
Figure 3-11 Growth assay of the second fusion strain CP43-D1/ Δ PsbA and the FtsH2-knockout derivative CP43-D1/ Δ PsbA/ Δ FtsH2.	77
Figure 3-12 Activity assays of the fusion strain CP43-D1/ Δ PsbA and CP43-D1/ Δ PsbA/ Δ FtsH2.....	79
Figure 3-13 Photoinhibition assay on the CP43-D1 fusion strain	82
Figure 3-14 Relative quantifications of the transcript-level expression of the CP43-D1 fusion gene....	83
Figure 3-15 Degradation of the CP43-D1 protein under extra-high light stress.....	85
Figure 3-16 Growth assay and western blotting of the CP43-D1 split strain, FuBH.....	87
Figure 3-17 PSII repair and the D1 expression level in the CP43-D1 split strain, FuBH.....	88
Figure 3-18 Genotyping the CP43-D1 and its suppressor FuBH.	95
Figure 3-19 Design of the break fusion mutants.....	97
Figure 3-20 Confirmation of break fusion suppressors.	99
Figure 3-21 Localisation of the D1-encoding gene in the genomes of split and corresponding suppressor mutants.	101

Figure 4-1 Structures of FtsH homologues.....	110
Figure 4-2 Multiplicity and evolution of FtsH proteases.....	112
Figure 4-3 Phylogenetic tree of FtsH proteases from phototrophic groups (top) compared to those from Type II (bottom left) and Type I (bottom right) reaction centre proteins reprinted from Cardona (157).	115
Figure 4-4 Unrooted phylogenetic tree of cyanobacterial FtsH.....	118
Figure 4-5 Structural conservation of FtsH protease sampled from 55 phyla of bacteria.	125
Figure 4-6 Conservation and comparison of the cyanobacterial FtsH paralogous groups.....	128
Figure 4-7 Complementation assay of FtsH4 in the FtsH2 null <i>Synechocystis</i>	131
Figure 4-8 Different conformational statuses of FtsH complexes and monomers.....	136
Figure 4-9 A structure-based working model of FtsH protease complex.	139
Figure 5-1 Complementation growth assay and photoinhibition assay of the Δ Psb29 strain.	148
Figure 5-2 Close-up views of Δ Psb29, Δ FtsH2 and WT growth under different metabolic backgrounds.	149
Figure 5-3 Growth assay of Δ Psb29 and Δ Psb29 suppressor mutants.	151
Figure 5-4 Data mining of the DUF760 family and sequence conservation of the cyanobacterial DUF760 proteins.	154
Figure 5-5 Phylogeny and sequence similarity of the Psb29/THF1 family.	157
Figure 5-6 Conservation analyses of Psb29/THF1.	159
Figure 5-7 Close-up views of the conserved face of Psb29/THF1 identified by the ConSurf analysis.	161

List of tables

Table 2-1 Oligonucleotide primers used in this project.....	32
Table 2-2 <i>Synechocystis</i> sp. PCC 6803 strains used in this study	35
Table 2-3 Supplements in BG-11 medium	38
Table 3-1 List of SNPs detected in the FuBH and its background strains.	90
Table 3-2 The quality report of the genome assemblies assessed by QUAST toolkit	94
Table 4-1 A list of some reported substrates of FtsH and homologues	107
Table 4-2 List of cyanobacteria that might have lost one or more FtsH during evolution.	121
Table 5-1 List of SNPs detected in Δ Psb29 suppressors	152

Abbreviations

1-D	one-dimensional
2-D	two-dimensional
ADP	adenosine-5' - diphosphate
ATP	adenosine-5'-triphosphate
bp	base pairs
β -DM	n-dodecyl- β -D-maltoside
Chl	chlorophyll
CN	clear-native
CTD	carboxyl-terminal domain
DCBQ	2,6-dichloro- <i>p</i> -benzoquinone
DCMU	3- (3,4-dichlorophenyl)-1,1-dimethylurea
ECL	enhanced chemiluminescence
EtOH	ethanol
FtsH	filament temperature sensitive H
Gya	billion (10^9) years ago
kDa	kilodalton
LB	Luria-Bertani medium
LMM	low molecular mass
<i>m</i> -AAA	matrix facing mitochondrial FtsH family protease
NADP ⁺	the oxidized form of nicotinamide adenine dinucleotide phosphate
NADPH	the reduced form of nicotinamide adenine dinucleotide phosphate
OEC	oxygen-evolving complex
OCP	orange carotenoid protein

P680	primary electron donor of photosystem II
P700	primary electron donor of photosystem I
PAGE	polyacrylamide gel electrophoresis
PCC	Pasteur Culture Collection
PCR	polymerase chain reaction
PDB	protein data bank
PSI	photosystem I
PSII	photosystem II
qPCR	quantitative polymerase chain reaction
RC	reaction centre
RC47	PSII sub-complex containing D1, D2 and CP47 but lacking CP43
RC43	PSII sub-complex containing D1, D2 and CP43 but lacking CP47
RCC	PSII reaction centre complex
ROS	reaction oxygen species
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
THF1	thylakoid formation 1
TW	terra (10^{12}) watts
v/v	volume per volume
w/v	weight per volume
WT	wild type

Chapter 1. General introduction

1.1. Oxygenic photosynthesis and the global energy and food crisis

In oxygenic photosynthesis, solar energy is utilised to split water and fix inorganic carbon. This reaction has been fundamental in shaping both the biosphere and atmosphere since ca 2.5 Gya (1, 2). Photosynthesis absorbs solar energy, oxidises the atmosphere and causes climate change (3). As the major primary producer, photosynthetic organisms play a vital role in feeding the food chain. Since the emergence of agriculture from the Neolithic Revolution, the human race secured a sustainable source of food mainly by the domestication of crops. This utilisation of photosynthesis successfully fostered early human civilisation and ultimately to its present unprecedented prosperity. Fossil fuels, the energy inheritance of ancient photosynthesis, have been utilised to enhance the productivity of agriculture, in the form of artificial fertilisers, herbicides, pesticides and mechanisation, which contributed to the so-called “Green Revolution” (4) (Figure 1-1). This extra energy input from fossil fuels vastly released the potential of agricultural production, and the success of modern agriculture naturally allows more population to be fed. From 1960 to 2010, the world consumption of fossil fuels increased by approximately 364% (5); meanwhile, the world population increased by approximately 225% (<https://esa.un.org/unpd/wpp/>) (Figure 1-1).

With the current trajectory of human population growth (Figure 1-1, dotted line), the demand for a sustainable source of energy has never been greater. The technological advances of modern agriculture that started with the Green Revolution have made it possible to provide food for a global human population of 7.5 billion in 2017. This is projected to grow to 9.7 billion by 2050 (6). To feed such a large population, a

tremendous amount of energy must be made available as soon as possible. Currently, fossil fuels are the sole energy source capable of meeting such a scale of demand. However, fossil fuels are not without limitations. With the growth of the human population, the expansion of settlements must inevitably compete with the agronomic land use, and the benefits from fossil fuel consumption in agriculture become increasingly marginal (7). Moreover, excessive combustion of fossil fuels produces greenhouse gases that cause climate change (8, 9) as well as various environmental pollutants (10-12). Despite mankind's development of advanced technologies to enable humans to "play God" with the environment, our biological and ecological characteristics are not dissimilar to other Earth residents, in terms of the reliance on a fine-tuned biosphere and the resilience to abrupt climate or ecological changes. Therefore, the search for a sustainable source of energy that is both carbon-neutral and environmentally-friendly is a necessary endeavour for humanity, in order to resolve the global energy and food crisis.

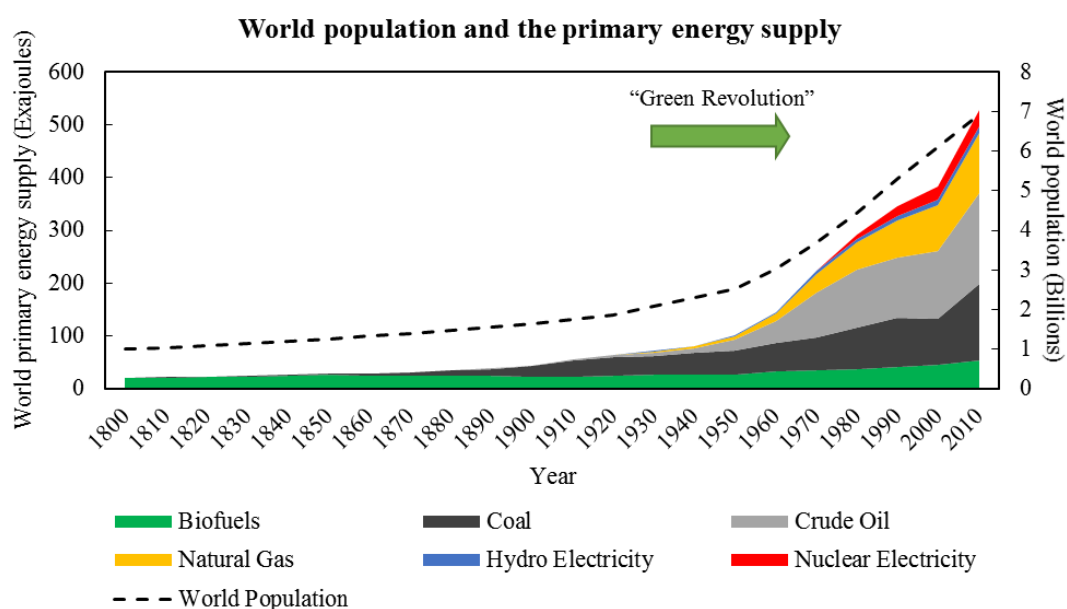


Figure 1-1 World population and the primary energy supply.

The data are from Smil (5). The green arrow indicates the agricultural "Green Revolution" from the 1930s to 1960s (4).

One potential solution to the energy crisis is to improve the efficiency of oxygenic photosynthesis. It is estimated that photosynthetic organisms capture solar energy at a rate of 200 TW globally (13), 11 times greater than the rate of energy consumption by humankind (17.7 TW in 2012 (14)). Nevertheless, the harvesting of energy by oxygenic phototrophs is relatively inefficient and achieved with a global energy conversion efficiency of approximately 0.2% (13).

A number of components of oxygenic photosynthesis have been proposed as targets for improving photosynthetic efficiency (15). Through genetic engineering, increased photosynthetic efficiency has led to as high as 20% increase in biomass in field-grown crop plants (16). A major target for enhancing photosynthetic efficiency is to enhance the tolerance of crop plants to light stress. Light causes the photodamage to the photosystem II (PSII) complex, which functions as the light-driven water:plastoquinone oxidoreductase in photosynthetic electron transport. However, photoautotrophs have evolved a number of strategies to mitigate the effects of photodamage, including an efficient PSII repair cycle, which maintains the functionality of PSII and removes damaged PSII complexes that act as a source of reactive oxygen species (ROS). Improving the PSII repair process is therefore one potential route to increase the efficiency of photosynthesis.

Recent studies by Nixon and colleagues have established a working model of PSII repair which emphasises the role of the FtsH proteases in degrading damaged PSII proteins (17-21). The studies described in Chapter 3 of this PhD dissertation test the current FtsH model of PSII repair, by investigating the assembly and repair of a novel type of PSII complex in which the N-terminus of D1 is fused to the C-terminus of the CP43 chlorophyll-binding complex. Chapter 4 of this dissertation describes the application of bioinformatics to investigate the early divergence of the FtsH proteases

involved in PSII repair. Chapter 5 discusses the structural conservation and potential role of Psb29/THF1, a highly conserved protein found exclusively in oxygenic photosynthetic organisms (22), which was recently shown to be involved in the accumulation of FtsH proteases in plants and cyanobacteria (22-24).

1.2. PSII: the engine of life

1.2.1. PSII initiates the energy flow of oxygenic photosynthesis

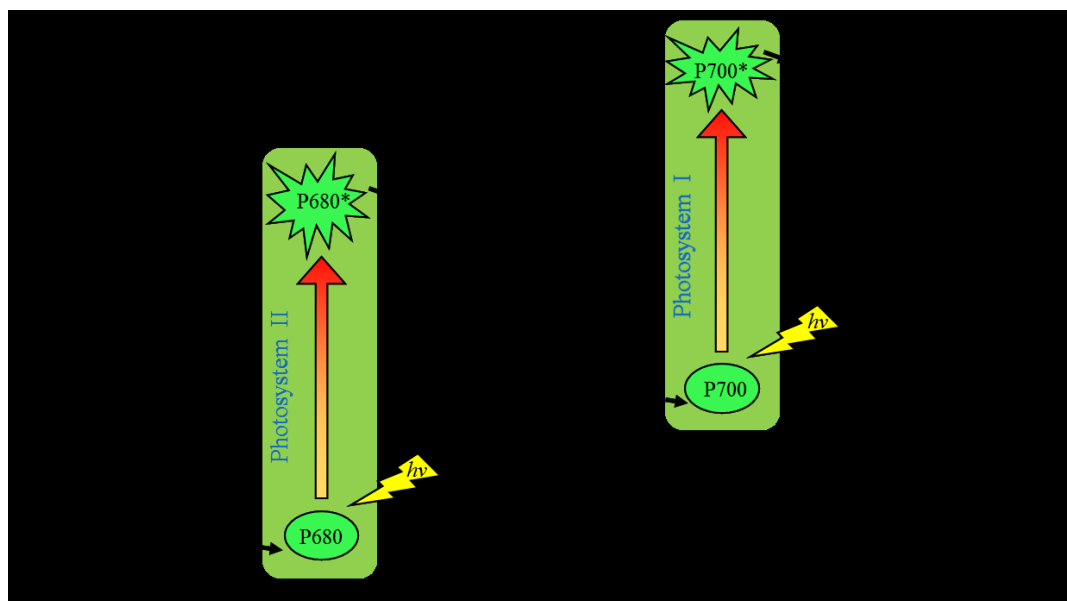


Figure 1-2 The “Z-scheme” of the electron transfer during the oxygenic photosynthesis.

The electron transport pathway from water to NADP^+ , or the “Z-scheme”, adapted from <http://www.life.illinois.edu/govindjee/ZSchemeG.html>. Designation in the diagram: Mn, the manganese complex; Y_Z , the redox active tyrosine Z in PSII; P680, the redox active PSII reaction centre chlorophyll molecules, the 680 number is its absorption maximum in the red part of the visible spectrum (680 nm); P680*, the excited state of the P680; Pheo, pheophytin; Q_A and Q_B , plastoquinone molecules bound in PSII; PQ, the pool of detached free plastoquinone; PC, plastocyanin; P700, the redox active PSI reaction centre chlorophyll molecules, its absorption spectrum peaks at 700 nm; P700*, the excited state of P700; A_0 , the primary electron acceptor chlorophyll molecule of PSI; A_1 , a phylloquinone molecule (vitamin K); F_X , F_A and F_B are three separate immobile iron-sulfur protein centres; FD, ferredoxin; FNR, ferredoxin- NADP^+ oxidoreductase.

In the so-called “light-dependent” reactions of photosynthesis, the energy of solar quanta is absorbed and transferred by a series of electron transitions and movements within the photosynthetic apparatus (22). Oxygenic photosynthetic electron transport comprises a series of redox reactions that extract electrons from the oxidation of water (23-25) (“water splitting”), ultimately reducing NADP^+ to NADPH, as described by the

classic “Z-scheme” established half a century ago (Figure 1-2) (26, 27). During this process, a proton electrochemical gradient is formed across the thylakoid membrane which drives ATP synthesis (28). Upon illumination, P680, a photochemically heterogeneous chlorophyll species (29), is excited to a higher energy state, P680*, which is then oxidised to P680⁺ by transfer of an electron to a nearby pheophytin molecule. The highly oxidising P680⁺ species subsequently oxidises a neighbouring tyrosine residue (30, 31), Yz, which in turn oxidises a Mn₄CaO₅ cluster. After the metal cluster has accumulated four oxidising equivalents (25, 32, 33), two bound water molecules are oxidised to one dioxygen molecule, releasing four electrons and four protons. This redox cycle continuously extracts electrons from water, exporting them out of PSII via pheophytin, plastoquinones (Q_A and Q_B) to the downstream acceptors, the cytochrome *b₆f* complex, plastocyanin and the photosystem I (PSI) complex. Within PSI, the electrons are “elevated” again to a higher reducing potential via excitation of P700 to P700*. Electrons are then delivered to the iron-sulfur cluster, ferredoxin and eventually passed to the ferredoxin:NADP⁺ reductase to produce NADPH, which provides the reducing power needed for carbon fixation (34).

1.2.2. Architecture of cyanobacterial PSII

Photosystem (PSII) is a dimeric membrane protein-pigment complex with each monomer consisting of at least 99 cofactors (35-37). The structures of PSII are highly conserved between cyanobacteria, red algae and green plants although there are some differences in the peripheral light-harvesting systems (38) and the composition of the PSII extrinsic subunits (39). The first X-ray structure of a photosynthetic reaction centre was determined from a purple bacterium (40, 41) and the first X-ray structure of cyanobacterial PSII was obtained with a resolution of 3.8 Å (42). Since then there has been a progressive improvement in the quality of PSII crystal structures so that the

structural resolution now stands at 1.9 Å (35-37). The most recent crystal structures of cyanobacterial PSII reveal that each PSII monomer contains 19-20 protein subunits, 35-36 chlorophylls, 11-12 β-carotenes, two pheophytins, 2-3 plastoquinones, two hemes, one non-heme iron, 25 lipids, a Mn₄CaO₅ cluster and two nearby chloride ions (37, 43).

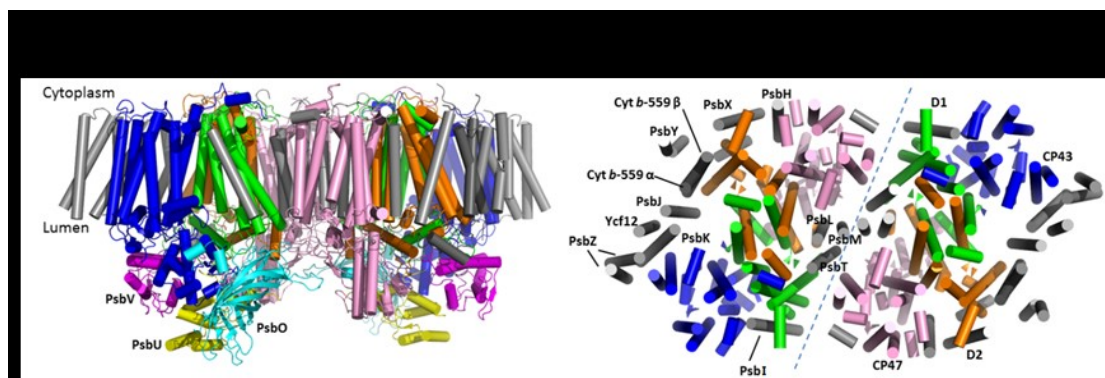


Figure 1-3 Organisation of the protein components of cyanobacterial PSII complex.

Adapted from the *Thermosynechococcus elongatus* PSII structure (PDB: 4V62) (36). Panel A presents the overall organisation of 20 protein factors, 3 extrinsic proteins are labelled as PsbV, PsbU and PsbO in one monomer; Panel B is the top-view from cytoplasm showing the arrangement of transmembrane helices of 17 intrinsic proteins, the 4 main scaffold proteins are labelled on the right-side monomer, the 13 low molecular weight subunits are labelled on the left counterpart. For clarity, all cofactors are omitted. Colour assignments: CP43-blue, D1-green, CP47- pink, D2-orange, α- and β-subunits of cytochrome b₅₅₉ and PsbH, I, J, K, L, M, T, X, Y, Z, Ycf12-gray, PsbO-cyan, PsbU-yellow, PsbV-magenta.

The 17 transmembrane proteins form a spatially and functionally coordinated complex that binds the various cofactors. D1 and D2 are subunits that constitute the photochemical reaction centre, located at the heart of PSII, each containing five transmembrane helices. D1 provides seven out of nine amino-acid ligands for the Mn₄CaO₅ cluster (37), which is the oxygen-evolving centre responsible for splitting water. The D1-D2 reaction centre subcomplex binds six chlorophylls, two pheophytins, two plastoquinones and two β-carotenes. The two sets of five transmembrane helices of D1 and D2 and the associated cofactors are arranged around a pseudo-twofold axis, providing a physical path for electron transfer within PSII. The remaining chlorophylls and β-carotenes of PSII are mainly bound to the core antenna proteins CP43 and CP47, each consisting of six transmembrane helices. CP43 and CP47 bind 13 chlorophylls and

16 chlorophylls, respectively (35, 44). The CP43 and CP47 pigment-protein complexes function as an antenna network to funnel excitation energy from absorbed photons to the reaction centre. The PSII core complex also contains 13 low molecular mass (LMM) intrinsic proteins: PsbH, I, J, K, L, M, T, X, Y and Z, Ycf12 (Psb30), and the α - and β -subunits of cytochrome *b*₅₅₉ (PsbE, F), each containing one to two transmembrane helices. The functions of these LMM proteins are likely to include assisting in the assembly and repair of PSII complexes. For instance, PsbI is an early assembly partner of D1 and assists in CP43 binding to PSII (45). Additionally, the PSII complex contains at least 25 integral lipids which are essential in maintaining structural integrity and play a role in the assembly and repair of PSII. It is thought that integral lipids may also be important for the release of dioxygen from PSII, the product of water-splitting (46).

Three extrinsic proteins, PsbO, PsbU and PsbV, are bound to the luminal side of cyanobacterial PSII. One of their functions is to shield the Mn₄CaO₅ cluster from reduction by external reductants (47). In cyanobacteria, two other luminal extrinsic proteins are present, CyanoP and CyanoQ, which are homologues of PsbP and PsbQ, the equivalent proteins found in the chloroplasts of higher plants (47-49). CyanoQ is not included in currently available PSII crystal structures, although the association between CyanoQ and PSII is well-evidenced (48, 50). A recent chemical cross-linking study suggests that CyanoQ interacts with both PsbO and CP47 at the interface of the two monomeric PSII complexes (51). Moreover, a variant of PSII, containing multiple copies of CyanoQ and lacking PsbU and PsbV has recently been observed, indicating that CyanoQ may undertake a role in the assembly or turnover of PSII (52). The structural and functional association between CyanoP and PSII have been difficult to elucidate as CyanoP is not retained during standard PSII preparations (53). Knockout of CyanoP appears to have no significant impact on the assembly and function of

mature PSII (53, 54), suggesting that the role of CyanoP is different to those of other extrinsic subunits. A recent *in vitro* study has suggested that CyanoP binds to the C-terminal region of D2 (55). Further evidence for the association between CyanoP and PSII *in vivo* has been demonstrated with a recent pull-down study, which has indicated a physical interaction between CyanoP and an early PSII assembly intermediate, comprising cytochrome *b*₅₅₉ and D2, and weak interaction with CP43 is also observed (56). This interaction between CyanoP and D2 shows a physiological significance in the prompt formation of PSII RC. Notably, in higher plants, PsbP, the homolog of CyanoP, binds to a position corresponding to the binding sites of PsbV and PsbU in cyanobacterial PSII (38).

1.2.3. Cyanobacteria as model organisms for PSII studies

The cyanobacteria is a phylum of bacteria and is the only known prokaryotic group capable of oxygenic photosynthesis. It is thought that the emergence of cyanobacteria has been instrumental to the Great Oxygenation Event (GOE), in which biologically produced molecular oxygen became a major component of the atmosphere, drastically altering the course of evolution of life on the planet Earth (1, 3). It is generally accepted that the common ancestor of all eukaryotic phototrophs arose through the acquisition of cyanobacterial-derived photosynthetic apparatus, the so-called “primary endosymbiosis” event (57-59). Cyanobacteria are therefore evolutionary cousins of chloroplast organelles in eukaryotic phototrophs. Cyanobacteria are standard model organisms in the investigation of photosynthesis, particularly for the structural and functional characterization of PSII. Cyanobacteria and higher plants possess markedly different light harvesting system (60). Nevertheless, the arrangements of PSII protein subunits in cyanobacteria and plants are nearly identical, as evidenced by recent high-resolution PSII structures at 1.9 Å (37) and 3.2 Å (38) respectively, with the exception

of PsbW that is only present in higher plants and algae (61). Moreover, cyanobacteria have a number of features amenable to laboratory investigation, including straightforward cultivation, ease of genetic manipulation and genetic simplicity.

Since its first isolation from fresh water in Oakland, California by R. Kunisawa in 1968 (62), the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis*) has become the most studied cyanobacterial model. *Synechocystis* was the first photosynthetic organism to have its genome sequenced (63), and at least 11 substrain genomes have now been sequenced (64-67). In spite of the fact that *Synechocystis* is highly polyploid (68), a typical *Synechocystis* genome is approximately 3.9 Mb in size, comprising one chromosome and seven plasmids which, in total, contain 3726 genes (NCBI assembly: GCA_000340785.1). A widely-studied model is the glucose-tolerant *Synechocystis* substrain (hereafter referred to as WT-G) (64, 69) which has the advantageous capability of utilising glucose as an alternative carbon source in the absence of functional PSII. Hence, WT-G and its derived strains serve as powerful tools for the characterization of PSII. Furthermore, *Synechocystis* is amenable to straightforward genetic transformation, as it is naturally susceptible to spontaneous exogenous DNA uptake (70).

Thermosynechococcus elongatus (hereafter referred to as *T. elongatus*) is another popular cyanobacterium used in photosynthesis research and was first isolated from a Japanese hot spring (71). Being thermophilic, with an optimal growth temperature of approximately 57°C (71), *T. elongatus* proteins are typically highly stable and therefore favoured in crystallisation experiments. Indeed, crystallisation of *T. elongatus* PSII proteins, either from direct purification from *T. elongatus* cultures or recombinant expression in other hosts such as *E. coli*, led to the first crystal structures of cyanobacterial PSII (42) and PSII-related proteins (35, 48, 72-75). The closely related

Thermosynechococcus vulcanus has also been a crucial tool in studying the structure of PSII (37, 76-78). Notably, in contrast to *Synechocystis*, *T. elongatus* is generally considered an obligate photolithoautotrophic prokaryote, although recently it has been shown that *T. elongatus* might also be capable of utilising organic carbon sources (79). Nevertheless, mutagenesis of essential photosynthetic proteins in *T. elongatus* is still challenging.

1.3. Photoinhibition and the maintenance of PSII

1.3.1. Mechanisms of PSII photoinhibition

While absorption of light quanta is essential for water-splitting and electron transfer in PSII, light is damaging to PSII. At low irradiances, PSII activity is proportional to light intensity (Figure 1-4). However, after achieving maximum PSII activity, a further increase in light intensity results in inhibition of PSII. This repression of PSII activity by excessive light is generally known as chronic or irreversible photoinhibition.

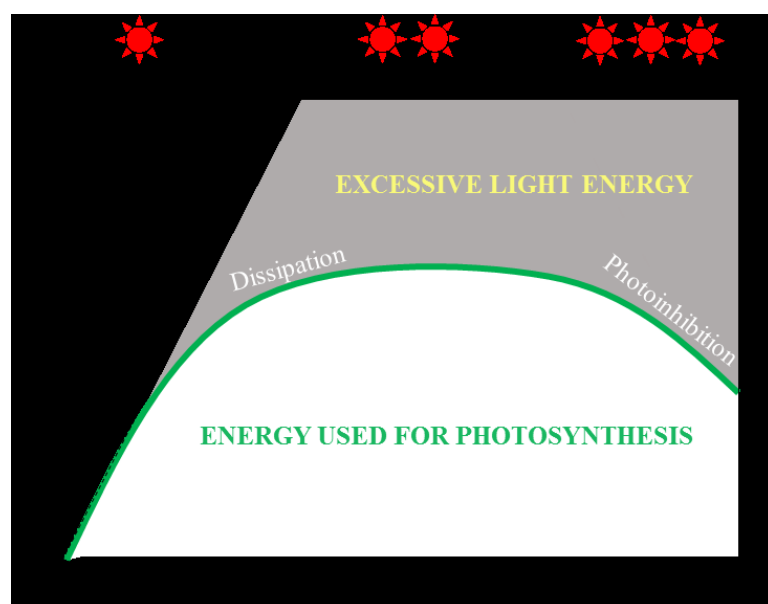


Figure 1-4 Light response curve for PSII activity.

The diagram is adapted from Yamamoto (80). The dotted line represents the energy input, and the grey area indicates the energy that is not photosynthetically productive.

Photoinhibition is considered to be a direct consequence of an imbalance between the rate of photodamage and rate of recovery of PSII functionality. Oxidative photodamage mainly affects the D1 subunit of PSII (81, 82), as it has a much shorter half-life than other PSII components. This is reflected in the rapid turnover of D1, which has a turnover rate about 5-, 8- and 10-fold greater than those of D2, CP43 and CP47 respectively (83).

During photoinhibition, damage to the D1 subunit by ROS is elevated. ROS can be generated through a number of donor-side and acceptor-side photoinhibition mechanisms. In the donor-side photoinhibition (84, 85), lack of coordination between $P680^+$ reduction and water oxidation leads to the formation of hydrogen peroxide (H_2O_2). This incomplete oxidation of water and reduction of $P680^+$ is probably due to damage of the OEC through loss of a manganese ion (86). The resulting H_2O_2 is further oxidised by $Tyr_Z^\bullet$ or the chlorophyll cation Chl_Z^{++} to form the anionic superoxide radical ($O_2^{\bullet-}$) (87, 88). Alternatively, H_2O_2 is reduced by free manganese to form the hydroxyl radical ($HO^\bullet$) (89). Moreover, the delayed reduction of $P680^+$ consequently increases the chance of stochastic oxidative damage (84). In acceptor-side photoinhibition (90-92), over-reduction of the plastoquinone pool leads to charge recombination which promotes the generation of $^3P680$ (93). Notably, charge recombination also occurs under low light, accompanied by a decrease in the rate of PSII excitation (94). $^3P680$ can react with triplet oxygen to produce the highly damaging singlet oxygen ($^1O_2$) species.

Impaired turnover of PSII also accounts for photoinhibition. The removal of photodamaged D1 is synchronised with its replacement by newly synthesised D1. Therefore the availability of newly synthesised D1 is a major rate-limiting constraint

for PSII recovery. ROS, a by-product of photosynthesis or photoinhibition, not only damages PSII components but also represses the expression of D1 at the transcriptional and translational levels (95, 96), which directly impedes the repair of PSII. Moreover, in some cyanobacteria, different D1 paralogs coexist (97), and their incorporation into PSII adjust the oxygen-evolving activity and electron transfer properties (98-100). The availability of various D1 paralogs and their incorporation in the turnover process make possible a prompt adaptation strategy to react upon photoinhibition by switching between different versions of PSII.

1.3.2. The step-wise modular assembly model of PSII

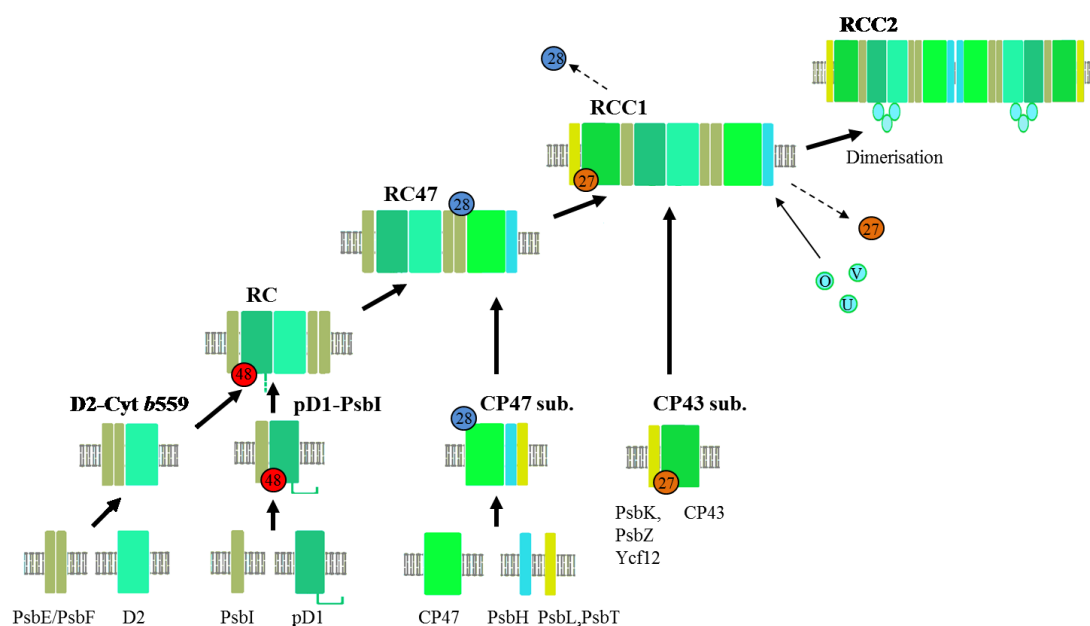


Figure 1-5 An illustration of the step-wise modular assembly model of cyanobacterial PSII.

The step-wise assembly model is adapted from Nixon *et al.* (19). The number 48, 28 and 27 refer to assembly factors Ycf48, Psb28 and Psb27 respectively. Single letters O, U, V represent the extrinsic proteins PsbO, PsbU and PsbV, respectively. Protein subunits are labelled underneath, and assembly intermediates are labelled on top. Low-molecular-mass subunits except PsbH are indicated with the undifferentiated yellow column. “CP47 sub.” and “CP43 sub.” refer to CP47 subcomplex and CP43 subcomplex respectively.

The currently accepted model of cyanobacterial PSII assembly is the “step-wise modular assembly model” (19, 101), based on evidence obtained from studies investigating *Synechocystis* sp. PCC 6803 mutants with one or more PSII subunits

knocked-out. In the absence of certain subunits, the *de novo* assembly of PSII is stalled, leading to the accumulation of partially-assembled PSII subcomplexes (45, 102, 103). Through purification and analysis of the composition of these subcomplexes, the assembly steps, involving incorporation of the core subunits (D1, D2, CP43 and CP47) into the corresponding assembly intermediates (RC, RC47, RCC1 and RCC2) have been elucidated (Figure 1-5).

According to the current model, PSII in cyanobacteria is assembled from pre-assembled sub-complexes or modules, in a step-wise manner. The assembly process is initiated from a subcomplex consisting of cytochrome *b*₅₅₉ (PsbE, PsbF) and D2, of which the cytochrome *b*₅₅₉ is an essential prerequisite for the accumulation of D2 (103). The D1 precursor, pD1, together with PsbI, then incorporates into the D2-Cytochrome *b*₅₅₉ subcomplex. The next intermediate form of D1, intermediate D1 (iD1) is formed through the C-terminal cleavage of pD1, mediated mainly by the protease CtpA (C-terminal protease). Notably, knock-out of PsbI does not prevent the accumulation of the RC complex, though its absence destabilises the later binding of CP43 (45). Assembly factors PratA and Ycf48 likely participate in the processing and stabilisation of pD1 (102, 104, 105). Ycf48 also physically interacts with Sll0933 (or PAM68 in *Arabidopsis*) which is important for the accumulation of CP47 and CP43, respectively (102, 106).

After the formation of the RC complex, a CP47 subcomplex comprising PsbH and some other low molecular mass (LMM) subunits is then incorporated (107) to form the next assembly intermediate, RC47 (108). Assembly factor Psb28 (or Psb28-1, to differentiate from Psb28-2 encoded by *slr1739*) has been found to associate with RC47 from the cytoplasmic side (109, 110) and occupies a position proximal to the heme of cytochrome *b*₅₅₉, the Q_B site and the cavity on PSII where phycobilisome inserts (111).

Interestingly, it has been demonstrated that RC47 is capable of driving electron flow from tyrosine Y_Z to Q_A but is incapable of oxidation of water due to the absence of the Mn₄CaO₅ cluster (112).

Subsequently, a CP43 subcomplex composed of CP43, PsbK, PsbZ and Ycf12 (107, 113) binds to RC47 to form the monomeric PSII complex, RCC1. During this step, Sll0933 stabilises the CP43 subcomplex. The assembly factor Psb27 also binds to CP43 from the luminal side (113, 114), and this occurs after the dissociation of Psb28 from RC47 (111). Subsequent dissociation of Psb27 is needed for attachment of extrinsic subunits PsbO, U and V (115-117). Although knockout of Psb27 in *Synechocystis* does not abolish photosynthetic activity, the recovery of PSII from photoinhibition, as well as the photoactivation of PSII, is significantly affected in the absence of Psb27. It is conceivable that Psb27 aids the photoactivation by excluding extrinsic proteins to allow for efficient access of manganese to form the Mn₄CaO₅ cluster. Although not observed in crystal structures, CyanoQ also binds to PSII after the dissociation of Psb28 (50, 52), and it is likely that this occurs prior to the attachment of PsbU and V (52). *De novo* assembly of dimeric PSII (RCC2) is completed with the dimerization of monomeric PSII (RCC1).

1.3.3. The FtsH-mediated PSII repair model

Oxygenic photoautotrophs have evolved a sophisticated repair mechanism to maintain the functionality of PSII. Upon damage, D1 is rapidly removed from PSII and replaced with a new copy. This involves the partial disassembly of PSII, selective and synchronised replacement of D1, and the subsequent reassembly of PSII. During this process, the FtsH protease complex plays a crucial role. FtsH is a type of membrane-anchored ATP-dependent metalloprotease (118, 119). It is highly conserved, with different numbers of homologues in various species, including four in *Synechocystis*.

According to the “FtsH-only model” (Figure 1-6) applied to *Synechocystis*, light-induced damage of D1 triggers the partial disassembly of PSII. Subsequently, the damaged D1 subunit is selectively degraded by the FtsH2/3 protease complex (120) possibly from the N-terminus (121). The degradation of damaged D1 is synchronised with the insertion of a newly synthesised copy of D1. It is speculated that the subsequent reactivation of the RC47 complex is similar to the corresponding steps in PSII *de novo* assembly (Figure 1-5). Once a new D1 is incorporated into RC47, the CP43 subcomplex is reattached to form the PSII monomer, and the Mn_4CaO_5 cluster is subsequently assembled in a process known as ‘photoactivation’ and shielded by extrinsic proteins. The process is completed with the formation of a functional PSII dimer.

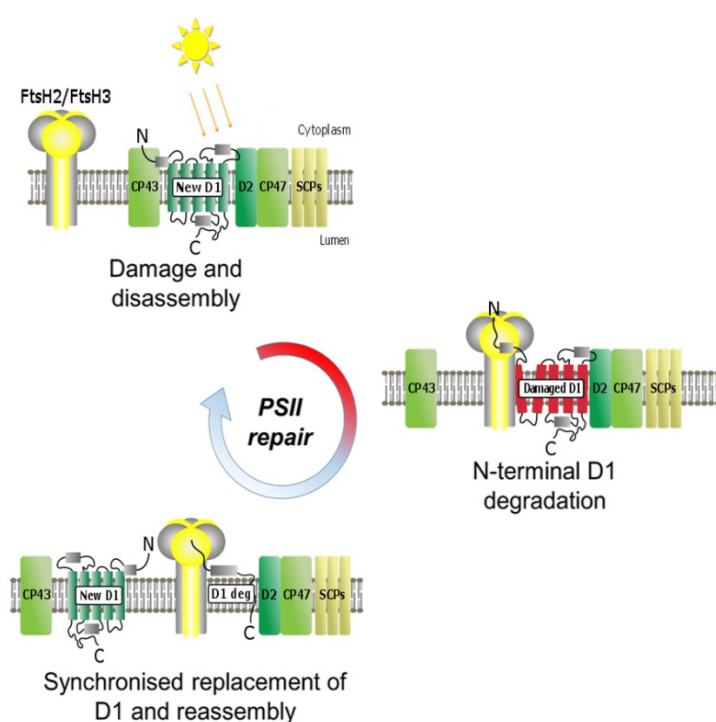


Figure 1-6 Current model of PSII repair adapted from Nixon *et al.* (19).

SCPs represent small CAB (chlorophyll a/b binding)-like proteins.

1.4. Project scope

Why PSII has not evolved to give rise to a PSI like reaction centre complex in which the antenna and reaction subunits are fused is unclear. Given the recent model for PSII

repair (19) and the evolutionary relationship between PSII and PSI (122-125), it is hypothesised in this thesis that the maintenance of separate CP43 and D1 subunits confers an evolutionary advantage for prompt and economic PSII repair, and this hypothesis is tested.

In this work, the physiological importance of the detachment of CP43 from PSII complex in the repair process of PSII has been studied, and its evolutionary significance is discussed (see Chapter 3 for a detailed introduction and discussion). Moreover, based on a phylogenetic analysis of the FtsH protease family across the tree of life (see Chapter 4), it is proposed that the FtsH-mediated repair mechanism might have been influencing the evolution of oxygenic photosynthesis from an early stage. In addition, the structural conservation and potential interacting sites of Psb29/THF1, a protein solely found in oxygenic phototrophs, and recently discovered to be an interacting partner of FtsH proteases involved in PSII repair (73), is discussed.

Chapter 2. Materials and methods

2.1. Biological materials

2.1.1. *E. coli* strain and growth condition

The *E. coli* strain DH5 α was used to propagate plasmid DNA. Cells grown on LB 1.5% (w/v) agar plates were incubated in a static incubator (Astell Hearsen, UK) with the temperature set to 37°C. Cells in liquid Luria-Bertani (LB) medium (1% (w/v) NaCl, 1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract) were grown on an orbital shaker (Innova 4400 incubator-shaker, New Brunswick Scientific, UK) placed inside the static incubator with the speed set to 200 rpm. 100 μ g/ml of ampicillin was used for the selective growth in either solid or liquid medium.

2.1.2. Oligonucleotide primers used in this work

Oligonucleotides were purchased from Sigma-Aldrich, UK. For general PCR, the final concentration of primer in a 20 μ l reaction is 1 μ M. Table 2-1 lists the oligomers involved in this work.

Table 2-1 Oligonucleotide primers used in this project

#	Name	Sequence (5' - 3')	Notes
P1	P1 (CF)	ACGGCCAGTGAATTCGAGCTGCTAGCTACG CAAGAGGATTTG	CP43-D1 fusion plasmid sequencing
P2	P2	GGATCGAGCTCTTAACCGTTGACAGCAGG	
P3	P3	CTATGACCATGATTACGCCAAGCTGTGCGAC TGCAAGATTGATAGACAGAG	
P4	I3	TGGGTTGTGGTGCTCTGTTA	
P5	I4	CCACTTTGTCCTTGGCTTCT	
P6	CR	CCACCACCTCCGTCGAGGTCAGGCATGAAC	
P7	S7	ATACTACCAACTGGTAAGGA	
P8	I5	TGGTAACCTCCTCCTTGGTG	
P9	S11	GGCATTGCGTTCGTGCATTACTTC	

P10	S6	ACTTTGTTTATAGGGCGACTG	split-fusion
P11	S2	AGCAGATTACGGTGACGATC	
P12	S4	CCGGTTAACTTTCGGTCGGT	
P13	InsF	CACGTTGTTTCATGCCTGACCTCGACTAAGG AGGTGGTGGATCCACAACGAC	
P14	InsR	GTCGTTGTGGATCCACCACCTCCTTAGTCG AGGTCAGGCATGAACAACGTG	
P15	PoiF	CACGTTGTTTCATGCCTGACCTCGACTAAGG AGGTGGTGGAAATGACAACGACTCTCCAACAGCGCG	
P16	PoiR	CGCGCTGTTGGAGAGTCGTTGTCATTCCAC CACCTCCTTAGTCGAGGTCAGGCATGAACAACGTG	
P17	DelF	CACGTTGTTTCATGCCTGACCTCGACTAATG ACAACGACTCTCCAACAGCGCG	
P18	DelR	CGCGCTGTTGGAGAGTCGTTGTCATTAGTC GAGGTCAGGCATGAACAACGTG	
P19	SubF	CACGTTGTTTCATGCCTGACCTCGACTAATT ATAACCAAATGACAACGACTCTCCAACAGCGCG	
P20	SubR	CGCGCTGTTGGAGAGTCGTTGTCATTTGGT TATAATTAGTCGAGGTCAGGCATGAACAACGTG	qPCR
P21	psbBF	GTACCGGTGAATCCGCATTG	
P22	psbBR	TGGCCCGTCAGACCATAGG	
P23	psaBF	CTCCAACCGAAGTTCCGTCC	
P24	psaBR	CAACCAACGTGTTGACCCC	
P25	psbA23F	TTCGGTACCTTGATGATC	
P26	psbA23R	ACCAGAGATGATGTTGTTA	
P27	psbCF	CCCAGGTGTCGTATACAC	
P28	psbCR	ACCAGATGACCAACATCA	
P29	psbDF	GTTCGAAATTTCCCGTCTG	
P30	psbDR	GGGTTCAAAGTCCAGTTG	
P31	rps1F	GCGTCTTCAATGTCAATG	
P32	rps1R	TGGTGGAAAGGGAAATAC	
P33	1638F	AGATACGGTTGGAAACCATGGCT	suppression mutation candidate <i>slr1638</i>
P34	1638R	TCCTTAGCCATGAACCGTCAG	

P35	F-inA3	ATCAAGAATACGGCGGTGGC	structural variation check; only present in strain S4
P36	RywlC	TCAATCCCCGTCAAGACCAGAC	structural variation check; plus P37, positive in strain S3
P37	F-spec	ATCCAGCTAAGCGGAACTG	structural variation check; plus P1, positive in strain S4
P38	RpsbA3	TGACATCGACGGTATCCGTG	near <i>psbA3</i> locus
P39	RspeA	TCCGAGGATTTGAGCGACTG	structural variation check; from <i>speA</i> upstream
P40	F-insSP	ATGTCGGTTGGTTCGGTACC	
P41	F-chlR	ATGAAAGACGGTGAGCTGGTG	structural variation check; plus P39, positive in strain S6
P42	F-A2UP	ACCCAGGGACAATGTGACC	structural variation check; plus CR, positive in strain S12
P43	CP43 fwd (P1)	ACGGCCAGTGAATTCGAGCTGCTAGCTACG CAAGAGGATTTG	Primers used to construct the pSS1 fusion plasmid
P44	Lfcp43 rev end	GTTCATGCCTGACCTCGAC	
P45	D1 fwd	TGACCTCGACGGAGGTGGTGGATCCATGAC AACGACTCTCAA	
P46	D1 rev	GGATCGAGCTCTTAACCGTTGACAGCAG	
P47	Gent fwd	CAACGGTTAAGAGCTCGATCCTGTTCGCGC AGGC	
P48	Gent rev	TCAATCTCGAGATCCTAGAAGATTCACCAC GTCACTA	
P49	RF fwd	CTTCTAGGATCTCGAGATTGAGACTTTTCT GATTTTGC	
P50	RF rev	CTATGACCATGATTACGCCAGTCGACTGCA AGATTGATAGACAGAG	

2.1.3. Plasmids used in this work

This original plasmid containing the CP43-D1 fusion protein encoding gene was constructed by previous lab member Catherine Hogg. Briefly, four fragments containing the *psbC* with its upstream-flanking region, the *psbA2* without start codon but a flexible linker, a gentamicin-resistance cassette and the *psbC* downstream-flanking region were amplified separately by PCR. The gentamicin resistance cassette was amplified from a pBS plasmid using primers P47/P48. The other three fragments were amplified using WT-P genomic DNA as the template and primers P43/P44, P45/P46 and P49/P50, respectively. These fragments along with a pUC19 vector digested with SacI and HindIII were connected via In-fusion PCR using In-fusion Cloning Kit (Clontech, UK). The resulting plasmid is designated as “pSS1” after sequencing confirmation. In addition, four types of mutations were introduced into pSS1 using site-directed mutagenesis strategy. The four plasmids are pSS1-Sub, pSS1-Poi, pSS1-Ins and pSS1-Del, corresponding to the transgenic strains Sub, Poi, Ins and Del, respectively (Table 2-2), details of these mutations are explained in Figure 3-19.

2.1.4. Cyanobacteria strains and growth conditions

2.1.4.1. *Synechocystis* strains

Table 2-2 *Synechocystis* sp. PCC 6803 strains used in this study

#	Strain	Description	Source /Reference
S1	WT-P*	Wild-type strain cultivated in Prof Peter Nixon's group at Imperial College	This work; sequenced (67)

	*In this work, all “WT” controls refer to the strain WT-P, except for the His-tagged PSII control, in which case “WT” PSII is purified from a CP47 C-terminal His-tagged strain derived from WT-P.		
S2	ΔD1	A <i>psbA</i> triple deletion strain	Nixon <i>et al.</i> (126)
S3	CP43-D1	ΔD1 transformed with the pSS1 construct	This work
S4	FuBH	A suppressor strain derivated from CP43-D1	This work
S5	ΔPsbA	A <i>psbA</i> triple deletion generated using a marker-less strategy	Nagarajan <i>et al.</i> (127)
S6	CP43-D1/ΔPsbA	ΔPsbA transformed with the pSS1 construct	This work
S7	CP43-D1/ΔpsbA/ΔFtsH2	FtsH2 knockout in the CP43-D1/ΔPsbA strain	This work
S8	ΔFtsH2	FtsH2 knockout in the WT-P strain	Silva <i>et al.</i> (21)
S9	Sub (Sub1)	ΔPsbA transformed with a modified version of the pSS1 construct, see Results Chapter.	This work
	Sub61sp	A suspicious suppressor strain that needs further characterisation	This work
S10	Poi (Poi1)	ΔPsbA transformed with a modified version of the pSS1 construct, see Results Chapter.	This work
	Poi51sp	A suspicious suppressor strain that needs further characterisation	This work

S11	Ins (Ins5)	Δ PsbA transformed with a modified version of the pSS1 construct, see Results Chapter.	This work
S12	Ins-SP (Ins70sp)	A suppressor strain derivated from the Ins strain	This work
S13	Del (Del5)	Δ PsbA transformed with a modified version of the pSS1 construct, see Results Chapter.	This work
S14	Del-SP (Del61sp)	A suppressor strain derivated from the Del strain	This work
S15	Δ Psb29	A <i>psb29</i> knockout in the WT-P strain	The Δ Psb29camA strain as in Beckova <i>et al.</i> (73)
S16	Δ Psb29- SP-b	A suppressor strain derivated from Δ Psb29	This work
S17	Δ Psb29- SP-c	A suppressor strain derivated from Δ Psb29	This work

2.1.4.2. Routine growth conditions

Synechocystis sp. PCC 6803 strains were cultivated on BG11 1.5% (w/v) bacto-agar plates or grown in liquid BG11 media (BG11 basic mineral medium supplemented with 0.3% (w/v) sodium thiosulphate, 10 mM N-tris[hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES-KOH) pH = 8.2) and supplements (Table 2-3) were added where applicable. Plates were restreaked every two to four weeks. Liquid culture below 100 ml was grown in sterile, filter-capped tissue culture flask. Flasks were placed

in an orbital shaker incubator with speed set to 120 rpm. Liquid culture above 100 ml was grown in a glass vessel, stirred by a magnetic stirrer and bubbled with sterile air filtered by a 0.2 μm pore filter (Midisart 2000, Sartorius Limited, UK). Cells were incubated in a temperature-controlled room set at 29°C and illuminated with 8 to 90 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ of white light provided by arrays of GroLux fluorescent light tubes (Sylvania, UK).

For long term storage, cells from plates or liquid cultures were suspended in 1 ml of liquid BG11 supplemented with 15% (v/v) glycerol and 5 mM glucose, flash frozen in liquid nitrogen and kept at -80°C.

Table 2-3 Supplements in BG-11 medium

Supplements	Final concentration	Description
Glucose	5 to 10 mM	Heterotrophic carbon source
Kanamycin	25-50 $\mu\text{g}/\text{ml}$	Selection pressure
Chloramphenicol	30 $\mu\text{g}/\text{ml}$	Selection pressure
Spectinomycin	50 $\mu\text{g}/\text{ml}$	Selection pressure
Erythromycin	15 $\mu\text{g}/\text{ml}$	Selection pressure
Gentamycin	5 $\mu\text{g}/\text{ml}$	Selection pressure
DCMU	10 μM	PSII inhibitor

2.1.5. Estimation of cell concentration of liquid *E. coli* and *Synechocystis* sp.

PCC 6803 cultures

The optical density of liquid *E. coli* culture was determined at 600 nm (OD_{600}), whereas the optical density of liquid *Synechocystis* sp. PCC 6803 culture was measured at 730 nm (OD_{730}). A 1-ml disposable cuvette was used with a Shimadzu spectrophotometer (UV-1601, Shimadzu, Japan). An OD_{730} of 1 for *Synechocystis* sp. PCC 6803 WT-P

culture corresponds to approximately 5.7×10^7 ml based on serial-dilution cell counting in this work.

2.2. DNA and RNA techniques

2.2.1. DNA transformation of cells

*2.2.1.1. Transformation of chemically competent *E. coli* cells*

For one transformation reaction, 10 to 50 μ l of chemically competent (128) *E. coli* was thawed on ice, gently mixed with 1 to 2 μ l (~ 1 μ g) of plasmid DNA and incubated for 10 min on ice. The mixtures were then heat shocked for 45 s at 42°C and incubated for 2 min on ice. 1 ml of liquid LB medium were added and the cells incubated for 1 hour under vigorous shaking at 37°C. Then 200 μ l of the mix was directly plated onto LB 1.8% (w/v) agar plates supplemented with a suitable antibiotic. The plates were incubated at 37°C overnight.

*2.2.1.2. Transformation of *Synechocystis* sp. PCC 6803*

Recipient cells grown in the liquid BG11 medium were harvested in exponential growth phase (OD_{730} around 0.4 to 0.8) by centrifugation (4,000 g, 15 minutes, 25°C) and resuspended in fresh BG11 medium yielding a cell suspension with a final OD_{730} around 5.0. For each transformation, 200 μ L of the concentrated cells were mixed with 1 to 10 μ g of recombinant plasmid DNA. This mixture was incubated at 29°C under constant white light illumination ($8 \mu\text{E} \cdot \text{m}^{-2} \text{s}^{-1}$) for 4 to 6 hours with occasional agitation. The transformation mix was then plated onto a 2 μ m cellulose nitrate membrane filter (Schleicher & Schuell MicroScience GmbH, Germany) on a BG11 1.5% (w/v) agar plate supplemented with 5 mM glucose. After 1 to 2 days, the filter was transferred to a new plate containing suitable supplement (s) for the selective growth. Transformation plates were kept under 8 to $20 \mu\text{E} \cdot \text{m}^{-2} \text{s}^{-1}$ of illumination unless high light intensity was

used as the selective pressure. Resistant colonies appeared after 10 to 12 days and were restreaked into single colony 2 to 3 times for segregation. PCR analyses were then performed to confirm the segregation of expected transformants.

2.2.2. Extraction and purification of nucleic acids

2.2.2.1. Plasmid DNA preparation from E. coli

Mini-scale plasmid DNA preparations from *E. coli* were performed using the QIAprep Spin Miniprep Kit following the manufacturer's protocol. Briefly, a single colony was picked and inoculated into 3 ml of LB liquid medium containing the appropriate antibiotic (s). The *E. coli* culture was then incubated at 37°C, 200 rpm shaking overnight. Cells were then harvested from 1.5 ml culture by centrifugation in a microfuge (model 5410, Eppendorf AG, Germany) at 1,300 rpm for 1min. The cell pellet was resuspended in 50 µl of P1 buffer (Plasmid-Midi-Kit, Qiagen, UK). After adding 50 µl of P2 buffer (Plasmid-Midi-Kit, Qiagen, UK), the sample was gently inverted 6 times immediately to lyse the cells efficiently without disturbing the genomic DNA. The mixture was then left at room temperature for 5 min, and 50 µl of ice-cold P3 buffer (Plasmid-Midi-Kit, Qiagen, UK) was added. After 6 times of gentle inversion, it was left on ice for 5 min. Then the precipitate was spun down in the microfuge at 1,300 rpm for 5 min. The supernatant was transferred to a new tube, and 1 ml of 100% ethanol was added to precipitate plasmid DNA. The sample was incubated at room temperature for 5 min, and then DNA was pelleted in the microfuge at 1,300 rpm for 10 min. The supernatant was removed, and 200 µl of 70% ethanol was added to wash the DNA. 1 min later, the ethanol was removed, and the DNA sample was air dried until all the ethanol had evaporated. Finally, 50 µl of Elution buffer (Plasmid-Midi-Kit, Qiagen, UK) was added to dissolve the plasmid DNA. The sample was stored at -20°C.

*2.2.2.2. Genomic DNA extraction from *Synechocystis* sp. PCC 6803*

High-quality genomic DNA extraction from *Synechocystis* was performed using the ZR Fungal/Bacterial Miniprep Kit (Zymo Research Co. USA) per the manufacturer's protocol.

2.2.2.3. RNA extraction

Before RNA extraction, all equipment and working bench were cleaned with RNaseZAP (Sigma Life Science, UK) to reduce RNase contamination. RNA extraction from *Synechocystis* was performed using the TRIzol reagent (Thermo Fisher Scientific, UK) and RNeasy Mini Kit (Qiagen, UK) according to a modified method (129). Briefly, TRIzol reagent was used to break cell and inactivate RNase activity. The RNeasy Mini Kit was used to purify RNA. After purification, RNase-Free DNase Set (Qiagen, UK) was used following manufacturer's protocol to remove potential DNA contamination.

2.2.2.4. Determination of nucleic acids concentration and quality

The concentrations of DNA and RNA were measured with NanoDrop (Thermo Scientific, UK). The 260/280 ratio represents the quality of nucleic acids. The pass threshold for the quality check is ~1.8 for DNA and ~2.0 for RNA. A low ratio normally suggests contamination.

2.2.3. Amplification and analyses of nucleic acids

2.2.3.1. DNA polymerase chain reaction (PCR)

The thermostable DNA polymerases Taq DNA Polymerase (New England Biolabs Limited, UK) were used in routine PCR amplification of DNA fragments of interest according to the manufacturer's protocol. The Failsafe™ PCR buffer (Cam-bio Limited, UK) was used to prepare PCR master mix. For general amplification purpose, the reaction consisted of an initial denaturation step at 98°C for 30 s and 30 cycles of 98°C

for 7 s (denaturation), 60°C for 15 s (annealing) and 72°C for 2 to 5 min depending on the size of the target product. The final extension step was performed at 72°C for 7 min. The PCR program consisted of an initial denaturation step at 94°C for 2 min and 30 to 35 cycles of 94°C for 10 s (denaturation), 55 to 60°C for 30 s (primer annealing) and 68°C for 1 min every 1 kb of target product (primer extension). The final extension step was performed at 68°C for 10 min.

2.2.3.2. Agarose gel electrophoresis

Agarose gel electrophoresis allows the separation of DNA fragments according to their sizes. An 0.7 to 1.5% (w/v) agarose gel, depending on the separation range of interest, was prepared with molecular grade agarose (Bioline, UK) dissolved in Tris-acetate-EDTA (TAE) buffer (40 mM Tris-acetate, 1 mM EDTA pH = 8.0) and 1:10,000 SYBR® safe stain (Life Technologies, USA). A DNA sample was mixed with 6x DNA loading buffer (40% (w/v) sucrose, 0.25% (w/v) Orange G), and then loaded into one well of the agarose gel. The molecular marker that used to determine the DNA fragments size is 2-log DNA marker ladder (New England BioLabs Limited, UK). The gel electrophoresis was carried out in TAE buffer at 100 to 130 V for 20 to 30 min in a horizontal mini gel system (PerfectBlue™ Mini Gel system; Peqlab Limited, Germany). DNA was visualised by the BioDoc-ITTM System (UVP; CA, USA).

2.2.3.3. Reverse transcription and quantitative PCR (qPCR)

1 µg of purified RNA was used as the template for cDNA synthesis with the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, UK), following the manufacturer's protocol. The reaction mixture of the qPCR included the Fast SYBR® Green Master Mix (Applied Biosystems, UK), 250 nM of each primer and 25 pg/µl of each template. Three technical replicates were applied. Two to three independent runs were repeated. The qPCR was performed on an ABI7500 Real-Time PCR Systems

(Applied Biosystems, UK). The following thermal cycling parameters were used: denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing and extension at 60°C for 1 min. Melting curve analysis was conducted using the default setting. Fold change in gene expression was calculated according to an improved $2^{-\Delta\Delta C_T}$ method (130).

$$\Delta\Delta C_T = (C_{T, \text{target gene, tested strain}} - C_{T, rpsI, \text{tested strain}}) - (C_{T, \text{target gene, WT}} - C_{T, rpsI, \text{WT}}),$$

The $C_{T, \text{target gene, tested strain}}$ is the cycle number of the amplification of target gene reaching a set threshold, using cDNA from tested strain as template. Five fragments (~ 150 bp) from *Synechocystis* genes *rpsI*, *psbA2,3*, *psbC*, *psbB* and *psaB* were selected for the quantification, and the amplification efficiency was validated before the calculation. The gene *rpsI* was chosen as the endogenous reference gene according to the literature (121).

2.2.3.4. DNA sequencing

General DNA sequencing was performed by Beckman Coulter Genomics Limited, UK. Samples were prepared following the instruction from the service provider.

2.2.3.5. Genome sequencing

For genome sequencing, a single colony of *Synechocystis* was inoculated and grown to liquid culture, and genomic DNA was extracted. Library preparation was performed by Ms Mia Jaffe from Professor Gavin James Sherlock's group at Stanford University. The Nextera DNA Library Prep Kit (Illumina, USA) was used to generate sequencing library, and the average size of all sequenced libraries was 516 bp.

2.3. Protein biochemistry techniques

2.3.1. Thylakoid membrane and PSII purification from *Synechocystis*

2.3.1.1. Small scale thylakoid membrane preparation

For each sample, 50 ml of liquid culture grown to exponential growth phase (OD_{730} between 0.6 – 0.8) was harvested by centrifugation at 9,000 g for 1 min. Harvested cells were resuspended in 500 μ l KPN buffer (40 mM K-phosphate, pH = 8.0, 100 mM NaCl) and mixed with 200 μ l of nylon beads (Sigma-Aldrich), Cocktail EDTA-free protease inhibitor (Roche) was added to the KPN buffer according to manufacturer's instruction. Cell-beads mixture was then disrupted by strong vortex for 1 min following 1 min ice incubation, and the disruption was repeated 3 times. Unbroken cells and glass beads were pelleted at 9,000 g for 1 min, and the supernatant was then collected into a 1.5 ml centrifuge tube. Again, centrifugation at 9,000 g for 1 min and then the supernatant was transferred to another new 1.5 ml centrifuge tube to remove beads and unbroken cells. The crude thylakoid membrane was then pelleted from the supernatant by centrifugation at 13,000 g for 25 min. The thylakoid pellet was then resuspended in 50 μ l KPN buffer. All procedures were carried in a cold room in the dark. For solubilization, 1% n-dodecyl- β -maltoside (β -DM) was added to thylakoid samples. After 1 min incubation on ice, soluble fraction was collected by centrifugation at 13,000 g for 25 min. For solubilization, 1% n-dodecyl- β -maltoside (β -DM) was added to thylakoid samples. After 1 min incubation on ice, soluble fraction was pelleted by centrifugation at 13,000 g, 4°C for 25 min.

2.3.1.2. Large scale thylakoid membrane preparation and His-tagged PSII purification

Typically, 30 L cultures of *Synechocystis* were harvested in the logarithmic growth phase (OD_{730} between 0.7 and 0.9) by cell concentration in a crossflow filtration system

(Sartorius) and then centrifugation at 8,000 rpm for 10 min. Pelleted cells were then resuspended in 50 ml washing buffer (50 mM MES (2- (N-morpholino)ethanesulphonic acid)/NaOH, pH 6.5) with protease inhibitor (cOmplete, EDTA-free protease inhibitor cocktail tablet, Roche). Then, again, the cells were pelleted and washed in the same buffer. Cell disruption was similar to the small scale crude membrane preparation described above except using a bigger adapter for the bead-beater (88 mL) (Bio-spec products). The cells were broken in darkness using 8 pulses of 15 s each, with 5 min cooling intervals. Unbroken cells and beads were removed by centrifugation at 4,000 rpm for 5 min. Thylakoids were pelleted by centrifugation at 100,000 g (Ti70 rotor, Beckman) and resulting thylakoid pellet was resuspended in washing buffer to 1 mg/mL Chl *a* for solubilisation. A 10% stock of n-dodecyl- β -D-maltoside (β -DM) (Calbiochem) was added dropwise to the thylakoid membrane to a 1% final concentration. Extraction was performed on a rotator in the dark for 10 min at 0°C. Insolubilized membranes were removed by centrifugation at 150,000 g for 30 min at 4°C (Ti70 rotor, Beckman). Solubilized thylakoid membranes were mixed with Ni²⁺-NTA agarose resin (Qiagen) previously equilibrated with column buffer (50 mM MES/NaOH, pH 6.0, 25% (v/v) glycerol, 20 mM CaCl₂, 5 mM MgCl₂, 5 mM imidazole, 0.03% (v/v) β -DM). The solubilized extract was diluted to 0.15 mg Chl/ml with column buffer and mixed in the dark at 4°C for 1 hour with a certain volume of NTA resin. The resin was washed five times with 1.5 column volume of column buffer, and then His-tagged proteins were eluted by gradient imidazole-washing of concentrations 50, 100, 200 and 300 mM.

2.3.1.3. Chlorophyll extraction and quantification

Chlorophyll concentration of the liquid *Synechocystis* sp. PCC 6803 culture and isolated crude membrane isolation were measured, typically in duplicates. In the case of a liquid culture, 1 ml of cells was centrifuged (9,000 g, 1 min, room temperature)

and the pellet resuspended in 1 ml of 100% methanol. Crude membrane extract was diluted 200 times by adding 5 µl of the crude membrane to 995 µl of 100% methanol. Chlorophyll was extracted for 5 min and the sample subsequently centrifuged maximum speed, 1 min, room temperature. The sample was transferred to a 1-ml disposable cuvette and measured with a Shimadzu spectrophotometer (UV-1601, Shimadzu, Japan) calibrated with 100% methanol between the wavelengths of 600 and 800 nm. The chlorophyll a content was then estimated (131) using the formula: $(A_{666} - A_{750}) \times 12.61 = [\text{chlorophyll a}]$ in µg/ml.

2.3.2. Polyacrylamide gel electrophoresis (PAGE)

2.3.2.1. Sodium dodecyl sulphate gel electrophoresis (SDS-PAGE)

Unless specifically mentioned, thylakoid samples equivalent to 0.5 µg of chlorophyll a per lane for SDS-PAGE analysis were mixed with sample buffer (final concentration: 62.5 mM Tris/HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 0.02% (w/v) bromophenol blue and 5% (v/v) β-mercaptoethanol) and incubated at room temperature for 1 hour, then samples were centrifuged at 13,000 g for 1 min to remove insoluble fraction. The soluble fraction was loaded on a 12% Tris-Tricine gel pre-made according to the published protocol (132). Electrophoresis was driven by 80 to 120 V of voltage for 60 to 90 minutes. After electrophoresis, the gel was stained with Quick Coomassie Stain (Generon, UK) according to the manufacturer's protocol.

2.3.2.2. One-dimensional clear-native polyacrylamide gel electrophoresis (1-D CN-PAGE)

Electrophoresis was carried using the NativePAGE™ Bis-Tris/HCl/HCl pre-cast gel (Life Technologies, UK) or freshly prepared native gels (120). For the freshly prepared native gel, anode buffer is 50 mM Bis-Tris/HCl, pH 7.0; cathode buffer is 50 mM Tricine, 15 mM Bis-Tris/HCl, pH 7.0, 0.05% sodium deoxycholate, 0.03% n-Dodecyl

β -D-maltoside (β -DM). Samples were mixed with native sample buffer (final concentration: 50 mM Bis-Tris/HCl, pH 6.5; 7% glycerol, 0.01% Ponceau S). Electrophoresis was conducted at 8 mA per gel for 2 to 3 h at 4°C protected from light. Finished Native PAGE was imaged using LAS-3000 system (Fujifilm, Japan) and R670 filter was selected for chlorophyll detection. Colour image was taken by a common scanner.

2.3.2.3. *Two-dimensional CN-Native/SDS-PAGE*

After the native PAGE and images being taken, lanes of the native gel were cut into individual slices and incubated in the denaturing buffer (2% DTT, 3% SDS, 10% glycerol, 50 mM Tris/HCl, pH 7.0, 6 M Urea) for 45 to 60 min. After denaturation, the second dimension slice was assembled into a prepared 12.5% Tris-Tricine SDS-PAGE gel as described above, and the stacking gel was prepared after the second dimension was loaded. Electrophoresis was performed same as normal SDS-PAGE.

2.3.3. **Western blotting and semi-quantification**

After SDS-PAGE, gels were removed from the cassette, rinsed twice in distilled water and then blotted to PVDF membrane (pore size 0.2 μ m) using either the iBlot® semi-dry blotting system (Life Technologies, USA) following manufacturer's protocol or traditional wet transfer method as described below.

After the transfer, PVDF membrane was blocked for 1 h in 5% milk powder dissolved in PBS-T buffer (145 mM NaCl, 7.5 mM Na₂HPO₄, 2.5 mM NaH₂PO₄ and 0.1% Tween 20). The membrane was then washed 3 times, 10 min each, in PBS-T buffer. Primary antibody was diluted 5,000 to 10,000 fold in PBS-T buffer. Incubation with the primary antibody was carried out overnight at 4°C. Membrane was then washed 3 times, 20 min each, in PBS-T buffer to remove the primary antibody. Incubation with secondary

antibody was performed at room temperature for 1h. After that, the membrane was washed 3 times, 10 min each, in PBS-T and then washed another 2 times, 10 min each, in PBS buffer (145 mM NaCl, 7.5 mM Na₂HPO₄, 2.5 mM NaH₂PO₄). Enhanced chemiluminescence was then performed. Membrane was incubated for 1 min in a 1:1 mixture of ECL reagent A (100 mM Tris/HCl pH = 8.3, 0.4 mM p-coumaric acid (90 mM stock solution in DMSO), 2.5 mM luminol (250 mM stock solution in DMSO)) and ECL reagent B (100 mM Tris/HCl pH = 8.3, 5.3 mM H₂O₂) and visualised using LAS-3000 system (Fujifilm, Japan). Semi-quantitative densitometry was performed using ImageJ software.

2.3.4. Sample preparation for mass spectrometry

2.3.4.1. In-gel tryptic digestion

Separated protein bands of interest were excised from SDS-PAGE after staining and sliced into about 1mm cubes. Gel stain was removed by incubating gel pieces with 100 µl of 50 mM Ambic solution (ammonium bicarbonate, pH 8.4 adjusted with 20% ammonia) followed by 100 µl of ACN solution (acetonitrile) at room temperature for 5 min. Destained gel pieces were dried in a vacuum centrifuge and soaked in 10mM dithiothreitol (DTT) dissolved in 50 mM Ambic solution and incubated at 56°C for 30 min. After reduction, gel pieces were washed with ACN solution and dried in a vacuum centrifuge. Carboxymethylation was performed by re-swell dried samples in 200 µl of 55 mM iodoacetic acid dissolved in 50 mM Ambic solution and incubating at room temperature in the dark. After removal of iodoacetic acid solution, gel pieces were washed with 500 µl of 50 mM Ambic solution for 15 min. Gel pieces were shrunk again with 200 µl of ACN for 5 min and dried in a vacuum centrifuge. A working solution of sequencing grade modified trypsin (Promega, UK) was prepared according to the manufacturer's instruction. Dried gel pieces were re-swelled in 20 µl of working

solution containing in total 0.5 µg of trypsin and incubated at room temperature for 15 min. ACN buffer was added to the mixture to cover gel pieces, and the mixture was incubated at 37°C overnight. On the next day, the supernatant containing hydrophilic peptides was transferred to a new tube, 50 µl of 0.1% trifluoroacetic acid solution was added to the gel pieces (to halt the digestion) and incubated at 37°C for 10 min. Another 100 µl of ACN solution was added and incubated at 37°C for 15 min. The supernatant was pooled with the previous hydrophilic peptides supernatant. 0.1% trifluoroacetic acid solution treatment can be repeated to increase yield. The final solution was reduced to 10 to 30 µl in a vacuum centrifuge and ready for mass spectrometry analysis.

2.3.4.2. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

Mass spectrometry was performed at the CISBIO mass spectrometry core facility, managed by Dr. Paul Hitchen, at Imperial College London.

2.4. Physiological analyses

2.4.1. Growth experiment under light stress

2.4.1.1. Light stress

Synechocystis cells on plates were incubated in a temperature controlled room set at 29°C. Illumination was from white light fluorescent tubes. Different intensities were set by adjusting the number of working light tubes (controlling range 8 to 45 µE·m⁻² s⁻¹) or shortening the distance of exposure (controlling range 45 to 120 µE·m⁻² s⁻¹). The exact light intensity and temperature were confirmed by a QRT1 Quantitherm light meter (Hansatech Instruments, UK) before each experiment. Highlight in this work refers to 90 µE·m⁻² s⁻¹, medium light refers to 20 µE·m⁻² s⁻¹, and low light refers to 8 µE·m⁻² s⁻¹.

2.4.1.2. Serial dilution spot growth assay

For each strain, a single colony was inoculated to 10 ml of liquid medium. Cells were grown to exponential phase, OD₇₃₀ of 0.3 to 0.6 approximately. Cells were then diluted to OD₇₃₀ of 0.1 using the BG11 medium. 10 µl of such starting culture was added to 90 µl of BG11 medium to make to 10⁻² dilution. Likewise, 10⁻³ and 10⁻⁴ dilutions were prepared. 2.5 µl of each serial dilution was then dropped onto a BG11 agar plate with or without supplement. The plates were grown under specified conditions for at least 7 days before images were taken by a digital camera or dissection microscopy.

2.4.2. Physiological measurements

2.4.2.1. Fluorescence decay kinetics

Fast-induction kinetics of chlorophyll fluorescence in whole cells was performed on a JTS-10 pump-probe spectrometer (Bio-Logic, France) according to the manufacturer's instruction. Cells from liquid culture grew to late-exponential phase (OD₇₃₀ of ~ 0.9) were harvested and resuspended to OD₇₃₀ of 0.5 and treated with or without 10 µM of DCMU. Cells were dark-adapted for 15 min before the measurements. After the onset of a 300 µs of strong actinic flash, chlorophyll fluorescence was recorded in a logarithmic time series between 50 µs to 15s.

2.4.2.2. Oxygen evolution measurements

Oxygen evolution rate was measured using a Hansatech DW2 oxygen electrode managed by Oxy-Lab 1 system (Hansatech Instruments Limited, UK). The temperature of the reaction chamber was controlled by a Grant W14 circulating water bath (Grant Instruments Limited, UK). Before the measuring, the system was calibrated according to the manufacturer's instructions.

For the activity measurement of whole cells, liquid culture was normalised to OD₇₃₀ approximate 1, and 2 to 3 ml of liquid culture was kept separately in the dark or frozen for later measurements of the chlorophyll a concentration. 1 ml of liquid culture was examined for the oxygen evolution in the presence of 2 mM 2,6-dichloro-1,4-benzoquinone (DCBQ; Eastman Kodak Co., USA; in EtOH), a photosystem II QA electron acceptor and 1 mM potassium ferricyanide (K₃Fe (CN)₆), a DCBQ oxidising agent. Both reagents were added under stirring to the cell suspension to a total volume of 1 ml into the DW2 chamber. After the oxygen signal was stabilised for at least 30 s in the dark, cells were exposed to actinic light illumination at 2300 µE·m⁻² s⁻¹ for 30 s to 1 min, and oxygen evolution was recorded.

2.4.2.3. Photosynthesis light-saturation curve

A nonrectangular hyperbola-based model (133) was selected to analyse photosynthetic light response data using the pre-compiled datasheet by Lobo *et al.* (134):

$$P_N = \frac{\phi_{(I_0)} \times I + P_{gmax} - \sqrt{(\phi_{(I_0)} \times I + P_{gmax})^2 - 4\theta \times \phi_{(I_0)} \times I \times P_{gmax}}}{2\theta} - P_D$$

where: I – the photosynthetic light intensity; $\phi_{(I_0)}$ – the quantum yield at $I = 0$; P_N – the net photosynthetic rate; P_D – the dark respiration photosynthetic rate; P_{gmax} – the asymptotic estimate of the maximum gross photosynthetic rate; θ – the convexity (dimensionless); $I_{(50)}$ – the light saturation point at which $P_N + P_D = 50\%$ of P_{gmax} .

2.4.2.4. Photoinhibition assay

For each strain, 80 ml of liquid culture was grown under 10 µE·m⁻² s⁻¹ of light until OD₇₃₀ near 0.8. Cells were harvest and resuspended in fresh liquid medium to OD₇₃₀ of 1. Liquid culture was supplemented with or without a final concentration of 250 µM apramycin before the application of 220 µE·m⁻² s⁻¹ of highlight or 550 µE·m⁻² s⁻¹ of

extra-high light. The high-intensity light was from a homemade cool white LED panel. Environment temperature was controlled at 29 to 30°C. 25 ml of liquid culture was grown in sterile, air-filter capped tissue culture flasks on an orbital shaker at 166 rpm. Samples were taken at half an hour to 2-hour intervals for oxygen evolution or fluorescence measurements. Samples for thylakoid extraction were frozen immediately. Chlorophyll concentration of first and last samples was measured for assessing potential variation due to cell propagation.

2.5. Bioinformatic techniques

2.5.1.1. Genome de novo assembly and alignment

Genome de novo assembly was performed under the guidance of this published protocol (135), and the QUAST (136) toolkit was used to examine the quality of assemblies. Briefly, the sequencing quality was examined by FastQC. Short reads were assembled by Velvet (137) and VelvetOptimiser, and assembled contigs were reorganised by Mauve (138) using *Synechocystis* sp. PCC 6803 genome GenBank ID BA000022.2 as reference. For genome annotation, the ordered contigs were written into a single contig with ACT (139) and annotated with Prokka (140). These steps were proceeded using a pipeline script written with python and run on Ubuntu 16 operating system. Mauve was also used for genome comparison. Structural variations around the genes of interest were checked by Sanger sequencing.

2.5.1.2. Genome resequencing and SNP calling

Resequencing analysis was carried out by mapping reads to the genomes of wild-type and selected mutants using the Geneious10 software. The threshold of calling polymorphisms was set to coverage greater than 30 and variant frequency greater than

70%. Reported polymorphisms were manually checked between different background and mutation strains, and shortlisted mutations were checked using Sanger sequencing.

2.5.2. Evolutionary analyses of the FtsH protein family

2.5.2.1. Construction of *FtsH* sequence datasets

6427 protein sequences containing the M41 peptidase domain were retrieved from the Pfam 30.0 database (141) under the entry PF01434 on 14th Oct 2016. Sequences lacking the AAA domain (entry PF00004) were removed using the HMMER tool (142), yielding in total 6082 sequences belonging to 73 bacterial phyla and candidate phyla and 378 eukaryote species. It is noteworthy that the Pfam database is based on the manually and algorithmically curated UniProt Reference Proteomes database (www.uniprot.org/proteomes/), in which annotation errors, although rare, do exist. For example, we have previously found misinterpretation of the start codon in at least 9 entries of the Psb29 protein within the database. Besides, the Pfam database, as well as the HMMER method, is not sensitive to convergent evolution at the molecular level (143, 144). Therefore compositional bias cannot be excluded. Other than the above limitations, this dataset covered over 3100 genome-sequenced species spanning the domain Bacteria and Eukarya. The presence of FtsH homologues in Archaea was assessed in a local copy of 210 *Archaea* reference proteomes. Only one FtsH sequence, A0A0M0BK70, was found in “miscellaneous Crenarchaeota group archaeon SMTZ-80”, and its phylogenetic proximity to FtsH protein H1XNZ9 from *Caldithrix abyssi* DSM 13497 suggests that it is likely obtained via horizontal gene transfer (HGT) from bacteria. Overall, this dataset was considered as a comprehensive representation of FtsH homologues in the context of the tree of life. For the comprehensive dataset of cyanobacterial FtsH, a copy of 103 cyanobacteria reference proteomes was interrogated and downloaded from the UniProt Proteomes database on 10th Oct 2016. Similar in the

construction of the large dataset of FtsH, in total 417 sequences were retrieved for cyanobacteria by searching for AAA domain and an M41 protease domain.

2.5.2.2. Sequence alignment and phylogenetic analysis

The 6082 FtsH sequences were aligned using MAFFT version 7 programme with the “L-INS-I” setting applied (145). Gaps within the alignment were then removed by the TrimAl (146) tool using the “gappyout” strategy, 570 characters were retained after the trimming process. 417 cyanobacterial FtsH sequences were processed similarly, and 611 characters were retained in the final alignment. Phylogeny inference of the 6082 sequences dataset was carried out using the CIPRES Science Gateway super-computer server (<https://www.phylo.org/>). FastTree programme, an approximate maximum likelihood method, was adopted using the default setting for the inference. Phylogeny of the 417 cyanobacterial FtsH was inferred through ETE3 toolkit using PhyML method (145, 147), applied amino acid substitution model was JTT/GTR (Jones-Taylor-Thornton/Generalised Time Reversible), branch support adopted aLRT (approximate likelihood ratio test). The resulting unrooted trees were organised and beautified with iTOL (148). The phylogenetic tree shown in Figure 4-3 was inferred based on a manually collected dataset of 312 sequences from different types of phototrophs. Sequences were aligned, trimmed and submitted to PhyML 3.0 (147) for phylogenetic inference.

2.5.2.3. Structural conservation analysis

270 FtsH sequences were manually selected from the UniProt database covering 55 bacterial phyla or candidate phyla based on a recently inferred tree of life (149). Similar as the cyanobacterial FtsH dataset, this dataset was aligned, trimmed and inferred through ETE3 toolkit using MAFFT, TrimAl and PhyML. The resulting phylogenetic

tree was submitted to ConSurf server for the structural conservation analysis. FtsH structures with PDB code 2DHR and 3KDS were used as the model.

2.5.3. Conservation analyses of Psb29/THF1

2.5.3.1. Construction of Psb29/THF1 dataset

Psb29 sequences were retrieved by blasting Psb29 from *Synechocystis* (*sll1414* gene product) against UniProt KnowledgeBase Reference proteomes (<http://www.uniprot.org>). The cut-off threshold was empirically set to 1e-4 after manually examining the resulting hits. 103 records were from cyanobacteria, 84 from higher plants, 11 from green algae, 12 from red algae and one from a virus that infects the green alga *Chlorella* sp. strain NC64A.

2.5.3.2. Sequence alignment and phylogenetic analysis

Sequence alignment and phylogenetic analysis of the Psb29/THF1 dataset was similar to that of FtsH datasets as described above. Briefly, sequence alignment of the 211 sequences was performed using MAFFT version 7 programme with the “G-INS-I” setting applied (145). Gaps within the alignment were trimmed by TrimAl using the “gappyout” method (146), and then the alignment was subjected to maximum likelihood-based phylogenetic inference using PhyML (147). ETE3 toolkit (150) was used to automate the above process; the PhyML setting was “+G+I+F, 4 classes and aLRT (approximate likelihood ratio test) branch supports, default models JTT/GTR”. The resulted unrooted tree was visualised with iTOL (148). Subsets of 103 cyanobacterial and 84 plant Psb29 homolog sequences were clustered according to their phylogeny. The trimmed alignments used in the conservation analysis were subjected to identity and similarity calculations using MatGAT (151).

2.5.3.3. *Structural conservation analysis*

Similar as in the FtsH evolutionary analyses, the evolutionary conservation was analysed using ConSurf2016 server (152). The above MAFFT alignment was trimmed of columns containing above 90% gaps; columns corresponding to the chloroplast transit peptide domain of *Arabidopsis thaliana* THF1, predicted by ChloroP 1.1 Server (153), were also removed.

Chapter 3. Repair and the evolution of photosystem II

3.1. Introduction

3.1.1. Structural similarity between two types of photosynthetic reaction centres

Two types of photosynthetic reaction centres have been defined based on spectroscopy comparisons and structural studies (122-124, 154-158): The Type I reaction centre found in phototrophs within the bacteria phyla Chlorobi, Firmicutes and Acidobacteria; and the Type II reaction centre found in phototrophs within the bacteria phyla Proteobacteria, Chloroflexi and Gemmatimonadetes. Uniquely, cyanobacteria and the chloroplasts of photosynthetic eukaryotes possess two types of reaction centres, and PSII is the only reaction centre, a type II, that generates oxygen. Moreover, except the aphotic classes *Melainabacteria* and *Sericytochromatia* (159), cyanobacteria are near-monophyletic in the capability of oxygenic photosynthesis, while all other phototrophic bacteria are numerical minorities (based on the available genomes in public database) within the corresponding phyla. Despite the low sequence identity between Type I and Type II reaction centres, the two types of reaction centres have been widely accepted as descendants of a common ancestor (122, 158, 160). PSI and PSII are examples of Type I and Type II reaction centres, respectively. As shown in Figure 3-1, CP43 has 6 transmembrane helices, and their spatial arrangement resembles that of the N-terminal 6 transmembrane helices of PsaA (or PsaB), the core subunit of PSI. The arrangement of the 5 transmembrane helices of D1 resembles that of the C-terminal 5 transmembrane helices of PsaA (or PsaB). This structural similarity also applies to the situation between CP47-D2 and PsaB (or PsaA), the other core subunit of PSI (124, 125).

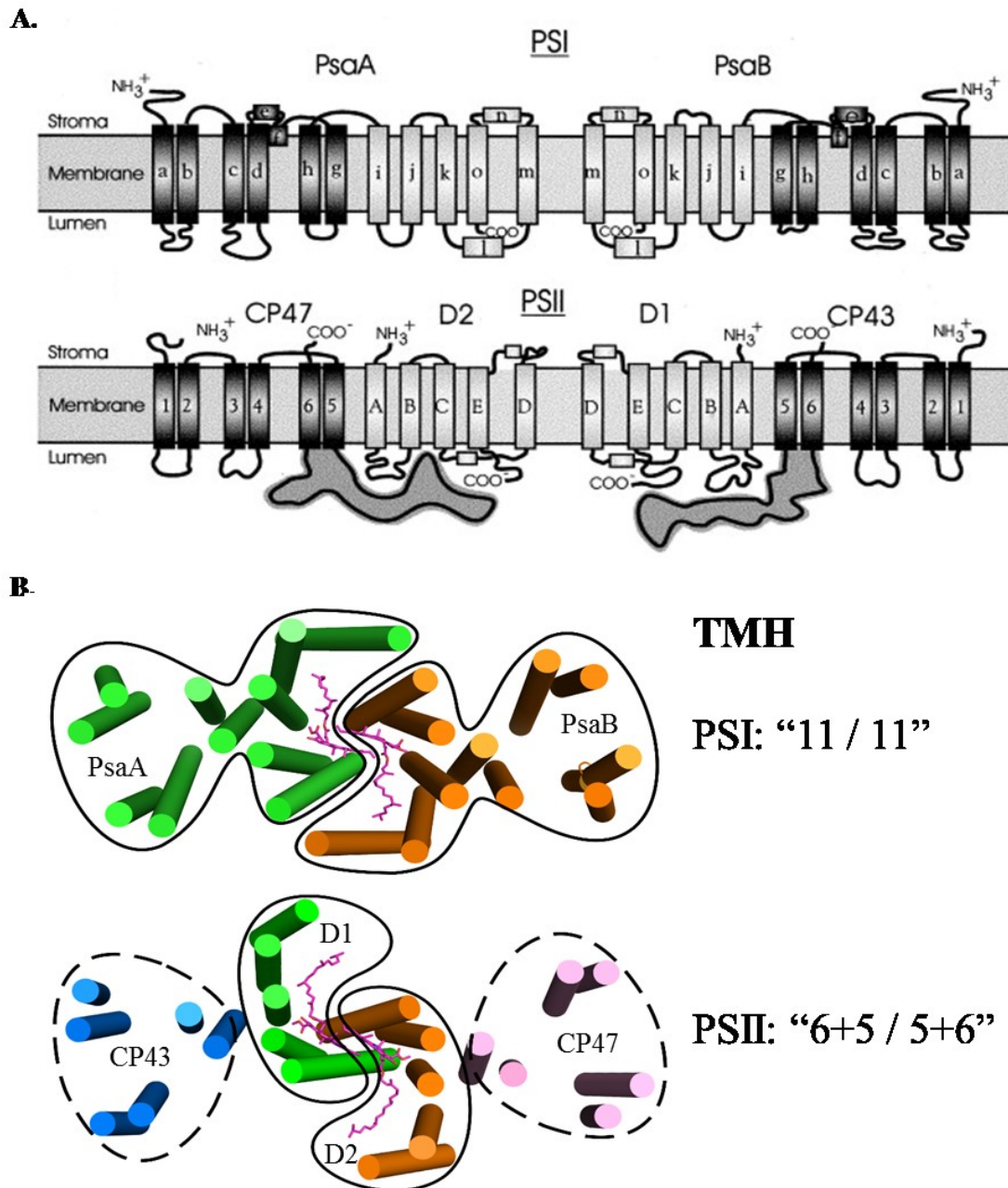


Figure 3-1 Comparison of the arrangement of PSI and PSII transmembrane helices.

A. A diagrammatic representation from Barber *et al.* ⁽¹⁶¹⁾ showing the arrangement of transmembrane helices of PSI and PSII. B. The spatial arrangement of the transmembrane helices based on crystal structures of *Thermosynechococcus vulcanus* PSII (PDB: 3WU2) and *Synechococcus elongatus* PSI (PDB:1JB0) ⁽¹⁶²⁾. Only transmembrane helices are shown for clarity. D1 and D2 are coloured same as PsaA and PsaB to indicate similarity. The special pair pigments in PSI and PSII are coloured in magenta.

3.1.2. The evolutionary controversy of two types of reaction centres

The evolutionary precedence of different types of photosynthetic reaction centres (123, 158, 163-169) and the timing of oxygenic photosynthesis (2, 160, 170-174) relative to the great oxygenation event (GOE) has long been a topic of intense debate. Although it

is widely accepted that Type I and Type II reaction centres descend from a common ancestor, there is not yet a consensus on which type of reaction centre can best resemble this speculated ancestor. Different scenarios have been proposed: a Type-I (163, 165, 167, 168) ancestor, a Type-II (164, 169) ancestor or an ancestor in-between the two types (123, 158, 166). It is known that the most significant burst of atmospheric oxygen, the GOE, is dated between approximately 2.4 to 2.35 Gya based on geochemical evidence (175, 176). Cyanobacteria, which possess two types of reaction centres, are believed responsible for the massive oxygenation of atmosphere starting from this period. Therefore, the divergence of two types of reaction centres at least predates the GOE. On the other hand, multiple lines of evidence have suggested the presence of oxygenic photosynthesis before the GOE, ranging from shortly before to hundreds of millions of years before the GOE (1, 174, 177). The time from the earliest oxygenic photosynthesis to the GOE might represent a crucial period when a competitive oxygen-evolving mechanism evolved. If this gap is minor, in other words, the earliest oxygenic photosynthesis appeared shortly before the GOE; then it would suggest that oxygenic photosynthesis is a consequence of fast evolutionary events such as the lateral acquisition of genes required for oxygen evolution (159, 178, 179). In the so-called “fusion theory”, it is believed that cyanobacteria obtained PSI and PSII from different reaction centre ancestors via horizontal gene transfer (170, 179). In this scenario, the Type I and Type II reaction centres must have undergone some independent evolution before they eventually meet in cyanobacteria to work on the generation of oxygen.

In contrast, if this gap is major, meaning the origin of oxygenic photosynthesis is as early as 3.8 to 3.5 Gya (180-183), it is then more reasonable to deduce that this primitive oxygenic phototroph is not competitive until GOE. Accordingly, the success of cyanobacteria would be most likely a consequence of “slow” vertical rather than “fast”

horizontal gene transfer. The gap between the emergence of oxygenic photosynthesis and GOE could be a period of slow optimisation of the primitive oxygenic machinery. The selective loss (160, 179, 184) and cyanobacteria origin (171, 185) theories favour the early origin of Cyanobacteria.

3.1.3. A potential role of repair in the evolution of oxygenic photosynthesis

Despite the difficulty in deducing the evolution of different reaction centres due to the extremely low sequence identity (158, 160), the repair mechanism of PSII might provide valuable information on the evolution of oxygenic photosynthesis. Firstly, the drastic photochemistry of water-splitting, a source of reactive oxygen species, strongly implies the necessity of some protective or repair mechanism co-evolving with PSII. Secondly, the sacrificial turnover (157) of the reaction centre subunit, D1, is a mechanism unique to PSII, comparing with all other types of reaction centres. In Type I reaction centre and anoxygenic Type II reaction centres, photodamage is prevented at an early stage and, once the reaction centre is damaged, the turnover of reaction centre subunit is much slower (186, 187). Thirdly, the FtsH-mediated repair mechanism of PSII might have evolved earlier than the extant PSII. It has been shown that, in the absence of CP47, D2 can be promptly degraded by FtsH (188), resembling the D1 turnover in the repair of PSII. The FtsH-dependent degradation of D1 and D2 strongly implies that such a repair process might have existed in the homodimeric ancestor of PSII (157). Moreover, it is predicted that primitive water oxidation likely evolved in such a homodimeric reaction centre (189), which, apparently, would render evolutionary advantage to a repair mechanism in this speculated homodimeric ancestor.

3.2. A CP43-D1 fusion PSII is assembled and active

3.2.1. Construction of the CP43-D1 fusion PSII

3.2.1.1. Design of the CP43-D1 fusion

Although the structural conformation of the N-terminal tail of D1 is not resolved in current PSII crystal structures (35-37), the accessibility of this region has been proposed to play a crucial role in the repair process of PSII (121, 190). In this work, by comparing the structural similarity between PSII and PSI, we hypothesise that the presence of CP43 and D1 as separate subunits reflects a selective advantage in replacing damaged D1 and so maintaining water oxidation. To test this hypothesis, the C-terminus of CP43 and the N-terminus of D1 were fused together with a single repeat of the classic flexible linker (Gly-Gly-Gly-Gly-Ser)_n (191-193) (codon: GGAGGTGGTGGATCC); the resulting fusion protein, in theory, imitates the architecture of PsaA or PsaB in the PSI complex (Figure 3-2). The PSII complexes containing this fusion subunit were then examined for functionality and repair efficiency.

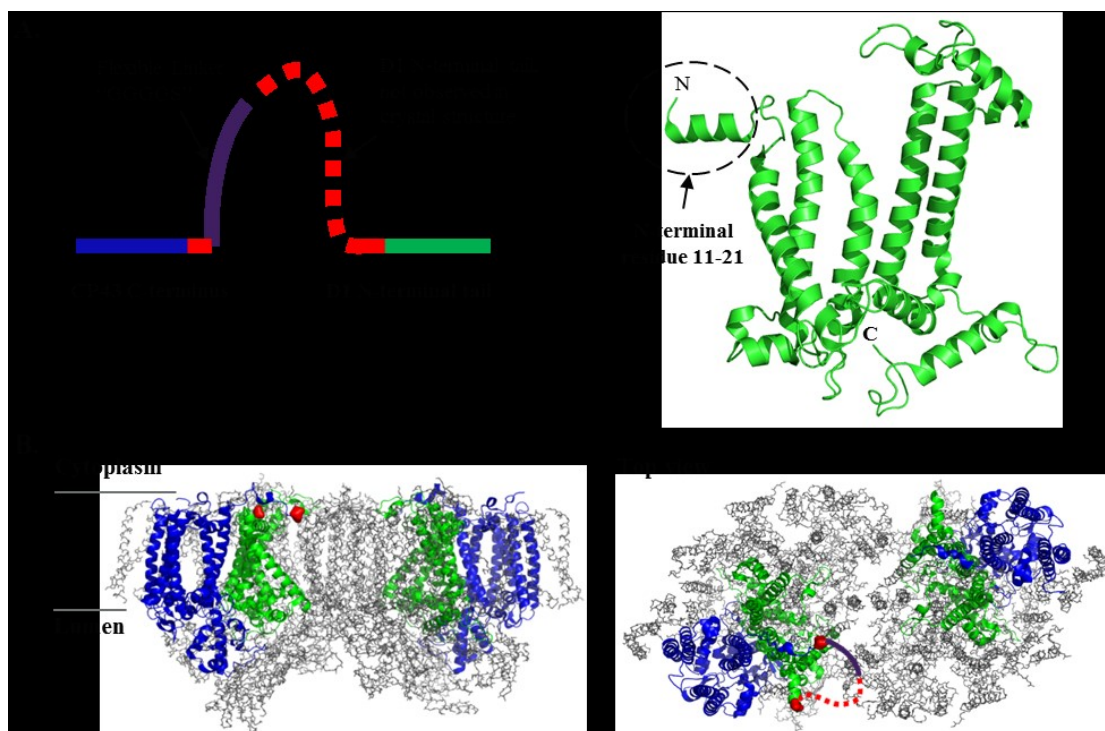


Figure 3-2 Structural view of the CP43-D1 fusion design.

A. The left panel is a diagram showing the fusion strategy. The right panel shows the approximate position of the N-terminal region of D1. B. The relative position between the C-terminus of CP43 and the N-terminus of D1. PSII structure model: PDB 3WU2.

3.2.1.2. Construction of the CP43-D1 fusion mutant

The CP43-D1 fusion plasmid was constructed by a previous group member Catherine Hogg. As shown in Figure 3-3B, the CP43-D1 fusion gene marked with a gentamycin-resistance cassette is flanked by the upstream and downstream region of *psbC* and inserted into the pUC19 vector via *SacI* and *HindIII* sites. The CP43-D1 fusion plasmid, here named pSS1, was used to transform a recipient *Synechocystis* strain (31, 127) in which all three paralogous *psbA* genes were deleted. Successful transformants were selected on plates supplemented with gentamycin and glucose, and the segregation of the CP43-D1 fusion mutant was confirmed by PCR (Figure 3-3C) and confirmation of

the correct fusion by Sanger sequencing of the PCR product from CP43-D1 fusion shown in Figure 3-3C.

In this work, two D1-null recipient strains were adopted in parallel, and these two strains were constructed independently and have different genetic backgrounds. The first D1-knockout strain (31), here designated as “ Δ D1”, was His-tagged at the C-terminus of CP47 which would allow the affinity purification of PSII from its transformant “CP43-D1”; the second D1-knockout strain (127) was constructed based on a marker-less strategy, which would facilitate further transgenic manipulations. To differentiate with the Δ D1 strain, this second marker-less D1-null strain was designated as “ Δ PsbA”, and its derivatives include the fusion strain “CP43-D1/ Δ PsbA”, the FtsH2-knockout fusion strain “CP43-D1/ Δ PsbA/ Δ FtsH2” and the break fusion strains described in Section 3.5.2. It is needed to mention that the strain CP43-D1/ Δ PsbA is the counterpart of strain CP43-D1, both the two strains are null in all three *psbA1*, *psbA2* and *psbA3* genes but differ in the genetic backgrounds.

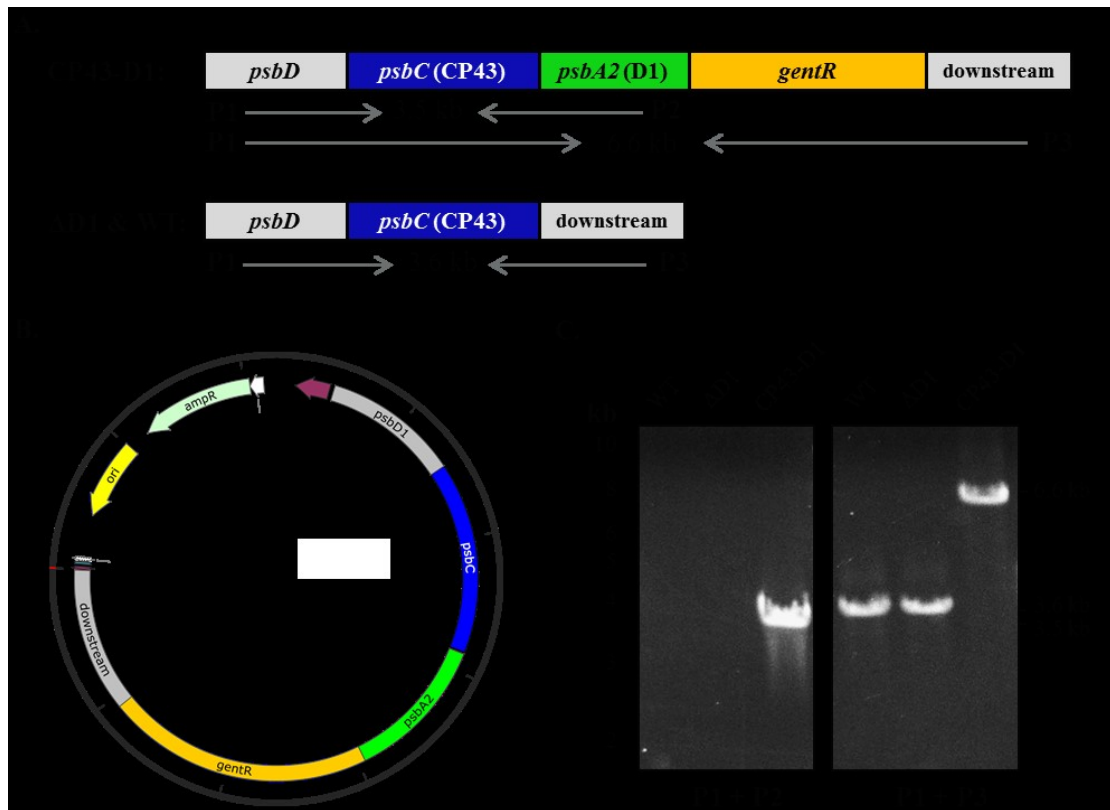


Figure 3-3 Plasmid map of the CP43-D1 fusion construct and the segregation of transformant.

A. A schematic view showing the location of the CP43-D1 insertion. The *gentR* represents the gentamicin-resistant cassette. B. Plasmid map of pSS1. C. PCR confirmation of the successful segregation of the CP43-D1 strain at the $\Delta D1$ background. P1, P2, P3 represent the primers used for the PCR screening, and their positions are indicated in Figure A. Size of target PCR fragment was indicated on the right.

3.2.1.3. The CP43-D1 fusion protein is expressed in *Synechocystis*

Immunoblotting experiments were performed to check the expression of the CP43-D1 protein and, if expressed, whether the CP43-D1 fusion protein accumulated in the fusion strain, or had been cleaved into separate CP43 and D1 subunits. The theoretical molecular mass of the CP43-D1 fusion protein is about 75 kDa. Thylakoid membrane proteins purified from the CP43-D1 fusion strain were separated by SDS-PAGE and probed with different antibodies. As shown in Figure 3-4, a clear D1 signal was detected at about 75 kDa in the thylakoid membrane sample of each of four independent CP43-D1 fusion mutants. No smaller D1 fragment was detected in the CP43-D1 fusion mutants. The CP43 antibody did not detect the '75 kDa' protein in the mutant thylakoid

samples, possibly because of the fusion interfering with the antigen at the C-terminus of CP43.

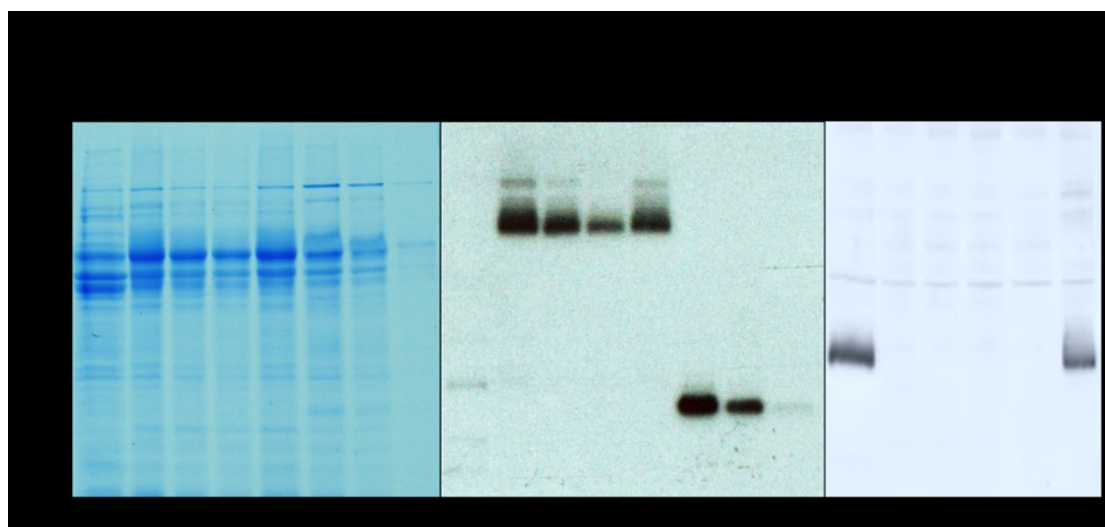


Figure 3-4 Western blotting of the CP43-D1 protein in thylakoid preparation.

Δ D1 is the negative control, WT is a positive control, a serial dilution of WT was loaded for a semi-quantification reference. CP43-D1-a, CP43-D1-b, CP43-D1-c and CP43-D1-d are four individual colonies selected for screening.

3.2.2. Purification and composition analysis of the CP43-D1 fusion PSII

3.2.2.1. Preparation of the CP43-D1 fusion PSII

PSII from the CP43-D1 strain was isolated using nickel affinity chromatography via the C-terminal six-histidine tag on CP47 and subjected to SDS-PAGE electrophoresis and immunoblotting. As shown in Figure 3-5, the CP43 antibodies were now able to detect the “75 kDa” fusion protein, probably due to increased abundance compared with the thylakoid samples (Figure 3-4). The extrinsic PsbO subunit was detected at similar levels in both WT and fusion PSII complexes, indicating assembly of this extrinsic subunit on the donor side of the fusion PSII complex.

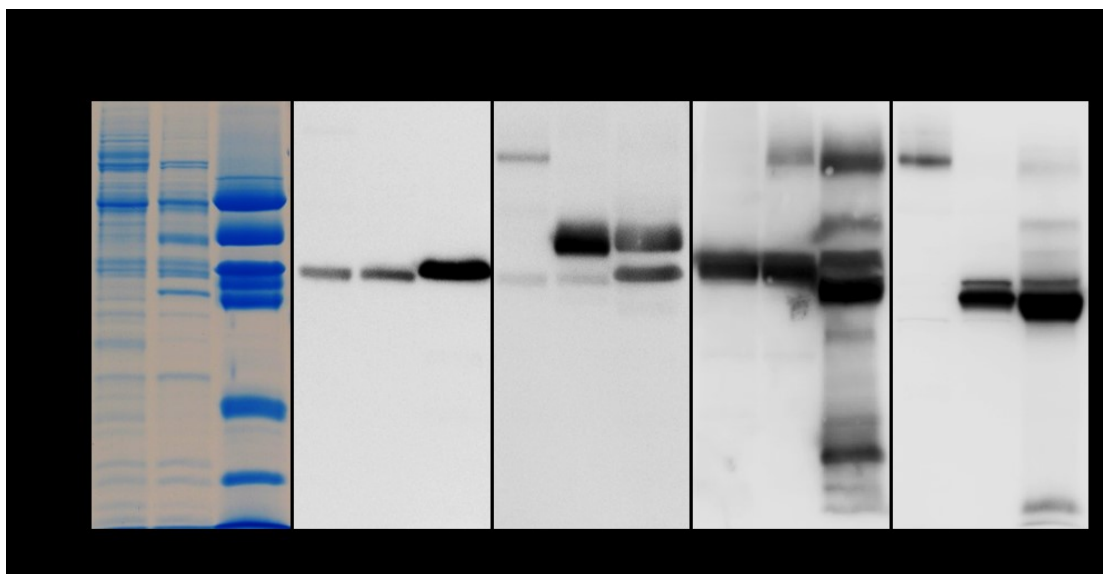


Figure 3-5 Western blotting of the CP43-D1 PSII.

Dimeric PSII preparation from *T. elongatus* (D-PSII) was included as a positive control. “Coomassie staining” is the SDS-PAGE separation of the purified PSII. “ α PsbO”, “ α D2” and “ α D1” are the Western blotting of PsbO, D2 and D1, respectively. “ α CP43, α PsbO” is the reprobe by CP43 after the western blotting of PsbO.

The protein composition of the His-tagged PSII sample was analysed by mass spectrometry (Supplementary Table 1). As shown in Figure 3-6, in both WT and CP43-D1 fusion strain, the major components, above 10 kDa, of PSII were all separated by SDS-PAGE. In band No.2 of the fusion PSII, both CP43 and D1 were detected and gave the highest score in all matches. Also, in the approximate location of CP43, bands No.6 and No.7, both gave a poorly-scored match of CP43, likely from minor CP43-D1 fragmentation. Similarly, no significant D1 match was found near 30 kDa in the fusion PSII.

By comparing the SDS-PAGE profiles of WT-PSII and fusion-PSII, several upregulated components, or contaminants, were detected in the fusion PSII preparation. The function of Sll1530 is not clear; Sll0018 plays a role in glycolysis as fructose 1,6-bisphosphate aldolase; Slr1051 is the enoyl-[acyl-carrier-protein] reductase FabI, an essential component in the synthesis of fatty acids, playing a role in the thermal stability of PSII (194). ClpC is the ATPase subunit of the ATP-dependent Clp protease.

Although it is interesting to find that CP43-D1 is also detected with ClpC which might indicate a quality control pathway of this fusion protein, the possibility of contamination cannot be excluded. FtsH2, FtsH3 and FtsH4 were detected in both WT and CP43-D1, probably because of physical interaction with His-tagged PSII (21). Psb29 was also found in the PSII preparation in both WT and CP43-D1. Although previously showed to copurify with PSII (195), Psb29 is now known to physically interact with the FtsH2/3 complex in *Synechocystis* (73); therefore its presence might be attributed to the presence of FtsH2/3 proteases. At the migration position corresponding to the WT CP43, two bands in CP43-D1 gave matches to Sll1214 and Sll1212, with theoretical sizes of 42.2 kDa and 41.3 kDa, respectively. Sll1214 is involved in the magnesium-protoporphyrin IX monomethylester (oxidative) cyclase reaction and is indispensable for chlorophyll synthesis (196). Sll1212, a GDP-D-mannose 4,6-dehydratase, is reported to negatively respond to a block in photosynthetic electron transport (197). CpcB and CpcA, the α/β subunits of phycocyanin, a soluble detachable antenna protein bound to dimeric PSII, were significantly higher in the fusion PSII. Ycf85, together with CmR (Chloramphenicol O-acetyltransferase, the gene product of the resistance cassette), were also detected. Ycf85 is a sulphate ABC transporter; its possible role in PSII is unknown. Psb27 was also found in the CP43-D1 preparation, very likely from the PSII assembly or repair intermediates lacking the extrinsic PsbO, PsbU and PsbV subunits (115, 117).

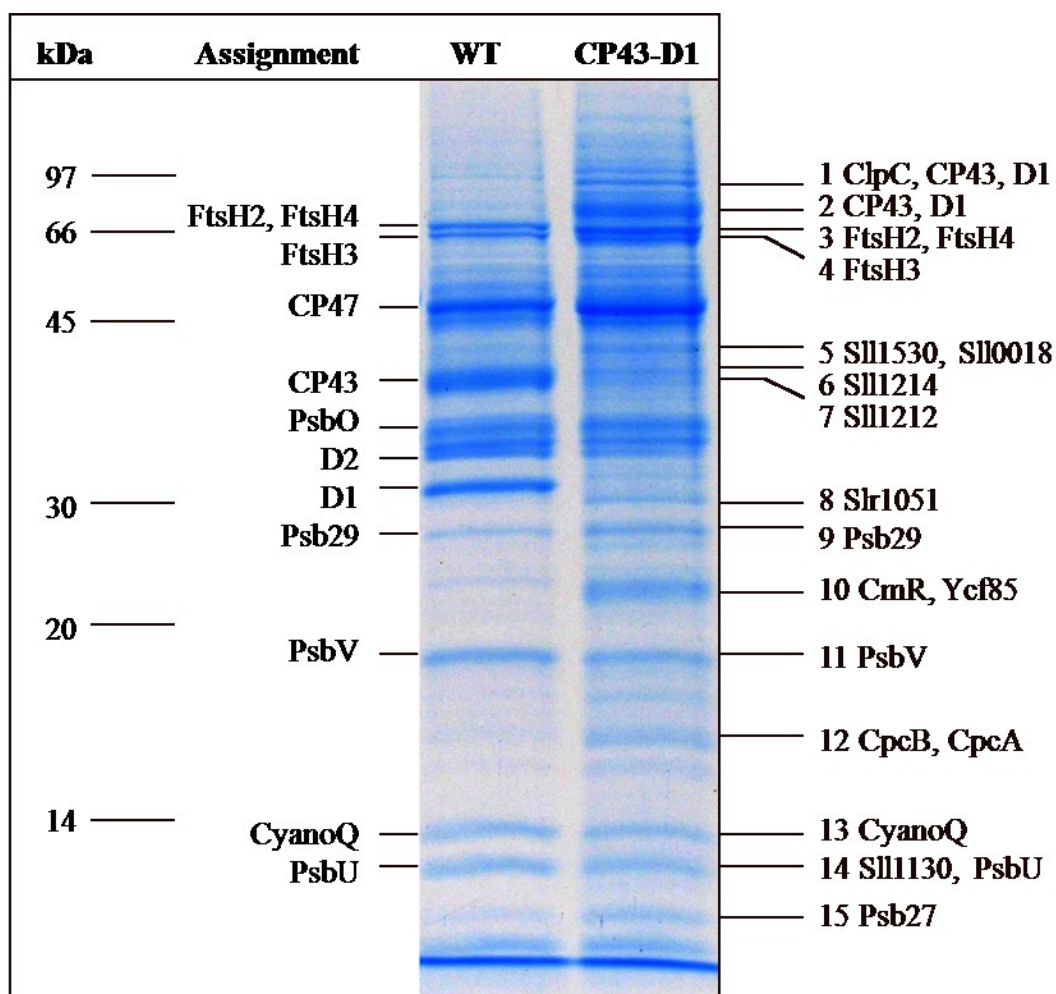


Figure 3-6 Components of the CP43-D1 fusion PSII analysed by mass spectrometry.

For the WT reference, components of PSII were confirmed by mass spectroscopy and western blotting, except that CP47 and Psb29 were assigned empirically⁽¹⁹⁵⁾. For Psb29, the corresponding band in CP43-D1 was confirmed by mass spectrometry. For the CP43-D1 fusion, top 1-3 scored matches were listed, a full list of all matches is in Supplementary Table 1. His-tagged PSII preparation was used in this analysis.

3.2.2.2. Dimerization of the CP43-D1 fusion PSII

Clear native gel electrophoresis (Figure 3-7A) revealed that both monomeric and dimeric PSII could be assembled in the CP43-D1 fusion strain. However, the ratio of monomer to dimer PSII in the fusion strain was slightly higher than that in WT. Also, a significant amount of unassembled or partially assembled proteins were detected in the CP43-D1 fusion strain, which might be due to less-efficient assembly or repair in the fusion strain. The unassembled fraction in the fusion strain gave a significant

fluorescence signal, suggesting the presence of chlorophyll-binding complexes, most likely unassembled His-tagged CP47.

2-D gels (Figure 3-7B) combined with mass spectrometric analysis confirmed that the fusion protein was incorporated into both the monomeric and dimeric PSII complexes. The accumulation of CP47 and PsbH, main components of the CP47 subcomplex during PSII assembly, in the unassembled or partial assembled fraction resembled the knockout of CP43 in *Synechocystis* (198). Sll1130, a negative regulator of thermotolerance (199), was found in the same fraction. Besides, two states of PSII monomer, differing in the attachment of CyanoQ, were clearly distinguished, with the similar ratio of the two states in WT and CP43-D1 strain suggesting unimpaired binding of cyanoQ to the fusion PSII complex. The previously mentioned CmR (Chloramphenicol O-acetyltransferase) was confirmed in the unassembled or partially assembled fraction and is most likely a contaminant. Overall, this evidence proved that the fusion of CP43 and D1 still allowed the assembly of PSII.

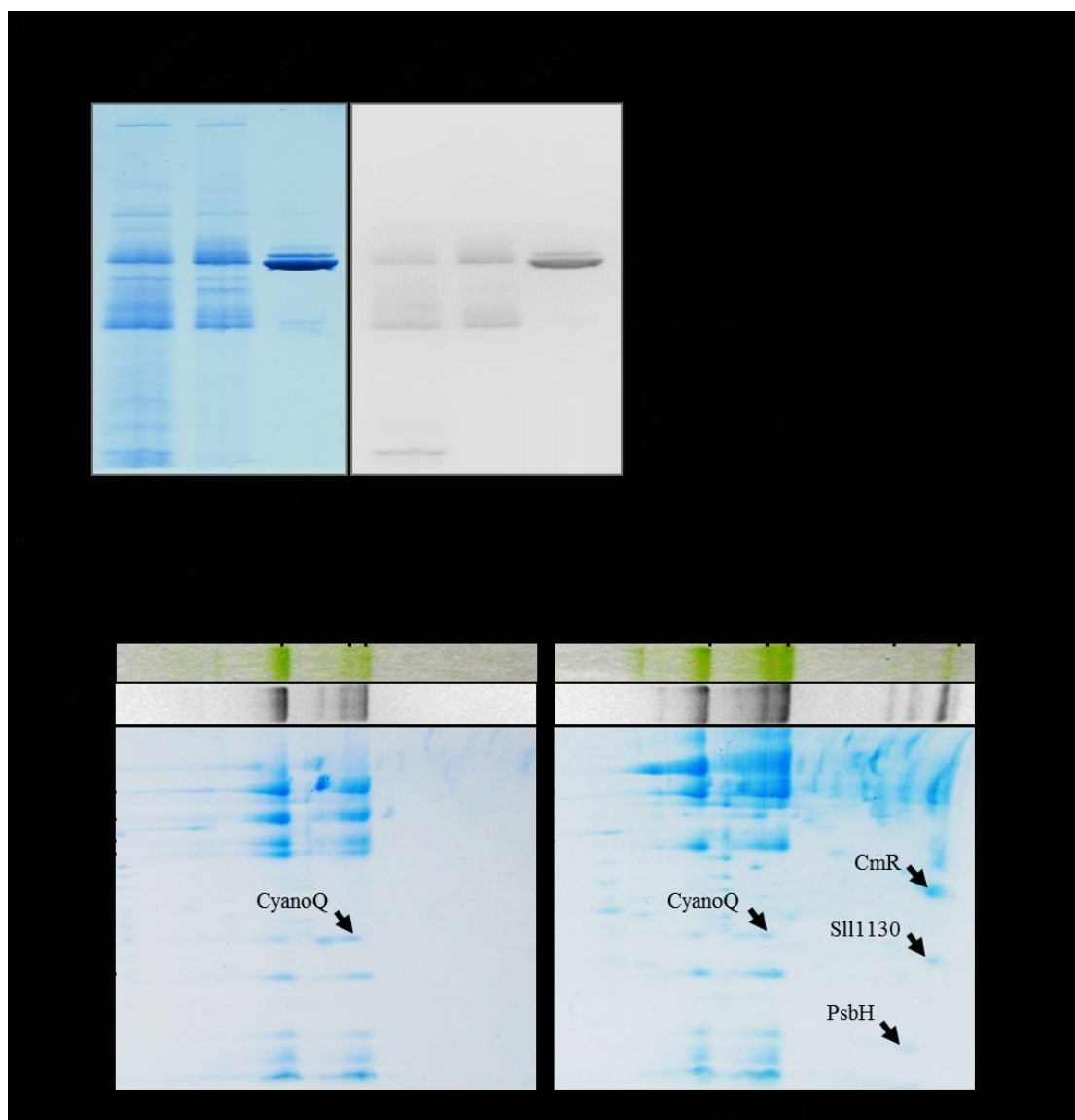


Figure 3-7 Clear-native and 2D PAGE analyses of the CP43-D1 fusion PSII.

His-tagged purified PSII was analysed. A. Coomassie staining visualises protein components while fluorescence detection above 670nm specifically visualises chlorophyll-binding components. B. Arrows indicate some of the components confirmed by mass spectrometry.

3.2.3. Characteristics of the CP43-D1 fusion strain

3.2.3.1. Photoautotrophic growth and sensitivity to highlight

The growth of the CP43-D1 fusion strain was examined to assess the activity of photosynthesis. A Δ FtsH2 mutant of *Synechocystis*, which is impaired in PSII repair (21), and the recipient strain Δ D1 were included as controls. As shown in Figure 3-8A, the CP43-D1 fusion strain was able to grow photoautotrophically at low to medium

light ($8 - 20 \mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$) but not at high light ($90 \mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$), indicating the photosynthetic machinery in the fusion strain is functional but less robust. For the liquid cultures (Figure 3-8B) grown photoautotrophically under normal light, the doubling time of the CP43-D1 fusion strain was approximately 24 hours, compared to about 12 h for WT. When glucose was supplemented, the growth rate of the CP43-D1 fusion strain was elevated to a level close to WT, which suggested that the major growth-limiting factor of the CP43-D1 fusion strain is the efficiency of photosynthesis. Overall, the growth assay revealed that the CP43-D1 fusion strain was able to perform photosynthesis at a repressed efficiency but was more susceptible to light stress.

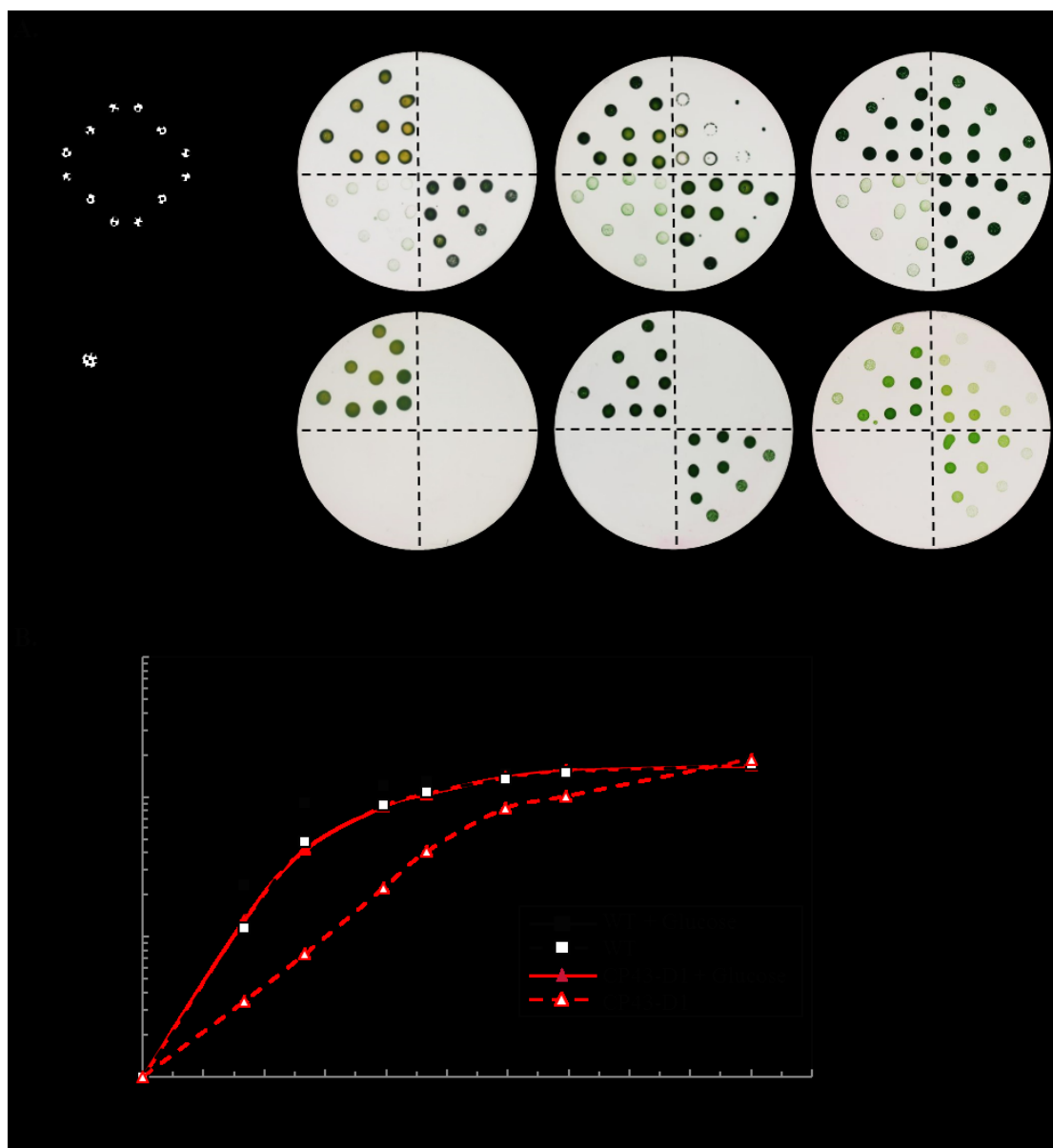


Figure 3-8 Growth assay and growth curve.

A. Each spot grew from a 5 μ l of serial diluted liquid culture, the initial OD₇₃₀ is 1. BG11 plates with or without additional carbon source were kept at different light intensity as labelled. B. All liquid cultures were grown under 30 μ E·m⁻² s⁻¹ of light.

3.2.3.2. *Oxygen evolution and electron transfer properties of the CP43-D1 fusion strain*

The functionality of the fusion PSII was assessed by comparing the oxygen evolution rate and electron transfer properties of PSII in the CP43-D1 fusion and WT strains. Measurements of oxygen evolution rate were conducted by me in two different labs, and similar results were observed. As shown in Figure 3-9A, oxygen evolution rate of

the CP43-D1 fusion strain was about $52.7 \pm 24.1\%$ of the WT activity. Light energy absorbed by photosynthetic apparatus can drive photochemical reactions that leading to the production of stable energy currency such as ATP and NADPH; excess light energy can also be dissipated as heat or re-emitted as fluorescence radiation (200, 201). Chlorophyll fluorescence has been widely used as an indicator of photosynthetic energy conversion since the discovery of the Kautsky effect (202). PSII absorbs light and drives the electron flow from water to plastoquinone. When the integrity of the electron transfer path inside PSII is altered, the energy flow towards photochemical reaction decreases and that towards fluorescence radiation increases. Besides, DCMU can block the energy flow from Q_A to Q_B hence the photochemical utilisation of light energy, supplement of DCMU therefore could maximise the chlorophyll fluoresce which is proportional to PSII. Chlorophyll fluorescence decay measurements (Figure 3-9B), in the presence of 10 μ M of DCMU, showed that the levels of PSII in the CP43-D1 fusion strain were reduced to 60-70% of WT level and that the rate of charge recombination between Q_A^- and the donor side of PSII (203, 204) were similar in WT and CP43-D1. The similarity in rates indicates that the Mn cluster was fully assembled in the fusion PSII complex as PSII lacking the Mn cluster shows much faster charge recombination rates (205). Figure 3-9C illustrates the fast and medium phase of fluorescence relaxation which are general indicators of the electron transfer from Q_A^- to the bound or newly recruited (from quinone pool) Q_B (91, 203, 206). Collectively, these data indicate that the PSII complexes in the CP43-D1 behave similarly to the WT PSII.

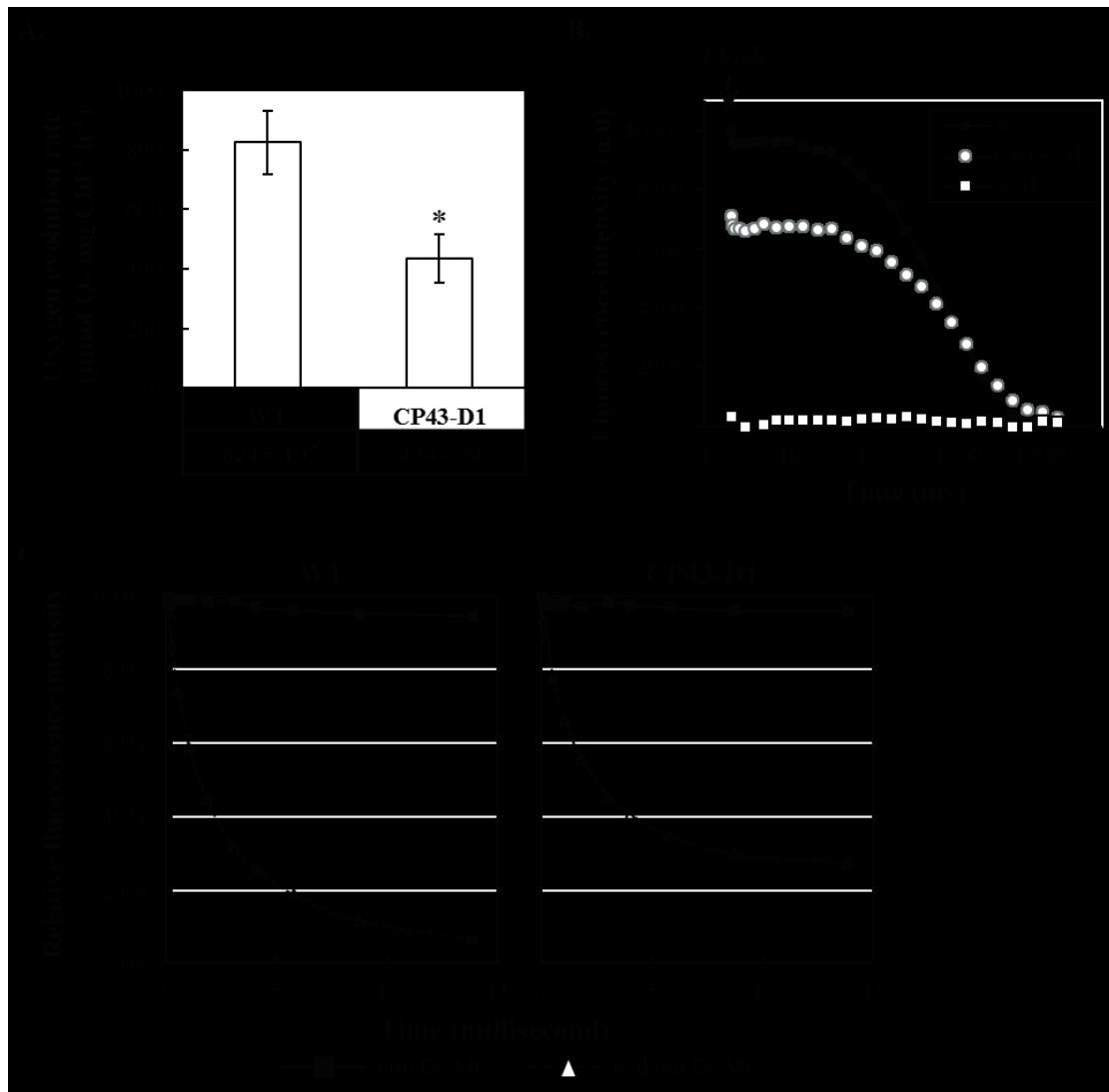


Figure 3-9 Oxygen evolution activity and electron transfer property.

A. Comparison of oxygen evolving activities of WT and CP43-D1. Data were from three biological replicates with each data point containing two to three technical replicates; error bar represents standard deviation. Significance is checked by a T-test. *, $p < 0.05$. B. A representative measurement of the fluorescence decay of WT and CP43-D1 in the presence of 10 μM of DCMU. Same amount of cells were used for measurements based on the same OD_{730} . Values of different strains were normalised to the same baseline. C. Fluorescence relaxation measurement in the presence and absence of 10 μM of DCMU. The duration of measurement is enough to comprise the fast and middle phases of fluorescence relaxation according to literature (91, 203, 206).

3.2.4. Characterisation of a parallel CP43-D1 fusion strain and its Δ FtsH2

derivative

3.2.4.1. PSII composition of the CP43-D1/ Δ PsbA and CP43-D1/ Δ PsbA/ Δ FtsH2

strains

Although the feasibility of fusing CP43 and D1 was confirmed in the above CP43-D1 strain using Δ D1 as the recipient, further genetic manipulations in this CP43-D1 strain were hindered due to the lack of workable genetic markers. Five different resistance cassettes were used in the CP43-D1 strain: the resistance cassettes of chloramphenicol, kanamycin and spectinomycin for the gene knockout of *psbA1*, *psbA2* and *psbA3*, respectively (31); the erythromycin-resistance cassette for His-tagging the CP47 protein (107); and the gentamycin-resistance cassette adopted in this work for the selection of the CP43-D1 fusion mutant.

A second CP43-D1 fusion mutant, here designated as “CP43-D1/ Δ PsbA”, was constructed using “ Δ PsbA”, in which *psbA1* and *psbA3* were disrupted without the presence of a selectable marker, as a recipient strain. An FtsH2-knockout CP43-D1 derivative designated as “CP43-D1/ Δ PsbA/ Δ FtsH2” was also constructed to assess the repair of the CP43-D1 fusion PSII, by interrupting the *slr0228* gene with a chloramphenicol-resistance cassette (21). Segregation of mutants were confirmed by PCR (Figure 3-10A, right panel) and Sanger sequencing.

As illustrated in the D1 immunoblot shown in Figure 3-10A, the main D1 signal in the CP43-D1/ Δ PsbA strain and its FtsH-knockout derivative CP43-D1/ Δ PsbA/ Δ FtsH2 was the size of a CP43-D1 fusion protein, suggesting the CP43-D1 fusion was established and fully segregated in the Δ PsbA recipient background. No distinguishable difference regarding the size and potential cleavage of the fusion protein was observed between the two parallel fusion strains, namely CP43-D1/ Δ PsbA and CP43-D1. 2D PAGE

showed both monomeric and dimeric PSII present in thylakoid membranes of CP43-D1/ Δ PsbA and CP43-D1/ Δ PsbA/ Δ FtsH2 (Figure 3-10B). Two bands corresponding to the FtsH2 and FtsH3 were missing in CP43-D1/ Δ PsbA/ Δ FtsH2, in accord with the previous finding that FtsH2 is essential for the accumulation of FtsH2/3 protease complex (120).

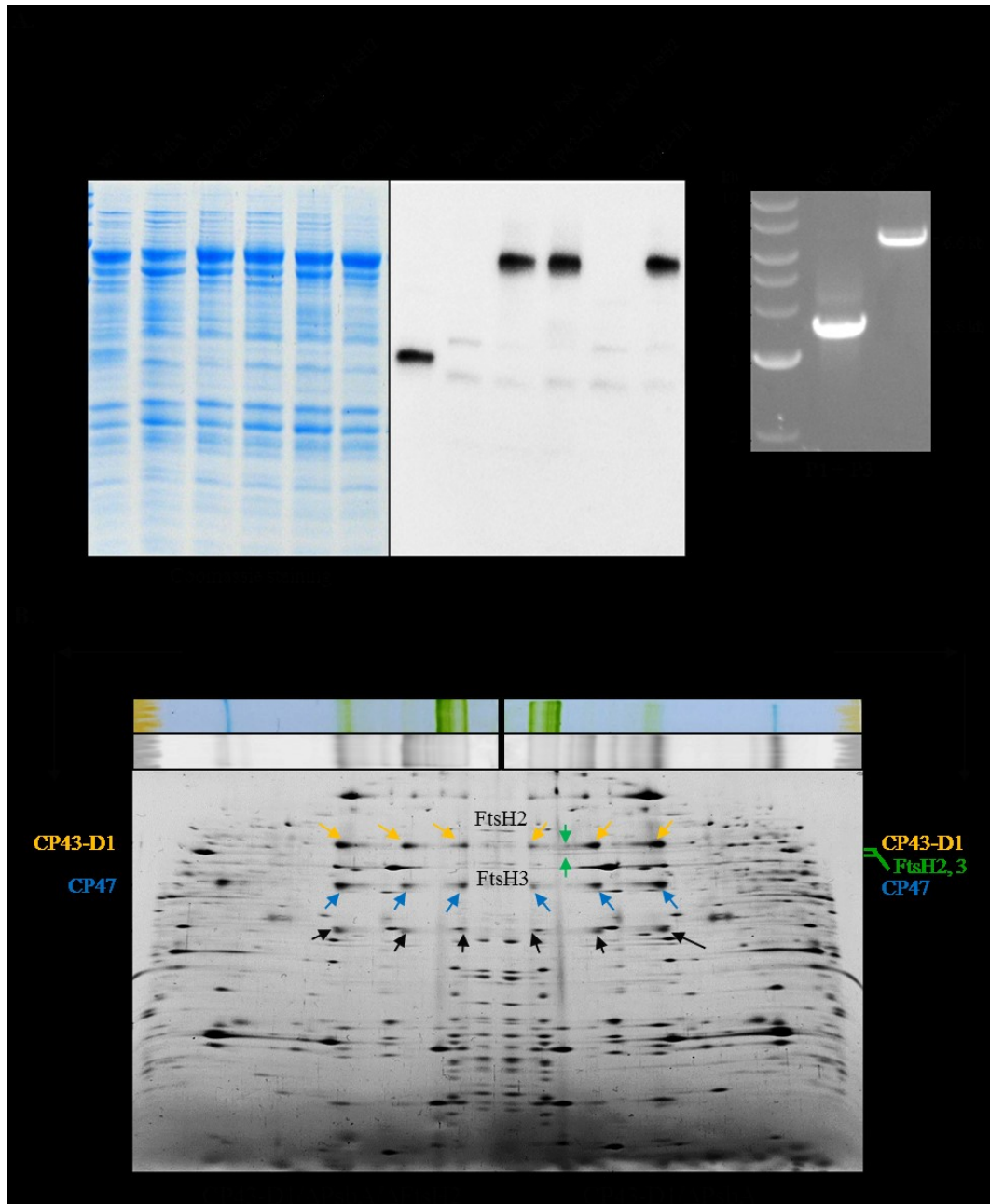


Figure 3-10 Detection of CP43-D1 fusion PSII in CP43-D1/ Δ PsbA and CP43-D1/ Δ PsbA/ Δ FtsH2.

A. The left panel shows the SDS-PAGE and Western blotting of the CP43-D1 protein in two strains. The right panel shows the PCR confirmation of the segregation of the CP43-D1/ Δ PsbA strain, the annealing

site of primers are illustrated in Figure 3-3A. B.2D Clear-native/SDS-PAGE of the CP43-D1 fusion PSII complexes. The second dimension was Sypro-stained. Protein bands corresponding to FtsH2 and FtsH3 are indicated by green arrows; yellow arrows point to CP43-D1 fusion protein; blue arrows point to CP47 protein; black arrows point to D2 protein. Proteins were empirically designated based on the high reproducibility of this method as shown in previous studies (113, 207, 208).

3.2.4.2. Photoautotrophic growth and light sensitivity

Similar to the first fusion strain CP43-D1, the second fusion strain CP43-D1/ Δ PsbA was able to grow photoautotrophically under medium to low light but not highlight (Figure 3-11). Knockout of FtsH2 in the fusion strain did not eliminate photosynthetic growth in strain CP43-D1/ Δ PsbA/ Δ FtsH2. However, under photoautotrophic condition, the light sensitivity of strain CP43-D1/ Δ PsbA/ Δ FtsH2 was aggravated to a level similar to the strain Δ FtsH2 (Figure 3-11).

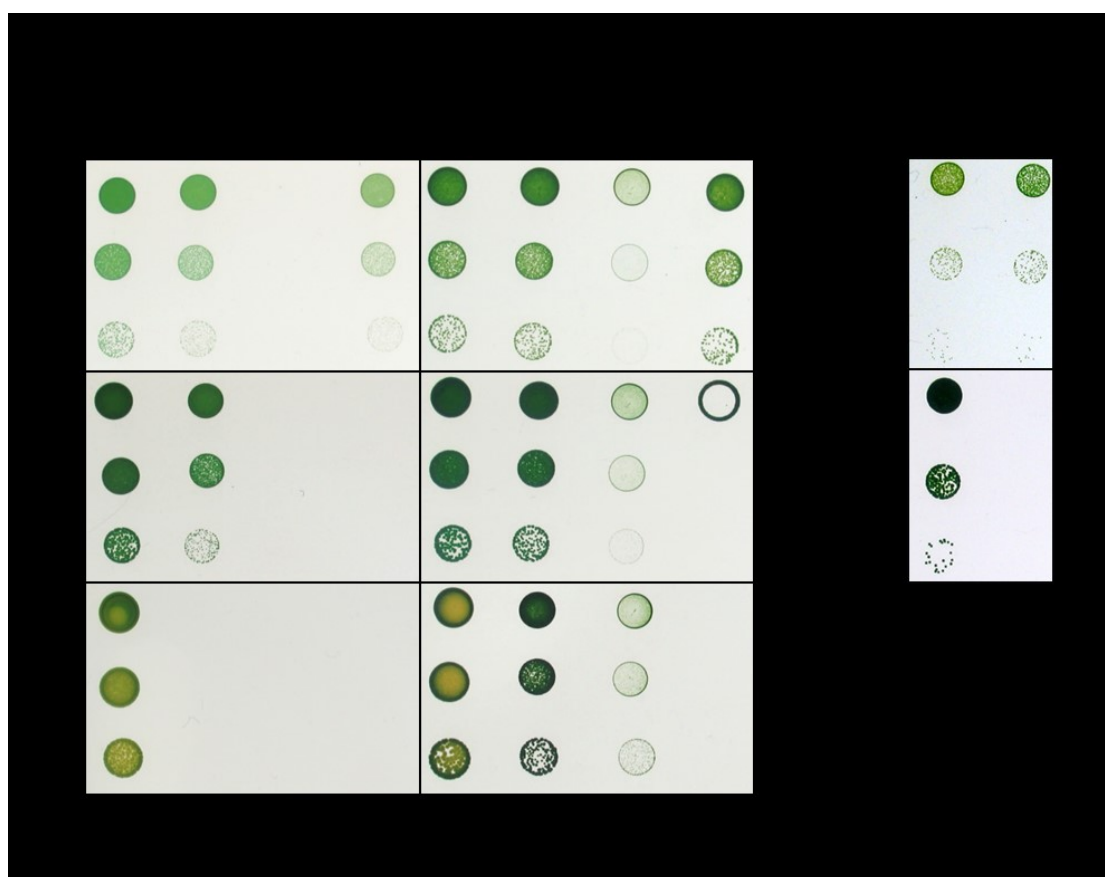


Figure 3-11 Growth assay of the second fusion strain CP43-D1/ Δ PsbA and the FtsH2-knockout derivative CP43-D1/ Δ PsbA/ Δ FtsH2.

Each spot grew from 2.5 μ l of serial diluted liquid culture, and the initial OD₇₃₀ is 0.1.

3.2.4.3. *Light saturation curve of oxygen evolution and chlorophyll fluorescence*

decay of PSII

The light saturation curve of oxygen evolution was measured in WT and strain CP43-D1/ Δ psbA to assess the PSII activity, and the data were fitted using a nonrectangular hyperbola-based model (133). As illustrated in Figure 3-12A, the dotted line represents the fitting result of the light-saturation curve of PSII activity in WT, and the dash-dot line represents that for CP43-D1/ Δ psbA. Based on the fitting curve, the P_{gmax} (asymptotic estimate of the maximum gross photosynthetic rate) is 844 $\mu\text{mol O}_2 \text{ mg Chl}^{-1} \text{ h}^{-1}$ for WT, and 655 $\mu\text{mol O}_2 \text{ mg Chl}^{-1} \text{ h}^{-1}$ for CP43-D1/ Δ psbA; The $I_{(50)}$ (light intensity at which photosynthetic activity reaches 50% of P_{gmax}) is 1148 $\mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$ for WT and 1666 $\mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$ for CP43-D1/ Δ psbA.

Profiles of Q_A^- reoxidation were also measured and compared between WT and CP43-D1/ Δ psbA (Figure 3-12C) by fluorescence relaxation after a saturating flash (91). No significant difference was observed, indicating the electron transfer path within PSII is intact in both strains. In addition, knockout of FtsH2 in the fusion strain showed no significant change in the PSII activity (Figure 3-12B).

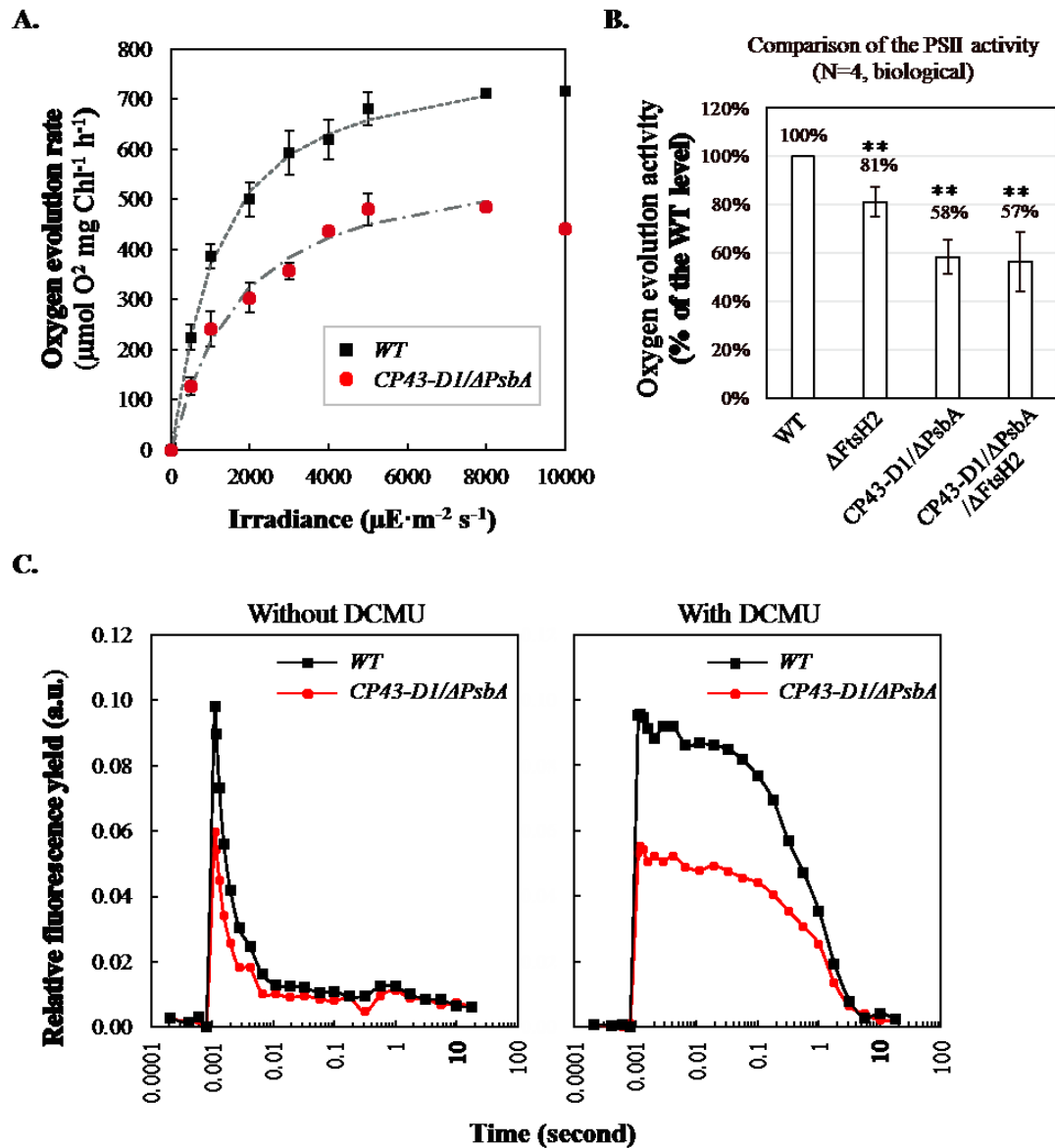


Figure 3-12 Activity assays of the fusion strain CP43-D1/ΔPsbA and CP43-D1/ΔPsbA/ΔFtsH2.

A. Light saturation curve of oxygen evolution. Data were produced from one biological replicate; each point was from two to three technical replicates. Error bar showed the standard deviation. B. Oxygen evolution rate was determined via a measuring light of intensity 2,000 μE·m⁻² s⁻¹. Data were produced from 4 biological replicates. Error bar represented standard deviation. Significance is checked by a T-test. **, p<0.001. C. Chlorophyll fluorescence decay in the presence and absence of 10 μM of DCMU.

3.3. Repair of the fusion PSII is defective

3.3.1. PSII repair is defective in the fusion strain

In WT *Synechocystis*, PSII is continuously photodamaged under any light condition, and an effective repair mechanism has evolved to maintain PSII functionality. The impact of fusing together CP43 and D1 on the repair process of PSII was investigated

using a photoinhibition assay. Photoinhibition rate is a combined effect of rate of photodamage and rate of PSII repair; the principle of the photoinhibition assay is to disentangle the photodamage rate from PSII repair rate, hence provide insight to the fitness of the PSII repair machinery. The photodamage rate can be measured by applying a protein synthesis inhibitor which blocks the repair process, and in tandem with the knowledge of photoinhibition rate, the PSII repair rate can be elucidated. As illustrated in Figure 3-13A, in the presence of the protein synthesis inhibitor apramycin, oxygen evolution activities of WT and three mutant strains showed a similar rate of damage to PSII, with around 10% of initial activity present in each strain after 4 h illumination. In the absence of apramycin, a significant difference was observed between WT and the repair-defective control, $\Delta FtsH2$. WT maintained about 67% of initial activity after 4 h of exposure, comparing to about 35% in $\Delta FtsH2$. Numerically, the FtsH2 protease was believed to contribute approximately 48% $((67\%-35\%/67\%))$ of the oxygen evolution in the 4-hour-treated WT by the PSII repair process. Indeed, the D1 degradation profiles (Figure 3-13B) revealed that removal of photodamaged D1 in $\Delta FtsH2$ was about 40% $((100-63)/(100-6))$ of the D1 removal in WT. The fusion strain CP43-D1/ $\Delta PsbA$, however, showed similar susceptibility to photoinhibition as $\Delta FtsH2$, 29% of initial activity was maintained in CP43-D1/ $\Delta PsbA$ after 4 h, indicating defective compensation mechanisms to counteract the photodamage of PSII, such as PSII repair and *de novo* assembly. By comparing the degradation profiles of the D1 and CP43-D1 proteins, the fusion strain CP43-D1/ $\Delta PsbA$ was more similar to $\Delta FtsH2$ than WT, and no significant difference in the degradation pattern of the CP43-D1 protein was noticed between CP43-D1/ $\Delta PsbA$ and its FtsH2-knockout derivative, CP43-D1/ $\Delta PsbA/\Delta FtsH2$. These results suggested that the FtsH-mediated PSII repair mechanism was incapable of efficient repair of the CP43-D1 fusion PSII.

Interestingly, a fragment of a size slightly larger than CP43 was detected by the CP43 antibody in CP43-D1/ Δ PsbA/ Δ FtsH2 but not CP43-D1/ Δ PsbA. This indicated a cleavage event of the CP43-D1 protein, possibly through a less efficient degradation pathway. Such cleavage might also exist in CP43-D1/ Δ PsbA, but the resulting fragments were degraded by FtsH2.

A difference in compensating for PSII photodamage via *de novo* assembly was also noticed between Δ FtsH2 and CP43-D1/ Δ PsbA/ Δ FtsH2. During the 4-hour photoinhibition treatment, Δ FtsH2 samples treated with or without protein synthesis inhibitor showed a significant difference in the oxygen evolution rate. This difference is attributed to the *de novo* assembly of PSII, as indicated by the upregulated D1 content in the 4-hour non-treated Δ FtsH2 sample to 129% of the initial level (Figure 3-13B). While in CP43-D1/ Δ PsbA/ Δ FtsH2, no significant difference in oxygen evolution activity was observed between the treated and non-treated samples; this is likely due to the reduced promoter strength of the *psbD1/psbC* operon compared to *psbA2* and consequent reduction in the rate of synthesis of the CP43-D1 protein compared to D1. The level of the D1 and CP43-D1 transcripts will be discussed later in Section 3.3.2.

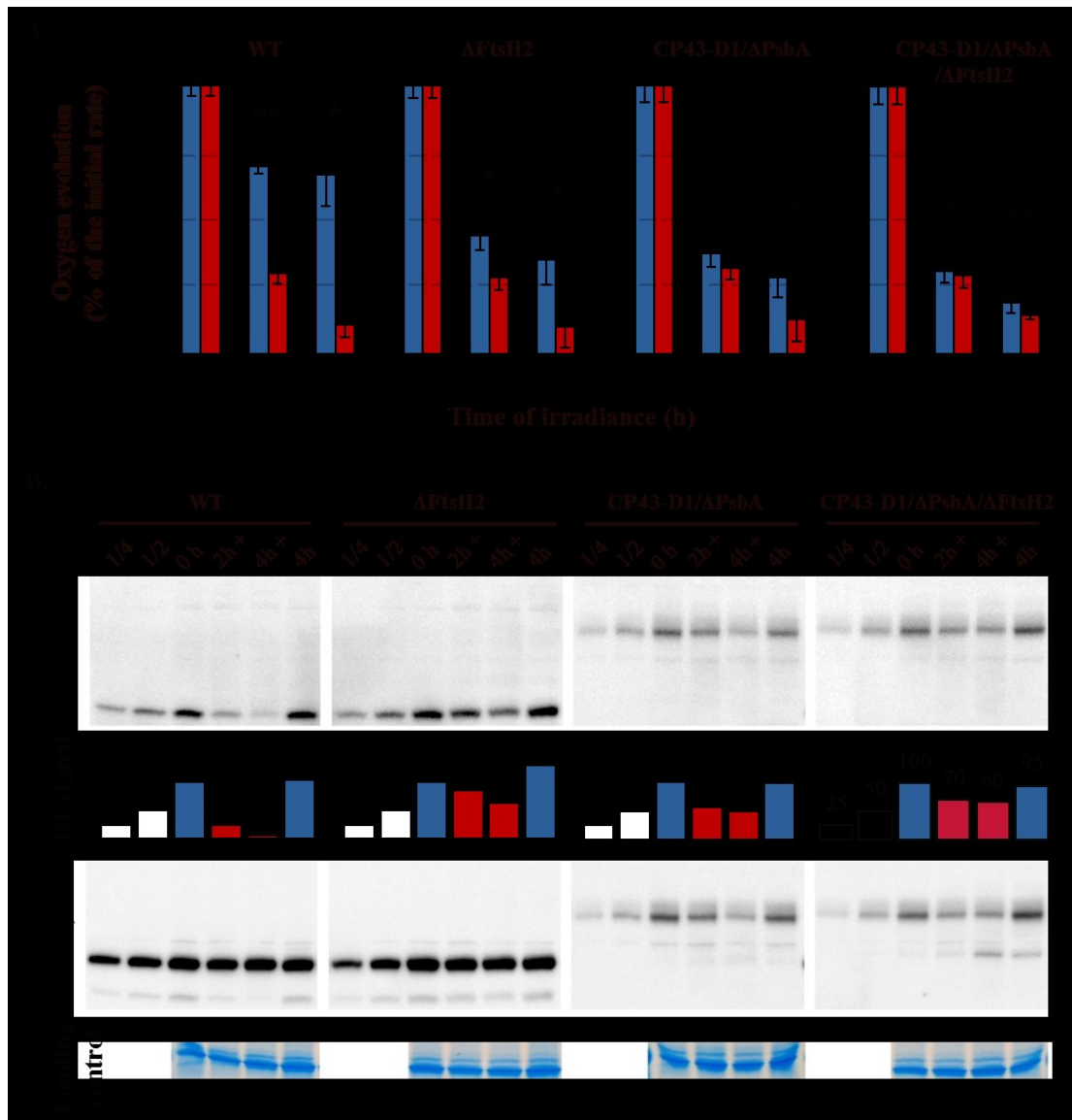


Figure 3-13 Photoinhibition assay on the CP43-D1 fusion strain

The colour scheme: blue means untreated and red treated with protein synthesis inhibitor, 200 μ M of apramycin. A. Time course of oxygen evolution activity of different strains subjected to photoinhibitory light of 200 μ E·m⁻² s⁻¹. Each data point represented 2 to 3 biological replicates, error bar showed the standard deviation. Significance is checked by a T-test. *, $p < 0.05$. **, $p < 0.001$. NS, not significant at a threshold of 0.05. B. Time course of protein degradation revealed by western blotting, the experiment was repeated biologically 2 to 3 times and proved reproducible. For each time point, 0.5 μ g of Chl a was loaded, 1/4 and 1/2 loading of the 0 h sample were included as the reference for semi-quantification. Symbol “+” indicated samples treated with protein synthesis inhibitor. Densitometry of the D1 and CP43-D1 signals was quantified in the percentage of the 0 h level and visualised as a histogram. Detection of the CP43 protein was reprobed on the PVDF membrane for D1 detection. Coomassie staining was showed as the equal loading control. Star symbol indicated the potential degradation band of interest.

3.3.2. The expression level of the CP43-D1 fusion gene

The transcript level of the CP43-D1 fusion gene, as well as genes encoding CP47 and PsaB, a PSI subunit, was assessed by quantitative PCR (qPCR) using the $\Delta\Delta C_T$ method. As shown in Figure 3-14, transcript levels of *psbA2,3* (the *psbA2* segment of the fusion gene, here named as *psbA2,3* due to the fact that primers cannot differentiate between *psbA2* and *psbA3*) were significantly reduced in the two fusion strains compared with WT, suggesting that lower mRNA level might be a constraint to the accumulation of CP43-D1 proteins in the fusion strains. This difference in the *psbA2,3* transcripts between the CP43-D1 fusion and WT strains is most likely due to the difference in the promoter. The transcript levels of *psbB* and *psbC* were enhanced in both fusion strains, likely a compensating mechanism responding to decreased level of PSII biogenesis. The transcription level of *psaB* was also enhanced. It is known that defective PSII activity can lead to the decrease of PSI content (209). Here, the increase of *psaB* mRNA may suggest increased oxidative stress due to the insufficient repair of PSII.

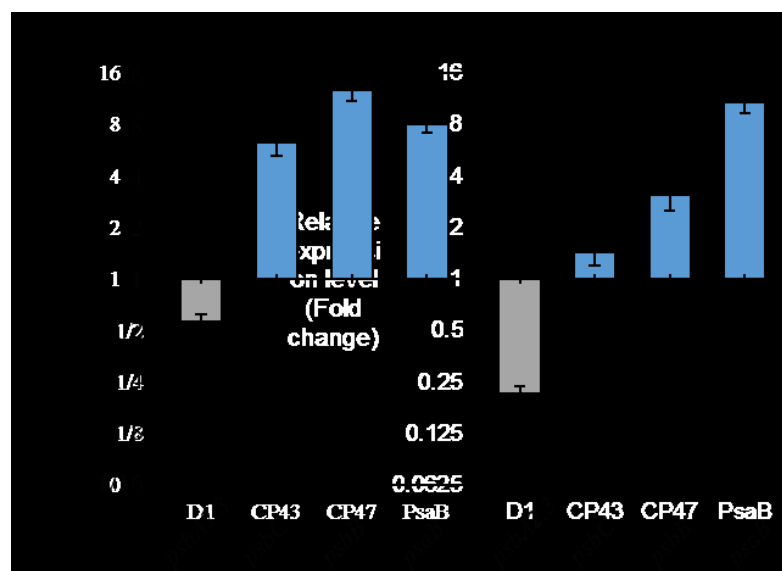


Figure 3-14 Relative quantifications of the transcript-level expression of the CP43-D1 fusion gene.

Relative quantification was performed using $\Delta\Delta C_T$ method, and the *rps1* gene was selected as the endogenous reference. Data were from one assay consisting of 4 - 6 technical replicates, the experiment was biologically repeated 2 times. Error bar indicated standard deviation. Cells for the extraction of RNA and cDNA synthesis were grown under $20 \mu E \cdot m^{-2} s^{-1}$ of white light in liquid BG11 medium supplemented with 5 mM of glucose.

3.3.3. The CP43-D1 fusion protein can be degraded under extreme highlight stress

Although less repairable upon photoinhibition, the CP43-D1 protein proved to be degradable *in vivo* under certain conditions such as enhanced high light stress. When cells were subjected to extreme illuminance, oxygen evolution activity of WT was severely depressed (Figure 3-15, right bottom corner). In the presence of protein synthesis inhibitor, significant degradation of the D1 and CP43-D1 protein was observed not only in WT but also in CP43-D1/ Δ PsbA and Δ FtsH2, although Δ FtsH2 showed the lowest level of D1 degradation. Interestingly, a fragment detected by CP43 antibody was observed in CP43-D1/ Δ psbA (Figure 3-15, arrow), likely the same product as in CP43-D1/ Δ PsbA/ Δ FtsH2 (Figure 3-13B, star symbol). Taken together, these results showed the CP43-D1 protein could be degraded *in vivo* by an unclear mechanism in response to the extreme light condition. However, it is not known if the same mechanism is behind the degradation of the CP43-D1 protein and the degradation of the D1 protein in Δ FtsH2 under extreme highlight.

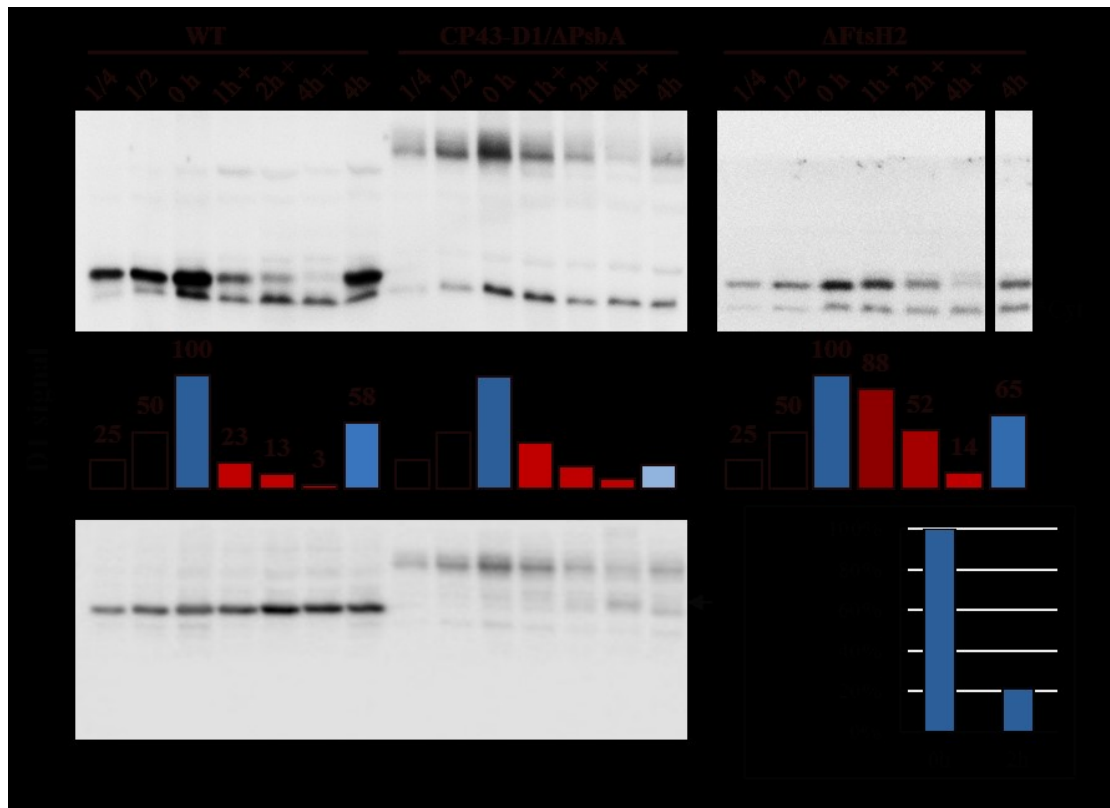


Figure 3-15 Degradation of the CP43-D1 protein under extra-high light stress.

The colour scheme: blue means untreated and red treated with protein synthesis inhibitor, 200 μ M of apramycin. For each time point, 0.5 μ g of Chl a was loaded, 1/4 and 1/2 loading of the 0 h sample were included as the reference for semi-quantification. Symbol “+” indicated samples treated with protein synthesis inhibitor. Densitometry of the D1 and CP43-D1 signals was quantified in the percentage of the 0 h level and visualised as a histogram. Detection of the CP43 protein was reprobed on the PVDF membrane for D1 detection. The arrow indicates a potential degradation product. The star symbol indicates signal from the peroxidase reaction between Cytochrome *f* and luminescence substrate⁽¹⁹⁸⁾.

3.4. Split for survival: genomic study of fusion suppressors

3.4.1. FuBH, a CP43-D1 split highlight suppressor

3.4.1.1. Segregation and growth assay of FuBH

It is of interest and evolutionary importance to see how cells acclimate to environmental stress via genetic regulation. A suppressor strain of CP43-D1, namely FuBH, was identified through continuous growth under elevated light condition ($\sim 40 \mu\text{E} \cdot \text{m}^{-2} \text{s}^{-1}$). As shown in Figure 3-16A, unlike the strain CP43-D1, FuBH was capable of photoautotrophic growth even under high light. Western blotting showed a “D1” protein with a size similar to a WT D1 protein, indicating a break of the CP43-D1 protein. Considering that there was no intact CP43-D1 detected in strain FuBH, this

suppression was suspected to be due to mutation of the CP43-D1 fusion gene. Several PCR amplifications starting from both ends of the CP43-D1 fusion gene were successful until near the flexible linker region (Figure 3-16B), after which abnormal PCR fragments were detected. Furthermore, whole genome sequencing revealed a translocation event (between position CR and S7 in Figure 3-16B) that split the fusion gene: a segment after the *psbC* gene was swapped to another location in the genome (see Section 3.4.1.4 for details). Notably, PCR using primers I4/S11 still amplified at low level a fragment only present in the intact CP43-D1 encoding gene; based on the genomic sequencing (described in Section 3.4.1.4), this product likely results from annealing of a forward ssDNA from residual *psbC* with a reverse ssDNA from *psbA2* at *psbA3* locus followed by amplification. Primers I4 and S11 amplified in WT a fragment of size 0.1 kb, which is likely to be a nonspecific product as the theoretical size of the correct product is 1.2 kb.

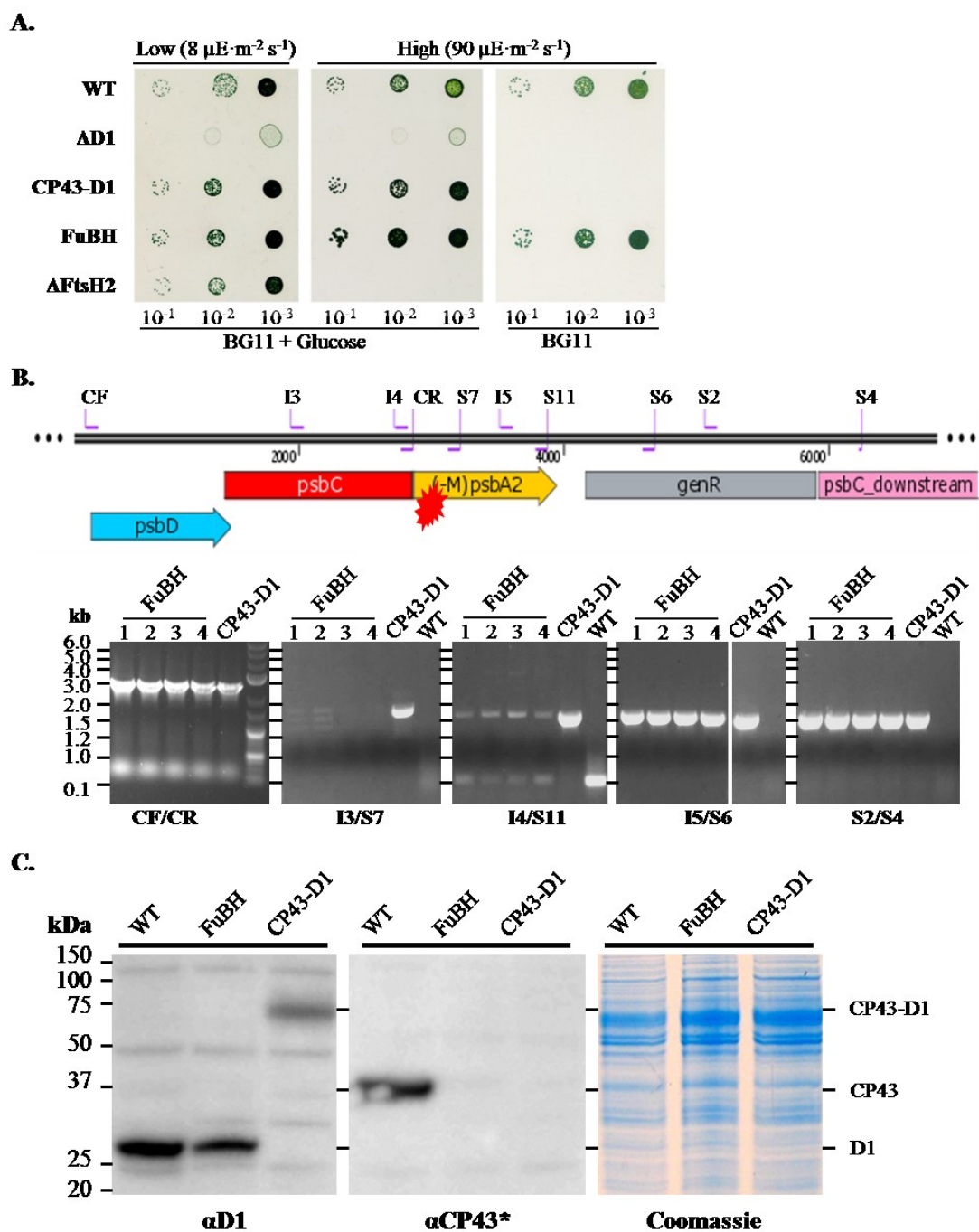


Figure 3-16 Growth assay and western blotting of the CP43-D1 split strain, FuBH.

A. Spot growth assay as described in previous Sections. B. Diagram showing where the translocation occurred to split the fusion gene and the PCR attempts in aim for locating the site of the split, all PCR products except CF/CR were examined on a single large size agarose gel with two rows. Red explosion symbol indicates the site of fusion gene split. C. Western blotting detection of D1 and CP43 in the thylakoid samples of WT, FuBH and CP43-D1.

3.4.1.2. PSII repair is restored in FuBH

A photoinhibition assay was performed to assess the PSII repair process in the suppressor FuBH. As illustrated in Figure 3-17A, upon photoinhibition, the degradation

of the D1 protein in FuBH was faster than the fusion protein in CP43-D1, and at a similar level as in the WT (Figure 3-13). However, the transcript level of *psbA2,3* was still significantly lower than the WT level, excluded the possibility that the lower expression level had led to the high light sensitivity of the fusion strains. These results complemented the conclusion in Section 3.3.1 and supported the conclusions that the split between CP43 and D1 benefits PSII repair. Interestingly, the transcription level of *psbC* is significantly decreased in the FuBH strain comparing to that in WT. One possible explanation could be enhanced post-transcriptional regulation if the split mutation leads to an abnormal *psbC* mRNA. Another possibility is altered post-translational feedback if the split mutation leads to a more stable form of CP43.

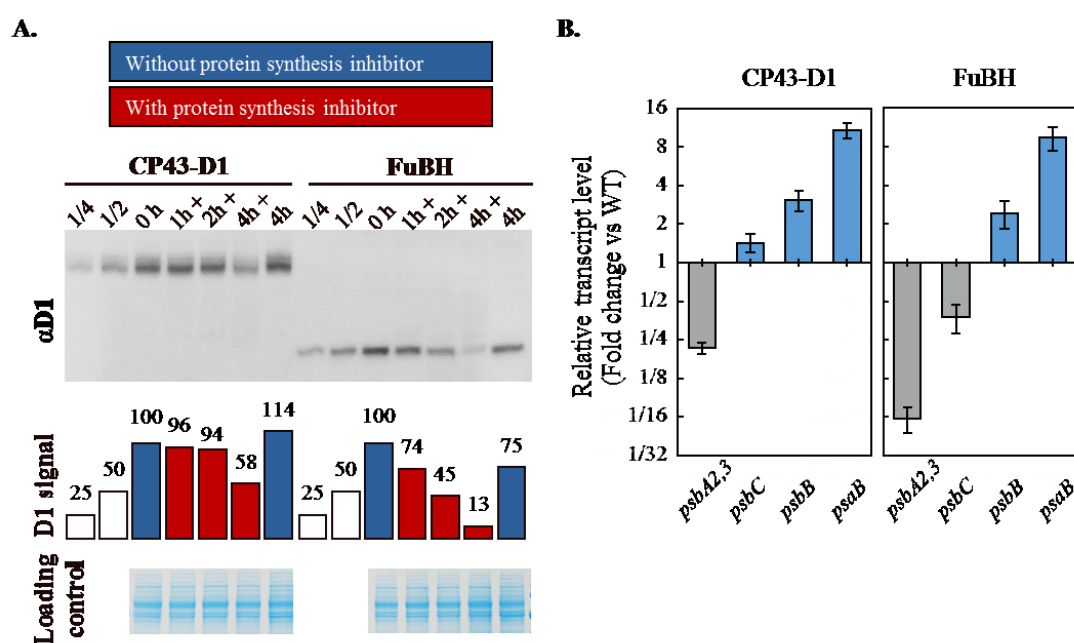


Figure 3-17 PSII repair and the D1 expression level in the CP43-D1 split strain, FuBH.

The colour scheme: blue means untreated and red treated with protein synthesis inhibitor, 200 μ M of apramycin. A. Photoinhibition assay as described previously in Figure 3-13. B. The relative quantification of the D1o level in strain FuBH. Error bar represents standard deviation. Data are based on 4-6 technical replications from one experiment.

3.4.1.3. Genome resequencing and SNP analysis of FuBH

Potential mutations that give rise to the suppressor FuBH were firstly assessed via a genome resequencing analysis. By mapping the sequencing data to the *Synechocystis* sp. PCC 6803 reference genome (GenBank: BA000022.2, “Kazusa” substrain (210)), a few interesting sites of single nucleotide polymorphism (SNP) were revealed (Table 3-1). In accord with previous resequencing works on *Synechocystis* (64-66, 211), 19 sites of SNP (grey coloured), which are considered as database error (DE), or “mistakes” of the Kazusa strain genome, as described by Tajima *et al.* (211), were also found in WT, Δ D1, CP43-D1 and FuBH. Two previously reported DE sites at genome position 3096187 and 3110343, however, were not found in the above four strains, consistent with the previously reported genome of the “PCC-M” substrain (65). The WT strain referred in this work (WT-P) is a glucose-tolerant *Synechocystis* sp. PCC 6803 strain cultivated in the laboratory led by Professor Peter Nixon at Imperial College London. Among the SNP events in WT-P, the substitution in *rpl3* and frame shift in *slr1609* were also reported in a recent subculture from WT-P (67). The mutants, Δ D1, CP43-D1 and FuBH, are derived from the WT-P strain. By comparing the SNP profiles of each strain, a signature of 9 mutations was found in Δ D1, CP43-D1 and FuBH (Table 3-1), which confirms that FuBH is not a contaminant strain. Two unique variants were also found in FuBH, however with no obvious connection to the split of CP43-D1 protein. A substitution of *slr0105* in FuBH would introduce the mutation of G76R in Slr0105. Although the role of Slr0105 has not yet been characterised in *Synechocystis*, it is orthologous to cyanobacterial cobyrinic acid a,c-diamide synthase and chromosome partitioning ATPase according to CyanoBase (<http://genome.microbedb.jp/cyanobase>). Overall, the split of CP43-D1 was unlikely to be due to SNP events.

Table 3-1 List of SNPs detected in the FuBH and its background strains.

The average sequencing coverage was between 70 to 90 times, and the threshold for SNP detection was set to above 35 times of coverage, and above 90% of variant frequency. *A signature of 9 mutations found in Δ D1, CP43-D1 and FuBH.

No.	Start	End	Size	Change	Codon Change	AA Change	Protein Change	CDS	Product / Database error
<i>Variants identified in all WT-P background strains (WT, ΔD1, CP43-D1, FuBH)</i>									
1	386406	386406	1	T -> A	GTT -> GAT	V -> D	Substitution	slr1084	Hypothetical protein
2	842060	842060	1	C -> T	CGG -> CAG	R -> Q	Substitution	rpl3	50S ribosomal protein L3
3	943495	943495	1	G -> A	GTC -> ATC	V -> I	Substitution	psaA	Considered as database error (DE)
4	1012958	1012958	1	G -> T					DE
5	1200306	1200306	1	C -> A	CCA -> CAA	P -> Q	Substitution	slr1862	Hypothetical protein
6	1211634	1212076	443	Unalignable			Deletion or transposition	sll1774	Hypothetical protein
7	1213710	1213733	24	Unalignable			Deletion or transposition		
8	1364187	1364187	1	A -> G	TTG -> CTG		None	pyrF	DE
9	1454955	1455444	490	Unalignable			Deletion or transposition	slr1712	Hypothetical protein
10	1819782	1819782	1	A -> G	CTT -> CTC		None	psbA3	DE
11	1819788	1819788	1	A -> G	TCT -> TCC		None	psbA3	DE
12	2092571	2092571	1	A -> T			Truncation	sll0422	DE
13	2198893	2198893	1	T -> C	TTA -> TTG		None	sll0142	DE
14	2204584	2204584	1	(G)9 -> (G)8			Frame Shift	gspF	General secretion pathway protein F
	2204584	2204584	1	(G)9 -> (G)8			Frame Shift	pilC	Pilin biogenesis protein
15	2301721	2301721	1	A -> G	AAG -> GAG	K -> E	Substitution	slr0168	DE
16	2350285	2350286	0	+A					DE
17	2360246	2360247	0	+C			Frame Shift	slr0364	DE
18	2409244	2409244	1	-C			Frame Shift	sll0762	DE
19	2419399	2419399	1	-T			Frame Shift	ycf22	DE
20	2544044	2544045	0	+C			Frame Shift	ssl0787	DE
21	2602717	2602717	1	C -> A	CAC -> CAA	H -> Q	Substitution	slr0468	DE
22	2602734	2602734	1	T -> A	ATT -> AAT	I -> N	Substitution	slr0468	DE
23	2748897	2748897	1	C -> T					DE
24	2817601	2817751	151	Unalignable			Deletion or transposition		
25	3063738	3066047	2310	Unalignable			Deletion or transposition	sll0790; sll0789; sll0788	Sensory transduction histidine kinase; OmpR subfamily; Hypothetical protein
26	3110189	3110189	1	G -> A					DE
27	3142651	3142651	1	A -> G	CTT -> CTC		None	sps	DE
28	3260096	3260096	1	(C)7 -> (C)6					DE
<i>Variants identified in ΔD1, CP43-D1 and FuBH*</i>									
29	2048412	2048414	3	GAA -> TTC					
30	2048426	2048426	1	T -> A					
31	2049580	2049581	2	TC -> AT	TTC -> TAT	F -> Y	Substitution	slr1636	Hypothetical protein
32	2354058	2354058	1	T -> A	GTT -> GTA		None	slr0364	Hypothetical protein
33	2354649	2354649	1	T -> A	GTT -> GTA		None	slr0364	Hypothetical protein
34	2547027	2547027	1	C -> T	GTG -> ATG	V -> M	Substitution	sll0410	Putative esterase
35	2817740	2817740	1	C -> T					
36	2817748	2817748	1	C -> A					

37	2817751	2817751	1	C -> T						
<i>Strain-specific variants: WT-P</i>										
38	488230	488230	1	T -> G	TTC -> TGC	F -> C	Substitution	slr1609	Long-chain-fatty-acid CoA ligase	
39	779341	779342	0	(T)8 -> (T)9			Frame Shift	zam	Acetazolamide conferring resistance protein zam	
40	1822198	1822198	1	(G)8 -> (G)7			Frame Shift	sll1389	Hypothetical protein	
41	3110114	3110114	1	G -> T	GCC -> GCA		None	sll0666	Transposase	
<i>Strain-specific variants: FuBH</i>										
42	1869205	1869205	1	G -> A	GTG -> GTA		None	gspD	General secretion pathway protein D	
43	2978733	2978733	1	G -> A	GGG -> AGG	G -> R	Substitution	slr0105	Hypothetical protein	

3.4.1.4. Locating the suppression mutation by genome *de novo* assembly

The SNP analysis provided information on the genetic background of the CP43-D1 fusion strains but was not sufficient to clarify the genomic location of genes. The genome sequencing reaches a depth of average 70 to 90 times, allowing the genome *de novo* assembly to be performed to locate the *psbA2* segment of the gene encoding CP43-D1. The four genomes were assembled to contig level, and Table 3-2 shows a quality report of the assemblies. Contigs containing mutations of interest were analysed through sequence alignment and database blast to deduce the suppression mutation. Mutations leading to structural variation were confirmed by PCR and Sanger sequencing of the PCR product of FuBH shown in Figure 3-18B. As illustrated in Figure 3-18A, in the FuBH genome, a segment of the CP43-D1 fusion gene containing an intact *psbA2* and 68 bp downstream *psbA2* in the fusion cassette was found in the $\Delta psbA3$ locus, consistent with the D1 protein detected in FuBH (Figure 3-16). It should be mentioned that four nucleotide differences between *psbA2* and *psbA3*, at positions 36, 468, 936 and 969, allow differentiation of *psbA2* and *psbA3*. The start codon of the relocated *psbA2* gene in FuBH was likely obtained via homologous recombination with the 384 bp residual *psbA3* inherited from the $\Delta D1$ strain (126). Notably, a truncated version of the CP43-D1 fusion gene still exists, and likely a truncated CP43-D1 protein

is expressed. Moreover, the CP43 antibody failed to detect the potential truncated CP43-D1 protein, suggesting the antigen recognition site of CP43 is still masked in FuBH (Figure 3-16C). The gentamycin-resistance cassette was not read by assembled genome contigs. However, its existence was confirmed by PCR (Figure 3-16B, S2/S4) and antibiotic selective growth (data not shown). In the *psbC-psbA2* locus, the sequence after position 331 of *psbA2* (numbering of the intact *psbA2*) was replaced by an Ω fragment (212) from the knockout-insertion of $\Delta psbA3$. Coincidentally, primer S7 binds to position 330-349 of *psbA2*, which explains why the PCR using primers I3/S7 was not successful in Figure 3-16B. The precise position where the Ω fragment transposition ends in the *psbC* locus of the FuBH is not clear due to the nature of the assembled contig. However, the continuity and integrity of gentamycin resistance cassette were confirmed by previous PCR with primers I5/S6 and S2/S4 (Figure 3-16B).

The size of the DNA fragment that left *psbC* locus is estimated, according to the PCR and genome sequencing data, to below 321 bp (It is the length of the segment from position 332 to the annealing site of primer I5). For the same reason, the precise position where the transposed *psbA2* gene, together with the gentamycin resistance cassette, ends in the *psbA3* locus is unknown. Transpositions of the above mentioned two segments (dash line box in Figure 3-18A) was likely the key event to give rise to the strain FuBH. However, the underlying mechanism of the transposition is unclear. Two segments of size 732 bp and 716 bp (light green and dark green, respectively, in Figure 3-18A) show over 95% sequence identity, and both of them contained a 309 bp integrase sequence with only one base pair of difference (Supplement 3). Apart from these homologous regions which could enable recombination, a genetic mobile element consisted of *slr0856* and *slr0857* lies downstream of the CP43-D1 locus, which might

confer potential genetic mobility to the nearby sequence. These features complicate the deduction that one transposition mechanism leads to FuBH.

The structural variation was confirmed by PCR (Figure 3-18B) using primer P37, which anneals to the spectinomycin-resistance cassette used for the knockout of *psbA3*, and primers P35 (genome position 7707-7608) or P36 (genome position 1820536-1820515), respectively. Using primers P35/P36, a 1.0 kb fragment in FuBH but not CP43-D1 or WT was successfully amplified; this result is consistent with the interpretation based on the genome sequencing. Two bands lower than 1.0 kb were also noticed in the CP43-D1 strain and are likely nonspecific products. Overall, these data suggest that a transposition event led to the splitting of the CP43-D1 gene.

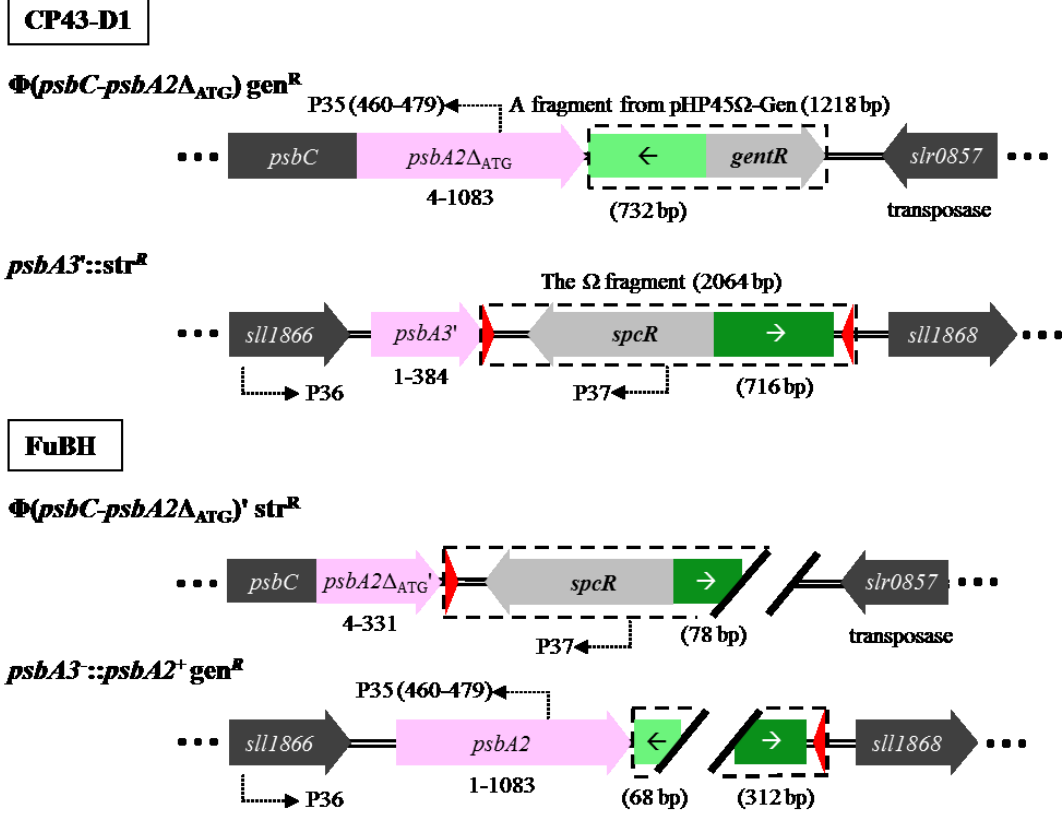
Table 3-2 The quality report of the genome assemblies assessed by QUAST toolkit

	WT (Nixon)	CP43-D1	FuBH	ΔD1
Reference: BA000022.2, “Kazusa”; GC-content 47.72%; 3573470 bp; 3196 genes (chromosome)				
GC (%)	47.53	47.53	47.53	47.52
Total aligned length (% of WT)	3508643 (98.2%)	3511192 (98.3%)	3510571 (98.2%)	3508705 (98.2%)
Genes	3066 + 28 part	3068 + 33 part	3068 + 31 part	3065 + 31 part
contigs	115	107	107	102
Largest contig	386457	388047	387993	388022
Largest alignment	386457	246471	254603	246477
Misassemblies	3	6	6	7
NG50	114278	130033	130037	129949
LG50	9	8	8	9
NGA50	98359	104724	104824	98487
Duplication ratio	1	1.002	1.002	1.002
N's per 100 kbp	22.61	23.31	22.41	19.83
mismatches per 100 kbp	1.51	1.79	1.74	2.22

All statistics are based on contigs of size over 500 bp. Misassemblies: the number of positions in the contigs (breakpoints) that satisfy one of the following criteria: 1, the left flanking sequence aligns over 1 kbp away from the right flanking sequence on the reference; 2, flanking sequences overlap on more than 1 kbp; 3, flanking sequences

align to different strands. NG50: contig length that using longer or equal length contigs produces 50% of the bases of the reference genome. LG50: number of contigs of length at least NG50. NGA50: NG50 where the lengths of aligned blocks are counted instead of contig lengths. Duplication ratio: the total number of aligned bases in the assembly divided by the total number of aligned bases in the reference genome. N's per 100 kbp: the average number of uncalled bases (N's) per 100,000 assembly bases.

A.



B.

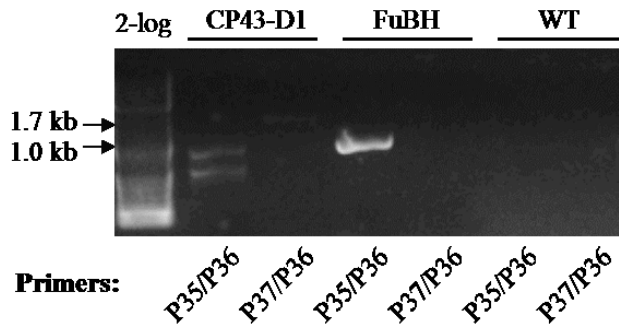


Figure 3-18 Genotyping the CP43-D1 and its suppressor FuBH.

A. Two segments of genes of interest were shown. DNA fragments of *psbA2* and *psbA3* were indicated by pink arrows with size labelled underneath. The apostrophe indicates truncated genes. *psbA2 Δ_{ATG}* represents the *psbA2* gene without the start codon, corresponding to position 4-1083; *psbA3'* represents the residual part of the *psbA3* gene from the recipient D1 triple deletion mutant Δ D1; *psbA2 Δ_{ATG} '* represents the residual *psbA2* gene at the site of CP43-D1 fusion, after a suppressor mutation. Antibiotic resistance was shown as a grey arrow. The *gentR* and *spcR* represent antibiotic resistances to gentamicin and spectinomycin, respectively. Genes adjacent to the genes of interest were showed as dark grey arrow to indicate the orientation of the two segments. Dashed box was used to emphasise the fragment originated from the construction of mutations. The invert repeats of the Ω fragment (212), of a size 146 bp, were shown as red triangles. The light and dark green boxes showed two homologous regions of size 732 bp and 716 bp, respectively. These two regions are in opposite orientation as indicated by the black

and white small arrows. Slash indicated the end of an assembled sequence contig. Dots indicated the omitted upstream and downstream sequences. Besides, *slr0587* encodes a transposase which could play a role in gene transposition. B. PCR confirmation of the translocation event in FuBH. The primer binding sites were indicated in Figure A with dot-arrow lines, “2-log” is the DNA ladder and the expected sizes of PCR product is 1.7 kb for P37/P36 in CP43-D1 and 1.0 kb for P35/P36 in FuBH. Primers P37/P36 were used as positive control, and a faint 1.7 kb band can be seen in CP43-D1 amplified with P37/P36.

3.4.2. Mock-evolution: mutagenesis-assisted CP43-D1 split suppressors

3.4.2.1. Design of break fusions

To further study spontaneous mutations in the CP43-D1 fusion strain that would give rise to a free D1 protein, the CP43-D1 fusion gene was artificially split through mutagenesis approaches. Four types of split were designed (Figure 3-19A): Ins, a stop codon was inserted to produce CP43 only; Poi, a stop codon was added to *psbC*, and a start codon was added to restore the intact *psbA2* gene; Del, the flexible linker region was replaced by a frame-shifted stop/start codon; Sub, add stop codon for CP43 expression, add 10 bp upstream sequence of the *psbA2* locus and the start codon to *psbA2*. Strains containing each of these modified CP43-D1 split genes were successfully constructed using Δ PsbA as the recipient strain. PCR was used to confirm segregation (Figure 3-19B) and Sanger sequencing the mutation.

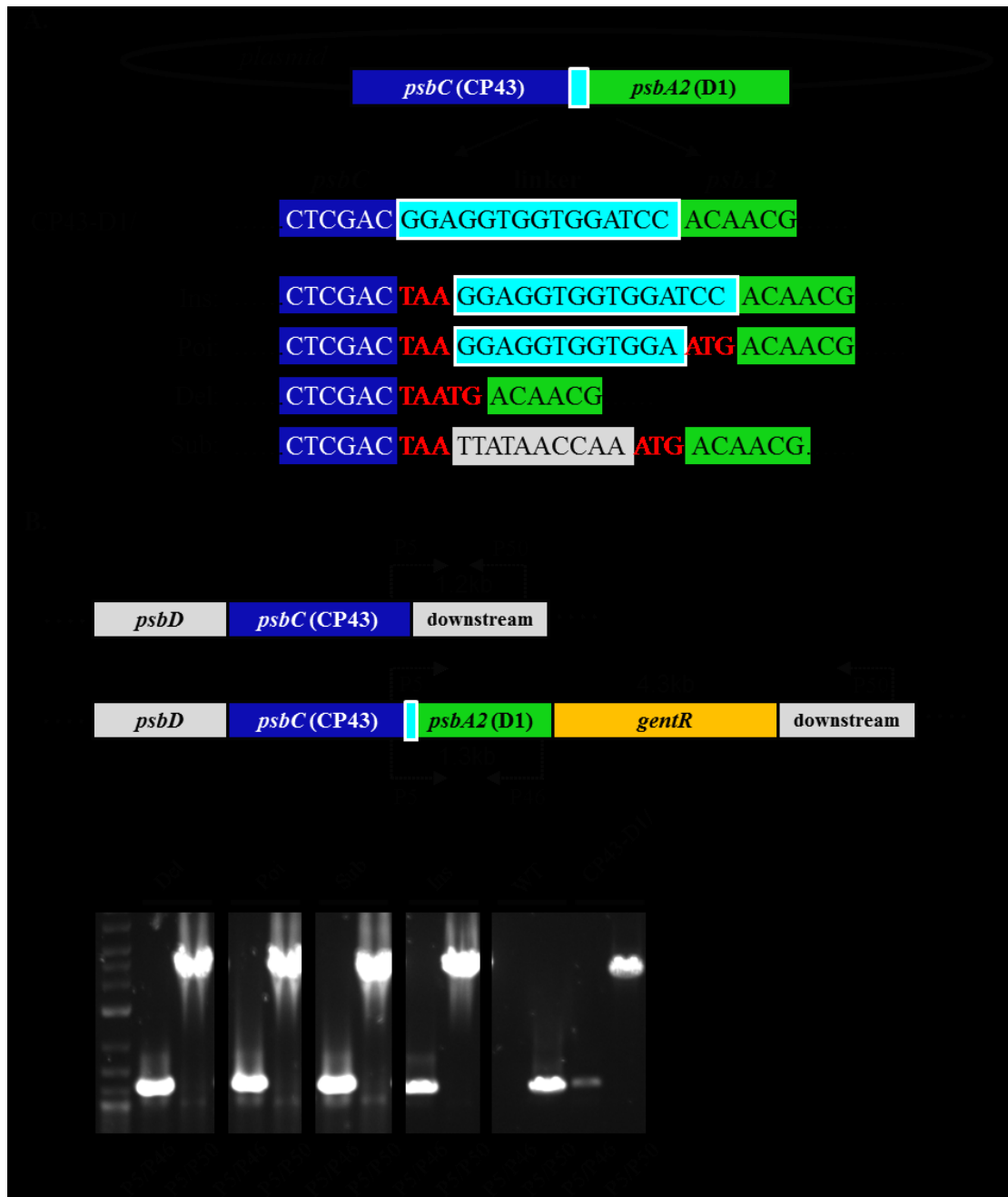


Figure 3-19 Design of the break fusion mutants.

A. The illustration of the sequence designs of the break fusion mutants. B. The upper panel shows the annealing positions of PCR primers used to confirm the successful segregation of designed mutants. The bottom panel shows the PCR results. WT and CP43-D1/ Δ psbA were used as a control.

3.4.2.2. Screening of CP43-D1 split suppressors

As shown in Figure 3-20A, the CP43-D1 split strains, Del, Ins, Sub-1, Sub-2, Poi-1 and Poi-2, all grew in the presence of glucose but failed to grow under photoautotrophic growth conditions under normal and high light irradiances.

By continuously restreaking mutants on BG11 plus 5 mM of glucose and gradually increasing the light intensity between each restreaking, two types of suppressor strains were obtained: Ins-SP from the Ins strain and Del-SP from the Del strain. As shown in Figure 3-20A, both suppressors can even grow photoautotrophically under high light conditions. Western blotting analysis showed no D1 and less CP43 in the Ins strain but close-to-WT levels in the Ins-SP strain. All three designs of the split suppressor strain with added start codon “ATG”, namely Sub, Poi and Del, showed the presence of D1, suggesting that there might be ribosome binding sites near the 3' of *psbC* in the three strains. In the case of Del, synthesis of D1 indicates a case of ribosome reattachment. Moreover, by genome sequencing, no suppression mutation of the CP43-D1 split gene was observed in Sub and Poi strains.

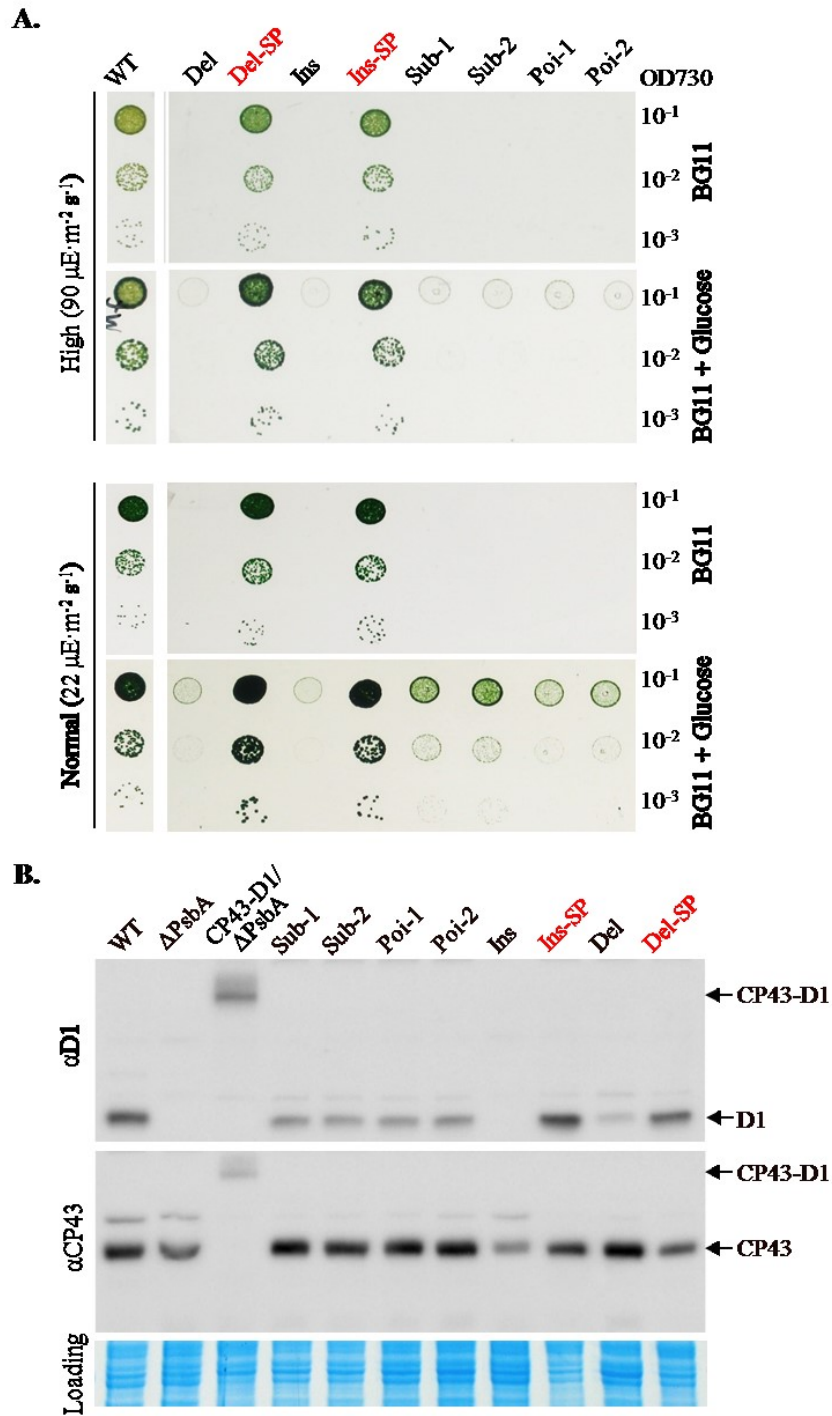


Figure 3-20 Confirmation of break fusion suppressors.

A. Growth assay of the CP43-D1 split mutants and two derived suppressor mutants (labelled in red). The growth was photographed on the 8th day. B. Western blotting shows the expression of a “split” D1 protein in the suppressor mutants. The bottom panel is Coomassie staining for loading control.

3.4.2.3. Genome localisation of the split CP43 and D1

Using the same approach as in Section 3.4.1.4, suppression mutations in Ins-SP and Del-SP were analysed. Surprisingly, the same type of suppression was observed in the

two suppressors as found in FuBH. As shown in Figure 3-21A, the CP43-D1 split gene likely recombined with the 174 bp *in situ* residual *psbA2* fragment, to give rise to an intact *psbA2* followed by the rest of the CP43-D1 split gene sequences; however, the end of the downstream segment is still not yet clarified. Similarly, the residual *psbA2* segment including the chloramphenicol-resistance cassette, utilised for the construction of $\Delta psbA2$, transposed to the CP43-D1 split gene locus; again the precise end is not yet solved. Figure 3-21B shows the PCR confirmation of the suppressor mutations. Interestingly, the Ins-SP and Del-SP seem not fully segregated, probably due to insufficient evolutionary pressure or time of evolution to fully discard the designed break fusion genes. Comparing to the FuBH, Ins-SP and Del-SP were constructed from a marker-less D1 triple deletion strain, and the residual *psbA2* fragment was half the size of the *psbA3* fragment found in the $\Delta D1$ strain that gave rise to FuBH. Also, the composition of the chloramphenicol-resistance cassette of $\Delta psbA2$ was simpler than the Ω fragment of $\Delta psbA3$. Nevertheless, a similar transposition event took place in the two strains in response to light stress.

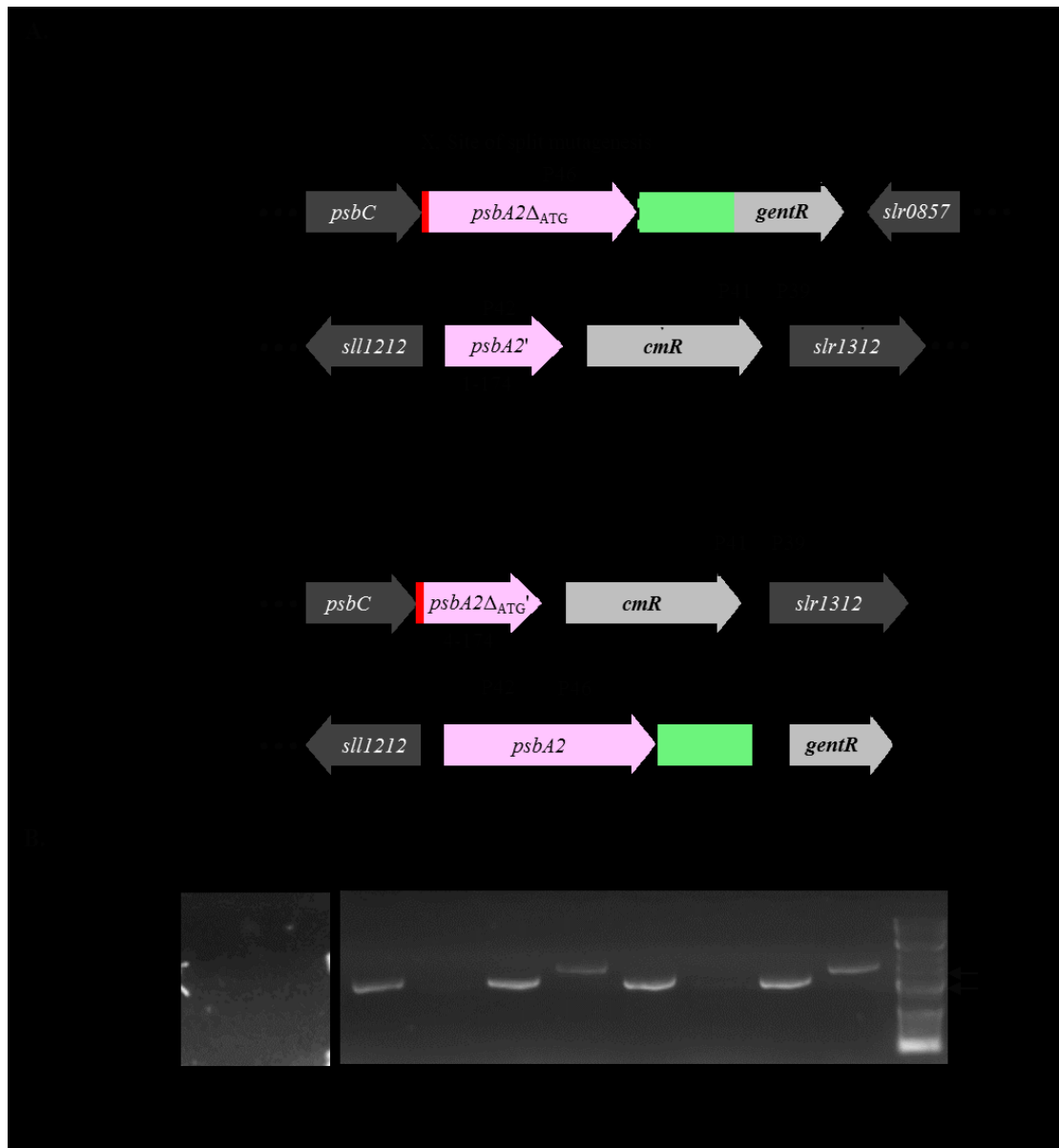


Figure 3-21 Localisation of the D1-encoding gene in the genomes of split and corresponding suppressor mutants.

A. *psbA2* Δ_{ATG} represents the *psbA2* gene without the start codon, corresponding to position 4-1083; *psbA2'* represents the residual part of the *psbA2* gene from the recipient D1 triple deletion strain Δ PsbA; *psbA2* Δ_{ATG}' represents the residual *psbA2* gene at the anticipated site of CP43-D1 break, after a suppressor mutation. The *gentR* and *cmR* represent antibiotic resistances to gentamicin and chloramphenicol, respectively. *The P46 primer anneals partially (18 of 29 bp) to the *psbA2* locus in WT, but the PCR using P42/P46 in WT did not give a corresponding band. B. PCR confirmation of the translocation event in Ins-SP/Del-SP. The primer binding sites were indicated in Figure A with dot-arrow lines, “2-log” is the DNA ladder and the expected sizes of PCR product is 1.5 kb for P42/P46 in Ins-SP/Del-SP and 1.0 kb for P41/P39 in Ins and Del.

3.5. Discussion

3.5.1. Incorporation of CP47 and CP43 into PSII need not be sequential

In this work, I show by fusing the C-terminus of CP43 to the N-terminus of D1 that a CP43-D1 fusion protein, which to some extent resembles a Type I reaction centre subunit such as PsaA or PsaB, can be synthesised and assembled to form oxygen-evolving PSII complexes in *Synechocystis*. Analysis of His-tagged PSII preparations by western blot and mass spectrometry confirmed that the composition of this CP43-D1 fusion PSII is similar to that of a wild-type PSII complex (Figure 3-5, 3-6). Native/SDS 2D-PAGE of the purified PSII showed incorporation of this CP43-D1 protein into both monomeric and dimeric PSII (Figure 3-7). Activity assays further showed this fusion PSII is active, and the whole cell oxygen evolution activity of CP43-D1 is at least greater than 50% of the WT level. Taken together, these results showed that a “Type-I like” reaction centre subunit is capable of assembling an oxygen-evolving PSII complex.

According to the step-wise modular assembly of wild-type PSII (19, 101), synthesis of a CP43-D1 fusion subunit would now reverse the order of addition of CP47 and CP43 to the PSII RC complex. My results suggest that the attachment of the CP43 subcomplex after the CP47 subcomplex to RC is therefore not obligatory (101, 107) which raises the possibility that it might also occur at a low level in WT.

It remains unclear whether all the LMM (low molecular mass) subunits remain bound to CP43 in the CP43-D1 fusion. These include PsbK, PsbZ and Ycf12 located on the periphery of CP43 in WT PSII (35, 37). PsbI, however, is in proximity to both D1 and CP43 in the wild-type PSII (35, 37). Given the fusion of CP43 and D1, it is possible that binding of PsbI has been affected. PsbI has been shown to be dispensable for the formation of the RC complex and mature PSII but is important for stabilising

unassembled D1 and the subsequent binding of CP43 (45, 213); such a role suggests that loss of PsbI is unlikely to be a constraint in the assembly process when CP43 and D1 are already fused. Besides, in the wild-type PSII, the localisation of PsbH relative to CP47 is similar to the localisation of PsbI relative to CP43; and PsbH stabilises the binding of CP47 subcomplex (214), similar to the role of PsbI in stabilising the binding of CP43 (45, 213). However, in contrast, PsbH is attached to CP47 before the formation of the RC47 subcomplex (107), while PsbI mainly binds to D1 instead of CP43 at an early stage (45). Given the results of this work, it is reasonable to postulate a short-lived assembly intermediate “RC43” comprising mainly D1, D2 and CP43. As discussed above, the early interaction between PsbH and CP47 likely renders higher stability of the RC47 subcomplex relative to the speculated RC43 subcomplex. Besides, unlike the assembly intermediate RC47 which does not have all the ligands required for binding the OEC, the speculated RC43 possesses a complete set of ligands for OEC binding. Premature photoactivation of the OEC might have therefore have selected against the formation of RC43.

Because Psb27 associates with the CP43 subcomplex in wild type (113, 114), and it is implicated in the assembly of the manganese cluster (115), it was of interest to examine whether, during the assembly of the CP43-D1 fusion PSII, Psb27 interacts with CP43-D1 protein to prevent the premature association of extrinsic proteins and formation of OEC (115). Although mass spectrometry data (Figure 3-6) showed the presence of Psb27 in the preparation of His-tagged CP43-D1 fusion PSII, it is not clear to which complex the Psb27 binds. In a CP43-D1/ Δ CP47 mutant analysed by 2-D PAGE and western blotting (unpublished data by collaborators from Professor Josef Komenda's lab at Centre Algatech, Institute of Microbiology, the Czech Academy of Sciences), Psb27 does not co-migrate with the RC43 subcomplex, while the PSII assembly factor

HliD (high light inducible protein D) and its interacting partner Ycf39, as well as Ycf48, do co-migrate with RC43. The HliD-Ycf39 complex is involved in the assembly of RC (207) and is likely responsible for cotranslational synthesis and integration of chlorophyll into nascent PSII subunits (215). Moreover, it has been shown that Psb27 mainly binds to the monomeric PSII reaction centre complex under normal growth condition (113); upon photoinhibition, Psb27 dissociates from the reaction centre complex, but its binding to the uncomplexed CP43 is increased. Importantly, this uncomplexed CP43 is mainly attributed to the photoinhibition treatment (113). Taken together, it is speculated that Psb27 specifically binds to CP43 during the repair but not a *de novo* assembly of PSII. Further experiments are needed to locate Psb27 in the CP43-D1 fusion PSII.

3.5.2. Efficient repair selects against the fusion of CP43 and D1

Importantly PSII complexes containing the the CP43-D1 fusion protein are more prone to photoinhibition (Figure 3-13). I have also shown that the FtsH protease which degrades D1 of wild-type PSII is incapable of efficiently degrading photodamaged CP43-D1 fusion protein, in line with the previous finding that an exposed N-terminus of D1 is essential for D1 degradation (121). As discussed above, a fusion version of CP43 and D1, or even CP47 and D2, would simplify the assembly process of such a sophisticated complex as PSII, and it has been shown in this work that the fusion of CP43 and D1 is feasible. Why therefore is a fusion PSII not found in nature? The defective degradation of CP43-D1 protein in response to photodamage of oxygenic photosynthesis may provide a reasonable answer. Furthermore, the mutations that are responsible for the acquisition of high light resistance through the splitting of the CP43-D1 fusion occurred independently and similarly in three genetic backgrounds, namely FuBH from the CP43-D1 strain, Ins-SP from the Ins strain and Del-SP from the Del

strain. Although the selectable markers used for mutagenesis might be behind such transposition events (Figure 3-18A, 3-21A), the evolutionary advantage of a separated D1 is evidenced.

In the absence of CP47, FtsH2 can degrade the D2 protein of the PSII RC (188), suggesting such a repair mechanism might have existed in the homodimeric Type-II reaction centre ancestral to the heterodimeric PSII. This speculated PSII ancestor might represent a key stage in the evolution of oxygenic photosynthesis, during which the innovation of water oxidation occurred (189). If this is the case, such a repair mechanism will contribute significant evolutionary advantage to this early water-oxidation species near, if not before, the GOE. Furthermore, a question can be asked: If the homodimeric PSII did exist and carry out water oxidation and the repair of reaction centre, when did this repair mechanism arise? Chapter 4 provides a tentative answer to this question.

Chapter 4. An evolutionary view of the cyanobacterial FtsH proteases

4.1. Introduction

4.1.1. FtsH proteases and oxygenic photosynthesis

A hallmark of oxygenic photosynthesis is the presence of photoprotective mechanisms to prevent or repair the damage caused by the reaction centres operating in the presence of oxygen (216, 217). As discussed in the general introduction, FtsH plays a crucial role in oxygenic photosynthesis through the repair of PSII. Meanwhile, the multiplicity of thylakoid/chloroplast FtsH proteases in oxygenic phototrophs has long been noticed (218) and has been linked to possible differences in the capability to degrade D1, membrane localisation (219-223) and subunit composition of complexes (219, 224-226). Early phylogenetic attempts (219, 224, 226) using a limited sequence dataset suggested that the FtsH subunits required for PSII repair exist in two main forms, denoted Type A and Type B (227). Given the importance of FtsH-mediated PSII repair, it is plausible that the evolution of Type A/B FtsH could contain relevant information regarding the evolution of oxygenic photosynthesis. Besides, the evolutionary relationship between Type A/B FtsH and other thylakoid FtsH paralogs, and the relation between FtsH multiplicity and oxygenic photosynthesis are far from clarified. In this Chapter, a comprehensive phylogenetic analysis was carried out to address the evolution of the FtsH subunits specific for PSII repair.

4.1.2. Diverse function of FtsH proteases

FtsH is a membrane-integrated protease universally found in eubacteria and eukaryotic organelles. It plays a crucial role in the homoeostasis of soluble and membrane proteins. FtsH is the only indispensable protease in *E. coli* and most enterobacteria (228, 229)

with dysfunction of the *ftsH* gene in *E. coli* leading to a lethal imbalance of the lipopolysaccharide and phospholipids levels. FtsH is also involved in the downregulation of the σ^{32} (RpoH) transcriptional regulator which upregulates the transcription of many heat shock genes (230-232). This FtsH-mediated degradation requires binding of DnaK and DnaJ chaperones to σ^{32} (233, 234). FtsH is also involved in protein quality control by degrading the ssrA-tagged soluble and membrane proteins (235, 236) involved in translational surveillance. Another substrate of FtsH of fundamental importance is the bacterial SecYEG complex (Sec61 $\alpha\beta\gamma$ complex in mammals) responsible for protein translocation across the membrane. FtsH not only degrades uncomplexed SecY (237) but also removes complexed SecY as well as SecE from a malfunctioned translocon (238). The latter activity, destroying a translocon, is a potential suicide mechanism that requires fine tuning. Table 4-1 lists some of the substrates of FtsH reported in the literature.

Table 4-1 A list of some reported substrates of FtsH and homologues

Substrate	Substrate function	Dependence on auxiliary protein	Reference
LpxC, KdtA	lipopolysaccharides (LPS) biosynthesis		(228, 239, 240)
σ^{32} (RpoH)	upregulates the transcription of many heat shock genes	Require the binding of DnaK and DnaJ chaperones to σ^{32}	(230-232)
λ Xis, λ CII and λ CIII	lysogeny of phage λ		(241-243)
IscS	cysteine desulfurase, iron-sulfur cluster formation		(244)
ssrA-tagged proteins	Removal of incompletely translated polypeptide	Soluble, no; Membrane-anchored, likely yes;	(235, 236)

SecY, SecE	SecYEG complex (Sec61 $\alpha\beta\gamma$ complex in mammals), protein translocations across the membrane		(237) (238)
YccA	Modulator of FtsH protease YccA	Soluble, no; Membrane-anchored, HflK/C;	(238, 245)
AtpB	subunit a of the F ₀ sector of H ⁺ -ATPase		(246)
PspC	phage shock protein C, involved in transcription regulation		(247)
DadA	involved in deamination of D-amino acids		(244)
FdoH	a subunit of one type of formate dehydrogenase involved in adaptation to aerobic conditions		(244)
YfgM	Unknown function		(244, 248)
HSP21	Plant thermomemory		(249)
D1, D2	Maintenance of photosystem II		(20, 21, 188)

4.1.3. Common structure of FtsH complexes

A monomeric FtsH subunit is composed of an N-terminal membrane-spanning domain followed by a single AAA⁺ (ATPases Associated with diverse cellular Activities) domain and a C-terminal protease domain exposed to the soluble phase (250, 251). The membrane-spanning region, composed of one to two transmembrane segments (251, 252), is known to be variable across the tree of life (253, 254); the soluble region, however, is relatively more conserved. High-resolution crystal structures (255-257) of the soluble region and EM structures of the intact enzyme (120, 258) have revealed a

conserved hexameric ring-shaped structure for both the homocomplex and heterocomplex forms. The membrane-anchored N-terminal part of FtsH, despite its evolutionary complexity (254), likely favours formation of a hexameric ring structure as evidenced in *E. coli* FtsH and human Afg3L2, another FtsH homolog (253, 254). Crystal structures of the FtsH AAA+ domain from *E. coli* and *Thermus thermophilus* showed a homohexameric ring structure of the FtsH complex (259, 260); later this homohexameric organisation was also observed in the entire cytosolic region based on crystal structures of FtsH from *T. thermophilus* and *Thermotoga maritima* (255-257). Lee *et al.* revealed by cryo-EM a heterohexameric structure of the mitochondrial m-AAA complex composed of the Yeast FtsH homologues, Yta10 and Yta12 (258). However, due to low resolution, the subunit organisation cannot be clearly defined in this heterocomplex. In *Synechocystis*, by GST-tagging one of the FtsH subunits, Boehm *et al.* revealed an alternating organisation of FtsH2 and FtsH3 in a heterohexamer (120). A different subunit organisation might exist in the plant counterpart of the FtsH2/3 heterocomplex, based on the unequal stoichiometry of the two types of FtsH subunit (219, 261).

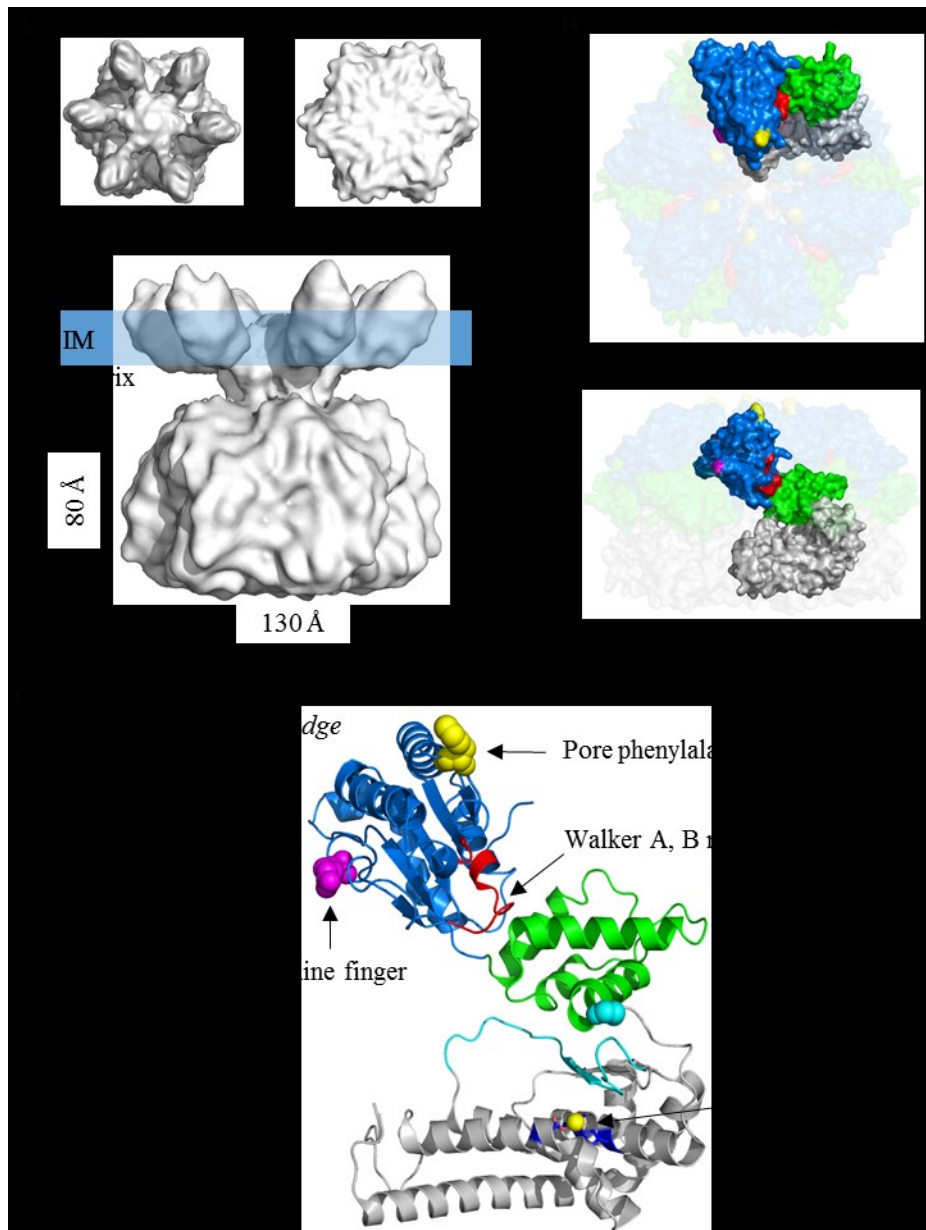


Figure 4-1 Structures of FtsH homologues.

A. The cryo-EM structure of the full-length mitochondrial m-AAA protease complex from yeast (EMD: 1712). Approximate dimensions of the main cytosolic body are indicated. Light blue indicates the membrane-embedded region. IMS: intermembrane space. IM: inner mitochondrial membrane. B. The crystal structure of the cytosolic region of *apo*-FtsH protease complex from *T. maritima* (PDB: 3KDS). C, cartoon view of the monomer *apo*-FtsH showing different structural segments. Colour scheme of B and C: yellow, pore residues; magenta, arginine finger; red, Walker motifs; green, helical subdomain; grey, protease domain; blue, catalytic zinc ligands; the individual yellow sphere adjacent to blue coloured ligands, zinc ion. Besides, cyan indicates suspicious regions of structural importance described later, including a conserved glycine, sphere view, and a flexible “lid helix” on top of the zinc.

Figure 4-1A shows the cryo-EM structure of the full-length mitochondrial m-AAA protease complex. The dimensions of this complex are similar to those of the FtsH2/3

complex in *Synechocystis* (120). Figure 4-1B shows the *apo*-FtsH homocomplex from *T. maritima* (257). In the presence or absence of bound nucleotide, a dramatic inter-domain movement occurs between the AAA+ and protease domains (Figure 4-1C), and an approximate 20° shift occurs between the N-terminal “wedge” and the C-terminal “helical” subdomains within the AAA+ domain (Figure 4-1C) (257). Hydrolysis of bound ATP is catalysed by the arginine finger from an adjacent subunit. The conserved pore phenylalanine (Phe 234 in *T. maritima* FtsH) is exposed to the intermediate space in-between the cytosolic main body and membrane, where substrates are believed to enter the ATPase ring (260). The protease domain of FtsH shows a unique fold compared to other zinc metalloproteases, and its proteolysis activity is independent of the AAA+ domain (256). The zinc ion is pyramidally coordinated by the two histidine residues of the HEXXH motif and a proximal aspartic acid residue.

4.2. A phylogenetic survey of FtsH

*Excerpts from “Section 4.2 A phylogenetic survey of FtsH” has been submitted for publication to a special issue in *Photosynthetica* under the title “**Divergence of Photosystem II-specific FtsH proteases at the dawn of oxygenic photosynthesis**”, authored by Shengxi Shao, Tanai Cardona and Peter J. Nixon from Imperial College London.

4.2.1. Early diversification of FtsH proteases

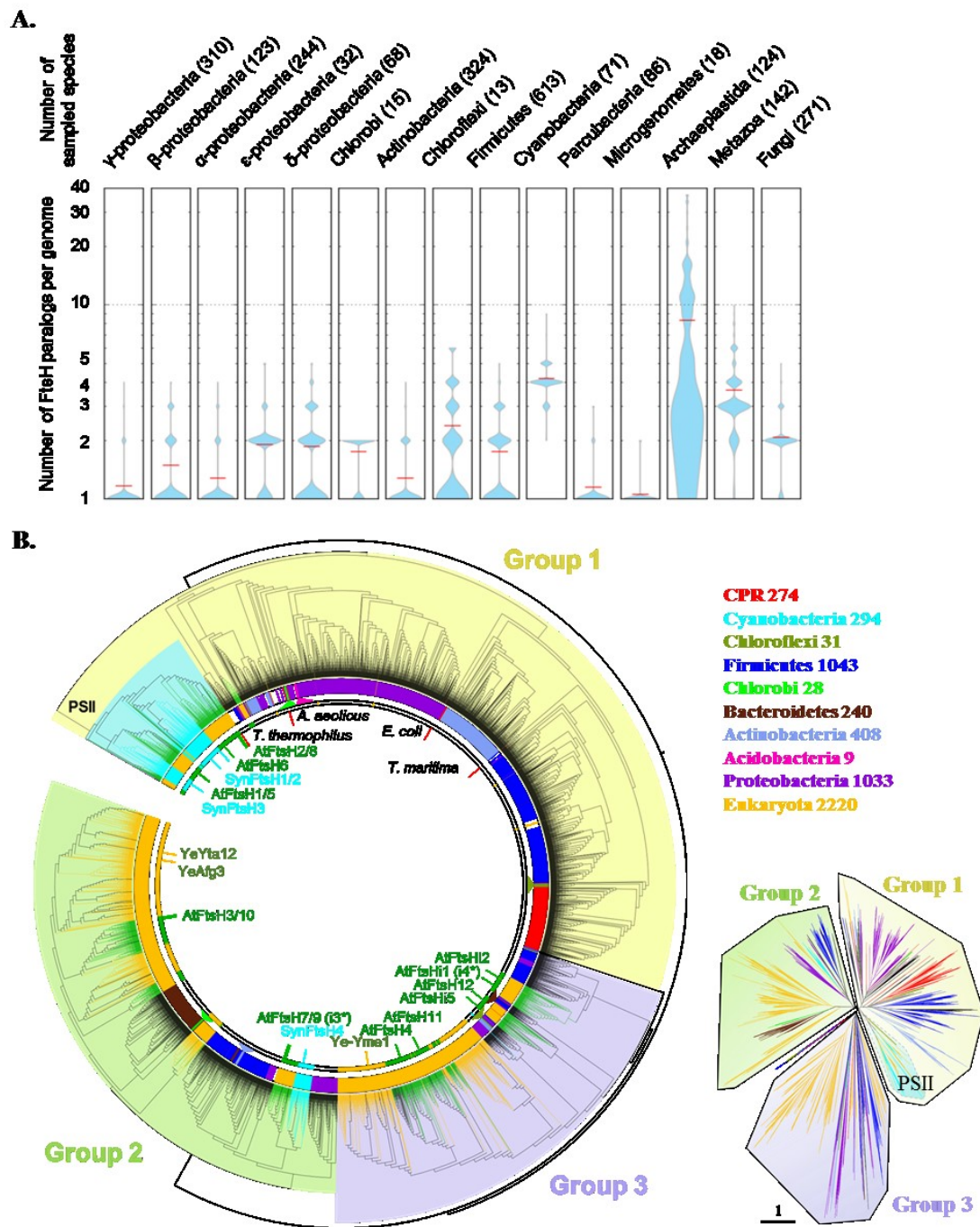


Figure 4-2 Multiplicity and evolution of FtsH proteases.

A. FtsH copy number in proteomes from major bacteria phyla or classes. Red lines indicate the mean value of the number of FtsH per genome and the light blue curve represents the frequency of a given number of FtsH per genome per clade. B. The phylogeny of the FtsH family. The left panel shows a circular cladogram of the phylogeny, and the right corner shows the unrooted tree. The three proposed groups of orthologs are coloured accordingly; the branch containing well-characterised FtsH proteases responsible for PSII repair in *Arabidopsis* and *Synechocystis* is also highlighted (PSII). Major taxonomic groups are colour-coded in the thick inner circle. CPR represents “candidate phyla radiation” (149). Eukaryotic and cyanobacterial FtsH proteases are shown in the thin inner circle to emphasise photosynthetic eukaryotes and Cyanobacteria. Representative FtsH proteases from Yeast (Ye),

Arabidopsis (At) and *Synechocystis* (Syn) are colour-labelled. Four FtsH proteases with structural information available are indicated by the red lines and species name in black.

A comprehensive dataset comprising 6082 FtsH homologues from over 3100 genome-sequenced species was constructed using the Pfam database. The number of FtsH paralogs per sampled species was assessed in some of the major taxonomic groups (Figure 4-2A). The multiplicity of FtsH proteases was observed in all major groups. Cyanobacteria showed the highest average number of FtsH homologues among prokaryotic groups with approximately 4 FtsH homologues per sequenced genome. Photosynthetic eukaryotes showed the highest multiplicity, with more than 8 FtsH proteases per genome, while animals showed an average of about 4 and fungi about 2. FtsH multiplicity in other prokaryote groups was also common: however, the number of FtsH paralogs per genome was in most cases no greater than 3 (Figure 4-2A).

The phylogenetic analysis confirmed the universality of FtsH proteases in the domain Bacteria. As illustrated in Figure 4-2A and 4-2B, at least three potential orthologous groups were resolved and denoted Group 1, Group 2, and Group 3. Each group reproduced a branch topology consistent with previous phylogenetic and phylogenomic studies on the diversification of bacteria (262-264). The appearance of the the three independent trees, namely the similar radiation of major bacteria phyla within each of the three groups, suggests that FtsH proteases were likely present in the last common ancestor of the domain Bacteria and that the three groups may have originated from ancestral gene duplication events early during the evolution of bacteria. Group 1 FtsH showed the greatest diversity and was almost universally found across bacteria. Most strains with a single FtsH protease were more likely to have retained a Group 1 FtsH.

Notably, within Group 1, an early diverging, well-resolved branch (support 0.983), contained FtsH sequences exclusively found in Cyanobacteria and photosynthetic

eukaryotes; these included the AtFtsH1/2/5/8 subunits from *Arabidopsis* and SynFtsH2/3 from *Synechocystis* sp. PCC 6803, all known to be involved in PSII repair. The origin of this group seemed to predate the radiation of FtsH orthologs from other bacteria phyla, including those of anoxygenic phototrophic bacteria within the same group. This early divergence was not found in other FtsH proteases of Cyanobacteria and photosynthetic eukaryotes that are less relevant to PSII protection and maintenance, for example, AtFtsH7/9 and SynFtsH4 located in Group 2. Furthermore, most eukaryotic FtsH sequences were located in Group 2 and Group 3, except photosynthetic eukaryotes, which additionally contained Group 1 FtsH acquired from the primary cyanobacterial endosymbiont. Notably, a case of horizontal gene transfer of a Group 2 FtsH sequence from an ancestral eukaryote to the ancestor of the phylum Bacteroidetes was detected, and this seemed to be the dominant FtsH of this phylum.

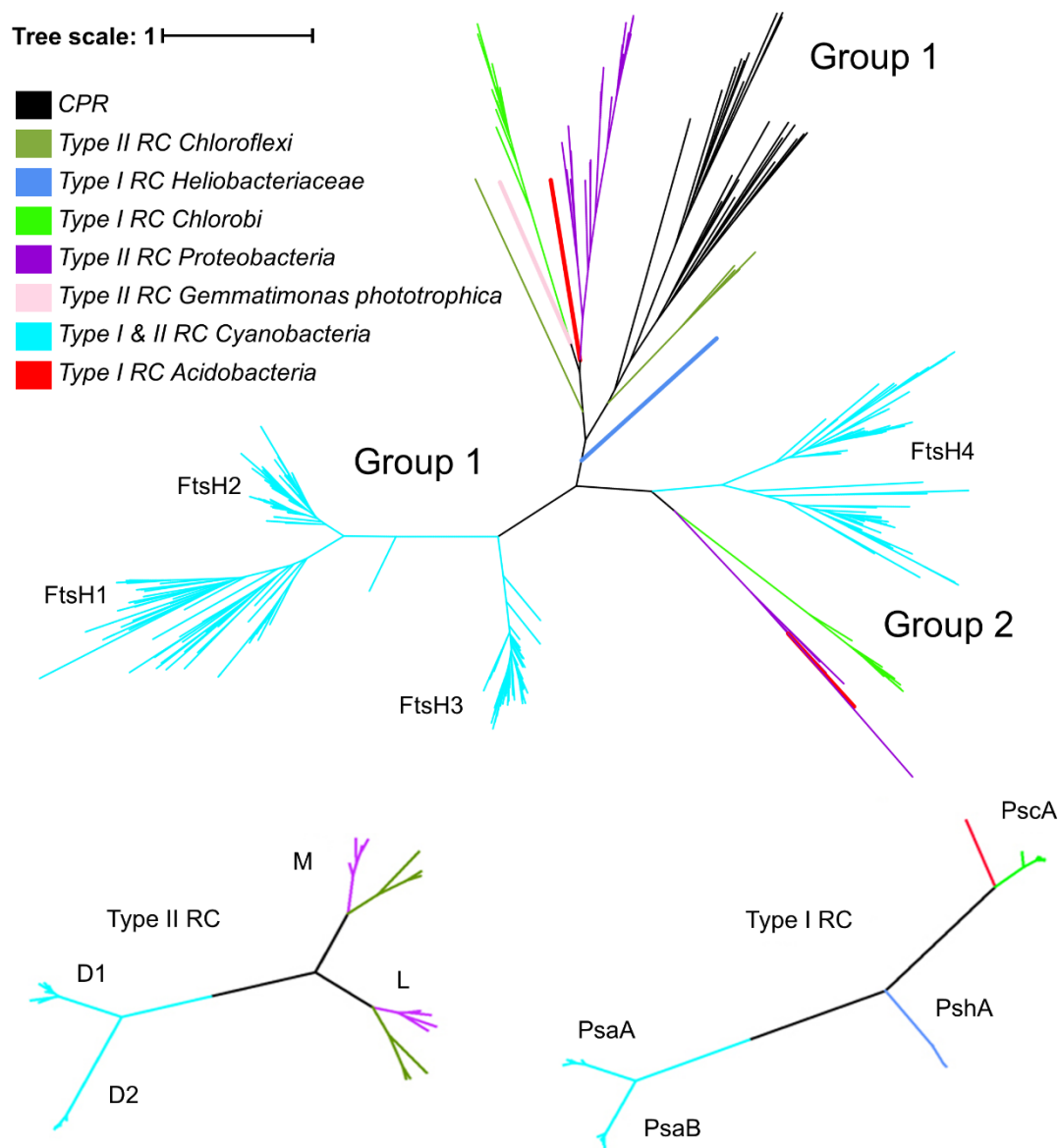


Figure 4-3 Phylogenetic tree of FtsH proteases from phototrophic groups (top) compared to those from Type II (bottom left) and Type I (bottom right) reaction centre proteins reprinted from Cardona (160).

A small number of sequences from a distant non-phototrophic clade were added to improve phylogenetic signal. The bar indicates amino acid substitutions per site.

Figure 4-3 shows a Maximum Likelihood phylogenetic tree calculated using the FtsH proteases of known phototrophic bacteria. On the left, the FtsH involved in PSII repair make a monophyletic group. The genome of *Heliobacterium modesticaldum* encoded a single FtsH branching basally among other Group 1 FtsH proteases of anoxygenic phototrophic bacteria. Proteobacteria, Acidobacteria, and Chlorobi had both Group 1

and Group 2 FtsH proteases. In both Group 1 and Group 2 the sequence from the phototrophic acidobacterium, *Chloracidobacterium thermophilum* branched out as a sister group to the Proteobacteria, and the Chlorobi sequences branched prior to the Acidobacteria and Proteobacteria split. This pattern is also observed in Figure 4-2 which also contained non-phototrophic representatives of the same phyla, and is consistent with previous phylogenetic (171, 184, 265-267) and phylogenomic (262-264, 268-273) studies that have repeatedly confirmed the Acidobacteria as a sister clade of the Proteobacteria, or branching within the Proteobacteria as a sister clade of the Deltaproteobacteria. At the same time, the nearness of the Chlorobi to Acidobacteria and Proteobacteria is also consistent with the Chlorobi-Bacteroidetes-Fibrobacteres supergroup bifurcating prior to the radiation of the Proteobacteria (263, 264, 268, 270, 273). Similarly, the phylogenetic proximity of the Gemmatimonadetes to the Chlorobi has also been demonstrated (264, 274), which is consistent with this group obtaining Type II reaction centres via horizontal gene transfer from the Proteobacteria (274).

When the phylogeny of FtsH from phototrophic bacteria was compared to the phylogeny of Type I and Type II reaction centre proteins (275, 276) (Figure 4-3), it was observed that the trees for the reaction centre proteins follow an identical topology to that of the FtsH sequences. For example, oxygenic Type II RC subunits, cyanobacterial D1 and D2, branch out before the divergence of L and M subunits of the anoxygenic Type II reaction centres of the Chloroflexi and Proteobacteria. Similarly, the PSII specific FtsH proteases (FtsH2/3) branch out before the divergence of Group 1 FtsH in Chloroflexi and Proteobacteria. Cyanobacterial PsaA and PsaB, the core subunits of Photosystem I, also branch out before PshA and PscA of anoxygenic homodimeric Type I reaction centre proteins of Heliobacteria, Acidobacteria, and Chlorobi (160, 277); with heliobacterial PshA branching out before PscA. This early branching of PSII and

PSI is also mirrored in the tree of Group 1 FtsH proteases of phototrophs, with PSII-specific FtsH sequences branching out before those present in anoxygenic phototrophs containing Type I reaction centres. In Group 1 FtsH the heliobacterial sequence also diverged before the split of Acidobacteria and Chlorobi. The situation is also repeated if cyanobacterial FtsH4 is compared to that of Group 2 FtsH retained in phototrophs taking into account that Heliobacteria and phototrophic Chloroflexi lack a Group 2 FtsH.

The similarity between Type I and Type II reaction centre protein evolution and that of the FtsH proteases confirms that reaction centre proteins have mostly been inherited via vertical descent from ancestral populations of phototrophic bacteria, with the notable exception of the Gemmatimonadetes phototrophic bacteria, as explained above. This result is in agreement with a birth-and-death model of protein evolution in which new proteins evolve by repeated gene duplication events, but some of the duplicated genes are maintained in the genome for long periods of time while others are eventually lost (278, 279).

4.2.2. Classification of cyanobacterial FtsH paralogs

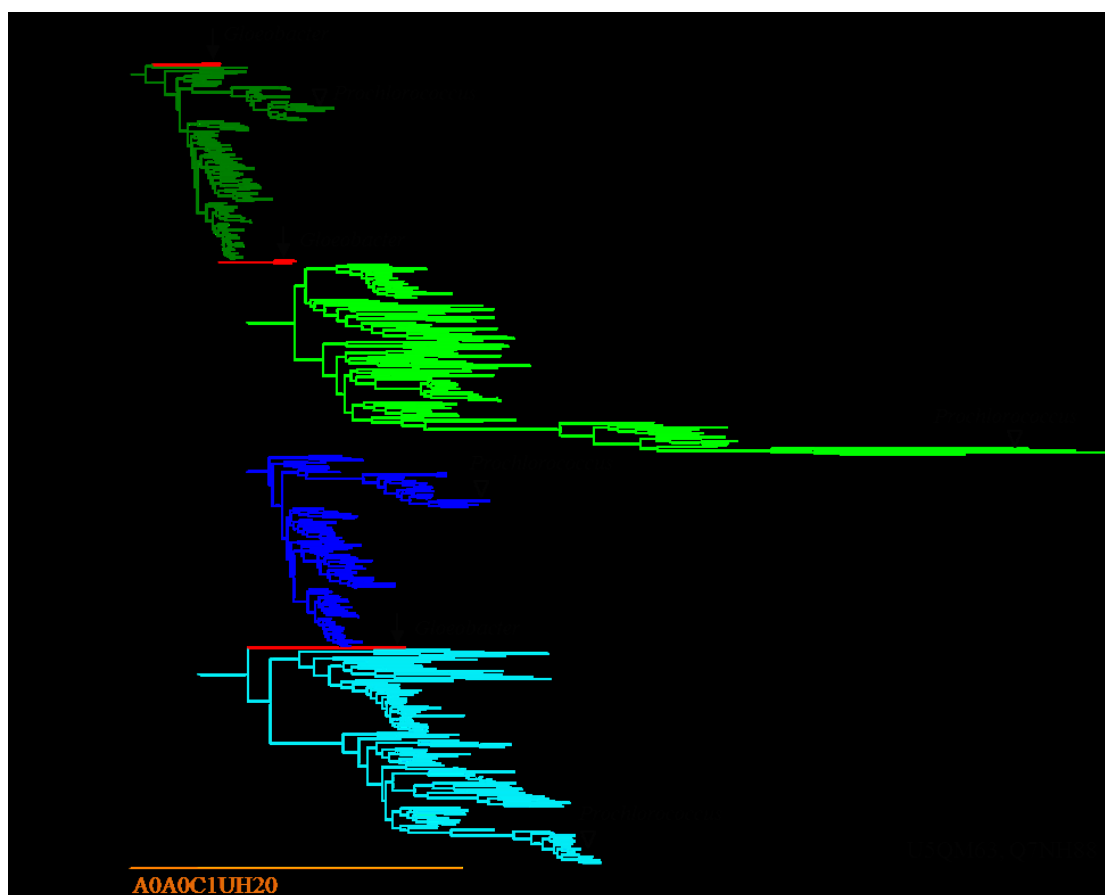


Figure 4-4 Unrooted phylogenetic tree of cyanobacterial FtsH.

Black arrows point to *Gloeobacter* FtsH. Open triangles point to *Prochlorococcus* FtsH. Uncommon branches are labelled by the UniProt Entry of the protein but see the main text. **Gloeobacter* FtsH diverged from the cyanoFtsH1/2 branches is considered as both cyanoFtsH1 or cyanoFtsH2 in the counting.

Previous studies (18, 21, 120, 280, 281) have shown that in *Synechocystis* the degradation of D1 during the repair process of PSII is carried out by an FtsH heterocomplex made of FtsH2 and FtsH3. Similar functions have been identified in their orthologs from green algae (225) and plants (20, 224). In this work, FtsH2 and FtsH3 were phylogenetically positioned together as an early branch of Group 1 (Figure 4-2B). It is, therefore, reasonable to expect that uncharacterized orthologs of FtsH2 and FtsH3 in other cyanobacteria and photosynthetic eukaryotes will also be involved in PSII repair. Figure 4-4 shows a phylogenetic analysis of all FtsH found in 103

sequenced strains of Cyanobacteria. This phylogeny was inferred to classify the diversity of cyanobacterial FtsH homologues into distinctive orthologous groups and identify species lacking specific types of FtsH. As shown in Figure 4-4, four distinctive groups were clearly resolved, renamed after *Synechocystis*'s FtsH1/2/3/4 as cyanoFtsH1/2/3/4.

Figure 4-4 shows that the last common ancestor of extant cyanobacteria had three FtsH paralogs. One ancestral to cyanoFtsH1 and cyanoFtsH2, a second one which was ancestral to cyanoFtsH3, and a third one which was ancestral to cyanoFtsH4. Figure 4-4 also reveals that the duplication event that gave rise to cyanoFtsH1 and cyanoFtsH2 occurred after the divergence of the genus *Gloeobacter* which is known for its primitivity in the composition of photosynthetic proteins and the lack of thylakoid membrane (282, 283). Each of the four groups is consistent with the known diversification of cyanobacteria, featuring the early branch of *Gloeobacter* (arrow) and long branches for the relatively more rapid evolving clades of the marine *Synechococcus* and *Prochlorococcus* (open triangle) (284-286). The substitution rates (branch length) of different groups are noticeably different, with cyanoFtsH3 displaying the slowest rate and cyanoFtsH4 the highest, suggesting different evolutionary pressures upon them.

Species or strains missing one or more of the four types of FtsH are listed in Table 4-2. Overall, the majority (>99%) of cyanobacteria have a set of cyanoFtsH1/2/3 indicating the crucial role of these three types. The only species lacking cyanoFtsH3 is *Crocospaera watsonii* WH 8501. However, a sequence was later found in this organism (UniProt accession number: Q4BUC6) that is orthologous to cyanoFtsH3, but lacking the M41 peptidase, and therefore this sequence did not meet the selection criteria as described in Materials and Methods. An unusual species is the recently

described and sequenced *Neosynechococcus sphagnicola* (287), which seems only to have a cyanoFtsH3. Although it was isolated from an unusual environment, a peat bog (284), it is needed to be mentioned that other peat bog cyanobacteria like *Synechococcus* sp. PCC 7502, *Pseudanabaena* sp. PCC 7429 and *Gloeocapsa* sp. PCC 73106 do possess the expected complement of four FtsH paralogs. Also, *Synechococcus* sp. strain JA-2-3B'a (2-13), isolated from hot spring microbial mat (288) was found to lack cyanoFtsH2.

Among those strains that lack FtsH paralogs, the most common version lost was found to be cyanoFtsH4. This relatively common loss of FtsH4 is in line with previous mutagenesis studies in *Synechocystis* that have shown that a knock-out of the FtsH4 did not result in a noticeable phenotypic change, suggesting that the role of cyanoFtsH4 may be redundant or compensated by the other paralogs (289). A noticeable feature of some of the species lacking cyanoFtsH4 is the reduced complexity of metabolism and reduced genome due to symbiotic interactions, such as *Atelocyanobacterium thalassa* (290) and *Richelia intracellularis* (291).

Three FtsH sequences found in cyanobacterial genomes showed an unusual positioning in Figure 4-4, with very long branches. Sequence A0A0C1UH20 was retrieved from the genome of the heterocystous cyanobacterium *Hassalia byssoidea* and probably represents a case of horizontal gene transfer from an uncharacterized bacterium. It was later analysed (data not shown) and placed near the mitochondria-targeted At-FtsH3/10 and FtsH from phylum Bacteroidetes. In addition to this, sequences U5QM63 and Q7NH88 were found in the genomes of *Gloeobacter kilaueensis* and *Gloeobacter violaceus* respectively and did not give the best BLAST hit to any other cyanobacterial sequence. In the phylogeny shown in Figure 4-2B, they were placed in Group 3. Either they represent ancestral cyanobacterial FtsH paralogs now lost in all other strain of

cyanobacteria, or an ancient event of horizontal gene transfer to the last common ancestor of *G. kilaueensis* and *G. violaceus* from a distantly related bacterium of uncharacterized phyla of bacteria. Given the universality and relative high degree of conservation of FtsH across bacteria, and adding to this the fact that early branching genus *Gloeobacter* is not a particularly fast evolving clade of cyanobacteria, it seems unlikely that the divergent position of U5QM63 and Q7NH88 is due to unusually high rates of evolution in comparison to other cyanobacteria FtsH sequences.

Table 4-2 List of cyanobacteria that might have lost one or more FtsH during evolution.

U5QN63 from *Gloeobacter kilaueensis* JS1 and Q7NHF9 from *Gloeobacter violaceus* (strain PCC 7421) are considered as the equivalent of both cyanoFtsH1 and cyanoFtsH2 due to their evolutionary position discussed above. Species with completely sequenced genomes are indicated in bold.

Species	cyanoFtsH1	cyanoFtsH2	cyanoFtsH3	cyanoFtsH4	Note
<i>Neosynechococcus sphagnicola</i>			A0A098TRH5		peat bog (284)
<i>Synechococcus</i> sp. strain JA-2-3B'a (2-13)	Q2JHR8		Q2JNP0		hot spring microbial mat (288)
<i>Aliterella atlantica</i> <i>CENA595</i>	A0A0D8ZU81	A0A0D8ZT40	A0A0D8ZWS3		
<i>Anabaena</i> sp. 90	K7WSA3	K7WS23	K7VZY8		
<i>Atelocyanobacterium thalassa</i> (isolate ALOHA)	D3ENM5	D3EPJ8	D3EQB0		endosymbiont
<i>Candidatus Synechococcus spongiarum</i> 142	A0A0U1QKZ0	A0A0G8B1F9	A0A0G8AVK4		
<i>Candidatus Synechococcus spongiarum</i> SP3		A0A0G2HLH7	A0A0G2J4Y2	A0A0G2IVX6, K9UQL4	
<i>Chrysosporum ovalisporum</i>	A0A0P1C082	A0A0P1BUU7	A0A0P1BVD0		
<i>Crocospira watsonii</i> WH 8501	Q4C3U9	Q4BUM7		Q4BY73	
<i>Cyanobacterium aponinum</i> (strain PCC 10605)	K9Z414	K9Z622	K9Z6W1		

<i>Cyanobacterium</i> <i>endosymbiont of Epithemia</i> <i>turgida</i> isolate EtSB Lake <i>Yunoko</i>	A0A077JFW7	A0A077JK69	A0A077JIP6		endosymbiont
<i>Cyanobacterium stanieri</i> strain PCC 7202	K9YKE4	K9YIN6	K9YQT6		
<i>Geitlerinema</i> sp. PCC 7407	K9SC27	K9S5X2	K9SA34		
<i>Geminocystis</i> sp. NIES- 3708	A0A0D6AE16	A0A0D6ABU5	A0A0D6AGY3		
<i>Oscillatoria acuminata</i> PCC 6304	K9TDN1	K9TGY5	K9TBZ2		
<i>Phormidium</i> sp. OSCR	A0A0P8BT41	A0A0P8C7Y9	A0A0P8BW82		microbial mat (292)
<i>Richelia intracellularis</i>	X5JUF5	X5JFW6	X5JRC0		extracellular symbiont
<i>Richelia intracellularis</i> HH01	M1X0E5	M1X2X0	M1WZS3		Same as above
<i>Richelia intracellularis</i> HM01	M1WNU4	M1WNE7	M1WPH5		Same as above
<i>Rubidibacter lacunae</i> KORDI 51-2	U5DKP8	U5DJI8	U5DKU6		

4.2.3. Multiplicity of FtsH in photosynthetic eukaryotes

Figure 4-2B shows that the last common ancestor of all eukaryotes likely inherited two bacterial FtsH paralogs, one from Group 2 and one from Group 3. Furthermore, Figure 4-2B shows that photosynthetic eukaryotes retained from the cyanobacterial primary endosymbiont at least 3 FtsH paralogs, one of these possibly branching prior to the divergence of cyanoFtsH1/2, or at a point in time when these two had not had enough time to diverge; a second one was homologous to cyanoFtsH3; and a third one was homologous to cyanoFtsH4. This pattern is consistent with recent phylogenetic

evidence suggesting an early branching cyanobacterium as the primary endosymbiont (293). In addition, photosynthetic eukaryotes seem to have independently acquired some other FtsH paralogs, ancestral to AtFtsH1/i2 and AtFtsH5/12, from a third (or more) bacterial donor, closely related to the phylum Firmicutes. This is not surprising as previous studies have shown that early evolving eukaryotes acquired genes from a broad range of bacterial origins beyond that of the mitochondria (294) and plastid ancestors (295). Among all clades in the tree of life, photosynthetic eukaryotes have the largest multiplicity of FtsH subunits (Figure 4-2A), suggesting that, after the establishment of the mitochondria and chloroplast, the ancestral FtsH subunits underwent several duplication events.

As seen in Figure 4-2 it is likely that both organelle ancestors likely possessed FtsH from different orthologous groups. For example, the fungi *Saccharomyces cerevisiae* has 3 FtsH homologues: YeYta12, YeAfg3 and YeYme1, all reported to be targeted to the mitochondria (296). *Arabidopsis thaliana* has 15 FtsH homologues, from these AtFtsH3/4/10 are targeted to the mitochondria, and the other 12 are targeted to the chloroplast. In contrast, AtFtsH11 is thought to be found in both the mitochondria and the chloroplast (219, 222, 223, 297, 298), but its presence in the mitochondria is controversial (299). Figure 4-2B shows that mitochondria-targeted YeYta12/YeAfg3 from *S. cerevisiae* and AtFtsH3/10 from *A. thaliana*, which are closely positioned in Group 2, have a common origin, although the gene duplication events that gave rise to the divergence of AtFtsH3 and AtFtsH10, and the divergence of YeYta12 and YeAfg3 are distinct. YeYme1 from *S. cerevisiae* is closely related to AtFtsH4 and AtFtsH11, all clustering in Group 3. Thus, from a phylogenetic perspective, it would seem as if AtFtsH11 was of mitochondrial origin and was later co-opted to support chloroplast function.

AtFtsH1/5 appear to have originated from cyanoFtsH3, and AtFtsH2/6/8 from an ancestral cyanoFtsH1/2. These data are in line with experimental data showing a common role for SynFtsH2/3 and AtFtsH1/2/5/8 in PSII repair (20, 120, 227, 280, 300). AtFtsH6 is involved in regulating acquired thermotolerance, or “thermomemory”, by degrading the plastidial heat shock protein HSP21 (249). *In vitro* evidence also suggests AtFtsH6 degrades LHC II, the light-harvesting complex of PSII (301). However, this potential role is not observed *in vivo* (302). Lastly, AtFtsH7/9, in turn, have greater similarity to cyanoFtsH4, which has been localised in the thylakoid membrane and possibly cytoplasmic membrane in GFP-tagging experiments carried out in *Synechocystis* (158), but whose function is poorly characterised. The presence of AtFtsH7/9 in the chloroplast envelope is suggested by proteomic evidence (303, 304). Previous studies have shown that at least AtFtsH1/2/5/7/8/9 are all targeted to the chloroplast (224), which is consistent with their cyanobacterial origins.

From the set of FtsH paralogs found in photosynthetic eukaryotes, AtFtsH12, AtFtsHi1, AtFtsHi2 and AtFtsHi5 have no evolutionary counterparts in cyanobacteria, but instead, they show clear proximity to FtsH subunits found in the phylum Firmicutes that cluster in Group 3. A distinct feature of AtFtsHi1/i2/i5 is an altered catalytic zinc ion binding site, which is observed in all FtsH homologues within the AtFtsHi1/i2-comprising green clade in Figure 4-2; most but not all (e.g. FtsH12) FtsH within the AtFtsH12/i5-comprising green clade in Figure 4-2 also have the zinc binding site disabled. It has been reported that AtFtsHi1/i2/i5 are essential for the biogenesis and division of chloroplast (298, 305). However, their precise function and mechanism still remain to be characterised.

Both the red alga *Cyanidioschyzon merolae* and the green alga *Chlamydomonas reinhardtii* contain 5 FtsH homologues and lack counterparts of AtFtsHi1/i2/i5/12. Two

FtsH are positioned in the PSII-specific clades of AtFtsH1/5 and AtFtsH2/8/6, respectively; one corresponds to the AtFtsH3/10; one corresponds to the AtFtsH4/11. Notably, *Chlamydomonas reinhardtii* has a counterpart of the AtFtsH7/9, or SynFtsH4, while *Cyanidioschyzon merolae* has lost it but instead possesses an additional counterpart to AtFtsH3/10. This observation would suggest that the diversification of FtsH seen in some land plants postdates the divergence of higher plants and algae.

4.3. Structural characteristics of the cyanobacterial FtsH

4.3.1. The structural conservation of FtsH in domain Bacteria

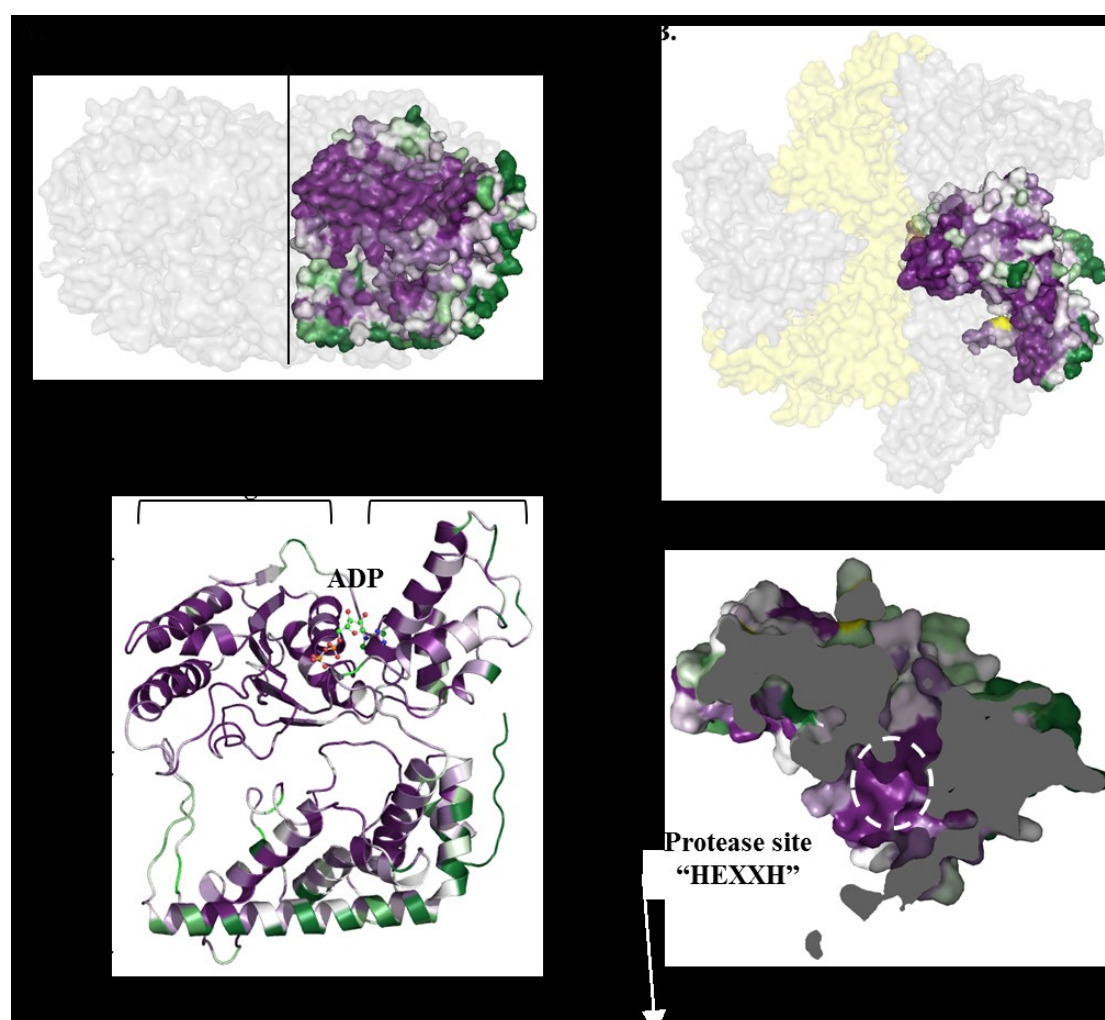


Figure 4-5 Structural conservation of FtsH protease sampled from 55 phyla of bacteria.

The crystal structure model submitted for ConSurf analysis was *T. thermophilus* FtsH (PDB: 2DHR). Purple represents highly conserved regions while green represents poorly conserved regions. Figure A shows a top view, while Figure B shows a side view. A single monomer is highlighted while the other

five as shown with transparency. The two yellow-coloured monomers highlight the three-fold symmetric structure. Dashed circle highlighted the protease activity site “HEXXH”.

As shown in Figure 4-2, the FtsH proteins found in photosynthetic organisms, especially those involved in PSII repair, are phylogenetically distant to the FtsH proteins of known crystal structures. Therefore, before applying any structural information in studying the FtsH proteins from phototrophs, we need firstly a reference of structural conservation of the FtsH protein family. A Consurf analysis was performed to assess structural conservation of the cytosolic region of all types of bacterial FtsH using 270 sequences randomly selected from species covering 55 bacterial phyla or candidate phyla. As shown in Figure 4-5, the AAA+ domain of bacterial FtsH is well-conserved (purple) with only a small highly divergent area (green) on the surface towards the membrane or exposed to the cytoplasm. The surface that interacts with the adjacent AAA domain is strictly conserved, which suggests that a ring-shaped structure and even the hexameric complex structure are a universal feature of all bacterial FtsH. The protease domain, however, was more divergent except the strictly conserved protease tunnel comprising the “HEXXH” motif. Based on the above observation, it is reasonable to deduce that the AAA+ domain is fundamentally important for the structure and function of FtsH. On the other hand, the variability of the protease domain underlies the multiplicity of FtsH and possibly substrate specificity, as well as the specific interactions between FtsH paralogs within the same organism. It should be emphasised that this deduction is based on the cytosolic region of FtsH.

The formation of heterocomplexes in cyanobacteria and chloroplast and the formation of heterocomplexes in the mitochondria are probably the result of convergent evolution. The yeast mitochondrial FtsH paralogs YeYta12 and YeAfg3 adopt a heterohexameric structure, the so-called m-AAA protease complex, where “m” refers to the matrix. Although SynFtsH3 and SynFtsH1/2, as well as their orthologs in the chloroplast, also

form a heterohexameric structure, the heterocomplexes in cyanobacteria and chloroplasts are evolutionarily distant from the mitochondrial heterocomplex, as the YeYta12/YeAfg3 proteases emerged from Group 3, while SynFtsH1/2/3 emerged from Group 1. The presence of heterocomplexes within different orthologous groups indicates a convergent evolution from homocomplex to heterocomplex. Intriguingly, Lee *et al.* showed that the mutation of just two residues in the protease domain was enough to enable YeYta12 to form a homocomplex (258). It is also noteworthy that FtsH forms homocomplexes in most bacteria that have been studied (255, 256, 259, 306), including cyanobacteria (e.g. SynFtsH4 in *Synechocystis*), as well as mitochondria (e.g., YeYme1).

The ConSurf analysis revealed a high degree of conservation in the AAA+ domain of FtsH from all bacteria phyla or candidate phyla, while the protease domain is comparatively more diverse. This finding suggests that the unfoldase activity of the AAA+ ring of FtsH is fundamental to all FtsH homologues, while the specialisation of the protease domain is the key to the specialisation of different FtsH proteases. This is in line with the fact that AtFtsHi1/i2/i5 have lost the “HEXXH” motif but the AAA+ domain is not much altered. The membrane-anchoring domain of FtsH could also play an important role in localisation and substrate recognition (254), but this suggestion needs further evidence.

4.3.2. Cyanobacterial characteristics near the structurally conserved regions of FtsH

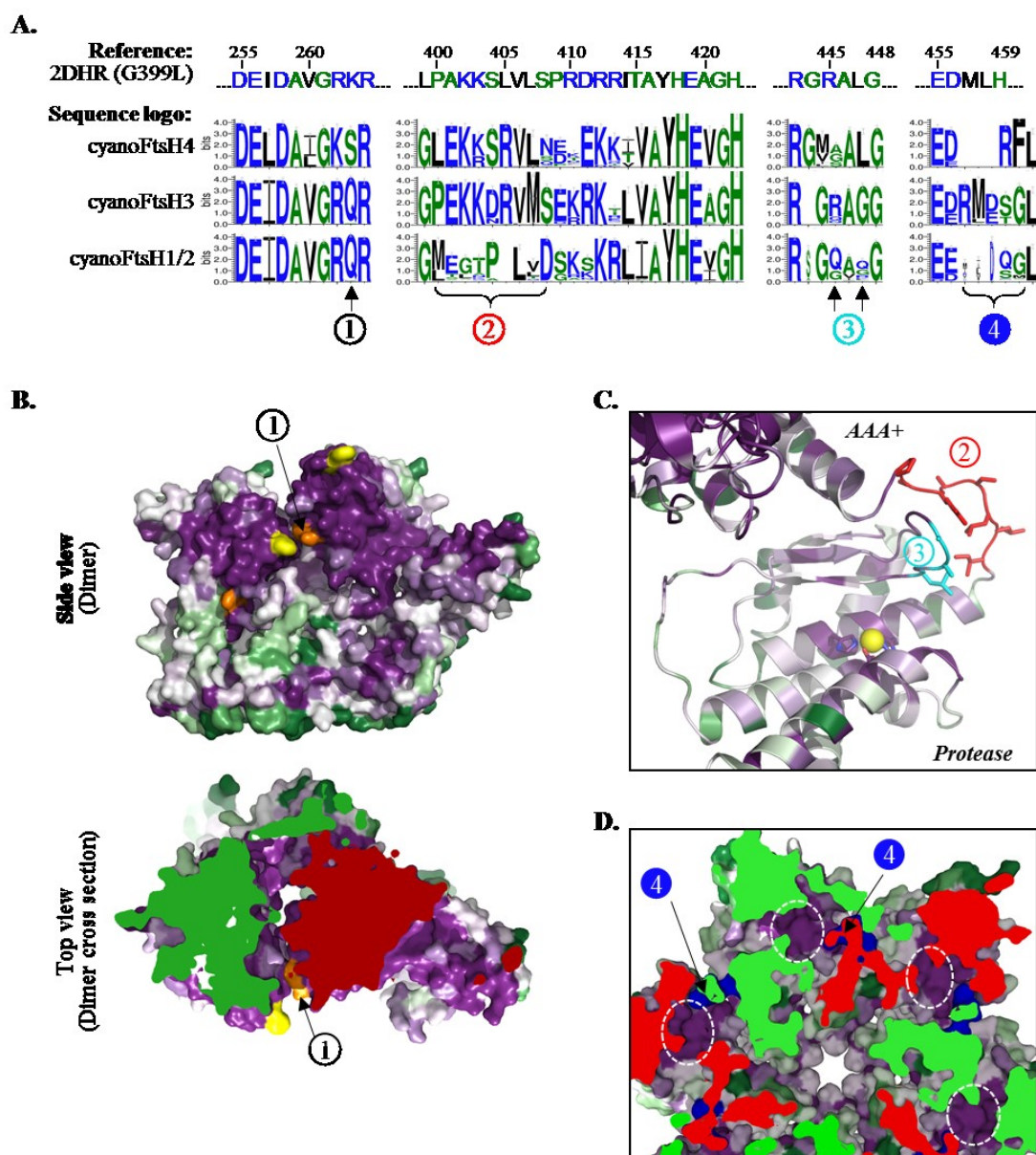


Figure 4-6 Conservation and comparison of the cyanobacterial FtsH paralogue groups.

A. Sequence logo profiles of four types of cyanoFtsH proteases. Y axis is the information content in bits of the composition of amino acids; the size of a character, single-letter code of amino acid, represents how conserved that amino acid is at that position, the larger the letter, the more conserved it is. Colour scheme: green, neutral residues; black, hydrophobic residues; and blue, hydrophilic residues. The amino acid numbering is the same as in the crystal structure of *T. thermophilus* FtsH (PDB: 2DHR). Four sites of significant changes between groups were indicated with arrows or brackets. B. Position 1 was shown in orange to highlight its relative position to the pore phenylalanine shown in yellow. The interior of two monomers was coloured red and green respectively. C. Positions 2 and 3 are emphasised in red and cyan respectively, and selected residues were shown in stick format. The structure from *T. maritima* (PDB: 3KDS) was used for depiction. D. Top view (cross section) of Position 4 showed in the structure 2DHR. The conserved “HEXXH” motif was indicated by dashed white circles.

Following the phylogenetic typing of 417 cyanobacterial FtsH into four groups, the structural conservation of each group was examined via ConSurf analysis to identify conserved and divergent regions of each group. Sequence logo profiles were generated to allow the comparison of conserved regions. The cyanoFtsH1 and cyanoFtsH2 sequences were grouped for generating the sequence logo profiles due to their evolutionary proximity, as described above. Four sites that are well conserved within a paralogous group but significantly divergent between paralogous groups were discovered, implying that differences in functions may be governed by the specific structural differences in these sites. These four sites are shown in Figure 4-6A and are described below.

Position 263 (*T. thermophilus* FtsH numbering) is a strictly conserved glutamine in cyanoFtsH1, cyanoFtsH2, and cyanoFtsH3, but a serine in cyanoFtsH4. As illustrated in Figure 4-6B, this residue is predicted to lie on a highly conserved surface within a short distance (~ 9 Å in structure 2DHR from *T. thermophilus* (256)) to the pore phenylalanine from the adjacent subunit. A similar proximity to the pore residues was also observed in the structure 2CEA (from *T. maritima* (255)). In the hexameric structure of *Synechocystis* FtsH2/3 heterocomplex, this conserved glutamine would be present in all six monomers, and it is, therefore, likely to occupy a similar position close to the pore residues. This residue might participate in polypeptide substrate recognition or translocation.

Position 400-408 and position 445-447 are proximal to the flexible glycine (G399 in *T. thermophilus* (256, 306) or G404 in *T. maritima* (257)) and lid helix (257, 307) regions respectively (Figure 4-6C). These two regions are structurally flexible, and their intradomain movements are crucial to functionality (257, 306, 307). These two regions are spatially close to each other in the *apo*-FtsH structure (257) (Figure 4-6C).

Mutations in the lid helix lead to a decrease of not only protease activity but also the ATPase activity (307), strongly indicating the lid helix and flexible glycine interact with each other. CyanoFtsH4 possesses a conserved leucine at position 400 and 447, in contrast to cyanoFtsH3 which has a highly rigid proline at position 400 and flexible glycine at position 447. Mutagenesis studies have shown that reduced flexibility at position 399 facilitates crystallisation (256) and affects ATPase and protease activities (257, 306) in the homocomplex. It is noticed that cyanoFtsH1/2 also has this rigid proline, at position 404. On the contrary, cyanoFtsH4 does not have this proline nearby position 400. It is speculated that this proline might contribute to the formation of a heterocomplex, which is worthy of further investigation

Position 457-459 shows great variation among groups with the deletion or insertion of residues. Figure 4-6D shows that this region is likely to be in proximity to the protease active site of the adjacent subunit in all six monomers. Changes in length and composition likely alter accessibility or specificity of substrate processing.

These four sites represent attractive targets for mutagenesis studies, and studies in this direction might provide key information on the formation of FtsH heterocomplex and the structure and function of FtsH homocomplex in oxygenic phototrophs, specifically cyanoFtsH4 and its eukaryotic counterparts.

4.4. Preliminary work on the function of FtsH4 in *Synechocystis*

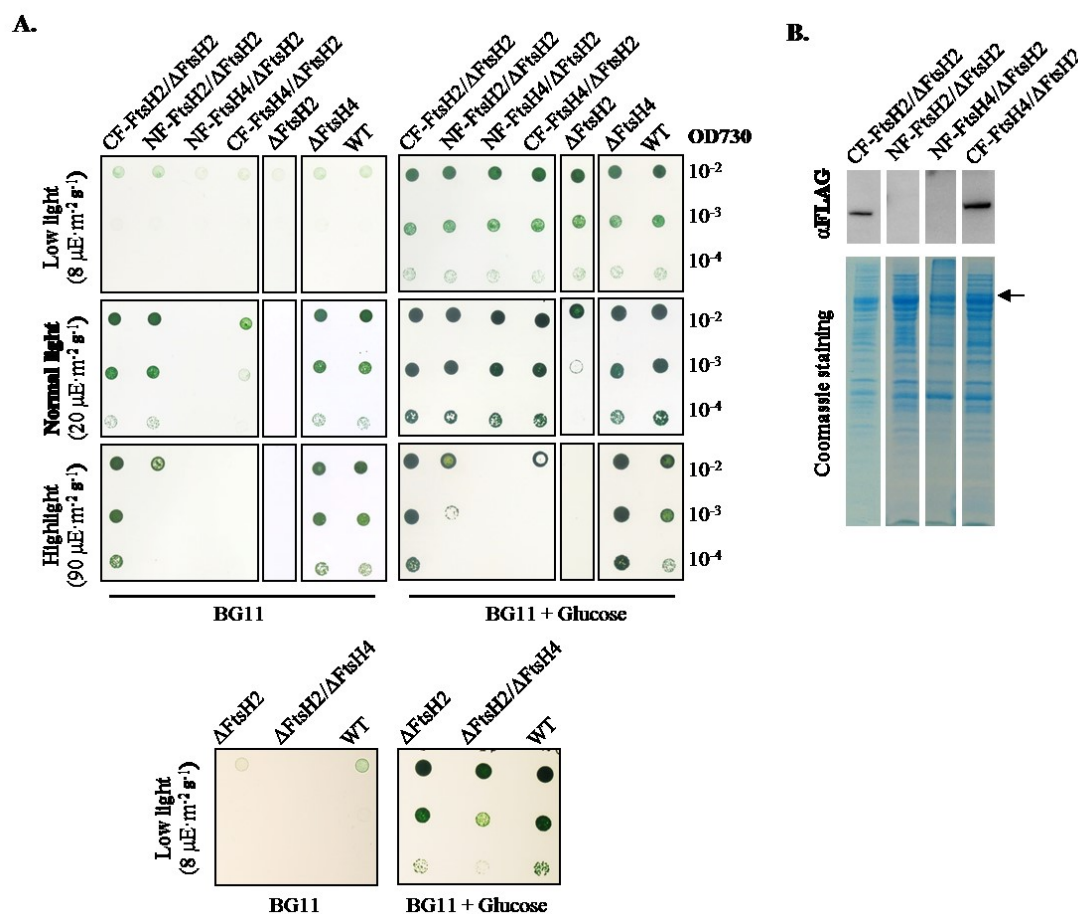


Figure 4-7 Complementation assay of FtsH4 in the FtsH2 null *Synechocystis*.

“CF-” represents “C-terminal Flag-tagged”, “NF-” represents “N-terminal Flag-tagged”. A. Growth and complementation assay of FtsH4 mutants. The upper panel and lower panel are from two independent 9-day growth assays. B. Western blotting of the Flag-tagged FtsH2 and FtsH4 proteins in the thylakoid membrane.

The physiological role of FtsH4 in *Synechocystis* is still unclear. To test for a possible physiological role in photoprotection, photoautotrophic growth of a $\Delta\text{FtsH2}/\Delta\text{FtsH4}$ double mutant (constructed by Dr. Jianfeng Yu at Imperial College London) was examined and compared to the growth of single mutants and WT. As shown in Figure 4-7A, the ΔFtsH4 single mutant grew as well as WT at low, normal and high light intensities, while the growth of the $\Delta\text{FtsH2}/\Delta\text{FtsH4}$ double mutant was more sensitive to low light than the ΔFtsH2 single mutant under mixotrophic growth condition. Under low light photoautotrophic growth condition, the growth of all strains was slower, and

no significant growth of the Δ FtsH2/ Δ FtsH4 double mutant was observed after 9 days. These data suggest that FtsH4 is dispensable for photosynthesis and might contribute to photoprotection, although it is not as important as FtsH2.

In the hope of finding potential substrates of FtsH4, a series of N-/C-terminal Flag-tagged FtsH4 mutant strains were constructed by Dr. Jianfeng Yu. Preliminary work on these mutants showed that by placing the *ftsH4* gene at the *psbA2* locus, overexpression of FtsH4 complemented to some extent the *ftsH2* deletion (Figure 4-7A). As shown in Figure 4-7B, the C-terminal Flag-tag is detected by Western blotting while the N-terminal Flag-tag is not. An overexpressed band at the approximate position of N-/C-terminal Flag-tagged proteins was observed in all cases.

Modification of the N-terminus of *E. coli* FtsH does not block the accumulation, functionality or oligomerization of FtsH (308). Therefore, it remains possible that the overexpression of Flag-tagged FtsH2 and FtsH4 is successful, but the N-terminal Flag-tag is cleaved. Indeed, N-terminal Flag-tagged FtsH2 complemented the FtsH2 null mutant, although some light sensitivity was still noticeable at high light (Figure 4-7A).

FtsH2 and FtsH4 are predicted to have two transmembrane segments according to the UniProtKB database. Therefore, the N-terminus is expected to be on the same side as the enzyme. In the FtsH4 overexpression mutants, only the C-terminal Flag-tagged mutant showed improvement in the tolerance of light. This increased tolerance might be due to improved PSII maintenance or a quantitative boost of unknown functions outside of PSII. Further experiments especially the Flag-tag pull-down assay are crucial for revealing the function of FtsH4.

4.5. Discussion

4.5.1. The overlooked evolution of FtsH

The multiplicity of FtsH in cyanobacteria and photosynthetic eukaryotes has long been noticed and discussed (218). In other bacteria, however, the multiplicity of FtsH is not clarified and sometimes neglected. For instance, FtsH has previously been considered a well-conserved single-orthologous group suitable for determining bacterial phylogeny (309, 310). Although there is a high degree of sequence conservation in the FtsH family, the FtsH protein family is not necessarily monophyletic during evolution.

In this work, it is shown that there are at least three different orthologous groups of FtsH in eubacteria and eukaryotes. Moreover, the multiplicity of FtsH in photosynthetic eukaryotes is likely attributed to the early endosymbiotic acquisition of organelles of different bacterial origins and later gene duplications. As discussed above, the first duplication event of FtsH predates the radiation of most bacteria phyla, which means different FtsH paralogs have been influencing and participating in the evolution of eubacteria and eukaryotes since an early stage. This can be reflected by the fact that the divergence of FtsH can resemble the divergence of photosynthetic reaction centres which is one of the earliest events during the evolution of life (184, 311). Importantly, FtsH proteases involved in photosystem II repair make a distinct clade branching out before the divergence of FtsH proteases found in all groups of anoxygenic phototrophic bacteria. This early branching is in consistent with the early origin of water oxidation chemistry (184, 189) rather than a late origin of oxygenic photosynthesis (159, 178).

Several structural models are now available for soluble portions of the FtsH complex (256, 257, 259, 306) which have greatly improved our understanding of this molecular machine. Here I show that those structurally well characterised FtsH (Figure 4-2B, black labels) proteins are in distant evolutionary positions to the FtsH responsible for

PSII repair. In addition, although heterocomplexes of FtsH have been observed in both mitochondria and chloroplast, they are evolutionarily distinct and are likely to be the consequence of convergent evolution. This work describes for the first time a clear vision of the evolutionary relationship between well-characterised FtsH and the PSII-specific FtsH, which is crucial for comparative studies on the PSII-specific FtsH heterocomplex. Here I show by ConSurf analysis that the AAA+ ring of the FtsH family is well-conserved throughout all bacteria. However, the protease domain, especially the interface between subunits, is diverse. By firstly classifying the cyanobacterial FtsH paralogs, a further comparison of the structural characteristics of these paralogous groups is made possible, and a few distinctive features are revealed.

4.5.2. Mutations might explain the diverse symmetries seen in FtsH crystal structures

The structural conservation profile of FtsH is useful in assessing the impact of mutations commonly deployed in structural studies of FtsH. Figure 4-8A shows three known crystal structures of FtsH, within them the protease domain of FtsH adopts a constant near-C6 symmetry irrelevant to the binding of nucleotide, while the ATPase domain, depending on the status of nucleotide-binding, varies between C2 (2CEA), C3 (2DHR) and C6 (3KDS) symmetries. Although current models of FtsH function are based on these structures, it should be mentioned that all of them harbour mutations. Intentional mutations are commonly introduced to facilitate the crystallisation process, by reducing structural flexibility or freezing enzymatic activity.

In the C3-symmetry structure from *T. thermophilus* FtsH (2DHR), an absolute conserved flexible glycine is mutated to a leucine G399L, in the hope of decreasing chain flexibility (256). This mutation is not favoured by the conservation analysis conducted in this Chapter. Although it is claimed neutral to the protease activity and

even positive to the ATPase activity in *T. thermophilus* (256), the same mutation reproduced in *T. maritima* (257), *T. thermophilus* and *A. aeolicus* FtsH (306) significantly depressed both protease and ATPase activities and affected oligomerization (257). Therefore this structure is likely not to be physiological.

In the C2-symmetry structure, 2CEA, the third zinc ligand is mutated, D500A, to block protease activity and neutral mutations K410L and K415A were introduced in the hope to improve crystallisation reproducibility. This three-mutation FtsH does form hexamers and preserves ATPase activity (255); it is therefore reasonable to claim that these mutations are neutral to physiological structure and C-2 symmetry is the general form of the FtsH structure. A recent structure from *A. aeolicus* FtsH also shows a C2-symmetry, although it harbours four unintentional mutations (I250M, F360L, K552R and E627G) and is not enzymatically active (306). Nevertheless, these mutations are not within the flexible linker and lid helix regions described in this work. Therefore it is likely that these mutations had not altered the symmetry of the complex.

In the C6-symmetrical *apo*-FtsH, 3KDS, a single mutation K207A was introduced into the Walker A motif to prevent ATP binding. This mutation is within a helix of the AAA wedge subdomain (Figure 4-1), so no significant impact on the structure of *apo*-FtsH is expected from this mutation. Collectively, the C2- and C6-symmetric structures likely depict more realistic sketches of FtsH homocomplex in action than the C3-symmetric structure due to less structural interference from the mutations.

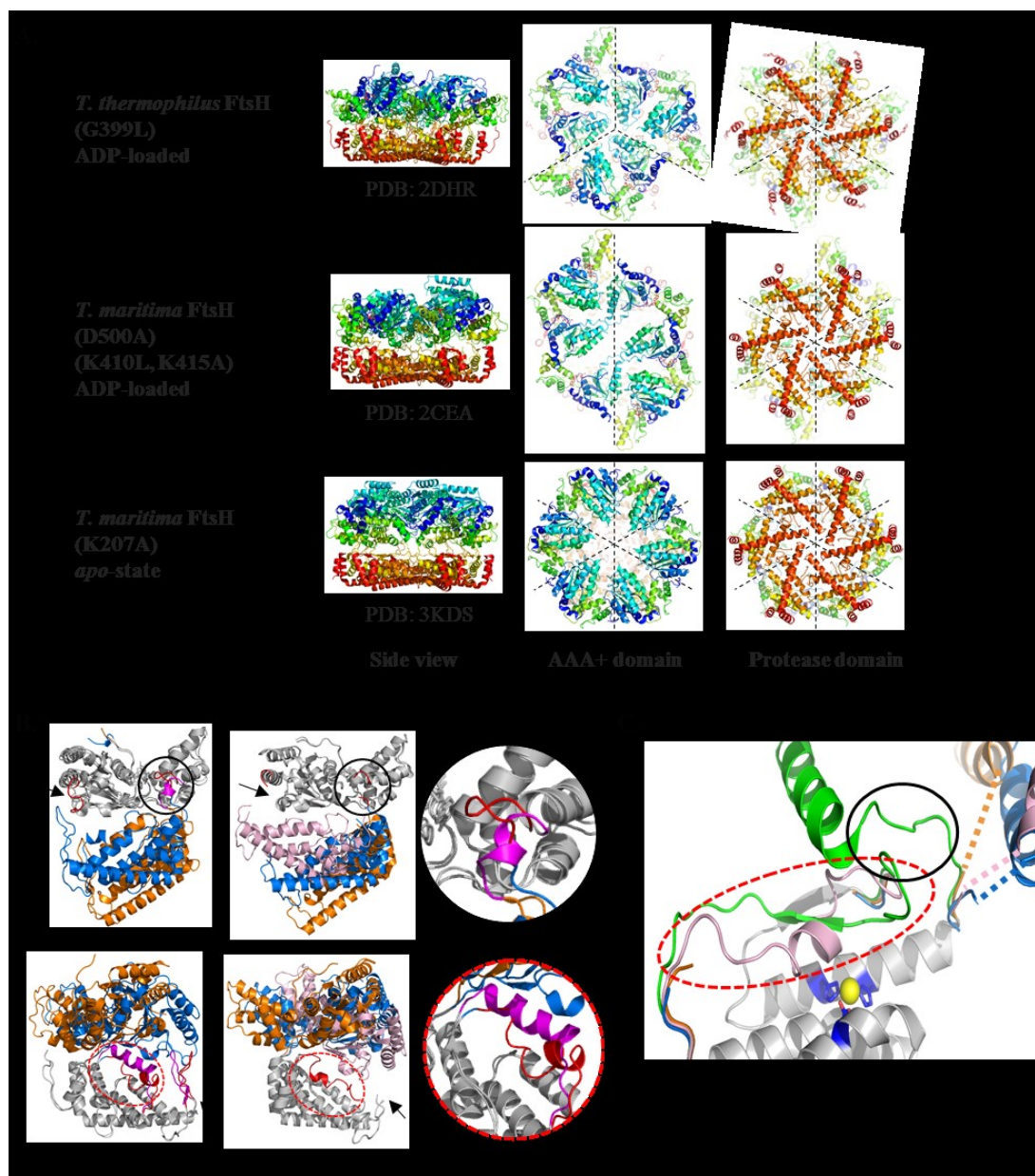


Figure 4-8 Different conformational statuses of FtsH complexes and monomers.

A. Three representative structures showing the different symmetries of FtsH complexes. Symmetric units are indicated by dash lines. B. Structural alignments of individual subunits of different conformations. C, close-up view showing the movement of the linker and lid helix regions from *apo*-FtsH to three ADP-bonded statuses. Dashed straight lines indicate disordered linker. The black line and red dash-line circles point to the same region as in Figure B. Colour scheme of B and C: orange, marine and pink indicate protease (top two views) or ATPase (bottom two views) domains of subunits of different conformations. The circled regions are showed separately in zoom view. Grey indicates the aligned and unchanged regions. Red and magenta are used to differentiate the structural movements.

4.5.3. The flexible linker and lid helix in the FtsH complex might interact

The flexible linker and lid helix, region No. 2 and No. 4 (Figure 4-6), respectively, are revealed in this work as two of the four well conserved characteristic regions that can

differentiate cyanoFtsH1/2, cyanoFtsH3 and cyanoFtsH4. Intriguingly, intradomain movement of these two regions are observed in different crystal structures (Figure 4-8B, C). These intradomain movements are revealed by structural alignment of the ATPase domain and protease domain, respectively, from subunits at different conformations.

In the ATPase domain, two sites are poorly resolved in all available structures likely due to a high degree of flexibility. One of them is the flexible linker connecting ATPase and protease domains, corresponding to M398-L405 of *T. thermophilus* FtsH. The structure of this region is only resolved in the above mentioned G399L mutant FtsH (2DHR) and the *apo*-FtsH, likely due to higher stability. Also, the G399L mutation likely abolishes a third structural form in 2DHR comparing to 2CEA (Figure 4-8B, top row, pink).

In the protease domain, the lid helix corresponds to V441-W460 of *T. thermophilus* FtsH, and it is likely an allosteric site responding to the binding of substrates (257). In the structure of *apo*-FtsH from *T. maritima* (3KDS), this region is mainly a 2-strand β sheet. When ADP is bound, a helical conformation is found in this region in 2 of 6 independent monomers (312). Additionally, in another ADP-bound structure, the lid helix shows a distinct conformational change between the so-called “open” and “close” configurations for substrate processing (256). The counterparts in FtsH structures from *A. aeolicus* (2DI4, 4WW0) and *T. maritima* (2CEA), however, are disordered therefore not resolved. Vostrukhina *et al.* argue that this lid helix, or “active-site switch β -strand” as they describe it, is also disordered in the G399L mutant FtsH (2DHR) by re-analysing its electron density (306). Mutagenesis experiments have shown that modifications in this region abolish the functionality of the protease (306, 307). Intriguingly, a G448P mutation in *T. thermophilus* FtsH disrupts not only the protease activity but also the

ATPase activity (307). It is possible that loss of ATPase activity in G448P is due to a direct or indirect interaction between the lid helix and flexible linker. In *apo*-FtsH (3KDS), the lid helix is shown in proximity to the flexible linker (Figure 4-8C).

As shown in Figure 4-6, the flexible linker and lid helix are two of the four distinguishing features among cyanobacterial FtsH. CyanoFtsH3 differs from cyanoFtsH1/2 and cyanoFtsH4 in that it possesses a highly rigid proline in the flexible linker region. CyanoFtsH4 has an extended lid helix comparing to cyanoFtsH1/2 and cyanoFtsH3. Based on the *apo*-FtsH structure (3KDS), this one residue extension of lid helix is in proximity to the zinc ion of the protease site, increasing the possibility of a nucleophilic attack at the lid helix. Indeed, self-degradation of *T. elongatus* cytosolic region of FtsH4 but not FtsH2, FtsH3 or FtsH2/3 mixture was noticed in unpublished work from my MRes thesis (313). Overall, further investigation on the role of these two elements is crucial for understanding the action of cyanobacterial FtsH.

4.5.4. A speculated action model of the FtsH protease complex

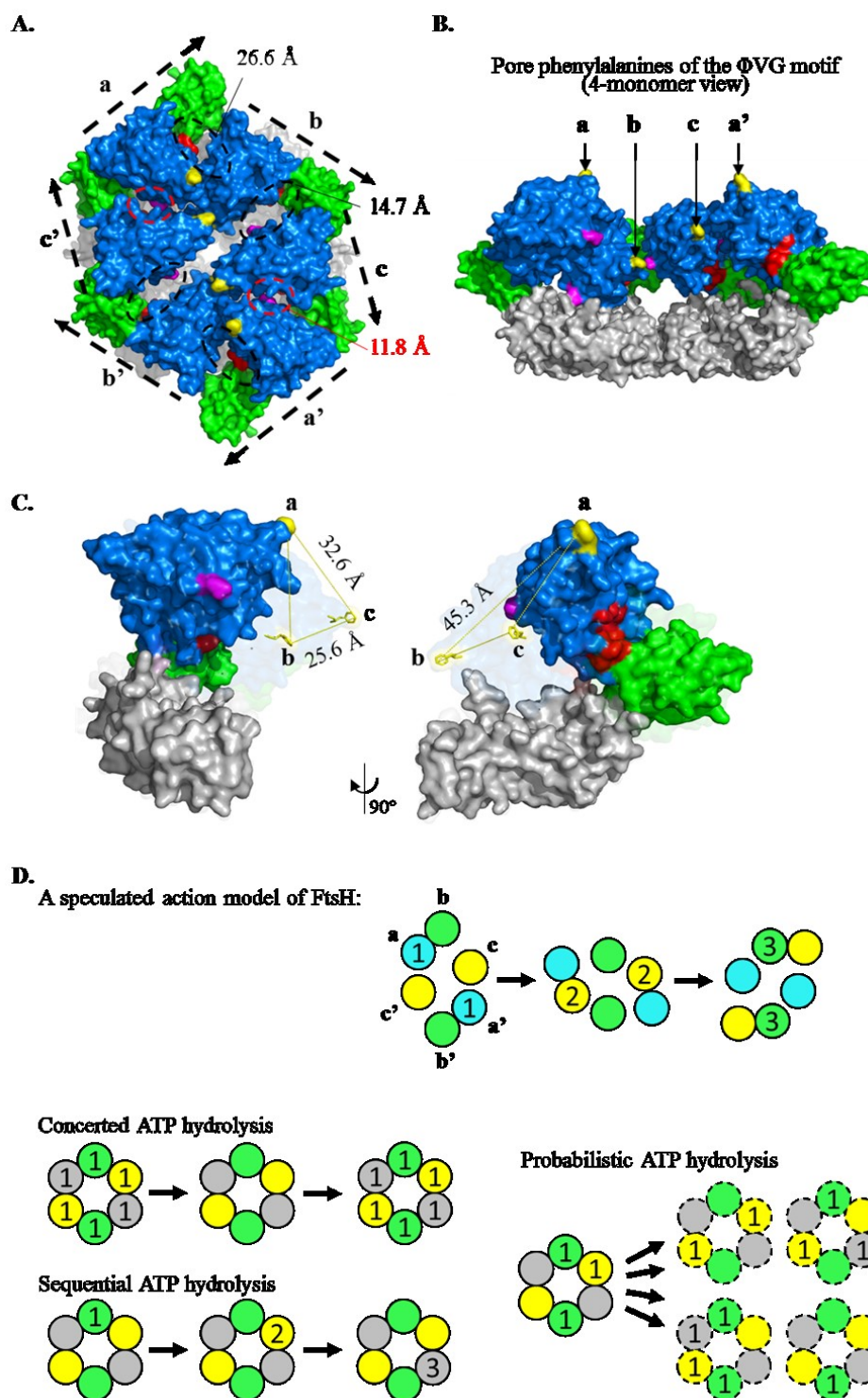


Figure 4-9 A structure-based working model of FtsH protease complex.

A. Top view of the ADP-bonded status of FtsH protease complex from *T. maritima* (PDB:2CEA). Dashed circles indicate sites of nucleotide - arginine finger pairs. Distance from the guanidine group of the arginine to the β -phosphate of bound ADP is indicated. Dashed arrow indicates the direction of potential ATPase domain movement. Colour scheme: yellow, pore residues; magenta, AAA-wedge subdomain; red, Walker motifs; green, AAA-helical subdomain; grey, protease

domain. B. A side view with two monomers removed for clarity. C. Two views showing the relative movement of ATPase domain, protease domain is taken as the reference for alignment, the other two statuses of monomer are shown in transparent, movement of the pore phenylalanine is indicated by yellow lines with distance labelled. D. A schematic view of the proposed action model of FtsH and the comparison with current models of ATP hydrolysis in AAA+ complex. Same subunits during the action are coloured in the same colour. Numbers 1, 2 and 3 represent the sequential sites of ATP hydrolysis during action. For non-sequential events, only number 1 is used to indicate the site of ATP hydrolysis. In the speculated action model of FtsH, subunits are also labelled with a, a', b, b', c and c' to correspond with the annotations in Figure A, B and C. The dash-line circles in the Probabilistic model indicate that the illustration of ATP hydrolysis sites is not meant to be comprehensive.

There are currently three models describing the action of FtsH complex induced by ATP hydrolysis: The concerted model, the stochastic firing and the sequential (or rotary) model (314, 315). The concerted model invokes a simultaneous structural change in all subunits. In this model, each step of the ATP hydrolysis, domain movement and ATP/ADP exchange would happen, or not happen, in all six subunits. From the structural point of view, this model favours a rigid C-6 symmetry of FtsH under physiological conditions, which is unlikely for nucleotide-bound FtsH based on the current structures (Figure 4-1, 4-8).

The stochastic model, in contrast, proposes that ATP hydrolysis is initiated independently among the subunits. One example of the stochastic model is the ATPase ClpX (316). When a series of ClpX monomers are deactivated by mutagenesis and fused within a hexamer, as little as one active monomer can initiate the hydrolysis of ATP (316). By covalently linking active and inactive ClpX monomers, ATP hydrolysis is still observed with activities proportional to the arrangement of active subunits (316). Notably, within this model, the conformation of the adjacent inactive monomer is still crucial for the activity of the working monomer. Therefore this model still requires, in addition to the active subunit, at least one additional subunit to provide assistance, most likely by providing an arginine figure.

The sequential model describes the unidirectional ATP hydrolysis events among subunits (317). For instance, the E1-hexameric helicase couples ATP hydrolysis to

DNA translocation in a sequential manner (317). The binding of ATP in one helicase subunit likely triggers the opening of the next subunit to ATP (318). In this way, the action of the hexameric ring relies on the cooperation of at least two adjacent subunits, and given sufficient ATP for binding, the hydrolysis of ATP can occur sequentially in one direction.

Despite the different symmetric architectures of FtsH (256, 257, 260), it is widely accepted that subdomain movements driven by ATP-hydrolysis underpin the action of the FtsH complex. Based on the above mentioned action models of AAA+ complexes, the crystal structure of *T. maritima* FtsH (255, 257) and the functions of well-characterised key residues (256, 257, 259, 260, 306, 307), I propose here a speculative action model of the FtsH complex. As shown in Figure 4-9A, given the precondition that the FtsH hexamer is fully loaded with nucleotides, subunits c/c' show the shortest distance from their bound ADP to the arginine finger of the adjacent subunit a/a', therefore, they have the highest chance to initiate ATP hydrolysis if ATP-bound FtsH is similarly in this conformation. The two subunits with the longest distance between their ADP and adjacent arginine finger, subunit a/a', are most likely in a conformation ready for ADP/ATP exchange as this "open" conformation is most accessible to the cytosol (or soluble phase). Meanwhile, the substrate-binding pore residue phenylalanine (residue 234 in *T. maritima* FtsH) of subunits a/a' is at the most outward position, therefore most approachable to potential substrates (Figure 4-9B). Subunits b/b' are likely one step away from the ATP-hydrolysis, although the distance from their bound nucleotide to the adjacent arginine finger is shorter than that of subunit a/a'. Upon the ATP-hydrolysis in subunit c/c', a mechanical force is generated that pushes the ATPase domain of subunit a/a' in a tangential direction. Consequently, the pore residues of subunit a/a' are driven by the mechanical force moving approximate 45.3 Å

inwards towards the axis of the complex, approaching the protease domain of an adjacent subunit (Figure 4-9C). This movement pulls substrate towards the protease domain and transforms subunit a/a' to the same conformation as subunit b/b'. Meanwhile, subunits b/b' are pushed forward to the same conformation as subunits c/c', where the next round of hydrolysis of ATP could occur. A movement of approximate 25.6 Å of the pore residue occurs in this conformational shift. In this model, the ATP hydrolysis events within a hexamer occur unidirectionally in the interval of one subunit. This speculated FtsH action model is neither simultaneous nor stochastic among six subunits, but rather partially sequential in the way that the motion of one subunit is reproduced successively by every other subunit in the hexamer. It should be mentioned that this model is based on the 6-ADP loaded conformation (250). In the absence of nucleotide-binding, a dramatic change of the conformation of FtsH monomer can be expected (257). However, other nucleotide-bound subunits might still follow this action model. To test this speculated model, a subunit-fusion approach can be applied to limit the number of subunits to be studied (316). And the flexible region (Figure 4-6C, 4-8C) could be important targets for mutagenesis to abolish the conformational change of individual monomers.

Chapter 5. Towards the role of Psb29/THF1

*Excerpts from Chapter 5 are in press for publication (73).

5.1. Introduction

The Psb29 subunit was first identified from analysis of a His-tagged PSII preparation isolated from a CP47-His tagged strain of *Synechocystis* sp. PCC 6803 (195). Substoichiometric levels of Psb29 were detected by N-terminal sequencing following SDS-PAGE separation. It was also shown that salt washing that can remove extrinsic proteins also dissociated Psb29. However, no evidence was given to exclude non-specific binding of Psb29 to the nickel column, so a role in PSII was unclear at that time (195).

In higher plants, the Psb29 homolog, THF1 (thylakoid formation 1), is found in the chloroplast envelope, stroma and thylakoid membrane (319, 320). Knockout of *THF1* leads to deficient thylakoid formation in *Arabidopsis* (320) with membrane vesicles from the chloroplast inner envelope failing to stack and to fuse to mature thylakoid membranes, leading to variegation in young leaves (320). However, as plants develop, this variegation is overcome. The knockout of *THF1* also suppresses the expression of plastidic genes and chlorophyll biosynthesis (321). Besides this, THF1 is also involved in regulating the dynamics of the PSII-LHCII supercomplex (322).

Knockout of *psb29* in *Synechocystis* and knockdown of *THF1* in *Arabidopsis* cause no significant changes to electron transfer within PSII, but PSII activity is light sensitive in both cases, and similar light-dependent increases of uncoupled proximal antenna fluorescence were observed (323). In *Oryza sativa*, mutation of the THF1 homolog, NYC4, severely suppresses the degradation of D1 and D2 subunits of PSII (324), which are usually conducted by FtsH proteases (19, 188). Furthermore, in a *thf1* null mutant of *Arabidopsis*, the Type B FtsH, which is responsible for PSII repair, is significantly

less accumulated (325, 326). The same is observed in *Synechococcus* sp. PCC 7942: deletion of *psb29* leads to decreased accumulation of FtsH (327). However, changes to individual FtsH were not clarified. A recent publication by Nixon and colleagues has provided evidence of a physical interaction between Psb29 and the FtsH2/3 heterocomplex (73), responsible for PSII repair (120). Overall, these results suggest a role of Psb29/THF1 in the FtsH-mediated PSII repair process.

In higher plants, THF1 is also known to be involved in stress responses and plant immunity. Deletion of *THF1* in *Arabidopsis* leads to an increased level of wound-induced oxylipins (328). The transcriptomic profile of cold-stressed *Populus simonii* shows a significant upregulation of THF1, along with PGR5 (involved in cyclic electron transfer (329)) and COP1 (involved in UV-B signalling (330)), suggesting that *THF1* is a cold-responsive gene (331). In wheat, THF1 (designated as ToxABP1) is reported to interact with Ptr ToxA (332), a proteinaceous toxin produced by the plant pathogen *Pyrenophora tritici-repentis*. Interestingly, the THF1 protein which interacts with the Ptr ToxA protein forms, predominantly, a 60-70 kDa complex, suggesting a homodimeric organisation (332). A similar pathological interaction is revealed in Tomato and *Arabidopsis*, in which non-host-specific phytotoxin coronatine might directly interact with THF1 to induce chlorosis (333).

It is important to note that THF1 participates in a G-protein signalling network involved in chloroplast development (319, 325, 334). In *Arabidopsis*, the ectopic expression of an active form of GPA1 (G protein α subunit), a plasma membrane-delimited weak GTPase, complements to some extent the chloroplast defects from the disruption of *THF1*, and partially recovers the decreased accumulation of FtsH (325). However, the ectopic activation of GPA1 does not rescue the defects in an *ftsh2 ftsh8* double-deletion mutant nor an *ftsh2* single-deletion mutant (325). The physical interaction between

GPA1 and THF1 was demonstrated by *in vivo* and *in vitro* experiments (319). Furthermore, D-glucose is transduced to GPA1 via the membrane sensor RGS1 (regulator of G-protein signalling protein 1) (334). It was shown that the level of THF1 is negatively regulated by D-glucose but not L-glucose, and overexpression of THF1 can complement the D-glucose hypersensitivity of root cells in *Arabidopsis* (319).

The crystal structure of *T. elongatus* Psb29 has recently been solved, and evidence for a direct interaction between Psb29 and the PSII-repair FtsH proteases in *Synechocystis* has been provided (73). This chapter examines the growth of a Psb29 knock out mutant of *Synechocystis* sp. PCC 6803 under various conditions, provides a preliminary analysis of suppressor mutants and then investigates the structural conservation of Psb29 and THF1, in the hope of providing information on potential interacting sites on Psb29/THF1.

5.2. Preliminary characterization of a *psb29* null mutant of *Synechocystis*

5.2.1. Accumulation of FtsH is affected in Δ Psb29

The function of Psb29/THF1 is still ill-defined. Recent studies suggest that Psb29/THF1 affects the accumulation of FtsH proteases in both plants and cyanobacteria (325, 327) and that Psb29 in *Synechocystis* physically interacts with the FtsH2/3 protease complex (73). The knockout of Psb29 significantly alters the accumulation of FtsH2/3 but has no drastic effect on FtsH4 (73). Here, the accumulation of FtsH2 and FtsH3 in a Psb29 knockout strain of *Synechocystis* in response to photoinhibitory illumination was investigated.

As shown in Figure 5-1A, knockout of FtsH2 in *Synechocystis* inhibits growth under normal and highlight ($\geq 20 \mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$). The ectopic expression of a C-terminal Flag-

tagged FtsH2 in the Δ FtsH2 background, allows the complementation of the Δ FtsH2 mutant, restoring its ability to grow under the normal light and highlight. In contrast, the Psb29 knockout strain, Δ Psb29, shows tolerance to normal light but is sensitive to high light ($90 \mu\text{E}\cdot\text{m}^{-2} \text{ s}^{-1}$). Complementation of Δ Psb29 with a C-terminal Flag-tag version of Psb29 restored the tolerance of strain CF-Psb29/ Δ Psb29 to highlight, indicating normal functionality of the C-terminally tagged Psb29 protein. However, the N-terminal Flag-tagging of Psb29 showed only minor improvement of light tolerance, suggesting an alteration to the functionality of Psb29. Furthermore, a difference in pigmentation was noticed between WT and CF-Psb29/ Δ Psb29 under high light stress. The latter is apparently more yellow-pigmented, suggesting unfitness probably due to minor interference on the functionality of Psb29. It should be mentioned that expression of the Flag-tagged target protein was driven by the strong *psbA2* promoter commonly used for protein overexpression in cyanobacteria (335, 336).

Intriguingly, unlike Δ FtsH2, the sensitivity of Δ Psb29 to highlight was not lethal. After a prolonged period of growth, colonies did appear in Δ Psb29 spots grown under high light. From these slow growing colonies, a few candidate suppressors were selected and subjected to screening for suppressor mutation as described later (Figure 5-3).

The level of D1 protein and FtsH2/3 in the Δ Psb29 strain was assessed by Western blotting (Figure 5-1B, lower panel). Under normal growth light (0 h), the accumulation of FtsH2 and FtsH3 was significantly reduced in the Δ Psb29 strain in comparison to the WT. When exposed to photoinhibitory light for 2 to 4 h, an increase of FtsH2 and FtsH3 signal was observed in both WT and Δ Psb29, which might account for the limited repair of PSII as shown by the compensation of oxygen evolution rate of Δ Psb29 in a photoinhibition assay (Figure 5-B, upper panel). Low levels increase in expression of FtsH2 and FtsH3 in response to high light were detected even in the presence of the

protein synthesis inhibitor (250 μ M apramycin). This effect is most likely due to ineffectual inhibition of protein synthesis, but the possibility also exists that high light triggers a spatial redistribution of FtsH3, which is found in both the cytoplasmic membrane (221) and thylakoid membrane (220, 221), and such a redistribution has been suggested for *Synechocystis* FtsH3 (220).

In the presence of the protein synthesis inhibitor, apramycin, D1 degradation is significantly affected in Δ Psb29. This is in accord with the decreased level of FtsH2/3, involved in D1 degradation (120). Overall, the photoinhibition assay reveals that effective PSII repair, in particular, the FtsH-mediated D1 removal, is impaired in Δ Psb29 and accumulation of FtsH2/3 is severely depressed. These results support the

idea that Psb29 participates in the PSII repair process by regulating the level of the FtsH2/3 protease complex.

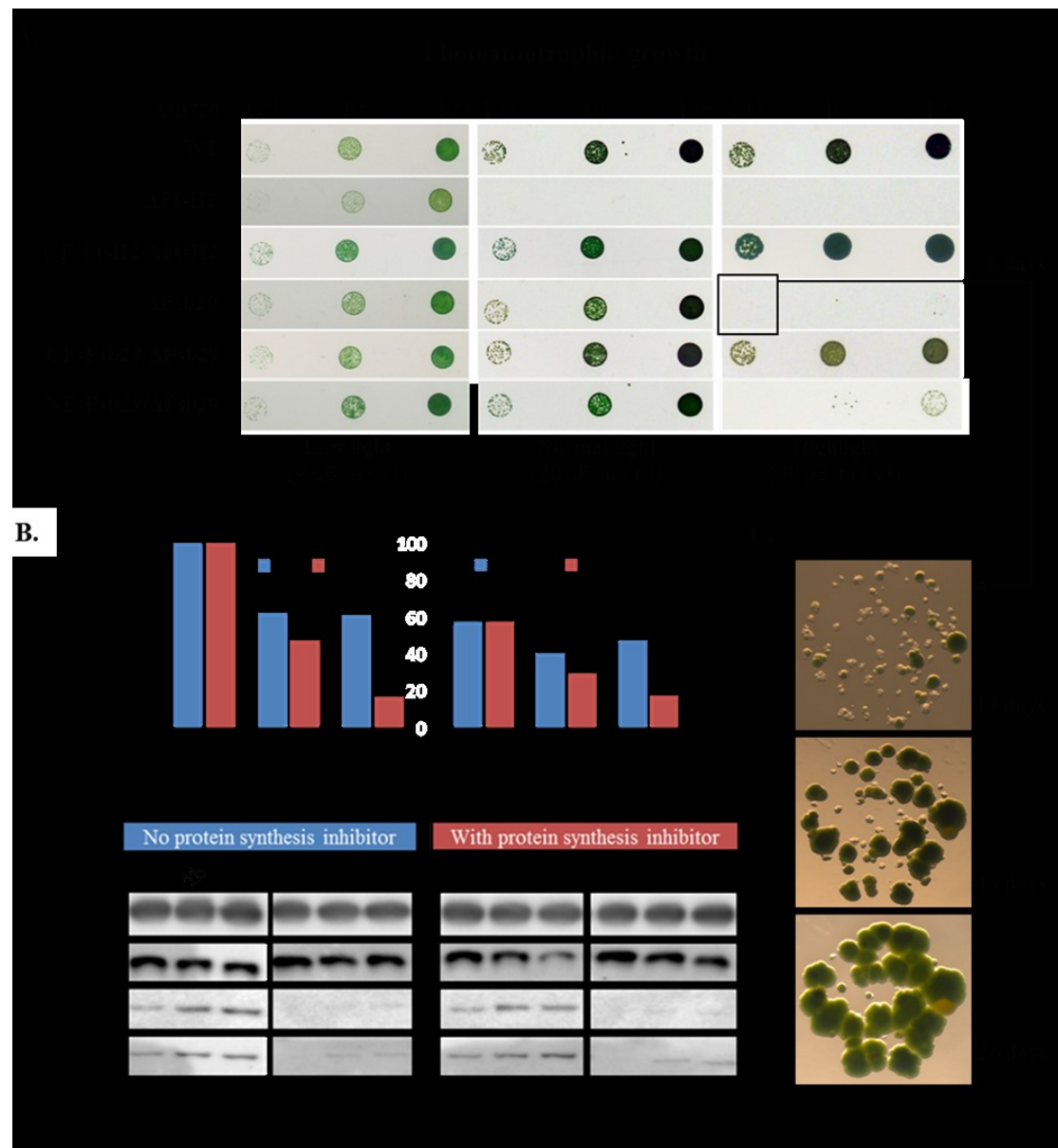


Figure 5-1 Complementation growth assay and photoinhibition assay of the ΔPsb29 strain.

A. Spot growth assay, 8-day growth. B. Photoinhibition assay and Western blotting detection of D1, FtsH2 and FtsH3 on samples at different stages of photoinhibition. “+” indicates the presence of 250 μM apramycin. Oxygen evolution data was from two to three technical replicates in one experiment, no biological replicate. C. A close-up observation of the prolonged spot growth of ΔPsb29 under highlight. The spot from Figure A (in rectangle) was observed for 13 to 26 days. Images were taken by Leica LAS AF microscopy system.

5.2.2. Mixotrophic defect/D-glucose sensitivity of Δ Psb29

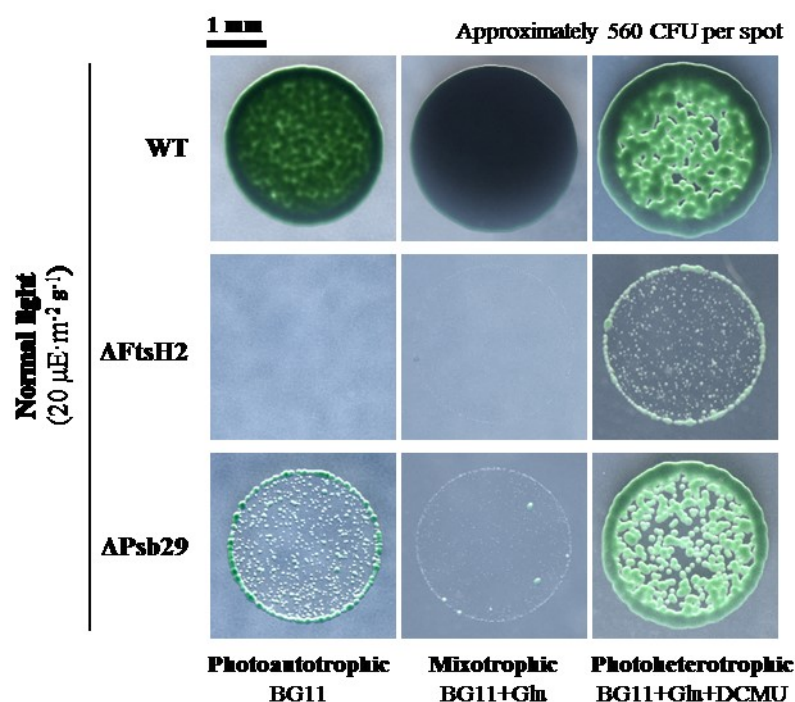


Figure 5-2 Close-up views of Δ Psb29, Δ FtsH2 and WT growth under different metabolic backgrounds.

2.5 μ L of OD₇₃₀ = 0.01 liquid culture was inoculated. Approximately 560 CFU counted in WT control. 10-day growth.

A noticeable feature of strain Δ Psb29 was the defective growth under mixotrophic conditions. As shown in Figure 5-2, the addition of 5 mM D-glucose significantly promotes the growth of WT *Synechocystis* but suppresses the growth of Δ Psb29 under normal light intensity, and completely inhibits it under high light (Figure 5-3). Under photoheterotrophic growth conditions, where photosynthetic electron transfer is blocked by DCMU, the “toxicity” of D-glucose was alleviated, suggesting that photosynthesis is one source of inhibition. A similar phenotype was also observed in strain Δ FtsH2. Δ FtsH2 can grow under the photoheterotrophic but not mixotrophic condition, suggesting that the source of “toxicity” is from photosynthesis under mixotrophic conditions. One reasonable explanation for the source of inhibition is the accumulation of photodamaged PSII which results in the formation of reactive oxygen

species (ROS). As previously described, the accumulation of FtsH2 is defective in Δ Psb29. Therefore similar physiological stress might underlie the observed phenotypes in Δ FtsH2 and Δ Psb29. However, an FtsH-independent route cannot be excluded at the current stage.

5.3. Analyses of suppressor mutations in Δ Psb29 *Synechocystis*

5.3.1. Phenotypes of Δ Psb29 suppressor mutants

Two independent suppressor strains, Δ Psb29-SP-b and Δ Psb29-SP-c, were segregated from Δ Psb29 subjected to light stress under photoautotrophic conditions. The *psb29* gene knockout was confirmed by PCR (Figure 5-3B) and antibiotic resistance (data not shown). Contrary to Δ Psb29, both suppressors grew mixotrophically even under high light. A noticeable change in pigmentation was observed in both suppressors strains under high light: the suppressor colonies appeared yellowish. The genomic DNA of the two suppressors was extracted and sequenced.

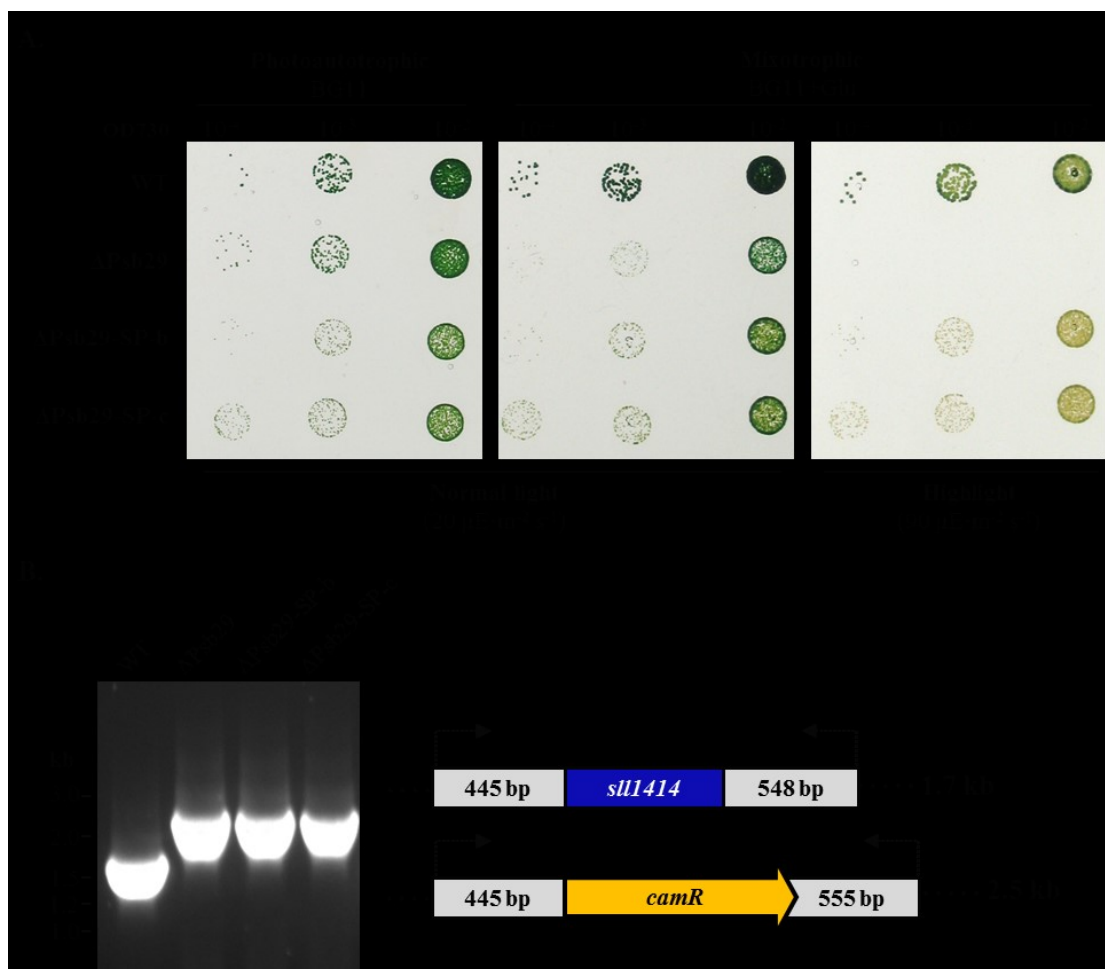


Figure 5-3 Growth assay of Δ Psb29 and Δ Psb29 suppressor mutants.

A. Δ Psb29-SP-b, Δ Psb29-SP-c are the two suppressors. 8-day growth was shown. B. PCR examination of candidate suppressors. Primers for *psb29* amplification amplify a fragment including the chloramphenicol resistance cassette in Δ Psb29.

5.3.2. Genotype of Δ Psb29 suppressor mutants

Genome resequencing was performed as described in Section 3.4.1.4. The “Kazusa” *Synechocystis* genome was used as the reference genome, and the genome sequencing data of Δ Psb29 (later designated as Δ Psb29-OCP; it is not the same strain as the above described Δ Psb29 due to an unintentional mutation), Δ Psb29-SP-b and Δ Psb29-SP-c were mapped to this reference genome for variant calling. The integrity of the *psb29* knockout mutants was firstly confirmed in the three genomes. After a quality filtration, six polymorphism sites (Table 5-1) were detected in Δ Psb29-SP-b, of which four polymorphism sites were also found in Δ Psb29, and therefore unlikely to be responsible

for the observed phenotype. A mutation was found in gene *sll1973* leading to a G429A mutation. Importantly, a truncation of gene *slr1638* was discovered. The analysis in Δ Psb29-SP-c showed a similar pattern of polymorphism, except the single nucleotide extension in the transposase-encoding gene *sll0172*, and the mutation in *sll1973* was absent. These two differences suggest that the suppressors are independent rather than clones. Taken together, the genome sequencing and quality filtration targeted the truncation in *slr1638* as the main candidate responsible for the suppressor mutation.

Table 5-1 List of SNPs detected in Δ Psb29 suppressors

No.	Start	End	Size	Change	Codon Change	AA Change	Result	CDS	Product / Notes
<i>SNPs found in ΔPsb29-SP-b (common SNPs shared with WT-P were ignored)</i>									
1	100086	100085	0	(A)6 -> (A)7			Frame Shift	me	Malic enzyme / Found in <i>APsb29</i>
2	1379640	1379640	1	G -> A	GCC -> GCT		None	sll1164	Hypothetical protein / Found in <i>APsb29</i>
3	1576164	1576164	1	G -> A	ACC -> ACT		None	sll1973	Glycerol-3-phosphate acyltransferase / Unique
4	1604181	1604181	1	(G)7 -> (G)6					Found in <i>APsb29</i>
5	1604605	1604607	3	CGC -> AAT					Found in <i>APsb29</i>
6	205059	2050529	1	C -> T		Q12*	Truncation	slr1638	Unknown function / Unique, phototroph
<i>SNPs found in ΔPsb29-SP-c (common SNPs shared with WT-P were ignored)</i>									
1	100086	100085	0	(A)6 -> (A)7			Frame Shift	me	Malic enzyme / Found in <i>APsb29</i>
2	1379640	1379640	1	G -> A	GCC -> GCT		None	sll1164	Hypothetical protein / Found in <i>APsb29</i>
3	1604181	1604181	1	(G)7 -> (G)6					Found in <i>APsb29</i>
4	1604605	1604607	3	CGC -> AAT					Found in <i>APsb29</i>
5	2050529	2050529	1	C -> T		Q12*	Truncation	slr1638	Unknown function / Unique, phototroph
6	2997812	2997812	1	T -> C			Extension	ssl0172	Transposase / Found in <i>APsb29</i>

Characterisation of the product of gene *slr1638*, with a predicted molecular mass of 12.4 kDa and an isoelectric point of 4.45, has not been reported in the literature to date. Preliminary information mining from this work suggests that this protein is likely to be of importance in oxygenic photosynthesis. In the Pfam database, the protein can be found under the entry DUF760 (PF05542) (DUF means domain of unknown function), which describes a family of proteins that are universally found in oxygenic photosynthetic organisms, and some poorly-defined parasite groups such as Alveolata (Figure 5-4A). In the UniProt reference proteome database which was interrogated on 11th February 2017, 107 out of 110 cyanobacteria had at least one DUF760 family protein. In total 205 proteins were retrieved. The sequence logo profile of these 205 sequences shows that the DUF760 protein is highly conserved among the 107 analysed cyanobacteria (Figure 5-4B). In *Arabidopsis*, the DUF760 domain is found in the chloroplastic UV-B-induced protein At3g17800, encoded by the *MEB5.2* gene (337, 338). The alignment of the DUF760 domains of Slr1638 and At3g17800 shows a similarity of 44.1%. It is, therefore, reasonable to deduce that the DUF760 family protein plays an important role in oxygenic photosynthetic organisms. Plasmids for complementation and knockout experiments of *slr1638* have been constructed by a group member for further analyses.

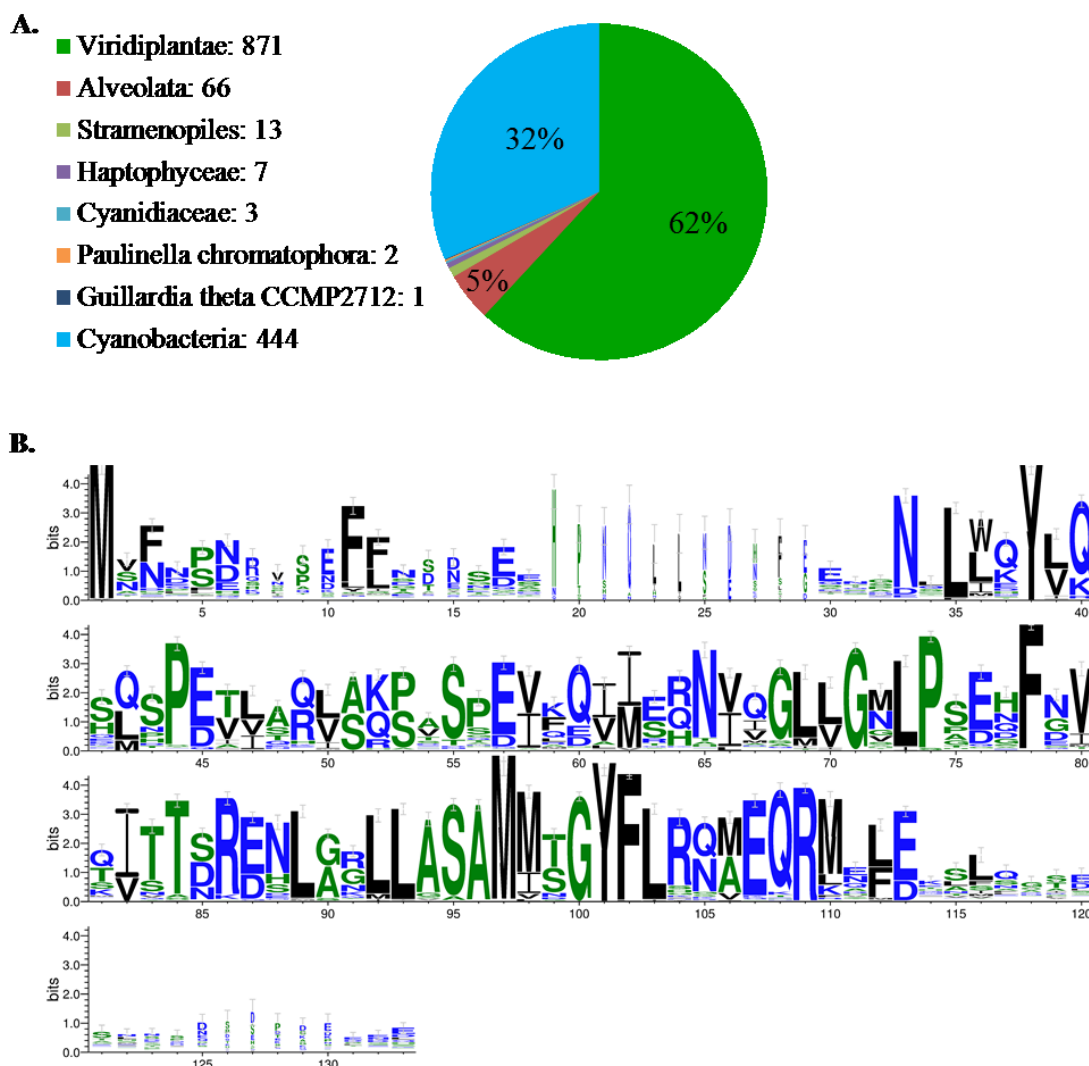


Figure 5-4 Data mining of the DUF760 family and sequence conservation of the cyanobacterial DUF760 proteins.

A. Distribution of the DUF760 family proteins. Data were retrieved from UniProt Reference Proteomes database on 1st Aug 2017. B. Sequence logo of 205 cyanobacterial DUF760 proteins. Y axis is the information content in bits of the composition of amino acids; the size of a character, single-letter code of amino acid, represents how conserved that amino acid is at that position, the larger the letter, the more conserved it is. Colour scheme: green - neutral, black - hydrophobic, blue - hydrophilic.

5.3.3. Δ Psb29-OCP: An OCP mutation in the sequenced Δ Psb29 strain

Orange carotenoid protein (OCP) plays an important role in quenching excess light energy in cyanobacterial non-photochemical quenching (339). OCP is a water-soluble protein found in most cyanobacteria except the α -clade which lack the light-harvesting phycobilisome (339). Photosynthetically inert blue-green light induces conformational

changes of the OCP from its inactive form OCP^o to the active form OCP^r (340). The latter then interacts with phycobilisome to dissipate excess light energy as heat (341). Notably, both the inactive and active forms of OCP can efficiently quench singlet oxygen (342).

Surprisingly, a mutation of the OCP (Orange carotenoid protein, encoded by *slr1963*) was found in the Δ Psb29 strain during the genome sequencing which was subsequently confirmed by Sanger sequencing. Four codons encoding residues LQPP (223 - 226) of OCP were deleted in the Δ Psb29 strain subjected to genomic DNA extraction and sequencing. It is important to mention that this deletion mutation was solely observed in the batch of Δ Psb29 strain subjected to genome sequencing, and was not present in the original strain characterised in this Chapter, as determined by Sanger sequencing of the PCR fragment amplified from the original cryostock (data not shown). Moreover, as shown in Figure 5-1A, introducing Flag-tagged Psb29 complements the deletion of Psb29.

The deletion mutation detected in OCP affects a region located in the interior of the C-terminal domain (CTD) of OCP which is in proximity to the carotenoid molecule, according to the *Synechocystis* OCP crystal structure (343). Notably, it has been reported that the OCP is constitutively active in the absence of the CTD (344). It is conceivable that the 4-residue deletion in CTD could cause a similar effect and was selected during serial restreaking of the Psb29 null mutant. Nevertheless, further clarification is needed to confirm the phenotypic effect of this LQPP deletion.

5.4. Conservation of the Psb29/THF1 family and its implication

5.4.1. Psb29/THF1 is closely related to oxygenic photosynthesis

Psb29/THF1 is universally found in oxygenic photosynthetic organisms. In the UniProt Reference Proteome database interrogated on 28th January 2017, a single copy of Psb29 was found in 103 out of 106 cyanobacterial proteomes. The three cyanobacteria without Psb29 are *Limnoraphis robusta* CS-951, *Leptolyngbya valderiana* BDU 20041 and *Cyanobium* sp. PCC 7001 (*Synechococcus* sp. PCC 7001). However, this observation is not meant to be conclusive, because the possibility exists, although low, that the lack of Psb29 in these three species is due to incomplete sequencing of the genome from which the proteomes were derived. 57 out of 58 plant proteomes show single or multiple copies of THF1. *Picea glauca*, whose genome is still only partially assembled, shows no THF1 homolog.

As shown in Figure 5-5, a total of 211 sequence homologues to Psb29 from *Synechocystis* (product of gene *sll1414*) were retrieved from the UniProtKB database (<http://www.uniprot.org>). 103 sequences are from cyanobacteria, 12 from red algae, 11 from green algae, 84 from plants and one from a virus that infects the green alga *Chlorella* sp. strain NC64A. The unrooted phylogenetic tree (Figure 5-5A) shows the evolutionary relationship between all examined Psb29/THF1 homologues. It is intriguing that the eukaryotic THF1 homologues are rooted in-between the *Gloeobacter* and *Prochlorococcus* branches. It is known that in a classic cyanobacterial phylogeny, *Gloeobacter* is the earliest diverged branch while *Prochlorococcus* is a later branch (284-286). Therefore, the root of the eukaryotic branch of Psb29/THF1 supports a cyanobacterial origin of the eukaryotic THF1, and the acquisition is likely to have occurred later than the diverging of *Gloeobacter* from the ancestor of the cyanobacteria. Moreover, Psb29/THF1 homologues in cyanobacteria and red algae show more

divergence than the homologues in green algae and plants. The THF1 homolog found in the algal virus was likely obtained from its host green alga at an early stage of the green algae speciation, as suggested by the early diverging branch.

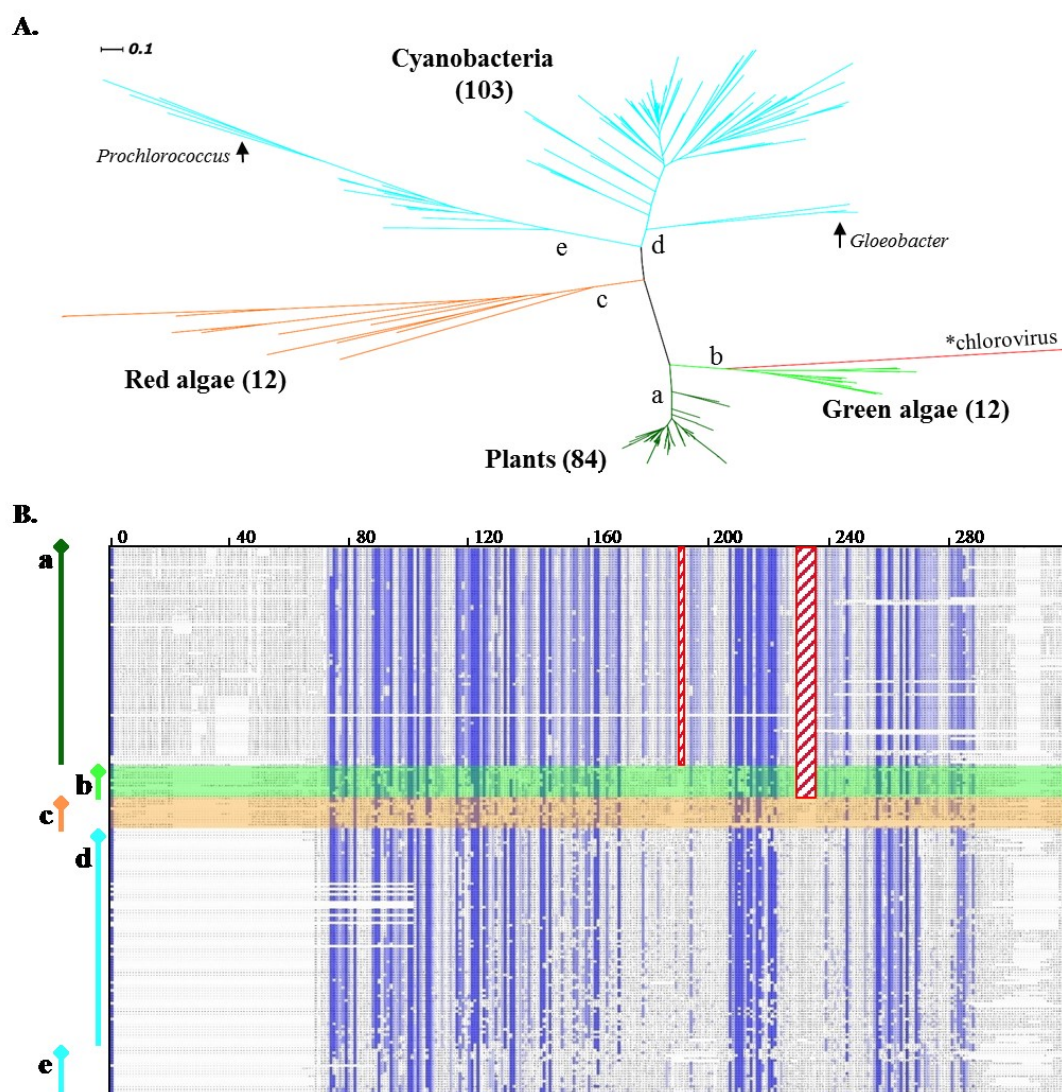


Figure 5-5 Phylogeny and sequence similarity of the Psb29/THF1 family.

A. Unrooted phylogenetic tree of 211 Psb29/THF1 sequences based on maximum-likelihood. Scale bar represents substitution rate. Colour scheme: cyan – cyanobacteria, branch “d” “e”; orange – red algae, branch “c”; green – green algae, branch “b”; red – algal virus, branch “b”; dark green – plant, branch “a”. The number of sequences of each group is indicated. The algal virus THF1 was considered as a green algal THF1. B. Overview of the sequence alignment of 211 Psb29/THF1 sequences. Sequences of green algae and red algae sequences are coloured in green and orange respectively. Blue shading represents sequence identity, with dark to light representing identity from high to low.

As shown in Figure 5-5B, the cyanobacterial Psb29 and the plant THF1 share a high degree of sequence similarity. Psb29 from *T. elongatus* shows a mean sequence

similarity of 59.2% with the other 102 cyanobacterial Psb29 sequences (subset “d” and “e”) and 53.7% with the 84 plant THF1 sequences (subset “a”). Distinguishable insertion/deletion events are observed in two relatively divergent regions of Psb29, corresponding to *T. elongatus* residues 121-122 and 151-154. Besides, both Psb29 and THF1 are less conserved at the C-terminal end from residue 204 onwards. 9 cyanobacterial sequences (A0ZEW1, A0A0D6KUW6, A0A0D6YP33, M1X5Z4, A0A139X4N1, A0A0C1N8T8, D4TV11, B4VS31, A0A0T7BTW0) are found to be lacking an N-terminal segment. However, by re-examining the original genome data, they are likely to be annotation errors due to the overlooking of the uncommon GTG start codon.

5.4.2. Structural conservation of the Psb29/THF1 family

Based on the sequence alignment of the 211 identified Psb29/THF1 sequences and the recently solved structure of Psb29 from *T. elongatus* (73) (Figure 5-6C), a few conserved structural features were revealed. As shown in Figure 5-7A, the residues F14 at helix a1, V35 and L39 at helix a2, G55 at helix a3, G138 at helix a7 and R133 at the beginning of helix a7 (*T. elongatus* numbering) are strictly conserved in all cyanobacterial and plant Psb29/THF1 sequences. F14, G55 and G138 are buried inside the protein and thus might play a role in helix packing. In contrast, V35, L39 and R133 are on the surface, and therefore could be potential sites for interaction with other proteins. A ConSurf analysis was performed on all Psb29/THF1 to assess the

conservation of the tertiary structure. A high degree of conservation is revealed on one side of the molecule (Figure 5-6D).

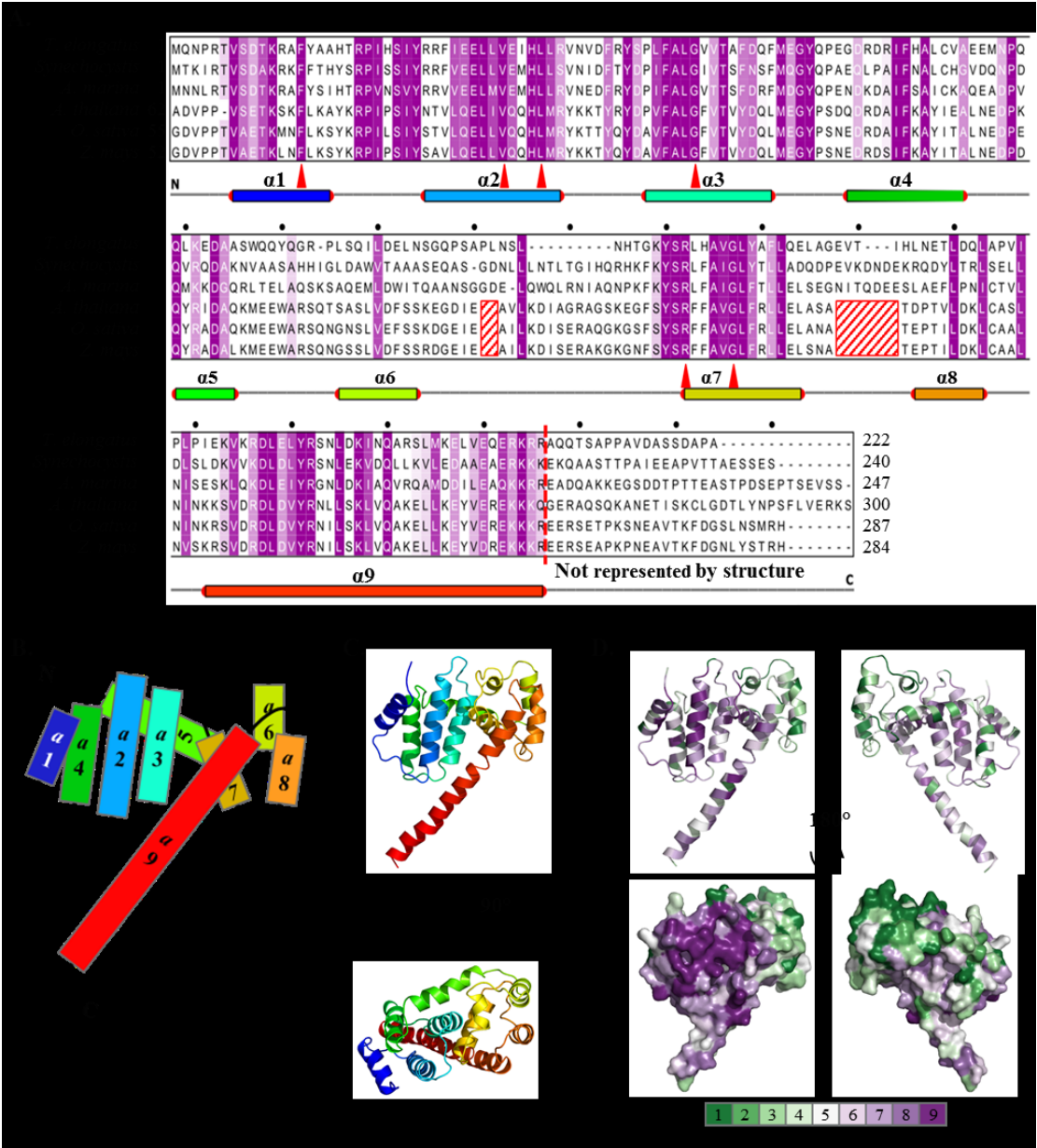


Figure 5-6 Conservation analyses of Psb29/THF1.

A, B: Primary and secondary structures of Psb29/THF1 based on the structure of *T. elongatus* Psb29, PDB: 5MLF. Purple columns from light to dark represent the sequence similarity from low to high., based on the BLOUSM62 score of the six selected sequences; predicted transit sequences of eukaryotic THF1 are omitted, and the numbering is sequence-specific. Six strictly conserved residues, based on the alignment of 211 sequences, are indicated by red triangles. Insertion/deletion events are indicated by red slash columns. C. The crystal structure of *T. elongatus* Psb29, PDB 5MLF. D. The ConSurf analysis of 211 Psb29/THF1 sequences. Colour scheme from green to purple represents conservation of amino acids from low to high.

The residues Y131, S132 and R133 form a structural “hub”, coordinating the highly conserved parts of helices α_2 , α_7 and α_9 (Figure 5-7B). This conserved “YSR” motif suggests the presence of similar structures among Psb29/THF1 homologues. Among the most conserved residues (Grade 9 in the ConSurf score (345), representing a similarity of approximately over 80%), 19 residues are spatially accessible to potential interacting partners. E31, R133 and D175 are within H-bonding distance to K11, E36 and Y131, respectively. These residues are very likely essential for stabilising the protein conformation. V7, V35 and L39 form an exposed hydrophobic patch, providing potential hydrophobic interacting sites. S132, E32, K185 and Q188 form a well-conserved shallow groove, providing a potential anchoring site for interacting partners. Another noticeable feature is a strictly conserved “G2G” contact buried inside the protein (Figure 5-7B). The residue G55 of helix α_3 is at the shortest distance of 3.4 Å to G138 of helix α_7 . Whether the proximity of the two glycines is of functional significance is unknown.

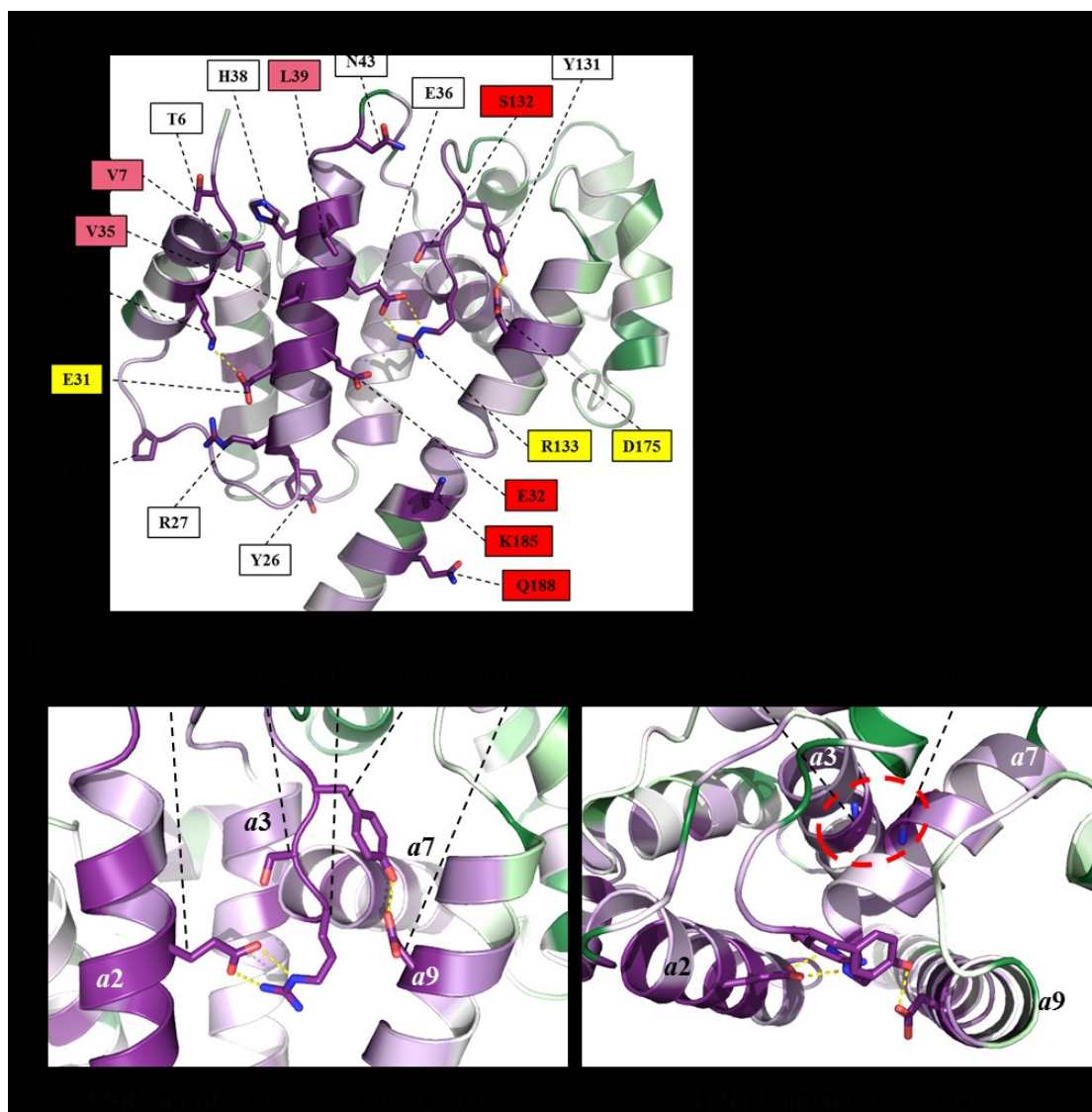


Figure 5-7 Close-up views of the conserved face of Psb29/THF1 identified by the ConSurf analysis.

19 of the most conserved residues that are not buried within the Psb29 structure are labelled and shown in stick form with red indicating oxygen atoms and blue nitrogen atoms. Polar intra-protein side-chain contacts are shown as yellow dashed lines. Red labels indicate a shallow groove formed by charged residues; yellow labels indicate residues possibly involved in both stabilising the structure and interacting with proteins; pink labels indicate potential hydrophobic contact sites.

5.5. Discussion

5.5.1. Psb29 likely provides site of diverse interaction

Psb29/THF1 is a highly conserved protein universally found in oxygenic phototrophs. Benefited from a recently solved crystal structure of Psb29 from *T. elongatus* (73), a structural conservation analysis was performed which points to a few structural features of potential importance to the functionality of Psb29/THF1. For example, previous

work revealed that the C-terminal region from amino acids 237 to 295 of THF1 from *N. benthamiana* is essential for the interaction of the protein with the coiled-coil domain of the I-2 like proteins which are involved in plant cell death (346). This region was previously predicted to be a coiled-coil domain (320), but the crystal structure showed that it formed a long helix (73). It is important to mention that the interaction between the THF1 C-terminus and the I-2 like protein was only verified by yeast-2-hybrid (346), and therefore does not rule out the possibility that other regions of THF1 can participate in protein-protein interactions. The hydrophobic residues V7, V35 and L39 on the highly conserved surface of the protein could provide potential interacting sites to hydrophobic amino acids or lipids. It is important to mention that mobile localisation of THF1 in the chloroplast (319, 320) might be a consequence of transport, but it is also possible due to different interaction through the above mentioned different conserved sites.

5.5.2. Two conserved motifs of structural significance

Some well-conserved structural features need further investigation. The conserved “YSR” motif (Figure 5-7B) might be crucial for all Psb29/THF1 because it not only mediates the contacts between helices α_2 , α_7 and α_9 to stabilise the general conformation of the protein, but also provides potential interacting sites at S132 and R133. Potential interactions with this “YSR” motif could induce conformational changes of Psb29/THF1. The strictly conserved “G2G” contact between helices α_3 and α_7 is unusual in the sense that glycines in proteins tend to introduce high flexibility and deform helices (347-349). Considering the two helices are comparatively conserved among species (Figure 5-7), it is unlikely that the “G2G” motif has an evolutionary advantage if the two helices are meant to be rigid. One possible explanation to the

strictly conserved “G2G” contact is that it may serve as a conformational hinge and its flexibility is of functional importance, therefore providing evolutionary advantages.

5.5.3. Psb29 and the accumulation of FtsH2

The N-terminal Flag-tagged Psb29 was found to be less capable of complementing Δ Psb29 (Figure 5-1A), suggesting that the region near the N-terminus is important for the functionality of Psb29, most likely by stabilising the FtsH heterocomplex. This finding is in line with the structural conservation analysis which shows a conserved surface near the N-terminus.

A role for Psb29 in regulating the accumulation of FtsH2 and FtsH3 was shown in this work. The physical interaction between Psb29 and the FtsH2/3 heterocomplex was separately reported (73). In the *Arabidopsis* chloroplast, knockout of *THF1* places a more drastic downregulating effect on AtFtsH2/8 than on AtFtsH5 (326). The similar preferential effect is noticed in *Synechocystis* (73), with FtsH2 being more affected than FtsH3 in two Δ Psb29 strains. Notably, the accumulation of FtsH4 was not significantly affected (73). It is known that the FtsH heterocomplex, which is responsible for PSII repair is composed of both Type A (AtFtsH1/5 in *Arabidopsis*, FtsH3 in *Synechocystis*) and Type B (AtFtsH2/8 in *Arabidopsis*, FtsH1/2 in *Synechocystis*) FtsH subunits. The absence of one type of FtsH can lead therefore to the instability of the other uncomplexed subunits (120, 219, 226, 227). However, it is not yet clear whether Psb29/THF1 interacts with both two types or just a single type of FtsH.

5.5.4. Psb29 and D-glucose sensitivity

A role for THF1 in mediating a D-glucose activated G-protein signalling pathway has been proposed, based on evidence from *Arabidopsis* (319, 325, 334). It has been shown that root growth of a *THF1* null mutant is depressed in the presence of exogenous D-

glucose (319). Intriguingly, similar hypersensitivity to D-glucose was found in Δ FtsH2 and Δ Psb29 *Synechocystis* strains (Figure 5-2). Moreover, by inhibiting photosynthesis with DCMU, this hypersensitivity was alleviated in both cases. The delayed growth of the Δ Psb29 *Synechocystis* cells (Figure 5-1A, C) resembles, to some extent, the defect in chloroplast development observed in *thf1* mutants (320, 325). These preliminary results raise the question whether a similar signalling path underlies the D-glucose sensitivity in cyanobacteria as in higher plants. Besides, a truncation of gene *slr1638* was detected in two Δ Psb29 suppressor strains that could survive highlight in the presence of D-glucose. Although the function of this gene is unknown, and more work is needed to confirm that it is responsible for suppression of the Δ Psb29 growth defect, its homologues are found to be universally present in oxygenic phototrophs. Slr1638 homologues in cyanobacteria show a high degree of sequence conservation (Figure 5-4). The UV-B upregulated *MEB5.2* gene in *Arabidopsis* also encodes a homologous domain of Slr1638. These findings suggest a conserved role for the DUF760 protein family in oxygenic photosynthesis, which is in need of further clarification.

Chapter 6. Conclusion and future works

PSII drives one of the most challenging yet fundamental biochemical reactions on the planet Earth – water splitting. To survive this harsh photochemical reaction, oxygenic phototrophs adopt a simple strategy - that of rapid D1 turnover involving “sacrificial” removal of damaged, or simply accessible D1 and replacement by newly synthesised functional D1. The current FtsH-mediated PSII repair model depicts an elegant repair mechanism by which damaged D1 is selectively and promptly replaced. In the works described in this PhD dissertation, I studied the putative impact of the adoption of this mechanism on the evolution of PSII. Moreover, I considered the phylogeny and evolution of FtsH proteases and investigated the relationship between these and the evolution of PSII. Finally, I analysed the structural conservation of Psb29, an interacting partner of the PSII-specific FtsH protease complex (73) in cyanobacteria and possibly other photosynthetic eukaryotes. Based on the above analysis, a conserved surface on this protein, as well as a number of structurally-significant sites, were proposed as targets for future functional studies of Psb29/THF1.

The FtsH-mediated repair mechanism can serve as a proxy of the early evolution of oxygenic photosynthesis. The investigation of the engineered CP43-D1 fusion PSII is pioneering in that it provides unprecedented evidence of the relatedness between Type-I and Type-II reaction centres. Construction of a CP43-D1/CP47-D2 double fusion PSII is currently in progress by other members of the Nixon group, with the aim of assembling a PSI-like water-splitting reaction centre. However, complexity exists in the mutagenesis design of such a double fusion, due to overlapping genes that encode CP43 and CP47. Nevertheless, the construction of the CP43-D1 fusion PSII suggests that such an oxygenic PSI-like PSII is highly feasible. The speculated PSI-like oxygenic reaction centre, if successfully constructed, would provide key information on the

evolution of oxygenic photosynthesis, and show that extant PSII is not necessarily the only design for oxygenic photosynthetic reaction centre. Moreover, recent studies have proposed that D1 and D2 possess not only similar structural features crucial for water oxidation (189, 277) but also show similar susceptibility to the proteolysis mediated by PSII-repair FtsH (188). Taken together, these results strongly imply a reaction centre that is prone to photodamage might have existed in a homodimeric PSII ancestor (157); and in such a primordial PSII, FtsH proteases were involved in the selective repair of reaction centre subunit ancestral to D1 and D2. Direct phylogenetic analysis on photosynthetic reaction centres is highly challenging due to the low sequence identity between different types of reaction centres (158, 160). However, it has been demonstrated that phylogenetic investigation of genes closely related to chlorophyll synthesis is a more feasible approach (171, 184) and that the evolution of these genes can serve as a proxy of the evolution of photosynthesis. The evolution of FtsH proteases involved in PSII repair, described in this work, also provides an important proxy for studying the evolution of oxygenic photosynthesis.

One direction requiring future investigation is the mechanism of CP43 detachment from the RC. It is speculated that lipids may serve as the candidate destabiliser that leads to the detachment of CP43. Twenty-five lipid molecules are known to be asymmetrically located in PSII (36), with more lipids located between D1 and CP43 than those located near D2 and CP47 within a monomer (36, 350, 351). It has been suggested that these lipids may be involved in the association of CP43 to PSII (350). Moreover, lipid peroxidation may cause oxidative damage to PSII subunits (352), which could trigger destabilisation of PSII. The fusion of CP43 and D1 has provided the opportunity to enrich the photodamaged CP43-D1 PSII for analysing the lipid composition and

modifications, as well as other direct consequences of photodamage, such as modifications to D1 and D2 (82).

As previously discussed, Psb29/THF1 is of crucial importance due to its putative central role in a signalling network yet to be fully elucidated. The interaction between FtsH and Psb29 (and likely THF1, although no direct evidence of physical interaction has been reported) is likely crucial for fulfilling the role of Psb29/THF1. The conserved sites on Psb29/THF1, identified in this work, can be examined through further mutagenesis studies, which is important for revealing the functions and protein-protein interactions of Psb29/THF1.

References

1. Buick R (2008) When did oxygenic photosynthesis evolve? *Philos Trans R Soc Lond B Biol Sci* 363 (1504):2731-2743.
2. Rasmussen B, Fletcher IR, Brocks JJ, & Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455 (7216):1101-1104.
3. Kopp RE, Kirschvink JL, Hilburn IA, & Nash CZ (2005) The paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* 102 (32):11131-11136.
4. Pingali PL (2012) Green revolution: impacts, limits, and the path ahead. *Proc Natl Acad Sci U S A* 109 (31):12302-12308.
5. Smil V (2016) *Energy Transitions: Global and National Perspectives* (Santa Barbara, California :Praeger, an imprint of ABC-CLIO, LLC,[2017]).
6. Anonymous (2015) *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables*, (Division UNDoEaSAP).
7. Godfray HC, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, & Toulmin C (2010) Food security: the challenge of feeding 9 billion people. *Science* 327 (5967):812-818.
8. Hoffert MI (2010) Climate change. Farewell to fossil fuels? *Science* 329 (5997):1292-1294.
9. Davis SJ, Caldeira K, & Matthews HD (2010) Future CO₂ emissions and climate change from existing energy infrastructure. *Science* 329 (5997):1330-1333.
10. Yan Y, Peng L, Li R, Li Y, Li L, & Bai H (2017) Concentration, ozone formation potential and source analysis of volatile organic compounds (VOCs) in a thermal power station centralized area: A study in Shuozhou, China. *Environ. Pollut.* 223:295-304.
11. Xu W, Wang F, Li J, Tian L, Jiang X, Yang J, & Chen B (2017) Historical variation in black carbon deposition and sources to Northern China sediments. *Chemosphere* 172:242-248.
12. Pirrone N, Cinnirella S, Feng X, Finkelman RB, Friedli HR, Leaner J, Mason R, Mukherjee AB, Stracher GB, Streets DG, & Telmer K (2010) Global mercury emissions to the atmosphere from anthropogenic and natural sources. *Atmospheric Chemistry and Physics* 10 (13):5951-5964.
13. Barber J & Tran PD (2013) From natural to artificial photosynthesis. *Journal of the Royal Society Interface* 10:20120984.
14. International Energy Agency (2014) Key World Energy Statistics 2014.
15. Murchie EH & Niyogi KK (2011) Manipulation of photoprotection to improve plant photosynthesis. *Plant Physiol.* 155 (1):86-92.
16. Kromdijk J, Glowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, & Long SP (2016) Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* 354 (6314):857-861.
17. Nixon PJ, Barker M, Boehm M, de Vries R, & Komenda J (2005) FtsH-mediated repair of the photosystem II complex in response to light stress. *J. Exp. Bot.* 56 (411):357-363.
18. Komenda J, Knoppova J, Krynicka V, Nixon PJ, & Tichy M (2010) Role of FtsH2 in the repair of Photosystem II in mutants of the cyanobacterium *Synechocystis* PCC 6803 with impaired assembly or stability of the CaMn₄ cluster. *Biochim. Biophys. Acta* 1797 (5):566-575.
19. Nixon PJ, Michoux F, Yu JF, Boehm M, & Komenda J (2010) Recent advances in understanding the assembly and repair of photosystem II. *Ann. Bot.* 106 (1):1-16.

20. Bailey S, Thompson E, Nixon PJ, Horton P, Mullineaux CW, Robinson C, & Mann NH (2002) A critical role for the Var2 FtsH homologue of *Arabidopsis thaliana* in the photosystem II repair cycle *in vivo*. *J. Biol. Chem.* 277 (3):2006-2011.
21. Silva P, Thompson E, Bailey S, Kruse O, Mullineaux CW, Robinson C, Mann NH, & Nixon PJ (2003) FtsH is involved in the early stages of repair of photosystem II in *Synechocystis* sp PCC 6803. *Plant Cell* 15 (9):2152-2164.
22. Knox RS (1996) Electronic excitation transfer in the photosynthetic unit: Reflections on work of William Arnold. *Photosynth Res* 48 (1-2):35-39.
23. Ruben S, Randall M, Kamen M, & Hyde JL (1941) Heavy oxygen (O¹⁸) as a tracer in the study of photosynthesis. *J. Am. Chem. Soc.* 63 (3):877-879.
24. Joliot P (1965) Reaction kinetics of coupled photosynthetic oxygen evolution. *Biochim. Biophys. Acta* 102 (1):116-134.
25. Kok B, Forbush B, & McGloin M (1970) Cooperation of charges in photosynthetic O₂ evolution-I. A linear four step mechanism. *Photochem. Photobiol.* 11 (6):457-475.
26. Hill R & Bendall FAY (1960) Function of the two cytochrome components in chloroplasts: A working hypothesis. *Nature* 186 (4719):136-137.
27. Walker DA (2002) The Z-scheme – down hill all the way. *Trends Plant Sci.* 7 (4):183-185.
28. Jagendorf AT & Uribe E (1966) ATP formation caused by acid-base transition of spinach chloroplasts. *Proc Natl Acad Sci U S A* 55 (1):170-177.
29. Durrant JR, Klug DR, Kwa SL, van Grondelle R, Porter G, & Dekker JP (1995) A multimer model for P680, the primary electron donor of photosystem II. *Proc Natl Acad Sci U S A* 92 (11):4798-4802.
30. Debus RJ, Barry BA, Sithole I, Babcock GT, & McIntosh L (1988) Directed mutagenesis indicates that the donor to P680⁺ in photosystem II is tyrosine-161 of the D1 polypeptide. *Biochemistry* 27 (26):9071-9074.
31. Metz JG, Nixon PJ, Rogner M, Brudvig GW, & Diner BA (1989) Directed Alteration of the D1 Polypeptide of Photosystem-II - Evidence That Tyrosine-161 Is the Redox Component, Z, Connecting the Oxygen-Evolving Complex to the Primary Electron-Donor, P680. *Biochemistry* 28 (17):6960-6969.
32. Joliot P, Barbieri G, & Chabaud R (1969) Un Nouveau Modele Des Centres Photochimiques Du Systeme II. *Photochem. Photobiol.* 10 (5):309-329.
33. Forbush B, Kok B, & McGloin MP (1971) Cooperation of charges in photosynthetic O₂ evolution-II. Damping of flash yield oscillation, deactivation. *Photochem. Photobiol.* 14 (3):307-321.
34. Shin M & Arnon DI (1965) Enzymic Mechanisms of Pyridine Nucleotide Reduction in Chloroplasts. *J. Biol. Chem.* 240:1405-1411.
35. Ferreira KN, Iverson TM, Maghlaoui K, Barber J, & Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. *Science* 303 (5665):1831-1838.
36. Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, & Saenger W (2009) Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* 16 (3):334-342.
37. Umena Y, Kawakami K, Shen JR, & Kamiya N (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* 473 (7345):55-U65.
38. Wei X, Su X, Cao P, Liu X, Chang W, Li M, Zhang X, & Liu Z (2016) Structure of spinach photosystem II-LHCII supercomplex at 3.2 Å resolution. *Nature* 534 (7605):69-74.
39. Ifuku K & Noguchi T (2016) Structural coupling of extrinsic proteins with the oxygen-evolving center in photosystem II. *Front Plant Sci* 7:84.
40. Michel H (1982) Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from *Rhodospseudomonas viridis*. *J. Mol. Biol.* 158 (3):567-572.

41. Deisenhofer J, Epp O, Miki K, Huber R, & Michel H (1985) Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3Å resolution. *Nature* 318 (6047):618-624.
42. Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, & Orth P (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409 (6821):739-743.
43. Murray JW, Maghlaoui K, Kargul J, Ishida N, Lai TL, Rutherford AW, Sugiura M, Boussac A, & Barber J (2008) X-ray crystallography identifies two chloride binding sites in the oxygen evolving centre of Photosystem II. *Energy Environ Sci* 1 (1):161-166.
44. Barber J (2008) Crystal structure of the oxygen-evolving complex of photosystem II. *Inorg. Chem.* 47 (6):1700-1710.
45. Dobakova M, Tichy M, & Komenda J (2007) Role of the PsbI protein in photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp PCC 6803. *Plant Physiol.* 145 (4):1681-1691.
46. Kern J & Guskov A (2011) Lipids in photosystem II: Multifunctional cofactors. *Journal of Photochemistry and Photobiology B-Biology* 104 (1-2):19-34.
47. Thornton LE, Ohkawa H, Roose JL, Kashino Y, Keren N, & Pakrasi HB (2004) Homologs of plant PsbP and PsbQ proteins are necessary for regulation of photosystem ii activity in the cyanobacterium *Synechocystis* 6803. *Plant Cell* 16 (8):2164-2175.
48. Michoux F, Boehm M, Bialek W, Takasaka K, Maghlaoui K, Barber J, Murray J, & Nixon P (2014) Crystal structure of CyanoQ from the thermophilic cyanobacterium *Thermosynechococcus elongatus* and detection in isolated photosystem II complexes. *Photosynthesis Res.* 122 (1):56-57.
49. Jackson SA, Hinds MG, & Eaton-Rye JJ (2012) Solution structure of CyanoP from *Synechocystis* sp. PCC 6803: new insights on the structural basis for functional specialization amongst PsbP family proteins. *Biochim. Biophys. Acta* 1817 (8):1331-1338.
50. Roose JL, Kashino Y, & Pakrasi HB (2007) The PsbQ protein defines cyanobacterial Photosystem II complexes with highest activity and stability. *Proc Natl Acad Sci U S A* 104 (7):2548-2553.
51. Liu H, Zhang H, Weisz DA, Vidavsky I, Gross ML, & Pakrasi HB (2014) MS-based cross-linking analysis reveals the location of the PsbQ protein in cyanobacterial photosystem II. *Proc Natl Acad Sci U S A* 111 (12):4638-4643.
52. Liu H, Weisz DA, & Pakrasi HB (2015) Multiple copies of the PsbQ protein in a cyanobacterial photosystem II assembly intermediate complex. *Photosynth Res* 126 (2-3):375-383.
53. Aoi M, Kashino Y, & Ifuku K (2014) Function and association of CyanoP in photosystem II of *Synechocystis* sp PCC 6803. *Research on Chemical Intermediates* 40 (9):3209-3217.
54. Jackson SA & Eaton-Rye JJ (2015) Characterization of a *Synechocystis* sp PCC 6803 double mutant lacking the CyanoP and Ycf48 proteins of Photosystem II. *Photosynthesis Res.* 124 (2):217-229.
55. Cormann KU, Bartsch M, Rogner M, & Nowaczyk MM (2014) Localization of the CyanoP binding site on photosystem II by surface plasmon resonance spectroscopy. *Front Plant Sci* 5:595.
56. Knoppová J, Yu J, Konik P, Nixon PJ, & Komenda J (2016) CyanoP is involved in the early steps of photosystem ii assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* 57 (9):1921-1931.
57. Giovannoni SJ, Turner S, Olsen GJ, Barns S, Lane DJ, & Pace NR (1988) Evolutionary Relationships among Cyanobacteria and Green Chloroplasts. *J. Bacteriol.* 170 (8):3584-3592.

58. Turner S, Pryer KM, Miao VPW, & Palmer JD (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* 46 (4):327-338.
59. Gould SB, Waller RR, & McFadden GI (2008) Plastid evolution. *Annu. Rev. Plant Biol.* 59:491-517.
60. Grossman AR, Bhaya D, Apt KE, & Kehoe DM (1995) Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu. Rev. Genet.* 29:231-288.
61. Garcia-Cerdan JG, Kovacs L, Toth T, Kereiche S, Aseeva E, Boekema EJ, Mamedov F, Funk C, & Schroder WP (2011) The PsbW protein stabilizes the supramolecular organization of photosystem II in higher plants. *Plant J.* 65 (3):368-381.
62. Stanier RY, Kunisawa R, Mandel M, & Cohen-Bazire G (1971) Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev* 35 (2):171-205.
63. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hiroseawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, & Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). *DNA Res.* 3 (3):185-209.
64. Kanesaki Y, Shiwa Y, Tajima N, Suzuki M, Watanabe S, Sato N, Ikeuchi M, & Yoshikawa H (2012) Identification of substrain-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803. *DNA Res.* 19 (1):67-79.
65. Trautmann D, Voss B, Wilde A, Al-Babili S, & Hess WR (2012) Microevolution in cyanobacteria: re-sequencing a motile substrain of *Synechocystis* sp. PCC 6803. *DNA Res.* 19 (6):435-448.
66. Ding Q, Chen G, Wang Y, & Wei D (2015) Identification of specific variations in a non-motile strain of cyanobacterium *Synechocystis* sp. PCC 6803 originated from ATCC 27184 by whole genome resequencing. *Int J Mol Sci* 16 (10):24081-24093.
67. Tichy M, Beckova M, Kopečna J, Noda J, Sobotka R, & Komenda J (2016) Strain of *Synechocystis* PCC 6803 with aberrant assembly of photosystem II contains tandem duplication of a large chromosomal region. *Front Plant Sci* 7:648.
68. Griesse M, Lange C, & Soppa J (2011) Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* 323 (2):124-131.
69. Williams JGK (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol.* 167:766-778.
70. Kufryk GI, Sachet M, Schmetterer G, & Vermaas WF (2002) Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiol. Lett.* 206 (2):215-219.
71. Yamaoka T, Satoh K, & Katoh S (1978) Photosynthetic activities of a thermophilic blue-green alga. *Plant Cell Physiol* 19 (6):943-954.
72. Bialek W, Wen S, Michoux F, Beckova M, Komenda J, Murray JW, & Nixon PJ (2013) Crystal structure of the Psb28 accessory factor of *Thermosynechococcus elongatus* photosystem II at 2.3 Å. *Photosynth Res* 117 (1-3):375-383.
73. Beckova M, Yu J, Krynicky V, Kozlo A, Shao S, Konik P, Komenda J, Murray JW, & Nixon PJ (2017) Structure of Psb29/Thf1 and its association with the FtsH protease complex involved in photosystem II repair in cyanobacteria. *Philos Trans R Soc Lond B Biol Sci* (In press).

74. Hellmich J, Bommer M, Burkhardt A, Ibrahim M, Kern J, Meents A, Muh F, Dobbek H, & Zouni A (2014) Native-like photosystem II superstructure at 2.44 Å resolution through detergent extraction from the protein crystal. *Structure* 22 (11):1607-1615.
75. Bommer M, Bondar AN, Zouni A, Dobbek H, & Dau H (2016) Crystallographic and computational analysis of the barrel part of the PsbO protein of photosystem II: carboxylate-water clusters as putative proton transfer relays and structural switches. *Biochemistry* 55 (33):4626-4635.
76. Kamiya N & Shen JR (2003) Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7 Å resolution. *Proc Natl Acad Sci U S A* 100 (1):98-103.
77. Kawakami K, Umena Y, Kamiya N, & Shen JR (2009) Location of chloride and its possible functions in oxygen-evolving photosystem II revealed by X-ray crystallography. *Proc Natl Acad Sci U S A* 106 (21):8567-8572.
78. Suga M, Akita F, Hirata K, Ueno G, Murakami H, Nakajima Y, Shimizu T, Yamashita K, Yamamoto M, Ago H, & Shen JR (2015) Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses. *Nature* 517 (7532):99-103.
79. Zilliges Y & Dau H (2016) Unexpected capacity for organic carbon assimilation by *Thermosynechococcus elongatus*, a crucial photosynthetic model organism. *FEBS Lett.* 590 (7):962-970.
80. Yamamoto Y (2016) Quality control of photosystem II: the mechanisms for avoidance and tolerance of light and heat stresses are closely linked to membrane fluidity of the thylakoids. *Front Plant Sci* 7:1136.
81. Lupinkova L & Komenda J (2004) Oxidative modifications of the Photosystem II D1 protein by reactive oxygen species: from isolated protein to cyanobacterial cells. *Photochem. Photobiol.* 79 (2):152-162.
82. Kale R, Hebert AE, Frankel LK, Sallans L, Bricker TM, & Pospisil P (2017) Amino acid oxidation of the D1 and D2 proteins by oxygen radicals during photoinhibition of Photosystem II. *Proc Natl Acad Sci U S A* 114 (11):2988-2993.
83. Yao DC, Brune DC, Vavilin D, & Vermaas WF (2012) Photosystem II component lifetimes in the cyanobacterium *Synechocystis* sp. strain PCC 6803: small Cab-like proteins stabilize biosynthesis intermediates and affect early steps in chlorophyll synthesis. *J. Biol. Chem.* 287 (1):682-692.
84. Tyystjarvi E (2008) Photoinhibition of Photosystem II and photodamage of the oxygen evolving manganese cluster. *Coord Chem Rev* 252 (3-4):361-376.
85. Eckert HJ, Geiken B, Bernarding J, Napiwotzki A, Eichler HJ, & Renger G (1991) Two sites of photoinhibition of the electron transfer in oxygen evolving and Tris-treated PS II membrane fragments from spinach. *Photosynth Res* 27 (2):97-108.
86. Hakala M, Tuominen I, Keranen M, Tyystjarvi T, & Tyystjarvi E (2005) Evidence for the role of the oxygen-evolving manganese complex in photoinhibition of Photosystem II. *Biochim. Biophys. Acta* 1706 (1-2):68-80.
87. Chen GX, Kazimir J, & Cheniae GM (1992) Photoinhibition of Hydroxylamine-Extracted Photosystem-II Membranes - Studies of the Mechanism. *Biochemistry* 31 (45):11072-11083.
88. Mano J, Takahashi M, & Asada K (2002) Oxygen evolution from hydrogen peroxide in photosystem II: flash-induced catalytic activity of water-oxidizing photosystem II membranes. *Biochemistry* 26 (9):2495-2501.
89. Pospisil P, Snrychova I, & Naus J (2007) Dark production of reactive oxygen species in photosystem II membrane particles at elevated temperature: EPR spin-trapping study. *Biochim. Biophys. Acta* 1767 (6):854-859.
90. Vass I & Cser K (2009) Janus-faced charge recombinations in photosystem II photoinhibition. *Trends Plant Sci.* 14 (4):200-205.

91. Vass I, Styring S, Hundal T, Koivuniemi A, Aro EM, & Andersson B (1992) Reversible and Irreversible Intermediates during Photoinhibition of Photosystem .2. Stable Reduced Qa Species Promote Chlorophyll Triplet Formation. *Proc Natl Acad Sci U S A* 89 (4):1408-1412.
92. Vass I & Styring S (1993) Characterization of Chlorophyll Triplet Promoting States in Photosystem-II Sequentially Induced during Photoinhibition. *Biochemistry* 32 (13):3334-3341.
93. Fischer BB, Hideg E, & Krieger-Liszkay A (2013) Production, detection, and signaling of singlet oxygen in photosynthetic organisms. *Antioxid. Redox Signal.* 18 (16):2145-2162.
94. Keren N, Berg A, VanKan PJM, Levanon H, & Ohad I (1997) Mechanism of photosystem II photoinactivation and D1 protein degradation at low light: The role of back electron flow. *Proc Natl Acad Sci U S A* 94 (4):1579-1584.
95. Constant S, Perewoska I, Alfonso M, & Kirilovsky D (1997) Expression of the *psbA* gene during photoinhibition and recovery in *Synechocystis* PCC 6714: inhibition and damage of transcriptional and translational machinery prevent the restoration of photosystem II activity. *Plant Mol. Biol.* 34 (1):1-13.
96. Nishiyama Y, Yamamoto H, Allakhverdiev SI, Inaba M, Yokota A, & Murata N (2001) Oxidative stress inhibits the repair of photodamage to the photosynthetic machinery. *EMBO J.* 20 (20):5587-5594.
97. Cardona T, Murray JW, & Rutherford AW (2015) Origin and evolution of water oxidation before the last common ancestor of the cyanobacteria. *Mol. Biol. Evol.* 32 (5):1310-1328.
98. Crawford TS, Hanning KR, Chua JP, Eaton-Rye JJ, & Summerfield TC (2016) Comparison of D1'- and D1-containing PS II reaction centre complexes under different environmental conditions in *Synechocystis* sp. PCC 6803. *Plant Cell Environ* 39 (8):1715-1726.
99. Murray JW (2012) Sequence variation at the oxygen-evolving centre of photosystem II: a new class of 'rogue' cyanobacterial D1 proteins. *Photosynthesis Res.* 110 (3):177-184.
100. Sugiura M, Azami C, Koyama K, Rutherford AW, Rappaport F, & Boussac A (2014) Modification of the pheophytin redox potential in *Thermosynechococcus elongatus* Photosystem II with PsbA3 as D1. *Biochim. Biophys. Acta* 1837 (1):139-148.
101. Komenda J, Sobotka R, & Nixon PJ (2012) Assembling and maintaining the Photosystem II complex in chloroplasts and cyanobacteria. *Curr. Opin. Plant Biol.* 15 (3):245-251.
102. Komenda J, Nickelsen J, Tichy M, Prasil O, Eichacker LA, & Nixon PJ (2008) The cyanobacterial homologue of HCF136/YCF48 is a component of an early photosystem II assembly complex and is important for both the efficient assembly and repair of photosystem II in *Synechocystis* sp PCC 6803. *J. Biol. Chem.* 283 (33):22390-22399.
103. Komenda J, Reisinger V, Muller BC, Dobakova M, Granvogl B, & Eichacker LA (2004) Accumulation of the D2 protein is a key regulatory step for assembly of the photosystem II reaction center complex in *Synechocystis* PCC 6803. *J. Biol. Chem.* 279 (47):48620-48629.
104. Rengstl B, Oster U, Stengel A, & Nickelsen J (2011) An intermediate membrane subfraction in cyanobacteria is involved in an assembly network for Photosystem II biogenesis. *J. Biol. Chem.* 286 (24):21944-21951.
105. Schottkowski M, Gkalypoudis S, Tzekova N, Stelljes C, Schunemann D, Ankele E, & Nickelsen J (2009) Interaction of the periplasmic PratA factor and the PsbA (D1) protein during biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 284 (3):1813-1819.

106. Rengstl B, Knoppova J, Komenda J, & Nickelsen J (2013) Characterization of a *Synechocystis* double mutant lacking the photosystem II assembly factors YCF48 and SII0933. *Planta* 237 (2):471-480.
107. Boehm M, Romero E, Reisinger V, Yu J, Komenda J, Eichacker LA, Dekker JP, & Nixon PJ (2011) Investigating the early stages of photosystem II assembly in *Synechocystis* sp. PCC 6803: ISOLATION OF CP47 AND CP43 COMPLEXES. *J. Biol. Chem.* 286 (17):14812-14819.
108. Boehm M, Yu J, Reisinger V, Beckova M, Eichacker LA, Schlodder E, Komenda J, & Nixon PJ (2012) Subunit composition of CP43-less photosystem II complexes of *Synechocystis* sp PCC 6803: implications for the assembly and repair of photosystem II. *Philos Trans R Soc Lond B Biol Sci* 367 (1608):3444-3454.
109. Dobakova M, Sobotka R, Tichy M, & Komenda J (2009) Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 149 (2):1076-1086.
110. Beckova M, Gardian Z, Yu J, Konik P, Nixon PJ, & Komenda J (2017) Association of Psb28 and Psb27 proteins with PSII-PSI supercomplexes upon exposure of *Synechocystis* sp. PCC 6803 to high light. *Mol Plant* 10 (1):62-72.
111. Weisz DA, Liu H, Zhang H, Thangapandian S, Tajkhorshid E, Gross ML, & Pakrasi HB (2017) Mass spectrometry-based cross-linking study shows that the Psb28 protein binds to cytochrome *b*₅₅₉ in Photosystem II. *Proc Natl Acad Sci U S A* 114 (9):2224-2229.
112. Rogner M, Chisholm DA, & Diner BA (1991) Site-Directed Mutagenesis of the PsbC-Gene of Photosystem-II - Isolation and Functional-Characterization of CP43-Less Photosystem-II Core Complexes. *Biochemistry* 30 (22):5387-5395.
113. Komenda J, Knoppova J, Kopečna J, Sobotka R, Halada P, Yu J, Nickelsen J, Boehm M, & Nixon PJ (2012) The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 158 (1):476-486.
114. Liu H, Huang RY, Chen J, Gross ML, & Pakrasi HB (2011) Psb27, a transiently associated protein, binds to the chlorophyll binding protein CP43 in photosystem II assembly intermediates. *Proc Natl Acad Sci U S A* 108 (45):18536-18541.
115. Roose JL & Pakrasi HB (2008) The Psb27 protein facilitates manganese cluster assembly in photosystem II. *J. Biol. Chem.* 283 (7):4044-4050.
116. Mabbitt PD, Wilbanks SM, & Eaton-Rye JJ (2014) Structure and function of the hydrophilic Photosystem II assembly proteins: Psb27, Psb28 and Ycf48. *Plant Physiol. Biochem.* 81:96-107.
117. Nowaczyk MM, Hebeler R, Schlodder E, Meyer HE, Warscheid B, & Rogner M (2006) Psb27, a cyanobacterial lipoprotein, is involved in the repair cycle of photosystem II. *Plant Cell* 18 (11):3121-3131.
118. Wagner R, Aigner H, & Funk C (2012) FtsH proteases located in the plant chloroplast. *Physiol. Plant.* 145 (1):203-214.
119. Langklotz S, Baumann U, & Narberhaus F (2012) Structure and function of the bacterial AAA protease FtsH. *Biochim. Biophys. Acta* 1823 (1):40-48.
120. Boehm M, Yu J, Krynicka V, Barker M, Tichy M, Komenda J, Nixon PJ, & Nield J (2012) Subunit organization of a *Synechocystis* hetero-oligomeric thylakoid FtsH complex involved in photosystem II repair. *Plant Cell* 24 (9):3669-3683.
121. Komenda J, Tichy M, Prasil O, Knoppova J, Kuvikova S, de Vries R, & Nixon PJ (2007) The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in *Synechocystis* sp PCC 6803. *Plant Cell* 19 (9):2839-2854.
122. Nitschke W & Rutherford AW (1991) Photosynthetic reaction centres: variations on a common structural theme? *Trends Biochem. Sci.* 16 (7):241-245.

123. Nitschke W, Mattioli T, & Rutherford AW (1996) The FeS-type photosystems and the evolution of photosynthetic reaction centers. *Origin and Evolution of Biological Energy Conversion*, ed Baltscheffsky H (Wiley-VCH), pp 177-203.
124. Rutherford AW & Nitschke W (1996) Photosystem II and the quinone–iron-containing reaction centers: comparisons and evolutionary perspectives. *Origin and evolution of biological energy conversion*, ed Baltscheffsky H (New York: VCH), pp 143-175.
125. Schubert WD, Klukas O, Saenger W, Witt HT, Fromme P, & Krauss N (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. *J. Mol. Biol.* 280 (2):297-314.
126. Nixon PJ, Trost JT, & Diner BA (1992) Role of the carboxy terminus of polypeptide D1 in the assembly of a functional water-oxidizing manganese cluster in photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803: assembly requires a free carboxyl group at C-terminal position 344. *Biochemistry* 31 (44):10859-10871.
127. Nagarajan A, Winter R, Eaton-Rye J, & Burnap R (2011) A synthetic DNA and fusion PCR approach to the ectopic expression of high levels of the D1 protein of photosystem II in *Synechocystis* sp. PCC 6803. *J Photochem Photobiol B* 104 (1-2):212-219.
128. Inoue H, Nojima H, & Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96 (1):23-28.
129. Lopez JA & Bohuski E (2007) Total RNA extraction with TRIZOL reagent and purification with QIAGEN RNeasy Mini Kit. (DGC, Indiana University, USA).
130. Rao X, Huang X, Zhou Z, & Lin X (2013) An improvement of the $2^{-\Delta\Delta CT}$ method for quantitative real-time polymerase chain reaction data analysis. *Biostat Bioinforma Biomath* 3 (3):71-85.
131. Lichtenthaler HK & Wellburn AR (1983) Determinations of total carotenoids and chlorophylls *a* and *b* of leaf extracts in different solvents. *Biochem. Soc. Trans.* 11 (5):591-592.
132. Schagger H (2006) Tricine-SDS-PAGE. *Nature Protocols* 1 (1):16-22.
133. Prioul JL & Chartier P (1977) Partitioning of transfer and carboxylation components of intracellular resistance to photosynthetic CO₂ fixation: a critical analysis of the methods used. *Ann. Bot.* 41 (174):789-800.
134. Lobo FD, de Barros MP, Dalmagro HJ, Dalmolin AC, Pereira WE, de Souza EC, Vourlitis GL, & Ortiz CER (2013) Fitting net photosynthetic light-response curves with Microsoft Excel - a critical look at the models. *Photosynthetica* 51 (3):445-456.
135. Edwards DJ & Holt KE (2013) Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation* 3 (1):2.
136. Gurevich A, Saveliev V, Vyahhi N, & Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8):1072-1075.
137. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18 (5):821-829.
138. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, & Perna NT (2009) Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* 25 (16):2071-2073.
139. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, & Parkhill J (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21 (16):3422-3433.
140. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (14):2068-2069.
141. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, & Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44 (D1):D279-285.

142. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* 7 (10):e1002195.
143. Zakon HH (2002) Convergent evolution on the molecular level. *Brain Behav. Evol.* 59 (5-6):250-261.
144. Mistry J, Finn RD, Eddy SR, Bateman A, & Punta M (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41 (12):e121.
145. Yamada KD, Tomii K, & Katoh K (2016) Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* 32 (21):3246-3251.
146. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25 (15):1972-1973.
147. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, & Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59 (3):307-321.
148. Letunic I & Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44 (W1):W242-W245.
149. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, HERNSDORF AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, & Banfield JF (2016) A new view of the tree of life. *Nature Microbiology* 1:16048.
150. Huerta-Cepas J, Serra F, & Bork P (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33 (6):1635-1638.
151. Campanella JJ, Bitincka L, & Smalley J (2003) MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* 4 (1):29.
152. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, & Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44 (W1):W344-350.
153. Emanuelsson O, Nielsen H, & von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8 (5):978-984.
154. Golbeck JH (1993) Shared thematic elements in photochemical reaction centers. *Proc Natl Acad Sci U S A* 90 (5):1642-1646.
155. Vermaas WF (1994) Evolution of heliobacteria: Implications for photosynthetic reaction center complexes. *Photosynth Res* 41 (1):285-294.
156. Allen JP & Williams JC (1998) Photosynthetic reaction centers. *FEBS Lett.* 438 (1-2):5-9.
157. Rutherford AW & Faller P (2003) Photosystem II: evolutionary perspectives. *Philos Trans R Soc Lond B Biol Sci* 358 (1429):245-253.
158. Sadekar S, Raymond J, & Blankenship RE (2006) Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core. *Mol. Biol. Evol.* 23 (11):2001-2007.
159. Soo RM, Hemp J, Parks DH, Fischer WW, & Hugenholtz P (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* 355 (6332):1436-1440.
160. Cardona T (2015) A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynthesis Res.* 126 (1):111-134.

161. Barber J, Morris E, & Buchel C (2000) Revealing the structure of the photosystem II chlorophyll binding proteins, CP43 and CP47. *Biochim. Biophys. Acta* 1459 (2-3):239-247.
162. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, & Krauss N (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411 (6840):909-917.
163. Gupta RS (2012) Origin and spread of photosynthesis based upon conserved sequence features in key bacteriochlorophyll biosynthesis proteins. *Mol. Biol. Evol.* 29 (11):3397-3412.
164. Xiong J, Fischer WM, Inoue K, Nakahara M, & Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. *Science* 289 (5485):1724-1730.
165. Nelson N (2013) Evolution of photosystem I and the control of global enthalpy in an oxidizing world. *Photosynthesis Res.* 116 (2-3):145-151.
166. Allen JF (2005) A redox switch hypothesis for the origin of two light reactions in photosynthesis. *FEBS Lett.* 579 (5):963-968.
167. Mulikidjanian AY & Galperin MY (2013) Time to scatter genes and a time to gather them: evolution of photosynthesis genes in bacteria. *Adv. Bot. Res.* 66:1-35.
168. Baymann D, Brugna M, Muhlenhoff U, & Nitschke W (2001) Daddy, where did (PS)I come from? *Biochim. Biophys. Acta* 1507 (1-3):291-310.
169. Cavalier-Smith T (2006) Rooting the tree of life by transition analyses. *Biology Direct* 1:19.
170. Fischer WW, Hemp J, & Johnson JE (2016) Evolution of oxygenic photosynthesis. *Annual Review of Earth and Planetary Sciences* 44 (1):647-683.
171. Sousa FL, Shavit-Grievink L, Allen JF, & Martin WF (2013) Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol Evol* 5 (1):200-216.
172. Ward LM, Kirschvink JL, & Fischer WW (2016) Timescales of oxygenation following the evolution of oxygenic photosynthesis. *Origins of Life and Evolution of Biospheres* 46 (1):51-65.
173. Dutkiewicz A, Volk H, George SC, Ridley J, & Buick R (2006) Biomarkers from Huronian oil-bearing fluid inclusions: An uncontaminated record of life before the Great Oxidation Event. *Geology* 34 (6):437-440.
174. Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J, & Buick R (2007) A Whiff of Oxygen Before the Great Oxidation Event? *Science* 317 (5846):1903-1906.
175. Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, & Beukes NJ (2004) Dating the rise of atmospheric oxygen. *Nature* 427 (6970):117-120.
176. Johnson JE, Webb SM, Thomas K, Ono S, Kirschvink JL, & Fischer WW (2013) Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Natl Acad Sci U S A* 110 (28):11238-11243.
177. Lyons TW, Reinhard CT, & Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506 (7488):307-315.
178. Shih PM, Hemp J, Ward LM, Matzke NJ, & Fischer WW (2017) Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* 15 (1):19-29.
179. Hohmann-Marriott MF & Blankenship RE (2011) Evolution of photosynthesis. *Annu. Rev. Plant Biol.* 62:515-548.
180. Allwood AC, Walter MR, Kamber BS, Marshall CP, & Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* 441 (7094):714-718.
181. Olson JM (2006) Photosynthesis in the Archean Era. *Photosynthesis Res.* 88 (2):109-117.

182. Schopf JW (2011) The paleobiological record of photosynthesis. *Photosynthesis Res.* 107 (1):87-101.
183. Rosing MT & Frei R (2004) U-rich Archean sea-floor sediments from Greenland - indications of > 3700 Ma oxygenic photosynthesis. *Earth Planet Sci Lett* 217 (3-4):237-244.
184. Cardona T (2016) Origin of bacteriochlorophyll a and the early diversification of photosynthesis. *PLoS One* 11 (3):e0151250.
185. Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, Dufresne A, Partensky F, Burd H, Kaznadzey D, Haselkorn R, & Galperin MY (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci U S A* 103 (35):13126-13131.
186. Tandori J, Hideg E, Nagy L, Maroti P, & Vass I (2001) Photoinhibition of carotenoidless reaction centers from *Rhodobacter sphaeroides* by visible light. Effects on protein structure and electron transport. *Photosynth Res* 70 (2):175-184.
187. Sonoike K (2011) Photoinhibition of photosystem I. *Physiol. Plant.* 142 (1):56-64.
188. Krynicka V, Shao SX, Nixon PJ, & Komenda J (2015) Accessibility controls selective degradation of photosystem II subunits by FtsH protease. *Nature Plants* 1:15168.
189. Cardona T, Sanchez-Baracaldo P, Rutherford AW, & Larkum A (2017) Molecular evidence for the early evolution of photosynthetic water oxidation. *bioRxiv.org* doi:<https://doi.org/10.1101/109447>.
190. Michoux F, Ahmad N, Wei ZY, Belgio E, Ruban AV, & Nixon PJ (2016) Testing the role of the N-terminal tail of D1 in the maintenance of photosystem II in Tobacco chloroplasts. *Front Plant Sci* 7:844.
191. Chen X, Zaro JL, & Shen WC (2013) Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 65 (10):1357-1369.
192. Hoedemaeker FJ, Signorelli T, Johns K, Kuntz DA, & Rose DR (1997) A single chain Fv fragment of P-glycoprotein-specific monoclonal antibody C219. Design, expression, and crystal structure at 2.4 Å resolution. *J. Biol. Chem.* 272 (47):29784-29789.
193. Trinh R, Gurbaxani B, Morrison SL, & Seyfzadeh M (2004) Optimization of codon pair use within the (GGGS)₃ linker sequence results in enhanced protein expression. *Mol. Immunol.* 40 (10):717-722.
194. Nanjo Y, Mizusawa N, Wada H, Slabas AR, Hayashi H, & Nishiyama Y (2010) Synthesis of fatty acids de novo is required for photosynthetic acclimation of *Synechocystis* sp. PCC 6803 to high temperature. *Biochim. Biophys. Acta* 1797 (8):1483-1490.
195. Kashino Y, Lauber WM, Carroll JA, Wang QJ, Whitmarsh J, Satoh K, & Pakrasi HB (2002) Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* 41 (25):8004-8012.
196. Peter E, Salinas A, Wallner T, Jeske D, Dienst D, Wilde A, & Grimm B (2009) Differential requirement of two homologous proteins encoded by *sll1214* and *sll1874* for the reaction of Mg protoporphyrin monomethylester oxidative cyclase under aerobic and micro-oxic growth conditions. *Biochim. Biophys. Acta* 1787 (12):1458-1467.
197. Hihara Y, Sonoike K, Kanehisa M, & Ikeuchi M (2003) DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 185 (5):1719-1725.
198. Komenda J, Tichy M, & Eichacker LA (2005) The PsbH protein is associated with the inner antenna CP47 and facilitates D1 processing and incorporation into PSII in the cyanobacterium *Synechocystis* PCC 6803. *Plant Cell Physiol* 46 (9):1477-1483.
199. Krishna PS, Rani BR, Mohan MK, Suzuki I, Shivaji S, & Prakash JS (2013) A novel transcriptional regulator, Sll1130, negatively regulates heat-responsive genes in *Synechocystis* sp. PCC6803. *Biochem. J.* 449 (3):751-760.

200. Butler WL & Strasser RJ (1977) Tripartite model for the photochemical apparatus of green plant photosynthesis. *Proc Natl Acad Sci U S A* 74 (8):3382-3385.
201. Bradbury M & Barker NR (1983) Analysis of the induction of chlorophyll fluorescence in leaves and isolated thylakoids: contributions of photochemical and non-photochemical quenching. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 220 (1219):251-264.
202. Kautsky H, Appel W, & Amann H (1960) Chlorophyll fluorescence and carbon assimilation. Part XIII. The fluorescence and the photochemistry of plants. *Biochem Z* 332:277-292.
203. Vass I, Kirilovsky D, & Etienne AL (1999) UV-B radiation-induced donor- and acceptor-side modifications of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochemistry* 38 (39):12786-12794.
204. Allahverdiyeva Y, Deak Z, Szilard A, Diner BA, Nixon PJ, & Vass I (2004) The function of D1-H332 in Photosystem II electron transport studied by thermoluminescence and chlorophyll fluorescence in site-directed mutants of *Synechocystis* 6803. *Eur. J. Biochem.* 271 (17):3523-3532.
205. Nixon PJ & Diner BA (1992) Aspartate-170 of the Photosystem-II Reaction Center Polypeptide-D1 Is Involved in the Assembly of the Oxygen-Evolving Manganese Cluster. *Biochemistry* 31 (3):942-948.
206. Crofts AR & Wraight CA (1983) The electrochemical domain of photosynthesis. *Biochim. Biophys. Acta* 726 (3):149-185.
207. Knoppova J, Sobotka R, Tichy M, Yu J, Konik P, Halada P, Nixon PJ, & Komenda J (2014) Discovery of a chlorophyll binding protein complex involved in the early steps of photosystem II assembly in *Synechocystis*. *Plant Cell* 26 (3):1200-1212.
208. Vass IZ, Kos PB, Knoppova J, Komenda J, & Vass I (2014) The cry-DASH cryptochrome encoded by the *sll1629* gene in the cyanobacterium *Synechocystis* PCC 6803 is required for Photosystem II repair. *J Photochem Photobiol B* 130:318-326.
209. Hager M, Hermann M, Biehler A, Krieger-Liszkay A, & Bock R (2002) Lack of the small plastid-encoded PsbJ polypeptide results in a defective water-splitting apparatus of photosystem II, reduced photosystem I levels, and hypersensitivity to light. *J. Biol. Chem.* 277 (16):14031-14039.
210. Kaneko T & Tabata S (1997) Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol* 38 (11):1171-1176.
211. Tajima N, Sato S, Maruyama F, Kaneko T, Sasaki NV, Kurokawa K, Ohta H, Kanesaki Y, Yoshikawa H, Tabata S, Ikeuchi M, & Sato N (2011) Genomic structure of the cyanobacterium *Synechocystis* sp. PCC 6803 strain GT-S. *DNA Res.* 18 (5):393-399.
212. Prentki P & Krisch HM (1984) In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* 29 (3):303-313.
213. Kawakami K, Umena Y, Iwai M, Kawabata Y, Ikeuchi M, Kamiya N, & Shen JR (2011) Roles of PsbI and PsbM in photosystem II dimer formation and stability studied by deletion mutagenesis and X-ray crystallography. *Biochim. Biophys. Acta* 1807 (3):319-325.
214. Komenda J (2005) Autotrophic cells of the *Synechocystis psbH* deletion mutant are deficient in synthesis of CP47 and accumulate inactive PSII core complexes. *Photosynthesis Res.* 85 (2):161-167.
215. Chidgey JW, Linhartova M, Komenda J, Jackson PJ, Dickman MJ, Canniffe DP, Konik P, Pilny J, Hunter CN, & Sobotka R (2014) A cyanobacterial chlorophyll synthase-HliD complex associates with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* 26 (3):1267-1279.
216. Krieger-Liszkay A (2005) Singlet oxygen production in photosynthesis. *J. Exp. Bot.* 56 (411):337-346.

217. Takahashi S & Badger MR (2011) Photoprotection in plants: a new light on photosystem II damage. *Trends Plant Sci.* 16 (1):53-60.
218. Sokolenko A, Pojidaeva E, Zinchenko V, Panichkin V, Glaser VM, Herrmann RG, & Shestakov SV (2002) The gene complement for proteolysis in the cyanobacterium *Synechocystis* sp. PCC 6803 and *Arabidopsis thaliana* chloroplasts. *Curr. Genet.* 41 (5):291-310.
219. Sakamoto W, Zaltsman A, Adam Z, & Takahashi Y (2003) Coordinated regulation and complex formation of YELLOW VARIEGATED1 and YELLOW VARIEGATED2, chloroplastic FtsH metalloproteases involved in the repair cycle of photosystem II in *Arabidopsis thylakoid* membranes. *Plant Cell* 15 (12):2843-2855.
220. Sacharz J, Bryan SJ, Yu J, Burroughs NJ, Spence EM, Nixon PJ, & Mullineaux CW (2015) Sub-cellular location of FtsH proteases in the cyanobacterium *Synechocystis* sp. PCC 6803 suggests localised PSII repair zones in the thylakoid membranes. *Mol. Microbiol.* 96 (3):448-462.
221. Krynicka V, Tichy M, Krafl J, Yu JF, Kana R, Boehm M, Nixon PJ, & Komenda J (2014) Two essential FtsH proteases control the level of the Fur repressor during iron deficiency in the cyanobacterium *Synechocystis* sp PCC 6803. *Mol. Microbiol.* 94 (3):609-624.
222. Urantowka A, Knorpp C, Olczak T, Kolodziejczak M, & Janska H (2005) Plant mitochondria contain at least two *i*-AAA-like complexes. *Plant Mol. Biol.* 59 (2):239-252.
223. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, & Millar AH (2004) Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16 (1):241-256.
224. Yu F, Park S, & Rodermel SR (2004) The *Arabidopsis* FtsH metalloprotease gene family: interchangeability of subunits in chloroplast oligomeric complexes. *Plant J.* 37 (6):864-876.
225. Malnoe A, Wang F, Girard-Bascou J, Wollman FA, & de Vitry C (2014) Thylakoid FtsH protease contributes to photosystem II and cytochrome *b₆f* remodeling in *Chlamydomonas reinhardtii* under stress conditions. *Plant Cell* 26 (1):373-390.
226. Yu F, Park S, & Rodermel SR (2005) Functional redundancy of AtFtsH metalloproteases in thylakoid membrane complexes. *Plant Physiol.* 138 (4):1957-1966.
227. Zaltsman A, Ori N, & Adam Z (2005) Two types of FtsH protease subunits are required for chloroplast biogenesis and Photosystem II repair in *Arabidopsis*. *Plant Cell* 17 (10):2782-2790.
228. Ogura T, Inoue K, Tatsuta T, Suzaki T, Karata K, Young K, Su LH, Fierke CA, Jackman JE, Raetz CR, Coleman J, Tomoyasu T, & Matsuzawa H (1999) Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli*. *Mol. Microbiol.* 31 (3):833-844.
229. Langklotz S, Schakermann M, & Narberhaus F (2011) Control of lipopolysaccharide biosynthesis by FtsH-mediated proteolysis of LpxC is conserved in enterobacteria but not in all gram-negative bacteria. *J. Bacteriol.* 193 (5):1090-1097.
230. Zhao K, Liu M, & Burgess RR (2005) The global transcriptional response of *Escherichia coli* to induced σ^{32} protein involves σ^{32} regulon activation followed by inactivation and degradation of σ^{32} in vivo. *J. Biol. Chem.* 280 (18):17758-17768.
231. Herman C, Thevenet D, D'Ari R, & Boulloc P (1995) Degradation of sigma 32, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proc Natl Acad Sci U S A* 92 (8):3516-3520.

232. Tomoyasu T, Gamer J, Bukau B, Kanemori M, Mori H, Rutman AJ, Oppenheim AB, Yura T, Yamanaka K, Niki H, Hiraga S, & Ogura T (1995) *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma (32). *EMBO J.* 14 (11):2551-2560.
233. Blaszcak A, Georgopoulos C, & Liberek K (1999) On the mechanism of FtsH-dependent degradation of the sigma 32 transcriptional regulator of *Escherichia coli* and the role of the Dnak chaperone machine. *Mol. Microbiol.* 31 (1):157-166.
234. Tomoyasu T, Ogura T, Tatsuta T, & Bukau B (1998) Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. *Mol. Microbiol.* 30 (3):567-581.
235. Herman C, Thevenet D, Boulloc P, Walker GC, & D'Ari R (1998) Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* 12 (9):1348-1355.
236. Hari SB & Sauer RT (2016) The AAA+ FtsH protease degrades an ssrA-tagged model protein in the inner membrane of *Escherichia coli*. *Biochemistry* 55 (40):5649-5652.
237. Kihara A, Akiyama Y, & Ito K (1995) FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* 92 (10):4532-4536.
238. van Stelten J, Silva F, Belin D, & Silhavy TJ (2009) Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY. *Science* 325 (5941):753-756.
239. Fuhrer F, Langklotz S, & Narberhaus F (2006) The C-terminal end of LpxC is required for degradation by the FtsH protease. *Mol. Microbiol.* 59 (3):1025-1036.
240. Katz C & Ron EZ (2008) Dual Role of FtsH in Regulating Lipopolysaccharide Biosynthesis in *Escherichia coli*. *J. Bacteriol.* 190 (21):7117-7122.
241. Leffers GG & Gottesman S (1998) Lambda Xis degradation in vivo by lon and ftsH. *J. Bacteriol.* 180 (6):1573-1577.
242. Kihara A, Akiyama Y, & Ito K (1997) Host regulation of lysogenic decision in bacteriophage lambda: Transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proc Natl Acad Sci U S A* 94 (11):5544-5549.
243. Herman C, Thevenet D, D'Ari R, & Boulloc P (1997) The HflB protease of *Escherichia coli* degrades its inhibitor lambda cIII. *J. Bacteriol.* 179 (2):358-363.
244. Westphal K, Langklotz S, Thomanek N, & Narberhaus F (2012) A trapping approach reveals novel substrates and physiological functions of the essential protease FtsH in *Escherichia coli*. *J. Biol. Chem.* 287 (51):42962-42971.
245. Kihara A, Akiyama Y, & Ito K (1998) Different pathways for protein degradation by the FtsH/HflKC membrane-embedded protease complex: An implication from the interference by a mutant form of a new substrate protein, YccA. *J. Mol. Biol.* 279 (1):175-188.
246. Akiyama Y, Kihara A, & Ito K (1996) Subunit a of proton ATPase F₀ sector is a substrate of the FtsH protease in *Escherichia coli*. *FEBS Lett.* 399 (1-2):26-28.
247. Singh S & Darwin AJ (2011) FtsH-dependent degradation of phage shock protein C in *Yersinia enterocolitica* and *Escherichia coli*. *J. Bacteriol.* 193 (23):6436-6442.
248. Bittner LM, Westphal K, & Narberhaus F (2015) Conditional proteolysis of the membrane protein YfgM by the FtsH protease depends on a novel N-terminal degron. *J. Biol. Chem.* 290 (31):19367-19378.
249. Sedaghatmehr M, Mueller-Roeber B, & Balazadeh S (2016) The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in *Arabidopsis*. *Nature Communication* 7:12439.

250. Tomoyasu T, Yamanaka K, Murata K, Suzuki T, Bouloc P, Kato A, Niki H, Hiraga S, & Ogura T (1993) Topology and subcellular localization of FtsH protein in *Escherichia coli*. *J. Bacteriol.* 175 (5):1352-1357.
251. Leonhard K, Herrmann JM, Stuart RA, Mannhaupt G, Neupert W, & Langer T (1996) AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria. *EMBO J.* 15 (16):4218-4229.
252. Klanner C, Prokisch H, & Langer T (2001) MAP-1 and IAP-1, two novel AAA proteases with catalytic sites on opposite membrane surfaces in mitochondrial inner membrane of *Neurospora crassa*. *Mol Biol Cell* 12 (9):2858-2869.
253. Ramelot TA, Yang Y, Sahu ID, Lee HW, Xiao R, Lorigan GA, Montelione GT, & Kennedy MA (2013) NMR structure and MD simulations of the AAA protease intermembrane space domain indicates peripheral membrane localization within the hexaoligomer. *FEBS Lett.* 587 (21):3522-3528.
254. Scharfenberg F, Serek-Heuberger J, Coles M, Hartmann MD, Habeck M, Martin J, Lupas AN, & Alva V (2015) Structure and evolution of N-domains in AAA metalloproteases. *J. Mol. Biol.* 427 (4):910-923.
255. Bieniossek C, Schalch T, Bumann M, Meister M, Meier R, & Baumann U (2006) The molecular architecture of the metalloprotease FtsH. *Proc Natl Acad Sci U S A* 103 (9):3066-3071.
256. Suno R, Niwa H, Tsuchiya D, Zhang X, Yoshida M, & Morikawa K (2006) Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* 22 (5):575-585.
257. Bieniossek C, Niederhauser B, & Baumann UM (2009) The crystal structure of *apo*-FtsH reveals domain movements necessary for substrate unfolding and translocation. *Proc Natl Acad Sci U S A* 106 (51):21579-21584.
258. Lee S, Augustin S, Tatsuta T, Gerdes F, Langer T, & Tsai FT (2011) Electron cryomicroscopy structure of a membrane-anchored mitochondrial AAA protease. *J. Biol. Chem.* 286 (6):4404-4411.
259. Krzywda S, Brzozowski AM, Verma C, Karata K, Ogura T, & Wilkinson AJ (2002) The crystal structure of the AAA domain of the ATP-dependent protease FtsH of *Escherichia coli* at 1.5 Å resolution. *Structure* 10 (8):1073-1083.
260. Niwa H, Tsuchiya D, Makyio H, Yoshida M, & Morikawa K (2002) Hexameric ring structure of the ATPase domain of the membrane-integrated metalloprotease FtsH from *Thermus thermophilus* HB8. *Structure* 10 (10):1415-1423.
261. Moldavski O, Levin-Kravets O, Ziv T, Adam Z, & Prag G (2012) The hetero-hexameric nature of a chloroplast AAA+ FtsH protease contributes to its thermodynamic stability. *PLoS One* 7 (4):e36008.
262. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, & Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311 (5765):1283-1287.
263. Jun SR, Sims GE, Wu GHA, & Kim SH (2010) Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution. *Proc Natl Acad Sci U S A* 107 (1):133-138.
264. Segata N, Bornigen D, Morgan XC, & Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications* 4:2304.
265. Bryant D, Liu Z, Li T, Zhao F, Klatt CG, Ward D, Frigaard NU, & Overmann J (2012) Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. *Functional Genomics and Evolution of Photosynthetic Systems*, Advances in Photosynthesis and Respiration, eds Burnap RL & Vermaas W (Springer, Dordrecht), Vol 33, pp 47-102.

266. Greening C, Carere CR, Rushton-Green R, Harold LK, Hards K, Taylor MC, Morales SE, Stott MB, & Cook GM (2015) Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proc Natl Acad Sci U S A* 112 (33):10497-10502.
267. Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, & Schleper C (2003) Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* 50 (2):563-575.
268. David LA & Alm EJ (2011) Rapid evolutionary innovation during an Archaeal genetic expansion. *Nature* 469 (7328):93-96.
269. Dutilh BE, Snel B, Ettema TJG, & Huynen MA (2008) Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* 25 (8):1659-1667.
270. Wu M & Eisen JA (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 9 (10):R151.
271. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, & Kuske CR (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* 75 (7):2046-2056.
272. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, & Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499 (7459):431-437.
273. Marin J, Battistuzzi FU, Brown AC, & Hedges SB (2017) The timetree of prokaryotes: New insights into their evolution and speciation. *Mol. Biol. Evol.* 34:437-446.
274. Zeng YH, Feng FY, Medova H, Dean J, & Koblizek M (2014) Functional Type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc Natl Acad Sci U S A* 111 (21):7795-7800.
275. Beanland TJ (1990) Evolutionary relationships between Q-Type photosynthetic reaction centers - Hypothesis-testing using parsimony. *J. Theor. Biol.* 145 (4):535-545.
276. Bryant DA, Costas AMG, Maresca JA, Chew AGM, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF, & Ward DM (2007) Candidatus Chloracidobacterium thermophilum: An aerobic phototrophic acidobacterium. *Science* 317 (5837):523-526.
277. Cardona T (2016) Reconstructing the origin of oxygenic photosynthesis: Do assembly and photoactivation recapitulate evolution? *Front Plant Sci* 7:257.
278. Nei M, Gu X, & Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A* 94 (15):7799-7806.
279. Nei M & Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39:121-152.
280. Komenda J, Barker M, Kuvikova S, de Vries R, Mullineaux CW, Tichy M, & Nixon PJ (2006) The FtsH protease slr0228 is important for quality control of photosystem II in the thylakoid membrane of *Synechocystis* sp PCC 6803. *J. Biol. Chem.* 281 (2):1145-1151.
281. Cheregi O, Sicora C, Kos PB, Barker M, Nixon PJ, & Vass I (2007) The role of the FtsH and Deg proteases in the repair of UV-B radiation-damaged Photosystem II in the cyanobacterium *Synechocystis* PCC 6803. *Biochim. Biophys. Acta* 1767 (6):820-828.

282. Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, & Tabata S (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res.* 10 (4):137-145.
283. Mares J, Hrouzek P, Kana R, Ventura S, Strunecky O, & Komarek J (2013) The primitive thylakoid-Less cyanobacterium *Gloeobacter* is a common rock-dwelling organism. *PLoS One* 8 (6):e66323.
284. Dvorak P, Hindak F, Hasler P, Hindakova A, & Poulickova A (2014) Morphological and molecular studies of *Neosynechococcus sphagnicola*, gen. et sp nov (Cyanobacteria, Synechococcales). *Phytotaxa* 170 (1):24-34.
285. Komarek J, Kastovsky J, Mares J, & Johansen JR (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86 (4):295-335.
286. Bombar D, Heller P, Sanchez-Baracaldo P, Carter BJ, & Zehr JP (2014) Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J* 8 (12):2530-2542.
287. Dvorak P, Casamatta DA, Poulickova A, Hasler P, Ondrej V, & Sanges R (2014) *Synechococcus*: 3 billion years of global dominance. *Mol. Ecol.* 23 (22):5538-5551.
288. Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, & Heidelberg JF (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* 1 (8):703-713.
289. Mann NH, Novac N, Mullineaux CW, Newman J, Bailey S, & Robinson C (2000) Involvement of an FtsH homologue in the assembly of functional photosystem I in the cyanobacterium *Synechocystis* sp PCC 6803. *FEBS Lett.* 479 (1-2):72-77.
290. Zehr JP, Bench SR, Carter BJ, Hewson I, Niazi F, Shi T, Tripp HJ, & Affourtit JP (2008) Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322 (5904):1110-1112.
291. Hilton JA, Foster RA, Tripp HJ, Carter BJ, Zehr JP, & Villareal TA (2013) Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nature Communication* 4:1767.
292. Cole JK, Hutchison JR, Renslow RS, Kim YM, Chrisler WB, Engelmann HE, Dohnalkova AC, Hu D, Metz TO, Fredrickson JK, & Lindemann SR (2014) Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front Microbiol* 5:109.
293. Ponce-Toledo RI, Deschamps P, Lopez-Garcia P, Zivanovic Y, Benzerara K, & Moreira D (2017) An early-branching freshwater cyanobacterium at the origin of plastids. *Curr. Biol.* 27 (3):386-391.
294. Pittis AA & Gabaldon T (2016) Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 531:101-104.
295. Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, Tandeau de Marsac N, Gugger M, Lockhart PJ, Allen JF, Brune I, Maus I, Puhler A, & Martin WF (2013) Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol Evol* 5 (1):31-44.
296. Janska H, Kwasniak M, & Szczepanowska J (2013) Protein quality control in organelles - AAA/FtsH story. *Biochim. Biophys. Acta* 1833 (2):381-387.
297. Chen J, Burke JJ, Velten J, & Xin Z (2006) FtsH11 protease plays a critical role in *Arabidopsis* thermotolerance. *Plant J.* 48 (1):73-84.

298. Lu X, Zhang D, Li S, Su Y, Liang Q, Meng H, Shen S, Fan Y, Liu C, & Zhang C (2014) FtsHi4 is essential for embryogenesis due to its influence on chloroplast development in *Arabidopsis*. *PLoS One* 9 (6):e99741.
299. Wagner R, von Sydow L, Aigner H, Netotea S, Brugiere S, Sjogren L, Ferro M, Clarke A, & Funk C (2016) Deletion of FtsH11 protease has impact on chloroplast structure and function in *Arabidopsis thaliana* when grown under continuous light. *Plant Cell and Environment* 39 (11):2530-2544.
300. Nishimura K, Kato Y, & Sakamoto W (2016) Chloroplast proteases: updates on proteolysis within and across suborganellar compartments. *Plant Physiol.* 171 (4):2280-2293.
301. Zelisko A, Garcia-Lorenzo M, Jackowski G, Jansson S, & Funk C (2005) AtFtsH6 is involved in the degradation of the light-harvesting complex II during high-light acclimation and senescence. *Proc Natl Acad Sci U S A* 102 (38):13699-13704.
302. Wagner R, Aigner H, Pruzinska A, Jankapaa HJ, Jansson S, & Funk C (2011) Fitness analyses of *Arabidopsis thaliana* mutants depleted of FtsH metalloproteases and characterization of three FtsH6 deletion mutants exposed to high light stress, senescence and chilling. *New Phytol.* 191 (2):449-458.
303. Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, & Rolland N (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. *Molecular & Cellular Proteomics* 2 (5):325-345.
304. Ferro M, Brugiere S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, & Rolland N (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* 9 (6):1063-1084.
305. Kadirjan-Kalbach DK, Yoder DW, Ruckle ME, Larkin RM, & Osteryoung KW (2012) FtsHi1/ARC1 is an essential gene in *Arabidopsis* that links chloroplast biogenesis and division. *Plant J.* 72 (5):856-867.
306. Vostrukhina M, Popov A, Brunstein E, Lanz MA, Baumgartner R, Bieniossek C, Schacherl M, & Baumann U (2015) The structure of *Aquifex aeolicus* FtsH in the ADP-bound state reveals a C₂-symmetric hexamer. *Acta Crystallogr D Biol Crystallogr* 71 (Pt 6):1307-1318.
307. Suno R, Shimoyama M, Abe A, Shimamura T, Shimodate N, Watanabe YH, Akiyama Y, & Yoshida M (2012) Conformational transition of the lid helix covering the protease active site is essential for the ATP-dependent protease activity of FtsH. *FEBS Lett.* 586 (19):3117-3121.
308. Makino S, Makino T, Abe K, Hashimoto J, Tatsuta T, Kitagawa M, Mori H, Ogura T, Fujii T, Fushinobu S, Wakagi T, & Matsuzawa H (1999) Second transmembrane segment of FtsH plays a role in its proteolytic activity and homo-oligomerization. *FEBS Lett.* 460 (3):554-558.
309. Bern M & Goldberg D (2005) Automatic selection of representative proteins for bacterial phylogeny. *BMC Evol. Biol.* 5:34.
310. Bern M, Goldberg D, & Lyashenko E (2006) Data mining for proteins characteristic of clades. *Nucleic Acids Res.* 34 (16):4342-4353.
311. Battistuzzi FU, Feijao A, & Hedges SB (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* 4:44.
312. Bieniossek C, Niederhauser B, & Baumann UM (2009) The crystal structure of apo-FtsH reveals domain movements necessary for substrate unfolding and translocation. *Proceedings of the National Academy of Sciences of the United States of America* 106 (51):21579-21584.

313. Shao S (2012) Structural and functional characterisation of the FtsH homologues from *Thermosynechococcus elongatus*. MRes (Imperial College London, London).
314. Lyubimov AY, Strycharska M, & Berger JM (2011) The nuts and bolts of ring-translocase structure and mechanism. *Curr. Opin. Struct. Biol.* 21 (2):240-248.
315. Nyquist K & Martin A (2014) Marching to the beat of the ring: polypeptide translocation by AAA plus proteases. *Trends Biochem. Sci.* 39 (2):53-60.
316. Martin A, Baker TA, & Sauer RT (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* 437 (7062):1115-1120.
317. Enemark EJ & Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442 (7100):270-275.
318. Yoshimoto K, Arora K, & Brooks CL, 3rd (2010) Hexameric helicase deconstructed: interplay of conformational changes and substrate coupling. *Biophys. J.* 98 (8):1449-1457.
319. Huang J, Taylor JP, Chen JG, Uhrig JF, Schnell DJ, Nakagawa T, Korth KL, & Jones AM (2006) The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in *Arabidopsis*. *Plant Cell* 18 (5):1226-1238.
320. Wang Q, Sullivan RW, Kight A, Henry RL, Huang JR, Jones AM, & Korth KL (2004) Deletion of the chloroplast-localized *Thylakoid formation1* gene product in *Arabidopsis* leads to deficient thylakoid formation and variegated leaves. *Plant Physiol.* 136 (3):3594-3604.
321. Wu W, Elsheery N, Wei Q, Zhang L, & Huang J (2011) Defective etioplasts observed in variegation mutants may reveal the light-independent regulation of white/yellow sectors of *Arabidopsis* leaves. *J Integr Plant Biol* 53 (11):846-857.
322. Huang W, Chen Q, Zhu Y, Hu F, Zhang L, Ma Z, He Z, & Huang J (2013) *Arabidopsis* thylakoid formation 1 is a critical regulator for dynamics of PSII-LHCII complexes in leaf senescence and excess light. *Mol Plant* 6 (5):1673-1691.
323. Keren N, Ohkawa H, Welsh EA, Liberton M, & Pakrasi HB (2005) Psb29, a conserved 22-kD protein, functions in the biogenesis of photosystem II complexes in *Synechocystis* and *Arabidopsis*. *Plant Cell* 17 (10):2768-2781.
324. Yamatani H, Sato Y, Masuda Y, Kato Y, Morita R, Fukunaga K, Nagamura Y, Nishimura M, Sakamoto W, Tanaka A, & Kusaba M (2013) NYC4, the rice ortholog of *Arabidopsis* THF1, is involved in the degradation of chlorophyll protein complexes during leaf senescence. *Plant J.* 74 (4):652-662.
325. Zhang LG, Wei Q, Wu WJ, Cheng YX, Hu GZ, Hu FH, Sun Y, Zhu Y, Sakamoto W, & Huang J (2009) Activation of the heterotrimeric G protein alpha-subunit GPA1 suppresses the FtsH-mediated inhibition of chloroplast development in *Arabidopsis*. *Plant J.* 58 (6):1041-1053.
326. Wu WJ, Zhu Y, Ma ZX, Sun Y, Quan Q, Li P, Hu PZ, Shi TL, Lo C, Chu IK, & Huang JR (2013) Proteomic evidence for genetic epistasis: ClpR4 mutations switch leaf variegation to virescence in *Arabidopsis*. *Plant J.* 76 (6):943-956.
327. Zhan J, Zhu X, Zhou W, Chen H, He C, & Wang Q (2016) Thf1 interacts with PS I and stabilizes the PS I complex in *Synechococcus* sp. PCC7942. *Mol. Microbiol.* 102 (4):738-751.
328. Gan Y, Li H, Xie Y, Wu W, Li M, Wang X, & Huang J (2014) THF1 mutations lead to increased basal and wound-induced levels of oxylipins that stimulate anthocyanin biosynthesis via COI1 signaling in *Arabidopsis*. *J Integr Plant Biol* 56 (9):916-927.
329. DalCorso G, Pesaresi P, Masiero S, Aseeva E, Schunemann D, Finazzi G, Joliot P, Barbato R, & Leister D (2008) A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in *Arabidopsis*. *Cell* 132 (2):273-285.

330. Yin R, Arongaus AB, Binkert M, & Ulm R (2015) Two distinct domains of the UVR8 photoreceptor interact with COP1 to initiate UV-B signaling in *Arabidopsis*. *Plant Cell* 27 (1):202-213.
331. Song YP, Chen QQ, Ci D, & Zhang DQ (2013) Transcriptome profiling reveals differential transcript abundance in response to chilling stress in *Populus simonii*. *Plant Cell Rep.* 32 (9):1407-1425.
332. Manning VA, Hardison LK, & Ciuffetti LM (2007) Ptr ToxA interacts with a chloroplast-localized protein. *Mol. Plant-Microbe Interact.* 20 (2):168-177.
333. Wangdi T, Uppalapati SR, Nagaraj S, Ryu CM, Bender CL, & Mysore KS (2010) A virus-induced gene silencing screen identifies a role for *Thylakoid Formation1* in *Pseudomonas syringae* pv tomato symptom development in tomato and Arabidopsis. *Plant Physiol.* 152 (1):281-292.
334. Grigston JC, Osuna D, Scheible WR, Liu C, Stitt M, & Jones AM (2008) D-Glucose sensing by a plasma membrane regulator of G signaling protein, AtRGS1. *FEBS Lett.* 582 (25-26):3577-3584.
335. Lindberg P, Park S, & Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metab. Eng.* 12 (1):70-79.
336. Bentley FK, Zurbriggen A, & Melis A (2014) Heterologous expression of the mevalonic acid pathway in cyanobacteria enhances endogenous carbon partitioning to isoprene. *Mol Plant* 7 (1):71-86.
337. Kalbina I & Strid A (2006) The role of NADPH oxidase and MAP kinase phosphatase in UV-B-dependent gene expression in *Arabidopsis*. *Plant Cell Environ* 29 (9):1783-1793.
338. Brosche M, Schuler MA, Kalbina I, Connor L, & Strid A (2002) Gene regulation by low level UV-B radiation: identification by DNA array analysis. *Photochem Photobiol Sci* 1 (9):656-664.
339. Kirilovsky D & Kerfeld CA (2016) Cyanobacterial photoprotection by the orange carotenoid protein. *Nat Plants* 2 (12):16180.
340. Wilson A, Punginelli C, Gall A, Bonetti C, Alexandre M, Routaboul JM, Kerfeld CA, van Grondelle R, Robert B, Kennis JTM, & Kirilovsky D (2008) A photoactive carotenoid protein acting as light intensity sensor. *Proc Natl Acad Sci U S A* 105 (33):12075-12080.
341. Wilson A, Ajlani G, Verbavatz JM, Vass I, Kerfeld CA, & Kirilovsky D (2006) A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria. *Plant Cell* 18 (4):992-1007.
342. Sedoud A, Lopez-Igual R, Ur Rehman A, Wilson A, Perreau F, Boulay C, Vass I, Krieger-Liszka A, & Kirilovsky D (2014) The cyanobacterial photoactive orange carotenoid protein is an excellent singlet oxygen quencher. *Plant Cell* 26 (4):1781-1791.
343. Leverenz RL, Sutter M, Wilson A, Gupta S, Thurotte A, Bourcier de Carbon C, Petzold CJ, Ralston C, Perreau F, Kirilovsky D, & Kerfeld CA (2015) A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. *Science* 348 (6242):1463-1466.
344. Leverenz RL, Jallet D, Li MD, Mathies RA, Kirilovsky D, & Kerfeld CA (2014) Structural and functional modularity of the orange carotenoid protein: distinct roles for the N- and C-terminal domains in cyanobacterial photoprotection. *Plant Cell* 26 (1):426-437.
345. Mayrose I, Graur D, Ben-Tal N, & Pupko T (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.* 21 (9):1781-1791.
346. Hamel LP, Sekine KT, Wallon T, Sugiwaka Y, Kobayashi K, & Moffett P (2016) The chloroplastic protein THF1 interacts with the coiled-coil domain of the disease resistance protein N' and regulates light-dependent cell death. *Plant Physiol.* 171 (1):658-674.

347. Jacob J, Duclohier H, & Cafiso DS (1999) The role of proline and glycine in determining the backbone flexibility of a channel-forming peptide. *Biophys. J.* 76 (3):1367-1376.
348. Serrano L, Neira JL, Sancho J, & Fersht AR (1992) Effect of alanine versus glycine in alpha-helices on protein stability. *Nature* 356 (6368):453-455.
349. Michels J, Appel E, & Gorb SN (2016) Resilin – The Pliant Protein. *Extracellular Composite Matrices in Arthropods*, eds Cohen E & Moussian B (Springer International Publishing, Cham), pp 89-136.
350. Mizusawa N & Wada H (2012) The role of lipids in photosystem II. *Biochim. Biophys. Acta* 1817 (1):194-208.
351. Loll B, Kern J, Saenger W, Zouni A, & Biesiadka J (2005) Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* 438 (7070):1040-1044.
352. Pospisil P & Yamamoto Y (2017) Damage to photosystem II by lipid peroxidation products. *Biochim. Biophys. Acta* 1861 (2):457-466.

Supplementary Table 1

Full list of the mass spectrometry results in Figure 3-6

No.	UniProt Entry	Description	Average Mass (Da)	Sequence Coverage (%)
1	Q55662	ATP-dependent Clp protease regulatory subunit	91232.5955	30.09
	F7UPA9	Photosystem II CP43 protein	50419.1604	14.13
	F7USQ4	Photosystem Q (B) protein	39953.5294	11.94
2	F7UPA9	Photosystem II CP43 protein	50419.1604	54.35
	F7USQ4	Photosystem Q (B) protein	39953.5294	27.78
	P73004	Long-chain-fatty-acid CoA ligase	78011.8775	47.84
3	F7USA2	ATP-dependent zinc metalloprotease FtsH	68554.3144	69.86
	F7UND9	ATP-dependent zinc metalloprotease FtsH	68257.4097	74.36
	F7ULX5	1-deoxy-D-xylulose-5-phosphate synthase	69618.0098	79.22
	F7UPA9	Photosystem II CP43 protein	50419.1604	37.61
	F7UL90	ATP-dependent zinc metalloprotease FtsH	67308.5159	62.66
	F7USQ4	Photosystem Q (B) protein	39953.5294	27.78
	P73913	Acetohydroxy acid synthase	68310.5579	66.83
	L8AG68	Acetolactate synthase 3 catalytic subunit	65353.0667	67.68
	Q55449	Slr0031 protein	67142.7984	45.21
	F7UN99	Photosystem I P700 chlorophyll a apoprotein A1	83181.9461	23.83
	F7UTG6	Sensory histidine kinase NblS/DspA/Dfr/Hik33	74536.4065	67.27
	H0PC53	Sensory histidine kinase NblS/DspA/Dfr/Hik33	74061.8278	66.01
	L8AEH7	Drug sensory protein A	74536.4065	67.27
	H0PHX9	Sensory histidine kinase NblS/DspA/Dfr/Hik33	74565.5153	67.27
	F7UM91	ATP-dependent zinc metalloprotease FtsH	72226.0713	9.92
	L8AG27	ATP-dependent zinc metalloprotease FtsH	69419.9419	10.28
	F7UMJ4	Photosystem Q (B) protein	40056.7303	13.61
	L8APG6	Photosystem Q (B) protein	35730.8316	15.12
	L8AHA5	Uncharacterized protein	70177.816	43.82
	P73606	Slr1855 protein	70177.816	42.34
	F7UNN6	Pyruvate kinase	63275.8141	51.1
	F7USZ7	Photosystem II CP47 protein	55960.8247	43.79
	F7UNA0	Photosystem I P700 chlorophyll a apoprotein A2	81408.1478	11.08
	P74454	Sll0147 protein	76675.8075	23.35
	P72653	Sll1049 protein	64529.7403	21.99
	L8ASF1	GMP synthase [glutamine-hydrolyzing]	61414.9415	30.63
	F7US81	GMP synthase [glutamine-hydrolyzing]	59386.6048	28.11
	L8AS59	Sulfur transferase	11950.6334	17.48
	F7UTJ0	Arginase	33927.2545	10.13
	P73359	Short-chain alcohol dehydrogenase family	71844.0773	20.85
	F7ULG6	Trigger factor	52726.3355	21.23

	P74747	Slr0601 protein	13538.6569	20.31
	F7UTR0	Threonine--tRNA ligase	68988.9662	25.7
	F7USM4	ThiG protein	70676.9862	18.9
	F7UNW6	ICFG protein	71309.8108	19.4
	F7UPM0	Aspartate--tRNA ligase	67441.5773	20.2
	F7UPK6	tRNA modification GTPase MnmE	50214.0496	19.3
4	F7UL90	ATP-dependent zinc metalloprotease FtsH	67308.5159	78.9
	F7UNN6	Pyruvate kinase	63275.8141	82.23
	L8ASF1	GMP synthase [glutamine-hydrolyzing]	61414.9415	49.45
	F7UNA0	Photosystem I P700 chlorophyll a apoprotein A2	81408.1478	15.32
	F7UTK7	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]	69896.8914	32.17
	F7UQR0	CTP synthase	62037.4124	56.52
	F7UT22	Flavoprotein	63854.4092	49.21
	F7UPA9	Photosystem II CP43 protein	50419.1604	42.61
	F7UN99	Photosystem I P700 chlorophyll a apoprotein A1	83181.9461	17.04
	F7UND9	ATP-dependent zinc metalloprotease FtsH	68257.4097	39.65
	P74569	Aspartokinase	63940.2921	44.5
	F7USA2	ATP-dependent zinc metalloprotease FtsH	68554.3144	23.92
	L8AH31	Uncharacterized protein	14928.5345	47.14
	F7USZ7	Photosystem II CP47 protein	55960.8247	40.04
	F7UPM0	Aspartate--tRNA ligase	67441.5773	29.55
	F7ULX5	1-deoxy-D-xylulose-5-phosphate synthase	69618.0098	48.44
	F7USQ4	Photosystem Q (B) protein	39953.5294	20.28
5	P74357	Slr1530 protein	44036.0145	80.57
	F7US55	Fructose-1_6-bisphosphate aldolase_ class II	39146.0846	90.81
	F7USD4	Citrate synthase	45004.6686	77.08
	P73128	Sulfolipid biosynthesis protein SqdB	43492.3095	75.2
	P72668	Slr0723 protein	41719.1159	41.6
	F7UTV7	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	41418.446	48.36
	P73735	NADH dehydrogenase	44662.994	81.19
	F7UM10	Serine hydroxymethyltransferase	46549.872	57.14
	F7UP27	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	44524.1364	56.08
	F7UNN2	30S ribosomal protein S1	36570.124	67.68
	L8AFK7	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	40021.9109	46.67
	F7UPA9	Photosystem II CP43 protein	50419.1604	52.17
	F7USZ7	Photosystem II CP47 protein	55960.8247	35.31
	P72924	Slr1025 protein	43261.5768	60.71
	F7UTZ1	N-acyl-L-amino acid amidohydrolase	42830.8211	36.64
	P73468	Cytoplasmic membrane protein for maltose uptake	45242.0929	61.03
	P74013	Mannosyltransferase B	45949.6953	35.84
	F7URR0	NAD (P)H-quinone oxidoreductase chain 4	57626.9082	25.52

	P74127	ABC transporter	40852.5306	53.53
	P72927	Succinate--CoA ligase	44168.4005	33.42
	P73369	SlI1971 protein	46347.8975	27.48
	Q55870	Slr0626 protein	44543.9339	32.16
	P72678	SlI0703 protein	45277.7398	51.38
	F7UPU8	D-fructose 1_6-bisphosphatase class 2/sedoheptulose 1_7-bisphosphatase	37596.7078	31.3
	F7USQ1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase	42559.3548	30.73
	Q55131	Slr0049 protein	44522.314	27.89
	P73524	SlI1350 protein	44737.1153	20.1
6	F7USQ1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase	42559.3548	67.6
	P74288	Hybrid sensory kinase	41588.666	56.79
	L8ATP9	Hybrid sensory kinase	41588.666	60.33
	P72864	Carboxysome formation protein	37807.8504	84.33
	Q55096	32.4kD protein	32616.8507	78.81
	P72909	Slr1073 protein	43933.1679	47.66
	F7UPX7	Phosphate acyltransferase	37517.1048	64.37
	P74294	PatA subfamily	44021.5236	34.74
	P73947	Slr1507 protein	41445.1614	36.39
	F7UPA9	Photosystem II CP43 protein	50419.1604	42.39
	P72788	Slr1788 protein	41254.2884	51.85
	P72676	NifS protein	42828.7457	59.34
	P73167	Slr1384 protein	42989.2286	32.23
	F7ULX7	Ycf48-like protein	37407.0069	24.56
7	P72586	GDP-D-mannose dehydratase	41449.8857	36
	F7ULJ0	Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (Phosphorylating)	36782.578	26
	P74294	PatA subfamily	44021.5236	29
	L8ASH2	PatA subfamily protein	43520.9712	29
	Q55689	Glucose-1-phosphate thymidyltransferase	42977.3589	26
	F7USQ1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase	42559.3548	30
	P74141	3-chlorobenzoate-3_4-dioxygenase	41234.5012	19
	P74348	SlI1534 protein	43348.1969	23
	P73947	Slr1507 protein	41445.1614	22
	P73566	Slr0882 protein	44557.1585	27
	P73085	Periplasmic binding protein component of an ABC type zinc uptake transporter	37047.6953	27
	P74450	Slr0151 protein	35031.0577	16
8	F7ULB6	Enoyl-[acyl-carrier-protein] reductase [NADH]	27683.8087	19
	L8AH68	Enoyl-[acyl-carrier-protein] reductase [NADH]	30133.6007	22
	M1M6R6	Enoyl-[acyl-carrier-protein] reductase [NADH]	29832.1906	22
	Q55474	SlI0513 protein	31267.5921	17

	L8ANN0	Uncharacterized protein	30621.8545	17
	F7UQC8	Serine acetyltransferase	27510.4257	21
	P73688	Cell-cell signalling protein_ C-factor	28559.8149	17
	P73370	Slr2070 protein	32106.1146	15
	P72907	Slr1071 protein	31049.4862	18
	F7ULQ6	Shikimate dehydrogenase	31157.1765	16
	P73093	Phycobilisome rod-core linker polypeptide CpcG	27450.0596	23
9	F7UPY3	Protein thfI	27132.6044	54.17
	F7UPJ9	Cyanophycinase	29563.7445	44.28
	F7UN82	Transcription termination/antitermination protein nusG	23473.8807	63.9
10	L8AU99	Chloramphenicol O-acetyltransferase	9145.5278	68.83
	P72710	ABC transporter	27919.1919	55.56
	Q55734	SlI0395 protein	23868.0209	81.13
	F7ULI2	NAD (P)H-quinone oxidoreductase subunit I	22672.6647	57.51
	P74574	SlI0625 protein	25864.3642	44
	F7UNT9	Apocytochrome <i>f</i>	35346.623	41.46
11	F7URC7	Cytochrome <i>c</i> -550	18000.1554	65
	F7UMZ5	30S ribosomal protein S5	18241.1169	64.74
	F7UMY4	50S ribosomal protein L13	16990.7215	62.25
	H0P3F0	50S ribosomal protein L13	16990.7215	62.25
	P73952	SlI1418 protein	20805.272	68.62
12	F7UMB8	Phycocyanin b subunit	18300.5385	72.09
	F7UMB7	Phycocyanin a subunit	17702.6329	69.75
	F7UQA6	NAD (P)H-quinone oxidoreductase subunit N	17687.5679	67.7
	F7UPI4	Allophycocyanin a chain	17469.8875	50.31
	P74295	Activator of photopigment and puc expression	17573.8633	28
	P74448	Slr0149 protein	18502.0292	62.18
	P72582	Slr0613 protein	18355.0979	39.31
13	P73048	SlI1638 protein	16535.0187	55.03
	Q55176	Slr0483 protein	16884.8708	29.53
	L8AIP5	30S ribosomal protein S6	13237.2979	34.51
14	P73488	SlI1130 protein	12991.1306	76.52
	F7UKB7	Photosystem II 12 kDa extrinsic protein	14303.2712	60.31
	F7US63	Ribulose biphosphate carboxylase small subunit	13413.0945	48.67
	F7UT96	Photosystem II reaction centre Psb28 protein	12590.3453	57.14
15	F7UR74	Photosystem II 11 kD protein	14975.0658	29.63
	F7UQX2	50S ribosomal protein L27	9448.6747	55.17

	F7URE8	Nitrogen regulatory protein P-II	12397.4035	65.18
	F7UK46	Carbon dioxide concentrating mechanism protein CcmK	11134.7194	50.49
	P74067	Ssl1498 protein	6399.3562	29.51
	F7UKF2	30S ribosomal protein S15	10373.0848	64.04
	F7ULX8	Cytochrome <i>b</i> ₅₅₉ subunit alpha	9448.6644	32.1

Supplement 2. Access to the interactive phylogenetic trees in

Figure 4-2

<http://itol.embl.de/tree/12931244106358221491342225>

Supplement 3. Sequences of the 732 and 716 bp homologous region shown in Figure 3-18A

The “732 bp” segment: (the integrase sequence is shown in light blue):

CCTGTTTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGA
 TCCTTGAGCCCTTGCCCTCCCGCACGATGATCGTGCCGTGATCGAAATCC
 AGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCCGTTCCA
 TACAGAAGCTGGGCGAACAACGATGCTCGCCTTCCAGAAAACCGAGGA
 TGCGAACCCTTCATCCGGGGTTCAGCACACCAGGCAAGCGCCGCGACGGC
 CGAGGTCTTCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTT
 GCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGACC
 GAAACCTTGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCGCT
 GCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAACGGATGAAGGCA
 CGAACCCTGACATAAGCCTGTTTCGGTTCGTAACTGTAATGCAAGTA
 GCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTG
 GTAACGGCGCAGTGGCGGTTTTTCATGGCTTGTTATGACTGTTTTTTGTAC
 AGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCTGA
 TGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACGATGTTA
 CGCAGCAGGGCAGTCGCCCTAAAACAAAGTTA

The “716 bp” segment: (the integrase sequence is shown in light blue; “///” in red shows a position where the assembled contig ends):

CCTGCTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGA
 TCCTTGAG///CCCTTGCCCTCCCGCACGATGATCGTGCCGTGATCGAAAT
 CCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCCGTTCC
 CATAAGAAGCTGGGCGAACAACGATGCTCGCCTTCCAGAAAACCGAG
 GATGCGAACCCTTCATCCGGGGTTCAGCACACCAGGCAAGCGCCGCGACG

GCCGAGGTCTTCCGATCTCCTGAAGCCAGGGCAGATCCGTGCACAGCACCC
TTGCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGA
CCGAAACCTTGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCG
CTGCTGCCCCAAGGTTGCCGGGTGACGCACACCGTGGAACGGATGAAGGC
ACGAACCCAGTGGACATAAGCCTGTTTCGGTTCGTAAAGCTGTAATGCAAGT
AGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGT
GGTAACGGCGCAGTGGCGGTTTTTCATGGCTTGTTATGACTGTTTTTTTGGG
GTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGG
TCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTC
GCCCTAAAACAAAGTTA